Title

Soil enzymes as indicators of soil function: A step toward greater realism in microbial ecological modeling

Journal

Global Change Biology, 28(5)

ISSN

1354-1013

Authors

Wang, Gangsheng
Gao, Qun
Yang, Yunfeng
et al.

Publication Date

2022-03-01

DOI

10.1111/gcb.16036

Primary Research Article

Soil enzymes as indicators of soil function: a step toward greater realism in microbial ecological modeling

Running Title: **Soil enzymes based ecological modeling**

Gangsheng Wang[1,2*], Qun Gao[3], Yunfeng Yang[3], Sarah E Hobbie[4], Peter B Reich[5,6], Jizhong Zhou[2,7,8*]

[1]Institute for Water-Carbon Cycles and Carbon Neutrality, and State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072 China

[2]Institute for Environmental Genomics, and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA

[3]State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China

[4]Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, MN 55108 USA

[5]Department of Forest Resources, University of Minnesota, St Paul, MN 55108 USA

[6]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales 2753, Australia.

[7]School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK 73019, USA

[8]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

*Corresponding authors:**

Gangsheng Wang, ORCID: 0000-0002-8117-5034. *E-mail address*: wang.gangsheng@gmail.com
Jizhong Zhou, ORCID: 0000-0003-2014-0564. *E-mail address*: jzhou@ou.edu

Co-authors:
Qun Gao: ORCID: 0000-0002-2148-5807. gaoqun998@163.com
Yunfeng Yang, ORCID: 0000-0001-8274-6196. yangyf@tsinghua.edu.cn
Sarah E Hobbie, ORCID: 0000-0001-5159-031X. shobbie@umn.edu
Peter B Reich, ORCID: 0000-0003-4424-662X. preich@umn.edu

Abstract

Soil carbon (C) and nitrogen (N) cycles and their complex responses to environmental changes have received increasing attention. However, large uncertainties in model predictions remain, partially due to the lack of explicit representation and parameterization of microbial processes. One great challenge is to effectively integrate rich microbial functional traits into ecosystem modeling for better predictions. Here, using soil enzymes as indicators of soil function, we developed a competitive dynamic enzyme allocation scheme and detailed enzyme-mediated soil inorganic N processes in the Microbial-ENzyme Decomposition (MEND) model. We conducted a rigorous calibration and validation of MEND with diverse soil C-N fluxes, microbial C:N ratios, and functional gene abundances from a 12-year CO_2×N grassland experiment (BioCON) in Minnesota, USA. In addition to accurately simulating soil CO_2 fluxes and multiple N variables, the model correctly predicted microbial C:N ratios and their negative response to enriched N supply. Model validation further showed that, compared to the changes in simulated enzyme concentrations and decomposition rates, the changes in simulated activities of eight C-N associated enzymes were better explained by the measured gene abundances in responses to elevated atmospheric CO_2 concentration. Our results demonstrated that using enzymes as indicators of soil function and validating model predictions with functional gene abundance in ecosystem modeling can provide a basis for testing hypotheses about microbially-mediated biogeochemical processes in response to environmental changes. Further development and applications of the modeling framework presented here will enable microbial ecologists to address ecosystem-level questions beyond empirical observations, toward more predictive understanding, an ultimate goal of microbial ecology.

Keywords: elevated CO_2, functional traits, metagenomics, microbial ecological modeling, microbial functional genes, nitrogen enrichment, predictive ecology, soil enzymes

1 INTRODUCTION

Projecting future carbon cycling and climate change scenarios is a grand challenge in ecology, and for society (Cavicchioli *et al.*, 2019). Microorganisms, acting as detritivores, plant symbionts, or pathogens, are critical in mediating ecosystem carbon (C) and nutrient cycling and consequently climate change (Bardgett *et al.*, 2008, Cavicchioli *et al.*, 2019). However, traditional biogeochemical and Earth system models (ESMs) do not explicitly consider the roles of microbial communities by assuming that microbes are in equilibrium with their environment (Schimel, 2013). Such classical models appear to work well for large scale patterns of bulk soil organic matter pools, but they may have reached their limits, particularly when depicting transient dynamics in the face of environmental changes (Wieder *et al.*, 2015). In the last decade, a considerable amount of effort has been devoted to explicitly integrating microbial communities and functions into microbial ecological models (e.g., Allison *et al.*, 2010, Davidson *et al.*, 2012, Manzoni *et al.*, 2016, Schimel & Weintraub, 2003, Sulman *et al.*, 2018, Tang & Riley, 2019, Wang *et al.*, 2013, Wieder *et al.*, 2015). Studies have shown that microbial-explicit models could more accurately represent the impacts of global change drivers, such as warming and priming effects (Wieder *et al.*, 2015). This calls for more mechanistic microbial ecological models to advance our understanding of soil microbial and biogeochemical responses to environmental changes.

Ecosystem models with carbon-nitrogen (C-N) coupled processes have elucidated substantial impacts on the carbon-climate feedbacks that are lacking from the C-only models, for instance, smaller sensitivity of land C uptake to temperature variation or increasing atmospheric CO_2 concentration (Thornton *et al.*, 2007). N availability is known to strongly influence microbial growth and C cycling (Cavicchioli *et al.*, 2019, Treseder, 2008), hence, multiple microbial-explicit models have accounted for C-N interactions (e.g., Abramoff *et al.*, 2017, Drake *et al.*, 2013, Gao *et al.*, 2020, Kyker-Snowman *et al.*, 2020, Schimel & Weintraub, 2003). However, limited attention has been paid to the explicit representation and parameterization of multiple differential microbial groups, particularly related to the inorganic N cycle (e.g., N mineralization and immobilization, nitrification, and denitrification) (Sulman *et al.*, 2018, Wang *et al.*, 2019). This impedes a comprehensive validation of complex C-N processes and their interactions as have been done for classical terrestrial C-N coupled models. Therefore, the introduction of mechanistic inorganic N cycling into microbial ecological models may provide new opportunities to pose and validate further hypotheses about coupled C-N cycling in response to environmental perturbations, especially elevated atmospheric CO_2 concentration (eCO_2) and enhanced N deposition (Abramoff *et al.*, 2017, Wieder *et al.*, 2015).

The absence of microbial communities in ecosystem models is primarily due to the extremely high diversity and complexity of microbial communities and the lack of appropriate strategies and frameworks for using microbial information in ecological modeling (Bailey *et al.*, 2018, Bardgett *et al.*, 2008, Gao *et al.*, 2020, Todd-Brown *et al.*, 2012, Wieder *et al.*, 2015). Because microbial communities under natural settings are extremely diverse and complex, functional traits-based approaches are very attractive and promising for explicitly accounting for the role of microbes in regulating biogeochemical cycles in ecosystem models (Falkowski *et al.*, 2008, Klausmeier *et al.*, 2020). However, one big challenge is how to extract and scale functional information to inform ecosystem modeling (Torsvik & Øvreås, 2002). This challenge has also become an important motivation to develop microbially-explicit models (Bailey *et al.*, 2018). Despite increasing interest in incorporating microbial functional traits into ecosystem models, it remains a major challenge to directly link genomes to global processes (Bailey *et al.*, 2018). However, it is viable to link genomes and processes at intermediate scales with integrated applications of powerful analytical and modeling techniques (Song *et al.*, 2017, Trivedi *et al.*, 2013). While representing a massive number of microbial taxa in models is impractical and unnecessary, owing to functional redundancy (Bailey *et al.*, 2018), grouping microbes and enzymes into simplified functional guilds is feasible and enables the parameterization of microbial ecological models (Chen & Sinsabaugh, 2021, Song *et al.*, 2017).

However, it remains challenging to develop microbially-explicit N transformation processes. First, the multi-step inorganic N reactions are regulated by intracellular enzymes that are located at cell membrane, cytoplasm, or periplasm (Fiencke & Bock, 2006, Schlesier *et al.*, 2016, Song *et al.*, 2017). These intracellular N enzymes differ from extracellular enzymes (e.g., ligninases and cellulases) and have little capability of acting on their own, leading to the concern in representing them in microbial ecological models. Second, an effective microbial or enzyme allocation scheme is warranted to handle diverse microbial communities associated with the multiple inorganic N processes. For instance, we have recently used GeoChip-based gene abundances (Shi *et al.*, 2019) to constrain the Microbial-ENzyme Decomposition (MEND) model, where we only considered extracellular C-degrading enzymes owing to the lack of detailed representation of microbially-mediated inorganic N reactions (Gao *et al.*, 2020, Guo *et al.*, 2020). In short, modeling efforts have not kept pace with the rapid advances in the microbial ecology of N relevant microorganisms and genes (Hu *et al.*, 2015).

Model parameterization through calibration and validation with field observations is arduous due to the limited available long-term experimental data and large uncertainties in measurements of the state variables, fluxes, and microbial community structure and functions, as well as uncertainties in model structure and simulations (Bradford *et al.*, 2016, Sulman *et al.*, 2018). Consistent with these large uncertainties in observations and model simulations, recent comparison of five soil C models with different representation of microbial and mineral processes revealed that existing traditional measurements (e.g., CO_2 fluxes and soil C contents) were insufficient to constrain or validate ecosystem models (Sulman *et al.*, 2018). To demonstrate the capability of microbially-explicit models, development of benchmarking with multiple datasets with a variety of microbial and omics data, especially for inorganic N cycling, is needed.

In this study, building on past work (Gao *et al.*, 2020), we attempted to improve the MEND model by developing a new microbially-mediated inorganic N module that uses relevant enzymes as indicators of soil function, with the proposition of a competitive dynamic enzyme allocation scheme. The new inorganic N module accounts for the important roles of intracellular enzymes in regulating several critical inorganic N transformations, including N fixation, nitrification, and the sequential denitrification reactions from nitrate (NO_3^-) to dinitrogen (N_2) (Xue *et al.*, 2016, Zhou *et al.*, 2012). In addition to several important observations (soil respiration, soil concentrations of ammonium and nitrate, and abundances of two functional gene groups targeting SOM decomposition) used in Gao et al. (Gao *et al.*, 2020), the new MEND model was further calibrated and validated with a variety of new data from that 12-year field experiment, called BioCON (Gao *et al.*, 2020), including SOC content, multiple inorganic N transformations, and the abundances of six functional gene groups important to inorganic N processes. We directly compared model outputs to the relative changes of the measured gene abundances in response to eCO_2. Our results indicated that explicitly using enzymes as soil function indicators in ecosystem models and validating model predictions with gene abundance data can provide a basis for better understanding and testing hypotheses about microbially-mediated biogeochemical processes under environmental changes.

2 MATERIALS AND METHODS

2.1 Overview of the MEND modeling framework

We developed an integrated microbial ecological modeling framework, consisting of several key components such as model development, sensitivity analysis, model calibration, validation, and uncertainty quantification (Fig. 1a). The new MEND model explicitly represents distinct microbial and enzyme groups responsible for C-N transformation processes. The Multi-Objective Parameter Sensitivity Analysis (MOPSA) was used to determine the relative importance of parameters in terms of multiple response objectives (i.e., variables). The sensitivity analysis forms the cornerstone of the Multi-Objective Parameter Stochastic Optimization (MOPSO) and validation procedure. The MOPSO approach aims to determine the values of those “free” parameters by calibrating the model against observations (Fig. 1b), where a stochastic optimization algorithm, the Shuffled Complex Evolution (SCE) (Duan *et al.*, 1992), is modularized and incorporated into the MEND model for automatically calibrating parameters. The SCE algorithm combines the strengths of several optimization strategies such as controlled random search, complex shuffling, and competitive evolution, which ensure that the parameter space is efficiently and thoroughly exploited (Duan *et al.*, 1992, Wang *et al.*, 2015). The MOPSO enables to fit multiple observational variables (soil respiration, C pools, microbial biomass, etc.) by minimizing the overall objective function as the weighted average of multiple objectives pertaining to these variables (Fig. 1b). We further validated the model and evaluated model performance against datasets not used for model calibration. The Uncertainty Quantification by Critical Objective Function Index (UQ-COFI) approach was developed to filter the parameter sets generated by the MOPSO procedure. These filtered parameter sets by UQ-COFI represented the posterior parameter space, which were used to drive multiple model runs to quantify uncertainties in response variables due to parametric uncertainty. We employed this integrated modeling framework to guide reliable model development and application.

A detailed description of model sensitivity analysis (MOPSA) and uncertainty quantification (UQ-COFI) are presented in Supporting Information Sections 3.3 and 3.6, respectively. In the following, we overview the new MEND model and its calibration and validation against multiple datasets.

2.2 New MEND model with a competitive dynamic enzyme allocation scheme

We incorporated a new N-associated module into the old MEND (MEND-old) model (Gao *et al.*, 2020, Wang *et al.*, 2021) (Supporting Information Fig. S1b; for comparison, the new MEND model (MEND-new) is shown in Fig. S1a, a copy of Fig. 2) by explicitly representing: (i) multiple key intracellular enzymes as indicators that catalyze nitrification, sequential denitrification, and nitrogen fixation processes; (ii) plant-microbial competition for inorganic N (NH_4^+ and NO_3^-); (iii) ammonium (NH_4^+) sorption; (iv) nitrate (NO_3^-) and nitrite (NO_2^-) leaching; and (v) N gases (NO, N_2O, and N_2) exchange between soil and the atmosphere. A reaction rate in the model may be modified by soil pH, soil temperature and moisture conditions (Supporting Information Figs. S2–S4). Details on MEND-new model and its state variables, governing equations, component fluxes and parameters are described in Supporting Information Sections 1–3 and Table S1–S6.

We used flexible stoichiometry (i.e., time-variant C:N ratio) for SOM and microbial biomass pools to represent microbial adaptation in response to the stoichiometric imbalance of available resources (Du *et al.*, 2018, Fanin *et al.*, 2017, Mooshammer *et al.*, 2014a, Mooshammer *et al.*, 2014b, Zechmeister-Boltenstern *et al.*, 2015). In addition to the three SOM-degrading enzyme functional groups, i.e., EP_O, EP_H and EM, we incorporated six new enzyme systems as indicators controlling inorganic N transformations (Fig. 2), i.e., nitrogenases (corresponding to functional genes of nifH), ammonia oxidases (amoA), nitrate reductases (narG/napA), nitrite reductases (nirS/nirK), nitric oxide reductases (norB), and nitrous oxide reductases (nosZ) (Xue *et al.*, 2016, Zhou *et al.*, 2012).

We proposed a competitive dynamic enzyme allocation scheme to deal with the synthesis of multiple enzyme groups (see Supporting Information Section 1.1.7). The enzyme allocation approach developed here is based on the synthetic results that enzyme activities are dependent on microbial biomass (Jian *et al.*, 2016) and substrate availability (Sinsabaugh *et al.*, 2014). A competitive allocation scheme is applied to the production of enzymes for each inorganic N transformation process, where the competitive allocation coefficient is the saturation level of an inorganic N substrate (i.e., the ratio of the substrate concentration to the corresponding half-saturation constant) relative to the sum of the saturation levels of all inorganic N substrates (Supporting Information Eq. 40).

2.3 Model calibration and validation

We implemented the MOPSO approach, based on the SCE algorithm (Duan *et al.*, 1992, Wang *et al.*, 2015), to calibrate selected model parameters according to the sensitivity analysis (Fig. 1b). We aimed to determine parameter values and their uncertainties by achieving high goodness-of-fits of model simulations against experimental observations, such as soil respiration (R_s), microbial heterotrophic respiration (R_h), microbial biomass carbon (MBC), and soil C and N pools and fluxes. Each objective evaluates the goodness-of-fit of a specific observed variable. The parameter optimization attempts to minimize the overall objective function (J) that is computed as the weighted average of multiple single-objectives (see Eqs. 67 in Supporting Information Section

3.4). Generally, equal weights are assigned to these objectives. However, a higher weight is recommended for a variable that is frequently measured or is vital to the research topic.

Different objective functions were used to quantify the goodness-of-fit for different variables (Supporting Information Section 3.4), depending on the measurement method and frequency. As per model validation (Refsgaard, 1997), we used datasets that were not involved in model calibration to evaluate model performance, where the same calibrated parameter values were used in model validation.

2.4 BioCON datasets for model calibration and validation

Since there is no gold-standard for validating model performance, it is a common practice to use published datasets in ecosystem and bioinformatic studies, which have advantages for model calibration and validation (Luo *et al.*, 2012, Ning *et al.*, 2020). Thus, we used experimental data (Table 1) from a well-designed, long-term multifactor free-air CO_2 enrichment experiment, BioCON (Biodiversity, CO_2, and N deposition) (45.4010° N, 93.2010° W) in Minnesota, USA (Reich & Hobbie, 2013). The BioCON experiment aims, among other goals, to elucidate how microbe-mediated feedbacks to soil respiration are affected by N addition (+4 g N m^{-2} yr^{-1}) and elevated atmospheric CO_2 (eCO_2, +180 ppm) (Adair *et al.*, 2009, Adair *et al.*, 2011). The BioCON soil is an Entisol, more specifically, a mixed, frigid Lamellic Udipsamments as per the USDA soil taxonomy (O'Geen *et al.*, 2017, Soil Survey Staff, 1999). This excessively drained soil, derived from glacial outwash with a coarse structure, has very poor development and a sandy texture (92–94% sand and 2–3% clay in the top 114 cm) (Kazanski *et al.*, 2021, O'Geen *et al.*, 2017). In summary, there were four CO_2×N treatments among 296 plots: ambient atmospheric CO_2 & ambient N supply (aCO_2-aN), eCO_2-aN, aCO_2 & enriched N supply (aCO_2-eN), and eCO_2-eN with each treatment having 74 plots (biological replicates). To examine the effects of plant diversity on ecosystem N cycling, the BioCON experiment also has (at each level of CO_2 and N) treatment plots planted with either 1, 4, 9, or 16 grassland species (Dijkstra *et al.*, 2007).

Estimates of daily GPP (gross primary production) values were obtained from a corrected 8-day GPP product based on the MODIS GPP (MOD17A2/MOD17A2H) (Gao *et al.*, 2020, Zhu *et al.*, 2018) and used to drive model simulations under the control treatment (aCO_2-aN). The GPP for the other three treatments was rescaled according to the general linear relationship between NPP (net primary production) and GPP (Gao *et al.*, 2020). Meanwhile, environmental datasets measured in each CO_2×N treatment were also used for model simulations for each treatment, including monthly soil pH, daily soil temperature and moisture.

Soil samples for microbial community analysis were collected from the 296 plots in August 2009. Each sample was a composite of five soil cores from each plot at a depth of 0–15 cm. Microbial DNA was extracted, hybridized with GeoChip arrays, and analyzed as described previously (Guo *et al.*, 2020, Tu *et al.*, 2014). The eCO_2 effect on the abundance of each functional gene (total abundance of all probes of this gene) was examined by the response ratio (Luo *et al.*, 2006):

$$RR = \ln\left({}^{x_T}/{x_C}\right) \qquad (1)$$

where *RR* is the response ratio (effect size) that quantifies the log-proportional change between the gene abundances of eCO_2 (x_T) and aCO_2 (x_C) samples.

The observed response ratios (*RR*s) between the gene abundances (GA_{obs}) of eCO_2 and aCO_2 were used as additional data to evaluate model-simulated enzyme concentrations (EC_{sim}), enzyme activities (EA_{sim}), or equivalent first-order reaction rates (FR_{sim}). As the Michaelis-Menten kinetics is used in the MEND model, the relationships among EC_{sim}, EA_{sim}, and FR_{sim} are described as follows:

$$EA_{sim} = Vd \cdot EC_{sim} \quad (2)$$

$$FR_{sim} = (Vd \cdot EC_{sim})/(K + S) = EA_{sim}/(K + S) \quad (3)$$

where EC_{sim} (mg C cm^{-3}), EA_{sim} (mg C cm^{-3} h^{-1}), and FR_{sim} (h^{-1}) are simulated enzyme concentration, enzyme activity, and the equivalent first-order reaction rate, respectively; *S* denotes the substrate (e.g., SOC) concentration; and the parameters *Vd* and *K* represent the specific enzyme activity (mg C mg^{-1} C h^{-1}) and the half-saturation constant (mg C cm^{-3}), respectively.

In summary, nine C-N response variables were involved in the calibration of MEND-new (Table 1): soil CO_2 flux (R_s), microbial biomass C (MBC), soil organic C (SOC), ammonium (NH_4^+), nitrate + nitrite ($NO_3^- + NO_2^-$), as well as the reference rates of net N mineralization ($FN_{mn\text{-}im}$), nitrification (FN_{nit}), biological N fixation (FN_{fix}), and plant N uptake (FN_{im_VG}). Among these variables, the literature-reported biological N fixation rates (including both symbiotic and non-symbiotic N fixation) (Cleveland *et al.*, 2013, 1999) and plant N uptake rates (Bessler *et al.*, 2012, Harty *et al.*, 2017, Reyes *et al.*, 2015) in grasslands were used as reference for model calibration. To examine the predictive power of the model, we only calibrated the model against the data under the control treatment (aCO_2-aN) and then applied the calibrated parameters to the other three treatments for model validation. To further investigate the model's capability in representing microbial and enzyme functional traits, we directly validated the model against literature-reported microbial C:N ratios (Xu *et al.*, 2013) and the measured response ratios of gene abundances (GA_{obs}).

3 RESULTS

Detailed results of model sensitivity analysis and uncertainty quantification are presented in Supporting Information Results 5.1 (with Fig. S5) and 5.2 (with Fig. S6), respectively. In the following, we focus on the key results with respect to model calibration, validation, and ecological insights.

3.1 Model calibration and validation of soil respiration and inorganic N processes

3.1.1 Model calibration and validation strategy in terms of the BioCON data

Based on the aforementioned sensitivity analysis and previous studies on the MEND model (Wang *et al.*, 2019, Wang *et al.*, 2015, Wang *et al.*, 2013), we selected 14 important parameters (Supporting Information Fig. S6) to conduct model calibration.

In the first step of calibration, we calibrated nine microbial physiological parameters by achieving high goodness-of-fits of model simulations against experimental observations, such as soil respiration (R_s), microbial biomass carbon (MBC), and soil organic carbon (SOC) (Table 1). We only compared the simulated mean values of MBC and SOC to the observed reference MBC and SOC, respectively, as we only had observations at one time point for each of them. In the overall objective function (Eqs. 67 in Supporting Information Section 3.4), the weights of 0.50, 0.25, and 0.25 were assigned to the objectives pertaining to R_s, MBC, and SOC, respectively, owing to far more data points available for R_s (284 data points) than for MBC and SOC. The 9 parameters (Supporting Information Table S5) included: (i) six parameters (V_g, α, K_D, Y_g, kY_g, γ) for microbial growth, maintenance, and mortality; and (ii) three parameters (p_{EP}, fp_{EM}, r_E) for enzyme production, turnover and decomposition of SOM.

As for the second step of calibration, we fixed the parameter values determined by the first step and calibrated five important inorganic N parameters (Supporting Information Table S5: VN_{fix}, VN_{nit}, VN_{denit}, VN_{plant}, and $Qmax_{NH4}$) by fitting observed concentrations of ammonium (NH_4^+) and nitrate + nitrite ($NO_3^- + NO_2^-$), as well as the reference rates of net N mineralization ($FN_{mn\text{-}im}$), nitrification (FN_{nit}), biological N fixation (FN_{fix}), and plant N uptake (FN_{im_VG}) (Table 1). In the overall objective function, higher weights were used for the objectives of NH_4^+ and $NO_3^- + NO_2^-$ than the remaining N variables. As a result, there were nine individual objective functions regarding the nine C-N response variables in the model calibration: the first three objective functions were used for the calibration of microbial physiological parameters and the remaining six variables were used for the parametrization of inorganic N transformation parameters (Table 1).

The model simulation period covered the 12-year observational period (1998–2009). Model simulations for each treatment were driven by the corresponding data: GPP, soil temperature and moisture, and inorganic N (NH_4^+ and NO_3^-) input. We used the MOPSO approach to calibrate model parameters with the data from the aCO_2-aN treatment. We then validated the model using the same set of model parameters calibrated for aCO_2-aN to simulate R_h and R_s, and soil inorganic N in the other three treatments (eCO_2-aN, aCO_2-eN and eCO_2-eN).

3.1.2 Model calibration and validation results of soil respiration

Our model calibration with aCO_2-aN data achieved good agreement between simulated and observed soil respiration (Fig. 3a, $R^2 = 0.60$), so did the model validation of soil respiration in the other three treatments (Fig. 3b, $R^2 = 0.56$–0.61). In addition, the percent bias ($|PBIAS|$) values of mean soil respiration were 3% for calibration and 3–11% for validation, suggesting that simulated mean soil respiration values were close to the observed ones in all four treatments. The simulated mean values of MBC and SOC were within the tolerances for MBC (10%) and SOC (5%), respectively, as expected in model simulations (Table 1).

3.1.3 Model calibration and validation results of soil ammonium and nitrate

In addition, the simulated mean soil NH_4^+ and (NO_3^-+NO_2^-) concentrations also agreed well with the observations in both model calibration and validation (Fig. 3c and 3d). Although model validation showed larger percent bias between simulated and observed values (|*PBIAS*| = 24–29% for NH_4^+ and 5–39% for NO_3^-+NO_2^-) than model calibration (2% for both NH_4^+ and 8% for NO_3^-+NO_2^-), the model validation of inorganic N concentrations could still be judged as satisfactory according to the 70% bias criterion for N modeling (Moriasi *et al.*, 2007). Simulated variation (i.e., average standard deviation (SD) = 0.20 gN m^{-2}) in soil NH_4^+ concentrations by the MEND-new model was also comparable to observed variation (average SD = 0.15 gN m^{-2}), which was also true for soil NO_3^-+NO_2^- (average SD = 0.074 and 0.070 gN m^{-2} for observed and simulated concentrations, respectively). For comparison, the simulated average SD values by the MEND-old model were 0.072 and 0.095 gN m^{-2} for soil NH_4^+ and NO_3^-+NO_2^-, respectively.

Generally, the simulated mean NH_4^+ and NO_3^- concentrations by MEND-new from this study showed much lower biases than MEND-old with simplified N processes as described in Gao et al. (2020), except for the NO_3^- validation under eCO_2-aN. The average |*PBIAS*| for NH_4^+ was reduced from 45% (MEND-old, 12–68% in range) to 21% (MEND-new, 2–28%), though the average |*PBIAS*| values for NO_3^- were similar between MEND-old (11–32% with an average of 18%) and MEND-new (5–39% with an average of 18%) (Fig. 3c and 3d). In order to account for the effects of the number of free (i.e., calibrated) model parameters on the model performance, we calculated the Akaike information criterion (AIC) of the two models (Goll *et al.*, 2012). The number of free model parameters for inorganic N processes was five for MEND-old (Gao *et al.*, 2020) and six for MEND-new (see Table S5), as most of the N-related parameters in MEND-new were determined as per literature. Compared to MEND-old, MEND-new had a slightly higher AIC under aCO_2-aN (Fig. 3c), but lower AIC under the other three treatments (Fig. 3d).

3.1.4 Model calibration and validation results of inorganic N fluxes

Biological N fixation and plant N uptake rates during model calibration and validation were generally in accordance with literature-reported data (Fig. 4). The simulated biological N fixation rates in all four treatments were comparable to the ranges for grasslands reported in the literature (Cleveland *et al.*, 2013, Cleveland *et al.*, 1999). The N fixation rates were significantly higher under the two eN treatments compared to those under the aCO_2-aN treatment (Fig. 4a). However, we did not observe statistically significant eCO_2 effects on the N fixation rates. Our simulated plant N uptake rates were generally between 15 and 30 g N m^{-2} yr^{-1}, which were within the range (10–40 g N m^{-2} yr^{-1}) observed in grasslands (Bessler *et al.*, 2012, Harty *et al.*, 2017, Reyes *et al.*, 2015). The plant N uptake rates were significantly lower under aCO_2-aN than those under the other three treatments, with the highest under eCO_2-eN and no significant difference between eCO_2-aN and aCO_2-eN or eCO_2-eN (Fig. 4b).

The simulated net N mineralization and nitrification rates were within the observed ranges in both model calibration and validation (Supporting Information Fig. S7). As mentioned in the methods, we did not expect simulated values to match the measured nitrification rates and net N mineralization rates as they represented reference rates or rough estimates. Our simulated net N mineralization rates were 57–85% (with a mean of 68%) of the reference rates, with the lowest simulated actual N mineralization rate under aCO_2-aN and the highest under the two eCO_2 treatments (Supporting Information Fig. S7a). The simulated nitrification rates accounted for 39–

54% (with a mean of 47%) of the reference values, with the lowest under the two ambient N treatments and the highest under the two enriched N treatments (Supporting Information Fig. S7b).

3.2 Model validation of microbial C:N ratios

Independent model validation showed that the microbial C:N ratios simulated by MEND-new conformed to the literature-reported mean value and the 95% confidence interval for grassland soils (Xu *et al.*, 2013), whereas MEND-old predicted much higher microbial C:N ratios (Fig. 5a). Though both models predicted lower microbial C:N under eN than aN (Fig. 5b and 5c), only the MEND-new model revealed a statistically significant decrease in the microbial C:N as a result of N addition (Fig. 5c). However, neither model demonstrated significant eCO_2 effect on the microbial C:N ratios (Supporting Information Fig. S8).

3.3 Model validation with functional gene abundances

We first compared the eCO_2 effects on enzymes simulated by the two models, i.e., MEND-old and MEND-new. To make the results comparable between the two models, gene abundances were not included in the calibration of MEND-old, matching what we did for MEND-new in this study. We only examined the oxidative enzymes (Fig. 5d) and hydrolytic enzymes (Fig. 5e) that are associated with the C cycle because only these two groups are included in both models. The response ratios (*RRs*) of simulated enzyme concentrations (EC_{sim}), enzyme activities (EA_{sim}), and the first-order reaction rates (FR_{sim}) by MEND-old were significantly higher than the response ratios of observed gene abundances (GA_{obs}). The simulated response ratios by MEND-new were generally lower than those by MEND-old, except for the FR_{sim} of hydrolytic enzymes under eN and the FR_{sim} of oxidative enzymes. In short, compared to MEND-old, the simulated response ratios by MEND-new were generally closer to the measured values. Particularly, only the EA_{sim} by MEND-new correctly reflected the negative response in the oxidative enzymes observed under eN (Fig. 5d).

We further evaluated the similarity or dissimilarity between MEND-new simulated and observed response ratios of all eight enzymes associated with the C and N cycling by the Wilcoxon signed rank test (Conover, 1998). The simulated response ratios consist of EC_{sim} (Fig. 6a), EA_{sim} (Fig. 6b), or FR_{sim} (Fig. 6c) for eight enzymes, whereas the observed response ratios include GA_{obs} for eight corresponding genes (Table 1 and Fig. 6).

The simulated results of response ratios indicate that the eCO_2 effects on the enzymes were more pronounced under aN than under eN, consistent with the responses in GA_{obs}, i.e., 50% CI = 0.03~0.06 under aN *versus* –0.02~ –0.01 under eN (Fig. 6). We also found that, among the three simulated variables (EC_{sim}, EA_{sim}, and FR_{sim}), only EA_{sim} responses were not significantly different from the responses of GA_{obs} under aN or eN according to the Wilcoxon signed rank test (Fig. 6b). Our results showed positive responses of EA_{sim} under aN for six out of eight enzymes, which concurred with the changes in GA_{obs}. However, the other two enzyme groups (NO and N_2O reductases) exhibited slightly negative response ratios (–0.019 and –0.003) when comparing eCO_2-aN to aCO_2-aN, which were not consistent with GA_{obs} responses (0.03 and 0.05). In addition, negative response ratios of EA_{sim} under eN were found for all enzymes except two groups (hydrolytic enzymes and NO_2^- reductases), which generally concurred with the changes in GA_{obs} under eN.

4 DISCUSSION

4.1 Ecosystem modeling with explicit enzymes as bioindicators

The MEND-new model developed here offers new capabilities to investigate microbial-enzyme mediated N fixation, nitrification, and denitrification, and plant-microbe competition for inorganic N, as well as inorganic N leaching and gaseous emissions processes, which adds additional features to the original MEND-old model (Gao *et al.*, 2020, Wang *et al.*, 2020). The oxidative and hydrolytic enzymes for depolymerizing SOM are actual molecules which are independently functional. However, the intracellular N enzymes (responsible for biological N fixation, nitrification, and denitrification) are not physical molecules and thus have little ability to function independently of a living cell (Fiencke & Bock, 2006, Schlesier *et al.*, 2016, Song *et al.*, 2017). Toward this end, we treat these inorganic N enzymes as simple bioindicators of likely activity of cellular level microbial physiology. Explicit representing these intracellular N enzymes in the model is more “pseudo-mechanistic” rather than “truly-mechanistic” (Hommel, 2020), but it provides a tractable way to capture complex biological dynamics of inorganic N cycling. Although enzyme-enabled representation of more detailed C-N transformation processes increases model complexity, it potentially offers important insights into microbial control over biogeochemical and the interactions between multiple physical, chemical, and biological processes.

In contrast with enzyme-based models like MEND, the gene-centric approach was developed for examining ocean N cycling, where the gene abundances can be directly modeled to mediate chemical reactions (Reed *et al.*, 2014). The gene-centric approach offers the advantage of direct comparison between modeled and measured gene abundances. However, currently there is no enough information available for identifying appropriate biomarker genes for a specific metabolic pathway (Reed *et al.*, 2014). In addition, for modeling a complex system with many processes, compared with the models characterized by enzyme groups, the number of genes may increase dramatically, resulting in difficulties and uncertainties in estimating a vast number of parameters for these genes. In terms of ecosystem-level modeling that relies on bulk concentrations, it is currently more feasible to adopt the strategy with aggregated enzyme groups than the gene-centric approach.

We also proposed a competitive dynamic enzyme allocation scheme to assist the incorporation of multiple enzyme systems. Here, 'dynamic' means the allocation of each enzyme group varies with time, and 'competitive' implies that multiple enzyme systems compete with each other as per the relative saturation levels of the corresponding substrates. Enzyme allocation problems have been studied theoretically (Müller *et al.*, 2014) or empirically (Sinsabaugh & Moorhead, 1994, Sinsabaugh *et al.*, 2002) based primarily on stoichiometric information (Allison *et al.*, 2011). These previous studies were generally focused on limited groups of enzymes (Averill, 2014, Müller *et al.*, 2014), in contrast to the eight enzyme systems regulated by our competitive dynamic enzyme allocation scheme. We realize that this allocation approach could not be directly evaluated as it is currently challenging to measure in situ production rates, particularly, of multiple enzyme systems. However, our model calibration and validation with a variety of inorganic N concentrations and fluxes indirectly demonstrated the applicability of this competitive dynamic enzyme allocation scheme, which was further supported by the model evaluation with measured gene abundance data.

4.2 Rigorous calibration and validation of microbial ecological models

Rigorous calibration and validation of microbial ecological models against observations is essential for assessing and refining models. However, finding appropriate datasets to validate microbial and enzymatic reactions in ecosystem models exhibits significant challenges. Treating inorganic N enzymes as indicators of soil function also allows the use of corresponding gene abundance data in ecosystem modeling, yet the relationship between enzymes and their coding genes is complicated (Bailey *et al.*, 2018). Here, we used gene abundance data for model validation instead of calibration, because we attempted to explore the possible relationships between simulated ecosystem functioning (i.e., enzyme concentrations, enzyme activities, or reaction rates) and gene abundance. We showed that the changes in enzyme activities, rather than enzyme concentrations and the first-order reaction rates, are better explained by the responses in gene abundances. This may be due to the inclusion of more (eight in this study vs. two in Gao et al. (2020)) enzyme systems and relevant gene abundance data, which could introduce larger variation in the data resulting in differential modeling performance in terms of multiple variables. Therefore, we need more paired measurements of gene abundances and process rates under long-term field conditions in various ecosystems. DNA-based functional gene abundances have been thought to integrate longer-term (hours to days or longer) microbial potential in the physicochemical environment (Petersen *et al.*, 2012, Rocca *et al.*, 2015). Thus, we infer that DNA-based functional gene abundance is likely a better predictive variable for enzyme activity than for enzyme concentration or reaction rate. In the MEND model, enzyme activity contains the information of both active enzymes and their specific activity. To this end, enzyme activity represents the potential enzyme-catalyzed biogeochemical rates not limited by substrate availability (Ouyang *et al.*, 2018, Petersen *et al.*, 2012), whereas substrate availability is considered in the actual reaction rate (i.e., FR_{sim} calculated by Eq. 3). This interpretation supports our results on stronger relationship between GA_{obs} and EA_{sim} than between GA_{obs} and the other two variables (EC_{sim} and FR_{sim}).

Very few studies have adopted gene abundance data in ecosystem or environmental modeling, where the model-data integration practices were often implemented for a short time period (e.g., 20 days) based on laboratory data (Gao *et al.*, 2020, Li *et al.*, 2017, Pagel *et al.*, 2016, Song *et al.*, 2017). Compared to these short-term laboratory-based modeling studies, it is likely more challenging to conduct gene-informed long-term (e.g., years to decades or longer) ecosystem modeling in the field, owing to complex spatiotemporal environmental conditions and large uncertainties in measurements, as demonstrated by the current study.

We also adopted the differential split-sample test to conduct a rigorous model calibration (for the baseline treatment aCO_2-aN) and validation (for the other three treatments under differential CO_2 and N supply), which has been considered as the best possible approach for model parameterization (Refsgaard, 1997) and helped to demonstrate the predictive power of the calibrated model. During this process, we implemented advanced model-data integration by combining a wide spectrum of observations ranging from conventional measurements (e.g., soil respiration fluxes, concentrations of $NH_4^+$and NO_3^-), to less frequently measured variables (e.g., all kinds of inorganic N fluxes and microbial biomass), and to rarely available gene abundance data of multiple enzyme systems that regulates SOM decomposition and inorganic N processes.

Simulation of some processes and properties were improved using our new modeling approach, while others were not. For example, we incorporated new data associated with N processes from the BioCON experiment into model calibration and validation, and compared to the BioCON results from the MEND-old model (Gao *et al.*, 2020), the simulated NH_4^+ and NO_3^- concentrations from this study were improved as indicated by much lower biases and generally lower AIC (except aCO_2-aN). By contrast, the model performance in soil respiration simulations was consistent between MEND-new (R^2 = 0.56–0.61) and MEND-old (R^2 = 0.53–0.61). Previous studies have demonstrated that the incorporation of more detailed biogeochemical processes might not necessarily improve modeling performance of soil respiration, as multi-objective model calibration aims to find a compromise between different objectives, such as various observed C-N pool sizes and process rates other than soil respiration (Bao *et al.*, 2012, Davidson *et al.*, 2012, Wang & Chen, 2012, Wang *et al.*, 2019). Such model-data integration with multiple datasets on diverse system processes is crucial for examining the model's capability in representing a multitude of soil biogeochemical processes. In addition, model goals are not limited to improving gross predictions but also gaining insights to underlying processes. Mechanistic understanding and representation of microbially-mediated biogeochemical processes would help depict ecosystem responses to diverse perturbations more confidently.

Our direct validation of simulated microbial C:N ratios exemplifies the predictive power of the MEND-new model, given that microbial C:N ratios were not included in model calibration. The near congruence between observed and simulated microbial C:N ratios indicated the substantive improvement of the MEND-new model over the old version through mechanistic representation of N processes including dynamic N mineralization-immobilization and the competitive N uptake between plants and microbes. In addition, the MEND-new model predicted decreased microbial C:N ratios under enriched N supply (Xiao *et al.*, 2018), owing to insignificant changes in microbial biomass C and significantly increased microbial biomass N.

It should be noted that the soil system studied in this study could be not well representative. The soil in the experimental site is a Typic Udipsamments that is minimally developed with no diagnostic horizons, and it could have little potential for stabilizing organic matter by mineral sorption or occlusion in aggregates (Schimel & Schaeffer, 2012, Six *et al.*, 2002). Mineral interactions and spatial processes could play a small role in regulating the processing of plant detritus or SOM, or of N cycling processes in this soil. In addition, anaerobiosis and anerobic micro-sites are likely uncommon in the coarsely textured soils with high saturated hydraulic conductivity (O'Geen *et al.*, 2017), which will certainly affect N (especially denitrification) dynamics, differently from other soils that are more developed with more texture structure (Kristensen *et al.*, 2000). Nevertheless, this simple soil system is a perfect test-bed in many ways for experimental and modeling ideas. However, the parameterization of the model might not be readily applicable to other soils that have a fine texture and/or aggregate development, in which microbe-substrate-mineral interactions regulate the functioning of the biological components of the soil system. More likely, the model might just need different parameterization or perhaps more sophisticated treatment of organic-mineral interactions. Testing this would be a natural next phase in evaluating the model's applicability in diverse soils and ecosystems.

In summary, this study presents substantive methodological and ecological advances relative to previous studies, including our recent publication (Gao *et al.*, 2020), in that (i) the MEND-new model now includes a more detailed representation of enzyme-catalyzed N transformation processes, with the addition of a competitive dynamic enzyme allocation scheme to tackle the synthesis of multiple enzyme systems; (ii) the model was calibrated against a variety of observed N fluxes and validated by gene abundances for six N-associated processes, indicating that the changes in enzyme activities, rather than enzyme concentrations and reaction rates, were better explained by the measured gene abundances in responses to eCO_2; and (iii) the MEND-new model's predictions agreed well with the literature in terms of microbial C:N ratios and decreased microbial C:N as a result of N addition, whereas the MEND-old model did not. Taken together, our results indicated that representing microbial-enzyme groups in ecosystem models is a potentially valuable step forward to develop robust predictive models that interpolate or extrapolate observed interactions among microbes and soil C-N cycling, likely bolstering confidence in the assessments and projections of carbon-climate feedbacks. Pertaining to model refinement, a comprehensive understanding of microbial communities and their roles in regulating specific C and nutrient processes is essential for successful incorporation of enzymes-based bioindicators in ecosystem modeling. The newly refined MEND model has the potential to provide a powerful avenue for understanding and testing hypotheses about microbially mediated soil biogeochemical processes under environmental changes.

Abbreviations

aCO_2: ambient atmospheric CO_2 concentration; BioCON (Biodiversity, CO_2, and Nitrogen); C: Carbon; DOM: Dissolved Organic Matter; EA: Enzyme Activity; EC: Enzyme Concentration; eCO_2: elevated atmospheric CO_2 concentration; FR: First-order reaction Rate; GA: Gene Abundance; GPP: Gross Primary Productivity; MARE: Mean Absolute Relative Error; MB: Microbial Biomass; MBC: Microbial Biomass Carbon; MB_A: Active Microbial Biomass; MB_D: Dormant Microbial Biomass; MEND: Microbial-ENzyme Decomposition model; MOM: Mineral-associated Organic Matter; MOPSA: Multi-Objective Parameter Sensitivity Analysis; MOPSO: Multi-Objective Parameter Stochastic Optimization; N: Nitrogen; PBIAS: percent bias; POM: Particulate Organic Matter; RR: Response Ratio; SCE: Shuffled Complex Evolution; SD: Standard Deviation; SOC: Soil Organic Carbon; SOM: Soil Organic Matter; SWC: Soil Water Content; SWP: Soil Water Potential; UQ-COFI: Uncertainty Quantification by Critical Objective Function Index.

ACKNOWLEDGEMENTS

The BioCON experiment, P.B.R. and S.E.H were supported by the United States Department of Agriculture (USDA) (Project 2007-35319-18305) through NSF-USDA Microbial Observatories Program, the U.S. National Science Foundation (NSF) Long-Term Ecological Research (DEB-0620652, DEB-1234162 and DEB-1831944, Long-Term Research in Environmental Biology (LTREB) grants DEB-1242531 and DEB-1753859, Biological Integration Institutes grant NSF-DBI-2021898, Ecosystem Sciences grant DEB-1120064, and Biocomplexity grant DEB-0322057,); as well as the U.S. Department of Energy Program for Ecosystem Research (DE-FG02-96ER62291). The data compilation is supported by National Science Foundation of China (NSFC 41825016) and the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0503) to Y.Y. The modeling work was supported by the Excellent Young Scientists Fund of NSFC to G.W. and the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Numbers DE-SC0004601, DE-SC0010715, DE-SC0014079, DE-SC0016247, and DE-SC0020163 and by the Office of the Vice President for Research at the University of Oklahoma, all to J.Z. We thank the reviewer, Dr. Joshua Schimel, for his insightful comments and suggestions.

DATA AVAILABILITY STATEMENT

The model code and data that support the findings of this study are openly available in GitHub at https://github.com/wanggangsheng/MEND.git. The BioCON experimental data can be freely accessed at https://www.cedarcreek.umn.edu/research/data.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

CONFLICT OF INTEREST STATEMENT

The authors declare they have no conflict of interest.

Author's contributions

All authors contributed intellectual input and assistance to this study and manuscript preparation. The original concept and modeling strategy were developed by G.W., J.Z., P.R. and S.H. Field experiments are maintained by P.R. and S.H. Model input data were compiled by Q.G., Y.Y., and G.W. The MEND modeling was developed and conducted by G.W. All data analysis and integration were guided by G.W. and J.Z. The paper was written by G.W. and J.Z., with help from Q. G., Y.Y., P.R., and S.H.

Figure legends:

Figure 1 Framework for developing the Microbial-ENzyme Decomposition (MEND) model. **(a)** MEND modeling framework. **(b)** Procedure of the Multi-Objective Parameter Stochastic Optimization (MOPSO), which directly incorporates the Shuffled Complex Evolution (SCE) algorithm into MEND model calibration.

Figure 2 Diagram of the Microbial-ENzyme Decomposition (MEND) model. R_a and R_h are autotrophic and heterotrophic respiration, respectively. POM_O and POM_H are particulate organic matter (POM) decomposed by oxidative (EP_O) and hydrolytic enzymes (EP_H), respectively. MOM is mineral-associated OM, which is decomposed by a mixed enzyme group EM. Dissolved OM (DOM) interacts with the active layer of MOM (QOM) through sorption and desorption. Litter enters POM_O, POM_H, and DOM. Microbes consist of active (MB_A) and dormant microbes (MB_D). DOM can be assimilated by MB_A. Inorganic N deposition and fertilization enter NH_4^+ and NO_3^- that can be immobilized by microbes and taken up by plant roots. NH_4^+ adsorption is also considered. N fixation, nitrification and denitrification are mediated by nitrogenases (nifH), ammonia oxidases (amoA, nxrA/B) and N-reductases (narG/napA, nirS/nirK, norB, nosZ), respectively. Inorganic N loss pathways include leaching (NO_3^- and NO_2^-) and gas emission (NO, N_2O, and N_2) from the soil to the atmosphere.

Figure 3 MEND model calibration and validation. (a) Soil respiration (R_s) calibration with ambient CO_2-ambient N (aCaN) data, **(b)** R_s validation with data from the other three treatments: elevated CO_2-aN (eCaN), aC-enriched N (aCeN), and eCeN. **(c)** Absolute value of percent bias (|*PBIAS*|, %) between simulated and observed mean for the calibration of ammonium (NH_4^+) and nitrate (NO_3^-, including both NO_3^- and NO_2^-) from aCaN. **(d)** |*PBIAS*| for the validation of NH_4^+ and NO_3^- from the other three treatments. Error bars in A represent the standard deviations. R^2 values in A and B denote the coefficient of determination. In **(c)** and **(d)**, the two models of MEND-old and MEND-new denote the old version of MEND model as described in Gao et al. (2020) and the new MEND model in this study, respectively. The two numbers in each facet of **(c)** and **(d)** denote the Akaike information criterion (AIC, lower is better) for the two models, respectively.

Figure 4 Comparison between simulated rates and literature-reported nitrogen flux rates. (a) Biological N fixation rate; the "Literature" data were from Cleveland et al. (1999, GBC) and Cleveland et al. (2013, PNAS), where the bars show the mean values and the error bar shows the value range. **(b)** Plant N uptake rate; the "Literature" data were from Bessler et al. (2012) and Reyes et al. (2015), where the error bar denotes the value range. The difference in simulated rates between paired treatments was tested by the Wilcoxon signed rank test. "*", "**", and "***" denote significant difference with *p-value* < 0.05, *p-value* < 0.01, and *p-value* < 0.001, respectively. "NS." means not significant.

Figure 5 Comparison of microbial C:N ratios and functional enzymes simulated by two models (MEND-old and MEND-new). (a) modeled versus literature-reported microbial C:N ratios (error bars denote the 95% confidence interval); **(b)** MEND-old modeled microbial C:N ratios under ambient N (aN) and enriched N (eN); **(c)** MEND-new modeled microbial C:N ratios under aN and eN; **(d)** elevated CO_2 (eCO_2) effect on oxidative enzymes; **(e)** eCO_2 effect on hydrolytic enzymes. MEND-old and MEND-new denote the old version of MEND model as described in Gao et al. (2020) and the new MEND model in this study, respectively. The "Literature" data in **(a)** were from Xu et al. (2013). The eCO_2 effects in the year of 2009 (**d** and **e**) are quantified by the response ratio (RR), which is defined as the logarithmic ratio of a variable under eCO_2 to that under ambient CO_2 (aCO_2) as per ambient N (aN) or enriched N (eN) treatment. The RRs are calculated pertaining to observed gene abundances (GA_{obs}), simulated enzyme concentrations (EC_{sim}, mg C cm^{-3} soil), simulated enzyme activities (EA_{sim}, mg C cm^{-3} C h^{-1}), and simulated first-order reaction rates (FR_{sim}, h^{-1}). The difference between paired data was tested by the Wilcoxon signed rank test. "*", "**", and "***" denote significant difference with *p-value* < 0.05, *p-value* < 0.01, and *p-value* < 0.001, respectively. "NS." means not significant.

Figure 6 Elevated CO_2 (eCO_2) effects on functional genes/enzymes quantified by the response ratio (RR) in the year of 2009. (a) RRs of observed gene abundances (GA_{obs}) *versus* simulated enzyme concentrations (EC_{sim}, mg C cm^{-3} soil), **(b)** RRs of GA_{obs} *versus* simulated enzyme activities (EA_{sim}, mg C cm^{-3} C h^{-1}), **(c)** RRs of GA_{obs} *versus* simulated first-order reaction rates (FR_{sim}, h^{-1}). The RR is defined as the logarithmic ratio of a variable under eCO_2 to that under ambient CO_2 (aCO_2) as per ambient N (aN) or enriched N (eN) treatment. Each boxplot includes eight RR values from eight genes (enzymes): two groups (oxidative and hydrolytic) for the decomposition of soil organic matter, nitrogenases (nifH), ammonia oxidases (amoA) and four N-reductases (narG/napA, nirS/nirK, norB, nosZ). The difference in RR between two variables was tested by the Wilcoxon signed rank test. "*", "**", and "***" denote significant difference with *p-value* < 0.05, *p-value* < 0.01, and *p-value* < 0.001, respectively. "NS." means not significant.

References

Abramoff RZ, Davidson EA, Finzi AC (2017) A parsimonious modular approach to building a mechanistic belowground carbon and nitrogen model. Journal of Geophysical Research: Biogeosciences, **122**, 2418-2434.

Adair EC, Reich PB, Hobbie SE, Knops JM (2009) Interactive effects of time, CO_2, N, and diversity on total belowground carbon allocation and ecosystem carbon storage in a grassland community. Ecosystems, **12**, 1037-1052.

Adair EC, Reich PB, Trost JJ, Hobbie SE (2011) Elevated CO_2 stimulates grassland soil respiration by increasing carbon inputs rather than by enhancing soil moisture. Global Change Biology, **17**, 3546-3563.

Allison SD, Wallenstein MD, Bradford MA (2010) Soil-carbon response to warming dependent on microbial physiology. Nature Geoscience, **3**, 336-340.

Allison SD, Weintraub MN, Gartner TB, Waldrop MP (2011) Evolutionary-Economic Principles as Regulators of Soil Enzyme Production and Ecosystem Function. In: *Soil Enzymology.* (eds Shukla G, Varma A) pp 229-243. Berlin Heidelberg, Springer-Verlag.

Averill C (2014) Divergence in plant and microbial allocation strategies explains continental patterns in microbial allocation and biogeochemical fluxes. Ecology Letters, **17**, 1202-1210.

Bailey VL, Bond-Lamberty B, Deangelis K *et al.* (2018) Soil carbon cycling proxies: understanding their critical role in predicting climate change feedbacks. Global Change Biology, **24**, 895-905.

Bao Z, Zhang J, Liu J *et al.* (2012) Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. Journal of Hydrology*,* **466**, 37-46.

Bardgett RD, Freeman C, Ostle NJ (2008) Microbial contributions to climate change through carbon cycle feedbacks. The ISME Journal*,* **2**, 805-814.

Bessler H, Oelmann Y, Roscher C *et al.* (2012) Nitrogen uptake by grassland communities: contribution of N_2 fixation, facilitation, complementarity, and species dominance. Plant and Soil*,* **358**, 301-322.

Bradford MA, Wieder WR, Bonan GB, Fierer N, Raymond PA, Crowther TW (2016) Managing uncertainty in soil carbon feedbacks to climate change. Nature Climate Change*,* **6**, 751-758.

Cavicchioli R, Ripple WJ, Timmis KN *et al.* (2019) Scientists' warning to humanity: microorganisms and climate change. Nature Reviews Microbiology*,* **17**, 569-586.

Chen J, Sinsabaugh RL (2021) Linking microbial functional gene abundance and soil extracellular enzyme activity: Implications for soil carbon dynamics. Global Change Biology*,* **27**, 1322-1325.

Cleveland CC, Houlton BZ, Smith WK *et al.* (2013) Patterns of new versus recycled primary production in the terrestrial biosphere. Proceedings of the National Academy of Sciences*,* **110**, 12733-12737.

Cleveland CC, Townsend AR, Schimel DS *et al.* (1999) Global patterns of terrestrial biological nitrogen (N_2) fixation in natural ecosystems. Global Biogeochemical Cycles*,* **13**, 623-645.

Conover WJ (1998) *Practical Nonparametric Statistics, 3rd Edition,* New York, John Wiley & Sons.

Davidson EA, Samanta S, Caramori SS, Savage K (2012) The Dual Arrhenius and Michaelis–Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. Global Change Biology, **18**, 371-384.

Dijkstra FA, West JB, Hobbie SE, Reich PB, Trost J (2007) Plant diversity, CO_2, and N influence inorganic and organic N leaching in grasslands. Ecology, **88**, 490-500.

Drake J, Darby B, Giasson M-A, Kramer M, Phillips R, Finzi A (2013) Stoichiometry constrains microbial response to root exudation-insights from a model and a field experiment in a temperate forest. Biogeosciences, **10**, 821-838.

Du Z, Weng E, Jiang L, Luo Y, Xia J, Zhou X (2018) Carbon–nitrogen coupling under three schemes of model representation: a traceability analysis. Geoscientific Model Development, **11**, 4399-4416.

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

Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science, **320**, 1034-1039.

Fanin N, Fromin N, Barantal S, Hättenschwiler S (2017) Stoichiometric plasticity of microbial communities is similar between litter and soil in a tropical rainforest. Scientific Reports, **7**, 12498.

Fiencke C, Bock E (2006) Immunocytochemical localization of membrane-bound ammonia monooxygenase in cells of ammonia oxidizing bacteria. Archives of Microbiology, **185**, 99-106.

Gao Q, Wang G, Xue K *et al.* (2020) Stimulation of soil respiration by elevated CO_2 is enhanced under nitrogen limitation in a decade-long grassland study. Proceedings of the National Academy of Sciences, **117**, 33317-33324.

Goll DS, Brovkin V, Parida BR *et al.* (2012) Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. Biogeosciences, **9**, 3547-3569.

Guo X, Gao Q, Yuan M *et al.* (2020) Gene-informed decomposition model predicts lower soil carbon loss due to persistent microbial adaptation to warming. Nature Communications, **11**, 4897.

Harty MA, Forrestal PJ, Carolan R *et al.* (2017) Temperate grassland yields and nitrogen uptake are influenced by fertilizer nitrogen source. Agronomy Journal, **109**, 71-79.

Hommel B (2020) Pseudo-mechanistic explanations in psychology and cognitive neuroscience. Topics in Cognitive Science, **12**, 1294-1305.

Hu H-W, Chen D, He J-Z (2015) Microbial regulation of terrestrial nitrous oxide formation: understanding the biological pathways for prediction of emission rates. FEMS Microbiology Reviews, **39**, 729-749.

Jian S, Li J, Chen J *et al.* (2016) Soil extracellular enzyme activities, soil carbon and nitrogen storage under nitrogen fertilization: A meta-analysis. Soil Biology and Biochemistry, **101**, 32-43.

Kazanski CE, Cowles J, Dymond S *et al.* (2021) Water availability modifies productivity response to biodiversity and nitrogen in long-term grassland experiments. Ecological Applications, **31**, e02363.

Klausmeier CA, Kremer CT, Koffel T (2020) Traits-based ecological and eco-evolutionary theory. In: *Theoretical Ecology: concepts and applications.* (eds Mccann KS, Gellner G) pp 161-194. Oxford, UK, Oxford University Press.

Kristensen HL, Mccarty GW, Meisinger JJ (2000) Effects of Soil Structure Disturbance on Mineralization of Organic Soil Nitrogen. Soil Science Society of America Journal*,* **64**, 371-378.

Kyker-Snowman E, Wieder WR, Frey SD, Grandy AS (2020) Stoichiometrically coupled carbon and nitrogen cycling in the MIcrobial-MIneral Carbon Stabilization model version 1.0 (MIMICS-CN v1.0). Geosci. Model Dev.*,* **13**, 4413-4434.

Li M, Qian W-J, Gao Y, Shi L, Liu C (2017) Functional enzyme-based approach for linking microbial community functions with biogeochemical process kinetics. Environmental Science & Technology*,* **51**, 11848-11857.

Luo Y, Hui D, Zhang D (2006) Elevated CO_2 stimulates net accumulations of carbon and nitrogen in land ecosystems: A meta-analysis. Ecology*,* **87**, 53-63.

Luo Y, Randerson JT, Abramowitz G *et al.* (2012) A framework for benchmarking land models. Biogeosciences*,* **9**, 3857-3874.

Manzoni S, Moyano F, Kätterer T, Schimel J (2016) Modeling coupled enzymatic and solute transport controls on decomposition in drying soils. Soil Biology and Biochemistry*,* **95**, 275-287.

Mooshammer M, Wanek W, Hämmerle I *et al.* (2014a) Adjustment of microbial nitrogen use efficiency to carbon: nitrogen imbalances regulates soil nitrogen cycling. Nature Communications*,* **5**, 3694.

Mooshammer M, Wanek W, Zechmeister-Boltenstern S, Richter AA (2014b) Stoichiometric imbalances between terrestrial decomposer communities and their resources: mechanisms and implications of microbial adaptations to their resources. Frontiers in Microbiology, **5**, Article 22.

Moriasi D, Arnold J, Van Liew M, Bingner R, Harmel R, Veith T (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, **50**, 885-900.

Müller S, Regensburger G, Steuer R (2014) Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. Journal of Theoretical Biology, **347**, 182-190.

Ning D, Yuan M, Wu L *et al.* (2020) A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. Nature Communications, **11**, 4717.

O'geen A, Walkinshaw M, Beaudette D (2017) SoilWeb: A Multifaceted Interface to Soil Survey Information. Soil Science Society of America Journal, **81**, 853-862.

Ouyang Y, Reeve J, Norton J (2018) Soil enzyme activities and abundance of microbial functional genes involved in nitrogen transformations in an organic farming system. Biology and Fertility of Soils, **54**, 1-14.

Pagel H, Poll C, Ingwersen J, Kandeler E, Streck T (2016) Modeling coupled pesticide degradation and organic matter turnover: From gene abundance to process rates. Soil Biology and Biochemistry, **103**, 349-364.

Petersen DG, Blazewicz SJ, Firestone M, Herman DJ, Turetsky M, Waldrop M (2012) Abundance of microbial genes associated with nitrogen cycling as indices of biogeochemical process rates across a vegetation gradient in Alaska. Environmental Microbiology, **14**, 993-1008.

Reed DC, Algar CK, Huber JA, Dick GJ (2014) Gene-centric approach to integrating environmental genomics and biogeochemical models. Proceedings of the National Academy of Sciences, **111**, 1879-1884.

Refsgaard JC (1997) Parameterisation, calibration and validation of distributed hydrological models. Journal of Hydrology, **198**, 69-97.

Reich PB, Hobbie SE (2013) Decade-long soil nitrogen constraint on the CO_2 fertilization of plant biomass. Nature Climate Change, **3**, 278.

Reyes J, Schellberg J, Siebert S, Elsaesser M, Adam J, Ewert F (2015) Improved estimation of nitrogen uptake in grasslands using the nitrogen dilution curve. Agronomy for Sustainable Development, **35**, 1561-1570.

Rocca JD, Hall EK, Lennon JT *et al.* (2015) Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. The ISME Journal, **9**, 1693-1699.

Schimel JP (2013) Microbes and global carbon. Nature Climate Change, **3**, 867-868.

Schimel JP, Schaeffer SM (2012) Microbial control over carbon cycling in soil. Frontiers in Microbiology, **3**, Article 348.

Schimel JP, Weintraub MN (2003) The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. Soil Biology and Biochemistry, **35**, 549-563.

Schlesier J, Rohde M, Gerhardt S, Einsle O (2016) A conformational switch triggers nitrogenase protection from oxygen damage by Shethna protein II (FeSII). Journal of the American Chemical Society, **138**, 239-247.

Shi Z, Yin H, Van Nostrand JD *et al.* (2019) Functional gene array-based ultrasensitive and quantitative detection of microbial populations in complex communities. MSystems, **4**, e00296-00219.

Sinsabaugh R, Moorhead D (1994) Resource allocation to extracellular enzyme production: a model for nitrogen and phosphorus control of litter decomposition. Soil Biology and Biochemistry, **26**, 1305-1311.

Sinsabaugh RL, Belnap J, Findlay SG *et al.* (2014) Extracellular enzyme kinetics scale with resource availability. Biogeochemistry, **121**, 287-304.

Sinsabaugh RL, Carreiro MM, Repert DA (2002) Allocation of extracellular enzymatic activity in relation to litter composition, N deposition, and mass loss. Biogeochemistry, **60**, 1-24.

Six J, Conant RT, Paul EA, Paustian K (2002) Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. Plant and Soil, **241**, 155-176.

Soil Survey Staff (1999) *Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys,* Washington, DC, United States Department of Agriculture, Natural Resources Conservation Service.

Song H-S, Thomas DG, Stegen JC *et al.* (2017) Regulation-structured dynamic metabolic model provides a potential mechanism for delayed enzyme response in denitrification process. Frontiers in Microbiology, **8**, 1866.

Sulman BN, Moore JA, Abramoff R *et al.* (2018) Multiple models and experiments underscore large uncertainty in soil carbon dynamics. Biogeochemistry, **141**, 109-123.

Tang J, Riley WJ (2019) A theory of effective microbial substrate affinity parameters in variably saturated soils and an example application to aerobic soil heterotrophic respiration. Journal of Geophysical Research: Biogeosciences, **124**, 918-940.

Thornton PE, Lamarque JF, Rosenbloom NA, Mahowald NM (2007) Influence of carbon-nitrogen cycle coupling on land model response to CO2 fertilization and climate variability. Global Biogeochemical Cycles, **21**, GB4018.

Todd-Brown KE, Hopkins FM, Kivlin SN, Talbot JM, Allison SD (2012) A framework for representing microbial decomposition in coupled climate models. Biogeochemistry, **109**, 19-33.

Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology, **5**, 240-245.

Treseder KK (2008) Nitrogen additions and microbial biomass: A meta-analysis of ecosystem studies. Ecology Letters, **11**, 1111-1120.

Trivedi P, Anderson IC, Singh BK (2013) Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. Trends in Microbiology, **21**, 641-651.

Tu Q, Yu H, He Z *et al.* (2014) GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. Molecular Ecology Resources, **14**, 914-928.

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

Wang G, Huang W, Mayes MA *et al.* (2019) Soil moisture drives microbial controls on carbon decomposition in two subtropical forests. Soil Biology and Biochemistry*,* **130**, 185-194.

Wang G, Huang W, Zhou G, Mayes MA, Zhou J (2020) Modeling the processes of soil moisture in regulating microbial and carbon-nitrogen cycling. Journal of Hydrology*,* **585**, 124777.

Wang G, Jagadamma S, Mayes MA, Schadt CW, Steinweg JM, Gu L, Post WM (2015) Microbial dormancy improves development and experimental validation of ecosystem model. The ISME Journal*,* **9**, 226-237.

Wang G, Li W, Wang K, Huang W (2021) Uncertainty quantification of the soil moisture response functions for microbial dormancy and resuscitation. Soil Biology and Biochemistry*,* **160**, 108337.

Wang G, Post WM, Mayes MA (2013) Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic analyses. Ecological Applications*,* **23**, 255-272.

Wieder WR, Allison SD, Davidson EA *et al.* (2015) Explicitly representing soil microbial processes in Earth system models. Global Biogeochemical Cycles*,* **29**, 1782-1800.

Xiao W, Chen X, Jing X, Zhu B (2018) A meta-analysis of soil extracellular enzyme activities in response to global change. Soil Biology and Biochemistry*,* **123**, 21-32.

Xu X, Thornton PE, Post WM (2013) A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. Global Ecology and Biogeography*,* **22**, 737-749.

Xue K, Yuan MM, Shi ZJ *et al.* (2016) Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming. Nature Climate Change*,* **6**, 595-600.

Zechmeister-Boltenstern S, Keiblinger KM, Mooshammer M, Peñuelas J, Richter A, Sardans J, Wanek W (2015) The application of ecological stoichiometry to plant–microbial–soil organic matter transformations. Ecological Monographs*,* **85**, 133-155.

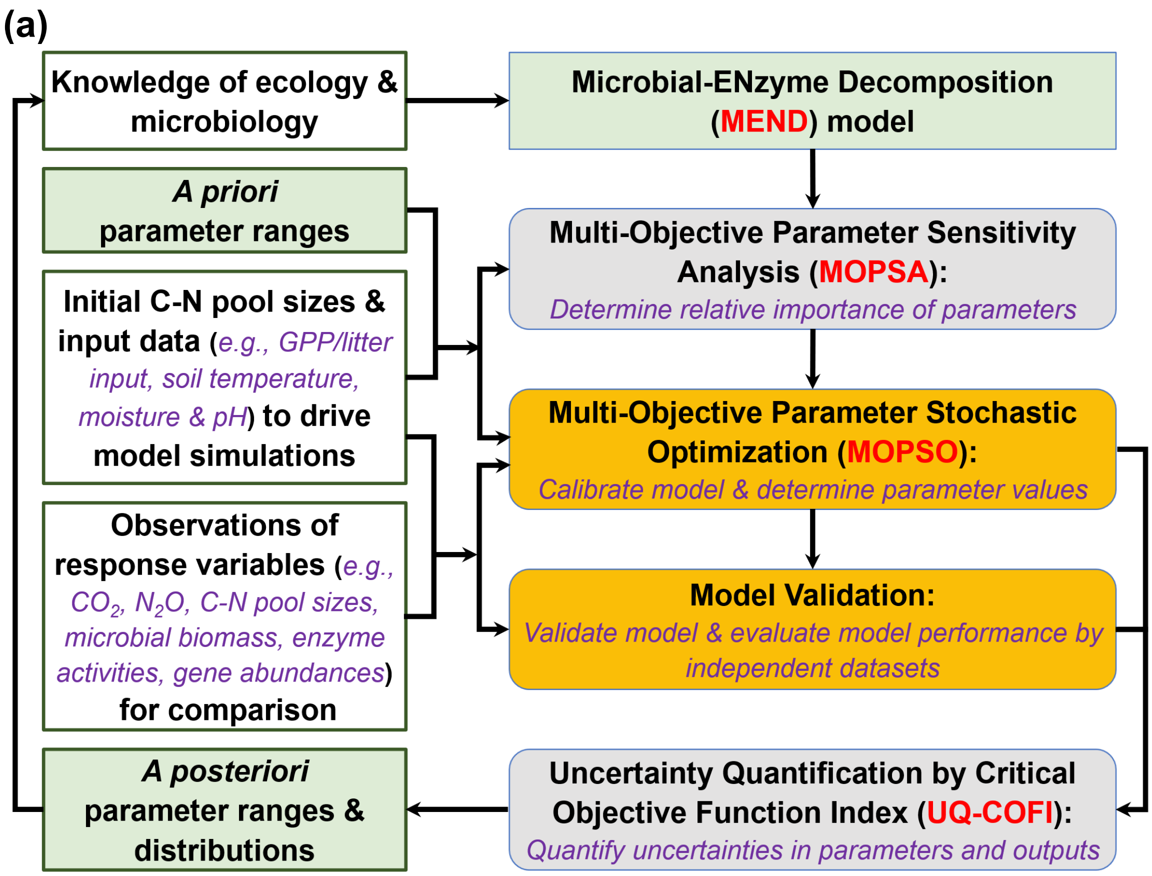
Zhou J, Xue K, Xie J *et al.* (2012) Microbial mediation of carbon-cycle feedbacks to climate warming. Nature Climate Change*,* **2**, 106-110.

Zhu X, Pei Y, Zheng Z *et al.* (2018) Underestimates of grassland gross primary production in MODIS standard products. Remote Sensing*,* **10**, 1771.

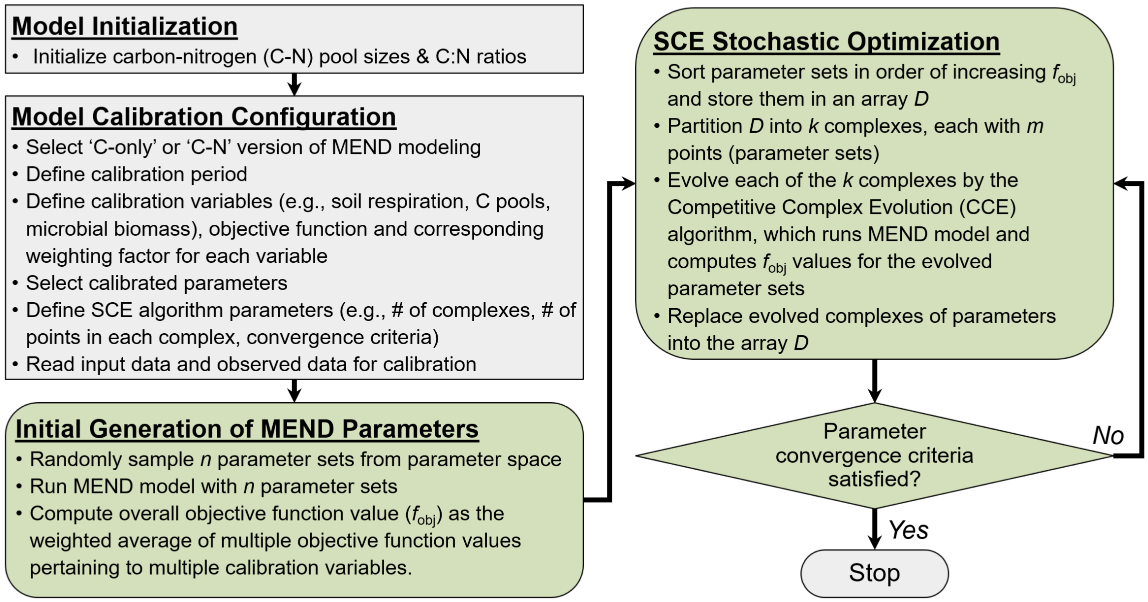
Table 1. BioCON data for MEND model calibration and validation

Response variable	Description	Objective Function	Number of data points
R_s (CO_2)	Soil respiration = root respiration (R_a) + microbial respiration (R_h)	$J_1 = R^2$	284
MBC	Microbial biomass carbon	$J_2 = MAREt$, tolerance = 10%	1
SOC	Soil organic carbon	$J_3 = MAREt$, tolerance = 5%	1
NH_4^+	Ammonium concentration	$J_4 = 0.8 \times \lvert PBIAS \rvert + 0.2 \times MARE$	8
$NO_3^- + NO_2^-$	Nitrate+Nitrite concentration	$J_5 = 0.8 \times \lvert PBIAS \rvert + 0.2 \times MARE$	8
$FN_{mn\text{-}im}$	Net N mineralization rate	$J_6 = MAREt$, tolerance = 0.5	10
FN_{nit}	Nitrification flux rate	$J_7 = MAREt$, tolerance = 0.9	10
FN_{fix}	N fixation flux rate	$J_8 = MAREt$, tolerance = 0.2	1
$FN_{im,VG}$	Plant uptake rate of N	$J_9 = MAREt$, tolerance = 0.5	1
EPO	Oxidative Enzyme	For model validation only: Compare simulated and observed Response Ratios (*RR*). Observed *RR* is the response ratio of omics-detected gene abundances between elevated CO_2 (eCO_2) and ambient CO_2 (aCO_2). Simulated *RR* is the response ratio of MEND-modeled enzyme concentrations, activities, or reaction rates between eCO_2 and aCO_2.	1
EPH	Hydrolytic Enzyme		1
ENH4	Ammonium oxidase		1
ENO3	Nitrate reductase		1
ENO2	Nitrite reductase		1
ENO	Nitric oxide reductase		1
EN2O	Nitrous oxide reductase		1
EN2	Nitrogenase		1

Notes: R^2 denotes the coefficient of determination, $|PBIAS|$ is the absolute value of the percent bias, *MARE* is the mean absolute relative error, *MAREt* is the *MARE* with a tolerance. See Supporting Information Eqs. 68–71 for a description of these criteria.



(b) Multi-Objective Parameter Stochastic Optimization (MOPSO)



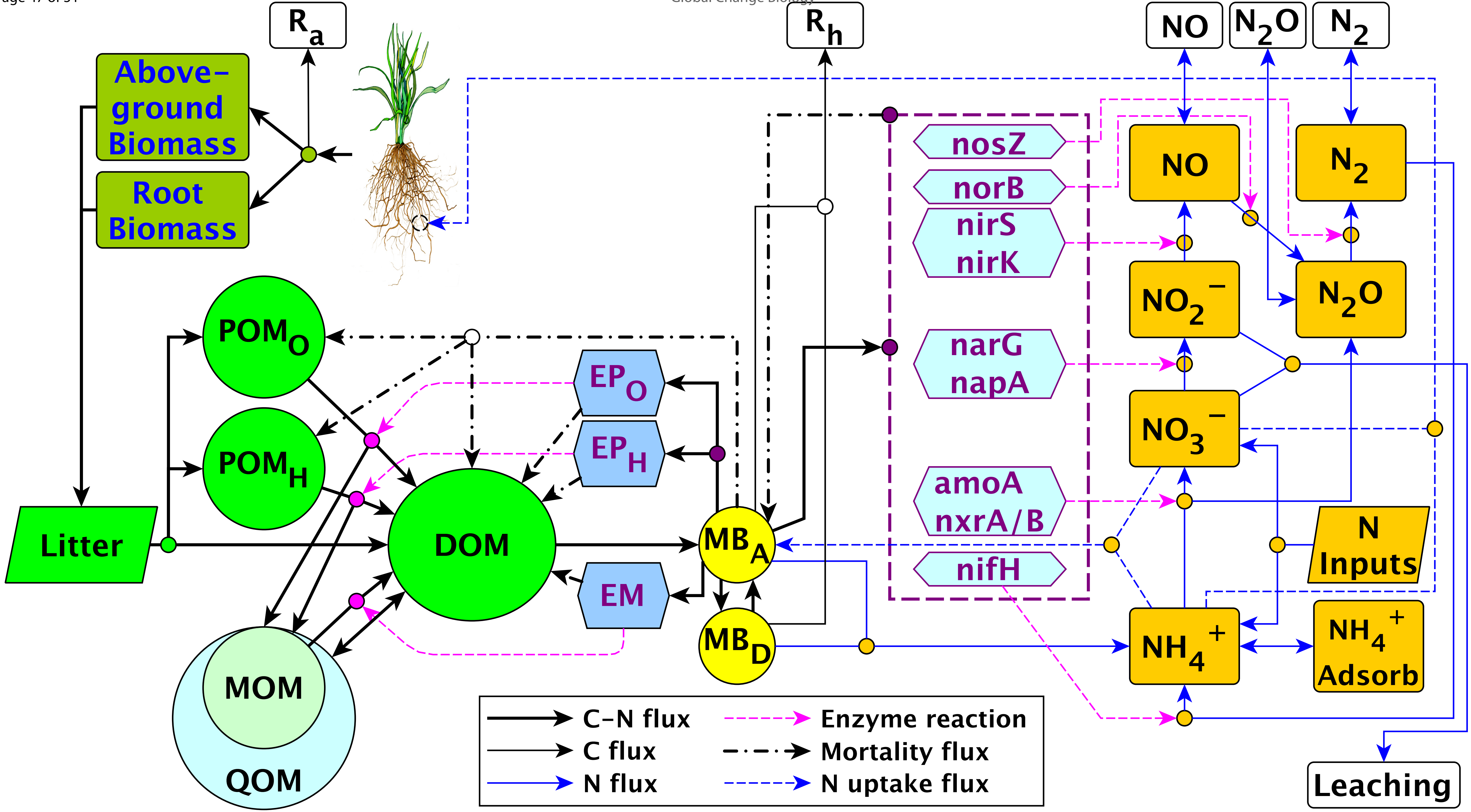

R_a
R_h
NO
N_2O
N_2
Above-ground Biomass
Root Biomass
nosZ
norB
nirS
nirK
narG
napA
amoA
nxrA/B
nifH
NO
N_2
NO_2^-
N_2O
NO_3^-
POM_O
POM_H
EP_O
EP_H
Litter
DOM
MB_A
MB_D
EM
N Inputs
NH_4^+
NH_4^+ Adsorb
MOM
QOM
Leaching
C-N flux
C flux
N flux
Enzyme reaction
Mortality flux
N uptake flux

Fig. 3

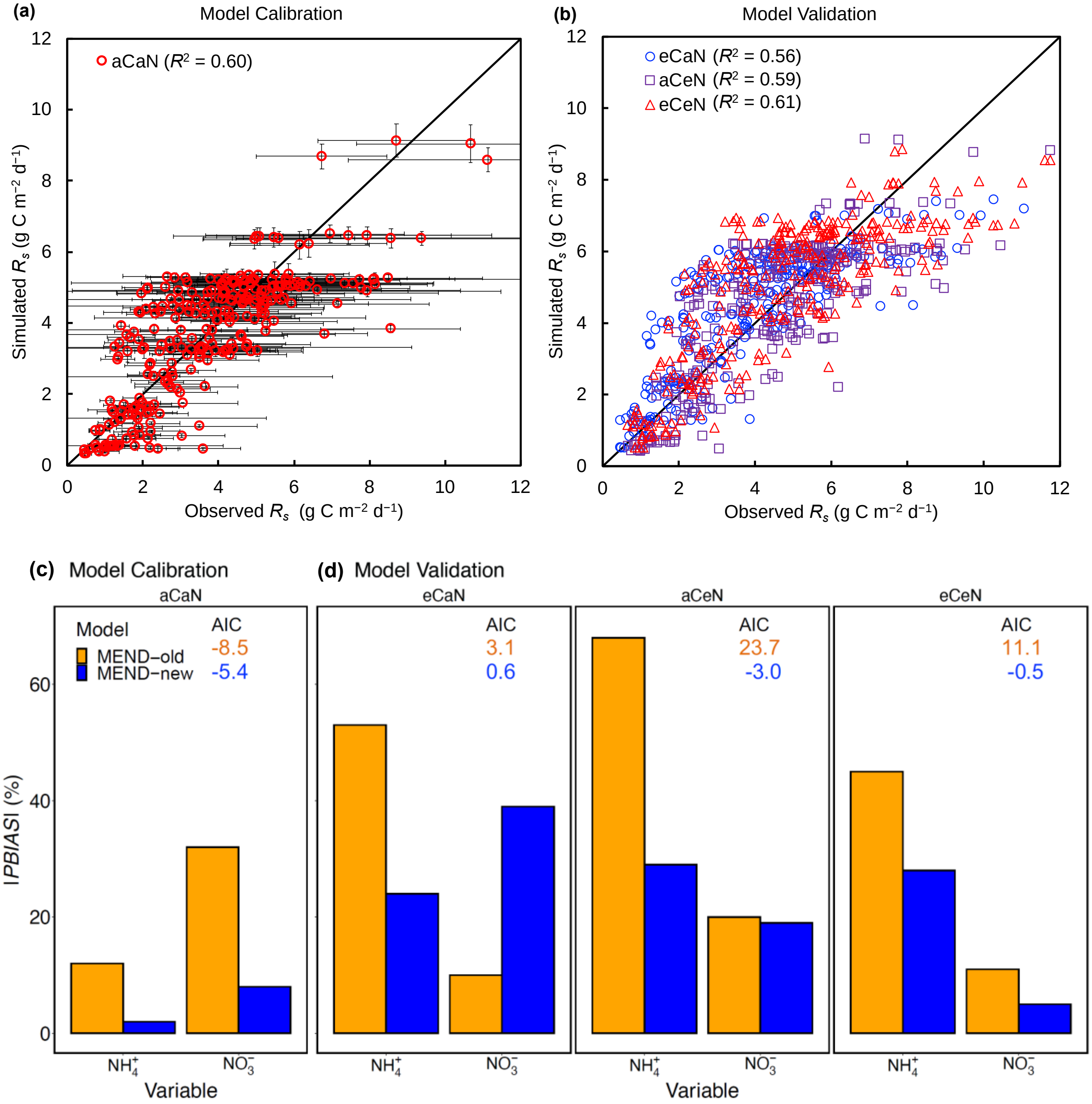

(a)
Model Calibration
aCaN (R^2 = 0.60)
Simulated R_s (g C m^{-2} d^{-1})
Observed R_s (g C m^{-2} d^{-1})
0
2
4
6
8
10
12
(b)
Model Validation
eCaN (R^2 = 0.56)
aCeN (R^2 = 0.59)
eCeN (R^2 = 0.61)
Simulated R_s (g C m^{-2} d^{-1})
Observed R_s (g C m^{-2} d^{-1})
(c) Model Calibration
aCaN
Model
MEND–old
MEND–new
AIC
-8.5
-5.4
(d) Model Validation
eCaN
AIC
3.1
0.6
aCeN
AIC
23.7
-3.0
eCeN
AIC
11.1
-0.5
|PBIAS| (%)
0
20
40
60
NH_4^+
NO_3^-
Variable

Fig. 4

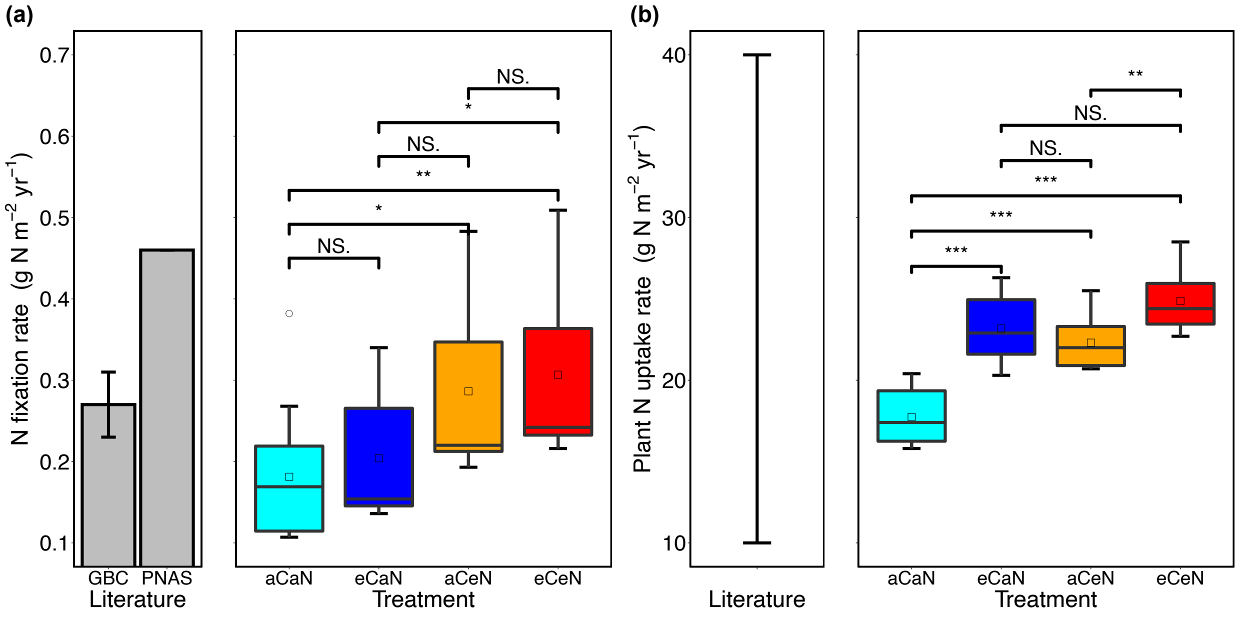

(a)
N fixation rate (g N m^{-2} yr^{-1})
0.7
0.6
0.5
0.4
0.3
0.2
0.1
GBC
PNAS
Literature
NS.
*
NS.
**
*
NS.
aCaN
eCaN
aCeN
eCeN
Treatment
(b)
Plant N uptake rate (g N m^{-2} yr^{-1})
40
30
20
10
Literature
**
NS.
NS.
***
***
***
aCaN
eCaN
aCeN
eCeN
Treatment

Fig. 5

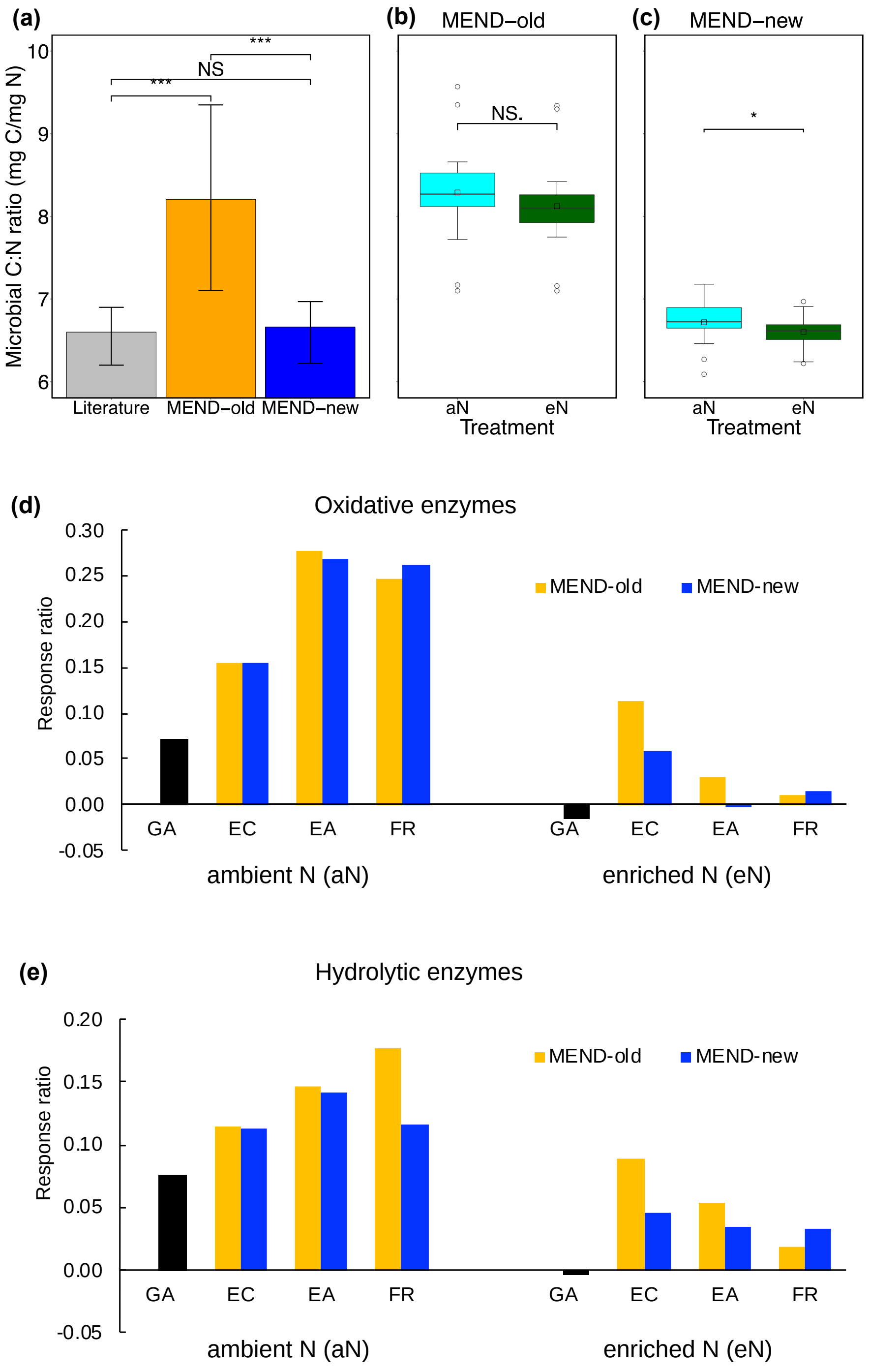
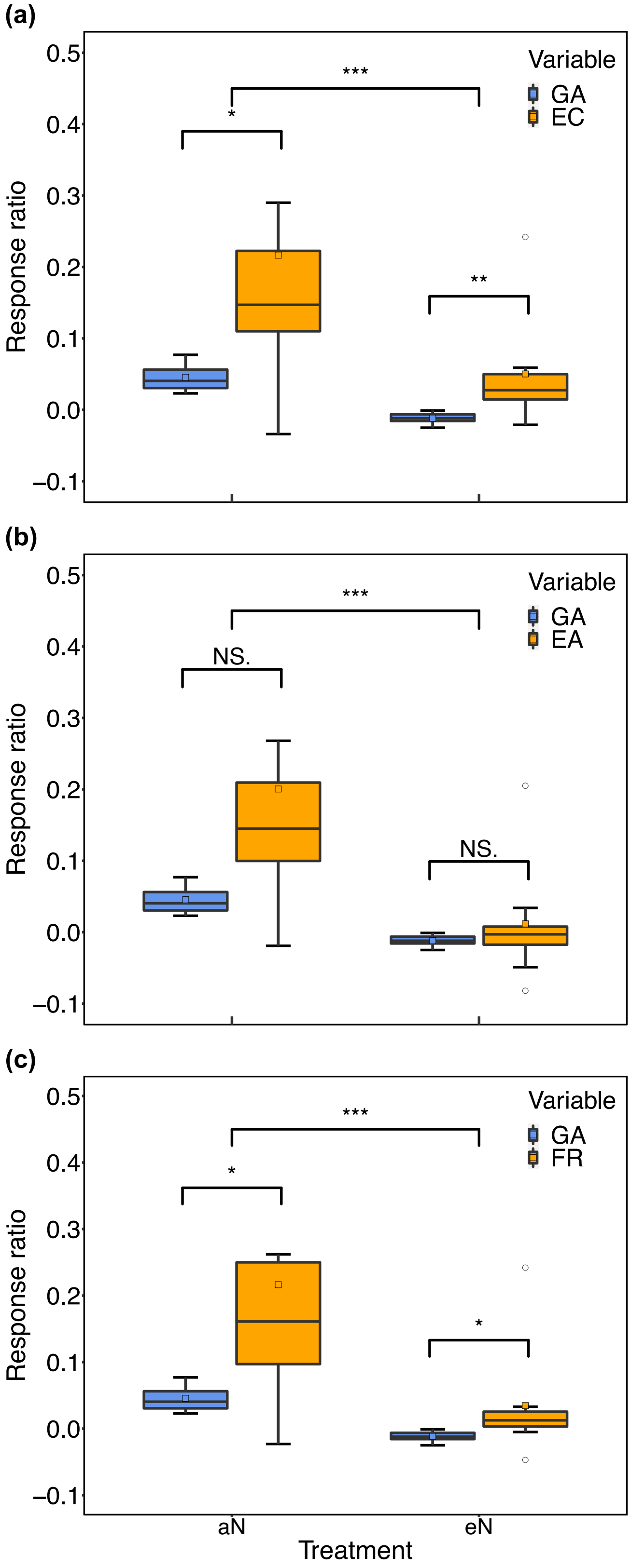

(a)
Microbial C:N ratio (mg C/mg N)
10
9
8
7
6
***
NS
***
Literature
MEND–old
MEND–new
(b)
MEND–old
NS.
aN
eN
Treatment
(c)
MEND–new
*
aN
eN
Treatment
(d)
Oxidative enzymes
Response ratio
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
MEND-old
MEND-new
GA
EC
EA
FR
GA
EC
EA
FR
ambient N (aN)
enriched N (eN)
(e)
Hydrolytic enzymes
Response ratio
0.20
0.15
0.10
0.05
0.00
-0.05
MEND-old
MEND-new
GA
EC
EA
FR
GA
EC
EA
FR
ambient N (aN)
enriched N (eN)

Fig. 6


(a)
Response ratio
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
***
*
**
Variable
GA
EC
(b)
Response ratio
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
***
NS.
NS.
Variable
GA
EA
(c)
Response ratio
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
***
*
*
Variable
GA
FR
aN
eN
Treatment