

---

# Applying the MEMS 3.0 Model Across Contrasting Grazing US Ecoregions and Management: Model Uncertainty and Evaluation

---

Rafael S. Santos , Emma K. Hamilton , Paige L. Stanley , Robert Clement , Rebecca Mitchell , Hao Yang , Lauren Hoskovec , [Isabella C. F. Maciel](#) , [Jason Rowntree](#) , [John D. Scasta](#) , Alejandro Romero-Ruiz , Stephen M. Ogle , [M. Francesca Cotrufo](#) , [Yao Zhang](#) \*

Posted Date: 3 March 2026

doi: 10.20944/preprints202603.0030.v1

Keywords: Bayesian calibration; Ecosystem model; SOM fractions; Adaptive grazing



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Applying the MEMS 3.0 Model Across Contrasting Grazing US Ecoregions and Management: Model Uncertainty and Evaluation

Rafael S. Santos <sup>1</sup>, Emma K. Hamilton <sup>2</sup>, Paige L. Stanley <sup>2</sup>, Robert Clement <sup>3</sup>, Rebecca Mitchell <sup>3</sup>, Hao Yang <sup>2</sup>, Lauren Hoskovec <sup>1</sup>, Isabella C. F. Maciel <sup>4</sup>, Jason Rowntree <sup>5</sup>, John D. Scasta <sup>6</sup>, Alejandro Romero-Ruiz <sup>7</sup>, Stephen M. Ogle <sup>1,8</sup>, M. Francesca Cotrufo <sup>2</sup> and Yao Zhang <sup>1,\*</sup>

<sup>1</sup> Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO, 80521, USA

<sup>2</sup> Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, 80521, USA

<sup>3</sup> Quanterra Systems, Centenary House, Peninsula Park, Exeter, EX2 7XE, UK

<sup>4</sup> Noble Research Institute, Ardmore, OK, 73401, USA

<sup>5</sup> Department of Animal Science, Michigan State University, East Lansing, MI, 48824, USA

<sup>6</sup> Department of Ecosystem Science and Management, University of Wyoming, Laramie, 82070, WY, USA

<sup>7</sup> Centre for Hydrogeology and Geothermics, University of Neuchâtel, Neuchâtel, Switzerland

<sup>8</sup> Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA, 80525

\* Correspondence: yao.zhang@colostate.com

## Abstract

Grazinglands store substantial soil organic carbon (SOC), yet their potential to act as net carbon (C) sinks depends on management-driven net ecosystem CO<sub>2</sub> exchange (NEE). Process-based models capable of representing contrasting management strategies are essential for evaluating mitigation potential. However, the uncertainty inherent in model outputs is often overlooked, which can limit the reliability of predictions. Here, we apply the MEMS 3.0 model to quantify uncertainty in SOC and NEE predictions and evaluate model performance across diverse U.S. grazing ecoregions. We conducted a Bayesian calibration with observations of SOC and its fractions, and annual NEE measurements from six NEON grazing sites. Model evaluation was performed with an independent dataset collected from four experimental sites located in Oklahoma, Michigan, and Wyoming USA under prescriptive and adaptive grazing management treatments. The model performed well in predicting baseline SOC and estimating weekly ecosystem fluxes across sites. Although annual NEE estimates exhibited some discrepancies relative to flux-tower observations, ~90% of measured fluxes fell within simulated posterior predictive intervals. Moreover, the model is consistent with the flux-observations, demonstrating no significant treatment differences between prescriptive and adaptive grazing treatments at the Oklahoma and Wyoming sites. We demonstrated that MEMS 3.0 can represent SOC dynamics and ecosystem fluxes in grazinglands across contrasting climates. Our results show that neglecting uncertainty in measured and simulated fluxes can lead to misleading model-data comparisons. These findings highlight the importance of uncertainty quantification for robust interpretation of predicted grazing management outcomes and for supporting credible climate change mitigation and C accounting frameworks.

**Keywords:** Bayesian calibration; Ecosystem model; SOM fractions; Adaptive grazing

---

## Highlights

- MEMS captures ecosystem SOC and NEE dynamics across diverse U.S. grazing lands
- The Bayesian framework used enabled the quantification of full model uncertainty
- Interpretation of predicted grazing outcomes can be limited without uncertainty quantification

- Quantifying model and measurement uncertainty is key for model-data comparison

## 1. Introduction

Climate change remains one of the greatest challenges of modern society, and drastically reducing atmospheric carbon dioxide (CO<sub>2</sub>) is needed to limit global warming below 1.5 °C [1,2]. Although agriculture has historically contributed to rising atmospheric CO<sub>2</sub> levels [3,4], it also offers some of the most scalable opportunities for CO<sub>2</sub> drawdown. Grazing lands are central to this potential as they represent one of the most extensive land uses on Earth covering ~25% of the planet's ice-free surface [5] and ~67% of global agricultural land [6]. Moreover, these systems are estimated to store 10-30% of global stock of soil organic carbon (SOC), which makes their CO<sub>2</sub> balance particularly important from a climate change perspective. Grazing management plays a crucial role in shaping whether these systems can act as net sources or sinks of CO<sub>2</sub> [7,8], representing a major opportunity to mitigate climate change [9]. However, fully harnessing this potential requires robust tools capable of accurately simulating grazing management regimes and predicting their effects on CO<sub>2</sub> fluxes across spatial and temporal scales. Such tools are crucial to inform sustainable land management and support climate policy and carbon (C) market frameworks [10].

Grazing management strategies strongly influence atmospheric CO<sub>2</sub> removal primarily by regulating plant growth via defoliation. By altering the light intercepted by plants, defoliation affects the allocation of photosynthates within the plant [11,12], ultimately affecting the quantity and quality of plant inputs to the soil [13,14]. Additionally, the timing, intensity, duration, and frequency under which pastures are grazed are critical management levers that regulate plant-soil feedbacks [15]. By shaping canopy structure, regrowth dynamics, and root turnover, these levers influence both the amount of atmospheric CO<sub>2</sub> assimilated by the vegetation and C input to the soil [15]. As a result, the optimization of these levers is expected to have major implications for CO<sub>2</sub> drawdown, for example, through SOC accrual along the soil profile [7,15]. Decision-support tools that can represent these management levers across contrasting ecoregions, while accurately quantifying their effects on CO<sub>2</sub> fluxes, are critical for assessing the effectiveness of grazing levers and for determining the spatial and temporal contexts in which they can deliver the most meaningful outcomes.

Process-based ecosystem models have been widely used to assess the potential outcomes of management strategies on net ecosystem CO<sub>2</sub> exchange (NEE) [16–18], providing critical insights to support decision-makers toward sustainable management practices [19]. Significant advancements have enabled the inclusion of certain grazing management levers in various models, such as continuous grazing at different levels of intensity (e.g., PaSim, DayCent, DNDC, APSIM), and, more recently, non-uniform grazing distributions associated with low or no rotational grazing management systems (MEMS 3.0) [20; under review], a critical step toward accurately representing grazing management strategies. Nevertheless, the application of such models across diverse ecoregions and management strategies remains a significant challenge. This is not only due to the difficulty of accurately representing complex animal-plant-soil-microbe interactions, but mostly because reliable model predictions require high-quality measured data for model initialization and for effective calibration, evaluation, and uncertainty estimation [10]. Ideally, independent time-series datasets should be used for calibration and evaluation, which improves the robustness of model testing [21]. In practice, however, such comprehensive datasets are often scarce, especially in grazing systems, which are inherently complex due to interactions between grazing animals and plants and their feedbacks on C dynamics. Consequently, fully integrated measurements linking biomass production, SOC dynamics, and greenhouse gas fluxes are seldom available or collected simultaneously. Recently coordinated monitoring efforts across ecoregions have been established to obtain a robust assessment of grazing outcomes on ecosystem dynamics [22; under review].

Despite major advances in process-based modeling, there is limited testing in grazing systems with atmospheric flux measurements that could provide further insights into uncertainty estimates. The eddy covariance (EC) technique is a well-established approach for quantifying temporal exchanges of CO<sub>2</sub> between land and atmosphere [23,24]. By leveraging high-frequency *in situ*

measurements, EC estimates NEE while inherently integrating animal, vegetation, and soil processes and their responses to environmental variability and management within the measurement footprint [25]. However, most models are evaluated with SOC stock changes, which are often only detectable over long measurement timeframes (i.e., several years to a decade), creating a major limitation to model verification as well as the evaluation of management change effectiveness based on SOC remeasurement protocols, hampering the implementation for soil C markets [26]. The few studies that integrated EC and modeling rely on EC fluxes primarily to assess model performance [16,17,27–29] or to support model calibration and/or subsequent evaluation [18,30–33] often without the use of fully independent datasets. However, none of them have quantified the uncertainty associated with model parameters and structure on simulating either SOC or NEE fluxes. Although this does not invalidate previous applications, overlooking model uncertainty limits the robustness and interpretability of simulated ecosystem C dynamics. Addressing this knowledge gap is critical to reducing uncertainties in model estimation of atmospheric CO<sub>2</sub> drawdown from grazing systems, which in turn enhances the reliability of model predictions of other ecosystem services (e.g., forage production, SOC storage, water dynamics, etc.), which frequently involve trade-offs relevant to management decisions.

Although process-based models are grounded in theoretical concepts and experimental evidence, they inevitably simplify reality due to incomplete representation of ecosystem processes and limited data availability. Consequently, simulations may be subject to imprecision and bias, contributing to overall model uncertainty [34]. Much of this uncertainty may arise from model parameters and structural components, which can be quantified using different methodological approaches [35–38]. Among these, Bayesian methods are widely regarded as a rigorous and transparent approach to evaluate model prediction uncertainty [39,40], especially for complex ecosystem models with numerous interacting parameters and processes [35,41]. Bayesian approaches provide a probabilistic framework in which unknown model parameters are estimated through their posterior distribution, which combines both the likelihood of the data given model parameters and the prior distribution of the model parameters [35]. The posterior distribution of the model parameters thus reflects how well the model fits the data while incorporating existing understanding of parameter values. To propagate structural uncertainty and parameter uncertainty in complex process-based models, Monte Carlo methods can be applied to quantify uncertainty in model predictions through iterative model simulation analyses [35,36,38]. The results incorporate uncertainty in model parameters and structure, providing a robust basis for deriving posterior predictive intervals in model predictions. Despite the wide range of applications of Bayesian methods across disciplines, uncertainty analysis in ecosystem modelling has remained limited to SOC simulations in croplands [34–37] and, to our knowledge, no study has yet quantified model uncertainty for SOC and NEE in grazing systems.

To address this gap, we quantified the uncertainty associated with parameters and model structure in the MEMS model version 3.0 [20; under review] by implementing a Bayesian framework using data from grassland sites in the National Ecological Observatory Network (NEON) [42]. Using a fully independent dataset from four grazing experimental sites located in Oklahoma, Michigan, and Wyoming, USA we initially tested the general model performance to predict soil temperature and soil water content, key regulators of ecosystem dynamics. Then, model evaluation was done using baseline SOC and ecosystem fluxes of treatments across grazing ecoregions by comparing model outputs with soil and EC measurements from the experimental sites. These sites span diverse ecological regions, characterized by varying vegetation types and densities, distinct climates, and a range of soil types, where livestock plays a vital role in shaping the landscape [22; under review]. We used MEMS 3.0 because it explicitly represents contrasting grazing management strategies through manipulation of key management levers and incorporates spatially non-uniform livestock distribution, which are features that are hypothesized to affect plant productivity and SOC dynamics [20; under review], and, consequently, overall NEE fluxes. Our overarching goals were to: i) assess the ability of MEMS 3.0 to represent plant dynamics across U.S. grazing ecoregions; ii) quantify the

uncertainty (i.e., parameters and structure) associated with MEMS 3.0 predictions of baseline SOC and NEE of grazing lands using a Bayesian framework; and iii) evaluate the model's performance to predict SOC in grazing lands and NEE under contrasting grazing management strategies against measurements from the experimental sites. By quantifying uncertainties in MEMS 3.0, simulating different grazing management strategies, and evaluating model predictions against independent datasets, we aim to provide a robust evidence-based approach to support researchers and decision-makers in improving C monitoring frameworks for grazing lands, which could aid in facilitating the large-scale adoption of grazing management strategies that enhance productivity while contributing to climate change mitigation.

## 2. Material and methods

### 2.1. The MEMS Model

MEMS is a process-based ecosystem model designed to simulate soil organic matter (SOM) dynamics that integrates comprehensive ecosystem functions to represent plant growth, nutrient cycling, hydrology, and grazing management across different environments and management conditions [29]. The MEMS model represents functionally distinct and quantifiable SOC pools, namely dissolved organic carbon (DOC), particulate organic carbon (POC), and mineral associated organic carbon (MAOC). It represents much of the state-of-the-art understanding of SOC dynamics [43], including the dual-pathway SOC formation framework [44], *in vivo* versus *ex vivo* microbial processing [45], and point-of-entry dynamics [46]. Decomposition products from microbial residues and plant debris contribute to distinct SOM pools, controlled by temperature sensitivity, mineral nitrogen (N) availability, substrate C/N ratios, and water availability, reflecting site-specific biogeochemical conditions [29]. The MEMS model incorporates soil C and N cycling across a multi-layered soil profile, where DOC and mineral N move vertically along the profile, and roots contribute with fresh inputs at depth. The model captures microbial-driven N cycling processes (immobilization and mineralization) that create feedback loops influencing plant growth and SOM dynamics. Soil temperature and moisture content are modeled daily across soil layers, which allows for high-resolution representation of ecosystem responses to environmental and management changes.

The MEMS model operates on a daily timestep and requires daily weather data, soil properties, plant characteristics, and management practices as key inputs [29]. It has the capability to simulate cropping systems [47] and perennial plants [48]. Plant phenology is controlled by accumulation of thermal time and adjusted by daylength. The model dynamically adjusts green-up and dormancy for perennial plants based on soil temperature, soil water, and daylength and dynamically calculates non-structural reserves stored in the reserve organs for green up and regrowth after defoliation. In addition, reserve organs maintain a balance between minimum and maximum N content, which is translocated between the reserve organs (post-seeding and prior to dormancy) and other plant organs to support regrowth and stress conditions. Standing dead biomass is also modeled for its shading effect which impacts light interception and plant production [49]. For grazing systems, the MEMS 3.0 model [20; under review] includes plant-animal interactions to simulate the non-uniform distribution of grazing animals using a spatial explicit structure. Grazing pastures are divided into spatial units, representing discrete areas where daily animal distribution is based on a habitat suitability index (HSI) that encompasses forage quantity and quality, water proximity, shading, and physical accessibility. It also provides an option to automatically move animals with a high rotation frequency (daily in small paddocks) resulting in a more uniform grazing distribution.

### 2.2. Data Used for Modeling

Measurements from six grassland sites within the NEON monitoring network [42] (Figure 1) and from four experimental sites in the 3M project (Metrics, Management, and Monitoring; fully described in Stanley et al., [22; under review] were used in our study. Together, these datasets represent the most comprehensive information currently available on SOC and NEE measurements

for grazing lands across the U.S. Although the datasets do not share the same set of variables, except for SOC and NEE, they were used because they are fully independent, ensuring a more robust model calibration and evaluation framework.

### 2.3. NEON Sites

We used data from the NEON grassland sites [Central Plains Experimental Range (CPER), Dakota Coteau Field Site (DCFS), Konza Prairie Biological Station (KONZ), Northern Great Plains Research Laboratory (NOGP), Oklahoma Agricultural Experiment Station (OAES), and Woodworth (WOOD)] for model calibration. All of these sites have been managed under cattle grazing and are instrumented with EC flux towers that continuously monitor ecosystem-atmosphere gas exchanges. To ensure high-quality model inputs, we prioritized the most reliable and consistent available data from these sites, which included MODIS gross primary productivity (GPP) estimates [50], measurements of SOC fractions (POC and MAOC) from Zhang et al. [29] along with bulk SOC and EC flux measurements from NEON [51]. Daily EC flux data were downloaded from the AmeriFlux portal [51–53] and processed by removing years with substantial missing data prior to calculating annual cumulative values. As part of data quality control, we excluded years in which the EC flux data showed unusually high and continuous GPP fluxes during winter, i.e., when plants are not photosynthetically active after confirming these anomalies using MODIS GPP data.

### 2.4. 3M Experimental Sites

The 3M project is a large collaborative grazing project aimed at assessing the ecosystem outcomes of grazing management across various U.S. grazing ecoregions. It encompasses experimental sites located in Oklahoma (OK-CR and OK-RR), Michigan (MI-LC), and Wyoming (WY-MR) [22; under review] (Figure 1). Each experimental site is constructed in a randomized block design, where each replicate block is split into two treatment plots: prescriptive (PR) and adaptive (AD) grazing management. Treatments were designed to reflect ecological outcomes of fixed PR grazing management versus flexible AD rotational grazing management. Each site had an EC flux tower installed within each treatment plot collecting data continuously. Experimental site characteristics are reported in Table 1, while a more detailed description including historical management is reported in the supplementary material. Although measurements are ongoing, this analysis of the MEMS 3.0 model includes data collected from 2022, when the experiment was started, through 2024. The data from the 3M project was used for testing key regulators of ecosystem dynamics and model evaluation.

**Table 1.** Site information for prescriptive (PR) and adaptive (AD) grazing management across experimental sites located in Oklahoma (OK-CR and OK-RR), Michigan (MI-LC), and Wyoming (WY-MR).

Site name	Location	Altitude (m)	Koppen's climate classification	Dominant plant	# of blocks	Grazing management	Plot size (ha)	Paddock size (ha)*	Average plot resting period (day)	Length of grazing event (day)	Stocking rate (AU ha <sup>-1</sup> )
OK-CR	Oklahoma - Coffey Ranch	250	Humid subtropical (Cfa)	Warm-season grasses (e.g., <i>Schizachyrium scoparium</i> ), but it also includes legumes and shrubs	3	PR	11.5	-	104	1-13	1.9-6.3
						AD	7.4	0.4-1.5	170	5-17	5-8.4
OK-RR	Oklahoma - Red River Grazing Facility	226	Humid subtropical (Cfa)	Warm-season grasses, particularly bermudagrass ( <i>Cynodon dactylon</i> ), with some presence of annuals and forbs	3	PR	9.7	-	76	2-11	4.3
						AD	9.7	0.97-1.94	178	1	20.5-46.7
MI-LC	Michigan - Lake City Research Center	340	Humid continental (Dfb)	Cool-season grasses (e.g., <i>Dactylis glomerata</i> , <i>Poa pratensis</i> , <i>Bromus inermis</i> )	3	PR	4	-	81	2-16	7.8
						AD	4	0.2-2	153	0.2-1	64-425
WY-MR	Wyoming - McGuire Ranch	2190	Arid cold steppe (BSk)	Cool-season grasses (e.g., <i>Pascopyrum smithii</i> , <i>Koeleria macrantha</i> , <i>Poa fendleriana</i> ) and sagebrush	5	PR	38.8-148.5	-	308	70-77	0.3
						AD	38.8-148.5	-	353	8-21	0.5-3.5

\*PR grazing management plots were treated as one large paddock. AU: animal unit based on 500 kg (Allen et al., 2011).

#### 2.4.1. Flux Tower Set-up and Measurements

Each treatment plot across the four experimental sites was equipped by Quanterra Systems Ltd, (UK) with standardized instrumentation: a sonic anemometer (Gill Windmaster), gas sampling enclosure (Quanterra 2301XX), weather station (Vaisala WXT535), four-component radiometer (Hukseflux NR01), soil probe (Meter TEROS11), and soil heat flux plate (Hukseflux HFP01) to monitor ecosystem fluxes, as well as atmospheric and soil climatic conditions. A tower was placed within a representative area of each plot chosen to minimize interference from adjacent non-target land cover types. To inform optimal placement, a flux footprint model [54] driven by ERA5 reanalysis data [55] was employed to ensure the flux source area primarily reflected the managed treatment area. The same suite of sensors was deployed across all sites. Taller towers (5.5 m) were installed at OK-CR, OK-RR, and WY-MR sites, where plots are larger, while shorter towers (3-3.7 m) were installed in MI-LC to accommodate smaller plots. Towers were designed to minimize obstruction of the radiometer's field of view. Radiometers were positioned for optimal southern exposure, and anemometers were aligned to true north. Each tower was powered by a single solar panel positioned 10-15 m east or west of the tower to minimize airflow disturbance. Soil sensors (10 cm depth) were installed 2-3 m south of the tower to measure soil temperature, soil heat flux and water content. Sensors were calibrated as needed. Wind and gas data were collected at 10 Hz to calculate NEE fluxes, while other variables were measured at lower frequencies and aggregated accordingly. Negative NEE values, i.e., obtained when ecosystem CO<sub>2</sub> uptake exceeds CO<sub>2</sub> efflux, represent net C drawdown from the atmosphere into the ecosystem. Gross primary productivity and ecosystem respiration (ER) were partitioned from NEE. For detailed methodology and data processing description, please see Stanley et al. [22; under review].

#### 2.4.2. Plant Aboveground Biomass and Leaf Area Index Measured Data

Measurements of plant biomass and leaf area index (LAI) were conducted every 28 days throughout the growing season within a 100 × 100 m grid, that was randomly established in each plot within the fetch area of the EC flux tower. The grid was divided into five transects (100 m long) spaced 20 meters apart. Biomass clippings and LAI measurements (n = 6) were collected at 5 m and 55 m along three of these transects. At each sampling, LAI was measured using a LI-COR LAI 2200, followed by aboveground biomass clipping at ground level within a 0.25 m<sup>2</sup> circular hoop, which included both new and residual material. All plant material was oven-dried at 60 °C for 72 hours and weighed to determine dry biomass. The period of data collection varied among sites due to differences in experiment start dates. At WY-LC, LAI was not measured because of low vegetation cover.

#### 2.4.3. Soil Sampling and Analysis

Baseline soil samples were collected in 2022, following a triangle design [56] with 12 equilateral (1-m side) triangles per experimental plot, distributed within the fetch area of the EC tower. A hydraulic coring system (in OK-CR, OK-RR, and WY-MR; Giddings Machine Company) or percussion core (in MI-LC; AMS Inc.) were used to extract three 1-m depth cores, one per each vertex of the triangle, which were sectioned into 0-15, 15-30, 30-50, and 50-100 cm intervals and composited by triangle and depths. Soil samples were sieved first through an 8-mm and then through a 2-mm mesh to remove coarse materials such as rocks, roots, and litter. These samples were oven-dried, weighed, and quantified for correction of soil bulk density, which was determined based on soil volume and mass at each depth interval following Poeplau et al. [57]. The 2-mm soil samples were characterized for texture and pH, and fractionated by size and density [for the quantification of POC (>1.85 g cm<sup>-3</sup>), coarse heavy associated organic carbon (CHAOC; <1.85 g cm<sup>-3</sup> and >53 μm), and MAOC (<1.85 g cm<sup>-3</sup> and <53 μm)], following Leuthould et al. [58].

Bulk soils and physical fractions were oven-dried, ground (<0.149 mm), and analyzed for total C by dry combustion using an elemental analyzer (EA; vario Isotope Cube, Elementar). Because the MEMS 3.0 model does not simulate the CHAOC fraction and studies suggest it is more similar to MAOC than POC [58,59], we combined CHAOC with MAOC for evaluating the modeled MAOC estimates [29,60].

#### 2.4.4. Soil Water Content Measurements

Soil water content (SWC) measurements were carried out using the TEROS 10 (METER Group, Pullman, WA) capacity sensors with factory calibration. The sensors were installed in a subset of three out of the 12 triangles. The selected triangles represented the range of topographic positions and vegetation cover at each site, ensuring that measurements captured local spatial variability. Data were recorded hourly at depths of 15, 40, and 60 cm in OK-CR, OK-RR, and MI-LC, and at 10, 20, and 30 cm in WY-MR [22; under review].

## 2.5. Plant Parameterization

Plant parameterization for the NEON sites was based on the initial parameter sets developed by Zhang et al. [29] using the MEMS 2.0 model. However, due to structural updates in the current model version, we re-parameterized dominant grass functional types (associated with different grass communities at each site). Plant parameters were either set to measurements from individual sites, including plant physical and chemical properties [61], or manually adjusted based on ranges reported in the literature to reproduce MODIS GPP estimates for each site.

For the experimental sites, we parameterized grasses for each site as the plant communities were different from each other. Based on the measured plant community composition [22; under review], the most representative species were identified for each location (Table 1). Literature derived values were used for some parameters while the others were manually adjusted by comparing modeled and measured plant aboveground biomass and LAI, where available.

## 2.6. Bayesian Framework: Model Calibration and Uncertainty Quantification

Measurements of POC, MAOC, and EC flux measurements from the six NEON grasslands sites were used for model calibration and quantification of uncertainty. Based on the global sensitivity analysis performed by Zhang et al. [29] using the NEON grassland sites, five parameters governing SOM decomposition, and therefore regulating C losses and storage, were identified as the most influential. These parameters included the potential depolymerization rate of particulate organic matter ( $k_{\text{POMDeplo}}^{\text{mer}}$ ), the temperature sensitivity coefficient for decomposition ( $\text{Coeff\_TemperatureEffOnDecomp}_1$ ), the exchangeable fraction of mineral-associated organic matter ( $\text{Frac\_MAOMExchangeable}$ ), the decay rate constant of mineral-associated organic matter ( $k_{\text{MAOMDecay}}$ ), and the sorption rate constant of dissolved organic matter ( $k_{\text{DOMSorp}}$ ). These five parameters were selected for Bayesian calibration, and their prior distributions were adopted from Zhang et al. [29]. Plant parameters were excluded from the Bayesian calibration for simplicity, as our primary objective was to constrain SOC-related parameters. We assumed that the plant parameters were sufficient to represent key vegetation features in the model. Consequently, plant parameters were calibrated separately prior to the uncertainty analysis.

To calibrate the selected parameters, we implemented a Differential Evolution Adaptive Metropolis (DREAM) algorithm similar to that described in Mathers et al. [36] and Zhang et al. [38]. The DREAM algorithm is a Markov chain Monte Carlo (MCMC) approach that uses multiple chains and adaptive sampling to explore complex high-dimensional parameter spaces [62]. The DREAM algorithm utilizes prior distributions on the parameter space and log likelihood values from the observed and simulated data to propose new parameter sets at each iteration. The likelihood functions were constructed to quantify the agreement between simulated and observed values of POC, MAOC, and annual NEE. As the NEON soil cores were divided into horizons of varying thickness, the measured POC and MAOC were converted to volumetric units ( $\text{kg C m}^{-3}$ ) to ensure consistency across depth intervals. Also, because these two fractions are correlated, we developed a combined likelihood function to jointly represent their covariance structure. This joint likelihood was formulated as a multivariate mixed-effects model with soil depth treated as a random effect, implemented using the “lme4” R package [63]. For annual NEE, a mixed-effects model was used with measurement year as a random effect. Assuming independence between annual NEE and the joint distribution of POC and MAOC, the overall log-likelihood for the Bayesian calibration was then computed as the sum of the log-likelihood of NEE and the joint log-likelihood of POC and MAOC. Model error (predicted value minus measured) of POC and MAOC was partitioned into variance between soil depths ( $\sigma^2_{\text{depth\_POC}}$  and  $\sigma^2_{\text{depth\_MAOC}}$ ) and unexplained residual variance ( $\sigma^2_{\text{resid\_SOC}}$ ). Similarly, model error of annual NEE was partitioned into variance between measurement years ( $\sigma^2_{\text{year\_NEE}}$ ) and unexplained residual variance ( $\sigma^2_{\text{resid\_NEE}}$ ).

The DREAM algorithm was executed using the R package “DREAM” [64] in R version 4.4.1 [65]. Eight chains were run with a maximum of 5,000 iterations per chain. Convergence of the Markov chains was evaluated using the Gelman-Rubin diagnostic statistic (R) [66], with convergence

assumed when  $R \leq 1.2$ , following Zhang et al. [38]. To ensure that parameter sampling reflected the stationary posterior distribution, the first 50% of iterations from each chain were discarded as burn-in, consistent with the recommendations of Vrugt et al. [62].

The last 200 parameter sets and corresponding hyperparameters (representing model error variances) from the DREAM posterior samples were used to quantify predictive uncertainty through Monte Carlo simulation following the approach of Mathers et al. [36]. This procedure propagates uncertainties in both the calibrated parameters and the hyperparameters [35], thereby accounting for variability arising from model parameters and imperfect representation of processes in its structure. For each parameter set, model simulations were conducted to generate distributions of predicted variables, from which the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles were extracted to represent the 95% posterior predictive interval for SOC, POC, MAOC, and NEE outputs. Predictions based on the single best-performing parameter set (i.e., the maximum posterior likelihood) were also produced and are presented as the model's best estimates for all output variables and used to generate goodness-of-fit statistics.

## 2.7. Model Set-up

### 2.7.1. NEON Sites

Simulations for the NEON sites were conducted using site specific data from Zhang et al. [29]. Soil chemical and physical attributes measured at each site were used as model inputs [61,67]. Weather data was obtained from the gridMET database [68]. Sites were assumed to be at steady state and a 1000-year spin-up was run to reach equilibrium assuming site measured SOC [61] as the reference condition. Historical fire events were scheduled according to the frequency reported in Guyette et al. [69] with each event assumed to be in the spring. Grasslands were subjected to grazing designed to mimic the impact of wild herbivores [70], during the spin-up period. Following spin-up, simulations were extended to match the temporal coverage of MODIS GPP data for each site. During this period, prescribed fire was scheduled following site-specific experimental records [71]. Because grazing management information specific to the EC tower footprint was not available, we applied the general site level grazing management (personal communication with NEON Field Science Manager Andrea C. E. Anteau) in our simulations.

### 2.7.2. 3M Experimental Sites

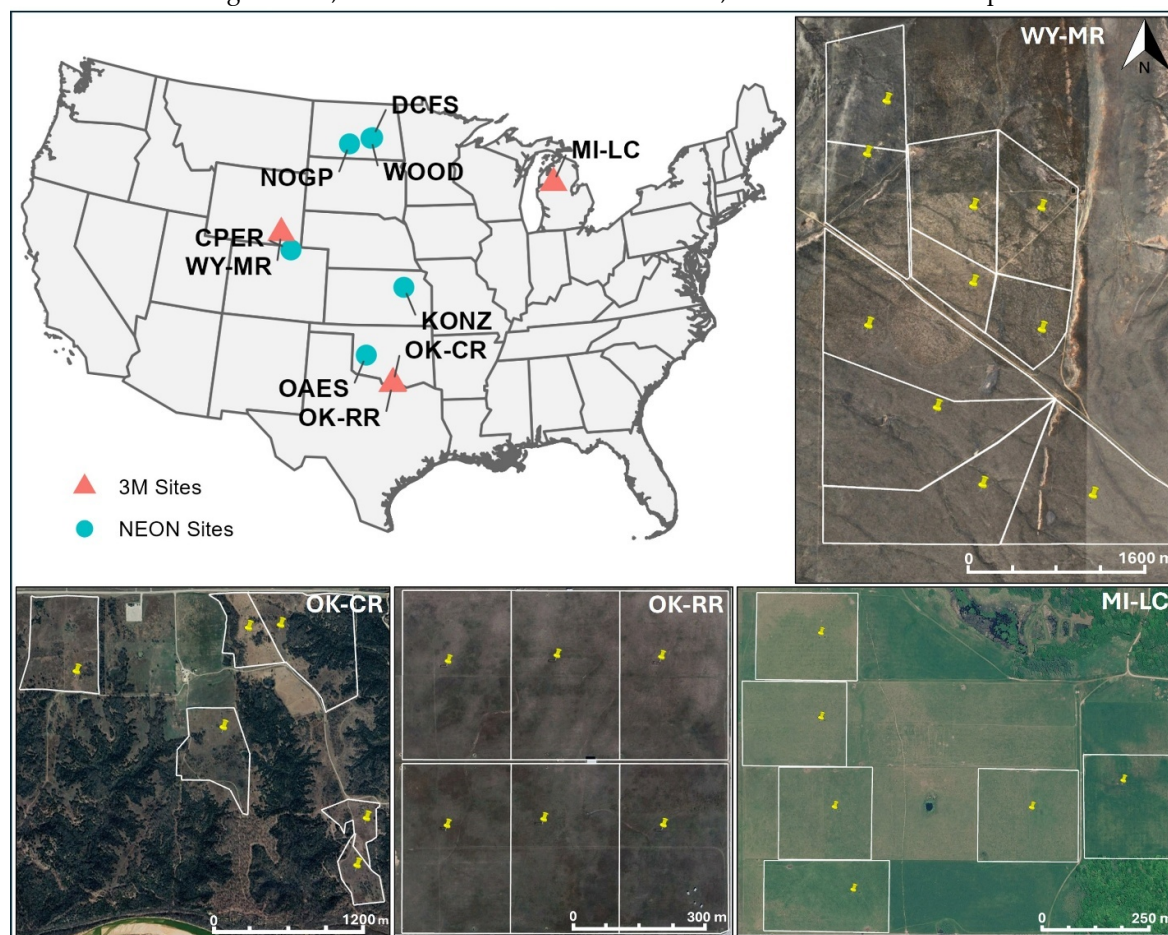
For the 3M experimental sites, model simulations were conducted for individual treatment plots at each site. Model soil inputs were based on measured soil data averaged at the plot level, along with soil hydraulic parameters estimated for each soil layer using the ROSETTA pedotransfer model [72,73]. Weather data for the pre-experimental period were obtained from the gridMET database [68]. For the experimental period, weather data collected by the EC flux towers were averaged to represent local climate conditions of each site, except WY-MR, where environmental contamination of rain gauge signals prevented reliable precipitation measurements. For this site, data from the gridMET database were used.

The simulations were conducted for three periods: non-managed native vegetation, spin-up of historical management, and the experimental period. The model was initialized by reaching a steady-state equilibrium in SOC stocks based on the native grasslands at each site through a 4000-year simulation. In this simulation, grazing intensity was set to mimic the impact of wild herbivores [70], and fire events were incorporated at frequencies specific to each ecological region [69]. For the pre-experiment spin-up period, long-term management information collected at each experimental site was used to construct the management input files. For the experimental period, specific details of the PR and AD grazing management, including the number of paddocks used for AD grazing management, stocking densities, and the dates of animal movement in and out of plots, were documented and used to define model inputs.

The MEMS 3.0 model is spatially explicit and divides the landscape into spatial units. The method developed in Zhang et al. [20; under review] was used to spatially divide the field boundary polygons into hexagons. To balance paddock size and computational efficiency, we allocated the spatial unit sizes as approximately 0.25, 0.36, 0.40, and 3.64 ha for OK-CR, OK-RR, MI-LC, and WY-MR, respectively (10 to 52 spatial units across the simulated sites). The minimal distance to drinking water source was calculated for each hexagon and used to estimate the HSI following Zhang et al. [20; under review]. This approach was based on animal movement predictions generated using a Lévy-walk model informed by GPS-collar data from Romero-Ruiz et al. [74]. As the MI-LC plots were relatively small (~4 ha), the distance to water was assumed to have a small effect on movement, and so the water index was set to 1.0 for all spatial units. The indexes for shading effect were derived using information from satellite imagery of Google Map [75] and consultations with site managers. Slope effects were not considered as elevation gradients across spatial units within each site were similar.

For consistent comparison between simulated and measured fluxes across sites, spatial units corresponding to the fetch area and exact placement of each EC flux tower (Figure 1) were selected at every experimental site.

d: index of agreement;  $R^2$ : coefficient of determination; RMSE: root mean square error.



**Figure 1.** Location of sites used for model calibration [National Ecological Observatory Network (NEON) sites: CPER, DCFS, KONZ, NOGP, OAES, and WOOD] and for model evaluation (3M sites: OK-CR, OK-RR, MI-LC, and WY-MR). White outlines delineate the block-level pasture areas, and pins indicate the locations of the eddy covariance flux towers at each plot across experimental sites. Satellite imagery from Google Earth Pro (© Google) accessed in November 2025.

## 2.8. Statistical Analysis

Model calibration and performance were evaluated using the following statistical metrics: BIAS,

$$BIAS = (\bar{P} - \bar{O})$$

index of agreement (d),

$$d = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2}$$

coefficient of determination ( $R^2$ ),

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - \hat{P}_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$$

and root mean square error (RMSE),

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}$$

where  $P$  refers to a predicted value  $i$ ,  $O$  refers to an observed value  $i$ ,  $n$  is the number of observations,  $\bar{P}$  is the mean of predicted values,  $\hat{P}_i$  refers to a fitted value  $i$  obtained from the linear regression between observed and predicted data, and  $\bar{O}$  represents the mean of observed values. BIAS measures the mean tendency of simulated values to be larger or smaller than observations; values near zero indicate little systematic error;  $d$  quantifies prediction accuracy relative to observed variability and ranges from 0 to 1, with values closer to 1 indicating excellent agreement between model predictions and observations, and closer to 0 indicates no agreement at all.  $R^2$  expresses the proportion of variance in the observed data explained by a linear relationship with the predictions and varies from 0 to 1, where larger values indicate stronger association; and RMSE summarizes the average magnitude of errors, with lower values indicating more accurate simulations. Statistical metrics were calculated using daily data for soil temperature and SWC, and weekly data for NEE, GPP, and ER.

### 3. Results

We first evaluate plant parameterization and vegetation dynamics, followed by model performance for soil temperature and soil water content. We then present the Bayesian calibration results and subsequently provide an independent evaluation of SOC dynamics and ecosystem fluxes at the 3M experimental sites.

#### 3.1. Plant Parameterization for the NEON and 3M Experimental Sites

Plant parameterization for the NEON grassland sites was evaluated through comparisons between simulated and MODIS-derived GPP. Across the NEON sites and multiple years, MEMS closely reproduced the seasonal dynamics and magnitude of MODIS GPP (Figure S1), as indicated by high agreement metrics (d: 0.84-0.95; R<sup>2</sup>: 0.67-0.86; Table 2). The timing of peak GPP was generally well aligned between simulated and MODIS for most sites. Nevertheless, year-specific deviations were observed, particularly at DCFS (BIAS = -6.49 g C m<sup>-2</sup> 8-day sum) and NOGP (BIAS = -6.54 g C m<sup>-2</sup> 8-day sum), where larger deviations from MODIS GPP occurred. Overall, MEMS consistently captured spatial and interannual variability in GPP across NEON grassland sites, with comparable RMSE values (5.46-11.17 g C m<sup>-2</sup> 8-day sum) among sites, indicating stable model performance.

**Table 2.** Statistical metrics used to evaluate plant parameterization for NEON sites (CPER, DCFS, KONZ, NOGP, OAES, WOOD) using gross primary productivity (GPP) and for 3M project experimental sites (OK-CR, OK-RR, MI-LC, WY-MR) using aboveground biomass (ABG) and leaf area index (LAI).

Dataset	Site name	Variable	Statistical metric			
			d	R <sup>2</sup>	BIAS	RMSE
NEON	CPER	GPP (g C m <sup>-2</sup> 8-day sum)	0.89	0.70	-2.62	5.46
	DCFS	GPP (g C m <sup>-2</sup> 8-day sum)	0.91	0.86	-6.49	11.17
	KONZ	GPP (g C m <sup>-2</sup> 8-day sum)	0.95	0.86	2.18	9.94
	NOGP	GPP (g C m <sup>-2</sup> 8-day sum)	0.84	0.79	-6.54	9.93
	OAES	GPP (g C m <sup>-2</sup> 8-day sum)	0.85	0.67	3.82	9.78
	WOOD	GPP (g C m <sup>-2</sup> 8-day sum)	0.95	0.82	-1.82	9.03
3M project	OK-CR	ABG (Mg ha <sup>-1</sup> )	0.67	0.21	0.17	0.99
		LAI (m <sup>2</sup> m <sup>-2</sup> )	0.79	0.41	-0.08	0.69
	OK-RR	ABG (Mg ha <sup>-1</sup> )	0.64	0.13	0.07	0.71
		LAI (m <sup>2</sup> m <sup>-2</sup> )	0.63	0.14	0.09	0.53
	MI-LC	ABG (Mg ha <sup>-1</sup> )	0.59	0.29	0.88	1.61
		LAI (m <sup>2</sup> m <sup>-2</sup> )	0.58	0.12	-0.39	1.75
	WY-MR	ABG (Mg ha <sup>-1</sup> )	0.50	0.05	0.04	0.15
		LAI (m <sup>2</sup> m <sup>-2</sup> )	-	-	-	-

d: index of agreement; R<sup>2</sup>: coefficient of determination; RMSE: root mean square error.

For the 3M experimental sites, plant parameterization was based on site-level measurements of the plant community, including aboveground biomass and LAI (Table 2). Simulated and observed aboveground biomass matched relatively well across sites (d: 0.50-0.67) with the model capturing mean values adequately but showing a slight tendency to overestimate observations (BIAS: 0.04-0.88 Mg ha<sup>-1</sup>). However, the model did not reproduce the observed variability in data across sites as indicated by the low values of R<sup>2</sup> (0.05-0.29). A similar pattern was observed for LAI (Table 2), though

performance was slightly better relative to aboveground biomass ( $d = 0.58-0.79$ ;  $R^2 = 0.12-0.41$ ), with stronger agreement at OK-CR and OK-RR, and the weakest performance at WY-MR.

### 3.2. Model Performance on Soil Temperature and Soil Water Content at the 3M Experimental Sites

To test the general performance of the model, we compared simulated against observed soil temperature and SWC (Table 3). The model consistently captured soil temperature changes over the measured period for all sites, as supported by the  $d$  values  $> 0.94$  and high  $R^2$  (0.93-0.95) (Table 3). However, the model showed a slight tendency to underestimate soil temperature (BIAS: ranged from -3.84 to -1.57 °C), especially at OK-RR, where the lowest temperature was observed (Table 2).

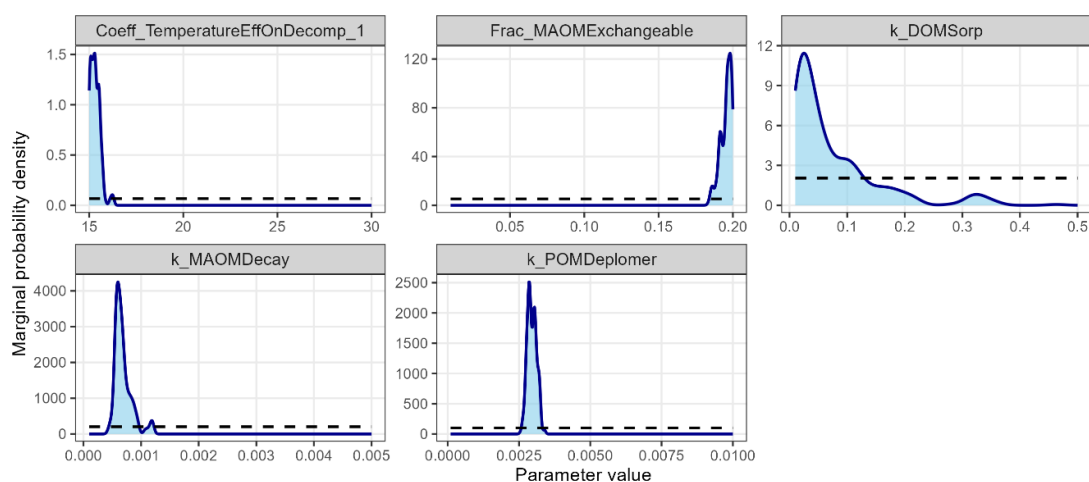
**Table 3.** Statistical metrics of data used for testing general model performance [soil temperature at 10 cm of depth, soil water content (SWC) at 15, 40 and 60 cm of depth], and model evaluation [weekly sum of net exchange ecosystem (NEE), gross primary productivity (GPP), and ecosystem respiration (ER)] across experimental sites located in Oklahoma (OK-CR and OK-RR), Michigan (MI-LC), and Wyoming (WY-MR).

Location	Statistical metric	General model performance				Model evaluation					
		Soil temperature <sub>10</sub> cm	SWC <sub>15</sub>	SWC <sub>40</sub>	SWC <sub>60</sub>	SOC	POC	MAOC	NEE	GPP	ER
			cm	cm	cm						
			°C	cm <sup>3</sup> cm <sup>-3</sup>							
OK-CR	$d$	0.95	0.90	0.92	0.91	0.88	0.90	0.47	0.80	0.94	0.94
	$R^2$	0.95	0.83	0.74	0.71	0.98	1.00	0.83	0.47	0.82	0.80
	BIAS	-2.76	0.03	0.01	-0.01	3.54	-1.48	6.01	1.82	-2.48	-0.66
	RMSE	3.72	0.06	0.04	0.04	3.87	1.54	6.35	7.57	10.23	7.11
OK-RR	$d$	0.94	0.86	0.86	0.86	0.96	0.65	0.25	0.68	0.80	0.73
	$R^2$	0.93	0.78	0.79	0.74	0.88	0.97	0.50	0.25	0.52	0.65
	BIAS	-3.84	-0.04	-0.03	-0.01	0.33	-1.71	2.66	-1.60	-5.76	-7.36
	RMSE	4.53	0.05	0.05	0.05	0.79	1.80	2.85	7.71	11.80	9.67
MI-LC	$d$	0.95	0.76	0.83	0.87	0.96	0.80	0.68	0.79	0.86	0.84
	$R^2$	0.94	0.54	0.54	0.59	0.95	0.86	0.88	0.54	0.68	0.79
	BIAS	-2.92	0.04	0.02	-0.01	2.90	1.30	4.39	6.51	3.02	9.54
	RMSE	3.52	0.06	0.05	0.05	3.75	4.60	4.94	13.51	18.80	15.81
WY-MR	$d$	0.98	0.60	0.66	0.77	0.79	0.42	0.72	0.79	0.93	0.86
	$R^2$	0.94	0.26	0.38	0.48	0.95	0.95	0.81	0.59	0.78	0.70
	BIAS	-1.57	-0.05	-0.05	-0.02	-1.53	-3.29	0.51	1.91	-0.45	1.46
	RMSE	2.83	0.07	0.06	0.04	3.14	3.41	2.20	3.82	3.70	3.08

The model's performance in simulating SWC had great variability across sites and depths (Table 3). Soil water content was generally slightly overestimated at 15 and 40 cm in OK-CR and MI-LC (BIAS: 0.01-0.04  $\text{cm}^3 \text{cm}^{-3}$ ), whereas it was slightly underestimated in OK-RR and WY-MR (BIAS ranged from -0.05 to -0.03  $\text{cm}^3 \text{cm}^{-3}$ ). The high values of  $d > 0.76$  for OK-CR, OK-RR, and MI-LC suggest the model performs well at these sites, compared to WY-MR, especially at surface soil layers ( $d$ : 0.60-0.66). At deeper layer, SWC estimates were closer to observations, with small biases observed across all sites (BIAS: -0.02 to 0.01  $\text{cm}^3 \text{cm}^{-3}$ ).

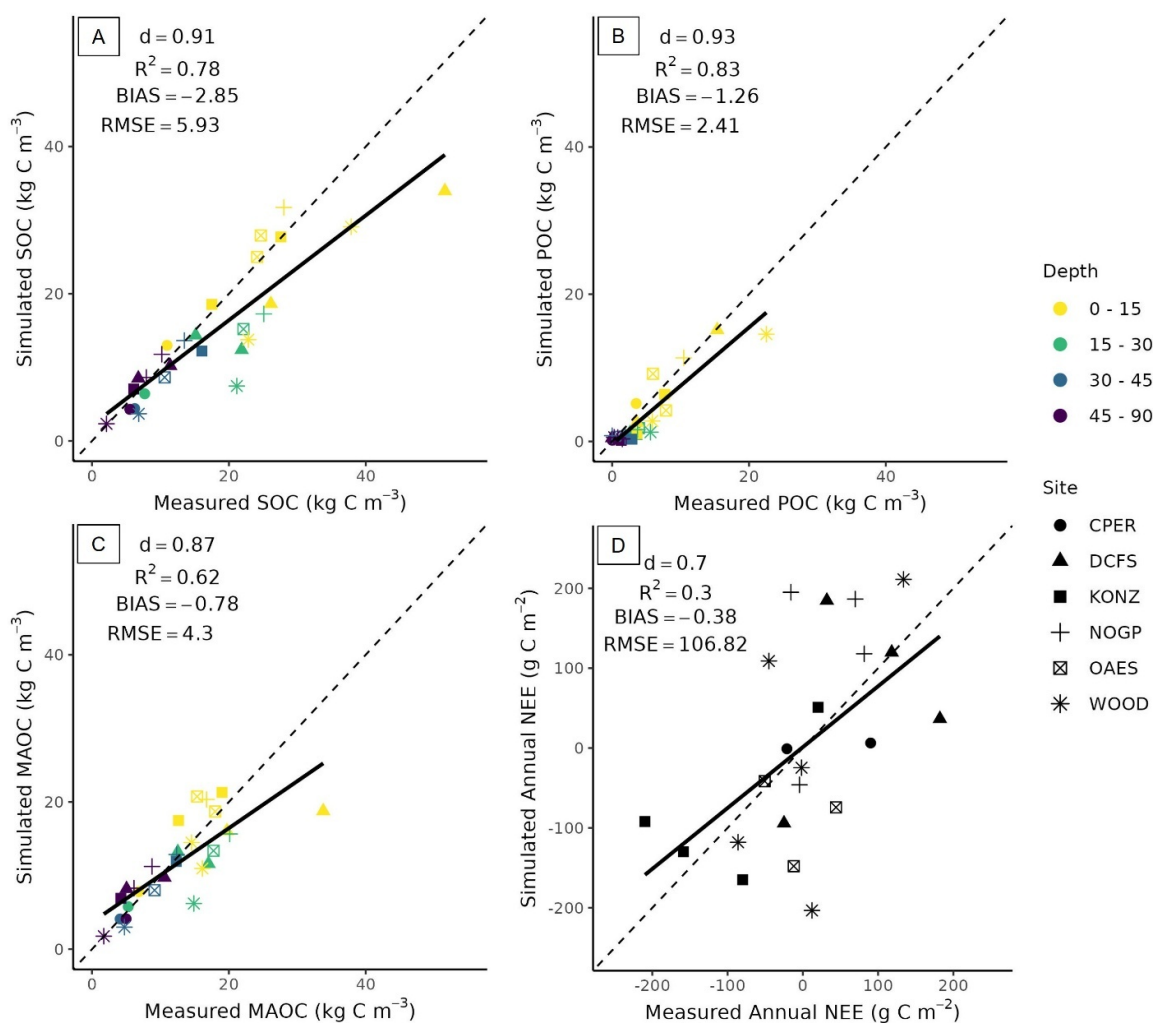
### 3.3. Bayesian Calibration Using the NEON Sites

Bayesian calibration substantially constrained most of the parameters (Figure 2), with posterior distributions markedly narrower than the priors. However, some parameters were concentrated close to the boundaries of the ranges. The `Coeff_TemperatureEffOnDecomp_1` was strongly constrained toward lower values (from 15.0 to 16  $^{\circ}\text{C}$ ), with most posterior density concentrated near the lower bound of the prior with peak at 15.3  $^{\circ}\text{C}$ . A similar trend was observed for `k_MAOMDecay`, which showed a narrow posterior (0.0004 to 0.0012) centered near 0.0006. In contrast, `Frac_MAOMExchangeable` converged near the upper bound of the prior range. The `k_POMDeplomer` exhibited a narrow posterior centered on a well-defined range (0.0029 to 0.0033), whereas `k_DOMSorp` remained comparatively weakly constrained, exhibiting substantial posterior uncertainty (0.0101 to 0.4634).



**Figure 2.** Posterior density (light blue area) and uniform priors (dashed black line) of the parameters calibrated for MEMS 3.0.

Overall, the model showed strong agreement with measurements for SOC dynamics across soil depths, as reflected in the high values of  $d$  (0.83-0.91; Figure 3A-C). Additionally, the model explained a substantial portion of the variance in observed values ( $R^2 = 0.62-0.83$ ) for bulk SOC, POC, and MAOC, though it tended to underestimate these variables (BIAS ranged from -0.78 to -2.85  $\text{kg C m}^{-3}$ ). For bulk SOC, agreement between observed and simulated values was particularly strong ( $d = 0.91$ ), although the model tended to underestimate values above  $\sim 38 \text{ kg C m}^{-3}$ , especially in surface layers at DCFs and WOOD (Figure 3A). Across SOM fractions, POC exhibited the best performance ( $d = 0.93$ ;  $R^2 = 0.83$ ;  $\text{RMSE} = 2.41 \text{ kg C m}^{-3}$ ), while MAOC showed slightly lower predictive accuracy ( $R^2 = 0.62$ ) (Figure 3B-C). Similarly to bulk SOC, underestimation in POC and MAOC was mainly driven by surface soil layers at WOOD and DCFs, respectively. For annual NEE, although the model reproduced the average magnitude reasonably well ( $d = 0.70$  and  $\text{BIAS} = 0.34 \text{ g C m}^{-2}$ ), annual prediction errors remained substantial. The model explained little of the observed interannual variability ( $R^2 = 0.30$ ), resulting in a relatively high RMSE (106.82  $\text{g C m}^{-2}$ ) (Figure 3D).

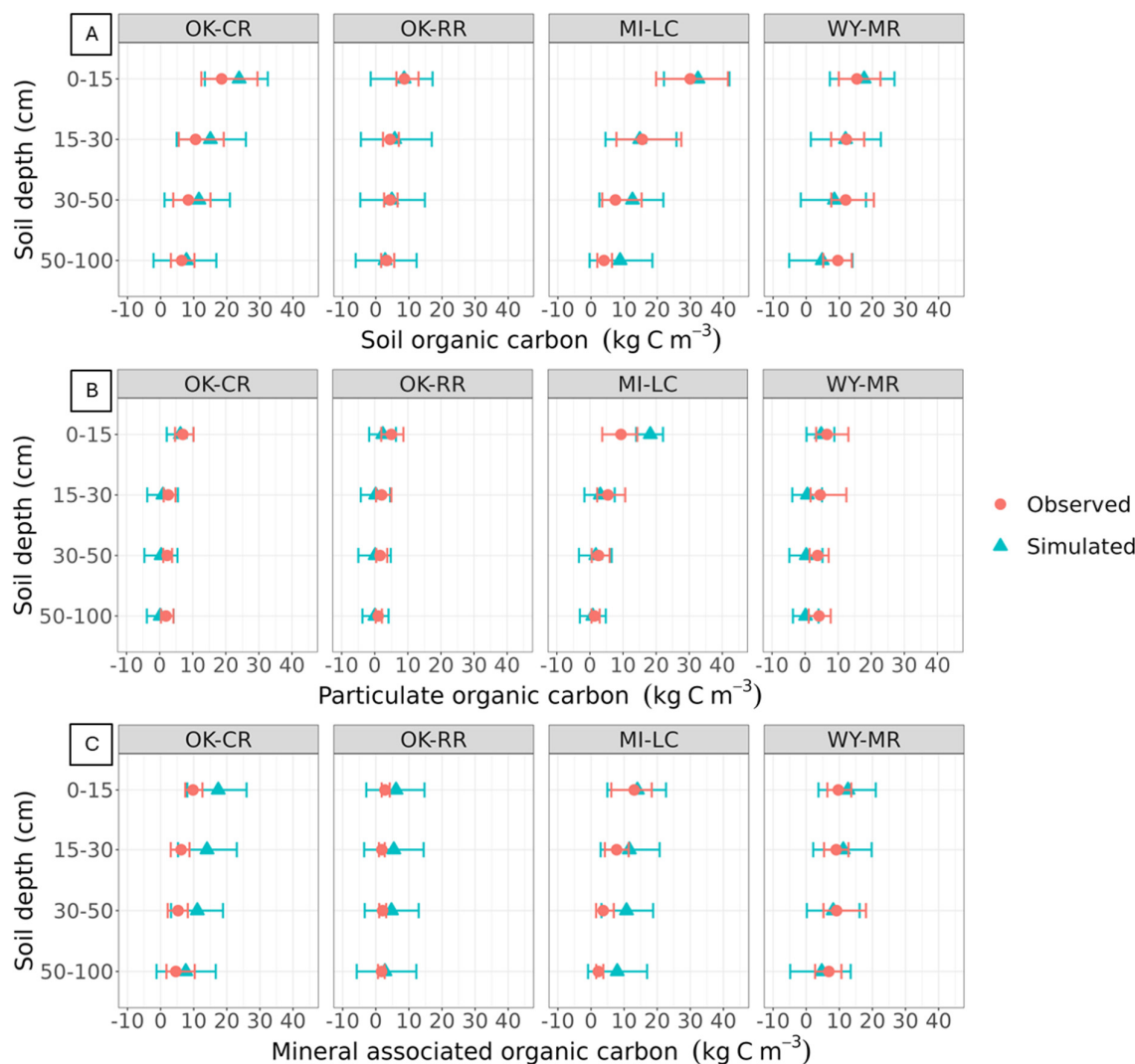


**Figure 3.** Simulated and observed values of bulk soil organic carbon (SOC; A), particulate organic carbon (POC; B), mineral-associated organic carbon (MAOC; C), and annual cumulative net ecosystem exchange (NEE; D) across six NEON sites (CPER, DCFS, KONZ, NOGP, OAES, and WOOD) used for model calibration.

### 3.4. Model Evaluation

#### 3.4.1. Soil Organic Carbon Dynamics at the 3M Experimental Sites

The model demonstrated high accuracy in simulating SOC dynamics along the soil profile across the experimental sites (Figure 4). For most of the sites, simulated SOC values strongly agreed with measured data, often falling within the 2.5<sup>th</sup>-97.5<sup>th</sup> percentile ranges of measured values (Figure 4A). This overall consistency is further supported by the high values of  $d$  (0.79-0.96) and  $R^2$  (0.88-0.98), and low values of BIAS and RMSE across most sites (Table 3). Nevertheless, SOC stocks were overestimated at OK-CR (BIAS: 3.54 kg C m<sup>-3</sup>) and MI-LC (BIAS: 2.90 kg C m<sup>-3</sup>), whereas they were underestimated at WY-MR (BIAS: -1.53 kg C m<sup>-3</sup>). The largest discrepancies occurred at the soil surface (<30 cm) at OK-CR, with simulated SOC approximately 1.4 times greater than measured values, in addition to simulations overestimating SOC in deeper soil layers by about twofold at MI-LC, and underestimating stocks in deeper layers by ~50% at WY-MR (Figure 4A).



**Figure 4.** Average observed and simulated values of bulk soil organic carbon (SOC; A), particulate organic carbon (POC; B), and mineral-associated organic carbon (MAOC; C) along the soil profile across experimental sites located in Oklahoma (OK-CR and OK-RR), Michigan (MI-LC), and Wyoming (WY-MR). Horizontal red bars represent the 2.5<sup>th</sup>-97.5<sup>th</sup> percentile ranges of observed values and horizontal teal bars represent the 95% posterior predictive intervals of simulated values. For simulated values, posterior predictive intervals were derived from the distribution of model outputs generated using the 200 parameter sets from the posterior of the Bayesian framework.

Regarding SOM fractions, the model generally performed better in reproducing POC (d: 0.42-0.90) than MAOC (d: 0.25-0.72) (Figure 4B-C; Table 3), with the largest discrepancies between observed and simulated POC values occurring at WY-MR (BIAS = -3.29 kg C m<sup>-3</sup>), particularly in the 50-100 cm layer (4.21 vs. 0.12 kg C m<sup>-3</sup>; Figure 4B). For MAOC, model performance was comparatively better at WY-MR (d: 0.72, R<sup>2</sup>: 0.81, BIAS: 0.51 kg C m<sup>-3</sup>), whereas larger deviations were observed at OK-CR (BIAS: 6.01 kg C m<sup>-3</sup>) and MI-LC (BIAS: 4.39 kg C m<sup>-3</sup>). Overall, simulated MAOC values were consistently greater than measured values across sites. At OK-CR, simulated MAOC concentrations were ~30% greater than measured values across the soil profile, whereas at OK-RR they were ~2.3 times greater. At MI-LC, the overestimation increased with depth and was most pronounced below 30 cm, where simulated MAOC values averaged approximately twice the observed concentrations (Figure 4C). At WY-MR, the largest discrepancy occurred in the surface layer, with simulated MAOC ~1.3 times greater than observations. However, the variability of these sites was captured by the

model as indicated by the high  $R^2$  (0.83-0.88) (Table 3). Despite these discrepancies, the model effectively captured overall SOC dynamics along the soil profile across the experimental sites (Figure 4; Table 3), as observed values consistently fell within the 95% posterior predictive intervals of modeled values. Across sites and fractions, 95% posterior predictive intervals ranged from -6.1 to 41.8  $\text{kg C m}^{-3}$  for simulated, while observed 2.5<sup>th</sup>-97.5<sup>th</sup> percentile ranged from 0.3 to 41.4  $\text{kg C m}^{-3}$  (Figure 4).

### 3.4.2. Ecosystem Fluxes at the 3M Experimental Sites

The model captured GPP temporal trends with reasonable accuracy across most experimental sites, as evidenced by the high  $d$  (0.80-0.94) and  $R^2$  values (0.68-0.82) (Table 3). However, differences in the magnitude of errors were observed across sites (RMSE: 3.70-18.8  $\text{g C m}^{-2} \text{ week}^{-1}$ ). At OK-RR, despite the high  $d$  value, the model showed limited ability to reproduce the EC variability ( $R^2 = 0.52$ ) and exhibited the greatest underestimation among sites (BIAS = -5.76  $\text{g C m}^{-2} \text{ week}^{-1}$ ). A GPP underestimation was observed also at OK-CR (BIAS = -2.48  $\text{g C m}^{-2} \text{ week}^{-1}$ ) and WY-MR (BIAS = -0.45  $\text{g C m}^{-2} \text{ week}^{-1}$ ), though to a lesser extent, whereas it was slightly overestimated at MI-LC (3.02  $\text{g C m}^{-2} \text{ week}^{-1}$ ; Table 3).

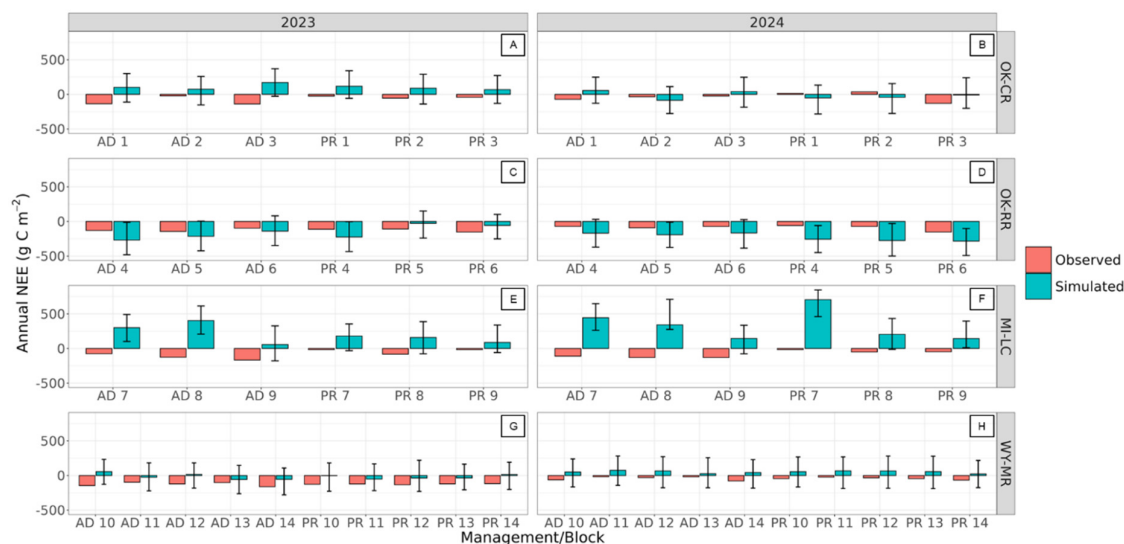
For ER, the model also reproduced temporal patterns with reasonable accuracy across experimental sites, supported by high  $d$  (0.73-0.94) and  $R^2$  (0.65-0.80). Model performance was stronger for WY-MR and OK-CR, where RMSE and BIAS were lowest, and comparatively weaker for OK-RR and MI-LC, where both metrics were higher (Table 3).

For NEE, when data were analyzed on a weekly basis (Table 3), the model reproduced the overall magnitude and direction of NEE reasonably well across most experimental sites, as indicated by the high values of  $d$  (0.68-0.80) and low BIAS (from -1.60 to 6.51  $\text{g C m}^{-2} \text{ week}^{-1}$ ) (Table 3). However, the relatively low  $R^2$  values (0.25-0.59) indicate that the model explained only a moderate fraction of the observed variability and was less accurate at reproducing the temporal dynamics of the EC fluxes.

Different plot-level discrepancies were observed between EC flux and simulated annual NEE across experimental sites (Figure 5). At OK-CR and MI-LC (Figure 5A, B, E, and F), the simulated annual NEE exhibited opposing patterns in the EC fluxes for most plots, irrespective of grazing management. At OK-RR and WY-MR (Figure 5C, D, G, and H), simulated and measured atmospheric flux values were more closely aligned, though overall the model tended to overestimate net C uptake at OK-RR across years and underestimate uptake at WY-MR in 2023. In 2024 the model exhibited an opposite trend in NEE at WY-MR compared to EC fluxes ( $\sim 53.5 \text{ g C m}^{-2}$  vs.  $\sim -41.9 \text{ g C m}^{-2}$ ). Overall, the annual simulated NEE across plots for these experimental sites varied from 56.9 to 704.6  $\text{g C m}^{-2}$  at MI-LC and from -84.2 to 171.4  $\text{g C m}^{-2}$  at OK-CR, compared with atmospheric EC flux values of -166.7 to -16.1  $\text{g C m}^{-2}$  and -137.8 to 36.8  $\text{g C m}^{-2}$ , respectively.

Even though simulated NEE values at the plot level differed from EC fluxes,  $\sim 90\%$  of EC fluxes fell within the 95% posterior predictive intervals of modeled values, while one plot at OK-RR and five plots at MI-LC fell outside these limits (Figure 5). The 95% posterior predictive intervals for simulated annual NEE ranged from -498.6 to 844.9  $\text{g C m}^{-2}$  across sites and years, with wider intervals in 2023 than in 2024.

Treatment differences in annual NEE exhibited substantial variability across experimental sites, and between observed and simulated estimates (Figure 6). For the observed NEE, treatment differences were negative at OK-CR (-37.6  $\text{g C m}^{-2}$ ), MI-LC (-85.2  $\text{g C m}^{-2}$ ) and WY-MR (-0.3  $\text{g C m}^{-2}$ ), whereas the difference was positive at OK-RR (8.7  $\text{g C m}^{-2}$ ). In contrast, simulated treatment differences showed an opposite pattern, with positive values at OK-CR (31.1  $\text{g C m}^{-2}$ ) and MI-LC (35.3  $\text{g C m}^{-2}$ ), and WY-MR (4.0  $\text{g C m}^{-2}$ ) and negative value at OK-RR (-4.1  $\text{g C m}^{-2}$ ). In both observed and simulated treatment differences, confidence intervals overlapped zero at all sites. The 95% confidence intervals ranged from -145.7 to 87.0  $\text{g C m}^{-2}$  for EC-based NEE, while those for simulated NEE varied from -114.9 to 173.2  $\text{g C m}^{-2}$ .



**Figure 5.** Annual eddy covariance flux tower and simulated net ecosystem exchange (NEE) across blocks under adaptive (AD) and prescriptive (PR) grazing management for 2023 and 2024 at experimental sites located in Oklahoma (OK-CR: A-B; and OK- RR: C-D), Michigan (MI-LC: E-F), and Wyoming (WY-MR: G-H). Vertical bars represent 95% posterior predictive intervals of simulated values. For simulated values, posterior predictive intervals were derived from the distribution of model outputs generated using the 200 parameter sets from the posterior of the Bayesian framework.

## 4. Discussion

### 4.1. Model Performance Representing Plant Growth Dynamics at NEON and 3M Experimental Sites

Proper representation of plant growth dynamics is essential for accurately simulating SOC dynamics and ecosystem fluxes, as it governs the magnitude and timing of soil C inputs. For the NEON sites, the model reproduced the general trends of MODIS GPP relatively well, which explained most of the variability across sites and years (Table 2). However, some deviations and relatively high RMSE values indicate increased variability in model performance. This variability likely reflects limited site- and year-specific management information for the NEON sites, as the generalized management assumptions used in our simulations may have not captured the actual grazing intensity or timing. Additional discrepancies also can arise from scale mismatches between MODIS GPP and model simulations since substantial variation in soil properties and grazing pressure can occur in a single MODIS pixel but remains unrepresented in point-scale simulations (Figure S1).

Measurements of aboveground biomass and LAI at the 3M experimental sites provided valuable information for plant parameterization, as these two variables are closely related and provide a more complete understanding of vegetation growth and productivity [76]. Although MEMS 3.0 adequately captured plant biomass across experimental sites, its ability to reproduce seasonal variability was more limited, as reflected in low  $R^2$  values across sites (Table 2). For example, the model captured aboveground dynamics very well in some years under both PR and AD grazing management at MI-LC, though it was less effective in others (Figure S2). While representing dominant plant functional types, an approach commonly used in other models (e.g., DayCent [18], PaSim [77], and ORCHIDEE [16]), is often adequate for grazing systems with relatively simple vegetation structure and limited functional diversity, this simplification can limit model performance in more diverse and structurally complex landscapes. For example, a study using two versions of the ORCHIDEE model reported  $d$  values of 0.20-0.45 for aboveground biomass and 0.10-0.56 for LAI in grazed grasslands of Europe [16], substantially lower than those obtained in our study (Table 2).

In fact, in a previous study where MEMS was applied to a monoculture of bahiagrass (*Paspalum notatum*), [48] achieved much higher model performance ( $R^2 = 0.72$ ), underscoring that mixed-species communities are inherently more challenging to represent due to their diversity of traits and growth strategies. Although improving species-level representation may enhance simulations of plant dynamics and reduce uncertainty in vegetation-related outputs [78], such approaches remain challenging due to the extensive species-specific data requirements and the need substantially more complex model architectures involved [79].

#### 4.2. Soil Temperature and Soil Water Content Dynamics at 3M Experimental Sites

Representative simulation of soil temperature and water dynamics are crucial for accurate ecosystem modeling as these factors strongly regulate plant growth, microbial activity, and the decomposition of SOM, ultimately shaping SOC dynamics and ecosystem fluxes. Our results indicated that both variables were well represented in the MEMS model (Table 3). However, soil texture and the lack of accurate simulation of snow dynamics may have influenced soil temperature and SWC, respectively.

Soil temperature was slightly underestimated, particularly during winter when observed temperatures approached or dropped below 0 °C across sites. This bias likely reflects the simple empirical approach used to estimate soil surface temperature in our model [80], which does not explicitly resolve the surface energy balance, especially under snow-covered conditions. As a result, inaccuracies in the surface boundary condition may propagate through the soil profile, leading to systematic biases in simulated soil temperatures. These effects appear to be amplified in sandy soils, which exhibited a stronger negative bias in simulated soil temperature (Table 3), consistent with previous MEMS 2.0 results [29] and evident when comparing OK-RR and OK-CR under similar climatic conditions but contrasting soil textures. In addition, uncertainties in the spatial and temporal distribution of snow likely exacerbated discrepancies at the coldest sites (MI-LC and WY-MR), consistent with earlier DayCent model simulations of winter grassland conditions at CPER in northern Colorado [80]. Despite these limitations, the model demonstrated overall high predictive accuracy for soil temperature, in line with previous MEMS applications [29,47].

MEMS 3.0 adequately captured the temporal dynamics and magnitude of SWC across most experimental sites (Table 3), suggesting that major hydrological processes governing soil moisture, including infiltration, evaporation, and drainage, were realistically represented. Except for WY-MR,  $d$  values were relatively high and the model explained a substantial fraction of the observed variability (Table 3). The poorer performance at WY-MR likely reflects challenges in representing snow dynamics, which strongly influence water balance [81] but are not fully captured by the model. Additionally, the rock content at WY-MR, which exceeds 13% in some soil samples, may have further contributed to discrepancies in simulated and observed soil water dynamics since the ROSETTA pedotransfer model used to estimate soil hydraulic parameters does not account for rock content. Lastly, our model does not explicitly represent slope position, topographic gradients, or lateral water movement, all of which can substantially influence soil moisture dynamics and likely contributed to discrepancies between observed and simulated values.

Despite these limitations, model performance was reasonable with relatively accurate predictions across all sites and layers, with small BIAS and RMSE values indicating minimal systematic deviation (Table 3). Overall, the model performed better than previous applications of MEMS at different sites, particularly in deeper soil layers where it explained a greater portion of the observed variability [29,47] and yielded lower RMSE values [29]. Compared with PaSim simulations of Mediterranean grasslands, MEMS 3.0 produced similar  $d$  values for topsoil SWC and explained more variability at OK-CR and OK-RR, but was less accurate at MI-LC and WY-MR [82]. Relative to topsoil APSIM simulations for the Mongolian steppe, MEMS exhibited RMSE values were on average ( $0.06 \text{ cm}^3 \text{ cm}^{-3}$ ) twice as large as those reported for APSIM [83].

#### 4.3. Model Calibration Using the NEON Sites

The posterior distributions of the calibrated parameters showed that the Bayesian calibration reduced uncertainty in parameter ranges, indicating improved parameter identifiability (Figure 2). The posterior distributions of the two decomposition rate parameters for MAOM (“k\_MAOMDecay”) and POM (“k\_POMDeplomer”) were close to a bell shape, suggesting that these parameters were well constrained by the observational data. In contrast, the posterior distributions of “Coeff\_TemperatureEffOnDecomp\_1”, “Frac\_MAOMExchangeable”, and “k\_DOMSorp” were skewed toward either the lower or upper boundaries of their prior ranges. This pattern suggests potential structural limitations in the model. Specifically, such “edge-hitting” behavior often indicates that the model is attempting to compensate for simplified or omitted process representations by pushing parameters toward extreme values to minimize the residual error between observations and predictions. These patterns are also often observed in other model studies [35,36], indicating that model improvements are needed.

For annual NEE of the NEON sites, model performance may have been influenced by limited information on grazing management within the flux tower fetch areas or by incomplete representation of grassland community physiological responses to grazing and other environmental drivers. This variability was particularly evident in the predicted GPP, as in some years MEMS predictions diverged from EC estimates, which instead aligned more closely with MODIS GPP estimate [50] (Figure S1). In some other years, MEMS GPP predictions agreed well with MODIS but not with the EC flux tower (Figure S1). These results may reflect the coarse spatial resolution of MODIS (representing broader landscape conditions and management practices) contrasting with the localized footprint of EC measurements.

In addition, the quality of the EC dataset itself may have contributed to the discrepancies observed. We noticed anomalously high GPP values during winter months when they are expected to be close to 0 (Figure S1), which would result in highly negative measured NEE fluxes. Although these clearly problematic data periods were excluded from our analysis, the remaining data may still contain uncertainties arising from potential issues such as errors in gap-filling algorithms, sensor biases, or other measurement-related artifacts.

#### 4.4. Model Performance Representing Soil Organic Carbon Dynamics at 3M Experimental Sites

Accurate representation of SOC dynamics is essential for capturing subsequent ecosystem fluxes as SOM decomposition accounts for a large fraction of ecosystem CO<sub>2</sub> exchange. Overall, the model reproduced baseline total SOC well, but larger discrepancies emerged in the partitioning of SOM fractions (Figure 4; Table 3). The strong agreement between simulated and observed SOC at most sites (Figure 4A), reflected by high  $d$  values, indicates that the model captured the dominant controls on SOM dynamics across contrasting climates and management histories, consistent with previous MEMS applications [29,48].

Simulated SOC values generally fell within the 2.5<sup>th</sup>-97.5<sup>th</sup> percentile ranges of measurements across depths, indicating consistency with observations (Figure 4A). At the same time, site-specific BIAS values reveal systematic tendencies toward SOC overestimation at OK-CR and MI-LC and weaker underestimation at WY-MR (Table 3), suggesting that some stabilization and decomposition processes may not yet be fully represented in the model. At OK-CR, discrepancies likely reflect underrepresented surface decomposition, whereas at MI-LC, SOC overprediction at depth may indicate insufficient representation of decomposition under buffered moisture and temperature conditions. At WY-MR, missing interactions between soil inorganic C and SOC and high calcium availability, which promote C accrual through enhanced soil aggregation, co-precipitation, and mineral associations [84], as well as potential contributions from pyrogenic C [85], may have contributed to the observed tendency of underestimation.

The model also showed limitations in reproducing the observed partitioning between POC and MAOC, with POC frequently slightly underpredicted and MAOC correspondingly overpredicted (Figure 4B-C; Table 3). This imbalance likely reflects missing or simplified stabilization mechanisms, including the lack of representation of POM protection by aggregate occlusion, of mineralogy

controlled organo-mineral interactions, and microbial feedbacks that are known to regulate C transfer and persistence [86]. In addition, incomplete C recovery during SOM fractionation can introduce a methodological discrepancy relative to model outputs [87], which does not account for this methodological limitation, and may further contribute to observed mismatches. Incorporating microbial-mediated and mineralogy-specific stabilization processes would likely improve the representation of SOM formation and stabilization processes [43,88], ultimately contributing to enhancing SOM fraction predictive performance.

Despite site-specific discrepancies between mean observed and simulated SOC values, the model 95% posterior predictive intervals encompassed all measurements (Figure 4), reflecting high uncertainty associated with SOC predictions. Although comparable uncertainty levels have been reported previously for cropland systems [35], explicit quantification of uncertainty in process-based SOC modeling across grazing systems remains largely unexplored, underscoring the relevance of this study. The magnitude of predictive uncertainty was, in some cases, comparable to the 2.5<sup>th</sup>-97.5<sup>th</sup> percentile ranges of the measurements themselves, highlighting the inherent spatial variability of SOC [89] and the associated challenges in constraining model simulations. These results reinforce the importance of explicitly accounting for predictive uncertainty to enable a more comprehensive interpretation of model performance, identify contexts in which predictions are most reliable, and identify potential limitations when applying SOC models across heterogeneous landscapes such as grazing lands. Addressing these uncertainties will require improved representation of context-dependent SOC controls [90,91] and detailed site-level data to better constrain parameterization and improve predictions across diverse ecoregions.

#### 4.5. Model Performance on Gross Plant Productivity and Ecosystem Respiration at 3M Experimental Sites

The model effectively represented ecosystem dynamics across contrasting climates and grazing management strategies as indicated by the strong agreement between simulated and observed weekly cumulative GPP and ER (Table 3). These results are consistent with Chang et al. [16] who simulated seven grazed European grassland sites under diverse plant compositions and reported RMSE values ranging from 1.10 to 2.38 g C m<sup>-2</sup> day<sup>-1</sup> (equivalent to 7.70-16.66 g C m<sup>-2</sup> week<sup>-1</sup>) for GPP and from 0.62 to 1.52 g C m<sup>-2</sup> day<sup>-1</sup> (equivalent to 4.34-10.64 g C m<sup>-2</sup> week<sup>-1</sup>) for ER using two versions of the ORCHIDEE model. Although our analysis is based on weekly cumulative values rather than seasonal-to-annual timescales, the comparable RMSE ranges reinforce the overall good performance of our model (Table 3).

Block-level differences in simulated and EC-derived GPP reflected spatial variability in plant growth and grazing effects within the 3M experimental sites (Figure S3). During winter, EC data suggested residual plant activity while the model simulated complete dormancy, likely because the model represents only the dominant grass species and does not capture limited growth by other functional groups (e.g., cool-season species or forbs). This discrepancy was most apparent at the warmer sites (OK-CR and OK-RR) where winter conditions do not strongly constrain plant growth, whereas simulated and EC-derived GPP agreed closely at colder sites where low temperatures and snow cover limit productivity (Figure S3E-H). Simulated GPP also exhibited greater short-term variability in response to grazing events than EC estimates, particularly at MI-LC, where block-level differences were most pronounced (Figure S3E-F). This possibly reflects variability in plot-level inputs (e.g., soil texture and soil hydraulic parameters), that directly affect plant productivity. In contrast, such differences were less evident at WY-MR, possibly because EC measurements integrate fluxes over a more homogeneous landscape (e.g., soil texture). It is also important to note that flux-tower GPP is derived from measured NEE and estimated ER, which introduces additional errors [92,93] and may partially explain the mismatch between simulated and EC-derived GPP (Figure S3; Table 3). At the same time, the good agreement between simulated and MODIS-derived GPP (Figure S4) illustrates that different observation methods carry their own sources of biases, and discrepancies do not necessarily reflect model error alone.

For ER, simulated and observed values generally matched well across sites, except at OK-RR and MI-LC, where the model underestimated and overestimated ER, respectively (Table 3), although block-level simulations were more consistent with each other (Figure S5). RMSE values at these sites approached or exceeded those reported for grazed grasslands in previous modeling studies, including ORCHIDEE simulations in Europe ([16]; equivalent to 4.34-10.64 g C m<sup>-2</sup> week<sup>-1</sup>) and DayCent simulations in semiarid grasslands in China ([94]; equivalent to 12.60 g C m<sup>-2</sup> week<sup>-1</sup>). Most of the mismatch occurred during the growing season (Figure S5C-F), suggesting that the model did not fully capture seasonal increases in respiratory fluxes, potentially related to temporal variability in soil respiration drivers (e.g., litter and feces inputs) or biomass allocation. Because ecosystem respiration integrates multiple plant and soil components, and independent measurements of these components are lacking, isolating the exact source of this error was not possible. Furthermore, ER from EC flux towers is estimated under the assumption that nighttime NEE approximates ecosystem respiration, thus any deviation from this assumption introduces errors and may contribute to the disagreement between modelled and flux tower-derived ER [92].

#### 4.6. Net Ecosystem Exchange at 3M Experimental Sites

While the model captured much of the observed weekly variability, minor mismatches in simulated GPP and ER accumulated over time, leading to a substantial divergence in annual NEE estimates (Table 3; Figure 5). This contrast highlights the strong sensitivity of NEE to small but persistent biases in its component fluxes (Figure 5). Because NEE represents the balance between two large and opposing fluxes ( $NEE = ER - GPP$ ), even small systematic errors in either component can be amplified at longer timescales. This likely explains why model performance was consistently higher for GPP and ER than for NEE (Table 3), a pattern also observed by Kirschbaum et al. [95] when comparing atmospheric flux observations with the CenW model simulations for grazed pastures in New Zealand and France.

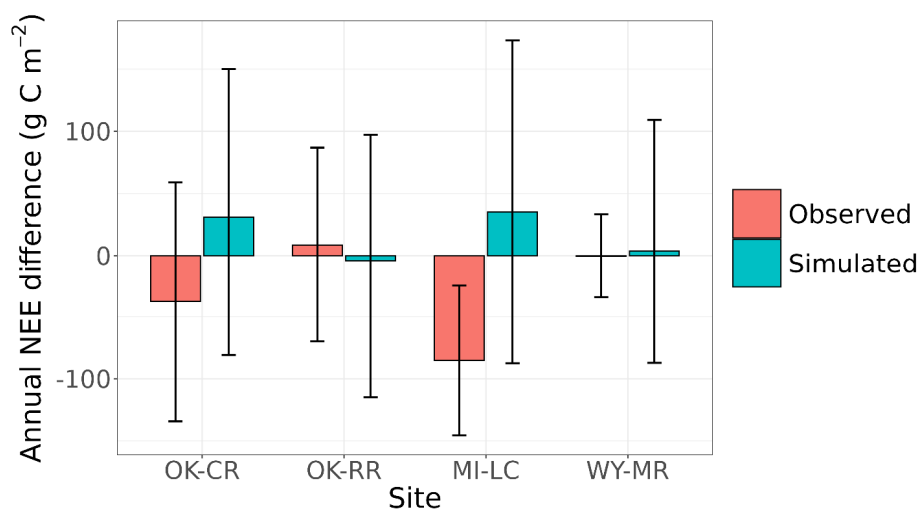
Capturing short-term NEE dynamics while exhibiting divergence in cumulative annual estimates has been widely documented across ecosystem models. For example, Takeda et al. [18] showed that DayCent-CABBI model reproduced the overall NEE trend over one year of measurements, but cumulative annual values diverged at most sites used for evaluation. Likewise, Chang et al. [16] reported notable mismatches between ORCHIDEE and EC flux observations across eleven European grasslands. Together, these studies reinforce that even when short-term dynamics are well captured, small biases can propagate into large annual discrepancies.

The NEE model-data mismatch may also reflect differences in instrumentation and flux-processing approaches between the NEON towers used for calibration and the Quanterra EC systems deployed at the 3M experimental sites and used for evaluation. For any EC system, installation location and instrumentation configuration coupled with anemometer coordinate rotations, lag effects, corrections for frequency response and density effects, along with varying approaches to gap-filling and quality control have potential to introduce systematic biases which are generally small on a 30-minute timescales but can accumulate over longer timescales [96–98]. Moreover, calibrating the model with NEON flux towers and validating with independent in-situ EC systems may have reduced predictive precision at individual experimental sites, though increased the robustness of the model evaluation [21].

Studies that calibrated models using subsets of local flux measurements, such as Falvo et al. [99] for MEMS 2.0 and Takeda et al. [18] for DayCent-CABBI, reported improved agreement between simulated and observed NEE. However, such improvements primarily reflect parameter optimization under the same environmental and management conditions used for evaluation, a practice that is frequently adopted due to data limitations but inherently limits rigorous model evaluation. By employing a strict independent calibration and evaluation Bayesian framework, our study provides a more robust assessment of model performance and uncertainty, enabling a clearer evaluation of model simulation reliability and accuracy across different ecoregions [21].

Despite discrepancies in annual NEE, modeled uncertainty generally overlapped with EC flux observations, limiting the ability to detect treatment effects even when mean differences appear to exist (Figure 5). However, this limitation is not restricted to the model, as uncertainties associated with measured EC-based annual NEE remain poorly constrained. Where they have been estimated, uncertainties ranged widely, for example from 79 to 127 g C m<sup>-2</sup> yr<sup>-1</sup> in a *Pinus* plantation [100] and from -140 to 130 g C m<sup>-2</sup> yr<sup>-1</sup> in Swiss grasslands [101], indicating that observational variability alone can be much larger than treatment differences.

This pattern becomes evident when examining annual NEE treatment differences for both observations and simulations (Figure 6). Although the observational confidence intervals shown here do not represent the full uncertainty associated with EC-based annual NEE, they capture the variability in NEE responses to grazing management and therefore allow a meaningful comparison with modeled uncertainty. Simulated annual mean treatment differences were often opposite in direction to the observed values. However, their confidence intervals largely overlapped across sites (Figure 6). This overlap indicates that the apparent differences in mean responses are not robust enough to support a clear divergence between modeled and observed treatment effects, but rather shows the combined effect of uncertainty in both model predictions and observations. These results underscore the importance of explicit uncertainty quantification and caution against interpreting modeled NEE predictions without consideration of associated observational uncertainty in grazing systems. This is crucial to provide a more robust foundation for model evaluation and to support decision-making based on realistic expectations of model performance across grazing lands.



**Figure 6.** Annual treatment [adaptive grazing (AD) and prescriptive grazing (PR)] differences in net ecosystem exchange (NEE) derived from eddy covariance flux tower observations and model simulations across experimental sites in Oklahoma (OK-CR and OK-RR), Michigan (MI-LC), and Wyoming (WY-MR). For each experimental site, treatment differences were calculated as mean AD NEE minus mean PR NEE. Vertical bars represent the 95% confidence intervals for the mean difference in observed data between the two treatments and 95% posterior predictive intervals of the mean difference in treatments for simulations. For observed values, confidence intervals were obtained using Welch's t-test, considering plots as replicates within each site. For simulated values, posterior predictive intervals were derived from the distribution of model outputs generated using the 200 parameter sets from the posterior of the Bayesian framework.

## 5. Conclusions

Studies that apply process-based models to predict ecosystem C dynamics and rigorously calibrate and evaluate them using independent datasets remain rare, particularly in grazing systems. This study advances the field by providing a Bayesian quantification of uncertainty in MEMS 3.0

predictions of both SOC and NEE, based on fully independent datasets, while jointly evaluating model performance against soil measurements and atmospheric flux observations across a broad climatic gradient in U.S. grazing lands.

Overall, MEMS 3.0 reproduced key ecosystem drivers (e.g., plant growth, soil temperature, and soil water content) across sites, providing a consistent basis for simulating SOC dynamics and ecosystem fluxes in grazing systems. However, both observed and simulated annual treatment differences in NEE exhibited large variability, posing challenges on accurately detecting grazing management effects.

Reducing uncertainty in model-based assessments of grazing systems will require advances on both observational and modeling fronts. Improved harmonization and longer time series of EC flux measurements will strengthen model calibration and evaluation. Also, explicitly accounting for uncertainty associated with EC flux measurements would provide a more robust basis for model-data comparisons. For the NEON sites, more detailed management records and plant measurements (e.g., biomass and LAI) within EC flux tower footprints would be particularly valuable for better constraining plant parameterization and linking EC fluxes to ecosystem changes.

Our study emphasizes the importance of explicitly accounting for modeling uncertainty as it provides a more robust framework for interpreting model results and sets realistic expectations regarding the strengths and limitations of process-based models applied to grazing systems. Ultimately, this is essential for evaluating grazing as a climate mitigation strategy and for developing credible C accounting, while simultaneously balancing livestock production goals.

**CRedit authorship contribution statement:** **Rafael S. Santos:** Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Emma K. Hamilton:** Data curation, Formal analysis, Investigation, Methodology, Software, Writing - review & editing. **Paige L. Stanley:** Methodology, Writing - review & editing. **Robert Clement:** Methodology, Writing - review & editing. **Rebecca Mitchell:** Methodology, Writing - review & editing. **Hao Yang:** Data curation, Writing - review & editing. **Lauren Hoskovec:** Methodology, Writing - review & editing. **Isabella C. F. Maciel:** Data curation, Methodology, Writing - review & editing. **Jason Rowntree:** Methodology, Funding acquisition, Project administration, Writing - review & editing. **John D. Scasta:** Data curation, Methodology, Writing - review & editing. **Alejandro Romero-Ruiz:** Methodology, Writing - review & editing. **Steven M. Ogle:** Methodology, Writing - review & editing. **M. Francesca Cotrufo:** Conceptualization, Investigation, Methodology, Funding acquisition, Project administration, Supervision, Writing - original draft, Writing - review & editing. **Yao Zhang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Project administration, Supervision, Visualization, Writing - original draft, Writing - review & editing.

**Declaration of Competing Interest:** Yao Zhang, Rafael S. Santos, Emma K. Hamilton, Steven Ogle, and M. Francesca Cotrufo receive a portion of the licensing fees associated with the MEMS model. All other authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

**Data availability:** Data will be made available on request.

**Acknowledgments:** This research was supported by the Foundation for Food and Agriculture Research (FFAR grant number: DSnew-000000028), the Noble Institute, Greenacres Foundation, and Butcherbox, the United States Department of Agriculture (USDA agreement number NR233A750004G050), and the Shell International Exploration and Production Inc. (contract n. CW517568). The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of our funders. We thank NEON Field Science Manager Andrea C. E. Anteau for providing site management information. We thank Erica L. Patterson for soil sampling. We thank Alexandria Kuhl, Glenn O'Neil, and Jeremiah Asher for providing soil water data. We thank Guilherme F. S. Congio, Hugh Aljoe, Jeff J. Goodwin, João P. Sacramento, Lautaro Garcia, Nicole M. Nimlos, and Timm Gergeni for assistance with site management information. We thank staff from Lake City Research Station, Noble Research institute, and McGuire Ranch in assistance with experimental sites maintenance and cattle management.

## References

1. IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi: 10.1017/9781009157926. .
2. Zhang C, Dong Q, Chu H, et al. Grassland Community Composition Response to Grazing Intensity Under Different Grazing Regimes. *Rangel. Ecol. Manag.* 71(2), 196–204 (2018).
3. Ren S, Terror C, Li J, Cao Y, Yang S, Liu D. Historical impacts of grazing on carbon stocks and climate mitigation opportunities. *Nat. Clim. Chang.* 14(4), 380–386 (2024).
4. Sanderman J, Hengl T, Fiske GJ. Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences* [Internet]. 114(36), 9575–9580 (2017). Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1706103114>.
5. Lal R. Digging deeper: A holistic perspective of factors affecting soil organic carbon sequestration in agroecosystems. *Glob. Chang. Biol.* 24(8), 3285–3301 (2018).
6. FAO. Land statistics 2001–2022 – Global, regional and country trends. FAOSTAT Analytical Briefs, No. 88. Rome. [Internet]. . Available from: <https://openknowledge.fao.org/handle/20.500.14283/cd1484en>.
7. Bai Y, Cotrufo MF. Grassland soil carbon sequestration: Current understanding, challenges, and solutions. *Science* (1979). [Internet]. 377(6606), 603–608 (2022). Available from: <https://www.science.org/doi/10.1126/science.abo2380>.
8. Mcsherry ME, Ritchie ME. Effects of grazing on grassland soil carbon: A global review. *Glob. Chang. Biol.* 19(5), 1347–1357 (2013).
9. Paustian K, Lehmann J, Ogle S, Reay D, Robertson GP, Smith P. Climate-smart soils. *Nature* [Internet]. 532(7597), 49–57 (2016). Available from: <http://www.nature.com/articles/nature17174>.
10. Bradford MA, Kuebbing SE, Polussa A, Sanderman J, Oldfield EE. Upstream data need to prove soil carbon as a climate solution. *Nat. Clim. Chang.* [Internet]. 15(10), 1013–1016 (2025). Available from: <https://www.nature.com/articles/s41558-025-02429-4>.
11. Dong M, de Kroon H. Plasticity in Morphology and Biomass Allocation in *Cynodon dactylon*, a Grass Species Forming Stolons and Rhizomes. *Oikos* [Internet]. 70(1), 99 (1994). Available from: <https://www.jstor.org/stable/3545704?origin=crossref>.
12. Turner CL, Seastedt TR, Dyer MI. Maximization of aboveground grassland production: the role of defoliation frequency, intensity, and history. *Ecological Applications.* 3(1), 175–186 (1993).
13. Milchunas DG, Varnamkhasti AS, Lauenroth WK, Goetz H. Forage quality in relation to long-term grazing history, current-year defoliation, and water resource. *Oecologia* [Internet]. 101(3), 366–374 (1995). Available from: <http://link.springer.com/10.1007/BF00328824>.
14. Pavlů V, Hejcman M, Pavlů L, Gaisler J, Nežerková P. Effect of continuous grazing on forage quality, quantity and animal performance. *Agric. Ecosyst. Environ.* [Internet]. 113(1–4), 349–355 (2006). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0167880905005141>.
15. Stanley PL, Wilson C, Patterson E, Machmuller MB, Cotrufo MF. Ruminating on soil carbon: Applying current understanding to inform grazing management. *Glob. Chang. Biol.* [Internet]. 30(3) (2024). Available from: <https://onlinelibrary.wiley.com/doi/10.1111/gcb.17223>.
16. Chang JF, Viovy N, Vuichard N, et al. Incorporating grassland management in ORCHIDEE: model description and evaluation at 11 eddy-covariance sites in Europe. *Geosci. Model Dev.* [Internet]. 6(6), 2165–2181 (2013). Available from: <https://gmd.copernicus.org/articles/6/2165/2013/>.
17. Lawton D, Leahy P, Kiely G, Byrne KA, Calanca P. Modeling of net ecosystem exchange and its components for a humid grassland ecosystem. *J. Geophys. Res. Biogeosci.* [Internet]. 111(G4) (2006). Available from: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2006JG000160>.
18. Takeda N, Rowlings D, Parton W, et al. Soil carbon sequestration potential in subtropical grasslands estimated by DayCent-CABBI. *Soil Science Society of America Journal.* 89(1) (2025).
19. Campbell EE, Paustian K. Current developments in soil organic matter modeling and the expansion of model applications: A review. *Environmental Research Letters*10(12) (2015).

20. Zhang Y, Santos RS, Hamilton EK, et al. Modeling the Effects of Livestock Rotation Frequency on Forage Production and Soil Carbon Using a Spatially Explicit Grazing Distribution [Internet]. (2026). Available from: <https://www.preprints.org/manuscript/202602.0923/v1>.
21. Le Noë J, Manzoni S, Abramoff R, et al. Soil organic carbon models need independent time-series validation for reliable prediction. *Commun. Earth Environ.* 4(1) (2023).
22. Stanley P, Kuhl A, Rowntree J, Maciel I. Merging Rigor and Relevance in Grazingland Research: A Comprehensive Social-Ecological Monitoring Approach. Under review. (2026).
23. Baldocchi D. Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method. *Glob. Chang. Biol.* 20(12), 3600–3609 (2014).
24. Eugster W, Merbold L. Eddy covariance for quantifying trace gas fluxes from soils. *SOIL* [Internet]. 1(1), 187–205 (2015). Available from: <https://soil.copernicus.org/articles/1/187/2015/>.
25. Baldocchi DD. How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Glob. Chang. Biol.* [Internet]. 26(1), 242–260 (2020). Available from: <https://onlinelibrary.wiley.com/doi/10.1111/gcb.14807>.
26. Oldfield EE, Lavalley JM, Blesh J, et al. Greenhouse gas mitigation on croplands: clarifying the debate on knowns, unknowns and risks to move forward with effective management interventions. *Carbon Manag.* [Internet]. 15(1) (2024). Available from: <https://www.tandfonline.com/doi/full/10.1080/17583004.2024.2365896>.
27. Gottschalk P, Wattenbach M, Neftel A, et al. The role of measurement uncertainties for the simulation of grassland net ecosystem exchange (NEE) in Europe. *Agric. Ecosyst. Environ.* 121(1–2), 175–185 (2007).
28. Vuichard N, Soussana JF, Ciais P, et al. Estimating the greenhouse gas fluxes of European grasslands with a process-based model: 1. Model evaluation from in situ measurements. *Global Biogeochem. Cycles.* 21(1) (2007).
29. Zhang Y, Lavalley JM, Robertson AD, et al. Simulating measurable ecosystem carbon and nitrogen dynamics with the mechanistically defined MEMS 2.0 model. *Biogeosciences.* 18(10), 3147–3171 (2021).
30. Hu T, Malone SL, Rumpel C, Chabbi A. Maximizing soil organic carbon stocks through optimal ploughing and renewal strategies in (Ley) grassland. *Commun. Earth Environ.* 5(1) (2024).
31. Puche N, Senapati N, Flechard CR, Klumpp K, Kirschbaum MUF, Chabbi A. Modeling carbon and water fluxes of managed grasslands: Comparing flux variability and net carbon budgets between grazed and mowed systems. *Agronomy.* 9(4) (2019).
32. Sándor R, Barcza Z, Hidy D, Lellei-Kovács E, Ma S, Bellocchi G. Modelling of grassland fluxes in Europe: Evaluation of two biogeochemical models. *Agric. Ecosyst. Environ.* 215, 1–19 (2016).
33. Thentu TL, Forster D, Li Y, et al. DNDC modelling of greenhouse gas exchange from a boreal legume grassland under organic and mineral nitrogen management. *J. Environ. Manage.* 390 (2025).
34. Ogle SM, Jay Breidt F, Easter M, Williams S, Killian K, Paustian K. Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. *Glob. Chang. Biol.* 16(2), 810–822 (2010).
35. Gurung RB, Ogle SM, Breidt FJ, Williams SA, Parton WJ. Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty. *Geoderma* [Internet]. 376, 114529 (2020). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0016706120310375>.
36. Mathers C, Black CK, Segal BD, et al. Validating DayCent-CR for cropland soil carbon offset reporting at a national scale. *Geoderma.* 438 (2023).
37. Ogle SM, Breidt FJ, Easter M, Williams S, Paustian K. An empirically based approach for estimating uncertainty associated with modelling carbon sequestration in soils. *Ecol. Modell.* 205(3–4), 453–463 (2007).
38. Zhang Y, Arabi M, Paustian K. Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. *Environmental Modelling and Software.* 126 (2020).
39. van Oijen M. Bayesian Methods for Quantifying and Reducing Uncertainty and Error in Forest Models. *Current Forestry Reports* [Internet]. 3(4), 269–280 (2017). Available from: <http://link.springer.com/10.1007/s40725-017-0069-9>.

40. Van Oijen M, Rougier J, Smith R. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol.* [Internet]. 25(7), 915–927 (2005). Available from: <https://academic.oup.com/treephys/article-lookup/doi/10.1093/treephys/25.7.915>.
41. Lu D, Ricciuto D, Walker A, Safta C, Munger W. Bayesian calibration of terrestrial ecosystem models: A study of advanced Markov chain Monte Carlo methods. *Biogeosciences.* 14(18), 4295–4314 (2017).
42. Hinckley ES, Bonan GB, Bowen GJ, et al. The soil and plant biogeochemistry sampling design for The National Ecological Observatory Network. *Ecosphere* [Internet]. 7(3) (2016). Available from: <https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecs2.1234>.
43. Rocci KS, Cotrufo MF, Ernakovich J, et al. Bridging 20 Years of Soil Organic Matter Frameworks: Empirical Support, Model Representation, and Next Steps. *J. Geophys. Res. Biogeosci.* [Internet]. 129(6) (2024). Available from: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023JG007964>.
44. Cotrufo MF, Soong JL, Horton AJ, et al. Formation of soil organic matter via biochemical and physical pathways of litter mass loss. *Nat. Geosci.* 8(10), 776–779 (2015).
45. Liang C, Schimel JP, Jastrow JD. The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol.* [Internet]. 2(8), 17105 (2017). Available from: <http://www.nature.com/articles/nmicrobiol2017105>.
46. Sokol NW, Sanderman J, Bradford MA. Pathways of mineral-associated soil organic matter formation: Integrating the role of plant carbon source, chemistry, and point of entry. *Glob. Chang. Biol.* 25(1), 12–24 (2019).
47. Zhang Y, King AE, Hamilton E, Cotrufo MF. Representing cropping systems with the MEMS 2 ecosystem model. *Agron. J.* [Internet]. 116(5), 2328–2345 (2024). Available from: <https://access.onlinelibrary.wiley.com/doi/10.1002/agj2.21611>.
48. Santos RS, Hamilton EK, Stanley PL, Paustian K, Cotrufo MF, Zhang Y. Simulating adaptive grazing management on soil organic carbon in the Southeast U.S.A. using MEMS 2. *J. Environ. Manage.* 365 (2024).
49. Santos RS, Hamilton EK, Stanley PL, Paustian K, Cotrufo MF, Zhang Y. Simulating adaptive grazing management on soil organic carbon in the Southeast U.S.A. using MEMS 2. *J. Environ. Manage.* 365 (2024).
50. Running S, Mu Q, Zhao M. MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V061 [Data set]. NASA Land Processes Distributed Active Archive Center. (2021).
51. NEON. NEON Eddy-Covariance Tower Data (DP4.00200.001) for sites US-xDC, US-xAE, US-xCP, US-xKZ, US-xNG, and US-xWD. Dataset accessed via AmeriFlux Portal - Lawrence Berkeley National Laboratory (<https://ameriflux.lbl.gov>). [Internet]. . Available from: <https://ameriflux.lbl.gov>.
52. Novick KA, Biederman JA, Desai AR, et al. The AmeriFlux network: A coalition of the willing. *Agric. For. Meteorol.* [Internet]. 249, 444–456 (2018). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0168192317303295>.
53. Pastorello G, Trotta C, Canfora E, et al. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci. Data* [Internet]. 7(1), 225 (2020). Available from: <https://www.nature.com/articles/s41597-020-0534-3>.
54. Kormann R, Meixner FX. An Analytical Footprint Model For Non-Neutral Stratification. *Boundary. Layer. Meteorol.* [Internet]. 99(2), 207–224 (2001). Available from: <https://link.springer.com/10.1023/A:1018991015119>.
55. Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society.* 146(730), 1999–2049 (2020).
56. Spencer S, Ogle SM, Breidt FJ, Goebel JJ, Paustian K. Designing a national soil carbon monitoring network to support climate change policy: a case example for US agricultural lands. *Greenhouse Gas Measurement and Management.* 1(3–4), 167–178 (2011).
57. Poelau C, Vos C, Don A. Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content. *SOIL* [Internet]. 3(1), 61–66 (2017). Available from: <https://www.soil-journal.net/3/61/2017/>.
58. Leuthold SJ, Haddix ML, Lavallee J, Cotrufo MF. Physical fractionation techniques [Internet]. In: Reference Module in Earth Systems and Environmental Sciences, Elsevier, 1–13 (2022). Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780128229743000677>.

59. Samson MÉ, Chantigny MH, Vanasse A, Menasseri-Aubry S, Angers DA. Coarse mineral-associated organic matter is a pivotal fraction for SOM formation and is sensitive to the quality of organic inputs. *Soil Biol. Biochem.* 149 (2020).
60. Cotrufo MF, Haddix ML, Mullen JL, Zhang Y, McKay JK. Deepening Root Inputs: Potential Soil Carbon Accrual From Breeding for Deeper Rooted Maize. *Glob. Chang. Biol.* 30(11) (2024).
61. NEON. Soil physical and chemical properties, Megapit (DP1.00096.001), RELEASE-2025. <https://doi.org/10.48443/0hnd-yj57>. Dataset accessed from <https://data.neonscience.org/data-products/DP1.00096.001/RELEASE-2025> on October 10, 2025. [Internet]. Available from: <https://data.neonscience.org/data-products/DP1.00096.001/RELEASE-2025>.
62. Vrugt JA, Ter Braak CJF. DREAM (D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrol. Earth Syst. Sci.* [Internet]. 15(12), 3701–3713 (2011). Available from: <https://hess.copernicus.org/articles/15/3701/2011/>.
63. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* [Internet]. 67(1) (2015). Available from: <http://www.jstatsoft.org/v67/i01/>.
64. Guillaume J, Andrews F. dream: DiffeRential Evolution Adaptive Metropolis. R package version 0.4-2. URL <http://CRAN.R-project.org/package=dream> [Internet]. CRAN: Contributed Packages (2012). Available from: <http://CRAN.R-project.org/package=dream>.
65. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (2025).
66. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* [Internet]. 7(4) (1992). Available from: <https://projecteuclid.org/journals/statistical-science/volume-7/issue-4/Inference-from-Iterative-Simulation-Using-Multiple-Sequences/10.1214/ss/1177011136.full>.
67. NEON. Soil physical and chemical properties, distributed initial characterization (DP1.10047.001), RELEASE-2025. <https://doi.org/10.48443/a4td-8k14>. Dataset accessed from <https://data.neonscience.org/data-products/DP1.10047.001/RELEASE-2025> on October 10, 2025. .
68. Abatzoglou JT. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology* [Internet]. 33(1), 121–131 (2013). Available from: <https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.3413>.
69. Guyette RP, Stambaugh MC, Dey DC, Muzika RM. Predicting Fire Frequency with Chemistry and Climate. *Ecosystems.* 15(2), 322–335 (2012).
70. Knapp AK, ohn Briggs JM, Collins SL, Hartnett DC, Johnson LC, Gene Towne E. The Keystone Role of Bison in North American Tallgrass Prairie. *Bioscience* [Internet]. (1999). Available from: <https://academic.oup.com/bioscience/article-lookup/doi/10.1525/bisi.1999.49.1.39>.
71. NEON. Site management and event reporting (DP1.10111.001), RELEASE-2025. <https://doi.org/10.48443/af3c-5w64>. Dataset accessed from <https://data.neonscience.org/data-products/DP1.10111.001/RELEASE-2025> on October 10, 2025. .
72. Beaudette D, Skovlin J, Roecker S, Brown A. soilDB: Soil Database Interface. R package version 2.8.13. (2025). Available from: <https://CRAN.R-project.org/package=soilDB>.
73. Schaap MG, Leij FJ, van Genuchten MTh. rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol. (Amst).* [Internet]. 251(3–4), 163–176 (2001). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022169401004668>.
74. Romero-Ruiz A, Rivero MJ, Milne A, et al. Grazing livestock move by Lévy walks: Implications for soil health and environment. *J. Environ. Manage.* 345 (2023).
75. Google. Google Maps Satellite Imagery [Data set]. [Internet]. (2025). Available from: <https://www.google.com/maps>.
76. Spehn EM, Joshi J, Schmid B, Diemer M, Körner C. Above-ground resource use increases with plant species richness in experimental grassland ecosystems. *Funct. Ecol.* [Internet]. 14(3), 326–337 (2000). Available from: <https://besjournals.onlinelibrary.wiley.com/doi/10.1046/j.1365-2435.2000.00437.x>.

77. Riedo M, Grub A, Rosset M, Fuhrer J. A pasture simulation model for dry matter production, and fluxes of carbon, nitrogen, water and energy. *Ecol. Modell.* [Internet]. 105(2–3), 141–183 (1998). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0304380097001105>.
78. Van Oijen M, Barcza Z, Confalonieri R, et al. Incorporating Biodiversity into Biogeochemistry Models to Improve Prediction of Ecosystem Services in Temperate Grasslands: Review and Roadmap. *Agronomy* [Internet]. 10(2), 259 (2020). Available from: <https://www.mdpi.com/2073-4395/10/2/259>.
79. Kipling RP, Virkajärvi P, Breitsameter L, et al. Key challenges and priorities for modelling European grasslands under climate change. *Science of The Total Environment* [Internet]. 566–567, 851–864 (2016). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0048969716310750>.
80. Parton W, Hartman M, Ojima D, Schimel D. DAYCENT and its land surface submodel: description and testing. *Glob. Planet. Change* [Internet]. 19(1–4), 35–48 (1998). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S092181819800040X>.
81. Maurer GE, Bowling DR. Seasonal snowpack characteristics influence soil temperature and water content at multiple scales in interior western U.S. mountain ecosystems. *Water Resour. Res.* 50(6), 5216–5234 (2014).
82. Pulina A, Lai R, Salis L, Seddaiu G, Roggero PP, Bellocchi G. Modelling pasture production and soil temperature, water and carbon fluxes in Mediterranean grassland systems with the Pasture Simulation model. *Grass and Forage Science.* 73(2), 272–283 (2018).
83. Wu L, Liu H, Liang B, et al. A process-based model reveals the restoration gap of degraded grasslands in Inner Mongolian steppe. *Science of the Total Environment.* 806 (2022).
84. Rowley MC, Grand S, Verrecchia ÉP. Calcium-mediated stabilisation of soil organic carbon. *Biogeochemistry* [Internet]. 137(1–2), 27–49 (2018). Available from: <http://link.springer.com/10.1007/s10533-017-0410-1>.
85. Glaser B, Amelung W. Pyrogenic carbon in native grassland soils along a climosequence in North America. *Global Biogeochem. Cycles.* 17(2) (2003).
86. Cotrufo MF, Lavalley JM. Soil organic matter formation, persistence, and functioning: A synthesis of current understanding to inform its conservation and regeneration. In: *Advances in Agronomy*, Academic Press Inc., 1–66 (2022).
87. Poeplau C, Don A, Six J, et al. Isolating organic carbon fractions with varying turnover rates in temperate agricultural soils – A comprehensive method comparison. *Soil Biol. Biochem.* [Internet]. 125(April), 10–26 (2018). Available from: <https://doi.org/10.1016/j.soilbio.2018.06.025>.
88. Manzoni S, Schimel JP. Advances in modelling soil microbial dynamics. *Soil Biol. Biochem.* 197 (2024).
89. Stanley P, Spertus J, Chiartas J, Stark PB, Bowles T. Valid inferences about soil carbon in heterogeneous landscapes. *Geoderma* [Internet]. 430, 116323 (2023). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0016706122006309>.
90. Hansen PM, Even R, King AE, Lavalley J, Schipanski M, Cotrufo MF. Distinct, direct and climate-mediated environmental controls on global particulate and mineral-associated organic carbon storage. *Glob. Chang. Biol.* [Internet]. 30(1), e17080 (2024). Available from: <http://www.ncbi.nlm.nih.gov/pubmed/38273571>.
91. Lavalley JM, Haddix ML, Swan A, Hoover JD, Cotrufo MF. Using aridity as an overarching factor to advance understanding of soil organic carbon storage at the continental scale. *Biogeochemistry.* 168(5) (2025).
92. Oikawa PY, Sturtevant C, Knox SH, Verfaillie J, Huang YW, Baldocchi DD. Revisiting the partitioning of net ecosystem exchange of CO<sub>2</sub> into photosynthesis and respiration with simultaneous flux measurements of <sup>13</sup>CO<sub>2</sub> and CO<sub>2</sub>, soil respiration and a biophysical model, CANVEG. *Agric. For. Meteorol.* [Internet]. 234–235, 149–163 (2017). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0168192316307432>.
93. Reichstein M, Falge E, Baldocchi D, et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Glob. Chang. Biol.* [Internet]. 11(9), 1424–1439 (2005). Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2486.2005.001002.x>.
94. Lei L, Xia J, Li X, et al. Water response of ecosystem respiration regulates future projection of net ecosystem productivity in a semiarid grassland. *Agric. For. Meteorol.* 252, 175–191 (2018).

95. Kirschbaum MUF, Puche NJB, Giltrap DL, Liáng LL, Chabbi A. Combining eddy covariance measurements with process-based modelling to enhance understanding of carbon exchange rates of dairy pastures. *Science of the Total Environment*. 745 (2020).
96. van Ramshorst JG V., Knohl A, Callejas-Rodelas JÁ, et al. Lower-cost eddy covariance for CO<sub>2</sub> and H<sub>2</sub>O fluxes over grassland and agroforestry. *Atmos. Meas. Tech.* [Internet]. 17(20), 6047–6071 (2024). Available from: <https://amt.copernicus.org/articles/17/6047/2024/>.
97. Schmidt A, Hanson C, Chan WS, Law BE. Empirical assessment of uncertainties of meteorological parameters and turbulent fluxes in the AmeriFlux network. *J. Geophys. Res. Biogeosci.* [Internet]. 117(G4) (2012). Available from: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2012JG002100>.
98. Liu Y, Stoy P, Chu H, et al. A tale of two towers: comparing NEON and AmeriFlux data streams at Bartlett Experimental Forest. *Agric. For. Meteorol.* [Internet]. 378, 110939 (2026). Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0168192325005581>.
99. Falvo G, Zhang Y, Abraha M, et al. Combining Eddy Covariance Towers, Field Measurements, and the MEMS 2 Ecosystem Model Improves Confidence in the Climate Impacts of Bioenergy With Carbon Capture and Storage. *GCB Bioenergy*. 17(3) (2025).
100. Oren R, Hsieh CI, Stoy P, et al. Estimating the uncertainty in annual net ecosystem carbon exchange: Spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements. *Glob. Chang. Biol.* 12(5), 883–896 (2006).
101. Ammann C, Flechard CR, Leifeld J, Neftel A, Fuhrer J. The carbon budget of newly established temperate grassland depends on management intensity. *Agric. Ecosyst. Environ.* 121(1–2), 5–20 (2007).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.