▎ **REVIEW** `OPEN ACCESS`

Combining Observations and Models: A Review of the CARDAMOM Framework for Data-Constrained Terrestrial Ecosystem Modeling

Matthew A. Worden[1] 📵 | T. Eren Bilir[2] 📵 | A. Anthony Bloom[2] 📵 | Jianing Fang[3] 📵 | Lily P. Klinek[4] 📵 |
Alexandra G. Konings[1] 📵 | Paul A. Levine[2] 📵 | David T. Milodowski[5] 📵 | Gregory R. Quetin[6] 📵 | T. Luke Smallman[5,7] 📵 |
Yinon M. Bar-On[8,9] 📵 | Renato K. Braghiere[2,8] 📵 | Cédric H. David[2] 📵 | Nina A. Fischer[10] 📵 | Pierre Gentine[3] 📵 |
Tim J. Green[5] 📵 | Ayanna Jones[11] 📵 | Junjie Liu[2,8] 📵 | Marcos Longo[12] 📵 | Shuang Ma[2,13] 📵 |
Troy S. Magney[14] 📵 | Elias C. Massoud[15] 📵 | Vasileios Myrgiotis[16] 📵 | Alexander J. Norton[17] 📵 | Nick Parazoo[2] 📵 |
Elahe Tajfar[2] 📵 | Anna T. Trugman[6] 📵 | Mathew Williams[5,7] 📵 | Sarah Worden[2] 📵 | Wenli Zhao[3] 📵 | Songyan Zhu[5,18] 📵

[1]Department of Earth System Science, Stanford University, Stanford, California, USA | [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA | [3]Department of Earth and Environmental Engineering, Columbia University, New York, New York, USA | [4]Department of Plant Sciences, University of California, Davis, California, USA | [5]School of Geosciences, University of Edinburgh, Edinburgh, UK | [6]Department of Geography, University of California, Santa Barbara, California, USA | [7]National Centre for Earth Observation, University of Edinburgh, Edinburgh, UK | [8]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, USA | [9]Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot, Israel | [10]School of Mathematics, University of Edinburgh, Edinburgh, UK | [11]Department of Atmospheric Sciences, Howard University, Washington, DC, USA | [12]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA | [13]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, USA | [14]Department of Forest Management, University of Montana, Missoula, Montana, USA | [15]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA | [16]UK Centre for Ecology and Hydrology, Midlothian, UK | [17]Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia | [18]School of Geography and Environmental Science, University of Southampton, Southampton, UK

Correspondence: Matthew A. Worden (mworden@stanford.edu)

ABSTRACT

The rapid increase in the volume and variety of terrestrial biosphere observations (i.e., remote sensing data and in situ measurements) offers a unique opportunity to derive ecological insights, refine process-based models, and improve forecasting for decision support. However, despite their potential, ecological observations have primarily been used to benchmark process-based models, as many past and current models lack the capability to directly integrate observations and their associated uncertainties for parameterization. In contrast, data assimilation frameworks such as the CARbon DAta MOdel fraMework (CARDAMOM) and its suite of process-based models, known as the Data Assimilation Linked Ecosystem Carbon Model (DALEC), are specifically designed for model-data fusion. This review, motivated by a recent CARDAMOM community workshop, examines the development and applications of CARDAMOM, with an emphasis on its role in advancing ecosystem process understanding. CARDAMOM employs a Bayesian approach, using a Markov Chain Monte Carlo algorithm to enable data-driven calibration

of DALEC parameters and initial states (i.e., carbon pool sizes) through observation operators. CARDAMOM′s unique ability to retrieve localized model process parameters from diverse datasets—ranging from in situ measurements to global satellite observations—makes it a highly flexible tool for analyzing spatially variable ecosystem responses to environmental change. However, assimilating these data also presents challenges, including data quality issues that propagate into model skill, as well as trade-offs between model complexity, parameter equifinality, and predictive performance. We discuss potential solutions to these challenges, such as reducing parameter equifinality by incorporating new observations. This review also offers community recommendations for incorporating emerging datasets, integrating machine learning techniques, strengthening collaboration with remote sensing, field, and modeling communities, and expanding CARDAMOM′s relevance for localized ecosystem monitoring and decision-making. CARDAMOM enables a deep, mechanistic understanding of terrestrial ecosystem dynamics that cannot be achieved through empirical analyses of observational datasets or weakly constrained models alone.

1 | Introduction

Terrestrial biogeochemical cycles are an integral part of the earth system, regulating the global carbon cycle, buffering fossil fuel emissions, and providing planetary-scale ecosystem services (Bonan 2008; Friedlingstein et al. 2023, 2024). Yet terrestrial ecosystems—together with human behavior and cloud feedbacks—remain among the most uncertain factors in decadal-to-centennial earth system forecasts (Intergovernmental Panel on Climate Change [IPCC] 2023). Over recent decades, fundamental insights into ecosystem processes have arisen from careful observations and theoretical progress, including leaf-to-planetary-scale observation of photosynthetic processes (e.g., Ryu et al. 2019; Baldocchi 2020), atmospheric greenhouse gas inversions (e.g., Peylin et al. 2013; Basu et al. 2020), and the formulation of novel paradigms of soil–plant-atmosphere biophysical process representations in numerical models (e.g., Bonan 2019). These efforts have produced increasingly complex global vegetation models (Fisher and Koven 2020), but their sophistication often depends on parameterizations that are weakly linked to observations and assumed to be largely invariant in space, limiting model accuracy (Bonan and Doney 2018; Famiglietti et al. 2021). As a result, parameter uncertainty contributes as much to model uncertainty as model structure does (Quetin et al. 2020; Smallman et al. 2021). Consequently, ecosystem observations—and their integration—are crucial for refining and validating models at the planetary scale (Eyring et al. 2019; Schimel and Carroll 2024).

Observations of the terrestrial biosphere have dramatically increased in the last several decades, offering datasets from remote sensing and in situ measurements (e.g., Worden et al. 2021). In situ (i.e., field-based) observing networks offer numerous overlapping complementary observations, including biomass, leaf area, plant traits, and also high temporal frequency information (hourly to daily), such as soil moisture measurements and energy, carbon, and water exchange estimates from eddy covariance (Pastorello et al. 2020). Compiled databases of in situ plant trait information (Kattge et al. 2020) and spatially continuous maps of interpolated soil and vegetation characteristics (Butler et al. 2017; Poggio et al. 2021) are also available. Satellite-based observations (e.g., Smallman et al. 2022; Williams 2022) provide the means to monitor the entire land surface with increasing temporal frequency, spatial granularity, and over an ever-growing time span. Satellite observations can offer insights into the biological and physical states of the land surface with appropriate calibration, thereby complementing in situ observations and enhancing our understanding of ecosystem dynamics (Sellers et al. 2018; Schimel et al. 2019). Despite their value, all observations have limitations in terms of errors, bias, uncertainties, and gaps in space and time, which often remain poorly understood or characterized (Hollinger and Richardson 2005; Loescher et al. 2006; Lasslop et al. 2008; Morrison 2016; Zhao et al. 2020; Araza et al. 2023). Furthermore, in situ measurements are spatially biased toward northern latitudes (Schimel et al. 2015) and have small spatial extents that create representativeness errors when scaling (Chu et al. 2021). These biases can propagate into satellite retrieval algorithms. Finally, satellite observations provide an incomplete picture of terrestrial ecosystem cycling. Observations of different components are typically temporally separate, and many key states (e.g., belowground biomass) and fluxes (e.g., respiration) remain unobservable from space and are challenging to quantify in situ.

Process-based models of the terrestrial biosphere provide a systemic view of terrestrial ecosystems, bridging the gap between observations in space and time. These models represent theories of ecosystem functioning (e.g., Bonan 2019; Ge et al. 2022) and, when combined with observations, allow us to improve our understanding of the Earth system. They also offer a vital capacity to explore climate, disturbance, and management scenarios to predict likely future responses of ecosystem states and services. These "what-if" scenarios have informed IPCC reports (IPCC 2023), underpinning international negotiations aimed at tackling climate change and supporting policymaking. However, most process-based models are weakly constrained by sparse observations from a limited number of sites, primarily using observational data for benchmarking and initializing conditions (e.g., Collier et al. 2018; Seiler et al. 2022; Braghiere et al. 2022). As a result, they often have weakly constrained parameters and lack uncertainty estimates (Hourdin et al. 2017; Bonan and Doney 2018; Famiglietti et al. 2021). Moreover, without adequate calibration, rigorous evaluation of competing hypotheses embedded in process-based models becomes impossible (e.g., Fisher and Koven 2020; Famiglietti et al. 2021; Smallman et al. 2021; Norton et al. 2023).

Multiple studies have hypothesized that the inability to effectively integrate observations with current models underpins much of the divergence in carbon cycling forecasts (e.g., Lovenduski and Bonan 2017; Lovenduski et al. 2019; Arora et al. 2020; Bonan et al. 2021; Bonan et al. 2024; Smallman et al. 2021). This hypothesis is a key motivation for advancing model-data

fusion approaches, which can enhance the representation of complex environmental dynamics and reduce uncertainties in ecosystem and climate projections (Schimel and Carroll 2024). Moreover, integrating models with observations allows for better characterization, explanation, and communication of uncertainties in both models and data (e.g., Keenan et al. 2011; Lahoz and Schneider 2014; Yin et al. 2020). The primary challenge lies in integrating observations with models in a manner that effectively informs the underlying processes (Schneider et al. 2017). This challenge has stimulated the development of a number of different model-data fusion approaches (e.g., Bloom et al. 2016; Scholze et al. 2016; Fox et al. 2018; Luo and Smith 2022). These approaches and their associated tradeoffs are discussed in more detail in Section 3.

Here, we review the CARbon DAta MOdel fraMework (CARDAMOM), a model-data fusion framework developed over 20 years across multiple institutions that draws on spatially and temporally explicit observations and ecological knowledge to infer ecosystem states, process parameterizations, and biogeochemical dynamics (Bloom and Williams 2015; Bloom et al. 2016). CARDAMOM's unique ability to infer parameters from observations across varied spatial and temporal scales enables a plant functional type–agnostic approach that captures ecosystem functional gradients from localized studies to broader regional and global analyses (e.g., Myrgiotis et al. 2021; Yang et al. 2022; Bloom et al. 2016). First, we describe the CARDAMOM framework and provide examples of its applications, showcasing how the integration of models and observations has been used to derive insights that neither could achieve independently. Second, we highlight CARDAMOM's unique capacities by contrasting them, where appropriate, with those of other terrestrial ecosystem modeling approaches that assimilate data. Third, we discuss the challenges involved in working with the CARDAMOM system. Finally, we explore future opportunities—including machine and/or deep learning techniques, novel satellite observations, and emergent constraints derived from in situ observing networks—to further enhance ecosystem analyses, extract ecological insights, support robust model development, and overcome limitations.

2 | Overview, Historical Development, and Applications

2.1 | Framework Overview

The CARDAMOM framework consists of three main components: (i) a process-based model of the terrestrial ecosystem, called the Data Assimilation Linked Ecosystem Carbon Model (DALEC) (Figure 1, green box); (ii) observation operators that link the process-based model to ecosystem measurements (Figure 1, blue box); and (iii) a Bayesian inference algorithm that integrates observations and their uncertainties, prior knowledge of parameter values or ranges, and prior knowledge of ecological function (referred to as ecological and dynamical constraints; hereafter EDC) to constrain model states and process parameters (Figure 1, purple box; Bloom et al. 2016; Bloom and Williams 2015; Yang et al. 2022). Working with CARDAMOM begins with a model-data fusion step (Figure 1), during which the three main components iteratively retrieve parameters and

initial states that align DALEC with observations and their associated uncertainties. Once CARDAMOM has retrieved ensembles of parameters consistent with the observations, these parameter ensembles can be used by DALEC in a forward run to generate data-constrained ensembles of states and fluxes (Figure 1, orange box). CARDAMOM is implemented in a modular manner so that the process-based model, observation operators, and Bayesian inference algorithm can be swapped for alternative approaches and are distinct enough for users to use each part individually.

2.1.1 | The Process-Based Model DALEC as the Backbone of CARDAMOM

CARDAMOM uses a suite of intermediate complexity process-based models collectively known as DALEC. The DALEC suite represents terrestrial ecosystems with a pool-based mass balance approach, where carbon, water, and energy states evolve through parameterized ordinary differential equations (Williams 2022). Its relatively simple structure makes it straightforward to develop and implement new theoretical approaches as new observations become available to test competing hypotheses (e.g., Famiglietti et al. 2021). Currently, more than a dozen published model versions exist, each featuring a different level of complexity and reflecting various theoretical assumptions about ecosystem functioning. These versions include different representations of photosynthesis (Quetin, Famiglietti, et al. 2023; Williams et al. 2005; Smallman and Williams 2019), hydrological processes (Spadavecchia et al. 2011; Bloom et al. 2016; Smallman and Williams 2019; Yang et al. 2022), respiration processes (Rowland et al. 2014; Quetin, Famiglietti, et al. 2023; Yin et al. 2020; Ma et al. 2023), and phenological processes (Smallman et al. 2017; Exbrayat et al. 2018; Norton et al. 2023; Bloom et al. 2020). Users select the most appropriate model from the DALEC suite and modify it as needed based on their research questions and available data, a process that also allows different model structures to be compared and evaluated against observations (e.g., Famiglietti et al. 2021; Worden et al. 2024). Although the DALEC models are the only process-based models integrated into CARDAMOM, other models could be incorporated by adapting them to accept input parameters and initial states from CARDAMOM's Markov Chain Monte Carlo (MCMC) sampler and to produce output states and fluxes comparable to observations via CARDAMOM's observation operators (Figure 1).

DALEC is driven by environmental forcing data. Forcing data varies depending on the model version but typically includes meteorological variables and exogenous processes. Examples of meteorological forcings include minimum and maximum air temperatures, incoming short-wave radiation, atmospheric CO_2 concentration, vapor pressure deficit, precipitation, and wind speed. Exogenous processes typically encompass disturbance or management activities such as fire (Exbrayat et al. 2018) or biomass extraction from forests (Smallman et al. 2017). Using these forcings, along with the parameters and initial states provided by the MCMC sampler, DALEC generates states and fluxes that can be compared to observations through an observation operator (Figure 1).
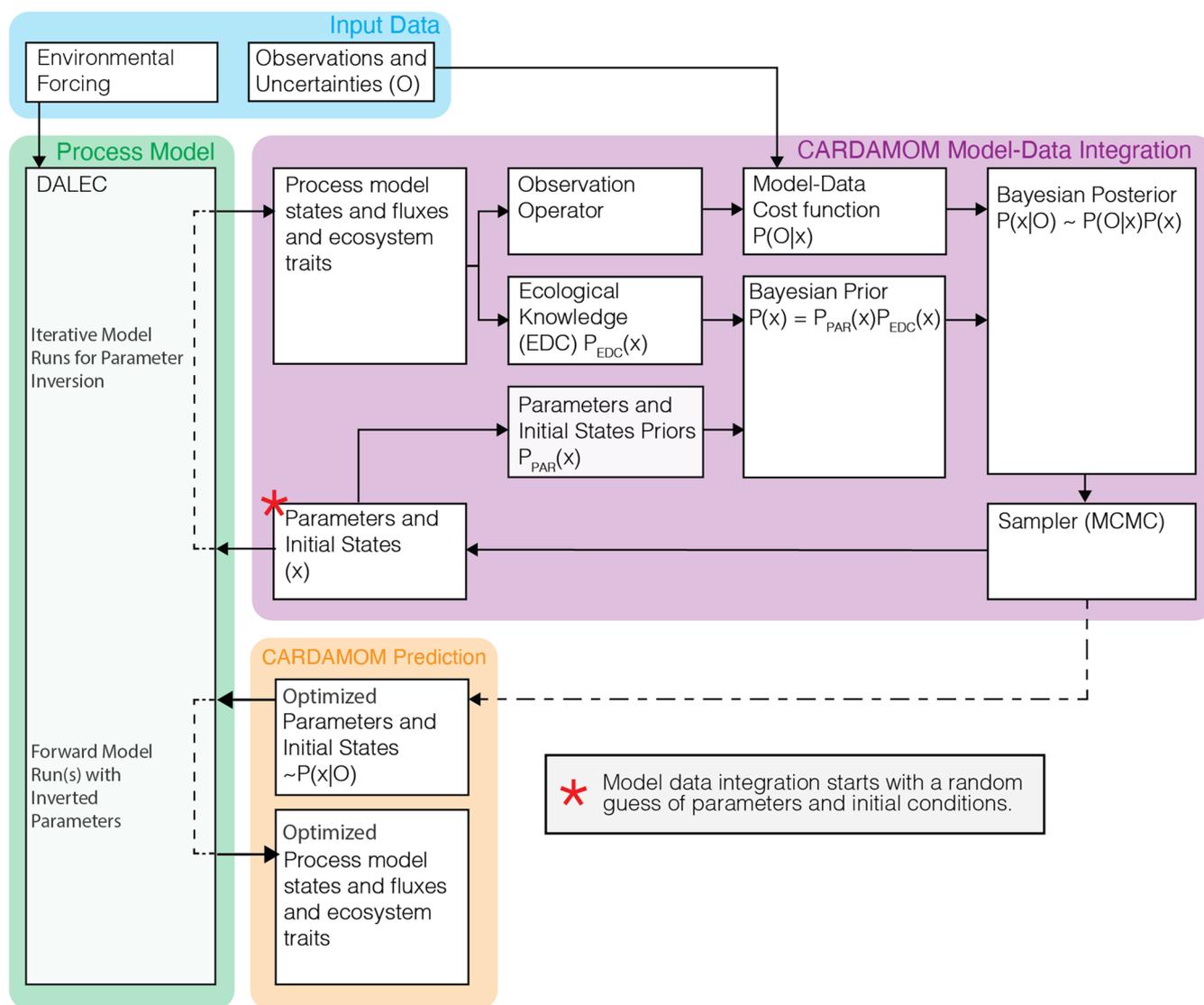
FIGURE 1 | Illustration of the CARbon DAta MOdel FraMework (CARDAMOM), describing its inputs (blue), its process-based model (green), the parameter and initial state optimization process (purple), and its outputs using optimized parameters (orange).

2.1.2 | Connecting Model States and Fluxes to Varied Observations With Flexible Resolution

CARDAMOM can assimilate any observation type, provided that a suitable observation operator (Figure 1) is developed to link the observation to DALEC states or fluxes. Possible observations include in situ measurements such as net and gross fluxes of gases (CO_2, water, and methane), biomass, foliage, leaf area index, fine roots, litter, and soil carbon (Smallman et al. 2017; Famiglietti et al. 2021; Yang et al. 2022; George-Chacon et al. 2023; Worden et al. 2024). CARDAMOM is equally versatile in assimilating satellite-based observations, including leaf area index, equivalent water thickness, solar-induced fluorescence, and net biosphere exchange (Bloom et al. 2020). Additionally, CARDAMOM's observation operators can be tailored to specific comparisons, such as linking solar-induced fluorescence with DALEC gross primary productivity or separating the monthly and interannual variability of net biosphere exchange (Quetin et al. 2020).

The temporal and spatial resolution at which CARDAMOM is applied is flexible. CARDAMOM has been applied at daily

(Myrgiotis et al. 2021; Revill et al. 2021), weekly (Rowland et al. 2014; Myrgiotis et al. 2020), and monthly (Bloom et al. 2016; Yang et al. 2022; Au et al. 2023) time steps. Spatial resolution varies widely, from multiple degree-scale grids (Bloom et al. 2016; Quetin, Famiglietti, et al. 2023) down to hectares and tens of meters (Stettz et al. 2022; Au et al. 2023; Myrgiotis et al. 2022).

CARDAMOM's flexible temporal resolution, observation operators, and optimized parameters and initial states enable it to reproduce ecosystem responses across multiple timescales. The model's temporal resolution (timestep) depends on the chosen granularity of driver datasets; however, observation operators allow for the assimilation of aggregated periods that extend beyond those driver timesteps. For example, CARDAMOM has assimilated long-term mean leaf area index and annual net biosphere exchange values (Bloom et al. 2020; Quetin et al. 2020), despite operating at a monthly timestep in the cited examples. This flexibility allows CARDAMOM to capture ecological dynamics across various temporal scales. At daily resolution, CARDAMOM has differentiated ecosystem responses to grazing versus cutting practices (Myrgiotis et al. 2021). At monthly

timesteps, it has captured ecosystem carbon dynamics, including net ecosystem exchange, when validated against withheld FLUXNET observations (Yang et al. 2022). At interannual timescales, it has accurately reproduced variability in net biosphere exchange and atmospheric CO_2 growth rates (Bloom et al. 2020; Worden et al. 2024; Bilir et al. 2025). CARDAMOM has only been applied once at centennial scales using monthly timesteps (Quetin, Famiglietti, et al. 2023), and not without some difficulty (computational burden; lack of constraints prior to the 2000s). Lastly, a key downside is that CARDAMOM's parameters are time invariant, which can limit its ability to capture changes in ecosystems due to land use and land cover change (Milodowski et al. 2023) or the acclimation of plant traits to shifting climate conditions (Cui et al. 2020; Quetin, Anderegg, et al. 2023; Famiglietti et al. 2024).

CARDAMOM retrieves spatially localized DALEC parameters and initial conditions (e.g., wood pool value on first time step) based on the observations it assimilates. When applied at site scale (e.g., flux towers), separate CARDAMOM model-data fusions are performed at each site to retrieve location-specific parameters (e.g., Yang et al. 2022). Similarly, CARDAMOM can be applied across spatially contiguous domains, with model-data fusion performed independently for each grid pixel, producing parameters and initial states specific to each location (e.g., Bloom et al. 2016). This approach allows parameters to vary across ecological gradients, capturing functional diversity as reflected in observations (e.g., Bloom et al. 2016; Smallman et al. 2017; Yin et al. 2020; Exbrayat et al. 2018; Smallman et al. 2021). Hence, unlike many large-scale models, CARDAMOM avoids reliance on Plant Functional Types (hereafter PFTs) to define vegetation function. This is advantageous because PFTs capture only a fraction of plant trait variability and fail to account for significant intra-PFT variability (Van Bodegom et al. 2012; Joswig et al. 2021; Konings et al. 2024).

2.1.3 | CARDAMOM Bayesian Model-Data Fusion

CARDAMOM retrieves information on DALEC parameters and initial states based on assimilated observations, taking into account their quantity, type, and uncertainty (Smallman et al. 2017; Quetin et al. 2020; Williams 2022). A key feature of CARDAMOM is its explicit handling of uncertainty, treating parameters and outputs as random variables represented by probability distributions using a Bayesian approach. This incorporates both prior ecological knowledge and observational constraints to retrieve optimal parameter ensembles, following Bayes' Theorem:

$$p(x \,|\, O) \propto p(x)p(O \,|\, x)$$

Where $p(x|O)$ is the posterior probability distribution of a DALEC parameter vector x, given a set of observations O. The $p(x)$ represents the prior knowledge of the parameter vector x, and $p(O|x)$ is the likelihood of the observations given the parameter ensemble.

The likelihood $p(O|x)$ is calculated using a cost function that compares observations (O) to their corresponding model state variables or fluxes generated from parameters x, scaled by an uncertainty term that accounts for both observational and model structural uncertainty (e.g., Bloom et al. 2016; Levine et al. 2025). In CARDAMOM, priors $p(x)$ are divided into parameter priors $p_{par}(x)$ and ecological dynamic constraint (EDC) priors $p_{EDC}(x)$ (Figure 1), which integrate previously gathered information (e.g., past studies, literature, expert knowledge) into the optimization process. Parameter priors $p_{par}(x)$ are typically defined as uniform or log-uniform distributions with bounds that exclude improbable or impossible values (e.g., negative turnover rates) but can be further constrained if parameter values are already known. EDC priors $p_{EDC}(x)$ embed ecological theory into model-data fusion (Bloom and Williams 2015), reducing the parameter space by rejecting or penalizing parameter combinations that produce outcomes inconsistent with ecological principles. Commonly applied EDC priors include constraints on carbon allocation (e.g., balancing foliage and fine root allocation), turnover rates (e.g., slower turnover for soil organic matter compared to litter), and biomass dynamics (e.g., preventing exponential decay in foliage biomass). The number and type of EDC priors vary by model. EDC priors are removed when alternate structures reduce net dimensionality and are often introduced when new processes are added to DALEC.

Given the multidimensional parameter space in a DALEC model, solving the full Bayes equation analytically is impractical. Instead, numerical algorithms approximate the posterior distribution. CARDAMOM typically uses Markov Chain Monte Carlo (hereafter MCMC)-based algorithms (e.g., Yang et al. 2022) but has also applied other samplers, such as simulated annealing (Myrgiotis et al. 2022). MCMC algorithms used in CARDAMOM include the Metropolis-Hastings algorithm (Yang et al. 2022) and the Differential Evolution MCMC algorithm (Braak 2006; Worden et al. 2024). A key advantage of MCMC-based parameter estimation approaches is that they quantify parametric uncertainty by generating a posterior distribution of parameters that reflects their likelihood given the data and model structure (Hill et al. 2012; Bloom et al. 2016).

2.2 | History of the CARDAMOM Framework: Designed for Data Assimilation

CARDAMOM and its design philosophy originated from efforts to simplify the soil–plant-atmosphere continuum model SPA (Williams et al. 1996), aligning it with the satellite observations and computational capabilities available at the time. This led to the development of an aggregate canopy model for gross primary productivity (Williams et al. 1997) and later the DALEC model (Williams et al. 2005). DALEC was designed to simulate carbon cycle dynamics in a manner optimized for data assimilation approaches (initially an ensemble Kalman filter) and to leverage the burgeoning availability of satellite observations (Williams et al. 2005). Building on DALEC, CARDAMOM was developed to explicitly infer parametric links between carbon model structures and observational data (Bloom and Williams 2015; Bloom et al. 2016; Quetin et al. 2020; Famiglietti et al. 2021; Smallman et al. 2021).

CARDAMOM and DALEC are continuously developed across multiple institutions (currently the University of Edinburgh, Jet Propulsion Laboratory, Stanford University, Columbia
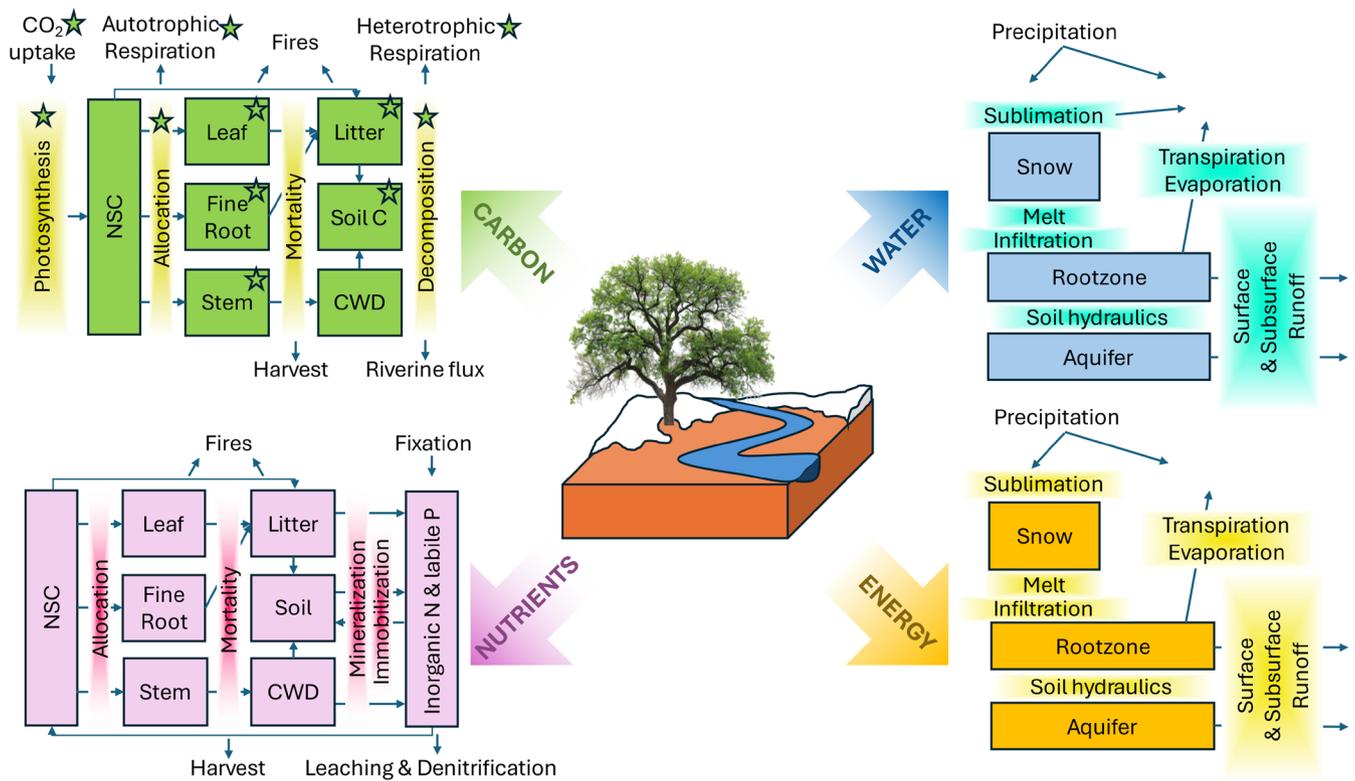
FIGURE 2 | Schematic illustrating the range of DALEC ecosystem modules explored within the CARDAMOM community. The four panels depict the land surface carbon cycle (green, upper-left), water cycle (blue, upper-right), nutrient cycle (purple, lower-left), and energy cycle (yellow, lower-right). Within the respective carbon, water, nutrient, and energy cycles, boxes denote state variables, and arrows denote fluxes. No single DALEC variant currently integrates all four cycles at the maximum level of complexity; instead, different versions combine selected modules and implement them in distinct ways to address specific scientific questions. Green stars mark the DALEC's original scope, which covered only a subset of the carbon cycle (green panel, upper left).

University, University of California Davis, and University of California Santa Barbara) to address emerging scientific challenges and incorporate an expanding range of observational constraints. Over time, new DALEC versions have been developed, spanning a range of complexities and process representations—from carbon cycle models (e.g., Williams et al. 2005) to those incorporating hydrological, nutrient, and/or energy cycles (e.g., Wang et al. 2018; Levine et al. 2025; Bilir et al. 2025)—creating a suite of DALEC models (Figure 2). Concurrently, the number and type of observations assimilated by CARDAMOM have increased (Bloom et al. 2016; Famiglietti et al. 2023).

2.3 | Applications to Extract Ecological Knowledge and Insight Into Model and Data Development

The CARDAMOM framework can be used for inferring ecosystem process information and for testing climate and management scenarios constrained by data. By assimilating localized observations, CARDAMOM can infer PFT-agnostic and location-specific ecosystem processes and traits at regional to global scales where direct measurement is infeasible. This capability has been leveraged in several ways, including for inferring carbon allocation, stocks, and turnovers (Bloom et al. 2016; Richardson et al. 2010; Rodríguez-Veiga et al. 2020), changes in carbon allocation under drought (Au et al. 2023; Worden et al. 2024), management practices in grasslands (Myrgiotis et al. 2021), arable crop development and harvest (Revill

et al. 2021), and temperature limitations on spring productivity in conifer forests (Stettz et al. 2022). Beyond capturing current ecosystem dynamics, CARDAMOM can also serve as a data-constrained prognostic tool, simulating ecosystem responses to variables such as climate, CO_2 levels, and disturbances. This is achieved by retrieving DALEC's parameters with present-day observations and then running DALEC with modified forcing data or parameter distributions. This approach has been used to assess how declining fires in tropical ecosystems enhance the land carbon sink (Yin et al. 2020), explore lagged or legacy effects on carbon fluxes (Bloom et al. 2020; Worden et al. 2024), quantify carbon-climate feedbacks in Alaskan wetlands (Ma et al. 2023), and partition the impacts of elevated CO_2 versus climate change on forests in the Yucatán, Mexico (George-Chacon et al. 2023), among others. These perturbation studies can be applied to both future and past projections. For example, one study quantified the impact of climate and CO_2 fertilization over the past century on carbon sink distribution and found that respiration, a location-specific optimized parameter, strongly influences whether a location acted as a source or sink (Quetin, Famiglietti, et al. 2023).

The design philosophy of CARDAMOM also makes it beneficial for model benchmarking. CARDAMOM's dual functionality allows it to be evaluated as a model (Yang et al. 2022; Quetin et al. 2020) while also serving as a coherent benchmark target when observations are unavailable, such as for permafrost soil residence times (Hugelius et al. 2024)
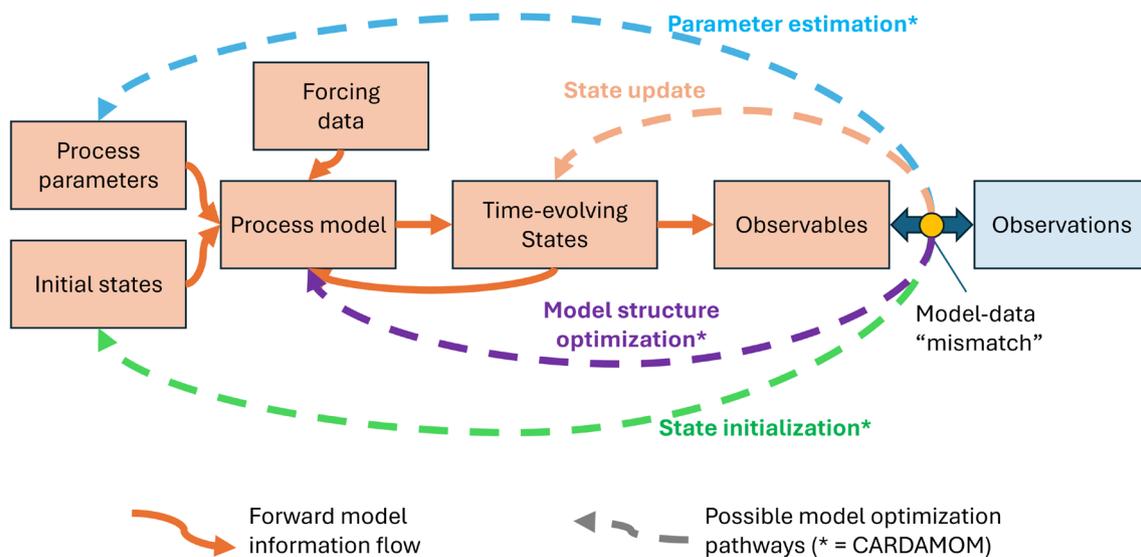
FIGURE 3 | Process-based model forward model information flow (orange solid arrows), linking parameters, initial conditions, and model forcing to observables. Model-data fusion approaches propagate information from a model-data "mismatch" assessment (e.g., a cost function and/or a probability evaluation) back to the model's configuration, through one or more model optimization "pathways" (dashed arrows). In contrast to other approaches, CARDAMOM optimizes time-invariant process parameters and initial conditions on a local (point or gridscale) basis. See Table 1 for comparison of the CARDAMOM model-data fusion approach relative to other state-of-the-art approaches.

or transit time distributions (Sierra et al. 2023). Alternative model structures can easily be compared to test hypotheses on process representation and its impact on modeled ecosystem sensitivity and predictive capacity (e.g., Famiglietti et al. 2021; Norton et al. 2023; Worden et al. 2024; Smallman et al. 2021). CARDAMOM outputs also provide valuable benchmark targets for other models by synthesizing multiple observation streams into an ecologically coherent representation of terrestrial ecosystems. This approach provides internally consistent estimates of carbon and water fluxes, stocks, and their associated uncertainties (Caen et al. 2022; Levine et al. 2023; Massoud et al. 2022). This capability is particularly important since different observational data streams may not be consistent with each other when compared naively. Examples of benchmarking include the use of CARDAMOM outputs in the international land model benchmarking (ILAMB) system (Collier et al. 2018; Friedlingstein et al. 2024), the validation of the Joint UK Land Environment Simulator (JULES) and Brazil's Integrated Land Surface Model (INLAND) land surface models across the Amazon (Slevin et al. 2017; Caen et al. 2022), and the validation of soil heterotrophic respiration in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Varney et al. 2022).

Finally, CARDAMOM can generate data-constrained outputs of specific carbon fluxes that inform broader analyses and serve as a testbed for experiments aimed at improving model formulation and model-data fusion. CARDAMOM outputs have been used to produce global satellite-driven estimates of heterotrophic respiration (Konings et al. 2019) and provide prior ranges for atmospheric inversion estimates of net biosphere exchange (Liu et al. 2021). Examples of CARDAMOM as a testbed for models and model-data fusion include assessing the impact of different analytical solvers, such as the Kalman filter versus genetic algorithms (Fox et al. 2009, a pre-CARDAMOM DALEC-based analysis), comparing model parameterization based on

plant functional types to those derived from environmental filtering approaches (Famiglietti et al. 2023), and demonstrating how parametric error can dominate uncertainty in forecasts of carbon stock changes across Brazil (Smallman et al. 2021).

3 | CARDAMOM's Spatially Explicit Parameters and Initial State Estimation Are Unique Among Data Assimilation Frameworks

CARDAMOM is part of a growing set of frameworks that integrate process-based models with ecosystem observations (Kaminski and Mathieu 2017). Across these frameworks, information generally flows in a similar pattern (Rayner et al. 2019; Figure 3). First, time-invariant parameters and initial states drive a terrestrial biosphere model. Next, observation operators sample time-varying states to generate "observables," which are compared with actual ecosystem observations. Finally, any mismatch between these observables and observations is used to refine the process-based model. Model-data fusion typically follows one or more of four pathways: (1) sequential state update, (2) time-invariant parameter estimation, (3) initial state estimation, and (4) model structure optimization (Figure 3).

The four pathways above offer various ways researchers can design frameworks (Table 1) to capture terrestrial biosphere processes accurately, incorporate observed ecosystem complexity, narrow parameter uncertainties, and produce data-informed predictions (Luo et al. 2011; Williams et al. 2009; Li et al. 2024). State-update approaches, such as the Community Land Model–Data Assimilation Research Testbed (CLM-DART; Fox et al. 2018, 2022), sequentially adjust model states to match observations but do not refine model parameters, potentially violating mass conservation. In contrast, parameter retrieval approaches enhance ecological learning by estimating time-invariant parameters that govern model processes.

TABLE 1 | Comparison of the CARDAMOM model-data fusion approach relative to other state-of-the-art systems which integrate models and data.

Framework	Model-data fusion algorithm or approach	Temporal and spatial resolution	State Update/initial state estimation	Process parameter estimation	Structure optimization	Prognostic with forcing data?	ESM integration
CLM-DART	Bayesian	Diurnal, gridded	State Update	No	No	No	No
CCDAS (BETHY, SINDBAD, etc....)	Bayesian, 4Dvar	Diurnal, gridded	No	Yes; PFT-wide	No	Yes	Yes
ORCHIDEE-DA	Gradient descent, genetic algorithm	Flexible	No	Yes; PFT-wide	No	Yes	Yes
PEcAn	Bayesian	Flexible	State Update	Yes; PFT-wide	No	Yes	No
CliMA-Land	Ensemble Kalman Filter	Flexible	Both	Yes; PFT-wide	No	Yes	Yes
DifferLand	Hybrid Model with gradient-based optimizer	Flexible	Initial only	Yes	Yes	Yes	No
CARDAMOM	Bayesian (MCMC)	Flexible	Initial only	Yes	Yes[a]	Yes	No

[a]Depending on how "structure" is defined, CARDAMOM can optimize model structure, such as retrieving different water stress or allocation functions using parametric equations (Levine et al. 2023; Worden et al. 2024).

These retrievals can be spatially integrated to estimate parameters at the plant functional type level, as in the Carbon Cycle Data Assimilation System (CCDAS; Wu et al. 2020; Scholze et al. 2016; Norton et al. 2018), Organising Carbon and Hydrology In Dynamic Ecosystems Data Assimilation (ORCHIDEE-DA; Kuppel et al. 2014; MacBean et al. 2022), and the Predictive Ecosystem Analyzer (PEcAn; LeBauer et al. 2013; Dokoohaki et al. 2021). They can also be spatially explicit, with parameters estimated locally, as in the Forest Biomass Assimilation and Regression (FöBAAR; Keenan et al. 2013), the Diagnostic Carbon Flux Model (DCFM; Xiao et al. 2014), and the Vegetation Photosynthesis and Respiration Model (VPRM; Mahadevan et al. 2008). The Climate Modeling Alliance (CliMA) Land, a more recent framework (Braghiere, Fisher, et al. 2021; Braghiere, Wang, et al. 2021; Wang et al. 2023), estimates both time-invariant parameters and initial states while also updating states. However, each of these approaches either operates only over smaller spatial extents when data assimilation is performed, or if run at a global scale, solves for PFT-wide parameters. Trade-offs between model complexity, assimilation methodology, data availability, and computational constraints can make each framework better suited to specific scientific or practical applications.

CARDAMOM offers spatially explicit, PFT-agnostic parameter and initial state optimization, carving out a distinct niche in model-data fusion. While previous frameworks have explored spatially explicit parameter optimization in situ (e.g., Keenan et al. 2013; Xiao et al. 2014; Raj et al. 2018), CARDAMOM is, to our knowledge, the only framework operationalized on a gridded, global scale that retrieves both parameters and initial states. This spatially explicit approach enables independent estimation of process parameters like photosynthetic capacity, leaf mass per area, allocation fractions, and drought sensitivities alongside corresponding initial states like initial foliar biomass. By retrieving parameters across space rather than relying on predefined PFT-level values (e.g., Bloom et al. 2016), CARDAMOM captures ecosystem functional gradients often overlooked by spatially integrated, PFT-based approaches (e.g., Exbrayat et al. 2018). The combined state–parameter estimation paradigm, along with ecological and dynamic constraints (Bloom and Williams 2015; see Section 2.1.3), also provides a pathway to bypass multidecadal-to-millennial spin-up times (e.g., Fang et al. 2015), which are both computationally expensive and depend on the assumption of initial equilibrium that may not be valid (Shi, M. et al., 2013).

CARDAMOM's ability to use data to constrain nonequilibrium initial conditions and spatially explicit parameters without relying on PFTs is unique relative to other data assimilation methods (Table 1). An exception is the recent DifferLand model (Fang and Gentine 2024). Built using DALEC and EDCs from CARDAMOM v2.3, DifferLand employs a hybrid physical–machine learning methodology that extends CARDAMOM's ability to optimize model structure by embedding neural networks within DALEC to replace existing empirical formulations. This includes, for example, replacing the soil water stress function used in gross primary productivity calculations with a neural network. The potential for hybrid and machine learning approaches to further extend CARDAMOM's capabilities is discussed in Section 5.4.

4 | Model-Data Fusion Challenges: Data Quality, Model, and Computation

4.1 | Data Quality Propagates Into Inference

CARDAMOM relies on documented uncertainties, which are input alongside their corresponding observations (Section 2.1.3). However, it is often unclear how uncertainties in, for example, large-scale dataset products should be scaled spatially and temporally when aggregated and used with CARDAMOM. Additionally, satellite and ecological network data are biased toward readily measurable features like canopy states, while biomass and dead organic matter—especially belowground components—often lack direct observational constraints. The uncertainty in these unobserved state variables is potentially substantial and can contribute significantly to forecast error within CARDAMOM (Smallman et al. 2021).

The interaction between these potentially large systematic errors in datasets, and inevitable structural errors in the model introduces a significant risk of biases being pushed into parameters that are less closely constrained by high-density observation streams (Cameron et al. 2022). Attribution of changes in the carbon cycle to specific driving processes (and thus the parameterization of these processes) is further hindered by an incomplete view of the drivers (Exbrayat et al. 2018). For instance, forest degradation (Ryan et al. 2012) and herbivory (Metcalfe et al. 2014; Hempson et al. 2015) are major drivers of carbon turnover in tropical ecosystems; yet their extent and intensity are not as well mapped as deforestation (Hansen et al. 2010) or burned area (Andela et al. 2017).

4.2 | Parameter Equifinality and Model Complexity

Equifinality is a challenge for all data assimilation approaches (Beven and Freer 2001). For a given set of observations, multiple combinations of model parameters and structures can produce simulations that fit ecosystem observations equally well (Famiglietti et al. 2021; Smallman et al. 2021). This limits model identifiability and becomes more pronounced as model complexity increases. While greater complexity (i.e., incorporating more processes) may enhance a model's ecological realism, without a corresponding increase in data constraints, parameter uncertainty rises, and forecast skill may decline (Famiglietti et al. 2021). Consequently, the guiding philosophy behind the development of DALEC models within CARDAMOM has been to adopt parsimonious, intermediate complexity models commensurate with available observations (Williams et al. 2005).

Developing and applying more complex models requires balancing ecological realism with the availability of observations needed to robustly characterize processes at relevant spatial and temporal scales. For example, dynamically partitioning heterotrophic respiration into aerobic and anaerobic components—important for understanding wetland carbon dynamics—requires appropriate observational constraints, such as methane emissions (Ma et al. 2023). Expanding the number and variety of observations assimilated by CARDAMOM is essential for improving model process representation and accuracy. However, the spatial and temporal resolution of these observations must align with model complexity. Coarser spatial scales and longer time steps (e.g., daily vs. monthly) may lack the necessary detail to constrain model processes effectively.

4.3 | Temporal and Spatial Scale Mismatches Reduce Interpretability in Data Assimilation

The effectiveness of CARDAMOM is fundamentally shaped by the temporal and spatial scales of the assimilated data. Temporal aggregation affects the process signatures embedded in observations and influences how environmental drivers are characterized and propagated through CARDAMOM's land surface representation (Jenkins and Watts 1968; Grieve et al. 2016). For example, submonthly temporal resolution may be required to resolve drivers of interannual variability, such as the onset timing of the wet season (Camberlin and Okoola 2003), or short-term management interventions (e.g., Myrgiotis et al. 2021).

Spatial resolution limitations present additional challenges. Coarse spatial scales average across diverse ecosystem types, degrading ecological information in retrieved parameters and potentially introducing bias into diagnostic analyses (Milodowski et al. 2023). In contrast, finer-scale observations can resolve distinct processes driving ecosystem function, such as land management and disturbance (Fisher and Koven 2020). Environmental change hotspots (Alencar et al. 2015; Brinck et al. 2017) and managed, mixed-use areas (Myrgiotis et al. 2022; Milodowski et al. 2023) are often characterized by heterogeneous, fragmented landscapes. The inability to resolve key management processes (e.g., arable harvest) has been identified as a factor contributing to discrepancies between bottom-up terrestrial carbon balance estimates from land surface models and top-down estimates from atmospheric inversions (White et al. 2019; Kondo et al. 2020), such as those often used to constrain CARDAMOM.

The spatial resolution of a CARDAMOM analysis is limited by its coarsest dataset. CARDAMOM assimilates data types with widely varying spatial resolutions—for example, net biosphere exchange derived from atmospheric inversions typically has a resolution > 100 km, while leaf area index or biomass data may have resolutions several orders of magnitude finer (10–500 m) and must be aggregated to use alongside net biosphere exchange. As a result, there is a trade-off between the spatial resolution needed to address a given research question and the dimensionality of available observational constraints, linking back to the equifinality challenge outlined in Section 4.2. An outstanding challenge is to resolve the ecological structure of heterogeneous landscapes while maximizing the use of available observation constraints from a multiscale satellite-based observation network. Besides acquiring higher-resolution observations, addressing this limitation could involve refining existing aggregation methods (e.g., stratifying assimilated data based on land use distributions; Milodowski et al. 2023) or developing new approaches, such as hierarchical models

or novel tiling techniques (e.g., Hill et al. 2011), that enable cross-scale integration of observations in CARDAMOM's likelihood calculation.

4.4 | Computational Cost

CARDAMOM currently uses MCMC-based algorithms, which can require significant computational effort. Each CARDAMOM location or pixel requires 10^5–10^9 model simulations to reach convergence. This computational cost is mitigated by DALEC's intermediate complexity, for which individual simulations are computationally inexpensive (Bloom and Williams 2015; Bloom et al. 2016; Famiglietti et al. 2021). As a result, single-site analyses can be performed on a standard workstation. Scaling up to large spatial domains, however, requires high-performance computing (HPC) or cloud-based architectures to run CARDAMOM at each point concurrently, introducing tradeoffs in the spatial and temporal resolution of regional to global analyses.

5 | Community Recommendations

5.1 | Incorporating New Observations and Processes

CARDAMOM's flexibility in assimilating different observation streams and its adaptable model structure make it well-suited to leveraging the expanding array of ecosystem observations from space (Figure 4). Globally, it can already assimilate datasets on net biosphere exchange (e.g., NASA Carbon Monitoring System Flux [CMS-Flux]; Liu et al. 2021), equivalent water thickness (e.g., Gravity Recovery and Climate Experiment [GRACE]; Watkins et al. 2015), fire carbon emissions (e.g., Measurements of Pollution in the Troposphere [MOPITT]; Jiang et al. 2017; Global Fire Emissions Database [GFED]; Giglio et al. 2013), soil organic carbon stocks (e.g., SoilGrids v2; Poggio et al. 2021; Harmonized World Soil Database [HWSD]; Hiederer et al. 2011), leaf area index (e.g., Moderate Resolution Imaging Spectroradiometer [MODIS]; Myneni et al. 2015), methane emissions (e.g., Wetland and Wetland CH_4 Emissions Model Intercomparison Project [WetCHARTs]; Ma et al. 2021), aboveground biomass (e.g., Xu et al. 2021; ESA Climate Change Initiative Biomass [ESA-CCI Biomass]; Santoro and Cartus 2024), and solar-induced fluorescence (e.g., Joiner et al. 2018; Sun et al. 2018).

A growing number of recent and forthcoming observational platforms offer opportunities to provide additional constraints for CARDAMOM (Table 2). These include the NASA–ISRO Synthetic Aperture Radar (NISAR; Kellogg et al. 2020), BIOMASS (Quegan et al. 2019), the Global Ecosystem Dynamics Investigation (GEDI; Dubayah et al. 2023), and the Radar Observing System for Europe L-band (ROSE-L; Davidson and Furnell 2021) missions. Their emphasis on calibration, error characterization, and repeated aboveground biomass estimates will enhance CARDAMOM's ability to capture biomass growth and disturbance dynamics. Additionally, the Copernicus Anthropogenic Carbon Dioxide Monitoring mission (CO2M; Noël et al. 2024) will provide new estimates

of column CO_2 and methane, enabling surface net carbon flux derivations that can be used to constrain CARDAMOM. CO2M, along with the FLuorescence EXplorer (FLEX; Moreno 2021), will also contribute new solar-induced fluorescence measurements. The recent Surface Water and Ocean Topography (SWOT) mission (Fu et al. 2024) is expected to deliver global river discharge estimates (Durand et al. 2023), which, on a monthly scale, can be converted to runoff maps (Pan and Wood 2013; David et al. 2019; Collins et al. 2024). These runoff maps could help constrain spatial and temporal runoff variability in CARDAMOM, which has historically been difficult to validate.

Hyperspectral satellite data offer new opportunities to constrain leaf properties. These constraints can improve modeling of phenology, gross primary productivity, and autotrophic respiration. Current observations of foliar nitrogen content from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS; Townsend et al. 2023), the Earth Surface Mineral Dust Source Investigation (EMIT; Green et al. 2020), and the Plankton, Aerosol, Cloud, ocean Ecosystem mission (PACE) will be complemented by the upcoming Copernicus Hyperspectral Imaging Mission for the Environment (CHIME; Rast et al. 2021) and Surface Biology and Geology (SBG; Schimel and Poulter 2022) missions (Table 2). When combined with already assimilated observations, these measurements can refine theoretical models linking ecosystem carbon and nitrogen cycles (e.g., Thomas and Williams 2014; Flack-Prain et al. 2019; Thomas et al. 2019). Hyperspectral data also enhance representations of leaf lifespan via leaf mass per area and photosynthetic capacity via leaf chlorophyll content. Additionally, hyperspectral signals correlate with mycorrhizal fungi (Sousa et al. 2021), which play a key role in plant nutrient acquisition and responses to CO_2 fertilization (Braghiere, Fisher, et al. 2021; Braghiere, Wang, et al. 2021; Braghiere et al. 2022). These fungi form symbiotic relationships with most plants, enhancing their nutrient acquisition—particularly nitrogen and phosphorus—which is essential for plant growth (Shi et al. 2023). These data present an opportunity to incorporate fungal dynamics into CARDAMOM, advancing our understanding of their influence on nutrient cycling and plant productivity.

Beyond exploiting the growing catalogue of remote sensing observations, CARDAMOM excels at extracting ecological insights from a broad range of observational networks and experimental platforms. For example, forest chronosequence data can constrain biomass accumulation rates in DALEC (Smallman et al. 2017; George-Chacon et al. 2023). Forest inventory data—such as from the Red Amazónica de Inventarios Forestales (RAINFOR) network (Malhi et al. 2002)—provide robust estimates of forest states. CARDAMOM can assimilate these data at the site level to refine allocation and turnover rates (Ge et al. 2022). Additionally, global plant trait databases, such as the Fine-Root Ecology Database (FRED; Iversen et al. 2017) and the TRY Plant Trait Database (TRY; Kattge et al. 2020), provide valuable constraints for prior parameter estimates and refined ecological dynamic constraints. Manipulation experiments involving drought or CO_2 adjustments offer further opportunities to refine vegetation models (Medlyn et al. 2015). By calibrating DALEC with CARDAMOM under both control and intervention scenarios, hypotheses about functional adaptation under
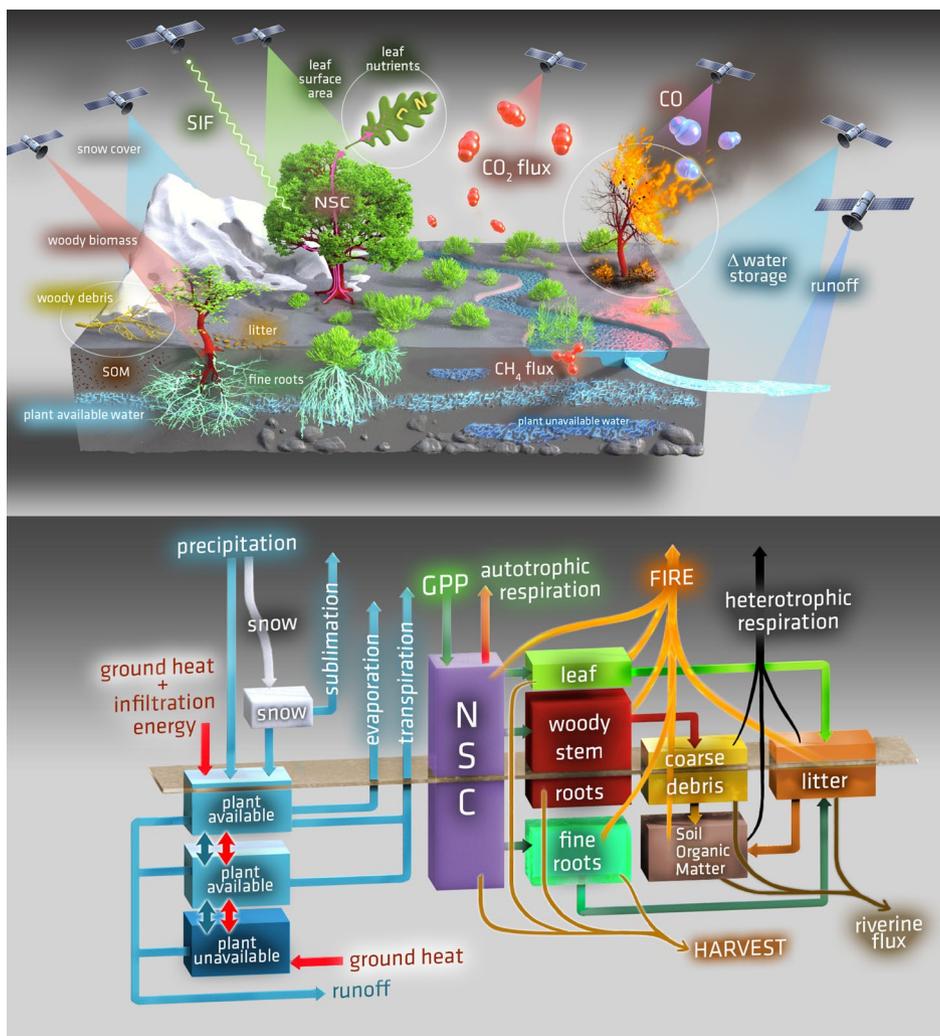
FIGURE 4 | Above figure shows current and future satellite-based observations used by CARDAMOM to constrain DALEC. The figure below presents a schematic representation of DALEC corresponding to the ecosystem shown in the figure above. Illustration by Victor O. Leshyk.

TABLE 2 | Possible new datasets for assimilation into CARDAMOM include existing, previously unassimilated datasets as well as those from future missions.

Mission	Measurement	Availability	Source
GEDI	Forest biomass	Available	Dubayah et al. (2023)
SWOT	Surface water extent and elevation	Available	Fu et al. (2024)
AVIRIS	Hyperspectral-derived plant traits	Available	Townsend et al. (2023)
EMIT	Hyperspectral-derived plant traits	Available	Green et al. (2020)
NISAR	Forest biomass	Recently launched	Kellogg et al. (2020)
BIOMASS	Forest biomass	Recently launched	Quegan et al. (2019)
FLEX	Solar-induced fluorescence and plant stress	2026	Moreno (2021)
CO2M	Total column CO_2 and CH_4. Solar-induced fluorescence	2027	Noël et al. (2024)
ROSE-L	Forest biomass	2027	Davidson and Furnell (2021)
CHIME	Hyperspectral-derived plant traits, and soil surface carbon and texture	2028	Rast et al. (2021)
SBG	Hyperspectral-derived plant traits	2032	Schimel and Poulter (2022)

stress can be tested, with predictive uncertainties propagated into ecosystem forecasts under climate and CO_2 change.

5.2 | Engagement With Wider Remote Sensing, Field, and Land Surface Modeling Communities

CARDAMOM does not exist in isolation, and its community must engage strongly with other research communities to maximize the possibility for ecological learning and supporting the creation of actionable information that can support decision-makers. Three directly relevant communities are remote sensing scientists, field scientists, and land surface modellers.

As new satellite missions and products emerge, fostering closer engagement with the remote sensing community remains a priority. CARDAMOM requires robust uncertainty estimates to guide assimilation effectively. Developing these estimates collaboratively with teams producing remote sensing products is essential (Labrière et al. 2023). CARDAMOM users would benefit significantly from a better understanding of how uncertainties scale with spatial aggregation and vary across time. Furthermore, observation operators that link DALEC state variables to observables (Reyes-Muñoz et al. 2024) may vary spatially and temporally due to factors such as ecological heterogeneity and human management, requiring input from remote sensing product creators. Additionally, the CARDAMOM community has substantial opportunities to shape future satellite missions by conducting data assimilation experiments with existing (Smallman et al. 2017, 2021) and synthetic datasets (Holtzman et al. 2023) to refine observation requirements and quantify the potential value of observations.

Close collaboration with field communities also offers an opportunity for improving process understanding from site to ecosystem scale. Efforts are increasing to establish networks of field inventory sites that provide consistent, repeated ecological observations (e.g., Malhi et al. 2002, 2021; Lopez-Gonzalez et al. 2011; The SEOSAW partnership 2021). These networks could benefit from systematic, data-informed analyses to generate and evaluate hypotheses relevant to the ecosystems they monitor (e.g., Medlyn et al. 2016).

The land surface modeling community supports national decision-makers (e.g., national carbon budgets) and international climate negotiations (e.g., IPCC reports) through large multimodel ensembles, such as the Trends in Global Net Land-Atmosphere Carbon Exchange project (TRENDY; Sitch et al. 2024). However, these models, due to their size and complexity, struggle to ingest observations for parameter updates and improved predictions. CARDAMOM offers a useful alternative by providing observationally informed insights into ecosystem states, dynamics, and traits, enabling more rigorous, systematic model evaluation, and identifying pathways for improvement (e.g., Caen et al. 2022; Williams et al. 2024). CARDAMOM already provides data-constrained carbon fluxes to TRENDY (e.g., Friedlingstein et al. 2023, 2024) and the Regional Carbon Cycle Assessment and Processes Phase 2 project (RECCAP2; Jones et al. 2023; Hugelius et al. 2024). Strengthening these connections will potentially expand CARDAMOM's usage and foster broader engagement between communities.

5.3 | Supporting Localized Ecosystem Monitoring, Management, and Decision-Making

CARDAMOM readily adapts to distinct terrestrial ecosystems, simulating responses to climate, management, and disturbance scenarios while propagating uncertainties for robust confidence estimates. It has already successfully modeled land management impacts in forestry (Smallman et al. 2017), arable ecosystems (Revill et al. 2021), and pastures (Myrgiotis et al. 2021, 2022). Additionally, its ability to assess shifting fire regimes, drought impacts, and landuse changes is highly relevant to land management and policy. When applied at appropriate spatial scales, these capabilities offer significant opportunities to support decision-makers and land managers with actionable insights for sustainable ecosystem management, such as simulating forest response to deforestation or disturbance (Smallman et al. 2022; Levy et al. 2022). The DALEC model structure also aligns with IPCC guidelines for United Nations Framework Convention on Climate Change (UNFCCC) national emissions reporting (IPCC 2006). By providing localized, observation-informed, and uncertainty-bounded carbon budgets, it offers an opportunity to enhance Tier 2 and Tier 3 land carbon balance reporting (Williams et al. 2024).

While CARDAMOM and DALEC show strong potential for supporting management, challenges remain. First, integrating CARDAMOM into operational frameworks requires collaboration with stakeholders (Melo et al. 2023). Second, capturing fine-scale land management patterns is difficult in large-scale studies relying on coarse spatial resolution (>1 km) observations (Milodowski et al. 2023). Third, robust frameworks for simulating landuse changes, especially in dynamic hotspots or over long periods, may require the development of temporal adjustments to model parameters to account for processes like biodiversity shifts (George-Chacon et al. 2023), which have not yet been common among CARDAMOM usage.

5.4 | Augmenting CARDAMOM With Machine Learning

The CARDAMOM framework is well-suited for integrating machine learning and deep learning approaches, while minimizing their drawbacks. Unlike purely data-driven techniques, DALEC models prognostically evolve carbon and water pools, enforcing mass conservation and allowing ecological memory effects to emerge in state-dependent processes (Bloom et al. 2020; Reichstein et al. 2019). CARDAMOM's ecological dynamic constraints and prior distributions eliminate many physically and ecologically infeasible parameter combinations, effectively reducing the dimensionality of the parameter space. This regularization allows the model to perform well even with sparse observations (Bloom and Williams 2015; Famiglietti et al. 2021). While traditional machine learning models often require thousands of training samples for robust performance, CARDAMOM analyses can achieve strong results with as few as 10–100 observations (Bloom et al. 2016; Shen et al. 2023). Building on these strengths, current and emerging machine learning strategies can complement and enhance CARDAMOM, including, for example,

new architectures using attention mechanisms (Vaswani et al. 2017), transfer learning (Immorlano et al. 2025), and hybrid model architectures (Fang and Gentine 2024). Below we describe four strategies through which machine learning techniques are currently or planned by the CARDAMOM community to be combined with CARDAMOM to enhance its utility (Figure 5).

First, machine learning can be used to fit unknown functional relationships within terrestrial biosphere models by replacing a prescribed model component with an embedded machine learning submodule, resulting in a hybrid model (Figure 5a). This approach is particularly effective for biophysical processes that are fast and have ample observational data (Reichstein et al. 2019; Braghiere et al. 2020; Massoud et al. 2023). The machine learning parameters (e.g., weights and biases in a neural network) can be calibrated alongside the physical parameters in a DALEC model, allowing for simultaneous quantification of structural and parametric uncertainties. For example, neural networks have been used to parameterize site-specific empirical water stress functions in a DALEC model, accounting for photosynthesis downregulation due to soil moisture limitation or high atmospheric demand (Fang and Gentine 2024). This is advantageous as uncertainties in the functional form of water stress contribute significantly to global carbon budget uncertainties (Trugman et al. 2018).

Second, DALEC can be coupled with a spatialization neural network to predict model parameters from spatial variables like climate and topography (Figure 5b; Reichstein et al. 2022; Bao et al. 2023; ElGhawi et al. 2023), though other approaches exist (Famiglietti et al. 2023). This coupled ML model (Fang et al. 2024) treats DALEC as a "physical layer" that learns parameters as a function of ecological and climate properties at each location. Parameter updates occur via gradients derived from the mismatch between DALEC simulations and observational data constraints, with tools like JAX, Julia, or PyTorch computing these gradients (Shen et al. 2023). This method offers two key benefits: (1) It preserves spatial covariations in model parameters imposed by model structures, forcing variables, and observational constraints, reducing overfitting compared to pixel-wise fitting with limited data; and (2) it incorporates the parameter sensitivity of the process-based model into the neural network through backpropagation, refining parameters while preventing infeasible combinations that could lead to instability or nonphysical outputs (Fang et al. 2024). Applying this approach to a light-use efficiency photosynthesis model has improved out-of-sample predictions, suggesting broader applicability in terrestrial biosphere modeling (Bao et al. 2023, 2024), including CARDAMOM applications.

Third, machine learning could downscale CARDAMOM outputs, generating higher-resolution ecological maps than the spatial resolution of CARDAMOM assimilations (Figure 5c). These fine-scale maps can support land management decisions and ecological service assessments (Yu et al. 2022). Currently, CARDAMOM's spatial resolution is constrained by computational capacity or the resolution of the coarsest dataset (see Section 4.3); machine learning offers a way to overcome these limitations by producing fine-grained estimates constrained by coarse-grained ones (Reichstein et al. 2019).

Fourth, machine learning techniques provide new ways to causally interpret (Bareinboim and Pearl 2016; Bareinboim et al. 2022; Massmann et al. 2021) CARDAMOM behavior (Figure 5d). Applying causal analysis to spatial CARDAMOM analyses can reveal biogeographic controls on the carbon cycle, leveraging CARDAMOM's ability to synthesize diverse observations into systemically consistent ecosystem representations. This integration helps capture potentially confounding
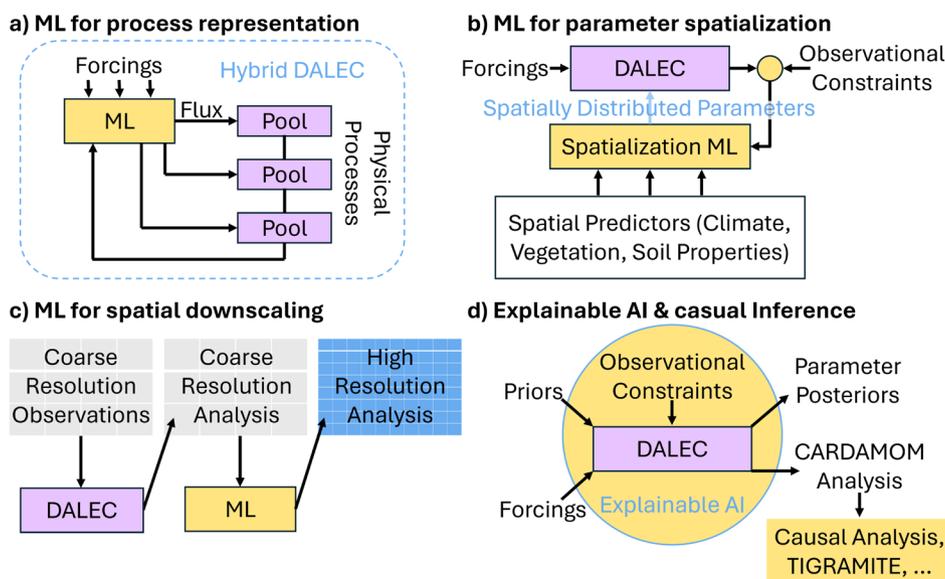

FIGURE 5 | Schematic showing different ways that machine learning (ML) can be incorporated into the CARDAMOM framework. (a) A hybrid model using machine learning to represent unknown functional relationships within a DALEC model; (b) A spatialization machine learning model used to simulate the relationships between observable climate, vegetation, and soil properties and the latent ecological variables in CARDAMOM; (c) machine learning used to produce high-resolution analyses of key ecological variables based on coarse-resolution CARDAMOM output; (d) Using interpretable AI and causal analysis tools in conjunction with CARDAMOM to understand the links between different ecological processes.

mechanistic pathways that individual observation streams may not fully resolve. Spatial CARDAMOM analyses can be combined with existing causal inference frameworks, such as Time-series Graph-based Measures of Information Transfer (TIGRAMITE; Runge 2018; Runge et al. 2023), to explore biogeographic variations. Applying causal analysis to a CARDAMOM representation of southern African woodlands demonstrated how interactions between climate, fire, and ecosystem traits drive spatial differences in carbon stocks, residence times, and fluxes (Williams et al. 2024).

As machine learning is incorporated into CARDAMOM, it would benefit from being benchmarked against Bayesian data assimilation-based versions in terms of predictive performance, parameter inference (e.g., spatial coherence and accuracy in retrieved ecosystem parameters across grids), computational efficiency (e.g., runtime and resource usage on high-performance computing clusters), uncertainty quantification capabilities (e.g., ensemble spread in model outputs and posterior distributions of parameters), and the ability to capture extreme events (e.g., droughts, floods, and heatwaves). Both versions would further benefit from broader benchmarks against existing process-based models (e.g., Quetin et al. 2020) as well as contemporary and emerging machine learning approaches (Dueben et al. 2022; Ullrich et al. 2025). Specific machine learning-based Earth system models for these comparisons could include random forest or neural network ensembles for estimating land surface fluxes (e.g., FLUXCOM; Jung et al. 2020), transformer-based architectures (e.g., Nguyen et al. 2023; Szwarcman et al. 2024), physics-informed neural networks (e.g., Wang et al. 2022), and foundation models pretrained on extensive climate datasets (e.g., Bodnar et al. 2025). While a standardized benchmark does not yet exist, Dueben et al. (2022) provide examples for machine learning-inspired evaluations, and Ullrich et al. (2025) mention potential standards for such benchmarks while encouraging leveraging existing process-based intercomparisons. Complementary comparisons against intercomparison frameworks could include the use of the International Land Model Benchmarking (ILAMB; Collier et al. 2018; already incorporated with the Bayesian version of CARDAMOM) and the Earth System Model Evaluation Tool (ESMValTool; Lauer et al. 2025).

6 | Conclusions

The potential for integrating new datasets and machine learning approaches into CARDAMOM presents opportunities to extract ecological insights across both site and global scales, as well as enhance model process representation. However, challenges such as data quality, computational costs, parameter equifinality, and model complexity must be understood to fully realize this potential. Placing CARDAMOM within the broader context of data assimilation frameworks reveals its unique strengths in spatially explicit, PFT-agnostic optimization of initial states and parameters on a global scale. To capitalize on these strengths, we recommend further collaboration with remote sensing, field, and land surface modeling communities to refine uncertainty estimates and deepen ecosystem understanding. These efforts can also deliver observation-informed, localized insights that support sustainable management and policy decisions. CARDAMOM's continued evolution will help bridge the gap between observations and process-based models, advancing our understanding of terrestrial ecosystem dynamics under a changing biosphere.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

Alencar, A. A., P. M. Brando, G. P. Asner, and F. E. Putz. 2015. "Landscape Fragmentation, Severe Drought, and the New Amazon Forest Fire Regime." *Ecological Applications* 25, no. 6: 1493–1505. https://doi.org/10.1890/14-1528.1.

Andela, N., D. C. Morton, L. Giglio, et al. 2017. "A Human-Driven Decline in Global Burned Area." *Science* 356, no. 6345: 1356–1362. https://doi.org/10.1126/science.aal4108.

Araza, A., S. De Bruin, L. Hein, and M. Herold. 2023. "Spatial Predictions and Uncertainties of Forest Carbon Fluxes for Carbon Accounting." *Scientific Reports* 13, no. 1: 12704. https://doi.org/10.1038/s41598-023-38935-8.

Arora, V. K., A. Katavouta, R. G. Williams, et al. 2020. "Carbon–Concentration and Carbon–Climate Feedbacks in CMIP6 Models and Their Comparison to CMIP5 Models." *Biogeosciences* 17, no. 16: 4173–4222. https://doi.org/10.5194/bg-17-4173-2020.

Au, J., A. A. Bloom, N. C. Parazoo, et al. 2023. "Forest Productivity Recovery or Collapse? Model-Data Integration Insights on Drought-Induced Tipping Points." *Global Change Biology* 29, no. 19: 5652–5665. https://doi.org/10.1111/gcb.16867.

Baldocchi, D. D. 2020. "How Eddy Covariance Flux Measurements Have Contributed to Our Understanding of *Global Change Biology*." *Global Change Biology* 26, no. 1: 242–260. https://doi.org/10.1111/gcb.14807.

Bao, S., L. Alonso, S. Wang, J. Gensheimer, R. De, and N. Carvalhais. 2023. "Toward Robust Parameterizations in Ecosystem-Level Photosynthesis Models." *Journal of Advances in Modeling Earth Systems* 15, no. 8: e2022MS003464. https://doi.org/10.1029/2022MS003464.

Bao, S., N. Carvalhais, J. Xu, et al. 2024. "Global Distribution Pattern in Characteristics of Gross Primary Productivity Response to Soil Water Availability." *Agricultural and Forest Meteorology* 372: 110701. https://doi.org/10.2139/ssrn.4789075.

Bareinboim, E., J. D. Correa, D. Ibeling, and T. Icard. 2022. "On Pearl's Hierarchy and the Foundations of Causal Inference." In *Probabilistic and Causal Inference*, edited by H. Geffner, R. Dechter, and J. Y. Halpern, 1st ed., 507–556. ACM. https://doi.org/10.1145/3501714.3501743.

Bareinboim, E., and J. Pearl. 2016. "Causal Inference and the Data-Fusion Problem." *Proceedings of the National Academy of Sciences* 113, no. 27: 7345–7352. https://doi.org/10.1073/pnas.1510507113.

Basu, S., S. J. Lehman, J. B. Miller, et al. 2020. "Estimating US Fossil Fuel CO_2 Emissions From Measurements of ^{14}C in Atmospheric CO_2." *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 24: 13300–13307. https://doi.org/10.1073/pnas.1919032117.

Beven, K., and J. Freer. 2001. "Equifinality, Data Assimilation, and Uncertainty Estimation in Mechanistic Modelling of Complex Environmental Systems Using the GLUE Methodology." *Journal of Hydrology* 249, no. 1–4: 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8.

Bilir, T. E., A. A. Bloom, A. G. Konings, et al. 2025. "Satellite-Constrained Reanalysis Reveals CO_2 Versus Climate Process Compensation Across the Global Land Carbon Sink." *AGU Advances*.

Bloom, A. A., K. W. Bowman, J. Liu, et al. 2020. "Lagged Effects Regulate the Inter-Annual Variability of the Tropical Carbon Balance." *Biogeosciences* 17, no. 24: 6393–6422. https://doi.org/10.5194/bg-17-6393-2020.

Bloom, A. A., J.-F. Exbrayat, I. R. Van Der Velde, L. Feng, and M. Williams. 2016. "The Decadal State of the Terrestrial Carbon Cycle: Global Retrievals of Terrestrial Carbon Allocation, Pools, and Residence Times." *Proceedings of the National Academy of Sciences* 113, no. 5: 1285–1290. https://doi.org/10.1073/pnas.1515160113.

Bloom, A. A., and M. Williams. 2015. "Constraining Ecosystem Carbon Dynamics in a Data-Limited World: Integrating Ecological "Common Sense" in a Model–Data Fusion Framework." *Biogeosciences* 12, no. 5: 1299–1315. https://doi.org/10.5194/bg-12-1299-2015.

Bodnar, C., W. P. Bruinsma, A. Lucic, et al. 2025. "A Foundation Model for the Earth System." *Nature* 641, no. 8065: 1180–1187. https://doi.org/10.1038/s41586-025-09005-y.

Bonan, G. B. 2008. "Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests." *Science* 320, no. 5882: 1444–1449. https://doi.org/10.1126/science.1155121.

Bonan, G. B. 2019. *Climate Change and Terrestrial Ecosystem Modeling.* Cambridge university press.

Bonan, G. B., and S. C. Doney. 2018. "Climate, Ecosystems, and Planetary Futures: The Challenge to Predict Life in Earth System Models." *Science* 359, no. 6375: eaam8328. https://doi.org/10.1126/science.aam8328.

Bonan, G. B., D. L. Lombardozzi, and W. R. Wieder. 2021. "The Signature of Internal Variability in the Terrestrial Carbon Cycle." *Environmental Research Letters* 16, no. 3: 34022. https://doi.org/10.1088/1748-9326/abd6a9.

Bonan, G. B., O. Lucier, D. R. Coen, et al. 2024. "Reimagining Earth in the Earth System." *Journal of Advances in Modeling Earth Systems* 16, no. 8: e2023MS004017. https://doi.org/10.1029/2023ms004017.

Braak, C. J. F. T. 2006. "A Markov Chain Monte Carlo Version of the Genetic Algorithm Differential Evolution: Easy Bayesian Computing for Real Parameter Spaces." *Statistics and Computing* 16, no. 3: 239–249. https://doi.org/10.1007/s11222-006-8769-1.

Braghiere, R. K., J. B. Fisher, K. Allen, et al. 2022. "Modeling Global Carbon Costs of Plant Nitrogen and Phosphorus Acquisition." *Journal of Advances in Modeling Earth Systems* 14, no. 8: e2022MS003204. https://doi.org/10.1029/2022MS003204.

Braghiere, R. K., J. B. Fisher, R. A. Fisher, et al. 2021. "Mycorrhizal Distributions Impact Global Patterns of Carbon and Nutrient Cycling." *Geophysical Research Letters* 48, no. 19: e2021GL094514. https://doi.org/10.1029/2021GL094514.

Braghiere, R. K., Y. Wang, R. Doughty, et al. 2021. "Accounting for Canopy Structure Improves Hyperspectral Radiative Transfer and Sun-Induced Chlorophyll Fluorescence Representations in a New Generation Earth System Model." *Remote Sensing of Environment* 261: 112497. https://doi.org/10.1016/j.rse.2021.112497.

Braghiere, R. K., M. A. Yamasoe, N. M. Évora do Rosário, H. Ribeiro da Rocha, J. de Souza Nogueira, and A. C. de Araújo 2020. "Characterization of the Radiative Impact of Aerosols on CO_2 and Energy Fluxes in the Amazon Deforestation Arch Using Artificial Neural Networks." *Atmospheric Chemistry and Physics* 20, no. 6: 3439–3458. https://doi.org/10.5194/acp-20-3439-2020.

Brinck, K., R. Fischer, J. Groeneveld, et al. 2017. "High Resolution Analysis of Tropical Forest Fragmentation and Its Impact on the Global

Carbon Cycle." *Nature Communications* 8, no. 1: 14855. https://doi.org/10.1038/ncomms14855.

Butler, E. E., A. Datta, H. Flores-Moreno, et al. 2017. "Mapping Local and Global Variability in Plant Trait Distributions." *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 51: E10937–E10946. https://doi.org/10.1073/pnas.1708984114.

Caen, A., T. L. Smallman, A. A. De Castro, et al. 2022. "Evaluating Two Land Surface Models for Brazil Using a Full Carbon Cycle Benchmark With Uncertainties." *Climate Resilience and Sustainability* 1, no. 1: e10. https://doi.org/10.1002/cli2.10.

Camberlin, P., and R. E. Okoola. 2003. "The Onset and Cessation of the "Long Rains" in Eastern Africa and Their Interannual Variability." *Theoretical and Applied Climatology* 75, no. 1: 43–54. https://doi.org/10.1007/s00704-002-0721-5.

Cameron, D., F. Hartig, F. Minnuno, et al. 2022. "Issues in Calibrating Models With Multiple Unbalanced Constraints: The Significance of Systematic Model and Data Errors." *Methods in Ecology and Evolution* 13, no. 12: 2757–2770. https://doi.org/10.1111/2041-210X.14002.

Chu, H., X. Luo, Z. Ouyang, et al. 2021. "Representativeness of Eddy-Covariance Flux Footprints for Areas Surrounding AmeriFlux Sites." *Agricultural and Forest Meteorology* 301: 108350. https://doi.org/10.1016/j.agrformet.2021.108350.

Collier, N., F. M. Hoffman, D. M. Lawrence, et al. 2018. "The International Land Model Benchmarking (Ilamb) System: Design, Theory, and Implementation." *Journal of Advances in Modeling Earth Systems* 10, no. 11: 2731–2754. https://doi.org/10.1029/2018MS001354.

Collins, E. L., C. H. David, R. Riggs, et al. 2024. "Global Patterns in River Water Storage Dependent on Residence Time." *Nature Geoscience* 17, no. 5: 433–439. https://doi.org/10.1038/s41561-024-01421-5.

Cui, E., E. Weng, E. Yan, and J. Xia. 2020. "Robust Leaf Trait Relationships Across Species Under Global Environmental Changes." *Nature Communications* 11, no. 1: 2999. https://doi.org/10.1038/s41467-020-16839-9.

David, C. H., J. M. Hobbs, M. J. Turmon, C. M. Emery, J. T. Reager, and J. S. Famiglietti. 2019. "Analytical Propagation of Runoff Uncertainty Into Discharge Uncertainty Through a Large River Network." *Geophysical Research Letters* 46, no. 14: 8102–8113. https://doi.org/10.1029/2019GL083342.

Davidson, M. W. J., and R. Furnell. 2021. "Rose-l: Copernicus l-Band Sar Mission." *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 872–873. https://doi.org/10.1109/IGARSS47720.2021.9554018.

Dokoohaki, H., B. D. Morrison, A. Raiho, S. P. Serbin, and M. Dietze. 2021. "A Novel Model–Data Fusion Approach to Terrestrial Carbon Cycle Reanalysis Across the Contiguous U.S Using SIPNET and PEcAn State Data Assimilation System v. 1.7.2." https://doi.org/10.5194/gmd-2021-236.

Dubayah, R. O., J. Armston, S. P. Healey, et al. 2023. *Global Ecosystem Dynamics Investigation (Gedi)gedi l4b Gridded Aboveground Biomass Density, Version 2. 1.* ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/2299.

Dueben, P. D., M. G. Schultz, M. Chantry, D. J. Gagne, D. M. Hall, and A. McGovern. 2022. "Challenges and Benchmark Datasets for Machine Learning in the Atmospheric Sciences: Definition, Status, and Outlook." *Artificial Intelligence for the Earth Systems* 1, no. 3: e210002. https://doi.org/10.1175/AIES-D-21-0002.1.

Durand, M., C. J. Gleason, T. M. Pavelsky, et al. 2023. "A Framework for Estimating Global River Discharge From the Surface Water and Ocean Topography Satellite Mission." *Water Resources Research* 59, no. 4: e2021WR031614. https://doi.org/10.1029/2021WR031614.

ElGhawi, R., B. Kraft, C. Reimers, et al. 2023. "Hybrid Modeling of Evapotranspiration: Inferring Stomatal and Aerodynamic Resistances Using Combined Physics-Based and Machine Learning." *Environmental Research Letters* 18, no. 3: 34039. https://doi.org/10.1088/1748-9326/acbbe0.

Exbrayat, J., T. L. Smallman, A. A. Bloom, L. B. Hutley, and M. Williams. 2018. "Inverse Determination of the Influence of Fire on Vegetation Carbon Turnover in the Pantropics." *Global Biogeochemical Cycles* 32, no. 12: 1776–1789. https://doi.org/10.1029/2018GB005925.

Eyring, V., P. M. Cox, G. M. Flato, et al. 2019. "Taking Climate Model Evaluation to the Next Level." *Nature Climate Change* 9, no. 2: 102–110. https://doi.org/10.1038/s41558-018-0355-y.

Famiglietti, C. A., T. L. Smallman, P. A. Levine, et al. 2021. "Optimal Model Complexity for Terrestrial Carbon Cycle Prediction." *Biogeosciences* 18, no. 8: 2727–2754. https://doi.org/10.5194/bg-18-2727-2021.

Famiglietti, C. A., M. Worden, L. D. L. Anderegg, and A. G. Konings. 2024. "Impacts of Climate Timescale on the Stability of Trait–Environment Relationships." *New Phytologist* 241, no. 6: 2423–2434. https://doi.org/10.1111/nph.19416.

Famiglietti, C. A., M. Worden, G. R. Quetin, et al. 2023. "Global Net Biome CO_2 Exchange Predicted Comparably Well Using Parameter–Environment Relationships and Plant Functional Types." *Global Change Biology* 29, no. 8: 2256–2273. https://doi.org/10.1111/gcb.16574.

Fang, J., K. Bowman, W. Zhao, X. Lian, and P. Gentine. 2024. "Differentiable Land Model Reveals Global Environmental Controls on Ecological Parameters." *arXiv.* https://doi.org/10.48550/ARXIV.2411.09654.

Fang, J., and P. Gentine. 2024. "Exploring Optimal Complexity for Water Stress Representation in Terrestrial Carbon Models: A Hybrid-Machine Learning Model Approach." *Journal of Advances in Modeling Earth Systems* 16, no. 12: e2024MS004308. https://doi.org/10.1029/2024MS004308.

Fang, Y., C. Liu, and L. R. Leung. 2015. "Accelerating the Spin-Up of the Coupled Carbon and Nitrogen Cycle Model in CLM4." *Geoscientific Model Development* 8, no. 3: 781–789. https://doi.org/10.5194/gmd-8-781-2015.

Fisher, R. A., and C. D. Koven. 2020. "Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems." *Journal of Advances in Modeling Earth Systems* 12, no. 4: e2018MS001453. https://doi.org/10.1029/2018MS001453.

Flack-Prain, S., P. Meir, Y. Malhi, T. L. Smallman, and M. Williams. 2019. "The Importance of Physiological, Structural and Trait Responses to Drought Stress in Driving Spatial and Temporal Variation in GPP Across Amazon Forests." *Biogeosciences* 16, no. 22: 4463–4484. https://doi.org/10.5194/bg-16-4463-2019.

Fox, A., M. Williams, A. D. Richardson, et al. 2009. "The REFLEX Project: Comparing Different Algorithms and Implementations for the Inversion of a Terrestrial Ecosystem Model Against Eddy Covariance Data." *Agricultural and Forest Meteorology* 149, no. 10: 1597–1615. https://doi.org/10.1016/j.agrformet.2009.05.002.

Fox, A. M., T. J. Hoar, J. L. Anderson, et al. 2018. "Evaluation of a Data Assimilation System for Land Surface Models Using clm4. 5." *Journal of Advances in Modeling Earth Systems* 10, no. 10: 2471–2494. https://doi.org/10.1029/2018MS001362.

Fox, A. M., X. Huo, T. J. Hoar, et al. 2022. "Assimilation of Global Satellite Leaf Area Estimates Reduces Modeled Global Carbon Uptake and Energy Loss by Terrestrial Ecosystems." *Journal of Geophysical Research: Biogeosciences* 127, no. 8: e2022JG006830. https://doi.org/10.1029/2022JG006830.

Friedlingstein, P., M. O'Sullivan, M. W. Jones, et al. 2023. "Global Carbon Budget 2023." *Earth System Science Data* 15, no. 12: 5301–5369. https://doi.org/10.5194/essd-15-5301-2023.

Friedlingstein, P., M. O'Sullivan, M. W. Jones, et al. 2024. "Global Carbon Budget 2024." https://doi.org/10.5194/essd-2024-519.

Fu, L., T. Pavelsky, J. Cretaux, et al. 2024. "The Surface Water and Ocean Topography Mission: A Breakthrough in Radar Remote Sensing of the Ocean and Land Surface Water." *Geophysical Research Letters* 51, no. 4: e2023GL107652. https://doi.org/10.1029/2023GL107652.

Ge, R., H. He, L. Zhang, et al. 2022. "Climate Sensitivities of Carbon Turnover Times in Soil and Vegetation: Understanding Their Effects on Forest Carbon Sequestration." *Journal of Geophysical Research: Biogeosciences* 127, no. 3: e2020JG005880. https://doi.org/10.1029/2020JG005880.

George-Chacon, S. P., T. L. Smallman, J. M. Dupuy, J. L. Hernández-Stefanoni, D. T. Milodowski, and M. Williams. 2023. "Isolating the Effects of Land Use and Functional Variation on Yucatán's Forest Biomass Under Global Change." *Frontiers in Forests and Global Change* 6: 1204596. https://doi.org/10.3389/ffgc.2023.1204596.

Giglio, L., J. T. Randerson, and G. R. Van Der Werf. 2013. "Analysis of Daily, Monthly, and Annual Burned Area Using the Fourth-Generation Global Fire Emissions Database (Gfed4)." *Journal of Geophysical Research: Biogeosciences* 118, no. 1: 317–328. https://doi.org/10.1002/jgrg.20042.

Green, R. O., N. Mahowald, C. Ung, et al. 2020. "The Earth Surface Mineral Dust Source Investigation: An Earth Science Imaging Spectroscopy Mission." 2020 IEEE Aerospace Conference, 1–15. https://doi.org/10.1109/AERO47225.2020.9172731.

Grieve, S. W. D., S. M. Mudd, D. T. Milodowski, F. J. Clubb, and D. J. Furbish. 2016. "How Does Grid-Resolution Modulate the Topographic Expression of Geomorphic Processes?" *Earth Surface Dynamics* 4, no. 3: 627–653. https://doi.org/10.5194/esurf-4-627-2016.

Hansen, M. C., S. V. Stehman, and P. V. Potapov. 2010. "Quantification of Global Gross Forest Cover Loss." *Proceedings of the National Academy of Sciences* 107, no. 19: 8650–8655. https://doi.org/10.1073/pnas.0912668107.

Hempson, G. P., S. Archibald, and W. J. Bond. 2015. "A Continent-Wide Assessment of the Form and Intensity of Large Mammal Herbivory in Africa." *Science* 350, no. 6264: 1056–1061. https://doi.org/10.1126/science.aac7978.

Hiederer, R., M. Köchy, and European Commission. 2011. "Global Soil Organic Carbon Estimates and the Harmonized World Soil Database." Publications Office. https://doi.org/10.2788/13267.

Hill, T. C., T. Quaife, and M. Williams. 2011. "A Data Assimilation Method for Using Low-Resolution Earth Observation Data in Heterogeneous Ecosystems." *Journal of Geophysical Research* 116, no. D8: D08117. https://doi.org/10.1029/2010JD015268.

Hill, T. C., E. Ryan, and M. Williams. 2012. "The Use of CO 2 Flux Time Series for Parameter and Carbon Stock Estimation in Carbon Cycle Research." *Global Change Biology* 18, no. 1: 179–193. https://doi.org/10.1111/j.1365-2486.2011.02511.x.

Hollinger, D. Y., and A. D. Richardson. 2005. "Uncertainty in Eddy Covariance Measurements and Its Application to Physiological Models." *Tree Physiology* 25, no. 7: 873–885. https://doi.org/10.1093/treephys/25.7.873.

Holtzman, N., Y. Wang, J. D. Wood, C. Frankenberg, and A. G. Konings. 2023. "Constraining Plant Hydraulics With Microwave Radiometry in a Land Surface Model: Impacts of Temporal Resolution." *Water Resources Research* 59, no. 11: e2023WR035481. https://doi.org/10.1029/2023WR035481.

Hourdin, F., T. Mauritsen, A. Gettelman, et al. 2017. "The Art and Science of Climate Model Tuning." *Bulletin of the American Meteorological Society* 98, no. 3: 589–602. https://doi.org/10.1175/BAMS-D-15-00135.1.

Hugelius, G., J. Ramage, E. Burke, et al. 2024. "Permafrost Region Greenhouse Gas Budgets Suggest a Weak co_2 Sink and ch_4 and $n_2 o$ Sources, but Magnitudes Differ Between Top-Down and Bottom-Up Methods." *Global Biogeochemical Cycles* 38, no. 10: e2023GB007969. https://doi.org/10.1029/2023GB007969.

Immorlano, F., V. Eyring, T. Le Monnier De Gouville, et al. 2025. "Transferring Climate Change Physical Knowledge." *Proceedings of the National Academy of Sciences of the United States of America* 122, no. 15: e2413503122. https://doi.org/10.1073/pnas.2413503122.

Intergovernmental Panel On Climate Change (IPCC). 2023. *Climate Change 2022—Impacts, Adaptation and Vulnerability: Working Group ii Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009325844.

IPCC. 2006. *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, edited by H. S. Eggleston, L. Buendia, K. Miwa, T. Ngara, and K. Tanabe. IGES.

Iversen, C. M., M. L. McCormack, A. S. Powell, et al. 2017. "A Global Fine-Root Ecology Database to Address Below-Ground Challenges in Plant Ecology." *New Phytologist* 215, no. 1: 15–26. https://doi.org/10.1111/nph.14486.

Jenkins, G. M., and D. G. Watts. 1968. *Spectral Analysis and Its Applications*. Holden-Day.

Jiang, Z., J. R. Worden, H. Worden, et al. 2017. "A 15-Year Record of CO Emissions Constrained by MOPITT CO Observations." *Atmospheric Chemistry and Physics* 17, no. 7: 4565–4583. https://doi.org/10.5194/acp-17-4565-2017.

Joiner, J., Y. Yoshida, Y. Zhang, et al. 2018. "Estimation of Terrestrial Global Gross Primary Production (Gpp) With Satellite Data-Driven Models and Eddy Covariance Flux Data." *Remote Sensing* 10, no. 9: 1346. https://doi.org/10.3390/rs10091346.

Jones, C. D., T. Ziehn, J. Anand, et al. 2023. "Reccap2 Future Component: Consistency and Potential for Regional Assessment to Constrain Global Projections." *AGU Advances* 4, no. 6: e2023AV001024. https://doi.org/10.1029/2023AV001024.

Joswig, J. S., C. Wirth, M. C. Schuman, et al. 2021. "Climatic and Soil Factors Explain the Two-Dimensional Spectrum of Global Plant Trait Variation." *Nature Ecology & Evolution* 6, no. 1: 36–50. https://doi.org/10.1038/s41559-021-01616-8.

Jung, M., C. Schwalm, M. Migliavacca, et al. 2020. "Scaling Carbon Fluxes From Eddy Covariance Sites to Globe: Synthesis and Evaluation of the FLUXCOM Approach." *Biogeosciences* 17, no. 5: 1343–1365. https://doi.org/10.5194/bg-17-1343-2020.

Kaminski, T., and P.-P. Mathieu. 2017. "Reviews and Syntheses: Flying the Satellite Into Your Model: On the Role of Observation Operators in Constraining Models of the Earth System and the Carbon Cycle." *Biogeosciences* 14, no. 9: 2343–2357. https://doi.org/10.5194/bg-14-2343-2017.

Kattge, J., G. Bönisch, S. Díaz, et al. 2020. "TRY Plant Trait Database—Enhanced Coverage and Open Access." *Global Change Biology* 26, no. 1: 119–188. https://doi.org/10.1111/gcb.14904.

Keenan, T. F., M. S. Carbone, M. Reichstein, and A. D. Richardson. 2011. "The Model–Data Fusion Pitfall: Assuming Certainty in an Uncertain World." *Oecologia* 167, no. 3: 587–597. https://doi.org/10.1007/s00442-011-2106-x.

Keenan, T. F., E. A. Davidson, J. W. Munger, and A. D. Richardson. 2013. "Rate My Data: Quantifying the Value of Ecological Data for the Development of Models of the Terrestrial Carbon Cycle." *Ecological Applications* 23, no. 1: 273–286. https://doi.org/10.1890/12-0747.1.

Kellogg, K., P. Hoffman, S. Standley, et al. 2020. "NASA-ISRO Synthetic Aperture Radar (NISAR) Mission." 2020 IEEE Aerospace Conference, 1–21. https://doi.org/10.1109/AERO47225.2020.9172638.

Kondo, M., P. K. Patra, S. Sitch, et al. 2020. "State of the Science in Reconciling Top-Down and Bottom-Up Approaches for Terrestrial CO_2 budget." *Global Change Biology* 26, no. 3: 1068–1084. https://doi.org/10.1111/gcb.14917.

Konings, A. G., A. A. Bloom, J. Liu, N. C. Parazoo, D. S. Schimel, and K. W. Bowman. 2019. "Global Satellite-Driven Estimates of Heterotrophic Respiration." *Biogeosciences* 16, no. 11: 2269–2284. https://doi.org/10.5194/bg-16-2269-2019.

Konings, A. G., K. Rao, E. L. McCormick, et al. 2024. "Tree Species Explain Only Half of Explained Spatial Variability in Plant Water Sensitivity." *Global Change Biology* 30, no. 7: e17425. https://doi.org/10.1111/gcb.17425.

Kuppel, S., P. Peylin, F. Maignan, et al. 2014. "Model–Data Fusion Across Ecosystems: From Multisite Optimizations to Global Simulations." *Geoscientific Model Development* 7, no. 6: 2581–2597. https://doi.org/10.5194/gmd-7-2581-2014.

Labrière, N., S. J. Davies, M. I. Disney, et al. 2023. "Toward a Forest Biomass Reference Measurement System for Remote Sensing Applications." *Global Change Biology* 29, no. 3: 827–840. https://doi.org/10.1111/gcb.16497.

Lahoz, W. A., and P. Schneider. 2014. "Data Assimilation: Making Sense of Earth Observation." *Frontiers in Environmental Science* 2: 16. https://doi.org/10.3389/fenvs.2014.00016.

Lasslop, G., M. Reichstein, J. Kattge, and D. Papale. 2008. "Influences of Observation Errors in Eddy Flux Data on Inverse Model Parameter Estimation." *Biogeosciences* 5, no. 5: 1311–1324. https://doi.org/10.5194/bg-5-1311-2008.

Lauer, A., L. Bock, B. Hassler, P. Jöckel, L. Ruhe, and M. Schlund. 2025. "Monitoring and Benchmarking Earth System Model Simulations With ESMValTool v2.12.0." *Geoscientific Model Development* 18, no. 4: 1169–1188. https://doi.org/10.5194/gmd-18-1169-2025.

LeBauer, D. S., D. Wang, K. T. Richter, C. C. Davidson, and M. C. Dietze. 2013. "Facilitating Feedbacks Between Field Measurements and Ecosystem Models." *Ecological Monographs* 83, no. 2: 133–154. https://doi.org/10.1890/12-0137.1.

Levine, P. A., A. A. Bloom, K. W. Bowman, et al. 2023. "Water Stress Dominates 21st-Century Tropical Land Carbon Uptake." *Global Biogeochemical Cycles* 37, no. 12: e2023GB007702. https://doi.org/10.1029/2023GB007702.

Levine, P. A., A. A. Bloom, M. A. Worden, et al. 2025. "Technical Documentation of the CARbon DAta-MOdel fraMework C Implementation (CARDAMOM-C)." https://cardamom-framework.github.io/CARDAMOM.

Levy, P., R. Clement, N. Cowan, et al. 2022. "Challenges in Scaling Up Greenhouse Gas Fluxes: Experience From the UK Greenhouse Gas Emissions and Feedbacks Program." *Journal of Geophysical Research – Biogeosciences* 127, no. 5: e2021JG006743. https://doi.org/10.1029/2021JG006743.

Li, X., F. Liu, C. Ma, et al. 2024. "Land Data Assimilation: Harmonizing Theory and Data in Land Surface Process Studies." *Reviews of Geophysics* 62, no. 1: e2022RG000801. https://doi.org/10.1029/2022RG000801.

Liu, J., L. Baskaran, K. Bowman, et al. 2021. "Carbon Monitoring System Flux Net Biosphere Exchange 2020(Cms-Flux Nbe 2020)." *Earth System Science Data* 13, no. 2: 299–330. https://doi.org/10.5194/essd-13-299-2021.

Loescher, H. W., B. E. Law, L. Mahrt, D. Y. Hollinger, J. Campbell, and S. C. Wofsy. 2006. "Uncertainties in, and Interpretation of, Carbon Flux Estimates Using the Eddy Covariance Technique." *Journal of Geophysical Research: Atmospheres* 111, no. D21: 2005JD006932. https://doi.org/10.1029/2005JD006932.

Lopez-Gonzalez, G., S. L. Lewis, M. Burkitt, and O. L. Phillips. 2011. "ForestPlots.Net: A Web Application and Research Tool to Manage and Analyse Tropical Forest Plot Data." *Journal of Vegetation Science* 22, no. 4: 610–613. https://doi.org/10.1111/j.1654-1103.2011.01312.x.

Lovenduski, N. S., and G. B. Bonan. 2017. "Reducing Uncertainty in Projections of Terrestrial Carbon Uptake." *Environmental Research Letters* 12, no. 4: 44020. https://doi.org/10.1088/1748-9326/aa66b8.

Lovenduski, N. S., G. B. Bonan, S. G. Yeager, K. Lindsay, and D. L. Lombardozzi. 2019. "High Predictability of Terrestrial Carbon Fluxes From an Initialized Decadal Prediction System." *Environmental Research Letters* 14, no. 12: 124074. https://doi.org/10.1088/1748-9326/ab5c55.

Luo, Y., K. Ogle, C. Tucker, et al. 2011. "Ecological Forecasting and Data Assimilation in a Data-Rich Era." *Ecological Applications* 21, no. 5: 1429–1442. https://doi.org/10.1890/09-1275.1.

Luo, Y., and B. Smith. 2022. *Land Carbon Cycle Modeling: Matrix Approach, Data Assimilation, & Ecological Forecasting.* 1st ed. CRC Press. https://doi.org/10.1201/9780429155659.

Ma, S., A. A. Bloom, J. D. Watts, et al. 2023. "Resolving the Carbon-Climate Feedback Potential of Wetland CO_2 and CH_4 Fluxes in Alaska." *Global Biogeochemical Cycles* 37, no. 9: e2022GB007524. https://doi.org/10.1029/2022GB007524.

Ma, S., J. R. Worden, A. A. Bloom, et al. 2021. "Satellite Constraints on the Latitudinal Distribution and Temperature Sensitivity of Wetland Methane Emissions." *AGU Advances* 2, no. 3: e2021AV000408. https://doi.org/10.1029/2021AV000408.

MacBean, N., C. Bacour, N. Raoult, et al. 2022. "Quantifying and Reducing Uncertainty in Global Carbon Cycle Predictions: Lessons and Perspectives From 15 Years of Data Assimilation Studies With the Orchidee Terrestrial Biosphere Model." *Global Biogeochemical Cycles* 36, no. 7: e2021GB007177. https://doi.org/10.1029/2021GB007177.

Mahadevan, P., S. C. Wofsy, D. M. Matross, et al. 2008. "A Satellite-Based Biosphere Parameterization for Net Ecosystem CO_2 Exchange: Vegetation Photosynthesis and Respiration Model (Vprm)." *Global Biogeochemical Cycles* 22, no. 2: 2006GB002735. https://doi.org/10.1029/2006GB002735.

Malhi, Y., C. Girardin, D. B. Metcalfe, et al. 2021. "The Global Ecosystems Monitoring Network: Monitoring Ecosystem Productivity and Carbon Cycling Across the Tropics." *Biological Conservation* 253: 108889. https://doi.org/10.1016/j.biocon.2020.108889.

Malhi, Y., O. L. Phillips, J. Lloyd, et al. 2002. "An International Network to Monitor the Structure, Composition and Dynamics of Amazonian Forests (Rainfor)." *Journal of Vegetation Science* 13, no. 3: 439–450. https://doi.org/10.1111/j.1654-1103.2002.tb02068.x.

Massmann, A., P. Gentine, and J. Runge. 2021. "Causal Inference for Process Understanding in Earth Sciences (No. arXiv:2105.00912)." *arXiv.* https://doi.org/10.48550/arXiv.2105.00912.

Massoud, E. C., A. A. Bloom, M. Longo, J. T. Reager, P. A. Levine, and J. R. Worden. 2022. "Information Content of Soil Hydrology in a West Amazon Watershed as Informed by GRACE." *Hydrology and Earth System Sciences* 26, no. 5: 1407–1423. https://doi.org/10.5194/hess-26-1407-2022.

Massoud, E. C., F. Hoffman, Z. Shi, et al. 2023. "Perspectives on Artificial Intelligence for Predictions in Ecohydrology." *Artificial Intelligence for the Earth Systems* 2, no. 4: e230005. https://doi.org/10.1175/AIES-D-23-0005.1.

Medlyn, B. E., M. G. De Kauwe, S. Zaehle, et al. 2016. "Using Models to Guide Field Experiments: *A Priori* Predictions for the CO_2 Response of a Nutrient- and Water-Limited Native Eucalypt Woodland." *Global Change Biology* 22, no. 8: 2834–2851. https://doi.org/10.1111/gcb.13268.

Medlyn, B. E., S. Zaehle, M. G. De Kauwe, et al. 2015. "Using Ecosystem Experiments to Improve Vegetation Models." *Nature Climate Change* 5, no. 6: 528–534. https://doi.org/10.1038/nclimate2621.

Melo, J., T. Baker, D. Nemitz, S. Quegan, and G. Ziv. 2023. "Satellite-Based Global Maps Are Rarely Used in Forest Reference Levels

Submitted to the UNFCCC." *Environmental Research Letters* 18, no. 3: 34021. https://doi.org/10.1088/1748-9326/acba31.

Metcalfe, D. B., G. P. Asner, R. E. Martin, et al. 2014. "Herbivory Makes Major Contributions to Ecosystem Carbon and Nutrient Cycling in Tropical Forests." *Ecology Letters* 17, no. 3: 324–332. https://doi.org/10.1111/ele.12233.

Milodowski, D. T., T. L. Smallman, and M. Williams. 2023. "Scale Variance in the Carbon Dynamics of Fragmented, Mixed-Use Landscapes Estimated Using Model–Data Fusion." *Biogeosciences* 20, no. 15: 3301–3327. https://doi.org/10.5194/bg-20-3301-2023.

Moreno, J. F. 2021. "The Fluorescence Explorer (Flex) Mission: From Spectral Measurements to High-Level Science Products." 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 115–118. https://doi.org/10.1109/IGARSS47720.2021.9553537.

Morrison, L. W. 2016. "Observer Error in Vegetation Surveys: A Review." *Journal of Plant Ecology* 9, no. 4: 367–379. https://doi.org/10.1093/jpe/rtv077.

Myneni, R., Y. Knyazikhin, and T. Park. 2015. "*Myd15a2h modis/aqua leaf area index/fpar 8-day l4 global 500m sin grid v006.*" [Dataset]. [Object]. https://doi.org/10.5067/MODIS/MYD15A2H.006.

Myrgiotis, V., E. Blei, R. Clement, et al. 2020. "A Model-Data Fusion Approach to Analyse Carbon Dynamics in Managed Grasslands." *Agricultural Systems* 184: 102907. https://doi.org/10.1016/j.agsy.2020.102907.

Myrgiotis, V., P. Harris, A. Revill, H. Sint, and M. Williams. 2021. "Inferring Management and Predicting Sub-Field Scale C Dynamics in UK Grasslands Using Biogeochemical Modelling and Satellite-Derived Leaf Area Data." *Agricultural and Forest Meteorology* 307: 108466. https://doi.org/10.1016/j.agrformet.2021.108466.

Myrgiotis, V., T. L. Smallman, and M. Williams. 2022. "The Carbon Budget of the Managed Grasslands of Great Britain—Informed by Earth Observations." *Biogeosciences* 19, no. 17: 4147–4170. https://doi.org/10.5194/bg-19-4147-2022.

Nguyen, T., J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover. 2023. "ClimaX: A Foundation Model for Weather and Climate." *arXiv.* https://doi.org/10.48550/ARXIV.2301.10343.

Noël, S., M. Buchwitz, M. Hilker, et al. 2024. "Greenhouse Gas Retrievals for the CO2M Mission Using the FOCAL Method: First Performance Estimates." *Atmospheric Measurement Techniques* 17, no. 8: 2317–2334. https://doi.org/10.5194/amt-17-2317-2024.

Norton, A. J., A. A. Bloom, N. C. Parazoo, et al. 2023. "Improved Process Representation of Leaf Phenology Significantly Shifts Climate Sensitivity of Ecosystem Carbon Balance." *Biogeosciences* 20, no. 12: 2455–2484. https://doi.org/10.5194/bg-20-2455-2023.

Norton, A. J., P. J. Rayner, E. N. Koffi, and M. Scholze. 2018. "Assimilating Solar-Induced Chlorophyll Fluorescence Into the Terrestrial Biosphere Model BETHY-SCOPE v1.0: Model Description and Information Content." *Geoscientific Model Development* 11, no. 4: 1517–1536. https://doi.org/10.5194/gmd-11-1517-2018.

Pan, M., and E. F. Wood. 2013. "Inverse Streamflow Routing." *Hydrology and Earth System Sciences* 17, no. 11: 4577–4588. https://doi.org/10.5194/hess-17-4577-2013.

Pastorello, G., C. Trotta, E. Canfora, et al. 2020. "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." *Scientific Data* 7, no. 1: 225. https://doi.org/10.1038/s41597-020-0534-3.

Peylin, P., R. M. Law, K. R. Gurney, et al. 2013. "Global Atmospheric Carbon Budget: Results From an Ensemble of Atmospheric CO_2 Inversions." *Biogeosciences* 10, no. 10: 6699–6720. https://doi.org/10.5194/bg-10-6699-2013.

Poggio, L., L. M. De Sousa, N. H. Batjes, et al. 2021. "SoilGrids 2.0: Producing Soil Information for the Globe With Quantified Spatial Uncertainty." *Soil* 7, no. 1: 217–240. https://doi.org/10.5194/soil-7-217-2021.

Quegan, S., T. Le Toan, J. Chave, et al. 2019. "The European Space Agency BIOMASS Mission: Measuring Forest Above-Ground Biomass From Space." *Remote Sensing of Environment* 227: 44–60. https://doi.org/10.1016/j.rse.2019.03.032.

Quetin, G. R., L. D. L. Anderegg, I. Boving, W. R. L. Anderegg, and A. T. Trugman. 2023. "Observed Forest Trait Velocities Have Not Kept Pace With Hydraulic Stress From Climate Change." *Global Change Biology* 29, no. 18: 5415–5428. https://doi.org/10.1111/gcb.16847.

Quetin, G. R., A. A. Bloom, K. W. Bowman, and A. G. Konings. 2020. "Carbon Flux Variability From a Relatively Simple Ecosystem Model With Assimilated Data Is Consistent With Terrestrial Biosphere Model Estimates." *Journal of Advances in Modeling Earth Systems* 12, no. 3: e2019MS001889. https://doi.org/10.1029/2019MS001889.

Quetin, G. R., C. A. Famiglietti, N. C. Dadap, et al. 2023. "Attributing Past Carbon Fluxes to Co_2 and Climate Change: Respiration Response to Co_2 fertilization Shifts Regional Distribution of the Carbon Sink." *Global Biogeochemical Cycles* 37, no. 2: e2022GB007478. https://doi.org/10.1029/2022GB007478.

Raj, R., C. Van Der Tol, N. A. S. Hamm, and A. Stein. 2018. "Bayesian Integration of Flux Tower Data Into a Process-Based Simulator for Quantifying Uncertainty in Simulated Output." *Geoscientific Model Development* 11, no. 1: 83–101. https://doi.org/10.5194/gmd-11-83-2018.

Rast, M., J. Nieke, J. Adams, C. Isola, and F. Gascon. 2021. "Copernicus Hyperspectral Imaging Mission for the Environment(Chime)." 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 108–111. https://doi.org/10.1109/IGARSS47720.2021.9553319.

Rayner, P. J., A. M. Michalak, and F. Chevallier. 2019. "Fundamentals of Data Assimilation Applied to Biogeochemistry." *Atmospheric Chemistry and Physics* 19, no. 22: 13911–13932. https://doi.org/10.5194/acp-19-13911-2019.

Reichstein, M., B. Ahrens, B. Kraft, et al. 2022. "Combining System Modeling and Machine Learning Into Hybrid Ecosystem Modeling." In *Knowledge-Guided Machine Learning*, edited by A. Karpatne, R. Kannan, and V. Kumar, 1st ed., 327–352. Chapman and Hall/CRC. https://doi.org/10.1201/9781003143376-14.

Reichstein, M., G. Camps-Valls, B. Stevens, et al. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566, no. 7743: 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Revill, A., V. Myrgiotis, A. Florence, et al. 2021. "Combining Process Modelling and Lai Observations to Diagnose Winter Wheat Nitrogen Status and Forecast Yield." *Agronomy* 11, no. 2: 314. https://doi.org/10.3390/agronomy11020314.

Reyes-Muñoz, P., D. Dkovács, K. Berger, et al. 2024. "Inferring Global Terrestrial Carbon Fluxes From the Synergy of Sentinel 3 & 5P With Gaussian Process Hybrid Models." *Remote Sensing of Environment* 305: 114072. https://doi.org/10.1016/j.rse.2024.114072.

Richardson, A. D., M. Williams, D. Y. Hollinger, et al. 2010. "Estimating Parameters of a Forest Ecosystem C Model With Measurements of Stocks and Fluxes as Joint Constraints." *Oecologia* 164, no. 1: 25–40. https://doi.org/10.1007/s00442-010-1628-y.

Rodríguez-Veiga, P., J. Carreiras, T. Smallman, et al. 2020. "Carbon Stocks and Fluxes in Kenyan Forests and Wooded Grasslands Derived From Earth Observation and Model-Data Fusion." *Remote Sensing* 12, no. 15: 2380. https://doi.org/10.3390/rs12152380.

Rowland, L., T. C. Hill, C. Stahl, et al. 2014. "Evidence for Strong Seasonality in the Carbon Storage and Carbon Use Efficiency of an Amazonian Forest." *Global Change Biology* 20, no. 3: 979–991. https://doi.org/10.1111/gcb.12375.

Runge, J. 2018. "Causal Network Reconstruction From Time Series: From Theoretical Assumptions to Practical Estimation." *Chaos (Woodbury, N.Y.)* 28, no. 7: 75310. https://doi.org/10.1063/1.5025050.

Runge, J., A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls. 2023. "Causal Inference for Time Series." *Nature Reviews Earth and Environment* 4, no. 7: 487–505. https://doi.org/10.1038/s43017-023-00431-y.

Ryan, C. M., T. Hill, E. Woollen, et al. 2012. "Quantifying Small-Scale Deforestation and Forest Degradation in African Woodlands Using Radar Imagery." *Global Change Biology* 18, no. 1: 243–257. https://doi.org/10.1111/j.1365-2486.2011.02551.x.

Ryu, Y., J. A. Berry, and D. D. Baldocchi. 2019. "What Is Global Photosynthesis? History, Uncertainties and Opportunities." *Remote Sensing of Environment* 223: 95–114. https://doi.org/10.1016/j.rse.2019.01.016.

Santoro, M., and O. Cartus. 2024. "*ESA Biomass Climate Change Initiative (Biomass_cci): Global Datasets of Forest Above-Ground biomass for the Years 2010, 2015, 2016, 2017, 2018, 2019, 2020 and 2021, v5* [Application/Xml]." NERC EDS Centre for Environmental Data Analysis. https://doi.org/10.5285/02E1B18071AD45A19B4D3E8ADAFA2817.

Schimel, D., R. Pavlick, J. B. Fisher, et al. 2015. "Observing Terrestrial Ecosystems and the Carbon Cycle From Space." *Global Change Biology* 21, no. 5: 1762–1776. https://doi.org/10.1111/gcb.12822.

Schimel, D., F. D. Schneider, and JPL Carbon and Ecosystem Participants. 2019. "Flux Towers in the Sky: Global Ecology From Space." *New Phytologist* 224, no. 2: 570–584. https://doi.org/10.1111/nph.15934.

Schimel, D. S., and D. Carroll. 2024. "Carbon Cycle–Climate Feedbacks in the Post-Paris World." *Annual Review of Earth and Planetary Sciences* 52, no. 1: 467–493. https://doi.org/10.1146/annurev-earth-031621-081700.

Schimel, D. S., and B. Poulter. 2022. "The Earth in Living Color—Nasa's Surface Biology and Geology Designated Observable." 2022 IEEE Aerospace Conference (AERO), 1–6. https://doi.org/10.1109/AERO53065.2022.9843640.

Schneider, T., S. Lan, A. Stuart, and J. Teixeira. 2017. "Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations." *Geophysical Research Letters* 44, no. 24: 12–396. https://doi.org/10.1002/2017GL076101.

Scholze, M., T. Kaminski, W. Knorr, et al. 2016. "Simultaneous Assimilation of SMOS Soil Moisture and Atmospheric CO_2 In-Situ Observations to Constrain the Global Terrestrial Carbon Cycle." *Remote Sensing of Environment* 180: 334–345. https://doi.org/10.1016/j.rse.2016.02.058.

Seiler, C., J. R. Melton, V. K. Arora, et al. 2022. "Are Terrestrial Biosphere Models Fit for Simulating the Global Land Carbon Sink?" *Journal of Advances in Modeling Earth Systems* 14, no. 5: e2021MS002946. https://doi.org/10.1029/2021MS002946.

Sellers, P. J., D. S. Schimel, B. Moore, J. Liu, and A. Eldering. 2018. "Observing Carbon Cycle–Climate Feedbacks From Space." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 31: 7860–7868. https://doi.org/10.1073/pnas.1716613115.

Shen, C., A. P. Appling, P. Gentine, et al. 2023. "Differentiable Modelling to Unify Machine Learning and Physical Models for Geosciences." *Nature Reviews Earth and Environment* 4, no. 8: 552–567. https://doi.org/10.1038/s43017-023-00450-9.

Shi, J., X. Wang, and E. Wang. 2023. "Mycorrhizal Symbiosis in Plant Growth and Stress Adaptation: From Genes to Ecosystems." *Annual Review of Plant Biology* 74, no. 1: 569–607. https://doi.org/10.1146/annurev-arplant-061722-090342.

Sierra, C. A., G. R. Quetin, H. Metzler, and M. Müller. 2023. "A Decrease in the Age of Respired Carbon From the Terrestrial Biosphere and Increase in the Asymmetry of Its Distribution." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 381, no. 2261: 20220200. https://doi.org/10.1098/rsta.2022.0200.

Sitch, S., M. O'Sullivan, E. Robertson, et al. 2024. "Trends and Drivers of Terrestrial Sources and Sinks of Carbon Dioxide: An Overview of the Trendy Project." *Global Biogeochemical Cycles* 38, no. 7: e2024GB008102. https://doi.org/10.1029/2024GB008102.

Slevin, D., S. F. B. Tett, J.-F. Exbrayat, A. A. Bloom, and M. Williams. 2017. "Global Evaluation of Gross Primary Productivity in the JULES Land Surface Model v3.4.1." *Geoscientific Model Development* 10, no. 7: 2651–2670. https://doi.org/10.5194/gmd-10-2651-2017.

Smallman, T. L., J.-F. Exbrayat, M. Mencuccini, A. A. Bloom, and M. Williams. 2017. "Assimilation of Repeated Woody Biomass Observations Constrains Decadal Ecosystem Carbon Cycle Uncertainty in Aggrading Forests." *Journal of Geophysical Research: Biogeosciences* 122, no. 3: 528–545. https://doi.org/10.1002/2016JG003520.

Smallman, T. L., D. T. Milodowski, E. S. Neto, G. Koren, J. Ometto, and M. Williams. 2021. "Parameter Uncertainty Dominates C Cycle Forecast Errors Over Most of Brazil for the 21st Century." *Earth System Dynamics Discussions* 2021: 1–52. https://doi.org/10.5194/esd-2021-17.

Smallman, T. L., D. T. Milodowski, and M. Williams. 2022. "From Ecosystem Observation to Environmental Decision-Making: Model-Data Fusion as an Operational Tool." *Frontiers in Forests and Global Change* 4: 818661. https://doi.org/10.3389/ffgc.2021.818661.

Smallman, T. L., and M. Williams. 2019. "Description and Validation of an Intermediate Complexity Model for Ecosystem Photosynthesis and Evapotranspiration: ACM-GPP-ETv1." *Geoscientific Model Development* 12, no. 6: 2227–2253. https://doi.org/10.5194/gmd-12-2227-2019.

Sousa, D., J. B. Fisher, F. R. Galvan, et al. 2021. "Tree Canopies Reflect Mycorrhizal Composition." *Geophysical Research Letters* 48, no. 10: e2021GL092764. https://doi.org/10.1029/2021GL092764.

Spadavecchia, L., M. Williams, and B. E. Law. 2011. "Uncertainty in Predictions of Forest Carbon Dynamics: Separating Driver Error From Model Error." *Ecological Applications* 21, no. 5: 1506–1522. https://doi.org/10.1890/09-1183.1.

Stettz, S. G., N. C. Parazoo, A. A. Bloom, et al. 2022. "Resolving Temperature Limitation on Spring Productivity in an Evergreen Conifer Forest Using a Model–Data Fusion Framework." *Biogeosciences* 19, no. 2: 541–558. https://doi.org/10.5194/bg-19-541-2022.

Sun, Y., C. Frankenberg, M. Jung, et al. 2018. "Overview of Solar-Induced Chlorophyll Fluorescence (Sif) From the Orbiting Carbon Observatory-2: Retrieval, Cross-Mission Comparison, and Global Monitoring for GPP." *Remote Sensing of Environment* 209: 808–823. https://doi.org/10.1016/j.rse.2018.02.016.

Szwarcman, D., S. Roy, P. Fraccaro, et al. 2024. "Prithvi-Eo-2. 0: A Versatile Multi-Temporal Foundation Model for Earth Observation Applications." *arXiv.* https://doi.org/10.48550/ARXIV.2412.02732.

The SEOSAW partnership. 2021. "A Network to Understand the Changing Socio-Ecology of the Southern African Woodlands (Seosaw): Challenges, Benefits, and Methods." *Plants, People, Planet* 3, no. 3: 249–267. https://doi.org/10.1002/ppp3.10168.

Thomas, R. Q., and M. Williams. 2014. "A Model Using Marginal Efficiency of Investment to Analyze Carbon and Nitrogen Interactions in Terrestrial Ecosystems (ACONITE Version 1)." *Geoscientific Model Development* 7, no. 5: 2015–2037. https://doi.org/10.5194/gmd-7-2015-2014.

Thomas, R. Q., M. Williams, M. A. Cavaleri, J.-F. Exbrayat, T. L. Smallman, and L. E. Street. 2019. "Alternate Trait-Based Leaf Respiration Schemes Evaluated at Ecosystem-Scale Through Carbon Optimization Modeling and Canopy Property Data." *Journal of Advances in Modeling Earth Systems* 11, no. 12: 4629–4644. https://doi.org/10.1029/2019MS001679.

Townsend, P., M. M. Gierach, C. Ade, et al. 2023. "Space-Based Imaging Spectroscopy and Thermal Pathfinder (Sister)sister: Aviris-Ng l2a Resampled Surface Reflectance and Uncertainty 30 m v002." *ORNL Distributed Active Archive Center.* https://doi.org/10.3334/ORNLDAAC/2191.

Trugman, A. T., D. Medvigy, J. S. Mankin, and W. R. L. Anderegg. 2018. "Soil Moisture Stress as a Major Driver of Carbon Cycle Uncertainty." *Geophysical Research Letters* 45, no. 13: 6495–6503. https://doi.org/10.1029/2018GL078131.

Ullrich, P. A., E. A. Barnes, W. D. Collins, et al. 2025. "Recommendations for Comprehensive and Independent Evaluation of Machine Learning-Based Earth System Models." *Journal of Geophysical Research: Atmospheres* 2, no. 1: e2024JH000496. https://doi.org/10.1029/2024JH000496.

Van Bodegom, P. M., J. C. Douma, J. P. M. Witte, J. C. Ordoñez, R. P. Bartholomeus, and R. Aerts. 2012. "Going Beyond Limitations of Plant Functional Types When Predicting Global Ecosystem–Atmosphere Fluxes: Exploring the Merits of Traits-Based Approaches." *Global Ecology and Biogeography* 21, no. 6: 625–636. https://doi.org/10.1111/j.1466-8238.2011.00717.x.

Varney, R. M., S. E. Chadburn, E. J. Burke, and P. M. Cox. 2022. "Evaluation of Soil Carbon Simulation in CMIP6 Earth System Models." *Biogeosciences* 19, no. 19: 4671–4704. https://doi.org/10.5194/bg-19-4671-2022.

Vaswani, A., N. Shazeer, N. Parmar, et al. 2017. "Attention Is All You Need." *arXiv.* https://doi.org/10.48550/ARXIV.1706.03762.

Wang, Y., R. K. Braghiere, M. Longo, et al. 2023. "Modeling Global Vegetation Gross Primary Productivity, Transpiration and Hyperspectral Canopy Radiative Transfer Simultaneously Using a Next Generation Land Surface Model—Clima Land." *Journal of Advances in Modeling Earth Systems* 15, no. 3: e2021MS002964. https://doi.org/10.1029/2021MS002964.

Wang, Y., P. Ciais, D. Goll, et al. 2018. "GOLUM-CNP v1.0: A Data-Driven Modeling of Carbon, Nitrogen and Phosphorus Cycles in Major Terrestrial Biomes." *Geoscientific Model Development* 11, no. 9: 3903–3928. https://doi.org/10.5194/gmd-11-3903-2018.

Wang, Y., L. Shi, X. Hu, W. Song, and L. Wang. 2022. "Multiphysics-Informed Neural Networks for Coupled Soil Hydrothermal Modeling." *Water Resources Research* 59, no. 1: e2022WR031960. https://doi.org/10.1029/2022WR031960.

Watkins, M. M., D. N. Wiese, D. Yuan, C. Boening, and F. W. Landerer. 2015. "Improved Methods for Observing Earth's Time Variable Mass Distribution With GRACE Using Spherical Cap Mascons." *Journal of Geophysical Research: Solid Earth* 120, no. 4: 2648–2671. https://doi.org/10.1002/2014JB011547.

White, E. D., M. Rigby, M. F. Lunt, et al. 2019. "Quantifying the UK'S Carbon Dioxide Flux: An Atmospheric Inverse Modelling Approach Using a Regional Measurement Network." *Atmospheric Chemistry and Physics* 19, no. 7: 4345–4365. https://doi.org/10.5194/acp-19-4345-2019.

Williams, M. 2022. "Global Carbon Cycle Data Assimilation Using Earth Observation." In *Land Carbon Cycle Modeling*, edited by Y. Luo and B. Smith, 1st ed., 225–235. CRC Press. https://doi.org/10.1201/9780429155659-34.

Williams, M., D. T. Milodowski, T. L. Smallman, et al. 2024. "Precipitation-Fire-Functional Interactions Control Biomass Stocks and Carbon Exchanges Across the World's Largest Savanna." *Biogeosciences* 22, no. 6: 1597–1614. https://doi.org/10.5194/egusphere-2024-2497.

Williams, M., E. B. Rastetter, D. N. Fernandes, et al. 1996. "Modelling the Soil-Plant-Atmosphere Continuum in a *Quercus–Acer* stand at Harvard Forest: The Regulation of Stomatal Conductance by Light, Nitrogen and Soil/Plant Hydraulic Properties." *Plant, Cell & Environment* 19, no. 8: 911–927. https://doi.org/10.1111/j.1365-3040.1996.tb00456.x.

Williams, M., E. B. Rastetter, D. N. Fernandes, M. L. Goulden, G. R. Shaver, and L. C. Johnson. 1997. "Predicting Gross Primary Productivity in Terrestrial Ecosystems." *Ecological Applications* 7, no. 3: 882–894. https://doi.org/10.1890/1051-0761(1997)007[0882:PGPPIT]2.0.CO;2.

Williams, M., A. D. Richardson, M. Reichstein, et al. 2009. "Improving Land Surface Models With FLUXNET Data." *Biogeosciences* 6, no. 7: 1341–1359. https://doi.org/10.5194/bg-6-1341-2009.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. 2005. "An Improved Analysis of Forest Carbon Dynamics Using Data Assimilation." *Global Change Biology* 11, no. 1: 89–105. https://doi.org/10.1111/j.1365-2486.2004.00891.x.

Worden, J., S. Saatchi, M. Keller, et al. 2021. "Satellite Observations of the Tropical Terrestrial Carbon Balance and Interactions With the Water Cycle During the 21st Century." *Reviews of Geophysics* 59, no. 1: e2020RG000711. https://doi.org/10.1029/2020RG000711.

Worden, M. A., C. A. Famiglietti, P. A. Levine, et al. 2024. "Inferred Drought-Induced Plant Allocation Shifts and Their Impact on Drought Legacy at a Tropical Forest Site." *Global Change Biology* 30, no. 5: e17287. https://doi.org/10.1111/gcb.17287.

Wu, M., M. Scholze, T. Kaminski, M. Voßbeck, and T. Tagesson. 2020. "Using SMOS Soil Moisture Data Combining CO_2 Flask Samples to Constrain Carbon Fluxes During 2010–2015 Within a Carbon Cycle Data Assimilation System (Ccdas)." *Remote Sensing of Environment* 240: 111719. https://doi.org/10.1016/j.rse.2020.111719.

Xiao, J., K. J. Davis, N. M. Urban, and K. Keller. 2014. "Uncertainty in Model Parameters and Regional Carbon Fluxes: A Model-Data Fusion Approach." *Agricultural and Forest Meteorology* 189: 175–186. https://doi.org/10.1016/j.agrformet.2014.01.022.

Xu, L., S. S. Saatchi, Y. Yang, et al. 2021. "Changes in Global Terrestrial Live Biomass Over the 21st Century." *Science Advances* 7, no. 27: eabe9829. https://doi.org/10.1126/sciadv.abe9829.

Yang, Y., A. A. Bloom, S. Ma, et al. 2022. "CARDAMOM-FluxVal Version 1.0: A FLUXNET-Based Validation System for CARDAMOM Carbon and Water Flux Estimates." *Geoscientific Model Development* 15, no. 4: 1789–1802. https://doi.org/10.5194/gmd-15-1789-2022.

Yin, Y., A. A. Bloom, J. Worden, et al. 2020. "Fire Decline in Dry Tropical Ecosystems Enhances Decadal Land Carbon Sink." *Nature Communications* 11, no. 1: 1900. https://doi.org/10.1038/s41467-020-15852-2.

Yu, T., Y. Pang, R. Sun, and X. Niu. 2022. "Spatial Downscaling of Vegetation Productivity in the Forest From Deep Learning." *IEEE Access* 10: 104449–104460. https://doi.org/10.1109/ACCESS.2022.3210218.

Zhao, Y., X. Chen, T. L. Smallman, S. Flack-Prain, D. T. Milodowski, and M. Williams. 2020. "Characterizing the Error and Bias of Remotely Sensed Lai Products: An Example for Tropical and Subtropical Evergreen Forests in South China." *Remote Sensing* 12, no. 19: 3122. https://doi.org/10.3390/rs12193122.