Parameter estimation in land surface models: challenges and opportunities with data assimilation and machine learning

Raoult, N. ORCID: https://orcid.org/0000-0003-2907-9456, Douglas, N. ORCID: https://orcid.org/0000-0002-3404-8761, MacBean, N. ORCID: https://orcid.org/0000-0001-6797-4836, Kolassa, J. ORCID: https://orcid.org/0000-0001-6644-8789, Quaife, T. ORCID: https://orcid.org/0000-0001-6896-4613, Roberts, A. G. ORCID: https://orcid.org/0009-0002-4274-7914, Fisher, R. ORCID: https://orcid.org/0000-0003-3260-9227, Fer, I., Bacour, C. ORCID: https://orcid.org/0000-0002-1913-3722, Dagon, K. ORCID: https://orcid.org/0000-0002-4518-8225, Hawkins, L., Carvalhais, N. ORCID: https://orcid.org/0000-0003-0465-1436, Cooper, E. ORCID: https://orcid.org/0000-0002-1575-4222, Dietze, M. C., Gentine, P. ORCID: https://orcid.org/0000-0002-0845-8345, Kaminski, T., Kennedy, D. ORCID: https://orcid.org/0000-0001-9494-3509, Liddy, H. M. ORCID: https://orcid.org/0000-0002-8666-0805, Moore, D. J. P. ORCID: https://orcid.org/0000-0002-6462-3288, Peylin, P. ORCID: https://orcid.org/0000-0001-9335-6994, Pinnington, E., Sanderson, B., Scholze, M. ORCID: https://orcid.org/0000-0002-3474-5938, Seiler, C. ORCID: https://orcid.org/0000-0002-2092-0168, Smallman, T. L. ORCID: https://orcid.org/0000-0002-0835-1003, Vergopolan, N.

JAMES | Journal of Advances in Modeling Earth Systems

REVIEW ARTICLE
10.1029/2024MS004733

Key Points:

- Data assimilation (DA) has been shown to be a powerful tool for reducing land surface model (LSM) parametric uncertainty
- Machine learning can facilitate parameter estimation by enhancing computational efficiency and replacing poorly represented processes
- Collaboration is key to advancing LSM calibration and DA, promoting knowledge exchange and standard methods

Correspondence to:

N. Raoult,
nina.raoult@ecmwf.int

Citation:

Raoult, N., Douglas, N., MacBean, N., Kolassa, J., Quaife, T., Roberts, A. G., et al. (2025). Parameter estimation in land surface models: Challenges and opportunities with data assimilation and machine learning. *Journal of Advances in Modeling Earth Systems*, *17*, e2024MS004733. https://doi.org/10.1029/2024MS004733

Received 4 OCT 2024
Accepted 20 SEP 2025

Parameter Estimation in Land Surface Models: Challenges and Opportunities With Data Assimilation and Machine Learning

Nina Raoult[1,2], Natalie Douglas[3], Natasha MacBean[4,5], Jana Kolassa[6,7], Tristan Quaife[3], Andrew G. Roberts[8], Rosie Fisher[9], Istem Fer[10], Cédric Bacour[11], Katherine Dagon[12], Linnia Hawkins[13], Nuno Carvalhais[14,15,16], Elizabeth Cooper[17], Michael C. Dietze[18], Pierre Gentine[13], Thomas Kaminski[19], Daniel Kennedy[12], Hannah M. Liddy[20,21], David J. P. Moore[22], Philippe Peylin[11], Ewan Pinnington[2], Benjamin Sanderson[9], Marko Scholze[23], Christian Seiler[24], T. Luke Smallman[25], Noemi Vergopolan[26,27], Toni Viskari[28], Mathew Williams[25], and John Zobitz[29]

[1]Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK, [2]European Centre for Medium-Range Weather Forecasts, Reading, UK, [3]National Centre for Earth Observation, Department of Meteorology, University of Reading, Reading, UK, [4]Department of Geography and Environment, Western University, London, ON, Canada, [5]Department of Biology, Western University, London, ON, Canada, [6]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [7]Science Systems and Applications, Inc., Lanham, MD, USA, [8]Computing and Data Sciences, Boston University, Boston, MA, USA, [9]CICERO, Olso, Norway, [10]Finnish Meteorological Institute, Helsinki, Finland, [11]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France, [12]NSF National Center for Atmospheric Research, Boulder, CO, USA, [13]Earth and Environmental Engineering Department, Columbia University, New York, NY, USA, [14]Max Planck Institute for Biogeochemistry, Jena, Germany, [15]Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, Caparica, Portugal, [16]ELLIS Unit Jena, Jena, Germany, [17]UK Centre for Ecology and Hydrology, Wallingford, UK, [18]Department of Earth & Environment, Boston University, Boston, MA, USA, [19]The Inversion Lab, Hamburg, Germany, [20]Columbia Climate School, Columbia University, New York, NY, USA, [21]NASA Goddard Institute for Space Studies, New York, NY, USA, [22]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA, [23]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden, [24]School of Environmental Studies, Queen's University, Kingston, ON, Canada, [25]School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh, UK, [26]Ken Kennedy Institute, Rice University, Houston, TX, USA, [27]Earth Environment and Planetary Sciences, Rice University, Houston, TX, USA, [28]European Commission, Joint Research Center (JRC), Ispra, Italy, [29]Department of Mathematics, Computer Science, and Data Science, Augsburg University, Minneapolis, MN, USA

Abstract Accurately predicting terrestrial ecosystem responses to climate change over long-timescales is crucial for addressing global challenges. This relies on mechanistic modeling of ecosystem processes through land surface models (LSMs). Despite their importance, LSMs face significant uncertainties due to poorly constrained parameters, especially in carbon cycle predictions. This paper reviews the progress made in using data assimilation (DA) for LSM parameter optimization, focusing on carbon-water-vegetation interactions, as well as discussing the technical challenges faced by the community. These challenges include identifying sensitive model parameters and their prior distributions, characterizing errors due to observation biases and model-data inconsistencies, developing observation operators to interface between the model and the observations, tackling spatial and temporal heterogeneity as well as dealing with large and multiple data sets, and including the spin-up and historical period in the assimilation window. We outline how machine learning (ML) can help address these issues, proposing different avenues for future work that integrate ML and DA to reduce uncertainties in LSMs. We conclude by highlighting future priorities, including the need for international collaborations, to fully leverage the wealth of available Earth observation data sets, harness ML advances, and enhance the predictive capabilities of LSMs.

Plain Language Summary Improving the accuracy of land surface models (LSMs) is crucial for reducing uncertainties in climate change projections. Parameter data assimilation (DA), which fine-tunes model parameters to better match observed data, is key to enhancing LSM performance. However, the complexity of LSMs poses challenges for global optimization. Advances in computational power, novel data sets, and machine learning (ML) offer promising solutions to improve LSMs. ML can streamline the DA process, handling large data sets and reducing computational demands. This article discusses the progress made in LSM parameter

estimation and the challenges faced by the community. We then discuss how ML can help address these challenges and outline future priorities. International collaboration, fostered by initiatives like the Analysis, Integration and Modeling of the Earth System Land DA Working Group and the International Land Model Forum, is essential for accelerating progress, facilitating knowledge exchange, and developing standardized methods for more accurate climate modeling.

1. Introduction and Premise

Our world faces unprecedented climate change, water scarcity, and food security challenges. To tackle these issues effectively, we need to predict the responses of terrestrial ecosystem dynamics to future global change at inter-annual to centennial timescales. This strongly relies on our ability to accurately model the underlying interactions between the Earth's surface and the atmosphere at the global scale (Watson-Parris, 2021). Such global-scale, mechanistic or process-based models of the terrestrial biosphere, are often embedded in Earth system models (ESMs) (wherein they are called land surface models—LSMs; Blyth et al., 2021). LSMs mathematically represent complex interacting ecosystem vegetation, carbon, water and energy cycling processes over half-hourly to centennial time scales. Thus, for a given atmospheric CO_2 or anthropogenic emissions scenario, LSMs are used to predict the response of terrestrial ecosystems to climate change, rising CO_2 and land use change, and the resultant feedbacks to climate. LSMs are indispensable tools not only for assessing climate change mitigation strategies—such as evaluating the effectiveness of nature-based solutions like reforestation in curbing rising CO_2 emissions—but also for broader applications in carbon monitoring systems. They often serve as prior estimates for atmospheric inversion studies, which are critical for quantifying global and regional land-atmosphere carbon exchanges.

LSMs are highly complex and subject to large uncertainties, both in terms of missing processes, inadequate or incorrect representation of processes, uncertainties in meteorological forcing, and poorly constrained parameters. Adding new processes to address model structural uncertainty often introduces additional parameters increasing parametric uncertainty. Furthermore, current LSMs differ significantly in how they represent many terrestrial processes (Gier et al., 2024; Green et al., 2024; Varney et al., 2024), reflecting differences in model structure, assumptions, and parameterization choices. As a result, their projections of ecosystem responses to future climate change often diverge. For example, LSMs disagree on the magnitude of the land carbon sink (Koven et al., 2022; Z. Shi et al., 2024), and the potential constraints on CO_2 fertilization due to water (Green et al., 2019) and nutrient (Davies-Barnard et al., 2022) limitations.

Parametric uncertainty remains one of the largest sources of uncertainty in all types of process-based land models, along with unrepresented processes (which we do not address in this paper). This is particularly true for long timescale predictions of carbon cycling, vegetation dynamics and climate-carbon cycle feedbacks (Booth et al., 2012; Dietze, 2017; Fisher et al., 2019; Smallman et al., 2021), where LSM predictions can be highly sensitive to parameter choices (Buotte et al., 2021; Exbrayat et al., 2014; Oberpriller et al., 2022; Zaehle, Friedlingstein, & Friend, 2010). Indeed, perturbing a single carbon flux parameter within its range of uncertainty can result in a wider spread in projected atmospheric CO_2 levels by the year 2100 than using different emissions scenarios (Booth et al., 2012). Additionally, while some parameters can be directly observed (e.g., photosynthetic capacity, wood density, hydraulic and thermal properties of snow and soil, bark thickness, tissue nutrient stoichiometry), many parameters either cannot be easily measured (e.g., rooting depth) or are essentially only "effective" parameters in that they have no physical meaning (e.g., water stress factor). Those parameters that can be directly measured are often only observable at scales that differ from the LSM model grid resolution (typically 0.5° or coarser). We urgently need to reduce parametric uncertainty to ensure we can utilize the full potential of LSMs—parameter optimization is one way to achieve this. Fortunately, we are in an era of unprecedented data availability, where different satellite products can help constrain specific processes and their associated parameters (Figure 1).

Historically, LSM parameters have been manually tuned—adjusted by trial-and-error or expert-elicitation to produce more realistic model behavior or to better fit a given important model variable to a given data set. However, manual tuning has increasingly become inefficient and unsustainable; the introduction of new model components or updates to model structure frequently disrupts previous tuning, requiring the entire process to be

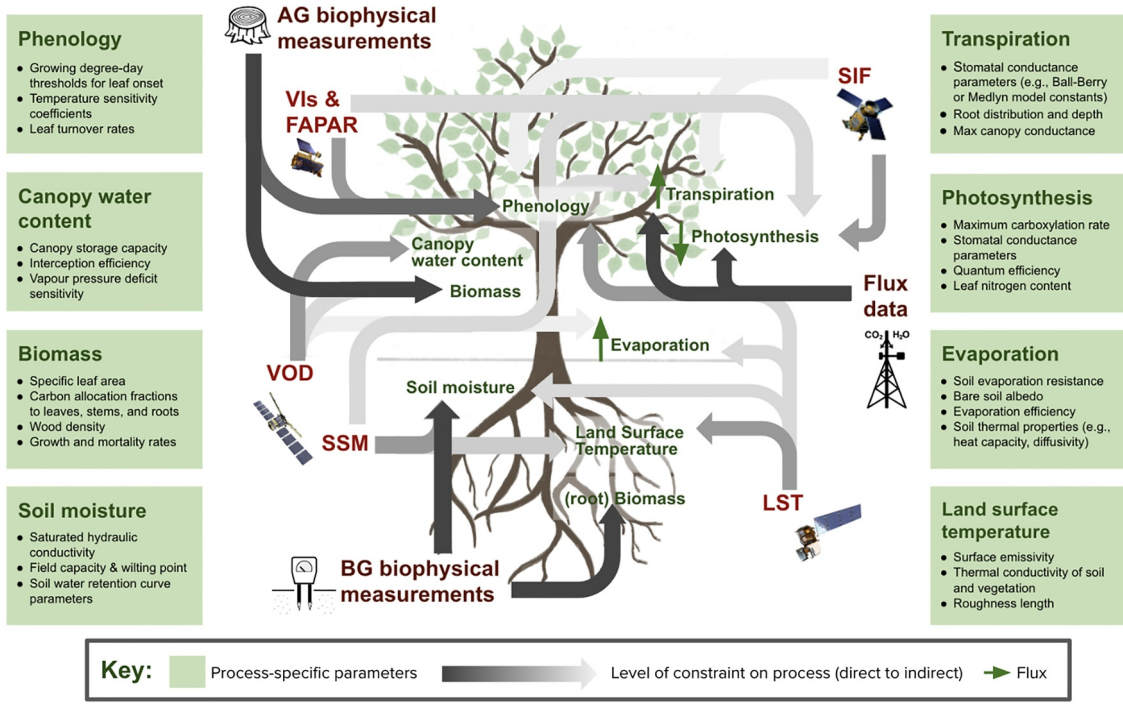

Figure 1. Conceptual schematic illustrating how various in situ and satellite products can be used to constrain different processes, thereby enabling the calibration of each process's parameters. The lists of parameters in each box are a handful illustrative examples; actual parameters may vary across different land surface models (LSMs) and configurations of a given LSM. Likewise, the examples of processes and observational products are not exhaustive, nor are the process-product relationships shown. In situ measurements are categorized into flux data and above- and below-ground (AG and BG) biophysical measurements, shown with examples of the type of processes they can help constrain. Satellite products shown include vegetation indices and fraction of absorbed photosynthetically active radiation, vegetation optical depth and surface soil moisture, solar-induced fluorescence, and land surface temperature. For an overview of the typical spatiotemporal scales of these observational data types, refer to Figure 2 in Scholze et al. (2017).

repeated—a task that becomes untenable as LSMs grow more complex. In the last two decades, improvements in computational power and data availability have made statistically robust parameter calibration feasible, prompting LSM groups to begin applying DA techniques. A selective history of these efforts—spanning early carbon cycle calibration systems to recent large-scale, multi-variable DA experiments—is summarized in Table 1.

DA methods are powerful as they allow observational data to be combined with numerical methods to optimize estimates of chosen variables (state and/or parameters) while accounting for uncertainties in both the model and the data (Rayner et al., 2019). In contrast to LSMs, the atmospheric and ocean components of Earth system models (ESMs) rely on fluid dynamic models governed by well-established physical laws—even if difficult to solve numerically—and many of their parameters are known or directly observable. DA activities in these components have thus far been heavily focused on numerical weather forecasting and reanalysis applications, for which estimating and correcting the optimal model state at each time step is the primary goal (de Rosnay et al., 2022; Hersbach et al., 2018; Zuo et al., 2019). In LSMs, however, which contain processes that are not as readily observable or instrumented, and run at timescales beyond sub-seasonal, parametric and structural uncertainties become increasingly important, eventually dominating their spread at multi-year and longer time horizons (Bonan & Doney, 2018; Draper, 2021; Luo et al., 2015). Additionally, LSM parameters are often linked to biological processes and organismal traits and are dependent on plant functional type (PFT) and so have a wide range of possible values. Characterizing and simplifying the wide variety of vegetation types, functional traits, and ecosystem processes into relatively few parameters is thus a challenge faced in LSM development that is less of an issue for atmospheric and ocean modeling.

While substantial progress in global DA experiments for complex LSM parameter estimation has been made, several challenges remain—most notably the high computational cost associated with utilizing the full high-dimensional parameter space (even when run in standalone mode). Groups are opting for computationally

Journal of Advances in Modeling Earth Systems 10.1029/2024MS004733

Table 1
Selective History of Data Assimilation Efforts Illustrating the Progression of Parameter Estimation in Land Surface Models

Period	Focus and developments	Key studies and citations	Data types used	Scale
1990s–early 2000s	Early global model calibration: optimizing vegetation and carbon cycle parameters in simple/intermediate complexity models	Knorr and Heimann (1995), Knorr and Heimann (2001), Kaminski et al. (2002), Rayner et al. (2005), Scholze et al. (2007)	Atmospheric CO_2, site CO_2 fluxes	Site-level/global
	Local flux tower data to optimize site-level photosynthesis, respiration, and energy flux	Y.-P. Wang et al. (2001), Sacks et al. (2006), Williams et al. (2005), Moore et al. (2008)	Eddy covariance flux tower data (CO_2, energy flux)	Site-level
Late 2000s	Intercomparison projects assessing DA methods: OptIC and REFLEX.	Trudinger et al. (2007), A. Fox et al. (2009)	Various site-scale flux measurements	Site to regional
	Parameter optimization in computationally expensive land models using DA	Santaren et al. (2007), Medvigy et al. (2009), Rayner (2010)	Flux tower CO_2, biomass	Site-level
Late 2000s–mid 2010s	Continued DA testing: site scale with fluxes, stocks, energy, water combinations	Braswell et al. (2005), Xu et al. (2006), Moore et al. (2008), Ricciuto et al. (2008, 2011), Medvigy et al. (2009), Richardson et al. (2010), Verbeeck et al. (2011), Weng et al. (2011), Weng and Luo (2011), Dietze et al. (2014), Keenan et al. (2013), Thum et al. (2017), Wutzler and Carvalhais (2014), Bloom and Williams (2015), Bloom et al. (2016), Bastrikov et al. (2018), Raczka et al. (2018)	CO_2 fluxes, biomass, energy and water fluxes	Site-level
	"Multi-site" experiments using FLUXNET network with various LSMs	Carvalhais et al. (2008, 2010), Knorr et al. (2010), Groenendijk et al. (2011), Kato et al. (2013), Xiao et al. (2014), M. Wu et al. (2018), Kuppel et al. (2012, 2014), Alton (2013), Raoult et al. (2016), Chaney et al. (2016), Post et al. (2017)	Multi-site flux towers (CO_2, energy), grouped by PFT	Site to regional
	Use of in situ CO_2 mole concentration data for regional/global CO_2 exchange	Kaminski et al. (2013), Peylin et al. (2016), Schürmann et al. (2016)	Atmospheric CO_2 mole fractions	Regional/global
				Regional/global

Journal of Advances in Modeling Earth Systems

10.1029/2024MS004733

Table 1
Continued

Period	Focus and developments	Key studies and citations	Data types used	Scale
Mid 2010s–early 2020s	Remote sensing constraints on vegetation dynamics, structure, and disturbance	MacBean et al. (2015), Knorr et al. (2010), Kaminski et al. (2012), Bacour et al. (2015), Forkel et al. (2014, 2019), Stöckli et al. (2008), Zobitz et al. (2014), MacBean et al. (2018), Bacour et al. (2019), Forkel et al. (2019), Norton et al. (2018, 2019), J. Wang et al. (2021), Knorr et al. (2025), Forkel et al. (2019), Shiklomanov et al. (2021)	Satellite: reflectance, normalized different vegetation index, solar induced fluorescence, aboveground biomass, burned area	
	Constraining carbon, water, and energy cycles with new satellite products	Scholze et al. (2016, 2019), Knorr et al. (2025), E. Pinnington et al. (2018, 2021), Raoult et al. (2021)	Satellite: XCO_2, VOD, soil moisture, snow cover, river flow	Regional/global
	Hydrological DA ensemble methods adopted for LSM joint parameter-state estimation	Moradkhani et al. (2005), Nie et al. (2011), Ait-El-Fquih et al. (2016), M. Lu et al. (2022)	Soil moisture, river flow, precipitation (in situ, remote sensing)	Site to catchment scale

Note. Further details on the history of parameter optimization in various land surface models are provided in Kaminski and Rayner (2008), Kaminski et al. (2010, 2013), Scholze et al. (2017), Rayner et al. (2019), Baatz et al. (2021), and MacBean, Bacour, et al. (2022).

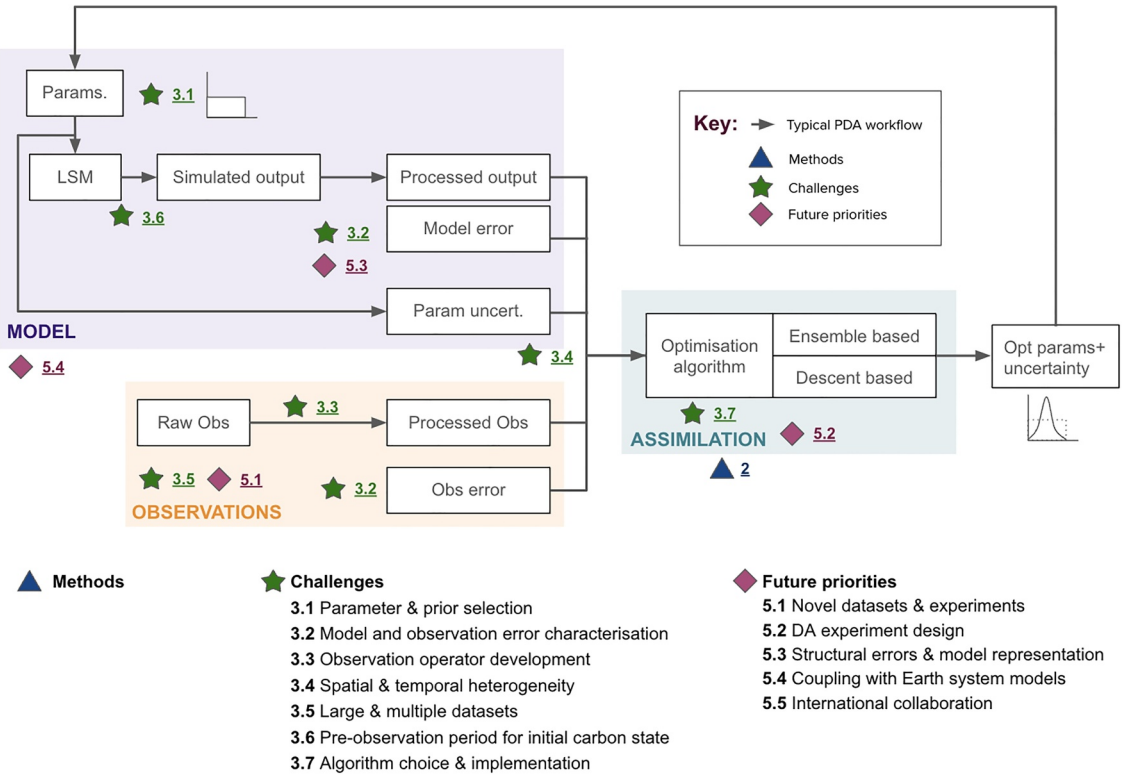

Figure 2. Schematic of the parameter data assimilation (PDA) workflow for land surface models, illustrating the interactions between model, observations, and assimilation components. Boxes show the main elements of the pipeline, with arrows indicating the typical PDA workflow. Methods (blue triangle, Section 2), challenges (green stars, Section 3) and future priorities (purple diamonds, Section 5) are mapped onto the relevant parts of the workflow, with numbering corresponding to subsections in the text. Section 4, which discusses how machine learning can facilitate the workflow is not shown on this figure, but is instead included as part of Figure 3 at the beginning of Section 4.

cheap inversion methods over extensive statistical parameter searches to calibrate their models. Currently, there is no overall community strategy for objective parameter estimation for complex LSMs. However, this rapidly expanding field offers a unique opportunity to learn from each other and address common technical challenges. Efforts to build a Land DA Community (https://land-da-community.github.io/) by the Analysis, Integration and Modeling of the Earth System (AIMES) Land DA Working Group (MacBean, Liddy, et al., 2022) and the International Land Model Forum (ILMF—https://hydro-jules.org/international-land-modeling-forum-ilmf) have precipitated this sharing of knowledge through online workshops and town halls. Capitalizing on this momentum is vital given the importance of this problem. The rapid advancements in machine learning (ML) and the increasing availability of global Earth observations and networks of in situ data create new opportunities for advancing land/Earth system modeling with the help of DA.

In this paper, we review the current state of LSM parameter estimation (Section 2), discuss key challenges and opportunities (Section 3), explore potential solutions via ML and computational advances (Section 4), and propose priorities for advancing the field given the urgent need for improved model projections (Section 5). To aid navigation, we also provide a graphical Table of Contents (Figure 2) that illustrates the parameter estimation pipeline and situates the structure of the paper within this framework. The focus of this paper is on parameter estimation for carbon-vegetation-water interactions, complementing Kumar et al. (2022)'s review paper on state estimation for vegetation and hydrology, and De Lannoy et al. (2022)'s paper on joint parameter-station estimation for the water cycle.

2. Data Assimilation Methods for Parameter Estimation in Land Surface Models

With the emergence of novel in situ and satellite observation data sets came the advent and development of techniques to combine them with models in a field of mathematics originally coined DA (Talagrand & Courtier, 1987). Data Assimilation (DA) techniques attempt to find an optimal balance between the knowledge

acquired from both the models and the observations whilst incorporating their respective uncertainty information. There are essentially two types of DA methods that differ in the nature of the temporal assimilation of the available observations; DA methods that assimilate all available observations over a given time window are known as batch (or offline/smoothers) techniques whereas those that incorporate the observations at the time they become available are referred to as sequential (or online/filters). We note here that there is some confusion in the community regarding the terminology used when describing DA methods, for example, the false dichotomies sometimes used between "variational versus sequential" and "optimization-based versus Bayesian"—these dichotomies have been marred over time with hybridization and the continual development of the techniques. Rayner et al. (2019) have made a significant effort to harmonize the notation by clarifying overlapping terminology within the community.

Although DA is primarily used in numerical weather forecasting to correct the model state—as this is the primary source of uncertainty on hourly to weekly timescales—in LSMs, DA is often employed to reduce parametric uncertainty against observational data, a process referred to here as parameter data assimilation (PDA). By calibrating parameters based on observational data, the problem of PDA in LSMs becomes the solution to the inverse problem (Tarantola, 1987, 2005): *find the parameter set* $\boldsymbol{\Theta}$ *given the observations* $\mathbf{y}$ *such that* $\mathbf{y} \cong \mathrm{G}(\boldsymbol{\Theta})$. In the context of PDA, G includes a mapping from parameters to states and propagates states through time via a forward model as well as an observation operator (Kaminski & Mathieu, 2017) that maps states to observation space. Typically, a unique solution to the exact inverse problem does not exist and often the logical step is to cast the approximate inverse problem into a loss minimization effort that locates the argument of a cost function that minimizes the discrepancy between $\mathbf{y}$ and $\mathrm{G}(\boldsymbol{\Theta})$. However, many techniques of this type only provide point estimates (i.e., a single solution), which have significant limitations when applied to LSM calibration. LSMs are inherently complex, involving many interacting processes, uncertain observations, and non-linear relationships. By focusing only on the best-fit parameters, point estimates ignore the range of plausible values that could explain the data equally well. This can lead to overconfident predictions, underestimating the variability and uncertainty in model outcomes, which is crucial for understanding the full spectrum of possible future climate scenarios. Instead, we want to be able to account for uncertainties in the model, data, and parameters, and reduce the uncertainty in the parameters by creating observationally constrained posterior distributions.

Hence, an approach more desirable for its ability to quantify the uncertainty in the estimated parameters and its inherent natural regularization, is the Bayesian approach. Bayesian methods include information on the prior distribution of the parameters $\mathrm{p}(\boldsymbol{\Theta})$ to define an entire posterior distribution:

$$\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y}) \propto \mathbf{p}(\mathbf{y}|\boldsymbol{\Theta})\,\mathbf{p}(\boldsymbol{\Theta}) \tag{1}$$

where $\boldsymbol{\Theta}$ is regarded as a random variable as opposed to a fixed value to be estimated. In this case, the maximum a posteriori estimate—the argument that maximizes the posterior distribution (i.e., its mode)—provides a point estimate for $\boldsymbol{\Theta}$ and is equivalent to a loss minimization estimate regularized with prior parameter information under Gaussian assumptions. Under such assumptions, maximizing the posterior distribution corresponds to minimizing the so-called variational cost function:

$$J(\boldsymbol{\Theta}) = 1/2\left[(G(\boldsymbol{\Theta}) - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathrm{G}(\boldsymbol{\Theta}) - \mathbf{y}) + (\boldsymbol{\Theta} - \boldsymbol{\Theta}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_{\mathrm{b}})\right], \tag{2}$$

where $\mathbf{R}$ and $\mathbf{B}$ are the model/data and prior error covariance matrices, respectively, and $\boldsymbol{\Theta}_{\mathrm{b}}$ is the prior parameter vector.

2.1. Methods for Minimizing Cost Functions

Methods commonly used to minimize cost functions, such as in Equation 2, often require numerical optimization due to their complex structure, particularly in LSM PDA. For this reason, and for computational prudence, parameter estimation attempts in LSM thus far have mostly employed gradient-descent techniques for minimization.

Gradient-descent methods require the gradient of the cost function; either exactly, which requires differentiating the entire LSM (see Section 3.7), or approximately when analytical differentiation is not possible or desirable. These methods can result in the identification of a local minimum, as opposed to the desired global one, because

they rely primarily on local gradient information, which may not capture the broader structure of the cost function. The most common gradient-based minimization method used in LSM PDA is the quasi-Newton algorithm L-BFGS-B (limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bound constraints—Byrd et al., 1995). This approach can leverage exact gradients derived from either the tangent linear (forward sensitivity propagation) or adjoint (backward sensitivity propagation) of the model as well as approximated gradients. Exact gradients can be obtained by hand or using automatic differentiation software (Gelbrecht et al., 2023; Griewank, 1989). While L-BFGS-B is powerful when exact gradients are available, practical challenges—such as the complexity and computational burden of maintaining the tangent linear/adjoint (see Section 3.7)—often necessitate alternatives. To address this, approximate gradient methods can be incorporated in place of exact gradients. One approach is to estimate gradients using finite differences, calculating the change in model output relative to changes in parameters. This method is especially useful for parameters related to threshold functions, such as those controlling phenology. However, the choice of perturbation size to be applied to each parameter individually is crucial, as inappropriate values can lead to inaccuracies and the number of model iterations required is potentially large compared to other techniques. In cases where gradient information including finite differences is difficult to obtain or unreliable, derivative-free methods offer a solution. The Nelder-Mead simplex algorithm (Nelder & Mead, 1965), for instance, iteratively adjusts a simplex (geometric shape) in parameter space to converge toward the minimum of a cost function, eliminating the need for direct gradient calculations or approximations. Additionally, more advanced approaches, such as the ensemble-based 4DVar (4DEnVar) algorithm proposed by C. Liu et al. (2008) use an ensemble of model trajectories to approximate gradient information via a control variable transform.

Although less commonly used in LSM PDA due to their computational intensity, global search methods provide an alternative to gradient-descent methods for cost function minimization. These techniques scan the full parameter space and in doing so avoid the pitfall of identifying a local minimum. These global search methods are categorized as Monte Carlo (MC), named so since they are methods that make use of repeated trials (or sampling) generated using random numbers (Owen, 2013). An example of one such method is the genetic algorithm (Goldberg & Holland, 1988; Haupt & Haupt, 2004), which is based on the laws of natural selection and belongs to the class of evolutionary algorithms.

Gradient descent and global search find point estimates but do not provide posterior uncertainty. The Hessian at the optimum can approximate this, but computing and inverting it can be costly, even with adjoint methods. Alternatives like sampling or variational inference can estimate uncertainty but add further computational cost.

2.2. Methods to Extract the Full Posterior Distribution

In contrast to methods that obtain point-estimates for the parameters, other approaches aim to extract useful information from the full posterior distribution $P(\boldsymbol{\Theta}|\mathbf{y})$, usually at a much higher computational expense and tend to be applied to computationally inexpensive LSMs, carbon cycle, and ecosystem models. Similarly to global search algorithms for objective function optimization, as opposed to gradient-descent methods, these techniques are often MC in nature and hence also derivative-free.

Techniques include importance sampling (Kloek & Van Dijk, 1978), a relatively simple approach that samples random values from the prior distribution $p(\boldsymbol{\Theta})$ and accumulates accepted parameterizations based on importance weights and aims to estimate expectations of interest such as mean, variance, etc. This approach can run into limitations when the problem becomes more complicated (e.g., dimensionality increases or target distribution gets more complex), as demonstrated by Ziehn et al. (2012). When the computational budget permits, Markov Chain MC (MCMC; Hastings, 1970) algorithms have emerged as the gold standard for quantifying uncertainty in the solution of Bayesian inverse problems. This class of iterative algorithms seeks to draw samples from the posterior distribution $P(\boldsymbol{\Theta}|\mathbf{y})$, which can in turn be used to estimate posterior statistics of interest. The cost of such comprehensive uncertainty quantification is that standard MCMC algorithms often require a large number ($>10^4$–10^7) of iterations that build on previously accepted values and so must be performed serially that is, not taking advantage of parallel high-performance computing. This essentially means that the full LSM must be run using a new parameter vector during each iteration, and while it is possible to run different information-sharing chains in parallel to accelerate sampling around a global optimum (Vrugt, 2016), within chain iterative model evaluations still precludes parallelization.

Particle filters offer an alternative to MCMC for sampling the posterior in time-evolving systems by representing it with particles updated at each new data point. Though computationally intensive and prone to degeneracy, they are effective for real-time tracking of system states and time-varying parameters. Similarly, the Ensemble Kalman Filter (EnKF; Evensen, 2003) uses an ensemble approach with Gaussian error assumptions and Kalman updates to efficiently estimate parameters. Many parameters in LSMs relate to biological processes and can vary over time due to acclimation, plasticity, adaptation, and evolution. While some studies explore seasonal parameter variability (Rowland et al., 2014; Verbeeck et al., 2011), most land model parameter estimation assumes fixed parameters. Consequently, particle and ensemble Kalman filters are rarely used in PDA (Speich et al., 2021) except for joint state-parameter estimation (e.g., H. Zhang et al., 2017).

Due to their computational demands and assumptions, the different methods outlined above have been applied to varying degrees across models and scales; representative applications are summarized in Table 2, as well as their advantages and disadvantages.

3. Challenges

3.1. Selecting Parameters and Their Prior Distributions

A big challenge in parameter estimation studies is defining the experiment, starting with selecting the parameters to be constrained. A common first step is to select from the (potentially quite large) number of model parameters to find a subset i.e. deemed the most influential on the variables to be optimized by some metric or method, while the rest of the parameters are fixed at their default values to reduce dimensionality. This challenge is amplified by large numbers of interconnected parameters influencing different parts of the model as parameters with strong enough covariances need to be considered jointly. Furthermore, the strong co-variations between parameters and forcing and boundary conditions further complicate the parameter selection process. It is vital to identify the key internal parameters that have the most impact on a given model output because (a) PDA techniques are computationally demanding, (unlike an adjoint computation) they scale with the number of parameters used in the optimization, and (b) due to the high degree of equifinality in most inverse problems (i.e., different parameter vectors giving the same fit to the observed data), attempting to estimate an excessive number of parameters can lead to overfitting and a severe degradation in model performance when the model is run in predictive mode. In other words, increasing model complexity for improved prediction is only justified when there are adequate observational constraints to its parameters (Famiglietti et al., 2021). Note that identifying key internal parameters is not a solution in itself to the equifinality issue—it is still possible to have only two key parameters and end up with equifinality. It is also important to reduce the risk of error compensation, where model biases are aliased onto parameters to which the optimized variable is less sensitive, which can occur when diverse data constraints with different observation influence are used in the optimization. This risk can be reduced by aligning observational constraints with physically related parameters (Bacour et al., 2015, 2023; Wutzler & Carvalhais, 2014).

In parameter sensitivity analyses, the choice of model variable, timescale, and evaluation metric (e.g., RMSE vs. mean bias if relying on a model-data comparison) fundamentally affects the parameter selection. The most common parameter sensitivity experiment is a one-factor-at-a-time parameter perturbation experiment. However, this does not account for covariance between parameters, which can vary along ecological tradeoffs and are known to strongly impact LSM outputs (Prihodko et al., 2008). One way to mitigate this is by using spatial pattern correlations to select parameters with low inter-correlation (Dagon et al., 2020). More robust methods, which provide a more comprehensive assessment of parameter influence, include local sensitivity analysis using adjoint models and global methods like Morris screening (Morris, 1991), Sobol variance decomposition (Saltelli et al., 2008; Sobol ′, 2001) and Fourier amplitude sensitivity tests (FAST; Cukier et al., 1973). These methods have been applied to wide range of LSMs including BEPS (Xing et al., 2023), CABLE (X. Lu et al., 2013), CLASSIC (Deepak et al., 2024), CLM4.5(FATES) (Massoud et al., 2019), JULES (Pianosi et al., 2017), Noah-MP (H. Wang et al., 2023) and ORCHIDEE (Dantec-Nédélec et al., 2017; Kuppel et al., 2012; Novick et al., 2022). However, these methods can be computationally intensive—e.g., Sobol often requires O(10,000) runs—or technically complex, especially when adjoint models are involved (see Section 3.7). Still, once an adjoint or ensemble is available, assessing the sensitivity of various model outputs becomes relatively straightforward.

In complex LSMs, even after selecting the most influential parameters, the large number of vegetation (e.g., 15 PFTs in ORCHIDEE) and soil texture classes (e.g., 13 USDA textural classes) used to represent the diversity of terrestrial ecosystems quickly increases the dimensionality of global calibrations, as each parameter can be varied

Table 2
Illustrative Applications of Inversion Methods for Parameter Estimation in Land Surface Models, With Representative Models, Key Studies, and Method Pros and Cons (Definition for the Different Process-Based Models Can Be Found in Table A1)

Method	Applications	Example models	Key studies	Advantages	Disadvantages
MCMC (Markov Chain Monte Carlo)	Computationally inexpensive land/carbon/eco models; isolated process modules	SIPNET, TECOS, FöBAAR, BETHY, DALEC, CARDAMOM, LPJ-GUESS	Fer et al. (2018), M. Liu et al. (2015), Sacks et al. (2006), Keenan et al. (2012), Knorr and Kattge (2005), Famiglietti et al. (2021), Bloom et al. (2016), Smallman et al. (2021), Jones et al. (2024), Kallingal et al. (2024)	– Full posterior distribution – Strong uncertainty quantification	– High computational cost – Rarely used in complex LSMs
4DVar	Complex LSMs with adjoint/ tangent linear models	ORCHIDEE, BETHY, JULES, BEPS	Bacour et al. (2015), Kaminski et al. (2012), Knorr et al. (2025), Kuppel et al. (2012), Raoult et al. (2016), Zhu, Wu, et al. (2024), Zhu, Xing, et al. (2024)	– Efficient gradient-based optimization – Enables posterior error estimation	– Requires adjoint models – Difficult implementation
4DEnVar	Avoids adjoint; ensemble-based gradient estimation	JULES	E. Pinnington et al. (2020)	– No adjoint needed – Hybrid flexibility	– Computationally intensive – Potentially less precise gradients
Nelder-Mead Simplex	Derivative-free local optimization	JULES	E. Pinnington et al. (2018)	– Simple to implement – Works without gradients	– Slow convergence – Prone to local minima
Finite Differences	Gradient approximation for 4DVar-like DA	ORCHIDEE, JULES	Bacour et al. (2019), Bastrikov et al. (2018), MacBean et al. (2015)	– No need for adjoint-Straightforward	– Computationally heavy – Accuracy depends on step size
EnKF (Ensemble Kalman Filter)	State and parameter estimation; soil moisture focus	CLM, SiB	W. Chen et al. (2015), D. Lü et al. (2011), Nie et al. (2011), Y. Shi et al. (2014, 2015), H. Zhang et al. (2017), Deng et al. (2016), Xiong et al. (2019)	– Real-time DA – Handles uncertainty – Popular in hydrology	– Produces time-varying parameters – Sensitive to ensemble properties
Monte Carlo type (GLUE, Genetic Algorithms, DREAM(zs))	Global search and uncertainty quantification at site level	JSBACH, ORCHIDEE, SiB, BEPS, CLM, LPJ-GUESS	Mäkelä et al. (2019), Bastrikov et al. (2018), Prihodko et al. (2008), Xing et al. (2023), Vrugt et al. (2009), Post et al. (2017), Bagnara et al. (2019)	– Flexible and model-agnostic – Global search methods	– Very expensive – Typically limited to site-level calibration

independently. One way to tackle this issue is to assume that the parameter differences among different groups vary proportionally and, therefore, optimize a parameter scaling factor instead of targeting each parameter per group (Fer et al., 2018; McNeall et al., 2024). However, for some plant traits, the "within functional type" uncertainty can be as large as the "across functional type" uncertainty (e.g., Trugman et al., 2020), possibly due to the traits being either weakly constrained by available data or genuinely plastic traits that vary spatially. In the latter case, this variability suggests that regionalizing parameters rather than using PFT-specific parameterizations may be more appropriate. As such, methods that allow for independent tuning of parameters within each PFT, or even regionalization of parameters, may be necessary. Scaling factors can also be used to target processes without needing to deeply explore detailed parameterizations.

Selecting parameters is only part of the challenge; defining realistic prior distributions is equally critical, as they shape the entire assimilation outcome. Parameter sensitivity itself often depends on these priors, yet for many parameters, our knowledge remains limited. A key difference between simple local sensitivity analyses—where all parameters might be varied by an arbitrary amount (e.g., ±10%)—and formal uncertainty partitioning methods, such as global sensitivity analyses (e.g., Morris or Sobol), is whether the input ranges meaningfully reflect prior knowledge before calibration. Such priors can come from empirical data, formal expert elicitation, or other informed sources (Dietze et al., 2014; LeBauer et al., 2013; Raczka et al., 2018). In model calibration, it can be common to assume uniform priors, either explicitly in Bayesian frameworks or implicitly by applying uniform parameter bounds in the optimization. This contrasts with variational DA methods like 4DVar (Equation 2), which typically use explicit Gaussian priors. Often, these uniform ranges are chosen based on informal "expert judgment" or ad hoc trial and error, rather than grounded in evidence. Where available, parameter uncertainties can be informed by in situ measurements (e.g., the TRY database; Kattge et al., 2020), or as a last resort, by applying simple percentage ranges (e.g., ±20% the operational value of the parameter). It is also essential to ensure priors are biogeophysically meaningful—avoiding negative values for inherently positive parameters, preserving dependencies between parameters, and respecting ecological plausibility (e.g., wood longevity exceeding foliage longevity).

While uniform priors can be practical, they rarely reflect true prior knowledge because they assume all values within the range are equally likely, while, in reality, some parameter values are often known a priori to be more plausible than others. More realistic alternatives include adopting probability distributions that better reflect expected parameter behavior, guided by structural constraints such as non-negativity. These can be implemented through transformations of Gaussian PDFs (Kaminski et al., 2012, 2013; Rayner et al., 2005; Ziehn et al., 2011), or informed by synthesizing trait data via meta-analyses and structured expert elicitation (Dietze, 2017; Dietze et al., 2014; LeBauer et al., 2013). Priors informed by trait data act as a kind of data fusion, combining multiple lines of evidence (e.g., Bloom & Williams, 2015). In practice, including prior trait constraints can radically alter which parameters must be estimated. For instance, highly sensitive but well-constrained traits (e.g., the parameter controlling the maximum rate of carboxylation—V_{cmax}) may need less focus than poorly constrained, low-sensitivity parameters that plausibly span multiple orders of magnitude and thus contribute more to overall model predictive uncertainty (Dietze, 2017; LeBauer et al., 2013). Parameter importance can also shift with the forecast horizon (e.g., weekly to decadal; Raczka et al., 2018).

Non-uniform priors can also embed known correlations and trade-offs within or across PFTs (Shiklomanov et al., 2018), though these are often ignored in practice by assuming zero covariances, which can lead to ill-posed inversions. Using informative priors also supports the iterative nature of Bayesian calibration, allowing posteriors to be recycled as new priors when new data become available, which offers significant computational and practical benefits. However, selecting alternative priors still be as problematic as choosing uniform ones, especially when true distributions are unknown, highlighting the need for formal prior predictive checks to validate assumptions (e.g., Bacour et al., 2023). Ultimately, no matter which method is used for parameter estimation, solutions only exist in the parameter space defined by the chosen parameters and their prior ranges (Williamson et al., 2013). Changing the number of parameters, their prior distributions, the observational data sets and/or the model process representation will yield different results due to new parameter interactions and the equifinality of solutions.

3.2. Characterization of Model and Data/Observation Errors

The state-of-the-art way to account for model and observation errors is through a Bayesian framework. However, properly characterizing these errors (especially data bias) can be a challenge and potential model-data biases are not always properly treated with this formalism (Cameron et al., 2022; MacBean et al., 2016). Model discrepancy, or model process error, refers to the inherent inability of a model to replicate observations (J.-L. Wu et al., 2023), stemming from factors such as missing processes, choice of process representation, ecosystem heterogeneity, stochastic processes (e.g., dispersal, recruitment, mortality, disturbance), biases in the model forcing data, uncertainties in the initial model state, and the resolution of numerical solvers. Observation error encompasses sampling variability, instrument inaccuracies, and any errors involved in deriving the data products making up the observations. Furthermore, observation error also usually includes a modeling step from the raw data measurement to any given physical quantity (see Section 3.3). Even though conceptually it is important to distinguish model error covariance and observational error covariance, the standard mathematical formalism expressed in Equation 2 uses as matrix R the sum of both these error covariances.

Although common, combining model error with observational error can lead to an overestimation of predictive uncertainty (van Oijen, 2017). Another approach to deal with model error is to ignore it (i.e., assume the model structure is correct), however, this means only the input uncertainty is propagated. A final approach is to treat model uncertainty as a separate parameter needing calibration. If a prior for the model error uncertainties can be specified explicitly, model and data error terms can theoretically be fitted separately. However, in practice, specifying an informative prior on the model error term is challenging due to incomplete theoretical understanding of the underpinning processes (Brynjarsdóttir & O'Hagan, 2014). Fortunately, it is often much easier to specify an informative prior on the observation error, as these are frequently reported in data products or estimable via sampling theory, and this is often useful to allow model error to be separately identifiable. We note, however, that model uncertainty is difficult to characterize, even when it is separately identifiable. Moreover, reported observational product uncertainty can be overconfident (D. Kennedy et al., 2024; R. E. Kennedy et al., 2024).

There are a number of arguments for keeping process and observation errors distinct. Process error propagates in space and time when making predictions, while observation error does not. Reducing process error requires improving model structure, whereas reducing observation error calls for better data quality. Furthermore, calibrating models with cost functions that rely only on fixed observation errors can distort parameter uncertainty and the relative weight given to each data stream, especially since model skill is not always proportional to measurement accuracy. The overall weight of a data stream also depends on the number of observations: a large volume of uncertain data can outweigh a small but precise data stream if not balanced properly, directly affecting the posterior estimates. There are clear examples where model uncertainty and measurement uncertainty for the same variable do not align. For example, at local scale, predictions of net ecosystem exchange (NEE) are often more uncertain than gross primary productivity (GPP), which only accounts for carbon uptake by photosynthesis. Yet eddy covariance GPP estimates carry extra uncertainty because they rely on an additional step to remove ecosystem respiration from NEE.

Quantifying both observation and model process error correlations, such as autocorrelated measurement error, presents an additional challenge. These correlations yield non-diagonal covariance structures, which are rarely well understood and are often ignored. Nevertheless, accounting for these correlated errors has been shown to improve DA results (Waller et al., 2016), for example, by increasing the information content of observations (Stewart et al., 2008). Since observation error correlations are more prevalent in dense observation networks (Bannister et al., 2020), strategies to mitigate not modeling them include observation thinning (reducing the number of observations assimilated in data-rich regions) and super-obbing (combining many observations into one (Lorenc, 1981)). Another common approach to inflate variances is to reduce the weight of observations in DA (Chevallier, 2007; Kuppel et al., 2013; Scholze et al., 2019; M. Wu et al., 2018). However, all these approaches are subjective and potentially reject meaningful information (Cameron et al., 2022).

Finally, addressing systematic errors in models and data is becoming increasingly crucial as the volume of data grows. With larger data sets, random errors tend to average out, leaving systematic errors to dominate. These errors have long been recognized by the LSM calibration community, such as when a model's ability to predict one variable worsens after assimilating data for another. However, the underlying causes and potential solutions have not been widely recognized. Since all models are approximations, systematic errors in both models and data require greater attention. To combat these biases, various approaches are emerging, ranging from incorporating simple linear bias correction factors in the cost function (Cameron et al., 2022; Fer et al., 2018) to more complex and flexible statistical models of bias, applied either within the assimilation process or post-hoc (M. C. Kennedy & O'Hagan, 2001; Oberpriller et al., 2021). Additionally, hybrid models that integrate ML with process-based models are being explored as a means to address these challenges (see Section 4.2).

Ultimately, interconnected efforts, including characterisating data errors together with the data providers, post-PDA analysis of remaining model-data discrepancies, multi-model PDA protocols that highlight relative model structural errors, and novel PDA algorithms are all valuable ways forward for distinguishing errors in the data from those in the model structure.

3.3. Developing Observation Operators

The term "observation operator" refers to any transformation of modeled quantities that enables comparison against observations (Kaminski & Mathieu, 2017). Note that often, what we call observations are themselves complex transformations of raw data measurements into physical quantities comparable to the LSM output. For

example, satellite-measure radiances at the top of the atmosphere can be translated into any number of land surface data products (such as leaf area index), through intricate retrieval processes (Pinty et al., 2007). Furthermore, these data are usually prepared in such a way that they are available on the model grid.

Sometimes, it is possible to prepare an observational product such that there is a one-to-one relationship between the model output and assimilated data, making the observation operator the identity matrix. More often, however, a more complex observation operator is needed, and its choicecan significantly impact DA results (Cooper et al., 2019). A common use of an observation operator is to bridge the spatial scale between model and observations— aggregating or interpolating the gridded observations and model outputs to compatible resolutions (Anderson et al., 2009; Huo et al., 2024; E. Pinnington et al., 2021; Raczka et al., 2021). More advanced spatial operators use weighted averaging to account for more detailed descriptions of the observation, such as modeling the point spread function of satellite data, or the footprint of an eddy-covariance flux measurement. For example, Vergopolan et al. (2020) developed a cluster-based operator mapping satellite Gaussian footprints to sub-grid scales of high-resolution LSMs, enabling assimilation of coarse soil moisture data alongside fine-scale observations (Vergopolan et al., 2021). E. M. Pinnington et al. (2017) partitioned flux tower data to assimilate separately in logged and unlogged forest stands.

Observation operators can also map surface fluxes to atmospheric concentrations via transport models, as in the assimilation of CO_2 flask measurements to constrain terrestrial biosphere models (Bacour et al., 2023; Kaminski et al., 2002, 2012; Knorr & Heimann, 1995; Peylin et al., 2016; Rayner et al., 2005, 2011; Scholze et al., 2007; M. Wu et al., 2020, 2024) and to evaluate simulated net CO_2 fluxes after optimizing against eddy-covariance data (Kuppel et al., 2014). For non-reactive gas species, it is sufficient to have data on atmospheric transport to drive the observation operator, but for reactive species such as CH_4, the process is more complex as atmospheric chemistry needs to be included.

Another key application is using observation operators to predict observed quantities that are not directly computed by the model itself, such as the assimilation of solar-induced fluorescence (SIF) data, which is typically used to constrain GPP. Observation operators of SIF range from simple linear relationships with GPP (Bloom et al., 2020; MacBean et al., 2018) to empirical and mechanistic models of canopy photochemistry and radiative transfer (Bacour et al., 2019; Knorr et al., 2025; R. Li et al., 2022; Norton et al., 2019). Another example is vegetation optical depth which has been used to constrain above-ground biomass and leaf area index (LAI; Knorr et al., 2025; Scholze et al., 2019).

Scholze et al. (2016, 2019) also developed observation operators to map model state variables onto retrievals (including root zone soil moisture) of surface soil moisture (SSM), which were also used by M. Wu et al. (2018, 2020, 2024). Since SSM is subject to large biases, such transformations are essential (see Section 3.2 on systematic errors). Methods such as cumulative density function matching (Reichle & Koster, 2004) or focusing on dynamics (e.g., dry down events, Raoult et al., 2021) are common when assimilating SSM. Dynamics-based assimilation is also often used with vegetation indices (VIs), FAPAR or LAI retrievals are normalized to estimate the seasonality of phenology instead of the absolute values (MacBean et al., 2015). The optimization then targets a reduced set of phenology-related parameters, rather than those related to photosynthesis (Bacour et al., 2015).

Forward modeling of remote sensed data—simulating satellite observations from LSM outputs via an observation operator—and then assimilating low level (minimally processed) satellite products (as in example of SIF above), is the alternative to assimilating high-level satellite products such as LAI or GPP (i.e., products that are derived from a model and/or highly processed). These products often embed assumptions inconsistent with those in LSMs. For instance, GPP retrievals frequently use light use efficiency models (e.g., MODIS GPP; Running et al., 2021), whereas LSMs typically employ enzyme-kinetic schemes (Collatz et al., 1992; Farquhar et al., 1980), and satellite-derived GPP estimates often rely on environmental drivers like downwelling shortwave radiation, which may differ from those used in the LSM. Furthermore, processing raw observations often introduces complex and unreported uncertainty correlations into the product, complicating the structure of the uncertainty covariance in the assimilation (Kaminski & Mathieu, 2017). Assimilating such products risks tuning model parameters to reflect retrieval assumptions rather than underlying biophysical processes. To avoid this, we recommend assimilating the lowest-level data possible (e.g., fAPAR or surface reflectance when considering GPP; Quaife et al., 2008), thereby minimizing structural mismatch and improving consistency between model and observation. Finally, satellite-derived estimates of variables like GPP or LAI often differ significantly, and any

single product is likely biased relative to the true state. As a result, discrepancies between high-level observations and LSM outputs can be difficult to interpret and quantify. This is why seasonal dynamics are often used instead of absolute values (as noted earlier), as they are less sensitive to such systematic biases.

Assimilating low-level products such as SIF or canopy reflectance (Quaife et al., 2008) often requires radiative transfer models, analogous to radiance assimilation in weather prediction. This approach allows attribution of systematic errors to the land model and observation operator itself. For example, Shiklomanov et al. (2021) enhanced the Ecosystem Demography v2 (ED2) model to predict full hyperspectral waveforms for airborne AVIRIS DA across eastern temperate US forests, later applied to tropical liana PFT development (Meunier et al., 2022). However, low-level products exhibit spatial and directional variability not fully resolved by LSMs, requiring compromises between model complexity, operator consistency, and accepting unresolved observation variability. For instance, directional effects of sun-sensor geometry in SIF and reflectance are not routinely modeled in global LSMs, which typically predict integrated fluxes. Strategies like space-time binning of SIF observations can reduce these effects, improving consistency. Practically, assimilation of raw data demands expertise in remote sensing and radiative transfer, underscoring the need for close collaboration between modelers and remote sensing scientists.

As observation operators become more complex, especially in the case of radiative transfer calculations, they also become more computationally expensive. This is a clear example of where ML may offer a unique opportunity within DA applications, as discussed in Section 4.3. To address the sensitivity of assimilation results with respect to the observation operators used, it is useful to develop community collections of observation operators. Examples are the Community Microwave Emission Model (CMEM, Drusch et al., 2009; de Rosnay et al., 2020; Holmes et al., 2008) and the Terrestrial Carbon Community Assimilation System (TCCAS), which provides, among others, several observation operators for SIF.

3.4. Tackling Spatial and Temporal Heterogeneity

The large variability in the surface properties of terrestrial ecosystems, arising from diverse climates, soil properties, and variations in plant and soil species composition, plasticity, and evolution, is an additional challenge in LSM parameter estimation. Calibration of the model at one location may not be applicable at another. Moreover, most LSMs are too computationally demanding to support calibration across large spatial domains. As such, it is important to develop strategies to ensure results offer a good compromise across different locations, as well as perform rigorous evaluation checks against data not used in the calibration.

A common approach to tackle this spatial heterogeneity is to perform "multi-site" optimizations, grouping sites and performing a single optimization over this group to obtain a more generic set of parameters. The multi-site approach has been shown to be very effective, at times out-performing site-specific optimizations (Knorr et al., 2010; Kuppel et al., 2012; Raoult et al., 2016). Another approach is to average the results of single-site optimizations. While usually less effective than multi-site optimizations, this is often a more practical solution and can still result in an improved parameter set. For example, Olivera-Guerra et al. (2024) found that the median values of optimized parameters improved simulated land-surface temperature performance.

Both approaches can be seen as end-members (all sites the same vs. all sites different) in a continuum of statistical independence across sites. While only recently applied to ecosystem model calibration (Dokoohaki et al., 2022; Fer, Shiklomanov, et al., 2021), hierarchical (or multilevel) models have long been used in ecology and other fields. A hierarchical model allows parameters to vary across groups (e.g., sites) but constrains that variation using higher-level statistical distributions. For example, site-specific parameters are drawn from an overarching distribution that describes variability across all sites. This structure captures the continuum from full pooling (same parameters everywhere) to no pooling (each site fully independent), enabling parameters to vary in space and time while sharing information—*borrowing strength*—across sites. By fitting across-site and within-site calibrations together, LSMs can reduce equifinality without forcing parameters to be identical everywhere. Hierarchical models also provide a formal way to account for greater uncertainty in out-of-sample predictions (where parameter vectors must be predicted) than in-sample ones. To date, most hierarchical calibrations assume simple random effects (site parameters drawn from one distribution), but there are opportunities to develop

models with explicit spatiotemporal covariances (nearby sites more similar) and site-level covariates that help explain and predict parameter differences.

A further alternative is the use of intermediate complexity models (e.g., DALEC; the DA Linked Ecosystem Carbon model), which, due to their reduced computational complexity, can retrieve parameters at the pixel scale utilizing spatially continuous information from Earth Observation (EO) data and thus derive unique information about the spatial variability of key underlying parameters, such as tissue residence times (Bloom et al., 2016) and the impact of fire (Exbrayat et al., 2018). The parameters and emergent ecosystem properties estimated from these models provide valuable insights into the spatial variability and magnitude of parameters. This can reduce the parameter space that needs to be searched when calibrating larger models. Furthermore, these optimized parameters may inform the calibration of more complex models and support exploration of their internal dynamics (Caen et al., 2022), although their direct transfer should be approached with caution due to potential structural differences between models.

Similarly, the interannual variability of atmospheric conditions means we also need to be careful which period is used for the assimilation. Ideally, we want to calibrate over multiple years to capture both the seasonal cycle and this interannual variability, while still retaining a number of years for evaluation (although using different sites for calibration and evaluation can help to relax this latter requirement). However, in practice, we are often limited by short time series (e.g., only a few years for some in situ experiments and recently launched satellite missions), data gaps, and the availability of meteorological forcing for corresponding periods, particularly for in situ data sets. Additionally, when transitioning from site-specific (tower-based meteorology) to regional or global simulations, the boundary conditions provided by meteorological forcing data (e.g., atmospheric reanalysis products) differ in both spatial and temporal resolution, which will change the effectiveness of the calibrated parameters. Therefore, parameters should ideally be calibrated using forcing data representative of the target scale, or with multi-site/multi-scale approaches that account for variability in input.

3.5. Dealing With Large and Multiple Observational Data Sets

Although EO instruments can provide global gridded data sets for model, fully exploiting these opportunities is challenging. Running experiments at the same resolution as the satellite products (e.g., 500 m MODIS resolution; Justice et al., 2002) is computationally intensive, and we do not always have access to matching meteorological forcing data. Furthermore, the resolution of EO products may also not align with the objectives of the experiment. When assimilating multiple EO data sets, conflicting spatial resolutions introduce further complexity, requiring choices about scaling—i.e., whether to upscale (aggregate) or downscale (interpolate) specific products (see Section 3.3). Generally, EO are scaled to match the chosen model grid, which is usually dictated by the resolution of the forcing data. Aggregating the satellite data to the model grid avoids spatial mismatches; however, this scaling can result in an over-generalization or loss of information. Moreover the propagation of observational uncertainty through the aggregation procedure is not straight forward, as spatial and temporal uncertainty correlations are typically not provided with the product. Biases resulting from spatial mismatches or errors resulting from data aggregation should be accounted for in the cost function, but in practice are rarely known.

Furthermore, the quality of EO data can differ hugely across different regions since they are impacted by atmospheric conditions (e.g., cloud cover) and topography, as well as the different data processing algorithms and calibration/validation strategies used to develop the different products. This can lead to regional and biome biases in the products that are very hard to circumvent due to measurement limitations, potentially generating structural model biases. Therefore, for many LSMs, it is common to select representative pixels for optimization (e.g., MacBean et al., 2015), although defining what is representative is a challenge in itself. Once selected, the representative pixel approach helps to (a) reduce the dimensionality of the problem, allowing for efficient and multi-data-stream calibrations, (b) focus on points with close to homogenous coverage to be able to calibrate class-specific parameters (e.g., plant functional types), and (c) define a different evaluation set of pixels with which to assess the optimizations, especially sites with additional ground data. After selecting representative pixels, multi-pixel optimizations are performed (as described in Section 3.4), focusing on estimating parameters for different ecosystem/edaphic conditions by spanning the various model plant functional types and soil textures all over the globe.

Another way to include more constraints to an optimization is by calibrating against multiple data streams. There is now an unprecedented wealth of in situ and EO data available, with even more satellite missions and in situ field measurement sites being planned (Balsamo et al., 2018; Ustin & Middleton, 2021). Different data streams offer information over different footprints and at different spatial and temporal resolutions offering unique opportunities to constrain different processes in the models. As LSMs become more complex through increased process representation and greater interconnectedness between the different terrestrial cycles (e.g., water, energy, carbon, nitrogen), multi-data stream optimizations are becoming paramount to provide adequate constraints since parameters are likely to impact different parts of the model. By optimizating only one specific data, we risk degrading the model's overall predictive capacity if some of the optimized parameters are loosely constrained (Bacour et al., 2015, 2023).

There are two possible approaches when assimilating multiple data streams. We can either calibrate against each data stream in turn, often referred to as "stepwise" assimilation, or include all data streams in one single optimization, known as "simultaneous" assimilation. Although mathematically equivalent when the posterior parameter uncertainties are properly estimated and propagated in the stepwise case (MacBean et al., 2016; Peylin et al., 2016), simultaneous assimilation is often preferable, since it ensures consistency (Kaminski et al., 2012) and avoids issues linked to accurately propagating the information gained about the parameter values from one step to the next. However, simultaneous optimizations may not always be practical, especially when running a computationally demanding LSM experiment, which is why the stepwise approach is often the pragmatic choice. In particular, there may be technical difficulties associated with the different number of observations for each data stream (which affects their influence in the cost function) and the characterization of error correlations between them (Bacour et al., 2023). Nevertheless, it must be emphasized that issues with unbalanced data streams are not solely due to imbalance but stem from the model's inability to accommodate both data sources when structural errors exist in either the model or the data (Oberpriller et al., 2021). In fact, properly quantifying and accounting for the uncertainty in the model structural error and data bias leads to better results than using ad-hoc methods such as reweighting different data streams (Cameron et al., 2022) (see Section 3.2).

3.6. Including the Pre-Observation Historical Period in the Assimilation to Better Constrain the Initial Carbon State

There is a large spread in model estimates of soil C content (Varney et al., 2024), which is due to poorly constrained, slow acting soil C cycling parameters (e.g., C allocation and turnover rates), poorly represented disturbance processes, and/or inaccurate historical climate and land use/cover change forcing data. As cumulative net carbon fluxes over long timescales depend on changes in soil C stock magnitudes (Le Noët al., 2023), biases in model predictions of the magnitude of soil C pools during the current observation period contribute to inaccurate current and future projections of carbon-climate feedbacks (Arora et al., 2020; Friedlingstein et al., 2023). Because many of the key parameters involved in predicting soil C pools act over long timescales, most LSM carbon cycle parameter DA studies at multi-site or global scale have included these slow-acting carbon cycle parameters in their optimizations (Peylin et al., 2016; Raoult et al., 2016; Schürmann et al., 2016). To do so would require including 100–1,000 s of years of the pre-observation historical period in the optimization (the length of which is dependent upon individual model structure, but likely includes both spin-up and transient runs). For most models this is computationally prohibitively expensive.

To make up for incorrect carbon pool magnitudes and the fact that including the pre-observation historical period in the assimilation is not yet feasible for computationally expensive LSMs, most past LSM carbon cycle parameter DA studies have included scalars on the initial C pools in the optimization, resulting in an improved fit to NEE and atmospheric CO_2 data (Carvalhais et al., 2008, 2010; Castro-Morales et al., 2019; Peylin et al., 2016; Schürmann et al., 2016). These scalars alter the initial carbon pool size at the start of the assimilation window. We also note that to date many of the carbon cycle assimilation studies have only performed a spin-up that brings the model to equilibrium immediately prior to the assimilation/observation window. They have not performed the full historical transient run with global change drivers i.e. now standard practice in global carbon cycle simulations (e.g., in the TRENDY protocol—Sitch et al., 2024). Thus, this scalar accounts for both the model structure, parameter, and forcing errors mentioned above that contribute to incorrect soil carbon pool magnitudes.

Other options for avoiding spin-up include directly initializing models with carbon stock observations, and including parameter calibration within iterative state DA approaches (e.g., A. M. Fox et al., 2018; A. Fox et al., 2022), as is common in hydrology model DA (e.g., W. Chen et al., 2015; H. Zhang et al., 2017). However, in all of these cases, calibrating the "right" model parameters to the "wrong" model pools is going to produce poor fits, complex sets of compensating errors, and potentially incorrect hypothesis testing around alternative model structures. In other words, adjusting initial carbon pools by including an initial C pool scalar or using joint state-parameter DA approaches to update the initial C state and the parameters, but without optimizing all the "slow" carbon cycle parameters to which the initial carbon pool magnitude is sensitive, is only useful if the purpose of the carbon cycle assimilation experiment is to update model estimates of *current* carbon budgets. If the desired goal is an accurate prediction of *future* carbon stock trajectories—for predicting carbon mitigation potentials or carbon-climate feedbacks under different scenarios of climate and disturbance trajectories—then simply adjusting initial carbon stocks via any method is insufficient. This is because in longer multi-decadal to centennial runs, any non-optimized C cycle parameters to which the initial carbon pools (and carbon pool trajectories) are sensitive will act to push the model back to a different equilibrium, resulting in an artificial trend in the modeled carbon pools (and resultant biases in carbon fluxes and land carbon sink estimates). Thus, for medium- to long-term projections of the land carbon sink and carbon-climate feedbacks, all parameters to which the initial carbon pools are sensitive need to be included in the optimizations. As the key C cycle parameters to which initial C pools are sensitive act over 100–1,000 s of years, this means that much longer time windows (depending on soil carbon model structure) will need to be included in the assimilation experiments (Raiho et al., 2021; Thum et al., 2017), despite the computational cost.

Methods for increasing the simulation speed (e.g., model emulation—see Section 4.2) will potentially solve the issue of prohibitive computational cost for these longer-term assimilation experiments. One opportunity for accelerating the spin-up is by adopting the matrix approach, where carbon balance equations are expressed as a single matrix equation without altering any processes of the original model, which has now been applied to multiple LSMs and used for both parameter sensitivity analyses and DA (Hararuk et al., 2014; Huang et al., 2018; Luo et al., 2022; Tao et al., 2020, 2024). Intermediate-complexity ecosystem models that optimize initial carbon conditions, rather than running a full spin-up and transient simulation, can help by providing more constrained priors for soil carbon pool turnover times and other parameters that strongly influence equilibrium or initial carbon stocks (Bloom et al., 2016).

In addition to longer assimilation time windows, assimilating time series of aboveground biomass or soil C stocks —in conjunction with carbon fluxes—is required to provide a useful additional constraint on carbon pools magnitude and trajectory (Thum et al., 2017). However, data on soil carbon stocks are relatively scarce compared to carbon fluxes, highly uncertain, lack timestamp metadata, and are often difficult to link to the conceptual carbon pools in many CENTURY-type models Parton et al. (1987), though this is changing (Abramoff et al., 2018). Additionally, these data sets often contain only one or a few time points. While assimilating some information on carbon stocks is better than not having any data, constraining long-term changes in C stocks will require multiple observations of both above- and belowground C stocks over time (Raiho et al., 2021) (or data representing rates of carbon cycling). Other data sets that may be useful in constraining long-term carbon stocks include nighttime and soil respiration data, or estimates of biomass derived from paleo fossil pollen records (e.g., Raiho et al., 2022) that so far have typically not been utilized in LSM DA studies. Just how long a time series we need to include to accurately estimate slow carbon cycle parameters will likely depend upon which parameters are important for estimating future carbon stock trajectories over the timescales of interest and the uncertainties associated with observations. More parameter sensitivity studies are needed to assess which slow carbon cycling parameters control carbon stock trajectories at different temporal scales (Huang et al., 2018; Raczka et al., 2018; Deepak et al., 2024). Ideally, these sensitivity studies should be performed with different scenarios of global change drivers, as changing inputs may alter the relative importance of slow carbon cycling parameters. The community can learn from the calibration and validation activities of soil biogeochemical models being approved for use in voluntary carbon markets (Mathers et al., 2023).

3.7. Choice and Implementation of Minimization Algorithms

To perform optimizations effectively, careful consideration must be given to the choice of algorithm and its implementation. As discussed in Section 2, various algorithms are available, each with distinct characteristics, such as local versus global optimization, each having different computational demands. Different DA algorithms make different assumptions on the parameters and this inevitably results in differing optimized parameter values. Additionally, every algorithm comes with a variety of configurable options. For instance, a Genetic Algorithm implementation by Scrucca (2013) offers a range of functions for parent selection (six options), crossover (five options), and mutation (three options), resulting in 90 possible combinations. Users can also adjust crossover and mutation probabilities. The success of the optimization process greatly depends on how the optimization is implemented, which may vary on a case-by-case basis. Systematically testing all possible combinations is unfeasible due to the large computational demand of an LSM. A more efficient approach is to use an emulator (see Section 4.1) rather than an LSM to find an optimal experimental design (Dagon et al., 2020); once the design has been identified, the optimization can be carried out using the LSM.

Furthermore, for gradient-based methods, implementing and maintaining the tangent linear or adjoint model is a huge challenge in LSM DA. For complex LSMs, which are historically coded in Fortran, developing the tangent linear and adjoint models can take years, even with automatic differentiation tools. This is because the model code must comply with the differentiation software without changing the model output. Moreover, to maximize the usefulness of the derivative code, it is important to expand the validity domain of the linear approximation by avoiding discontinuities or sharp transitions in the cost function (Kaminski & Mathieu, 2017; Kaminski et al., 2013). Examples include replacing look-up tables with their continuous formulations, reformulating minimum and maximum calculations to allow a smooth transition at boundaries (Schürmann et al., 2016), and developing smooth phenology schemes (Knorr et al., 2010). Once compliance with an automatic differentiation tool is achieved, it is recommended to develop that new modules with automatic differentiation in mind and to update the derivative code after each update of the underlying model and its observation operator. This approach is particularly crucial for models and development teams where variational DA is a high priority. This has been a challenge for complex community model like JULES and ORCHIDEE, where new processes are added approximately every six months. For JULES, the adjoint was developed for v2.2 (Raoult et al., 2016), while the model is at v7.3 at the time of writing. Similarly, the tangent linear exists for an older version of ORCHIDEE, (AR5), but predates major updates like the multi-layered soil hydrology scheme and nitrogen cycle. To address this, the ORCHIDEE DA team has been developing a preprocessing tool to allow automatic differentiation of any model version. On the other hand, owing to the high demand of up-to-date derivative code for variational DA applications, BETHY, with its comparatively lower complexity, has remained compatible with automatic differentiation for decades, providing efficient derivative code for the current model version. The same applies to its successor D&B (Knorr et al., 2025), used in the ESA-supported TCASS system, and to the Nanjing University Carbon Assimilation System (NUCAS; Zhu, Wu, et al., 2024). Alternatively, models written directly in an auto-differentiable language (Julia or python-JAX; see Section 5.4) can ease these challenges by making differentiable code easier to implement (Gelbrecht et al., 2023; Shen et al., 2023). However, developers must still ensure that the model's structure and parameterizations remain differentiable.

As discussed in Section 2, in the absence of the tangent linear or adjoint model, one can use finite differences. However, this necessitates the selection of an appropriate step size for accuracy and convergence speed, which will differ based on the sensitivities of the parameter estimated. Other methods to bypass the need for tangent linear and adjoint models include LAVENDAR's ensemble 4DVar approach (E. Pinnington et al., 2020) or the use of emulators, which can be used to either avoid gradient-based approaches in favor of MC ones, make numerical approximations of gradients viable, or both (e.g., Hamiltonian MCMC). However, these algorithms also come with a number of hyperparameters that need to be selected including the number of ensembles and convergence criteria.

4. Opportunities Through Machine Learning for Parameter Estimation

Despite the challenges and knowledge gaps discussed above, our community has never been in a better position to calibrate LSMs and rigorously diagnose their uncertainties. We now have access to large observational data sets at

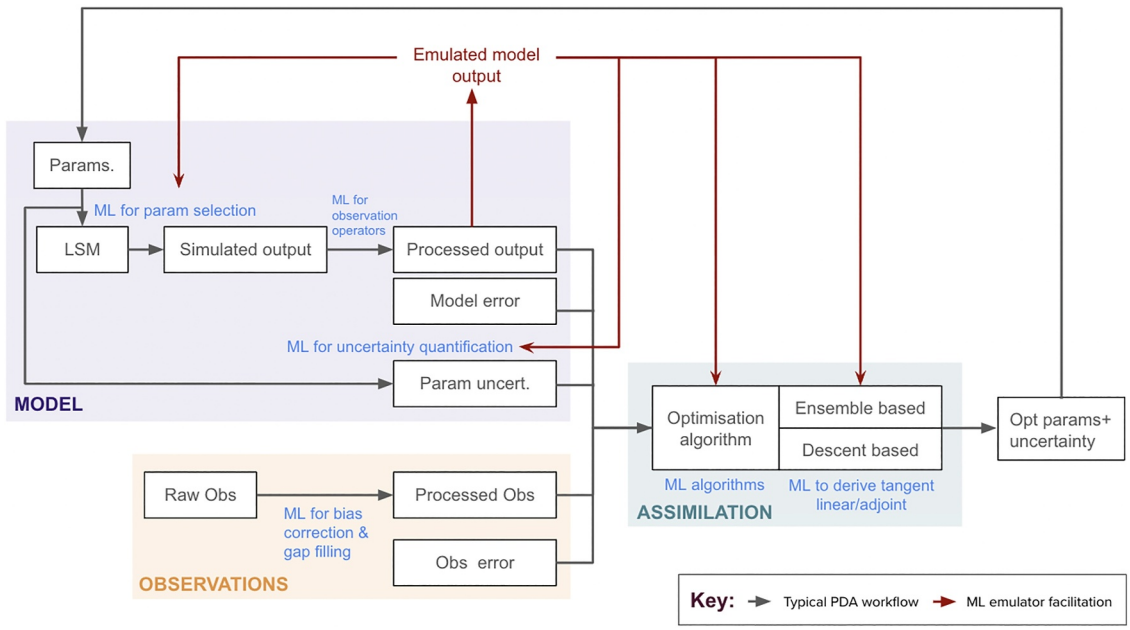

Figure 3. Examples of where machine learning (ML) can facilitate each part of the land surface model parameter data assimilation (PDA) workflow. The standard PDA workflow is represented by the gray arrows, where the processed model output and observations, along with their representative errors, are passed through to the assimilation system to derive optimal parameters and uncertainty. The ML facilitation is represented in two ways. First, by using emulators of the model output, represented by the red arrows. Second, more generally, the processes ML can assist with are written in blue—these are discussed more fully in the rest of this section.

high spatio-temporal resolutions and increased computational capacity and efficiency. These factors, combined with recent advances in ML, potentially allow us to make significant progress in model calibration.

The recent surge in ML has been evident in every aspect of society with the most relevant examples coming from numerical weather prediction (Lam et al., 2023) or remote sensing (Lary et al., 2016). These examples can help us identify ways in which ML can assist with land PDA. In this section, we specifically focus on how ML can be integrated into the land PDA process to help us address the current challenges and limitations outlined in previous sections (Figure 3). For example, ML can (a) reduce computational costs, making available optimization techniques that were previously prohibitively expensive, (b) overcome limitations posed by a lack of accurate process knowledge and inform the development of improved process knowledge, (c) expedite the optimization process, or (d) facilitate the use of multi-variate, sparse, or biased observations. With an increasing number of studies currently being published in the field of ML, we only provide a short overview of such applications and the most relevant literature for land PDA here. Given the diversity in both methodologies and applications of existing studies, evaluating the advantages and limitations of individual algorithms in a comprehensive manner remains a significant challenge. In this paper, we present a selection of sample applications of ML for parameter estimation, categorized into three distinct categories: (a) the use of ML to emulate the relationship between LSM parameters and its outputs or performance (Section 4.1), (b) the creation of "hybrid models" in which ML replaces or complements a component of a larger LSM by emulating the relationship between model inputs and observations for that component (Section 4.2), and (c) the use of ML to optimize the parameter estimation process itself (Section 4.3). This categorization aims to provide a rough guide for readers navigating the complexities of ML algorithm selection and implementation for PDA studies.

ML has the potential to significantly advance land PDA. The use of ML emulators can reduce the computational cost often associated with parameter calibration and ML techniques can improve the optimization process itself. Additionally, ML substitutes, trained on model inputs and observational outputs, can provide meaningful insights into model processes that are poorly captured or understood and thus contribute to improved parameterizations in process-based models. However, ML approaches are unlikely to fully replace process-based models due to several inherent limitations. These include susceptibility to issues such as equifinality (Kraft et al., 2022; Sawada, 2020), compensation for structural model errors, reduced accuracy outside of the training data range, and

a lack of physical constraints. We note that any ML approaches need substantial data to perform well and thus the ML components in the hybrid part need to be targeting processes for which data is plentiful. ML approaches often have a large number of parameters in their training which gives them a larger degree of flexibility that can compensate for errors in physical models, but can also lead to overfitting. Some of the above issues can be partially mitigated through the use of multi-variate independent observation constraints in the calibration. Given the limitations of both ML and process-based models, hybrid models that integrate the strengths of each are expected to play a central role in future LSM developments.

4.1. Parameter Perturbation Emulators

ML methods can improve the PDA process by mitigating the computational cost of high-complexity LSMs that has traditionally hindered the use of computationally demanding calibration techniques such as MCMC. By building an ML-based statistical relationship between input parameter settings and the LSM output or an aggregate of the LSM output (for instance over time or space), the LSM output can be estimated for a new set of input parameters. The statistical relationship serves as a computationally efficient surrogate model for the expensive LSM and is most frequently called an emulator (although this term is not exclusive to this application), while surrogate, meta-model, or reduced-order model are also used to refer to this tool. Indeed, emulators already have a rich history in climate sciences (Knutti et al., 2003; Sanderson et al., 2008; Watson-Parris, 2021). In the following, we provide a guide to key aspects for using emulators for PDA, including strategies for sampling the parameter space, the selection of the emulation method, the impact of computational cost reductions, and the emerging history matching (HM) technique. We also note that—as with all parameter estimation methods—PPEs can be subject to increased uncertainties when applied outside of the initial training data set.

4.1.1. Parameter Sampling Strategies

The training of an emulator requires an ensemble of LSM simulations with perturbations to the input parameters often called a perturbed parameter ensemble (PPE; e.g., D. Kennedy et al., 2024; R. E. Kennedy et al., 2024; McNeall et al., 2024). The design of the initial PPE depends on the intended use; for uncertainty quantification, it is often preferable to sparsely sample the entire parameter space using Latin hypercube sampling (McKay et al., 1979). However, for calibration applications discussed here, it can be more cost effective to use a non-random and targeted sampling strategy, such as active learning which tries to optimize the selection of the next sample (e.g., H. Zhao & Kowalski, 2022). Alternatively, an Ensemble Kalman Filter approach (Evensen, 2003) can be used to place the initial design points in regions of significant posterior mass to optimize the calibration process (e.g., Cleary et al., 2021). When building emulators for model calibration it can be particularly effective to treat this as an iterative design process, whereby an initial set of parameter vectors (e.g., Latin hypercube) is used to generate a rough idea of where in parameter space the optimum lies, then additional parameter vectors are sampled from this region, refining the emulator in a way conceptually similar to a nested grid in parameter space (Fer et al., 2018). How to optimally propose points in parameter space remains an important research question.

4.1.2. Emulation Methods

There are many ML methods appropriate for emulating the LSM response to parameter modifications. When it comes to the calibration problem specifically, an alternative to emulating the LSM output is to directly emulate the cost function (i.e., the response surface of model error as a function of parameter value) which is of much lower dimension and often much smoother that the model output itself (Y. Cheng et al., 2023, 2024; Dagon et al., 2020; Fer et al., 2018; Fer, Shiklomanov, et al., 2021).

Gaussian processes are commonly applied as they are well-suited to interpolate non-linear surfaces in data-scarce settings and moreover provide a measure of prediction uncertainty that can be used to quantify the emulator uncertainty. However, since the computational cost of Gaussian processes dramatically increases with the size of the data set, they are less feasible for larger data sets. A possible solution are sparse Gaussian processes (Baker et al., 2022), whereby a subset of coordinates representative of different parameter settings and forcing data regimes is selected under the assumption of no lateral information exchange between LSM grid cells. Another popular method for emulating LSMs are neural networks (NNs), as they are straightforward and fast to implement (Hatfield et al., 2021), with fast evaluation speeds and good predictive skill within the bounds of the training data.

However, NNs are sensitive to biases in the selection of the training data as well as the tuning of the algorithm hyperparameters, which means that they generally cannot extrapolate to scenarios beyond the training data or be transferred to new data sets without performance degradation (Shwartz-Ziv & Armon, 2022). Examples of NNs used to emulate LSMs include applications for carbon modeling parameter perturbation and uncertainty (D. Lu & Ricciuto, 2019), the emulation of CLM5 outputs given important biophysical parameter values (Dagon et al., 2020), and to emulate the ensemble mean of several urban LSMs (Meyer et al., 2022). While artificial NNs do not provide a probabilistic prediction, new methods are emerging such as neural processes (Garnelo et al., 2018) or randomized prior networks (Bhouri et al., 2023). Regression trees can also be extended to include probabilistic prediction such as with NGBoost (Duan et al., 2020) or XGBoost (Donnerer, 2024), as used, for example, to emulate ELM-FATES (L. Li et al., 2023). XGBoost has been shown to generally outperform NNs while requiring little parameter tuning and is able to achieve robust performance even when extrapolating to scenarios beyond the training data (Grinsztajn et al., 2022; Shwartz-Ziv & Armon, 2022). A disadvantage of tree-based methods is their slower evaluation speeds and the fact that they are not differentiable, which can limit their usability for certain applications (e.g., coupled DA, Hatfield et al., 2021). Long-Short Term Memory methods include memory mechanisms by leveraging long-term dependencies in the training data time series, allowing them to effectively emulate model processes across different time scales without performance loss at longer lead times, as is the case for XGBoost (e.g., Wesselkamp et al., 2024). This makes them particularly suited for the emulation of large-scale forecasting systems that encompass physical processes acting at different time scales (Datta & Faroughi, 2023; Guo et al., 2021; Wesselkamp et al., 2024).

4.1.3. History Matching

In the field of uncertainty quantification, one key method that can be used in combination with emulators and that is gaining traction for LSM parameter estimation is the so-called HM method (Hourdin et al., 2023). This method is not about finding the most likely parameter values, but rather ruling out implausible ones based on some given metrics (Williamson et al., 2013). An advantage of this more conservative approach is that it is less subject to equifinality, thus ensuring that the true parameters are contained in the posterior parameter distribution. Using emulators to facilitate computation, HM is commonly applied using successive iterations (also known as iterative refocusing) to reduce parameter space and retain the least implausible parameters. Like the cost function used in variational DA, the implausibility takes the observation and model structure errors into account. While these errors are still hard to determine (Peatier et al., 2023), HM gives a clear diagnostic if the errors are overestimated, by ruling out little to no parameter space. If the errors are underestimated, HM will rule everything out, suggesting the errors have been misspecified, whereas other optimization approaches would yield a solution every time. HM also allows the user to test many different metrics to see if parameters can capture specific features, similar to multi-objective optimizations, giving a clear diagnosis of model structure error. HM in combination with emulators has successfully been tested with some of the major high-complexity LSMs: CLM (Dagon et al., 2020), JULES (Baker et al., 2022; McNeall et al., 2024), and ORCHIDEE (Raoult, Beylat, et al., 2024), for example. These studies highlight how HM can be used to identify sensitive parameters, redefine ranges of variation and identify non-Gaussian relationships between parameters. This information could potentially be used to determine the prior error covariances (i.e., to set up the background error covariance matrix in variational DA) or provide ecological constraints to an optimization.

4.2. Hybrid Modeling

An alternative approach for enhancing land PDA through ML is the strategic substitution of physical model components with an ML approximation (Eyring et al., 2024), known as hybrid modeling. The appeal of the hybrid approach is that it can address known model inadequacies and computational bottlenecks in a targeted manner while retaining the use of physical process knowledge and constraints where they are reliable. For example, the hybrid approach can mitigate model structural errors, by replacing model processes that are missing or poorly understood with data-driven substitutes, assuming adequate data exists (Arsenault et al., 2018; Reichstein et al., 2019). At the same time, the hybrid approach can add physical constraints to the ML model components, thus maintaining physical consistency and interpretability (e.g., Beucler et al., 2021; Kraft et al., 2022; Reichstein et al., 2019). The hybrid approach offers a strategic combination of the advantages of ML and physically-based approaches and leverages their complementarity. Here we introduce three major applications of hybrid modeling

for land PDA: (a) substituting a specific model parameterization with an ML approximation, (b) deriving spatial parameterizations that better capture observed physical behavior, and (c) training on model-data residuals to predict process-model biases and characterize structural errors. Hybrid modeling has been implemented successfully in a number of LSM applications, including for streamflow (Yang et al., 2019), evapotranspiration (W. L. Zhao et al., 2019), subsurface flow (N. Wang et al., 2020), rainfall-runoff modeling (Xie et al., 2021), as well as more generally for the prediction of sea surface temperatures (de Bézenac et al., 2019), atmospheric convection (Gentine et al., 2018), and high impact weather events (McGovern et al., 2017).

4.2.1. Substitution of Uncertain or Missing Parameterizations and Processes

A common application of hybrid modeling is to improve the representation of complex processes that are poorly captured or missing completely in traditional LSMs. An example of this are human processes, which can be effectively accounted for with ML approaches trained in an aggregate manner (e.g., one NN trained on all locations) and using a combination of observations and process-model outputs to map observations into the model climatology (thus removing global biases). This approach retains the independent information on human processes i.e. inherent in the observations but typically removed in traditional bias correction approaches (e.g., Kumar et al., 2012) and has been used for example, to introduce the impact of irrigation and tile drainage in a physical LSM (Kolassa et al., 2017) as well as to predict water flow in urban areas from runoff estimates generated by a physical LSM (Assem et al., 2017). Hybrid modeling can also be used in cases when the naturally occurring physical processes are poorly understood. For example, Arsenault et al. (2018) used an ANN with a combination of remote sensing observations and model predicted states to generate improved estimates of snow depth within the Land Information System.

4.2.2. Observation Operators

One specific application of an ML-based process substitution is the use of ML-generated observation operators to translate satellite-observed radiances into model states or parameters (see challenges raised in Section 3.3). The use of ML techniques in this context has several advantages: (a) ML-based observation operators are relatively simple to implement compared to physically-based approaches, which often involve the inversion of radiative transfer models, (b) they are able to easily accommodate the simultaneous assimilation of multiple observation types, (c) they can inherently correct climatological biases between model and observations (see also section "Substitution of Uncertain or Missing Parameterizations and Processes"), and (d) they facilitate the assimilation of radiance observations rather than retrieval products, thus reducing errors stemming from possible inconsistencies between retrieval algorithm assumptions and models. ML-based observation operators can provide uncertainty estimates for their outputs, as they typically retain some observation data to assess performance after training. Due to these advantages, ML-based observation operators have been applied in several land DA studies, including for soil moisture (Kolassa et al., 2017; Rodríguez-Fernández et al., 2019), LAI (Durbha et al., 2007), snow water equivalent (Kwon et al., 2019), and as a combined forward model for soil moisture and LAI (Shan et al., 2022).

4.2.3. Improved Spatial Parameterizations

Hybrid modeling techniques have also been used successfully to generate model parameterizations that better capture the parameter spatial distribution and thus the observed physical behavior (Tao et al., 2020, 2024). Process-model parameterizations can be limited by observation sparsity, which can lead to ad hoc decisions when assigning parameter values globally. Similarly, many global LSMs significantly simplify biogeochemical and physical mechanisms into empirical parametric functions. Hybrid modeling can address these issues by mapping environmental variables into model parameters or using high-resolution, high-fidelity model simulations to derive new parameterizations for coarse-resolution models (e.g., Gentine et al., 2018). Bao et al. (2023) replaced the traditional PFT-based parameterization of a light use efficiency model with an ecosystem-property-based parameterization derived from a multi-layer perceptron NN to better capture the spatial variability of GPP within PFTs. Several studies have used a hybrid ML approach to improve the representation of evapotranspiration in LSMs, either by directly estimating evapotranspiration (W. L. Zhao et al., 2019) from observations or by inferring related prognostic variables, such as the stomatal and aerodynamic resistances (ElGhawi et al., 2023), or transpiration stress (Koppa et al., 2022). In each case, the hybrid model was able to learn unknown latent processes and thus outperform traditional physics-based schemes.

4.2.4. Model Error Identification/Characterization

Additionally, hybrid modeling implementations can serve as effective diagnostic tools to identify model errors. For an independently evaluated ML approximation, systematic differences between predictions from a physical model component and its ML counterpart can provide insights into missing or flawed model process representations as well as identify inadequate model parameters (e.g., McGovern et al., 2017), especially when the ML model is not only trained to represent the model outputs but uses other observational constraints in the learning phase. For example, Finn et al. (2023) and Gregory et al. (2023) used an ML trained on model-data residuals to predict model biases and characterize structural errors, while Gregory et al. (2024) extended this approach to implement an online bias correction within a DA framework. Similarly, Farchi et al. (2021, 2023) integrated a deep-learning step into a DA framework to create a hybrid model that dynamically learns and corrects model errors at each DA time step.

4.3. Optimization Process

Finally, ML can be integrated in land PDA to improve the optimization process itself. Since optimization is a key component to both ML and DA, there are many algorithms common to both fields including gradient-based and evolutionary algorithms (Section 2). Indeed, the strong mathematical similarities between ML and DA mean that both fields can learn from each other and share methodologies (Geer, 2021).

4.3.1. Optimisation Algorithm

ML approaches can be used to improve optimization algorithms themselves by helping speed up the search process and improve the quality of solutions (Song et al., 2019). Furthermore, ML can be used to automatically choose the setting of adjustable parameters found in some optimization algorithms. For example, clustering methods can be used to set the population size, crossover probability and mutation probability parameters in genetic algorithms (J. Zhang et al., 2007) and maintain population diversity. Tree-based random forest models have been used to dynamically construct, search, and prune the parameter space to efficiently optimize ML structure and hyperparameters (Akiba et al., 2019). ML techniques can also be used to choose the best-performing algorithm for a particular optimization problem (Kerschke et al., 2019). While the emerging ML methods are promising, they are very novel and—to the best of our knowledge—have not yet been applied to optimizing the parameter estimation algorithm hyperparameters themselves.

4.3.2. Differentiability

Finally, a novel and emerging use of ML is the use of large language models (e.g., ChatGPT). Modern open-source coding languages like Julia and Python through the Google JAX library (Bradbury et al., 2018) can be automatically differentiated to generate the tangent linear model (see Section 2). Many high-complexity LSMs are written in Fortran code; large language models can help translate Fortran code to more modern languages (Zhou et al., 2024), facilitating the derivative of such models. Alternatively, we can use NNs to emulate the tangent linear and adjoint models since NNs can be differentiated trivially (Hatfield et al., 2021).

4.3.3. Computational Cost Reduction

ML techniques significantly reduce computational costs of complex LSMs and DA systems, enabling previously prohibitively expensive parameter calibration approaches, such as MCMC methods that require numerous parameter distribution samples. For instance, Fer et al. (2018) showed ML emulators accelerated MCMC by over 100× for SIPNET and over 200,00× for the more complex ED2 model. Other studies (e.g., Sawada, Cleary et al., 2021) used emulators for Bayesian inversion, making costly MCMC sampling of the approximate posterior parameter distribution feasible. Torres-Rojas et al. (2022) used surrogate modeling with Pareto analysis to infer optimal LSM parameters at just 1% of the original computational cost. Coining the method "Calibrate, emulate, sample," Cleary et al. (2021) showed how the method could be successfully applied to models of different complexity, while other groups have also demonstrated the suitability of ensemble approaches for parameter selection (e.g., Couvreux et al., 2021).

Hybrid modeling can also reduce cost when used to replace computationally intensive model components. For instance, Sun et al. (2023) used bagging decision trees (an ensemble ML method based on Breiman (1996)) to

Table 3
Summary of Challenges Outlined in Section 3 and Their Machine Learning (ML) Opportunities

PDA challenge	ML opportunity
Selecting parameters and their prior distributions (Section 3.1)	
– Identifying which model parameters to optimize is challenging, due to high dimensionality and strong parameter covariances	– Emulators can reduce the computational demand of running models with many different parameter settings needed for sensitivity analyses (Section 4.1)
– Choosing prior distributions for parameters is crucial yet difficult, requiring detailed structural insights and data	– Emulators can be used to aid uncertainty quantification, for example, by facilitating history matching (Section 4.1)
Characterization of model and data/observation errors (Section 3.2)	
– Model errors are difficult to quantify due to uncertainties in process representation, missing processes, and the challenge of specifying an informative prior	– Hybrid modeling can be used to replace model processes that are missing or poorly understood, helping to diagnose model structural errors (Section 4.2)
– Quantifying data errors is tricky because of sampling variability, instrument inaccuracies, and complex error correlations that are often ignored	– ML methods can be used to generate improved estimates of derived quantities, thus reducing observation errors (Section 4.3)
Developing observation operators (Section 3.3)	
– Matching model outputs to observations require transformations that can introduce biases	– ML-generated observation operators can be used to directly translate satellite-observed radiances into model states or parameters (Section 4.3)
Tackling spatial and temporal heterogeneity (Section 3.4)	
– Variability in surface properties, driven by diverse climates, soils, and ecosystems, complicates parameter estimation across locations	– Hybrid modeling can be used to improve spatial parameterizations (Section 4.2)
– High computational demands make it difficult to calibrate LSMs across large spatial domains	– Emulators can help reduce the computational demand of running the model over large domains (Section 4.1)
– Temporal variability and short data series hinder the capture of both seasonal cycles and long-term trends	– Long Short-Term Memory encoder-decoder networks consider long-term dependencies and therefore may help capture seasonal and interannual trends (Section 4.1)
Dealing with large and multiple observational data sets (Section 3.5)	
– Products may be subject to regional biases due to varying data quality and processing methods	– ML-based observation operators are able to easily accommodate multiple observation types and remove global biases with respect to the model (Section 4.2)
– Assimilating multiple data streams in model calibrations presents challenges in consistency, error characterization, and balancing different data sources	
Including the historical period in the assimilation window (Section 3.6)	
– Spin-up and transient parts of model runs can be computationally demanding	– Hybrid modeling can be used to replace computationally costly parts of the model (Section 4.2)
Choice and implementation of minimization algorithms (Section 3.7)	
– Algorithms requiring a large number of model runs are computationally costly and therefore rarely applied to complex LSMs	– ML can enhance computational efficiency, enabling the use of algorithms that require numerous model runs (Section 4.1)
– For different algorithms, there can be a large number of configuration options and tuneable hyperparameters	– ML can help find the best configurations and hyperparameters to use when optimizing (Section 4.4)
– Maintaining tangent linear/adjoint models for gradient-based optimization in complex LSMs is challenging	– Large language models can be used to translate LSMs to modern coding languages that are easier to differentiate and can better exploit GPU. Alternatively, we can emulate the LSM using NNs, which are easily differentiable (Section 4.4)

emulate the spin-up phase of the ORCHIDEE model, which can normally account for up to 98% of total computation time. Similarly, Koppa et al. (2022) integrated deep learning with a physical LSM and satellite observations to estimate global evaporation, improving both predictive accuracy and computational efficiency.

5. Future Priorities in Land Model Parameter Estimation

Moving beyond the ML avenues outlined in the previous section and summarized in Table 3, here, we discuss the opportunities and future priorities where land PDA promises to have some large impacts, building on recent

successes. We argue that more funding for technical DA studies and software engineering support would significantly aid this work.

5.1. Testing Novel Data Sets and Experimental Configurations

In addition to the traditional data sets used to optimize LSM parameters (e.g., leaf area index, VIs, SIF, in situ carbon and water fluxes, atmospheric CO_2 mole fraction data, or soil moisture—see Section 1 and Figure 1), our data-rich world offers access to a wide array of novel and underutilized data streams enabling new and exciting constraints on multiple different processes in LSMs (as have been used for parameter DA in smaller scale ecosystem and ecology models). These include (to name a few):

- *Manipulation experiments*: For example, elevated CO_2 experiments can be used to constrain the fertilization effect at nitrogen-limited sites (Jiang et al., 2020; Mahmud et al., 2018; Raoult, Edouard-Rambaut, et al., 2024; Thomas et al., 2017).
- *Data about soil carbon stocks*: Data from the International Soil Carbon Network (Harden et al., 2018; Nave et al., 2016) and the global soil respiration database (Jian et al., 2021) can provide valuable insights. Similarly, soil radiocarbon measurements (C. R. Lawrence et al., 2020) can help constrain rates of soil carbon cycling (Z. Shi et al., 2020) and carbon isotope concentrations can be used to improve simulated soil organic matter decomposition (Mäkelä et al., 2022).
- *Tree ring data*: Annual biomass increments derived from tree ring widths can help infer carbon accumulation (Babst et al., 2014; Jeong et al., 2021). Similarly, tree ring isotopic data (carbon and oxygen) can act as constraints for leaf physiology and growth (Barichivich et al., 2021).
- *Other aboveground biomass products*: Products from the ESA BIOMASS mission (Quegan et al., 2019) help constrain carbon allocation and woody biomass turnover parameters (Smallman et al., 2021). Similarly, land-use and land-cover products (e.g., MapBiomas Collection 3.1, based on Landsat) can be used to create regrowth curves (Heinrich et al., 2021, 2023), which together with forest inventory data, can help constrain disturbance processes.
- *Additional remote sensing data sets*: New data sets, such as full-waveform lidar data from the Global Ecosystem Dynamics Investigation (GEDI) mission (Dubayah et al., 2020), can help constrain canopy structural parameters, including canopy height (Potapov et al., 2021). Similarly, improved observations of land surface temperature and total surface/groundwater content from GRACE instruments also can offer additional constraints on the energy and water cycles.
- *Trace gas flux measurements*: Carbonyl sulfide measurements (Whelan et al., 2018) can be used to constrain GPP and stomatal conductance (Abadie et al., 2023; Zhu, Xing, et al., 2024). There is also a growing number of nitrous oxide flux measurements (Nicolini et al., 2013), which can be used to calibrate LSMs that include nitrogen cycles. Methane flux measurements, such as those over peatlands (Salmon et al., 2022), can also be utilized to improve the representation of methane production processes.

As with all past carbon cycle DA studies, before novel data sets can be reliably used in a DA experiment, it will take time to test the best approaches for how to best use these data streams within a DA experimental framework. It should be standard practice to run synthetic DA experiments to test which types of observations provide the best constraint and which observational characteristics (temporal sampling interval, record length, observation uncertainty, choice of minimization algorithm and its configuration, etc.—Section 3.7) are required to retrieve the correct parameter values and/or reduction in uncertainty in associated variables, with the strong assumption that there is no modeling bias. Synthetic experiments, also known as "twin" experiments, use "pseudo data" that have been output from the model and modified according to known observational characteristics (see REFLEX and Optic experiments; Trudinger et al., 2007; A. Fox et al., 2009). As these data are model outputs, the "true" value of the parameters is known. Synthetic DA experiments can be conducted before data collection or instrument design to optimize sampling across space, time, and methodology.

5.2. Testing DA Experiment Configurations

In the introduction we outlined a number of studies in the early development of carbon cycle DA systems that aimed to better understand how different DA experimental set-up at site scale impacted posterior parameter values and their uncertainties. In recent years, such "technical" DA experiment configuration

studies have become less common; however, many questions about the most appropriate methods and approaches for parameter calibration using DA and ML remain. Existing studies have shown optimized estimates of various parameters are very dependent on the experimental setup and so few of the optimized parameter values are actually used in the operational version of each LSM—although this is something to strive for in future efforts. Indeed, even when calibrated parameters have been shown to improve model performance, getting them to be the new defaults in coupled models is non-trivial (Kyker-Snowman et al., 2022). In addition to the synthetic experiments that should be conducted when novel data sets are used (Section 5.1), additional tests of DA experimental configuration that are rarely performed (or rarely reported in the literature) but which are sorely needed include testing: (a) how parameters estimated at individual sites compare to parameters estimated when including multiple sites in the assimilation (Kuppel et al., 2012; Raoult et al., 2016) or using hierarchical approaches (Fer, Shiklomanov, et al., 2021; Tian et al., 2020) (see Section 3.4); (b) the utility of PFT dependent parameters versus alternative approaches for grouping parameters (e.g., regionally dependent PFTs—e.g., Bao et al., 2023; Dahlin et al., 2017); (c) how estimated parameters vary with the forcing data set used in the simulations; (d) how estimated values depend on which parameters and/or PFTs are optimized or which terms and observations to include in the cost function, (e) how estimated parameters vary in space and time within PFTs and what this tells us about missing processes, among other factors; (f) what the impact of record length is on estimated parameter values and uncertainties; (g) validation of estimated parameters over longer timescales than used in the assimilation; and (h) the impact on the posterior parameters of spatial mismatches between the data and model versus aggregating data to the resolution of the model. A critical test of any parameterization process is that the newly trained model must have improved predictive skill for independent data. For example, Famiglietti et al. (2021) demonstrated that different data combinations impact the resultant predictive skill and that the amount of data used in model calibration must be commensurate with the complexity of the model. Such technical tests are required each time a new process is optimized or a novel data set is used in the assimilation. Building DA frameworks to include this technical testing will give confidence in using estimated parameter values in operational versions of the models.

As parameter estimation methods and systems become more developed, we can run more experiments to quantify and reduce uncertainty due to poorly constrained parameters using different driving data sets and versions of the model that account for different representations of processes. In the wider climate science literature, there exist promising approaches to provide objective assessments of structural and parametric components of model error (Peatier et al., 2023). Moreover, the proposed move to more modular LSMs (Fisher & Koven, 2020) will also allow for different processes in the model to be isolated and calibrated sequentially, reducing the scale of parameter subspaces to be calibrated and enabling better testing of alternative hypotheses (e.g., different stomatal optimization theories) and facilitate collaboration across different modeling groups.

5.3. Identifying and Improving Structural Errors and Model Representation

By combining novel data and implementing testing DA approaches, as described in Sections 5.1 and 5.2, we can aspire not only to quantify and constrain parametric uncertainty and associated model predictive ability, but also to assess how this information influences our understanding of model processes. If we cannot match observations within the bounds of their known uncertainties by simply changing the parameter values or if "edge-hitting" posterior parameters are identified, this suggests that a process is poorly represented or missing from the model. This critical information is fed back to the model developers. The focus should then be on refining core processes driving ecosystem-scale carbon and water fluxes, understanding their underlying mechanisms, and testing their responses to global change, beyond just fitting historical data. We can also learn from the use of DA methods that produce time-varying parameters to determine if processes may be missing (e.g., Nie et al., 2011; Verbeeck et al., 2011; Xiong et al., 2019).

However, careful consideration is needed to avoid equating increased model complexity with progress. While addressing structural errors is vital, adding complexity does not always improve predictive power and may hinder validation and management. Balancing accuracy with parsimony is key. Model modifications should further be guided by the quality and availability of observational data to ensure new processes or parameters are well-

constrained. Without sufficient observational support, added complexity risks overparameterization and reduced predictive value.

5.4. Moving Toward Land Surface–Atmospheric Transport and Full Earth System Model Coupling With Uncertainty Quantification

Many studies have successfully constrained parameter uncertainty in LSMs, leading to reduced uncertainty in contemporary land-atmosphere carbon fluxes. However, this reduction in uncertainty has not been fully propagated to constrain future projections. There is a clear opportunity to take this extra step, enabling observationally constrained probabilistic statements about the future of the land biosphere.

Atmospheric CO_2 mole fraction measurements from tall towers worldwide have proven valuable for improving NEE predictions at regional to global scales within carbon cycle DA frameworks (Bacour et al., 2023; Castro-Morales et al., 2019; Kaminski et al., 2002, 2012, 2013; Knorr & Heimann, 1995; Koffi et al., 2012; Peylin et al., 2016; Rayner et al., 2005; Scholze et al., 2007, 2016; Schürmann et al., 2016). While atmospheric CO_2 data provide a direct constraint on net surface CO_2 exchange, reliable representation of terrestrial carbon sources and sinks ideally requires accurate simulations of the gross carbon fluxes. However, while global scale estimates of GPP are available for model evaluation or assimilation purposes (e.g., Joiner et al., 2018; Nelson et al., 2024) the currently available RECO products are still subject to large uncertainties. For instance, empirically upscaled RECO from eddy covariance measurements provided by FLUXCOM are inconsistent with inversion-based products in the tropics, possibly due to low sampling density in the region (Jung et al., 2020). In situ data are sparse and site history does not reflect larger-scale disturbance adequately. A key benefit of assimilating atmospheric CO_2 concentration data is that it provides one of the few large-scale (albeit indirect) constraints on RECO, which is strongly influenced by soil carbon stocks. Assimilating CO_2 data can therefore help improve how LSMs represent both soil carbon fluxes and stock trajectories, which is crucial for predicting the future carbon sink capacity of terrestrial ecosystems.

However, assimilating atmospheric CO_2 data requires coupling LSMs with atmospheric transport models in order to map the simulated land surface fluxes to atmospheric CO_2 concentrations at specified vertical levels (for station data) or integrated over the atmospheric column (for space-borne data). The observational constraints of atmospheric CO_2 data on LSM parameters is also more "diffuse" than when assimilating surface observations. This is due to the inclusion of additional modeling errors associated with the atmospheric model itself (physics and spatial/vertical discretization) and with the other CO_2 fluxes required as inputs (mainly ocean fluxes, fossil fuel emissions, and biomass burning). The coupling also presents technical and computational challenges. Compared to LSMs, the derivation of the tangent linear and adjoint models of atmospheric transport models is more straightforward (Kaminski et al., 1999; Meirink et al., 2008; Rödenbeck et al., 2003), but their implementation increases the computational load. One approach to overcome this issue is to use pre-calculated transport fields of the sensitivity of mean atmospheric concentrations at selected stations to the surface net CO_2 flux (see Peylin et al. (2016); or Bacour et al. (2023) for further details). However, this method has limited spatial and temporal coverage due to the finite time period of the precalculated sensitivities (estimating these sensitivities is also technically and computationally expensive). Assimilation of space-borne retrievals of XCO_2 (column-averaged carbon dioxide) with global coverage and pre-computed transport in SDBM and BETHY was demonstrated by Kaminski et al. (2010, 2017). Recent advances in the use of graphics accelerators (Chevallier et al., 2023) offer hope for significantly reducing computational times and enabling full coupling between LSMs and atmospheric transport models in the near future.

While coupling to an atmospheric transport model at least permits the use of atmospheric CO_2 data in parameter DA experiments, the ultimate goal for LSM parameter calibration is within a fully coupled ESM. This would allow representation of carbon-climate and land-atmosphere feedbacks within the optimizations. However, the difficulty of spinning up slow carbon pool processes and ocean circulation in ESMs remains a challenge for probabilistic coupled projections with ESMs (without flux corrections, Irvine et al., 2013). Additionally, to date, there has been limited assessment of whether posterior parameter values from offline DA experiments compare to estimated values from fully coupled runs (nor how estimated values vary when different offline climate reanalysis forcing products are used). As fully coupled simulations introduce systematic errors in atmospheric transport and land-atmosphere coupling, the likelihood of estimating similar parameter values is low, with significant potential

for compensating errors. To achieve this goal, LSM DA groups should learn from advances made in the numerical weather forecasting community (de Rosnay et al., 2022). As discussed at length in this review, while computational cost has so far been a prohibiting factor in achieving full ESM coupling, new ML techniques for model emulation (Section 4.1) (Watson-Parris et al., 2021) and automatic differentiation of model code (Gelbrecht et al., 2023; Griewank, 1989) should help considerably in alleviating this problem (see Section 3.7 for remaining challenges).

With increased computational power, we are acquiring the capability to run LSMs as ensembles rather than relying on a single realization, thereby enabling us to better capture the uncertainty of model predictions (Arora et al., 2023). Such efforts are already commonplace in ensembles of reduced complexity models (Sanderson, 2020; Smith et al., 2024), where large ensembles of future projections are computationally efficient due to the simplified model structures and reduced computational demands.

By sampling from the posterior distributions after a PDA experiment, we can generate ensemble simulations which can be used to explore future scenarios and idealized experiments (e.g., 1%/yr CO_2 concentration increase) and quantify constrained distributions of carbon-climate-CO_2 feedbacks. For example, by weighting the probability of each of the ensemble members, we can create probability density functions of future land carbon storage for different locations, thus narrowing the associated uncertainty of the future land sink and subsequently leading to more accurate calculations of carbon budget estimates. Although this can easily be done for simpler models where MCMC can be applied, for computational demanding models, there are two critical yet distinct questions in this area that need addressing. The first is how to generate joint posterior distributions for large models. This is demonstrated in Section 2 for ORCHIDEE and BETHY and may alternatively be addressed through emulators (Section 4.1). The second is how to intelligently select parameter vectors from those distributions. Currently, simple models might propagate uncertainty by using 100–1,000 ensemble members, but protocols like that used in the Coupled Model Intercomparison Project (CMIP) are not yet adopting such large ensembles, again due to computational expense and constraints on data storage. As a climate community, we should be striving to move toward using data-constrained ensemble simulations in CMIP or the TRENDY model intercomparison project (Sitch et al., 2024) to quantify uncertainties in model predictions reported in the Intergovernmental Panel on Climate Change eighth Assessment Report, the annual Global Carbon Budget and other emerging frameworks quantifying land carbon trajectories. Therefore, we must develop methods to maximize the propagation and partitioning of uncertainty with a limited number of ensemble runs. Constraining parameter uncertainty via improved DA and ML techniques should also help to reduce or exploit inter-model spread in CMIP and TRENDY, as model differences are likely partly attributable to variations in parameter values between models.

Other international frameworks that oversee policies and socioeconomic management of terrestrial carbon stocks —such as the voluntary carbon market and national emissions reporting for Nationally Determined Contributions under the Paris Agreement—already require estimates of model uncertainty; however, so far the models used in voluntary carbon market offset project verification tend to be of simple to intermediate complexity, and not full complexity LSMs. Better estimating uncertainty in LSMs via methods such as parameter DA should therefore facilitate their use in a wider range of policy and carbon management initiatives.

5.5. International Collaboration: Intercomparison Studies and Shared Toolboxes

Efforts by the Analysis, Integration and Modeling of the Earth System (AIMES) Land DA Working Group and the ILMF to build a Land DA Community have significantly advanced knowledge sharing through online workshops and town halls, highlighting the importance of continued collaboration. The goal is to facilitate cross-group interaction for DA methods training, knowledge exchange on technical DA developments and calibrated model intercomparison projects. The learning curve associated with learning DA for land surface modeling is steep. This is exacerbated by the lack of community-wide educational materials (although some resources exist, see https://land-da-community.github.io/training/ for a selective list, last accessed 27 August 2024). Understanding of DA methods is also hampered by the fact that technical studies testing different DA configurations are generally buried in supplementary materials or not published at all. Parameter DA system intercomparison studies would help to determine how much parameter uncertainty is contributing to the spread in model projections. This would signal to the wider LSM community that

parameter uncertainty quantification and reduction are needed to improve future projections of carbon-climate feedbacks and land-atmosphere interactions. One desirable outcome may be to create and share statistical toolboxes utilizing community cyberinfrastructure, for example, following the pioneering example of PEcAn (Predictive Ecosystem Analyzer; Fer, Gardella, et al., 2021; LeBauer et al., 2013), which offers a complete end-to-end informatic structure, as well as open-source LSM benchmarking tools (iLAMB: Collier et al., 2018; Seiler et al., 2022). Another key asset is the TCCAS, built around D&B (Knorr et al., 2025) and funded by ESA. It provides observation operators for diverse EO data streams, along with tangent and adjoint code for the core model and operators. Its modular design enables easy transfer of observation operators to other community models, enhancing interoperability and knowledge sharing. While LSMs with established DA systems may not switch to a community toolbox, such shared toolboxes will facilitate knowledge sharing, intercomparison studies and training of early career researchers. Simultaneously, if LSMs with established DA systems made more of their tools available within established community toolboxes, it would help reduce redundant research efforts and make the adoption of such tools easier. This is one of the big lessons we can learn as a community from the recent boom in ML. In addition to the improved hardware (e.g., GPUs), new algorithms and huge data sets, one of the reasons ML has been so successful is because the research has been done with a collaborative spirit and developed using open-source frameworks (e.g., TensorFlow, PyTorch, JAX).

6. Summary and Conclusion

Improving the accuracy of LSMs is of vital importance since land surface feedbacks on climate change represent one of the largest sources of uncertainty in climate change projections. Parameter DA is critical for enhancing the performance and reliability of these LSMs. This process involves determining the best estimates of model parameters, and their uncertainties, that best align the model outputs with observed data. Effective parameter estimation helps in capturing the complex dynamics of land-atmosphere interactions and improves the model's ability to simulate real-world phenomena. However, LSMs used to predict future climate scenarios (e.g., when coupled to ESMs) are complex in nature leading to many challenges when performing global scale optimizations. Nevertheless, advances in computational capability, novel data sets and emerging technologies offer promising avenues for improving parameter accuracy and model calibration.

Machine learning (ML) clearly has a pivotal part to play in the future of LSM DA, helping to streamline the assimilation process, manage large data sets and speed up otherwise computationally demanding processes. International collaboration is crucial in this endeavor, as shared knowledge and resources can significantly accelerate the advancement of LSM calibration and DA. Efforts to build a Land DA Community, such as those by the AIMES Land DA Working Group and the International Land Model Forum, have already made substantial progress in facilitating cross-group interactions. These collaborative platforms are essential for training, knowledge exchange, and the development of standardized methodologies, ultimately leading to more accurate LSMs.

Appendix A

The process-based models mentioned throughout the paper are listed in Table A1. This list covers a wide spectrum of land models ranging in complexity and computational demand, including land surface model (LSMs) that simulate interactions between carbon, water, and energy cycles, often incorporating other biogeochemical cycles (e.g., nitrogen cycling) and dynamic vegetation processes; stand-alone DGVMs that have more complex representation of vegetation demography (so-called vegetation demographic models, VDMs) but may not fully represent energy and hydrology components; and ecosystem models that primarily represent carbon cycling and

Table A1
References for the Process-Based Models Mentioned in This Article

Acronym	Full name	Model reference
BETHY	Biosphere Energy Transfer Hydrology	Knorr (2000)
BEPS	Biosphere Exchange Process Simulator	B. Chen et al. (2007), J. M. Chen et al. (2012)
CABLE	Community Atmosphere Biosphere Land Exchange	Kowalczyk et al. (2006)
CARDAMOM	CARbon DAta MOdel fraMework	Bloom et al. (2016), Smallman et al. (2021)
CLASSIC	Canadian Land Surface Scheme Including Biogeochemical Cycles	Melton et al. (2020)
CLM	Community Land Model	D. M. Lawrence et al. (2019)
D&B	DALEC and BETHY	Knorr et al. (2025)
DALEC	Data Assimilation Linked Ecosystem Carbon	Williams et al. (2005)
ED	Ecosystem Demography	Ma et al. (2022), Moorcroft et al. (2001)
ECLand	European Center for Medium-range Weather Forecasts Land model (based on CHTESSEL: Carbon-Hydrology Tiled Scheme for Surface Exchanges over Land)	Boussetta et al. (2021)
FATES	Functionally Assembled Terrestrial Ecosystem Simulator	Fisher et al. (2015), Koven et al. (2020)
FöBAAR	Forest Biomass, Assimilation, Allocation and Respiration	Keenan et al. (2012)
JULES	Joint UK Land Environment Simulator	Best et al. (2011), Clark et al. (2011)
JSBACH	Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg	Mauritsen et al. (2019), Reick et al. (2021)
LPJ-GUESS	Lund-Potsdam-Jena General Ecosystem Simulator	Smith (2007)
Noah	–	Ek et al. (2003)
ORCHIDEE	Organizing Carbon and Hydrology In Dynamic Ecosystems	Krinner et al. (2005), Vuichard et al. (2019), Zaehle, Friend, et al. (2010)
SDBM	Simple Diagnostic Biosphere Model	Knorr and Heimann (1995)
SiB	Simple Biosphere Model	Sellers et al. (1996)
SIPNET	Simplified Photosynthesis and Evapotranspiration	Braswell et al. (2005)
TECOS	Terrestrial ecosystem	Xu et al. (2006)

simple representations of vegetation and hydrology processes but may lack the full mechanistic representation of energy and hydrological processes or vegetation dynamics seen in LSMs and VDMs.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

This article discusses the challenges and priorities in the field of parameter estimation for land DA, and the opportunities offered by ML—it does not include the specific use of any particular software or results involving specific data products.

Acknowledgments
This manuscript is the result of efforts by the Analysis, Integration, and Modeling of the Earth System (AIMES) Land DA Working Group annual virtual workshops (https://aimesproject.org/ldawg/) and initiatives to build a Land DA Community (https://land-da-community.github.io). NR has been supported by a H2020 Marie Skłodowska-Curie Actions grant (no. 101026422). ND and TQ would like to acknowledge funding from the UKRI National Centre for Earth Observation, under the International Science Programme (NE/X006328/1). AR and MD acknowledge funding from NSF MSB

References

Abadie, C., Maignan, F., Remaud, M., Kohonen, K.-M., Sun, W., Kooijmans, L., et al. (2023). Carbon and water fluxes of the boreal evergreen needleleaf forest biome constrained by assimilating ecosystem carbonyl sulfide flux observations. *Journal of Geophysical Research: Biogeosciences*, *128*(7), e2023JG007407. https://doi.org/10.1029/2023jg007407

Abramoff, R., Xu, X., Hartman, M., O'Brien, S., Feng, W., Davidson, E., et al. (2018). The millennial model: In search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry*, *137*(1–2), 51–71. https://doi.org/10.1007/s10533-017-0409-7

Ait-El-Fquih, B., El Gharamti, M., & Hoteit, I. (2016). A Bayesian consistent dual ensemble Kalman filter for state-parameter estimation in subsurface hydrology. *Hydrology and Earth System Sciences*, *20*(8), 3289–3307. https://doi.org/10.5194/hess-20-3289-2016

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *arXiv [cs. LG]. arXiv*. http://arxiv.org/abs/1907.10902

Alton, P. B. (2013). From site-level to global simulation: Reconciling carbon, water and energy fluxes over different spatial scales using a process-based ecophysiological land-surface model. *Agricultural and Forest Meteorology*, *176*, 111–124. https://doi.org/10.1016/j.agrformet.2013.03.010

2406258 and MD further acknowledges funding from NASA CMS 80NSSC21K0965. KD, LH, and PG would like to acknowledge funding from the National Science Foundation (NSF) Science and Technology Center (STC) Learning the Earth with Artificial Intelligence and Physics (LEAP), Award # 2019625-STC. KD further acknowledges support from the U.S. Department of Energy, Office of Biological & Environmental Research (BER), under Lawrence Livermore National Lab subaward DE-AC52-07NA27344, Lawrence Berkeley National Lab subaward DE-ACE02-05CH11231, an Pacific Northwest National Lab subaward DE-AC05-76RL01830 and the National Science Foundation (NSF) National Center for Atmospheric Research (NCAR), which is a major facility sponsored by NSF under Cooperative Agreement No. 1852977. NM acknowledges support from the Wolfe-Western Fellowship At-Large for Outstanding Newly Recruited Research Scholars Endowment Fund. IF acknowledges funding from the Research Council of Finland (Grant 337552) and Horizon Europe, HORIZON-MISS-2022-SOIL-01-05. EC acknowledges funding from UKRI Grant NE/X019063/1, TK acknowledges support from ESA through contract 4000141232 within the Carbon Science Cluster and by the European Union's Horizon Europe Programme through Grant Agreement 101082194. TLS was funded by the National Centre for Earth Observation, under the LTSS programme (NE/R016518/1) and the UK's EO Climate Information Service (NE/X019071/1). MS acknowledges support from three Swedish strategic research areas: ModElling the Regional and Global Earth system (MERGE), the e-science collaboration (eSSENCE), and Biodiversity and Ecosystems in a Changing Climate (BECC). The authors would further like to thank the reviewers for their thoughtful and constructive feedback, their insights were invaluable in improving the clarity and quality of this work.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. https://doi.org/10.1175/2009BAMS2618.1

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, *17*(16), 4173–4222. https://doi.org/10.5194/bg-17-4173-2020

Arora, V. K., Seiler, C., Wang, L., & Kou-Giesbrecht, S. (2023). Towards an ensemble-based evaluation of land surface models in light of uncertain forcings and observations. *Biogeosciences*, *20*(7), 1313–1355. https://doi.org/10.5194/bg-20-1313-2023

Arsenault, K. R., Kumar, S. V., Geiger, J. V., Wang, S., Kemp, E., Mocko, D. M., et al. (2018). The Land surface Data Toolkit (LDT v7.2)—A data fusion environment for land data assimilation systems. *Geoscientific Model Development*, *11*(9), 3605–3621. https://doi.org/10.5194/gmd-11-3605-2018

Assem, H., Ghariba, S., Makrai, G., Johnston, P., Gill, L., & Pilla, F. (2017). Urban water flow and water level prediction based on deep learning. In *Machine learning and knowledge discovery in databases* (pp. 317–329). Springer International Publishing. https://doi.org/10.1007/978-3-319-71273-4_26

Baatz, R., Hendricks Franssen, H. J., Euskirchen, E., Sihi, D., Dietze, M., Ciavatta, S., et al. (2021). Reanalysis in earth system science: Toward terrestrial ecosystem reanalysis. *Reviews of Geophysics*, *59*(3), e2020RG000715. https://doi.org/10.1029/2020rg000715

Babst, F., Bouriaud, O., Alexander, R., Trouet, V., & Frank, D. (2014). Toward consistent measurements of carbon accumulation: A multi-site assessment of biomass and basal area increment across Europe. *Dendrochronologia*, *32*(2), 153–161. https://doi.org/10.1016/j.dendro.2014.01.002

Bacour, C., MacBean, N., Chevallier, F., Léonard, S., Koffi, E. N., & Peylin, P. (2023). Assimilation of multiple datasets results in large differences in regional-to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model. *Biogeosciences*, *20*(6), 1089–1111. https://doi.org/10.5194/bg-20-1089-2023

Bacour, C., Maignan, F., MacBean, N., Porcar-Castell, A., Flexas, J., Frankenberg, C., et al. (2019). Improving estimates of gross primary productivity by assimilating solar-induced fluorescence satellite retrievals in a terrestrial biosphere model using a process-based SIF model. *Journal of Geophysical Research: Biogeosciences*, *124*(11), 3281–3306. https://doi.org/10.1029/2019jg005040

Bacour, C., Peylin, P., MacBean, N., Rayner, P. J., Delage, F., Chevallier, F., et al. (2015). Joint assimilation of eddy covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model. *Journal of Geophysical Research: Biogeosciences*, *120*(9), 1839–1857. https://doi.org/10.1002/2015JG002966

Bagnara, M., Silveyra Gonzalez, R., Reifenberg, S., Steinkamp, J., Hickler, T., Werner, C., et al. (2019). An R package facilitating sensitivity analysis, calibration and forward simulations with the LPJ-GUESS dynamic vegetation model. *Environmental Modelling & Software: With Environment Data News*, *111*, 55–60. https://doi.org/10.1016/j.envsoft.2018.09.004

Baker, E., Harper, A. B., Williamson, D., & Challenor, P. (2022). Emulation of high-resolution land surface models using sparse Gaussian processes with application to JULES. *Geoscientific Model Development*, *15*(5), 1913–1929. https://doi.org/10.5194/gmd-15-1913-2022

Balsamo, G., Agusti-Panareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., et al. (2018). Satellite and in situ observations for advancing global Earth surface modelling: A review. *Remote Sensing*, *10*(12), 2038. https://doi.org/10.3390/rs10122038

Bannister, R. N., Chipilski, H. G., & Martinez-Alvarado, O. (2020). Techniques and challenges in the assimilation of atmospheric water observations for numerical weather prediction towards convective scales. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 1–48. https://doi.org/10.1002/qj.3652

Bao, S., Alonso, L., Wang, S., Gensheimer, J., De, R., & Carvalhais, N. (2023). Toward robust parameterizations in ecosystem-level photosynthesis models. *Journal of Advances in Modeling Earth Systems*, *15*(8), e2022MS003464. https://doi.org/10.1029/2022ms003464

Barichivich, J., Peylin, P., Launois, T., Daux, V., Risi, C., Jeong, J., & Luyssaert, S. (2021). A triple tree-ring constraint for tree growth and physiology in a global land surface model. *Biogeosciences*, *18*(12), 3781–3803. https://doi.org/10.5194/bg-18-3781-2021

Bastrikov, V., MacBean, N., Bacour, C., Santaren, D., Kuppel, S., & Peylin, P. (2018). Land surface model parameter optimisation using in situ flux data: Comparison of gradient-based versus random search algorithms (a case study using ORCHIDEE v1. 9.5. 2). *Geoscientific Model Development*, *11*(12), 4739–4754. https://doi.org/10.5194/gmd-11-4739-2018

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R., Ménard, C. B., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302. https://doi.org/10.1103/PhysRevLett.126.098302

Bhouri, M. A., Joly, M., Yu, R., Sarkar, S., & Perdikaris, P. (2023). Scalable Bayesian optimization with randomized prior networks. *Computer Methods in Applied Mechanics and Engineering*, *417*, 116428. https://doi.org/10.1016/j.cma.2023.116428

Bloom, A. A., Bowman, K. W., Liu, J., Konings, A. G., Worden, J. R., Parazoo, N. C., et al. (2020). Lagged effects regulate the inter-annual variability of the tropical carbon balance. *Biogeosciences*, *17*(24), 6393–6422. https://doi.org/10.5194/bg-17-6393-2020

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., & Williams, M. (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 of the United States of America*, *113*(5), 1285–1290. https://doi.org/10.1073/pnas.1515160113

Bloom, A. A., & Williams, M. (2015). Constraining ecosystem carbon dynamics in a data-limited world: Integrating ecological "common sense" in a model–data fusion framework. *Biogeosciences*, *12*(5), 1299–1315. https://doi.org/10.5194/bg-12-1299-2015

Blyth, E. M., Arora, V. K., Clark, D. B., Dadson, S. J., De Kauwe, M. G., Lawrence, D. M., et al. (2021). Advances in land surface modelling. *Current Climate Change Reports*, *7*(2), 45–71. https://doi.org/10.1007/s40641-021-00171-5

Bonan, G. B., & Doney, S. C. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, *359*(6375), eaam8328. https://doi.org/10.1126/science.aam8328

Booth, B. B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., et al. (2012). High sensitivity of future global warming to land carbon cycle processes. *Environmental Research Letters*, *7*(2), 024002. https://doi.org/10.1088/1748-9326/7/2/024002

Boussetta, S., Balsamo, G., Arduini, G., Dutra, E., McNorton, J., Choulga, M., et al. (2021). ECLand: The ECMWF land surface modelling system. *Atmosphere*, *12*(6), 723. https://doi.org/10.3390/atmos12060723

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., et al. (2018). JAX: Composable transformations of Python+NumPy programs (Version 0.3.13). Retrieved from http://github.com/google/jax

Braswell, B. H., Sacks, W. J., Linder, E., & Schimel, D. S. (2005). Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biology*, *11*(2), 335–355. https://doi.org/10.1111/j.1365-2486.2005.00897.x

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/bf00058655

Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, *30*(11), 114007. https://doi.org/10.1088/0266-5611/30/11/114007

Buotte, P. C., Koven, C. D., Xu, C., Shuman, J. K., Goulden, M. L., Levis, S., et al. (2021). Capturing functional strategies and compositional dynamics in vegetation demographic models. *Biogeosciences*, *18*(14), 4473–4490. https://doi.org/10.5194/bg-18-4473-2021

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing: A Publication of the Society for Industrial and Applied Mathematics*, *16*(5), 1190–1208. https://doi.org/10.1137/0916069

Caen, A., Smallman, T. L., de Castro, A. A., Robertson, E., von Randow, C., Cardoso, M., & Williams, M. (2022). Evaluating two land surface models for Brazil using a full carbon cycle benchmark with uncertainties. *Climate Resilience and Sustainability*, *1*(1), e10. https://doi.org/10.1002/cli2.10

Cameron, D., Hartig, F., Minnuno, F., Oberpriller, J., Reineking, B., Van Oijen, M., & Dietze, M. (2022). Issues in calibrating models with multiple unbalanced constraints: The significance of systematic model and data errors. *Methods in Ecology and Evolution*, *13*(12), 2757–2770. https://doi.org/10.1111/2041-210x.14002

Carvalhais, N., Reichstein, M., Ciais, P., Collatz, G. J., Mahecha, M. D., Montagnani, L., et al. (2010). Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints: Ecosystem C POOLS APART from equilibrium. *Global Change Biology*, *16*(10), 2813–2829. https://doi.org/10.1111/j.1365-2486.2010.02173.x

Carvalhais, N., Reichstein, M., Seixas, J., James Collatz, G., Pereira, J. S., Berbigier, P., et al. (2008). Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. *Global Biogeochemical Cycles*, *22*(2). https://doi.org/10.1029/2007GB003033

Castro-Morales, K., Schürmann, G., Köstler, C., Rödenbeck, C., Heimann, M., & Zaehle, S. (2019). Three decades of simulated global terrestrial carbon fluxes from a data assimilation system confronted with different periods of observations. *Biogeosciences*, *16*(15), 3009–3032. https://doi.org/10.5194/bg-16-3009-2019

Chaney, N. W., Herman, J. D., Ek, M. B., & Wood, E. F. (2016). Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. *Journal of Geophysical Research: Atmospheres*, *121*(22), 13218–13235. https://doi.org/10.1002/2016JD024821

Chen, B., Chen, J. M., & Ju, W. (2007). Remote sensing-based Ecosystem–Atmosphere Simulation Scheme (EASS)—Model formulation and test with multiple-year data. *Ecological Modelling*, *209*(2), 277–300. https://doi.org/10.1016/j.ecolmodel.2007.06.032

Chen, J. M., Mo, G., Pisek, J., Liu, J., Deng, F., Ishizawa, M., & Chan, D. (2012). Effects of foliage clumping on the estimation of global terrestrial gross primary productivity. *Global Biogeochemical Cycles*, *26*(1), GB1019. https://doi.org/10.1029/2010GB003996

Chen, W., Huang, C., Shen, H., & Li, X. (2015). Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation. *Advances in Water Resources*, *86*, 425–438. https://doi.org/10.1016/j.advwatres.2015.08.003

Cheng, Y., Wang, W., Detto, M., Fisher, R., & Shoemaker, C. (2024). Calibrating tropical forest coexistence in ecosystem demography models using multi-objective optimization through population-based parallel surrogate search. *Journal of Advances in Modeling Earth Systems*, *16*(8), e2023MS004195. https://doi.org/10.1029/2023ms004195

Cheng, Y., Xia, W., Detto, M., & Shoemaker, C. A. (2023). A framework to calibrate ecosystem demography models within earth system models using parallel surrogate global optimization. *Water Resources Research*, *59*(1), e2022WR032945. https://doi.org/10.1029/2022wr032945

Chevallier, F. (2007). Impact of correlated observation errors on inverted CO_2 surface fluxes from OCO measurements. *Geophysical Research Letters*, *34*(24), 24804. https://doi.org/10.1029/2007gl030463

Chevallier, F., Lloret, Z., Cozic, A., Takache, S., & Remaud, M. (2023). Toward high-resolution global atmospheric inverse modeling using graphics accelerators. *Geophysical Research Letters*, *50*(5), e2022GL102135. https://doi.org/10.1029/2022gl102135

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description–Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, *4*(3), 701–722. https://doi.org/10.5194/gmd-4-701-2011

Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., & Stuart, A. M. (2021). Calibrate, emulate, sample. *Journal of Computational Physics*, *424*, 109716. https://doi.org/10.1016/j.jcp.2020.109716

Collatz, G. J., Ribas-Carbo, M., & Berry, J. A. (1992). Coupled photosynthesis-stomatal conductance model for leaves of C4 plants. *Functional Plant Biology: FPB*, *19*(5), 519. https://doi.org/10.1071/pp9920519

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The international land model benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018ms001354

Cooper, E. S., Dance, S. L., García-Pintado, J., Nichols, N. K., & Smith, P. J. (2019). Observation operators for assimilation of satellite observations in fluvial inundation forecasting. *Hydrology and Earth System Sciences*, *23*(6), 2541–2559. https://doi.org/10.5194/hess-23-2541-2019

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., et al. (2021). Process-based climate model development harnessing machine learning: I. A calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, *13*(3), e2020MS002217. https://doi.org/10.1029/2020ms002217

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. *The Journal of Chemical Physics*, *59*(8), 3873–3878. https://doi.org/10.1063/1.1680571

Dagon, K., Sanderson, B. M., Fisher, R. A., & Lawrence, D. M. (2020). A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5. *Advances in Statistical Climatology Meteorology and Oceanography*, *6*(2), 223–244. https://doi.org/10.5194/ascmo-6-223-2020

Dahlin, K. M., Ponte, D. D., Setlock, E., & Nagelkirk, R. (2017). Global patterns of drought deciduous phenology in semi-arid and savanna-type ecosystems. *Ecography*, *40*(2), 314–323. https://doi.org/10.1111/ecog.02443

Dantec-Nédélec, S., Ottlé, C., Wang, T., Guglielmo, F., Maignan, F., Delbart, N., et al. (2017). Testing the capability of ORCHIDEE land surface model to simulate Arctic ecosystems: Sensitivity analysis and site-level model calibration. *Journal of Advances in Modeling Earth Systems*, *9*(2), 1212–1230. https://doi.org/10.1002/2016ms000860

Datta, P., & Faroughi, S. A. (2023). A multihead LSTM technique for prognostic prediction of soil moisture. *Geoderma*, *433*, 116452. https://doi.org/10.1016/j.geoderma.2023.116452

Davies-Barnard, T., Zaehle, S., & Friedlingstein, P. (2022). Assessment of the impacts of biological nitrogen fixation structural uncertainty in CMIP6 earth system models. *Biogeosciences*, *19*(14), 3491–3503. https://doi.org/10.5194/bg-19-3491-2022

de Bézenac, E., Pajot, A., & Gallinari, P. (2019). Deep learning for physical processes: Incorporating prior scientific knowledge. *Journal of Statistical Mechanics*, *2019*(12), 124009. https://doi.org/10.1088/1742-5468/ab3195

Deepak, R., Seiler, C., & Monahan, A. H. (2024). A global sensitivity analysis of parameter uncertainty in the CLASSIC model. In *Atmosphere-ocean* (pp. 1–13). https://doi.org/10.1080/07055900.2024.2396426

De Lannoy, G. J. M., Bechtold, M., Albergel, C., Brocca, L., Calvet, J.-C., Carrassi, A., et al. (2022). Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication. *Frontiers in Water*, *4*, 981745. https://doi.org/10.3389/frwa.2022.981745

Deng, C., Liu, P., Guo, S., Li, Z., & Wang, D. (2016). Identification of hydrological model parameter variation using ensemble Kalman filter. *Hydrology and Earth System Sciences*, *20*(12), 4949–4961. https://doi.org/10.5194/hess-20-4949-2016

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al. (2022). Coupled data assimilation at ECMWF: Current status, challenges and future developments. *Quarterly Journal of the Royal Meteorological Society*, *148*(747), 2672–2702. https://doi.org/10.1002/qj.4330

de Rosnay, P., Munoz-Sabater, J., Albergel, C., Isaksen, L., English, S., Drusch, M., & Wigneron, J. P. (2020). SMOS brightness temperature forward modelling and long-term monitoring at ECMWF. *Remote Sensing of Environment*, *237*, 111424. https://doi.org/10.1016/j.rse.2019.111424

Dietze, M. C. (2017). *Ecological forecasting*. Princeton University Press. https://doi.org/10.1515/9781400885459

Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., Kelly, R., et al. (2014). A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 286–300. https://doi.org/10.1002/2013jg002392

Dokoohaki, H., Morrison, B. D., Raiho, A., Serbin, S. P., Zarada, K., Dramko, L., & Dietze, M. (2022). Development of an open-source regional data assimilation system in PEcAn v. 1.7.2: Application to carbon cycle reanalysis across the contiguous US using SIPNET. *Geoscientific Model Development*, *15*(8), 3233–3252. https://doi.org/10.5194/gmd-15-3233-2022

Donnerer, C. (2024). *xgboost-distribution: Probabilistic prediction with XGBoost*. Github. Retrieved from https://github.com/CDonnerer/xgboost-distribution

Draper, C. S. (2021). Accounting for land model error in numerical weather prediction ensemble systems: Toward ensemble-based coupled land/atmosphere data assimilation. *Journal of Hydrometeorology*, *22*(8), 2089–2104. https://doi.org/10.1175/jhm-d-21-0016.1

Drusch, M., Holmes, T., de Rosnay, P., & Balsamo, G. (2009). Comparing ERA-40-based L-band brightness temperatures with Skylab observations: A calibration/validation study using the Community Microwave Emission Model. *Journal of Hydrometeorology*, *10*(1), 213–226. https://doi.org/10.1175/2008JHM990.1

Duan, T., Anand, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., & Schuler, A. (2020). NGBoost: Natural gradient boosting for probabilistic prediction. In H. D. Iii & A. Singh (Eds.), *Proceedings of the 37th international conference on machine learning* (Vol. 119, pp. 2690–2700). PMLR. Retrieved from https://proceedings.mlr.press/v119/duan20a.html

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., et al. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Egyptian Journal of Remote Sensing and Space Sciences*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

Durbha, S. S., King, R. L., & Younan, N. H. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, *107*(1–2), 348–361. https://doi.org/10.1016/j.rse.2006.09.031

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, *108*(D22), 8851. https://doi.org/10.1029/2002jd003296

ElGhawi, R., Kraft, B., Reimers, C., Reichstein, M., Körner, M., Gentine, P., & Winkler, A. J. (2023). Hybrid modeling of evapotranspiration: Inferring stomatal and aerodynamic resistances using combined physics-based and machine learning. *Environmental Research Letters*, *18*(3), 034039. https://doi.org/10.1088/1748-9326/acbbe0

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Exbrayat, J.-F., Pitman, A. J., & Abramowitz, G. (2014). Response of microbial decomposition to spin-up explains CMIP5 soil carbon range until 2100. *Geoscientific Model Development*, *7*(6), 2683–2692. https://doi.org/10.5194/gmd-7-2683-2014

Exbrayat, J.-F., Smallman, T. L., Bloom, A. A., Hutley, L. B., & Williams, M. (2018). Inverse determination of the influence of fire on vegetation carbon turnover in the pantropics. *Global Biogeochemical Cycles*, *32*(12), 1776–1789. https://doi.org/10.1029/2018gb005925

Eyring, V., Collins, W. D., Gentine, P., Barnes, E. A., Barreiro, M., Beucler, T., et al. (2024). Pushing the frontiers in climate modelling and analysis with machine learning. *Nature Climate Change*, *14*(9), 916–928. https://doi.org/10.1038/s41558-024-02095-y

Famiglietti, C. A., Smallman, T. L., Levine, P. A., Flack-Prain, S., Quetin, G. R., Meyer, V., et al. (2021). Optimal model complexity for terrestrial carbon cycle prediction. *Biogeosciences*, *18*(8), 2727–2754. https://doi.org/10.5194/bg-18-2727-2021

Farchi, A., Chrust, M., Bocquet, M., Laloyaux, P., & Bonavita, M. (2023). Online model error correction with neural networks in the incremental 4D-Var framework. *Journal of Advances in Modeling Earth Systems*, *15*(9), e2022MS003474. https://doi.org/10.1029/2022ms003474

Farchi, A., Laloyaux, P., Bonavita, M., & Bocquet, M. (2021). Using machine learning to correct model error in data assimilation and forecast applications. *Quarterly Journal of the Royal Meteorological Society*, *147*(739), 3067–3084. https://doi.org/10.1002/qj.4116

Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C 3 species. *Planta*, *149*(1), 78–90. https://doi.org/10.1007/BF00386231

Fer, I., Gardella, A. K., Shiklomanov, A. N., Campbell, E. E., Cowdery, E. M., De Kauwe, M. G., et al. (2021). Beyond ecosystem modeling: A roadmap to community cyberinfrastructure for ecological data-model integration. *Global Change Biology*, *27*(1), 13–26. https://doi.org/10.1111/gcb.15409

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., & Dietze, M. C. (2018). Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*, *15*(19), 5801–5830. https://doi.org/10.5194/bg-15-5801-2018

Fer, I., Shiklomanov, A., Novick, K. A., Gough, C. M., Altaf Arain, M., Chen, J., et al. (2021). Capturing site-to-site variability through Hierarchical Bayesian calibration of a process-based dynamic vegetation model. In *bioRxiv*. https://doi.org/10.1101/2021.04.28.441243

Finn, T. S., Durand, C., Farchi, A., Bocquet, M., Chen, Y., Carrassi, A., & Dansereau, V. (2023). Deep learning subgrid-scale parametrisations for short-term forecasting of sea-ice dynamics with a Maxwell elasto-brittle rheology. *The Cryosphere*, *17*(7), 2965–2991. https://doi.org/10.5194/tc-17-2965-2023

Fisher, R. A., & Koven, C. D. (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*(4), e2018MS001453. https://doi.org/10.1029/2018ms001453

Fisher, R. A., Muszala, S., Verteinstein, M., Lawrence, P., Xu, C., McDowell, N. G., et al. (2015). Taking off the training wheels: The properties of a dynamic vegetation model without climate envelopes, CLM4.5. *Geoscientific Model Development*, *8*(11), 3593–3619. https://doi.org/10.5194/gmd-8-3593-2015

Fisher, R. A., Wieder, W. R., Sanderson, B. M., Koven, C. D., Oleson, K. W., Xu, C., et al. (2019). Parametric controls on vegetation responses to biogeochemical forcing in the CLM5. *Journal of Advances in Modeling Earth Systems*, *11*(9), 2879–2895. https://doi.org/10.1029/2019ms001609

Forkel, M., Carvalhais, N., Schaphoff, S., von Bloh, W., Migliavacca, M., Thurner, M., & Thonicke, K. (2014). Identifying environmental controls on vegetation greenness phenology through model–data integration. *Biogeosciences*, *11*(23), 7025–7050. https://doi.org/10.5194/bg-11-7025-2014

Forkel, M., Drüke, M., Thurner, M., Dorigo, W., Schaphoff, S., Thonicke, K., et al. (2019). Constraining modelled global vegetation dynamics and carbon turnover using multiple satellite observations. *Scientific Reports*, *9*(1), 18757. https://doi.org/10.1038/s41598-019-55187-7

Fox, A., Huo, X., Hoar, T., Dashti, H., Smith, W. K., MacBean, N., 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*(8), e2022JG006830. https://doi.org/10.1029/2022JG006830

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., 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*(10), 1597–1615. https://doi.org/10.1016/j.agrformet.2009.05.002

Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., et al. (2018). Evaluation of a data assimilation system for land surface models using CLM4.5. *Journal of Advances in Modeling Earth Systems*, *10*, 2471–24942. https://doi.org/10.1029/2018MS001362

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., et al. (2023). Global Carbon Budget 2023. *Earth System Science Data*, *15*(12), 5301–5369. https://doi.org/10.5194/essd-15-5301-2023

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Ali Eslami, S. M., & Teh, Y. W. (2018). Neural processes. *arXiv [cs.LG]. arXiv*. Retrieved from http://arxiv.org/abs/1807.01622

Geer, A. J. (2021). Learning earth system models from observations: Machine learning or data assimilation? *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *379*(2194), 20200089. https://doi.org/10.1098/rsta.2020.0089

Gelbrecht, M., White, A., Bathiany, S., & Boers, N. (2023). Differentiable programming for Earth system modeling. *Geoscientific Model Development*, *16*(11), 3123–3135. https://doi.org/10.5194/gmd-16-3123-2023

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751. https://doi.org/10.1029/2018gl078202

Gier, B. K., Schlund, M., Friedlingstein, P., Jones, C. D., Jones, C., Zaehle, S., & Eyring, V. (2024). Representation of the terrestrial carbon cycle in CMIP6. *arXiv [physics.ao-ph]. arXiv*. https://doi.org/10.5194/egusphere-2024-277

Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, *3*(2/3), 95–99. https://doi.org/10.1023/a:1022602019183

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., & Gentine, P. (2019). Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature*, *565*(7740), 476–479. https://doi.org/10.1038/s41586-018-0848-x

Green, J. K., Zhang, Y., Luo, X., & Keenan, T. F. (2024). Systematic underestimation of canopy conductance sensitivity to drought by Earth System Models. *AGU Advances*, *5*(1), e2023AV001026. https://doi.org/10.1029/2023av001026

Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., & Zanna, L. (2023). Deep learning of systematic sea ice model errors from data assimilation increments. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003757. https://doi.org/10.1029/2023ms003757

Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., & Zanna, L. (2024). Machine learning for online sea ice bias correction within global ice-ocean simulations. *Geophysical Research Letters*, *51*(3), e2023GL106776. https://doi.org/10.1029/2023gl106776

Griewank, A. (1989). On automatic differentiation. Retrieved from https://www.researchgate.net/publication/2703247_On_Automatic_Differentiation

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? *arXiv [cs.LG]. arXiv*. Retrieved from http://arxiv.org/abs/2207.08815

Groenendijk, M., Dolman, A. J., Ammann, C., Arneth, A., Cescatti, A., Dragoni, D., et al. (2011). Seasonal variation of photosynthetic model parameters and leaf area index from global FluxNet eddy covariance data. *Biogeosciences*, *116*(G4), G04027. https://doi.org/10.1029/2011JG001742

Guo, Y., Yu, X., Xu, Y.-P., Chen, H., Gu, H., & Xie, J. (2021). AI-based techniques for multi-step streamflow forecasts: Application for multi-objective reservoir operation optimization and performance assessment. *Hydrology and Earth System Sciences*, *25*(11), 5951–5979. https://doi.org/10.5194/hess-25-5951-2021

Hararuk, O., Xia, J., & Luo, Y. (2014). Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 403–417. https://doi.org/10.1002/2013jg002535

Harden, J. W., Hugelius, G., Ahlström, A., Blankinship, J. C., Bond-Lamberty, B., Lawrence, C. R., et al. (2018). Networking our science to characterize the state, vulnerabilities, and management opportunities of soil organic matter. *Global Change Biology*, *24*(2), e705–e718. https://doi.org/10.1111/gcb.13896

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Hatfield, S., Chantry, M., Dueben, P., Lopez, P., Geer, A., & Palmer, T. (2021). Building tangent-linear and adjoint models for data assimilation with neural networks. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002521. https://doi.org/10.1029/2021ms002521

Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms* (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/0471671746

Heinrich, V. H. A., Dalagnol, R., Cassol, H. L. G., Rosan, T. M., de Almeida, C. T., Silva Junior, C. H. L., et al. (2021). Large carbon sink potential of secondary forests in the Brazilian Amazon to mitigate climate change. *Nature Communications*, *12*(1), 1785. https://doi.org/10.1038/s41467-021-22050-1

Heinrich, V. H. A., Vancutsem, C., Dalagnol, R., Rosan, T. M., Fawcett, D., Silva-Junior, C. H. L., et al. (2023). The carbon sink of secondary and degraded humid tropical forests. *Nature*, *615*(7952), 436–442. https://doi.org/10.1038/s41586-022-05679-w

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., et al. (2018). Operational global reanalysis: Progress, future directions and synergies with NWP|ERA Report Series. *ECMWF*. Retrieved from https://www.ecmwf.int/en/elibrary/80922-operational-global-reanalysis-progress-future-directions-and-synergies-nwp

Holmes, T., Drusch, M., Wigneron, J.-P., & de Jeu, R. (2008). A global simulation of microwave emission: Error structures based on output from ECMWF's operational Integrated Forecast System. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(3), 846–856. https://doi.org/10.1109/tgrs.2007.914798

Hourdin, F., Ferster, B., Deshayes, J., Mignot, J., Musat, I., & Williamson, D. (2023). Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections. *Science Advances*, *9*(29), eadf2758. https://doi.org/10.1126/sciadv.adf2758

Huang, Y., Zhu, D., Ciais, P., Guenet, B., Huang, Y., Goll, D. S., et al. (2018). Matrix-based sensitivity assessment of soil organic carbon storage: A case study from the ORCHIDEE-MICT model. *Journal of Advances in Modeling Earth Systems*, *10*(8), 1790–1808. https://doi.org/10.1029/2017MS001237

Huo, X., Fox, A., Dashti, H., Devine, C., Gallery, W., Smith, W., et al. (2024). Integrating state data assimilation and innovative model parameterization reduces simulated carbon uptake in the Arctic and Boreal region. *Journal of Geophysical Research: Biogeosciences*, *129*(9), e2024JG008004. https://doi.org/10.1029/2024JG008004

Irvine, P. J., Gregoire, L. J., Lunt, D. J., & Valdes, P. J. (2013). An efficient method to generate a perturbed parameter ensemble of a fully coupled AOGCM without flux-adjustment. *Geoscientific Model Development*, *6*(5), 1447–1462. https://doi.org/10.5194/gmd-6-1447-2013

Jeong, J., Barichivich, J., Peylin, P., Haverd, V., McGrath, M. J., Vuichard, N., et al. (2021). Using the International Tree-Ring Data Bank (ITRDB) records as century-long benchmarks for global land-surface models. *Geoscientific Model Development*, *14*(9), 5891–5913. https://doi.org/10.5194/gmd-14-5891-2021

Jian, J., Vargas, R., Anderson-Teixeira, K., Stell, E., Herrmann, V., Horn, M., et al. (2021). A restructured and updated global soil respiration database (SRDB-V5). *Earth System Science Data*, *13*(2), 255–267. https://doi.org/10.5194/essd-13-255-2021

Jiang, M., Medlyn, B. E., Drake, J. E., Duursma, R. A., Anderson, I. C., Barton, C. V. M., et al. (2020). The fate of carbon in a mature forest under carbon dioxide enrichment. *Nature*, *580*(7802), 227–231. https://doi.org/10.1038/s41586-020-2128-9

Joiner, J., Yoshida, Y., Zhang, Y., Duveiller, G., Jung, M., Lyapustin, A., et al. (2018). Estimation of terrestrial global Gross Primary Production (GPP) with satellite data-driven models and Eddy covariance flux data. *Remote Sensing*, *10*(9), 1346. https://doi.org/10.3390/rs10091346

Jones, S., Mercado, L. M., Bruhn, D., Raoult, N., & Cox, P. M. (2024). Night-time decline in plant respiration is consistent with substrate depletion. *Communications Earth & Environment*, *5*(1), 1–9. https://doi.org/10.1038/s43247-024-01312-y

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., et al. (2002). An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, *83*(1–2), 3–15. https://doi.org/10.1016/S0034-4257(02)00084-6

Kallingal, J. T., Scholze, M., Miller, P. A., Lindström, J., Rinne, J., Aurela, M., et al. (2024). Assimilating multi-site eddy-covariance data to calibrate the CH_4 wetland emission module in a terrestrial ecosystem model. *EGUsphere* (pp. 1–32). https://doi.org/10.5194/egusphere-2024-373

Kaminski, T., Heimann, M., & Giering, R. (1999). A coarse grid three-dimensional global inverse model of the atmospheric transport: 1. Adjoint model and Jacobian matrix. *Journal of Geophysical Research*, *104*(D15), 18535–18553. https://doi.org/10.1029/1999jd900147

Kaminski, T., Knorr, W., Rayner, P., & Heimann, M. (2002). Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Global Biogeochemical Cycles*, *16*(4), 1066. https://doi.org/10.1029/2001GB001463

Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., & Mathieu, P.-P. (2012). Consistent assimilation of MERIS FAPAR and atmospheric CO_2 into a terrestrial vegetation model and interactive mission benefit analysis. *Biogeosciences*, *9*(8), 3173–3184. https://doi.org/10.5194/bg-9-3173-2012

Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P. J., Zaehle, S., et al. (2013). The BETHY/JSBACH carbon cycle data assimilation system: Experiences and challenges. *Journal of Geophysical Research: Biogeosciences*, *118*(4), 1414–1426. https://doi.org/10.1002/jgrg.20118

Kaminski, T., & Mathieu, P.-P. (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*(9), 2343–2357. https://doi.org/10.5194/bg-14-2343-2017

Kaminski, T., & Rayner, P. J. (2008). Assimilation and network design. In H. Dolman, A. Freibauer, & R. Valentini (Eds.), *Observing the continental scale greenhouse gas balance of Europe (ecological studies)* (pp. 33–52). Springer-Verlag. https://doi.org/10.1007/978-0-387-76570-9-3

Kaminski, T., Scholze, M., & Houweling, S. (2010). Quantifying the benefit of A-SCOPE data for reducing uncertainties in terrestrial carbon fluxes in CCDAS. *Tellus Series B Chemical and Physical Meteorology*, *62*(5), 784–796. https://doi.org/10.1111/j.1600-0889.2010.00483.x

Kaminski, T., Scholze, M., Voßbeck, M., Knorr, W., Buchwitz, M., & Reuter, M. (2017). Constraining a terrestrial biosphere model with remotely sensed atmospheric carbon dioxide. *Remote Sensing of Environment*, *203*, 109–124. https://doi.org/10.1016/j.rse.2017.08.017

Kato, T., Knorr, W., Scholze, M., Veenendaal, E., Kaminski, T., Kattge, J., & Gobron, N. (2013). Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana. *Biogeosciences*, *10*(2), 789–802. https://doi.org/10.5194/bg-10-789-2013

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., et al. (2020). TRY plant trait database—Enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188. https://doi.org/10.1111/gcb.14904

Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W., & Richardson, A. D. (2012). Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. *Global Change Biology*, *18*(8), 2555–2569. https://doi.org/10.1111/j.1365-2486.2012.02684.x

Keenan, T. F., Davidson, E. A., Munger, J. W., & Richardson, A. D. (2013). Rate my data: Quantifying the value of ecological data for the development of models of the terrestrial carbon cycle. *Ecological Applications: A Publication of the Ecological Society of America*, *23*(1), 273–286. https://doi.org/10.1890/12-0747.1

Kennedy, D., Dagon, K., Lawrence, D. M., Fisher, R. A., Sanderson, B. M., Collier, N., et al. (2024). One-at-a-time parameter perturbation ensemble of the community land model, version 5.1. In *ESS open archive*. https://doi.org/10.22541/essoar.172745082.24089296/v1

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *63*(3), 425–464. https://doi.org/10.1111/1467-9868.00294

Kennedy, R. E., Serbin, S. P., Dietze, M. C., Andersen, H. E., Babcock, C., Baker, D. F., et al. (2024). Characterizing and communicating uncertainty: Lessons from NASA's Carbon Monitoring System. *Environmental Research Letters*, *19*(12), 123003. https://doi.org/10.1088/1748-9326/ad8be0

Kerschke, P., Hoos, H. H., Neumann, F., & Trautmann, H. (2019). Automated Algorithm Selection: Survey and perspectives. *Evolutionary Computation*, *27*(1), 3–45. https://doi.org/10.1162/evco_a_00242

Kloek, T., & Van Dijk, H. K. (1978). Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica: Journal of the Econometric Society*, *46*(1), 1. https://doi.org/10.2307/1913641

Knorr, W. (2000). Annual and interannual CO_2 exchanges of the terrestrial biosphere: Process-based simulations and uncertainties: CO_2 exchanges of the terrestrial biosphere. *Global Ecology and Biogeography: A Journal of Macroecology*, *9*(3), 225–252. https://doi.org/10.1046/j.1365-2699.2000.00159.x

Knorr, W., & Heimann, M. (1995). Impact of drought stress and other factors on seasonal land biosphere CO_2 exchange studied through an atmospheric tracer transport model. *Tellus Series B Chemical and Physical Meteorology*, *47*(4), 471–489. https://doi.org/10.1034/j.1600-0889.47.issue4.7.x

Knorr, W., & Heimann, M. (2001). Uncertainties in global terrestrial biosphere modeling, Part II: Global constraints for a process-based vegetation model. *Global Biogeochemical Cycles*, *15*(1), 227–246. https://doi.org/10.1029/1998GB001060

Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R., & Mathieu, P.-P. (2010). Carbon cycle data assimilation with a generic phenology model. *Journal of Geophysical Research*, *115*(G4), 125. https://doi.org/10.1029/2009jg001119

Knorr, W., & Kattge, J. (2005). Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biology*, *11*(8), 1333–1351. https://doi.org/10.1111/j.1365-2486.2005.00977.x

Knorr, W., Williams, M., Thum, T., Kaminski, T., Voßbeck, M., Scholze, M., et al. (2025). A comprehensive land-surface vegetation model for multi-stream data assimilation, D&B v1.0. *Geoscientific Model Development*, *18*(7), 2137–2159. https://doi.org/10.5194/gmd-18-2137-2025

Knutti, R., Stocker, T. F., Joos, F., & Plattner, G.-K. (2003). Probabilistic climate change projections using neural networks. *Climate Dynamics*, *21*(3), 257–272. https://doi.org/10.1007/s00382-003-0345-1

Koffi, E. N., Rayner, P. J., Scholze, M., & Beer, C. (2012). Atmospheric constraints on gross primary productivity and net ecosystem productivity: Results from a carbon-cycle data assimilation system. *Global Biogeochemical Cycles*, *26*(1), GB1024. https://doi.org/10.1029/2010gb003900

Kolassa, J., Reichle, R. H., Liu, Q., Cosh, M., Bosch, D. D., Caldwell, T. G., et al. (2017). Data assimilation to extract soil moisture information from SMAP observations. *Remote Sensing*, *9*(11), 1179. https://doi.org/10.3390/rs9111179

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., & Miralles, D. G. (2022). A deep learning-based hybrid model of global terrestrial evaporation. *Nature Communications*, *13*(1), 1912. https://doi.org/10.1038/s41467-022-29543-7

Koven, C. D., Arora, V. K., Cadule, P., Fisher, R. A., Jones, C. D., Lawrence, D. M., et al. (2022). Multi-century dynamics of the climate and carbon cycle under both high and net negative emissions scenarios. *Earth System Dynamics*, *13*(2), 885–909. https://doi.org/10.5194/esd-13-885-2022

Koven, C. D., Knox, R. G., Fisher, R. A., Chambers, J. Q., Christoffersen, B. O., Davies, S. J., et al. (2020). Benchmarking and parameter sensitivity of physiological and vegetation dynamics using the Functionally Assembled Terrestrial Ecosystem Simulator (FATES) at Barro Colorado Island, Panama. *Biogeosciences*, *17*(11), 3017–3044. https://doi.org/10.5194/bg-17-3017-2020

Kowalczyk, E. A., Wang, Y. P., Law, R. M., Davies, H. L., McGregor, J. L., & Abramowitz, G. S. (2006). *The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model*. CSIRO Marine and Atmospheric Research. https://doi.org/10.4225/08/58615C6A9A51D

Kraft, B., Jung, M., Körner, M., Koirala, S., & Reichstein, M. (2022). Towards hybrid modeling of the global hydrological cycle. *Hydrology and Earth System Sciences*, *26*(6), 1579–1614. https://doi.org/10.5194/hess-26-1579-2022

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., et al. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system: Dvgm for coupled climate studies. *Global Biogeochemical Cycles*, *19*(1), GB1015. https://doi.org/10.1029/2003gb002199

Kumar, S., Kolassa, J., Reichle, R., Crow, W., de Lannoy, G., de Rosnay, P., et al. (2022). An agenda for land data assimilation priorities: Realizing the promise of terrestrial water, energy, and vegetation observations from space. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003259. https://doi.org/10.1029/2022ms003259

Kumar, S. V., Reichle, R. H., Harrison, K. W., Peters-Lidard, C. D., Yatheendradas, S., & Santanello, J. A. (2012). A comparison of methods for a priori bias correction in soil moisture data assimilation: Bias correction in soil moisture data assimilation. *Water Resources Research*, *48*(3), 3515. https://doi.org/10.1029/2010wr010261

Kuppel, S., Chevallier, F., & Peylin, P. (2013). Quantifying the model structural error in carbon cycle data assimilation systems. *Geoscientific Model Development*, *6*(1), 45–55. https://doi.org/10.5194/gmd-6-45-2013

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., & Richardson, A. D. (2012). Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences*, *9*(10), 3757–3776. https://doi.org/10.5194/bg-9-3757-2012

Kuppel, S., Peylin, P., Maignan, F., Chevallier, F., Kiely, G., Montagnani, L., & Cescatti, A. (2014). Model–data fusion across ecosystems: From multisite optimizations to global simulations. *Geoscientific Model Development*, *7*(6), 2581–2597. https://doi.org/10.5194/gmd-7-2581-2014

Kwon, Y., Forman, B. A., Ahmad, J. A., Kumar, S. V., & Yoon, Y. (2019). Exploring the utility of machine learning-based passive microwave brightness temperature data assimilation over terrestrial snow in High Mountain Asia. *Remote Sensing*, *11*(19), 2265. https://doi.org/10.3390/rs11192265

Kyker-Snowman, E., Lombardozzi, D. L., Bonan, G. B., Cheng, S. J., Dukes, J. S., Frey, S. D., et al. (2022). Increasing the spatial and temporal impact of ecological research: A roadmap for integrating a novel terrestrial process into an Earth system model. *Global Change Biology*, *28*(2), 665–684. https://doi.org/10.1111/gcb.15894

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421. https://doi.org/10.1126/science.adi2336

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

Lawrence, C. R., Beem-Miller, J., Hoyt, A. M., Monroe, G., Sierra, C. A., Stoner, S., et al. (2020). An open-source database for the synthesis of soil radiocarbon data: International Soil Radiocarbon Database (ISRaD) version 1.0. *Earth System Science Data*, *12*(1), 61–76. https://doi.org/10.5194/essd-12-61-2020

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018ms001583

LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., & Dietze, M. C. (2013). Facilitating feedbacks between field measurements and ecosystem models. *Ecological Monographs*, *83*(2), 133–154. https://doi.org/10.1890/12-0137.1

Le Noë, J., Manzoni, S., Abramoff, R., Bölscher, T., Bruni, E., Cardinael, R., et al. (2023). Soil organic carbon models need independent time-series validation for reliable prediction. *Communications Earth & Environment*, *4*(1), 158. https://doi.org/10.1038/s43247-023-00830-5

Li, L., Fang, Y., Zheng, Z., Shi, M., Longo, M., Koven, C. D., et al. (2023). A machine learning approach targeting parameter estimation for plant functional type coexistence modeling using ELM-FATES (v2.0). *Geoscientific Model Development*, *16*(14), 4017–4040. https://doi.org/10.5194/gmd-16-4017-2023

Li, R., Lombardozzi, D., Shi, M., Frankenberg, C., Parazoo, N. C., Köhler, P., et al. (2022). Representation of leaf-to-canopy radiative transfer processes improves simulation of far-red solar-induced chlorophyll fluorescence in the Community Land Model version 5. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002747. https://doi.org/10.1029/2021MS002747

Liu, C., Xiao, Q., & Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Monthly Weather Review*, *136*(9), 3363–3373. https://doi.org/10.1175/2008mwr2312.1

Liu, M., He, H., Ren, X., Sun, X., Yu, G., Han, S., et al. (2015). The effects of constraining variables on parameter optimization in carbon and water flux modeling over different forest ecosystems. *Ecological Modelling*, *303*, 30–41. https://doi.org/10.1016/j.ecolmodel.2015.01.027

Lorenc, A. C. (1981). A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, *109*(4), 701–721. https://doi.org/10.1175/1520-0493(1981)109<0701:agtdms>2.0.co;2

Lu, D., & Ricciuto, D. (2019). Efficient surrogate modeling methods for large-scale Earth system models based on machine-learning techniques. *Geoscientific Model Development*, *12*(5), 1791–1807. https://doi.org/10.5194/gmd-12-1791-2019

Lü, H., Yu, Z., Zhu, Y., Drake, S., Hao, Z., & Sudicky, E. A. (2011). Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation. *Advances in Water Resources*, *34*(3), 395–406. https://doi.org/10.1016/j.advwatres.2010.12.005

Lu, M., Lu, S., Liao, W., Lei, X., Yin, Z., & Wang, H. (2022). Research and application of parameter estimation method in hydrological model based on dual ensemble Kalman filter. *Hydrology Research*, *53*(1), 65–84. https://doi.org/10.2166/nh.2021.272

Lu, X., Wang, Y.-P., Ziehn, T., & Dai, Y. (2013). An efficient method for global parameter sensitivity analysis and its applications to the Australian community land surface model (CABLE). *Agricultural and Forest Meteorology*, *182–183*, 292–303. https://doi.org/10.1016/j.agrformet.2013.04.003

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., et al. (2022). Matrix approach to land carbon cycle modeling. *Journal of Advances in Modeling Earth Systems*, *14*(7), e2022MS003008. https://doi.org/10.1029/2022ms003008

Luo, Y., Keenan, T. F., & Smith, M. (2015). Predictability of the terrestrial carbon cycle. *Global Change Biology*, *21*(5), 1737–1751. https://doi.org/10.1111/gcb.12766

Ma, L., Hurtt, G., Ott, L., Sahajpal, R., Fisk, J., Lamb, R., et al. (2022). Global evaluation of the Ecosystem Demography model (ED v3. 0). *Geoscientific Model Development*, *15*(5), 1971–1994. https://doi.org/10.5194/gmd-15-1971-2022

MacBean, N., Bacour, C., Raoult, N., Bastrikov, V., Koffi, E. N., Kuppel, S., 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*(7), e2021GB007177. https://doi.org/10.1029/2021gb007177

MacBean, N., Liddy, H., Quaife, T., Kolassa, J., & Fox, A. (2022). Building a land data assimilation community to tackle technical challenges in quantifying and reducing uncertainty in land model predictions. *Bulletin of the American Meteorological Society*, *103*(3), E733–E740. https://doi.org/10.1175/bams-d-21-0228.1

MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., et al. (2018). Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data. *Scientific Reports*, *8*(1), 1973. https://doi.org/10.1038/s41598-018-20024-w

MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., & Ciais, P. (2015). Using satellite data to improve the leaf phenology of a global terrestrial biosphere model. *Biogeosciences*, *12*(23), 7185–7208. https://doi.org/10.5194/bg-12-7185-2015

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geoscientific Model Development*, *9*(10), 3569–3588. https://doi.org/10.5194/gmd-9-3569-2016

Mahmud, K., Medlyn, B. E., Duursma, R. A., Campany, C., & De Kauwe, M. G. (2018). Inferring the effects of sink strength on plant carbon balance processes from experimental measurements. *Biogeosciences*, *15*(13), 4003–4018. https://doi.org/10.5194/bg-15-4003-2018

Mäkelä, J., Arppe, L., Fritze, H., Heinonsalo, J., Karhu, K., Liski, J., et al. (2022). Implementation and initial calibration of carbon-13 soil organic matter decomposition in the Yasso model. *Biogeosciences*, *19*(17), 4305–4313. https://doi.org/10.5194/bg-19-4305-2022

Mäkelä, J., Knauer, J., Aurela, M., Black, A., Heimann, M., Kobayashi, H., et al. (2019). Parameter calibration and stomatal conductance formulation comparison for boreal forests with adaptive population importance sampler in the land surface model JSBACH. *Geoscientific Model Development*, *12*(9), 4075–4098. https://doi.org/10.5194/gmd-12-4075-2019

Massoud, E. C., Xu, C., Fisher, R. A., Knox, R. G., Walker, A. P., Serbin, S. P., et al. (2019). Identification of key parameters controlling demographically structured vegetation dynamics in a land surface model: CLM4.5(FATES). *Geoscientific Model Development*, *12*(9), 4133–4164. https://doi.org/10.5194/gmd-12-4133-2019

Mathers, C., Black, C. K., Segal, B. D., Gurung, R. B., Zhang, Y., Easter, M. J., et al. (2023). Validating DayCent-CR for cropland soil carbon offset reporting at a national scale. *Geoderma*, *438*, 116647. https://doi.org/10.1016/j.geoderma.2023.116647

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018ms001400

McGovern, A., Elmore, K. L., Gagne, D. J., II, Haupt, S. E., Karstens, C. D., Lagerquist, R., et al. (2017). Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bulletin of the American Meteorological Society*, *98*(10), 2073–2090. https://doi.org/10.1175/bams-d-16-0123.1

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, *21*(2), 239–245. https://doi.org/10.2307/1268522

McNeall, D., Robertson, E., & Wiltshire, A. (2024). Constraining the carbon cycle in JULES-ES-1.0. *Geoscientific Model Development*, *17*(3), 1059–1089. https://doi.org/10.5194/gmd-17-1059-2024

Medvigy, D., Wofsy, S. C., Munger, J. W., Hollinger, D. Y., & Moorcroft, P. R. (2009). Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. *Journal of Geophysical Research*, *114*(G1), G01002. https://doi.org/10.1029/2008JG000812

Meirink, J. F., Bergamaschi, P., & Krol, M. C. (2008). Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: Method and comparison with synthesis inversion. *Atmospheric Chemistry and Physics*, *8*(21), 6341–6353. https://doi.org/10.5194/acp-8-6341-2008

Melton, J. R., Arora, V. K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E., & Teckentrup, L. (2020). CLASSIC v1.0: The open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM)—Part 1: Model framework and site-level performance. *Geoscientific Model Development*, *13*(6), 2825–2850. https://doi.org/10.5194/gmd-13-2825-2020

Meunier, F., Visser, M. D., Shiklomanov, A., Dietze, M. C., Guzmán Q, J. A., Sanchez-Azofeifa, G. A., et al. (2022). Liana optical traits increase tropical forest albedo and reduce ecosystem productivity. *Global Change Biology*, *28*(1), 227–244. https://doi.org/10.1111/gcb.15928

Meyer, D., Grimmond, S., Dueben, P., Hogan, R., & van Reeuwijk, M. (2022). Machine learning emulation of urban land surface processes. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002744. https://doi.org/10.1029/2021ms002744

Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A method for scaling vegetation dynamics: The ecosystem demography model. *Ecological Monographs*, *71*(4), 557–586. https://doi.org/10.1890/0012-9615(2001)071[0557:AMFSVD]2.0.CO;2

Moore, D. J. P., Hu, J., Sacks, W. J., Schimel, D. S., & Monson, R. K. (2008). Estimating transpiration and the sensitivity of carbon uptake to water availability in a subalpine forest using a simple ecosystem process model informed by measured net CO_2 and H_2O fluxes. *Agricultural and Forest Meteorology*, *148*(10), 1467–1477. https://doi.org/10.1016/j.agrformet.2008.04.013

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, *33*(2), 161–174. https://doi.org/10.1080/00401706.1991.10484804

Nave, L., Johnson, K., van Ingen, C., Agarwal, D., Humphrey, M., & Beekwilder, N. (2016). International Soil Carbon Network (ISCN) database v3-1. https://doi.org/10.17040/ISCN/1305039

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., et al. (2024). X-BASE: The first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X. In *EGUsphere*. https://doi.org/10.5194/egusphere-2024-165

Nicolini, G., Castaldi, S., Fratini, G., & Valentini, R. (2013). A literature overview of micrometeorological CH_4 and N_2O flux measurements in terrestrial ecosystems. *Atmospheric Environment*, *81*, 311–319. https://doi.org/10.1016/j.atmosenv.2013.09.030

Nie, S., Zhu, J., & Luo, Y. (2011). Simultaneous estimation of land surface scheme states and parameters using the ensemble Kalman filter: Identical twin experiments. *Hydrology and Earth System Sciences*, *15*(8), 2437–2457. https://doi.org/10.5194/hess-15-2437-2011

Norton, A. J., Rayner, P. J., Koffi, E. N., & Scholze, M. (2018). Assimilating solar-induced chlorophyll fluorescence into the terrestrial biosphere model BETHY-SCOPE v1.0: Model description and information content. *Geoscientific Model Development*, *11*(4), 1517–1536. https://doi.org/10.5194/gmd-11-1517-2018

Norton, A. J., Rayner, P. J., Koffi, E. N., Scholze, M., Silver, J. D., & Wang, Y.-P. (2019). Estimating global gross primary productivity using chlorophyll fluorescence and a data assimilation system with the BETHY-SCOPE model. *Biogeosciences*, *16*(15), 3069–3093. https://doi.org/10.5194/bg-16-3069-2019

Novick, K. A., Ficklin, D. L., Baldocchi, D., Davis, K. J., Ghezzehei, T. A., Konings, A. G., et al. (2022). Confronting the water potential information gap. *Nature Geoscience*, *15*(3), 158–164. https://doi.org/10.1038/s41561-022-00909-2

Oberpriller, J., Cameron, D. R., Dietze, M. C., & Hartig, F. (2021). Towards robust statistical inference for complex computer models. *Ecology Letters*, *24*(6), 1251–1261. https://doi.org/10.1111/ele.13728

Oberpriller, J., Herschlein, C., Anthoni, P., Arneth, A., Krause, A., Rammig, A., et al. (2022). Climate and parameter sensitivity and induced uncertainties in carbon stock projections for European forests (using LPJ-GUESS 4.0). *Geoscientific Model Development*, *15*(16), 6495–6519. https://doi.org/10.5194/gmd-15-6495-2022

Olivera-Guerra, L.-E., Ottlé, C., Raoult, N., & Peylin, P. (2024). Assimilating ESA-CCI land surface temperature into the ORCHIDEE land surface model: Insights from a multi-site study across Europe. https://doi.org/10.5194/egusphere-2024-546

Owen, A. B. (2013). Monte Carlo theory, methods and examples (Vol. *16*, pp. 19–22). Retrieved from https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=owV8q1cAAAAJ:tOudhMTPpwUC

Parton, W. J., Schimel, D. S., Cole, C. V., & Ojima, D. S. (1987). Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal. Soil Science Society of America*, *51*(5), 1173–1179. https://doi.org/10.2136/sssaj1987.03615995005100050015x

Peatier, S., Sanderson, B. M., & Terray, L. (2023). On the spatial calibration of imperfect climate models. In *EGUsphere*. https://doi.org/10.5194/egusphere-2023-2269

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., et al. (2016). A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. *Geoscientific Model Development*, *9*(9), 3321–3346. https://doi.org/10.5194/gmd-9-3321-2016

Pianosi, F., Iwema, J., Rosolem, R., & Wagener, T. (2017). A multimethod global sensitivity analysis approach to support the calibration and evaluation of land surface models. In *Sensitivity analysis in Earth observation modelling* (pp. 125–144). Elsevier. https://doi.org/10.1016/b978-0-12-803011-0.00007-0

Pinnington, E., Amezcua, J., Cooper, E., Dadson, S., Ellis, R., Peng, J., et al. (2021). Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data. *Hydrology and Earth System Sciences*, *25*(3), 1617–1641. https://doi.org/10.5194/hess-25-1617-2021

Pinnington, E., Quaife, T., & Black, E. (2018). Impact of remotely sensed soil moisture and precipitation on soil moisture prediction in a data assimilation system with the JULES land surface model. *Hydrology and Earth System Sciences*, *22*(4), 2575–2588. https://doi.org/10.5194/hess-22-2575-2018

Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., & Scoby, D. (2020). The land variational ensemble data assimilation framework: LAVENDAR v1.0.0. *Geoscientific Model Development*, *13*(1), 55–69. https://doi.org/10.5194/gmd-13-55-2020

Pinnington, E. M., Casella, E., Dance, S. L., Lawless, A. S., Morison, J. I. L., Nichols, N. K., et al. (2017). Understanding the effect of disturbance from selective felling on the carbon dynamics of a managed woodland by combining observations with model predictions. *Journal of Geophysical Research: Biogeosciences*, *122*(4), 886–902. https://doi.org/10.1002/2017JG003760

Pinty, B., Lavergne, T., Voßbeck, M., Kaminski, T., Aussedat, O., Giering, R., et al. (2007). Retrieving surface parameters for climate models from Moderate Resolution Imaging Spectroradiometer (MODIS)-Multiangle Imaging Spectroradiometer (MISR) albedo products. *Journal of Geophysical Research*, *112*(D10), D10116. https://doi.org/10.1029/2006JD008105

Post, H., Vrugt, J. A., Fox, A., Vereecken, H., & Hendricks Franssen, H.-J. (2017). Estimation of Community Land Model parameters for an improved assessment of net carbon fluxes at European sites: Estimation of CLM parameters. *Journal of Geophysical Research: Biogeosciences*, *122*(3), 661–689. https://doi.org/10.1002/2015jg003297

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., et al. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/j.rse.2020.112165

Prihodko, L., Denning, A. S., Hanan, N. P., Baker, I., & Davis, K. (2008). Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. *Agricultural and Forest Meteorology*, *148*(2), 268–287. https://doi.org/10.1016/j.agrformet.2007.08.006

Quaife, T., Lewis, P., Dekauwe, M., Williams, M., Law, B., Disney, M., & Bowyer, P. (2008). Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment*, *112*(4), 1347–1364. https://doi.org/10.1016/j.rse.2007.05.020

Quegan, S., Le Toan, T., Chave, J., Dall, J., Exbrayat, J.-F., Minh, D. H. T., 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

Raczka, B., Dietze, M. C., Serbin, S. P., & Davis, K. J. (2018). What limits predictive certainty of long-term carbon uptake? *Journal of Geophysical Research: Biogeosciences*, *123*(12), 3570–3588. https://doi.org/10.1029/2018jg004504

Raczka, B., Hoar, T. M., Duarte, H. F., Fox, A. M., Anderson, J. L., Bowling, D. R., & Lin, J. C. (2021). Improving CLM5.0 biomass and carbon exchange simulations across the Western U.S. using a data assimilation system. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2020MS002421. https://doi.org/10.1029/2020MS002421

Raiho, A. M., Nicklen, E. F., Foster, A. C., Roland, C. A., & Hooten, M. B. (2021). Bridging implementation gaps to connect large ecological datasets and complex models. *Ecology and Evolution*, *11*(24), 18271–18287. https://doi.org/10.1002/ece3.8420

Raiho, A. M., Paciorek, C. J., Dawson, A., Jackson, S. T., Mladenoff, D. J., Williams, J. W., & McLachlan, J. S. (2022). 8000-year doubling of Midwestern forest biomass driven by population-and biome-scale processes. *Science*, *376*(6600), 1491–1495. https://doi.org/10.1126/science.abk3126

Raoult, N., Beylat, S., Salter, J. M., Hourdin, F., Bastrikov, V., Ottlé, C., & Peylin, P. (2024). Exploring the potential of history matching for land surface model calibration. *Geoscientific Model Development*, *17*(15), 5779–5801. https://doi.org/10.5194/gmd-17-5779-2024

Raoult, N., Edouard-Rambaut, L.-A., Vuichard, N., Bastrikov, V., Lansø, A. S., Guenet, B., & Peylin, P. (2024). Using Free Air CO_2 Enrichment data to constrain land surface model projections of the terrestrial carbon cycle. *Biogeosciences*, *21*(4), 1017–1036. https://doi.org/10.5194/bg-21-1017-2024

Raoult, N., Ottlé, C., Peylin, P., Bastrikov, V., & Maugis, P. (2021). Evaluating and optimizing surface soil moisture drydowns in the ORCHIDEE land surface model at in situ locations. *Journal of Hydrometeorology*, *22*(4), 1025–1043. https://doi.org/10.1175/jhm-d-20-0115.1

Raoult, N. M., Jupp, T. E., Cox, P. M., & Luke, C. M. (2016). Land-surface parameter optimisation using data assimilation techniques: The adJULES system V1.0. *Geoscientific Model Development*, *9*(8), 2833–2852. https://doi.org/10.5194/gmd-9-2833-2016

Rayner, P. J. (2010). The current state of carbon-cycle data assimilation. *Current Opinion in Environmental Sustainability*, *2*(4), 289–296. https://doi.org/10.1016/j.cosust.2010.05.005

Rayner, P. J., Koffi, E., Scholze, M., Kaminski, T., & Dufresne, J.-L. (2011). Constraining predictions of the carbon cycle using data. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *369*(1943), 1955–1966. https://doi.org/10.1098/rsta.2010.0378

Rayner, P. J., Michalak, A. M., & Chevallier, F. (2019). Fundamentals of data assimilation applied to biogeochemistry. *Atmospheric Chemistry and Physics*, *19*(22), 13911–13932. https://doi.org/10.5194/acp-19-13911-2019

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., & Widmann, H. (2005). Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochemical Cycles*, *19*(2), GB2026. https://doi.org/10.1029/2004gb002254

Reichle, R. H., & Koster, R. D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, *31*(19), L19501. https://doi.org/10.1029/2004gl020938

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Reick, C. H., Gayler, V., Goll, D., Hagemann, S., & Heidkamp, M. (2021). The land component of the MPI Earth System Model: Documentation of version 3.2. https://pure.mpg.de/rest/items/item_3279802/component/file_3316522/content

Ricciuto, D. M., Butler, M. P., Davis, K. J., Cook, B. D., Bakwin, P. S., Andrews, A., & Teclaw, R. M. (2008). Causes of interannual variability in ecosystem–atmosphere CO_2 exchange in a northern Wisconsin forest using a Bayesian model calibration. *Agricultural and Forest Meteorology*, *148*(2), 309–327. https://doi.org/10.1016/j.agrformet.2007.08.007

Ricciuto, D. M., King, A. W., Dragoni, D., & Post, W. M. (2011). Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: Effects of constraining variables and data record length. *Journal of Geophysical Research*, *116*(G1), G01033. https://doi.org/10.1029/2010jg001400

Richardson, A. D., Williams, M., Hollinger, D. Y., Moore, D. J., Dail, D. B., Davidson, E. A., et al. (2010). Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. *Oecologia*, *164*(1), 25–40. https://doi.org/10.1007/s00442-010-1628-y

Rödenbeck, C., Houweling, S., Gloor, M., & Heimann, M. (2003). CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmospheric Chemistry and Physics*, *3*(6), 1919–1964. https://doi.org/10.5194/acp-3-1919-2003

Rodríguez-Fernández, N., de Rosnay, P., Albergel, C., Richaume, P., Aires, F., Prigent, C., & Kerr, Y. (2019). SMOS neural network Soil Moisture data assimilation in a Land surface model and atmospheric impact. *Remote Sensing*, *11*(11), 1334. https://doi.org/10.3390/rs11111334

Rowland, L., Hill, T. C., Stahl, C., Siebicke, L., Burban, B., Zaragoza-Castells, J., et al. (2014). Evidence for strong seasonality in the carbon storage and carbon use efficiency of an Amazonian forest. *Global Change Biology*, *20*(3), 979–991. https://doi.org/10.1111/gcb.12375

Running, S., Mu, Q., & Zhao, M. (2021). MODIS/Terra gross primary productivity 8-day L4 global 500m SIN grid V061 [Dataset]. *NASA EOSDIS Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MOD17A2H.061

Sacks, W. J., Schimel, D. S., Monson, R. K., & Braswell, B. H. (2006). Model-data synthesis of diurnal and seasonal CO_2 fluxes at Niwot Ridge, Colorado: Model-data synthesis of CO_2 fluxes. *Global Change Biology*, *12*(2), 240–259. https://doi.org/10.1111/j.1365-2486.2005.01059.x

Salmon, E., Jégou, F., Guenet, B., Jourdain, L., Qiu, C., Bastrikov, V., et al. (2022). Assessing methane emissions for northern peatlands in ORCHIDEE-PEAT revision 7020. *Geoscientific Model Development*, *15*(7), 2813–2838. https://doi.org/10.5194/gmd-15-2813-2022

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global sensitivity analysis: The primer*. Wiley & Sons. Retrieved from https://openlibrary.org/books/OL33376474M/Global_Sensitivity_Analysis

Sanderson, B. (2020). The role of prior assumptions in carbon budget calculations. *Earth System Dynamics*, *11*(2), 563–577. https://doi.org/10.5194/esd-11-563-2020

Sanderson, B. M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D. J., et al. (2008). Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. *Journal of Climate*, *21*(11), 2384–2400. https://doi.org/10.1175/2008JCLI1869.1

Santaren, D., Peylin, P., Viovy, N., & Ciais, P. (2007). Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France. *Global Biogeochemical Cycles*, *21*(2), GB2013. https://doi.org/10.1029/2006GB002834

Sawada, Y. (2020). Machine learning accelerates parameter optimization and uncertainty assessment of a land surface model. *Journal of Geophysical Research*, *125*(20), e2020JD032688. https://doi.org/10.1029/2020jd032688

Scholze, M., Buchwitz, M., Dorigo, W., Guanter, L., & Quegan, S. (2017). Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems. *Biogeosciences*, *14*(14), 3401–3429. https://doi.org/10.5194/bg-14-3401-2017

Scholze, M., Kaminski, T., Knorr, W., Blessing, S., Vossbeck, M., Grant, J. P., & Scipal, K. (2016). Simultaneous assimilation of SMOS soil moisture and atmospheric CO2 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

Scholze, M., Kaminski, T., Knorr, W., Voßbeck, M., Wu, M., Ferrazzoli, P., et al. (2019). Mean European carbon sink over 2010–2015 estimated by simultaneous assimilation of atmospheric CO2, soil moisture, and vegetation optical depth. *Geophysical Research Letters*, *46*(23), 13796–13803. https://doi.org/10.1029/2019gl085725

Scholze, M., Kaminski, T., Rayner, P., Knorr, W., & Giering, R. (2007). Propagating uncertainty through prognostic carbon cycle data assimilation system simulations. *Journal of Geophysical Research*, *112*(D17), D17305. https://doi.org/10.1029/2007JD008642

Schürmann, G. J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., et al. (2016). Constraining a land-surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System V1.0. *Geoscientific Model Development*, *9*(9), 2999–3026. https://doi.org/10.5194/gmd-9-2999-2016

Scrucca, L. (2013). GA: A package for genetic algorithms inR. *Journal of Statistical Software*, *53*(4), 1–37. https://doi.org/10.18637/jss.v053.i04

Seiler, C., Melton, J. R., Arora, V. K., Sitch, S., Friedlingstein, P., Anthoni, P., et al. (2022). Are terrestrial biosphere models fit for simulating the global land carbon sink? *Journal of Advances in Modeling Earth Systems*, *14*(5), e2021MS002946. https://doi.org/10.1029/2021ms002946

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., et al. (1996). A revised land surface parameterization (SiB2) for atmosphere–biosphere interaction. https://doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., et al. (2022). Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a Deep Neural Network. *Remote Sensing of Environment*, *279*, 113116. https://doi.org/10.1016/j.rse.2022.113116

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, *4*(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., & Yu, X. (2014). Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: A synthetic experiment. *Water Resources Research*, *50*(1), 706–724. https://doi.org/10.1002/2013WR014070

Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., & Yu, X. (2015). Parameter estimation of a physically based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment. *Advances in Water Resources*, *83*, 421–427. https://doi.org/10.1016/j.advwatres.2015.06.009

Shi, Z., Allison, S. D., He, Y., Levine, P. A., Hoyt, A. M., Beem-Miller, J., et al. (2020). The age distribution of global soil carbon inferred from radiocarbon measurements. *Nature Geoscience*, *13*(8), 555–559. https://doi.org/10.1038/s41561-020-0596-z

Shi, Z., Hoffman, F. M., Xu, M., Mishra, U., Allison, S. D., Zhou, J., & Randerson, J. T. (2024). Global-scale convergence obscures inconsistencies in soil carbon change predicted by Earth system models. *AGU Advances*, *5*(2), e2023AV001068. https://doi.org/10.1029/2023av001068

Shiklomanov, A. N., Cowdery, E. M., Bahn, M., Byun, C., Jansen, S., Kramer, K., et al. (2018). Does the leaf economic spectrum hold within plant functional types? A Bayesian multivariate trait meta-analysis. In *bioRxiv*. https://doi.org/10.1101/475038

Shiklomanov, A. N., Dietze, M. C., Fer, I., Viskari, T., & Serbin, S. P. (2021). Cutting out the middleman: Calibrating and validating a dynamic vegetation model (ED2-PROSPECT5) using remotely sensed surface reflectance. *Geoscientific Model Development*, *14*(5), 2603–2633. https://doi.org/10.5194/gmd-14-2603-2021

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *An International Journal on Information Fusion*, *81*, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Sitch, S., O'Sullivan, M., Robertson, E., Friedlingstein, P., Albergel, C., Anthoni, P., et al. (2024). Trends and drivers of terrestrial sources and sinks of carbon dioxide: An overview of the TRENDY project. *Global Biogeochemical Cycles*, *38*(7), e2024GB008102. https://doi.org/10.1029/2024gb008102

Smallman, T. L., Milodowski, D. T., Neto, E. S., Koren, G., Ometto, J., & Williams, M. (2021). Parameter uncertainty dominates C-cycle forecast errors over most of Brazil for the 21st century. *Earth System Dynamics*, *12*(4), 1191–1237. https://doi.org/10.5194/esd-12-1191-2021

Smith, B. (2007). LPJ-GUESS—An ecosystem modelling framework. https://www.baltex-research.eu/baltic2009/downloads/Literature/Specific/Smith_guess_software.pdf

Smith, C., Cummins, D. P., Fredriksen, H.-B., Nicholls, Z., Meinshausen, M., Allen, M., et al. (2024). fair-calibrate v1.4.1: Calibration, constraining and validation of the FaIR simple climate model for reliable future climate projections. In *EGUsphere*. https://doi.org/10.5194/egusphere-2024-708

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1–3), 271–280. https://doi.org/10.1016/s0378-4754(00)00270-6

Song, H., Triguero, I., & Özcan, E. (2019). A review on the self and dual interactions between machine learning and optimisation. *Progress in Artificial Intelligence*, *8*(2), 143–165. https://doi.org/10.1007/s13748-019-00185-z

Speich, M., Dormann, C. F., & Hartig, F. (2021). Sequential Monte-Carlo algorithms for Bayesian model calibration—A review and method comparison. *Ecological Modelling*, *455*, 109608. https://doi.org/10.1016/j.ecolmodel.2021.109608

Stewart, L. M., Dance, S. L., & Nichols, N. K. (2008). Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, *56*(8), 1521–1527. https://doi.org/10.1002/fld.1636

Stöckli, R., Rutishauser, T., Dragoni, D., O'Keefe, J., Thornton, P. E., Jolly, M., et al. (2008). Remote sensing data assimilation for a prognostic phenology model: Data assimilation and phenology modeling. *Journal of Geophysical Research*, *113*(G4), G01021. https://doi.org/10.1029/2008jg000781

Sun, Y., Goll, D. S., Huang, Y., Ciais, P., Wang, Y.-P., Bastrikov, V., & Wang, Y. (2023). Machine learning for accelerating process-based computation of land biogeochemical cycles. *Global Change Biology*, *29*(11), 3221–3234. https://doi.org/10.1111/gcb.16623

Talagrand, O., & Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory: Variational assimilation. I: Theory. *Quarterly Journal of the Royal Meteorological Society. Royal Meteorological Society (Great Britain)*, *113*(478), 1311–1328. https://doi.org/10.1002/qj.49711347812

Tao, F., Houlton, B. Z., Huang, Y., Wang, Y.-P., Manzoni, S., Ahrens, B., et al. (2024). Convergence in simulating global soil organic carbon by structurally different models after data assimilation. *Global Change Biology*, *30*(5), e17297. https://doi.org/10.1111/gcb.17297

Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., et al. (2020). Deep learning optimizes data-driven representation of soil organic carbon in Earth system model over the conterminous United States. *Frontiers in Big Data*, *3*, 17. https://doi.org/10.3389/fdata.2020.00017

Tarantola, A. (1987). Inversion of travel times and seismic waveforms. In *Seismic tomography* (pp. 135–157). Springer. https://doi.org/10.1007/978-94-009-3899-1_6

Tarantola, A. (2005). Back matter. In *Inverse problem theory and methods for model parameter estimation* (pp. 317–342). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717921.bm

Thomas, R. Q., Brooks, E. B., Jersild, A. L., Ward, E. J., Wynne, R. H., Albaugh, T. J., et al. (2017). Leveraging 35 years of *Pinus taeda* research in the southeastern US to constrain forest carbon cycle predictions: Regional data assimilation using ecosystem experiments. *Biogeosciences*, *14*(14), 3525–3547. https://doi.org/10.5194/bg-14-3525-2017

Thum, T., MacBean, N., Peylin, P., Bacour, C., Santaren, D., Longdoz, B., et al. (2017). The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites. *Agricultural and Forest Meteorology*, *234–235*, 48–65. https://doi.org/10.1016/j.agrformet.2016.12.004

Tian, X., Minunno, F., Cao, T., Peltoniemi, M., Kalliokoski, T., & Mäkelä, A. (2020). Extending the range of applicability of the semi-empirical ecosystem flux model PRELES for varying forest types and climate. *Global Change Biology*, *26*(5), 2923–2943. https://doi.org/10.1111/gcb.14992

Torres-Rojas, L., Vergopolan, N., Herman, J. D., & Chaney, N. W. (2022). Towards an optimal representation of sub-grid heterogeneity in land surface models. *Water Resources Research*, *58*(12), e2022WR032233. https://doi.org/10.1029/2022wr032233

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., et al. (2007). OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *Journal of Geophysical Research*, *112*(G2), G02027. https://doi.org/10.1029/2006JG000367

Trugman, A. T., Anderegg, L. D. L., Shaw, J. D., & Anderegg, W. R. L. (2020). Trait velocities reveal that mortality has driven widespread coordinated shifts in forest hydraulic trait composition. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8532–8538. https://doi.org/10.1073/pnas.1917521117

Ustin, S. L., & Middleton, E. M. (2021). Current and near-term advances in Earth observation for ecological applications. *Ecological Processes*, *10*(1), 1. https://doi.org/10.1186/s13717-020-00255-4

van Oijen, M. (2017). Bayesian methods for quantifying and reducing uncertainty and error in forest models. *Current Forestry Reports*, *3*(4), 269–280. https://doi.org/10.1007/s40725-017-0069-9

Varney, R. M., Friedlingstein, P., Chadburn, S. E., Burke, E. J., & Cox, P. M. (2024). Soil carbon-concentration and carbon-climate feedbacks in CMIP6 Earth system models. *Biogeosciences*, *21*(11), 2759–2776. https://doi.org/10.5194/bg-21-2759-2024

Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., & Ciais, P. (2011). Seasonal patterns of CO_2 fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. *Journal of Geophysical Research*, *116*(G2), G02018. https://doi.org/10.1029/2010jg001544

Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., & Wood, E. F. (2020). Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. *Remote Sensing of Environment*, *242*, 111740. https://doi.org/10.1016/j.rse.2020.111740

Vergopolan, N., Chaney, N. W., Pan, M., Sheffield, J., Beck, H. E., Ferguson, C. R., et al. (2021). SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US. *Scientific Data*, *8*(1), 264. https://doi.org/10.1038/s41597-021-01050-2

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software: With Environment Data News*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating markov chain Monte Carlo Simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Stimulation*, *10*(3), 273. https://doi.org/10.1515/IJNSNS.2009.10.3.273

Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., et al. (2019). Accounting for carbon and nitrogen interactions in the global terrestrial ecosystem model ORCHIDEE (trunk version, rev 4999): Multi-scale evaluation of gross primary production. *Geoscientific Model Development*, *12*(11), 4751–4779. https://doi.org/10.5194/gmd-12-4751-2019

Waller, J. A., Dance, S. L., & Nichols, N. K. (2016). Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quarterly Journal of the Royal Meteorological Society. Royal Meteorological Society (Great Britain)*, *142*(694), 418–431. https://doi.org/10.1002/qj.2661

Wang, H., Huo, X., Duan, Q., Liu, R., & Luo, S. (2023). Uncertainty quantification for the Noah-MP land surface model: A case study in a grassland and sandy soil region. *Journal of Geophysical Research: Atmospheres*, *128*(20), e2023JD038556. https://doi.org/10.1029/2023jd038556

Wang, J., Jiang, F., Wang, H., Qiu, B., Wu, M., He, W., et al. (2021). Constraining global terrestrial gross primary productivity in a global carbon assimilation system with OCO-2 chlorophyll fluorescence data. *Agricultural and Forest Meteorology*, *304–305*, 108424. https://doi.org/10.1016/j.agrformet.2021.108424

Wang, N., Zhang, D., Chang, H., & Li, H. (2020). Deep learning of subsurface flow via theory-guided neural network. *Journal of Hydrology*, *584*, 124700. https://doi.org/10.1016/j.jhydrol.2020.124700

Wang, Y.-P., Leuning, R., Cleugh, H. A., & Coppin, P. A. (2001). Parameter estimation in surface exchange models using nonlinear inversion: How many parameters can we estimate and which measurements are most useful? Nonlinear parameter estimation. *Global Change Biology*, *7*(5), 495–510. https://doi.org/10.1046/j.1365-2486.2001.00434.x

Watson-Parris, D. (2021). Machine learning for weather and climate are worlds apart. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *379*(2194), 20200098. https://doi.org/10.1098/rsta.2020.0098

Watson-Parris, D., Williams, A., Deaconu, L., & Stier, P. (2021). Model calibration using ESEm v1.1.0—An open, scalable Earth system emulator. *Geoscientific Model Development*, *14*(12), 7659–7672. https://doi.org/10.5194/gmd-14-7659-2021

Weng, E., & Luo, Y. (2011). Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. *Ecological Applications: A Publication of the Ecological Society of America*, *21*(5), 1490–1505. https://doi.org/10.1890/09-1394.1

Weng, E., Luo, Y., Gao, C., & Oren, R. (2011). Uncertainty analysis of forest carbon sink forecast with varying measurement errors: A data assimilation approach. *Journal of Plant Ecology*, *4*(3), 178–191. https://doi.org/10.1093/jpe/rtr018

Wesselkamp, M., Chantry, M., Pinnington, E., Choulga, M., Boussetta, S., Kalweit, M., et al. (2024). Advances in land surface model-based forecasting: A comparative study of LSTM, gradient boosting, and feedforward neural network models as prognostic state emulators. *arXiv [physics.ao-ph]. arXiv*. http://arxiv.org/abs/2407.16463

Whelan, M. E., Lennartz, S. T., Gimeno, T. E., Wehr, R., Wohlfahrt, G., Wang, Y., et al. (2018). Reviews and syntheses: Carbonyl sulfide as a multi-scale tracer for carbon and water cycles. *Biogeosciences*, *15*(12), 3625–3657. https://doi.org/10.5194/bg-15-3625-2018

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., & Kurpius, M. R. (2005). An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, *11*(1), 89–105. https://doi.org/10.1111/j.1365-2486.2004.00891.x

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., & Yamazaki, K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, *41*(7), 1703–1729. https://doi.org/10.1007/s00382-013-1896-4

Wu, J.-L., Levine, M. E., Schneider, T., & Stuart, A. (2023). Learning about structural errors in models of complex dynamical systems. *arXiv [physics.comp-ph]. arXiv*. http://arxiv.org/abs/2401.00035

Wu, M., Jiang, F., Scholze, M., Chen, D., Ju, W., Wang, S., et al. (2024). Regional responses of vegetation productivity to the two phases of ENSO. *Geophysical Research Letters*, *51*(8), e2024GL108176. https://doi.org/10.1029/2024gl108176

Wu, M., Scholze, M., Kaminski, T., Voßbeck, M., & Tagesson, T. (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

Wu, M., Scholze, M., Voßbeck, M., Kaminski, T., & Hoffmann, G. (2018). Simultaneous assimilation of remotely sensed Soil Moisture and FAPAR for improving terrestrial carbon fluxes at multiple sites using CCDAS. *Remote Sensing*, *11*(1), 27. https://doi.org/10.3390/rs11010027

Wutzler, T., & Carvalhais, N. (2014). Balancing multiple constraints in model-data integration: Weights and the parameter block approach. *Journal of Geophysical Research: Biogeosciences*, *119*(11), 2112–2129. https://doi.org/10.1002/2014jg002650

Xiao, J., Davis, K. J., Urban, N. M., & Keller, K. (2014). Uncertainty in model parameters and regional carbon fluxes: A model-data fusion approach. *Agricultural and Forest Meteorology*, *189–190*, 175–186. https://doi.org/10.1016/j.agrformet.2014.01.022

Xie, X., Meng, S., Liang, S., & Yao, Y. (2021). Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *Hydrology and Earth System Sciences*, *18*(10), 3923–3936. https://doi.org/10.1016/j.jhydrol.2021.127043

Xing, X., Wu, M., Zhang, W., Ju, W., Tagesson, T., He, W., et al. (2023). Modeling China's terrestrial ecosystem gross primary productivity with BEPS model: Parameter sensitivity analysis and model calibration. *Agricultural and Forest Meteorology*, *343*, 109789. https://doi.org/10.1016/j.agrformet.2023.109789

Xiong, M., Liu, P., Cheng, L., Deng, C., Gui, Z., Zhang, X., & Liu, Y. (2019). Identifying time-varying hydrological model parameters to improve simulation efficiency by the ensemble Kalman filter: A joint assimilation of streamflow and actual evapotranspiration. *Journal of Hydrology*, *568*, 758–768. https://doi.org/10.1016/j.jhydrol.2018.11.038

Xu, T., White, L., Hui, D., & Luo, Y. (2006). Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochemical Cycles*, *20*(2). https://doi.org/10.1029/2005GB002468

Yang, T., Sun, F., Gentine, P., Liu, W., Wang, H., Yin, J., et al. (2019). Evaluation and machine learning improvement of global hydrological model-based flood simulations. *Environmental Research Letters*, *14*(11), 114027. https://doi.org/10.1088/1748-9326/ab4d5e

Zaehle, S., Friedlingstein, P., & Friend, A. D. (2010). Terrestrial nitrogen feedbacks may accelerate future climate change: Carbon-nitrogen feedbacks affect climate. *Geophysical Research Letters*, *37*(1). https://doi.org/10.1029/2009gl041345

Zaehle, S., Friend, A. D., Friedlingstein, P., Dentener, F., Peylin, P., & Schulz, M. (2010). Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance: Nitrogen effects on global C cycling. *Global Biogeochemical Cycles*, *24*(1), GB1006. https://doi.org/10.1029/2009gb003522

Zhang, H., Hendricks Franssen, H.-J., Han, X., Vrugt, J. A., & Vereecken, H. (2017). State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. *Hydrology and Earth System Sciences*, *21*(9), 4927–4958. https://doi.org/10.5194/hess-21-4927-2017

Zhang, J., Chung, H. S.-H., & Lo, W.-L. (2007). Clustering-based adaptive crossover and mutation probabilities for genetic algorithms. *IEEE Transactions on Evolutionary Computation*, *11*(3), 326–335. https://doi.org/10.1109/TEVC.2006.880727

Zhao, H., & Kowalski, J. (2022). Bayesian active learning for parameter calibration of landslide run-out models. *Landslides*, *19*(8), 2033–2045. https://doi.org/10.1007/s10346-022-01857-z

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., et al. (2019). Physics-constrained machine learning of evapotranspiration. *Geophysical Research Letters*, *46*(24), 14496–14507. https://doi.org/10.1029/2019gl085291

Zhou, A., Hawkins, L., & Gentine, P. (2024). Proof-of-concept: Using ChatGPT to translate and modernize an Earth System model from fortran to Python/JAX. *arXiv [cs.DC]. arXiv*. http://arxiv.org/abs/2405.00018

Zhu, H., Wu, M., Jiang, F., Vossbeck, M., Kaminski, T., Xing, X., et al. (2024). Assimilation of Carbonyl Sulfide (COS) fluxes within the adjoint-based data assimilation system–Nanjing University carbon assimilation system (NUCAS v1.0). *Geoscientific Model Development*, *17*(16), 6337–6363. https://doi.org/10.5194/gmd-17-6337-2024

Zhu, H., Xing, X., Wu, M., Ju, W., & Jiang, F. (2024). Optimizing the terrestrial ecosystem gross primary productivity using carbonyl sulfide (COS) within a two-leaf modeling framework. *Biogeosciences*, *21*(16), 3735–3760. https://doi.org/10.5194/bg-21-3735-2024

Ziehn, T., Knorr, W., & Scholze, M. (2011). Investigating spatial differentiation of model parameters in a carbon cycle data assimilation system. *Global Biogeochemical Cycles*, *25*(2), GB2021. https://doi.org/10.1029/2010GB003886

Ziehn, T., Scholze, M., & Knorr, W. (2012). On the capability of Monte Carlo and adjoint inversion techniques to derive posterior parameter uncertainties in terrestrial ecosystem models: Comparison of inversion techniques. *Global Biogeochemical Cycles*, *26*(3). https://doi.org/10.1029/2011gb004185

Zobitz, J. M., Moore, D. J. P., Quaife, T., Braswell, B. H., Bergeson, A., Anthony, J. A., & Monson, R. K. (2014). Joint data assimilation of satellite reflectance and net ecosystem exchange data constrains ecosystem carbon fluxes at a high-elevation subalpine forest. *Agricultural and Forest Meteorology*, *195–196*, 73–88. https://doi.org/10.1016/j.agrformet.2014.04.011

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, *15*(3), 779–808. https://doi.org/10.5194/os-15-779-2019