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

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

[1]University of Exeter
[2]University of Reading
[3]Western University
[4]NASA Goddard Spaceflight Center
[5]Boston University
[6]CICERO Center for International Climate Research
[7]Finnish Meteorological Institute
[8]LSCE
[9]National Center for Atmospheric Research
[10]Columbia University
[11]Max Planck Institute for Biogeochemistry
[12]UK Centre for Ecology & Hydrology
[13]The Inversion Lab
[14]The University of Arizona
[15]ECMWF
[16]CICERO Centre for International Climate Research
[17]Lund University
[18]Queen's University
[19]University of Edinburgh
[20]Rice University
[21]Joint Research Centre
[22]Augsburg University

October 08, 2024

**Abstract**

Accurately predicting terrestrial ecosystem responses to climate change is crucial for addressing global challenges. This relies on mechanistic modelling 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 optimisation, focusing on carbon-water-vegetation inter-

actions, as well as discussing the technical challenges faced by the community. These challenges include identifying sensitive model parameters and their prior distributions, characterising 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 datasets, and including the spin-up and historical period in the assimilation window. We then 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 datasets, harness machine learning advances, and enhance the predictive capabilities of LSMs.

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

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

[1]Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK
[2]National Centre for Earth Observation, Department of Meteorology, University of Reading, Reading, UK
[3]Departments of Geography and Environment & Biology, Western University, Canada
[4]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[5]Science Systems and Applications, Inc., Lanham, MD, USA
[6]Computing and Data Sciences, Boston University, Boston, MA 02215 USA
[7]CICERO, Norway
[8]Finnish Meteorological Institute, Helsinki, Finland
[9]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Gif-sur-Yvette, France
[10]NSF National Center for Atmospheric Research, Boulder, CO, USA
[11]Earth and Environmental Engineering Department, Columbia University, New York, NY 10027, USA
[12]Max Planck Institute for Biogeochemistry, Hans Knöll Strasse 10, 07745 Jena, Germany;
[13]Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[14]ELLIS Unit Jena, Jena, Germany
[15]UK Centre for Ecology and Hydrology, Wallingford, UK
[16]Department of Earth & Environment, Boston University, Boston, MA 02215 USA
[17]Earth and Environmental Engineering Department, Columbia University, New York, NY 10027, USA
[18]The Inversion Lab, Hamburg, Germany
[19]NSF National Center for Atmospheric Research, Boulder, CO, USA
[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, 85721, USA
[23]European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK
[24]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden
[25]School of Environmental Studies, Queen's University, Kingston, ON, K7L3N6, Canada
[26]School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh, EH9 3FF
[27]Earth Environment and Planetary Sciences, Rice University, Houston, TX 77006, USA
[28]European Commission Joint Research Center, Ispra, 21027 VA, Italy
[29]Department of Mathematics, Statistics, and Computer Science, Augsburg University, Minneapolis, MN 55454 USA
*now at European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK

## Key points (max 140 characters each)

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

## Abstract (max 250 words)

Accurately predicting terrestrial ecosystem responses to climate change is crucial for addressing global challenges. This relies on mechanistic modelling 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 optimisation, 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, characterising 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 datasets, and including the spin-up and historical period in the assimilation window. We then 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 datasets, harness machine learning advances, and enhance the predictive capabilities of LSMs.

## Plain language summary (max 200 words)

Improving the accuracy of land surface models (LSMs) is crucial for reducing uncertainties in climate change projections. Parameter data assimilation, 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 optimisation. Advances in computational power, novel datasets, and machine learning (ML) offer promising solutions to improve these models. ML can streamline the data assimilation process, handling large datasets 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 machine learning 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 Data Assimilation Working Group and the International Land Model Forum, is essential for accelerating progress, facilitating knowledge exchange, and developing standardised methods for more accurate climate modelling.

# 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. This strongly relies on our ability to accurately model the underlying processes at the global scale. Such global-scale, mechanistic or process-based models of the terrestrial biosphere, often embedded in Earth system models (wherein they are called Land Surface Models – LSMs; Blyth et al., 2021), 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 (including emissions from land use change), 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 also indispensable tools in assessing climate change mitigation strategies, for example, to assess how effective nature-based solutions such as reforestation will be in curbing rising $CO_2$ emissions.

Representing all the requisite processes corresponding to interacting vegetation, biogeochemistry, water and energy cycles mechanistically (and accurately) in LSMs over a wide range of timescales, from sub-daily flux exchanges with the atmosphere to decadal-century timescales representative of changes in biomass and soil carbon pools required for carbon-climate feedbacks, is critical for robust and reliable projections (Watson-Parris, 2021). However, LSMs are highly complex and subject to large uncertainties, both in terms of missing processes, inadequate representation of processes, and poorly constrained parameters. Furthermore, when trying to address model structural uncertainty, implementing new processes tends to introduce additional parameters and, therefore, more parameter uncertainty. As a result, LSMs often diverge significantly in their representation of many terrestrial processes (Gier et al., 2024; Green et al., 2024; Varney et al., 2024). Consequently, their predictions of important ecosystem responses under future climate change scenarios often vary widely. For example, LSMs disagree on the magnitude of the land carbon sink (Koven et al., 2022; Shi et al., 2024), and the potential constraints on $CO_2$ fertilisation due to water (Green et al., 2019) and nutrient (Davies-Barnard et al., 2022) limitations.

Parametric uncertainty is one of the largest sources of uncertainty in all types of land models ( simple, intermediate and full complexity models), particularly for 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). Indeed, it has been shown for one LSM that even perturbing a single carbon flux related parameter within its range of uncertainty can result in a projection spread in atmospheric $CO_2$ by 2100 that is larger than running the model under different emissions scenarios (Booth et al., 2012). We urgently need to reduce this uncertainty to ensure we can utilise the full potential of LSMs—parameter optimisation is one way to achieve this.

Many processes in LSMs (as well as processes in ecosystem models, see Table A1 for all process-based models mentioned in the paper) are controlled by parameters that represent the

functioning of individual elements of the system. While some of these parameters can be directly observed (e.g. photosynthetic capacity, wood density, rooting depth, 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. Even those parameters that can be directly measured can often only be observed at scales that differ from the grid resolution of most global-scale LSM simulations (typically 0.5 degrees or greater). As a result, LSM predictions – particularly for vegetation and carbon cycle related processes – can be highly sensitive to parameter choices (in addition to model parameterisation or structural uncertainties) (Booth et al., 2012; Buotte et al., 2021; Exbrayat et al., 2014; Fisher et al., 2019; Oberpriller et al., 2022; Smallman et al., 2021; Zaehle, Friedlingstein, et al., 2010).

Historically, LSM parameters have simply been manually tuned (adjusted by hand to produce more realistic model behaviour or to better fit a given important model variable to a given dataset). Manual tuning of LSM parameters was often the only option given the required rapid pace of LSM development, the lack of available data at the correct scales for LSM parameter optimisation, or the computational demand of optimising the large number of parameters (typically >200) in LSMs with many complex, interacting processes. However, in the last two decades, the hurdles associated with performing rigorous LSM parameter optimisation (as opposed to tuning) have diminished to the point that it has become feasible: many datasets have become available at LSM-relevant scales, and the computational cost of running LSMs has decreased (although it remains a challenge – see Sect. 3). LSM groups have therefore started to optimise a selection of parameters using statistically robust data assimilation (DA) methods.

DA methods are powerful as they allow observational data to be combined with numerical methods to optimise estimates of chosen variables at the time of observations, either to update the state (state estimation) or to optimise internal parameters (parameter estimation) while accounting for uncertainties in both the model and the data (Rayner et al., 2019). However, the distinct requirements of LSMs compared to the atmospheric and ocean components of ESMs result in subtle but important differences in how DA techniques are applied. The atmospheric and ocean components of ESMs rely on fluid dynamic models, where the underlying fundamental laws are relatively well understood, even if complex to simulate, and many of the model parameters are known physical quantities that can be observed. Therefore, DA activities using atmospheric or ocean components of ESMs have thus far been heavily focused on numerical weather forecasting (NWP) 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, parametric and structural uncertainties dominate their spread (Bonan & Doney, 2018; Draper, 2021; Luo et al., 2015). LSM parameters are often linked to biological processes and organismal traits and are dependent on plant functional type (PFT). Therefore, these parameters have a wide range of possible values where they have been measured (in addition to a lack of data on parameters for some PFTs and the role of "effective" parameters as discussed above). Characterising and

simplifying the diversity of life into relatively few parameters is thus a challenge faced in LSM development that is less of an issue for atmospheric and ocean modeling.

Early efforts in global model calibration in the 1990s and 2000s focused on optimising vegetation and carbon cycle parameters of simplified or intermediate complexity land, carbon, or ecosystem models. These studies, such as Knorr & Heimann's (1995) work optimising parameters of the Simple Diagnostic Biosphere Model (SDBM) using site $CO_2$ measurements, laid the groundwork for DA-focused land model parameter optimisation. Knorr & Heimann's (1995) study was followed by further studies constrain carbon flux related processes in simple and intermediate complexity ecosystem models like BETHY (Rayner et al., 2005; Scholze et al., 2007), using frameworks referred to carbon cycle data assimilation systems (CCDASs). Parallel to this, there was significant progress in using local eddy-covariance flux tower measurements to optimise parameters related to photosynthesis, respiration, and energy flux in ecosystem models at the site level (e.g., Moore et al., 2008; Sacks et al., 2006; Y.-P. Wang et al., 2001; Williams et al., 2005). Two key intercomparison projects, OptIC and REFLEX, played a pivotal role in assessing various data assimilation techniques for parameter estimation in simple and intermediate complexity land, carbon cycle or ecosystem models (Fox et al., 2009; Trudinger et al., 2007).

Parameter optimisation of computational expensive land models using DA started in the late 2000s (Medvigy et al., 2009; Rayner, 2010; Santaren et al., 2007). These studies used similar data (*in situ* fluxes and biomass) and similar experimental configurations (site scale optimisations) as past studies with simple and intermediate complexity models but often with different DA methods due to the increase in computational expense of running much more complex models (Sect. 2). Building on the formative DA work with the SDBM (Kaminski et al., 2002) and BETHY models (Rayner et al., 2005), other LSM groups also started using global networks of *in situ* atmospheric $CO_2$ mole concentration data for constraining regional to global scale surface net $CO_2$ exchange (Kaminski et al., 2013; Peylin et al., 2016; Schürmann et al., 2016). Testing of DA configuration at site scale (data type, sampling interval, record length, and combinations of data - e.g., carbon fluxes and stocks or carbon fluxes) continued with all types of land models (Bastrikov et al., 2018; Bloom et al., 2016; Bloom & Williams, 2015; Braswell et al., 2005; Dietze et al., 2014; Keenan et al., 2013; Medvigy et al., 2009; Moore et al., 2008; Ricciuto et al., 2008, 2011; Santaren et al., 2014; Thum et al., 2017; Weng et al., 2011; Weng & Luo, 2011; Wutzler & Carvalhais, 2014; Xu et al., 2006). One example was the emergence of "multi-site" experiments – parameter estimation studies in which data from multiple sites (often grouped by PFT) were included simultaneously in the assimilation, with the retrieved parameters then compared to those from assimilations with only individual site data (see Sect. 3.4 for further discussion). These were initially performed against data from the global FLUXNET network for a range of intermediate and full complexity LSMs, including many LSMs used within ESMs (e.g., Carvalhais et al., 2008, 2010; Groenendijk et al., 2011; Kato et al., 2013; Knorr et al., 2010; Wu et al., 2018; Xiao et al., 2014, ORCHIDEE: Kuppel et al., 2012, 2014; JULES: Alton, 2013; Raoult et al., 2016; Noah: Chaney et al., 2016; CLM: Post et al., 2017). With the advent of satellite products, remote sensing indicators of vegetation dynamics (phenology and photosynthetic uptake) began to be employed to constrain model parameters at various spatial scales, including reflectance (Shiklomanov et al., 2021); vegetation indices

(Migliavacca et al., 2009; NDVI – MacBean et al., 2015), FAPAR (Bacour et al., 2015; Forkel et al., 2014, 2019; Kaminski et al., 2012; Knorr et al., 2010; Stöckli et al., 2008; Zobitz et al., 2014), solar-induced fluorescence (SIF; (Bacour et al., 2019; Forkel et al., 2019; Knorr et al., 2024; MacBean et al., 2018; Norton et al., 2018, 2019; J. Wang et al., 2021), aboveground biomass and burned area (Forkel et al., 2019). Over the past decade, parameter estimation has advanced to constrain the terrestrial carbon, water, and energy cycles simultaneously, driven by new remote sensing data on total column-integrated $CO_2$ fluxes ($XCO_2$), satellite-derived vegetation optical depth, soil moisture, snow cover, and river flow measurements, which have been successfully integrated, for example, into BETHY (Scholze et al., 2016), the new community D&B model developed by the European Space Agency (ESA)'s Carbon Cluster (Knorr et al., 2024); JULES (Pinnington et al., 2018, 2021), and ORCHIDEE (Raoult et al., 2021). Further details on the history of parameter optimisation in all types of land models are provided in Rayner (2010), Kaminski et al. (2013), Scholze et al. (2017), Rayner et al. (2019), Baatz et al. (2021), and MacBean, Bacour, et al. (2022).

While substantial progress in complex LSM parameter optimisation has been made (particularly for constraining parameters of short timescale vegetation dynamics and carbon fluxes, as described above), a number of challenges hindering objective calibration of the full high-dimensional LSM parameter space remain. Despite advances in the use of analytical techniques to dramatically reduce the time for LSM simulations (Luo et al., 2022; Sun et al., 2023), these highly complex models still have computational requirements – even for one global scale simulation – that are too high for efficient multi-site to global DA experiments. This is true even for "offline" simulations (i.e., LSM simulations forced with climate reanalysis data, as opposed to "online" cases when LSMs are run within the whole ESM). High dimensionality and computational cost make it difficult to calibrate LSMs using conventional statistical approaches like Markov Chain Monte Carlo. Methods used with simpler models often fail with LSMs due to their complexity. These challenges have also meant that LSMs currently struggle to fully leverage the large amount of data from ground networks and Earth observation platforms for calibration.

As an LSM community, thus far, we have no overall strategy for how to proceed towards a system that allows for objective parameter estimation. However, this field is rapidly expanding and we are in a unique position to learn from each other, especially in relation to the technical challenges we face with computational expensive LSM parameter DA. Efforts to build a Land Data Assimilation Community (https://land-da-community.github.io/) by the Analysis, Integration and Modeling of the Earth System (AIMES) Land Data Assimilation 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. Capitalising 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 modelling with the help of DA.

In this paper, we summarise the current state of parameter estimation in land surface modelling,

starting with DA methods, before outlining the different challenges and opportunities our community faces. We then highlight how some of these challenges can be potentially addressed by capitalising on emerging ML techniques and increasing computational capabilities. Finally, we propose future priorities for advancing the field given the urgent need for more accurate and precise LSM projections. We focus on the techniques and challenges related to optimising carbon-vegetation-water interactions in full complexity LSMs but also discuss parameter DA and ML methods applied to intermediate complexity land, carbon cycle and ecosystem models.

This paper complements Kumar et al. (2022) which addresses land surface model data assimilation in the context of state estimation, with a focus on vegetation and hydrology processes. A water cycle-focused perspective, tackling both state and parameter estimation, is offered by De Lannoy et al. (2022).

# 2. Data assimilation methods for parameter estimation in land surface models

LSMs have many parameters that need to be calibrated to accurately reflect the real world (ideally based on observations) and to increase confidence in their future projections. Expert knowledge and empirical measurements of some LSM parameters provide approximate values or their respective ranges. However, due to uncertainties in observations and processes, and the conceptual nature of most parameters, the exact values of LSM parameters are inherently difficult to determine. Instead, we make use of the abundance of observational data indirectly related to the parameters via the processes they are related to, and thus the problem of parameter estimation 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* $\mathbf{y} \cong G(\boldsymbol{\theta})$. In the context of parameter estimation, 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 minimisation effort that locates the argument of a cost function that minimises the discrepancy between $\mathbf{y}$ and $G(\boldsymbol{\theta})$. 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 regularisation, is the Bayesian approach. Bayesian methods include information on the prior distribution of the parameters p($\boldsymbol{\theta}$) to define an entire posterior distribution:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (1)$$

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 (MAP) estimate - the argument that maximises the posterior distribution (i.e., its mode) - provides a point estimate for $\boldsymbol{\theta}$ and is equivalent to a loss minimisation estimate regularised with prior parameter information under Gaussian assumptions. Under such assumptions, maximising the posterior distribution corresponds to minimising the so-called variational cost function:

$$J(\boldsymbol{\theta}) = ½[ (G(\boldsymbol{\theta}) - \mathbf{y})^{T}\mathbf{R}^{-1}(G(\boldsymbol{\theta}) - \mathbf{y}) + (\boldsymbol{\theta} - \boldsymbol{\theta}_{b})^{T}\mathbf{B}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_{b})], \quad (2)$$

where $\mathbf{R}$ and $\mathbf{B}$ are the model/data and prior error covariance matrix, respectively, and $\boldsymbol{\theta}_{b}$ are the prior parameter values.

With the emergence of novel ground and satellite observation sets came the advent and development of techniques to implement them in a field of mathematics originally coined Data Assimilation (DA) (Talagrand & Courtier, 1987). Along with the differences in the aforementioned approaches to solving the inverse problem, these methods also 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). There is some confusion in the community regarding the terminology used when describing DA methods, for example, the false dichotomies sometimes used between “variational and sequential” and “optimisation-based versus Bayesian” - these dichotomies have been marred over time with hybridisation and the continual development of the techniques. Rayner et al. (2019) have made a significant effort to harmonise the notation and clarify overlapping terminology within the community.

Although DA is primarily used in numerical weather forecasting to correct the model state, in LSMs, DA is often employed to reduce parametric uncertainty, a process referred to as parameter data assimilation (PDA). Techniques used in numerical weather forecasting can be adapted for parameter estimation in LSMs. One of the key methods is 4DVar, which involves minimising Eq. 2 (called 4DVar to contrast with 3DVar, where the observations are instead compared to a single model output at a time). The next part of this section looks a little deeper into methods used to reduce this cost function, as well as outlining alternative DA methods that extract the full posterior distribution.

**Methods for reducing cost functions:**

Methods commonly used to minimise the cost (e.g., Eq. 2) require numerical optimisation due to their complex structure and these can usually be grouped into local gradient-descent or global random search techniques. Although more computationally efficient, gradient-descent methods require the gradient of the cost function (either exact, which requires differentiating the entire LSM - see Sect. 3.7, or approximated when exact is not possible or desirable) and they can

result in the location of a local minimum. A common gradient-based minimisation method used in LSM parameter estimation 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. These gradients can be obtained by hand or using automatic differentiation software (Gelbrecht et al., 2023; Griewank, 1997). 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 Sect. 3.7)—often necessitate alternatives. To address this, approximate gradient methods can be employed. One approach is to estimate gradients using finite difference, 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. In cases where gradient information 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 towards the minimum of a cost function, eliminating the need for direct gradient calculations. Additionally, more advanced approaches, such as the ensemble-based 4DVar (4DEnVar) algorithm proposed by Liu et al. (2008) use an ensemble of model trajectories to approximate gradient information via a control variable transform.

Alternatively, global search methods can be used to minimise the cost function. These methods use techniques that try to scan the entire parameter space in some defined way to avoid this pitfall but often require heavy computational power to do so. These global search methods can be cateogrised as Monte Carlo (MC), since they are methods that make use of repeated trials (or sampling) generated using random numbers (Owen, 2013). An example of such a 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.

Although these gradient-descent and global search methods are very efficient in finding an optimal point-estimate of the parameters that minimise the given cost function, usually they do not directly offer information about the posterior error statistics. Nevertheless, it is possible to exploit information about the curvature of the cost function (via the Hessian) at the optimum to obtain such information, but this is typically more complicated than deriving gradient information and more costly in the case of global search.

**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($\boldsymbol{\theta}$|**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 optimisation, as opposed to gradient-descent methods, these techniques are often Monte Carlo in nature and hence also derivative-free (black-box).

Techniques include importance sampling (Kloek & Van Dijk, 1978), a relatively simple approach that samples random values from the prior and accumulates accepted parameterisations 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 Monte Carlo (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 (i.e., 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 parallelisation.

Particle filters provide an alternative to MCMC for sampling from the posterior distribution, particularly in time-evolving systems. They represent the posterior using a set of particles, updating them with each new data point. While computationally intensive and prone to particle degeneracy, particle filters are useful for real-time tracking of system states and time-varying parameters. However, many of the parameters in land surface models are linked to biological processes and thus are subject to change over time due to acclimation, phenotypic plasticity, adaptation and evolution. While some attempts have been made to explore the seasonal variability in parameters (Rowland et al., 2014; Verbeeck et al., 2011), the majority of the literature in land model parameter estimation so far operates on the assumption that parameters are fixed in time. As such, particle filters are rarely used in PDA (Speich et al., 2021) (unless part of joint state-parameter DA, for example, Zhang et al., 2017).

**Applications in LSMs**:
Due to the high number of required model evaluations, MCMC methods have primarily been applied to computationally inexpensive land, carbon cycle, and ecosystem models, or to calibrate isolated processes such as fitting parameters of a two-pool model of substrate dependence in plant respiration (Jones et al., 2024) or parameters of the wetlands CH4 emissions module in the second generation dynamic global vegetation model LPJ-GUESS (Kallingal et al., 2024). For example, MCMC methods have been used to estimate parameters of the Simplified PnET (SIPNET) ecosystem model (Fer et al., 2018; M. Liu et al., 2015; Sacks et al., 2006), TECOS (Xu et al., 2006), FöBAAR forest carbon cycle model (Keenan et al., 2012), BETHY (Knorr & Kattge, 2005) and the DALEC suite of intermediate complexity ecosystem models (Famiglietti et al., 2021; Keenan et al., 2011; D. Lu et al., 2017). DALEC is also at the heart of the cutting-edge CARbon DAta MOdel fraMework (CARDAMOM) where the full potential of MCMC-based carbon parameter estimation is performed (Bloom et al., 2016; Exbrayat, Smallman, et al., 2018; Smallman et al., 2021).

While computationally expensive LSMs build on this foundation, their complexity and parameter volume have made MCMC methods computationally prohibitive. Consequently, 4DVar has been the preferred approach for these models. When the tangent linear or adjoint models have been available (e.g., Bacour et al., 2015; Knorr et al., 2024; Kuppel et al., 2012; Raoult et al., 2016; Schürmann et al., 2016), these have been directly used to minimise the cost function and calculate the Hessian. Alternatively, the Nelder-Mead simplex algorithm (Pinnington et al., 2018), finite differences (Bacour et al., 2019; Bastrikov et al., 2018; MacBean et al., 2015) and 4DEnVar (Pinnington et al., 2020) have all been used to circumvent the need of such models. While some Monte Carlo approaches have been used to calibrate complex LSM parameters—either for global search methods to minimise the cost function or to extract the full posterior distribution—these are typically applied at the site scale and fall short of full global calibrations. Examples include the adaptive population importance sampler used to calibrate the JSBACH model (Mäkelä et al., 2019), the genetic algorithm used to calibrate ORCHIDEE (Bastrikov et al., 2018), and multichain MCMC method DiffeRential Evolution Adaptive Metropolis (DREAM(zs)) (Vrugt et al., 2009) used with CLM (Post et al., 2017) and LPJ-GUESS (Bagnara et al., 2019).

# 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 and the prior distributions over which they are allowed to vary. A common first step is to select from the (potentially quite large) number of model parameters, a subset that is deemed the most influential in some sense. The excluded parameters are then fixed at their nominal values, yielding a parameter space of reduced dimension. 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 i) PDA techniques are computationally demanding, scaling with the number of parameters used in the optimisation, and ii) due to the high degree of equifinality in most parameter spaces (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 at equifinality.

Which model output and metric is tested fundamentally affects the crucial parameter selection if relying primarily on sensitivity analysis. Furthermore, parameter sensitivity is often a function of the parameter prior distributions, about which for many parameters we may have poor knowledge. Indeed, a key distinction between a traditional sensitivity analysis, which may vary all parameters by the same arbitrary amount (e.g. +/- 10%), and an uncertainty partitioning analysis is whether the prior distributions accurately represent our knowledge about model parameters prior to calibration (direct data constraints, formal expert elicitation, etc.) (Dietze et al., 2014; LeBauer et al., 2013; Raczka et al., 2018).

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 solution to combat this is to use spatial pattern correlations as a metric for parameter selection to ensure that the parameters selected are not highly correlated (Dagon et al., 2020). More sophisticated methods include using the adjoint model to determine local sensitivities and global sensitivity methods such as Morris (Morris, 1991) and the variance-based Sobol (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 CABLE (Lu et al., 2013), CLASSIC (Deepak et al., 2024), CLM4.5(FATES) (Massoud et al., 2019), JULES (Pianosi et al., 2017), Noah-MP (Wang et al., 2023) and ORCHIDEE (Dantec-Nédélec et al., 2017; Novick et al., 2022). However, these methods can be hard to implement (see Sect. 3.7 for the discussion about adjoint models) or require a large number of model runs (e.g., O(10,000) for Sobol). Nevertheless, once the adjoint or ensemble exists, it is relatively easy to test the sensitivity of different model outputs.

In complex LSMs, even after selecting the most influential parameters, the large number of vegetation (e.g., 15 plant functional types 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 independently. One way to tackle this issue is to assume that the parameter differences among different groups vary proportionally and, therefore, optimise 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 localising parameters rather than using PFT-specific parameterisations may be more appropriate. As such, methods that allow for independent tuning of parameters within each PFT, or even localisation of parameters, may be necessary. Scaling factors can also be used to target processes without needing to deeply explore detailed parameterisations (e.g., Raoult et al., 2021) used a factor to scale the bare soil resistance to evapotranspiration parameterisation in ORCHIDEE).

Selecting parameters is only one part of the problem - choosing the prior distributions is equally important. In the existing LSM calibration literature, it is very common to assume uniform prior

distributions, either explicitly within Bayesian calibrations or implicitly when selecting uniform range restrictions within parameter estimation using a naive objective function (unlike, for example, classic variational DA techniques such as 4DVar which use an explicit Gaussian prior). In these cases, uniform ranges are often based on informal “expert judgment” or ad hoc trial and error. In some cases, parameter uncertainty ranges can be obtained from *in situ* measurements, such as the TRY database (Kattge et al., 2020). Alternatively, the range can be set based on the operational value of the parameter (e.g., ±20%) - although this should only be done as a last resort. When selecting ranges, extra considerations are needed to ensure that the ranges make physical sense (e.g., not sampling negative values if the parameter needs to be positive), that parameter dependencies are maintained (e.g., two parameters whose ratio should not surpass a given threshold, or multiple parameters that must sum to one) and that plausible relationships are retained (e.g., longevity of wood should be longer than that of foliage).

While uniform distributions are frequently chosen due to the lack of a more specific prior distribution, and often to ensure the range is broad enough to cover edge cases, this approach has significant drawbacks. Uniform priors rarely represent our actual prior knowledge of a system, as they imply that all values within a range are equally likely, but values even a little bit outside that range are impossible. In practice, parameter values in certain parts of parameter space are often known a priori to be more plausible than others. An alternative to assuming uniform prior distributions is to select from any of a plethora of other distributions, with such choices usually driven by a combination of structural constraints (e.g., using zero-bound distributions for non-negative parameters), formal syntheses and meta-analyses of trait data, and structured expert-elicitation exercises (Dietze, 2017; Dietze et al., 2014; LeBauer et al., 2013). However, selecting an inappropriate distribution can be as problematic as using a uniform distribution, especially given that the true prior distribution is often not well known at the start of the calibration process. This highlights the importance of conducting formal prior predictive checks to validate assumptions before proceeding.

Priors constructed from trait data, where available, can often be quite well constrained, acting as a form of data fusion (i.e. combining multiple constraints) and helping to constrain subsequent calibrations to biologically-plausible parts of parameters space. Indeed, accounting for prior trait knowledge can lead to very different conclusions about what parameters need to be included in a calibration, as there are cases where very sensitive parameters may be well constrained a priori (e.g., the parameter controlling the maximum rate of carboxylation - Vcmax) while in other cases much less sensitive, but unconstrained, parameters may plausibly span multiple orders of magnitude and thus contribute more to overall model predictive uncertainty (Dietze, 2017; LeBauer et al., 2013).

Informative non-uniform priors do not have to assume parameter independence; multivariate priors can be constructed to capture known correlation structures and trait trade-offs, both within- and across-PFTs (Shiklomanov et al., 2018). However, quantifying these correlations can be a challenge, and so error covariances are often omitted in PDA, neglecting natural parameter relationships. This simplification can result in an ill-posed inversion problem.

Finally, adopting informative non-uniform priors makes it easier to take advantage of the iterative nature of Bayesian inference, where the posteriors from one round of model calibration can be used as priors in the next round without requiring the recalibration of models to earlier data constraints. Not only does this greatly simplify the updating of model calibrations as new data becomes available, but it offers considerable computational advantages.

It is important to stress that no matter the method used for parameter estimation, solutions only exist in the parameter space defined by the parameter selection and authorised prior ranges (Williamson et al., 2013). Changing the number of parameters, their prior distributions, and/or the model process representation will require new calibrations since the solution may differ due to new parameter interactions and the equifinality of solutions.

## 3.2 Characterisation 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 characterising 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 (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 modelling step from the raw data measurement to any given physical quantity (see Sect. 3.3). Due to the difficulty in separating model and observation errors, they have often been combined in past studies. In fact, the mathematical formalisation commonly used in PDA assumes that observation errors include model errors, thereby treating model discrepancy as part of the observational error.

Although common, combining model error with data 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.

There are a number of arguments for keeping process and observation error distinct. Model process error propagates in space and time when making predictions, while observation error

does not. Additionally, addressing a large process error requires improving the model structure, while addressing a large observation error calls for improving data quality. Furthermore, calibrating models using cost functions that rely solely on fixed a priori observation errors can distort parameter uncertainty estimates as well as the relative weight assigned to different data constraints, as there's often no inherent reason to assume that model skill at predicting a variable is proportional to the accuracy of its measurement. Indeed, it is easy to point to examples where the uncertainty in our ability to model something differs in rank order from our ability to measure that same thing (e.g., at local scale, model predictions of net ecosystem exchange (NEE) are more uncertain than gross primary productivity (GPP: the flux of carbon absorbed into the land surface due to photosynthesis), but observations of GPP are more uncertain than 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 data assimilation 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 modelling 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 data assimilation (Chevallier, 2007; Kuppel et al., 2013). 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 datasets, random errors tend to average out, leaving systematic errors to dominate. These errors have long been recognised 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 appreciated. 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 (Kennedy & O'Hagan, 2001; Oberpriller et al., 2021). Additionally, hybrid models that integrate machine learning with process-based models are being explored as a means to address these challenges (see Sect. 4.2).

Ultimately, interconnected efforts, such as the characterisation of 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 in providing ways forward for discerning errors in data from those in model structure.

## 3.3 Developing observation operators

The term “observation operator” refers to any transformation of the modelled quantity used to allow comparison against observations (Kaminski & Mathieu, 2017). Note that what are often called observations are themselves complex transformations of raw data measurements used to estimate physical quantities comparable to the LSM output. For example, radiances observed by a satellite at the top of the atmosphere can be translated into any number of land surface data products, such as leaf area index. This processing can also be seen as a complex model, such as the inversion of a radiative transfer scheme. Furthermore, these data are usually prepared in such a way that they are available on the model grid.

In some cases, it is possible to assume a one-to-one relationship between the model output and assimilated data, in which case the observation operator is the identity matrix. However, in all other cases, an observation operator is required for DA, and the choice of observation operator can significantly impact the results (Cooper et al., 2019). A common use of an observation operator is to bridge the spatial scale between model and observations, either by aggregating the gridded observations to the resolution of the model or vice-versa (Pinnington et al., 2021). More complex examples of spatial scaling operators utilise a weighted averaging process to match a more detailed description of the observation, such as modelling the point spread function of satellite data, or the footprint of an eddy-covariance flux measurement. For example, Vergopolan et al. (2020) introduced a cluster-based observation operator that maps the Gaussian footprint of satellite observations to the sub-grid scale of high-resolution LSMs. This enables efficiently assimilating coarse soil moisture observations while bridging the spatial scale mismatch with fine-scale LSMs and ground observations (Vergopolan et al., 2021). In an application with flux tower data, Pinnington et al. (2017) partitioned the fluxes to observe different parts of the forest and run separate assimilation experiments for logged and unlogged forest stands.

In another example, atmospheric transport is used to map surface fluxes of gas species, such as $CO_2$, into atmospheric concentrations of that species at sampling points. In this way, flask measurements of $CO_2$ have been used to constrain parameters in models of the terrestrial biosphere (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) and to evaluate simulated net $CO_2$ fluxes after optimising against eddy-covariance data (Kuppel et al., 2014). For non-reactive species, it is sufficient to have data on winds 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.

Observation operators are also used to predict observed quantities that are not directly computed by the model itself. A recent example is the assimilation of SIF data, which is typically assumed to be a proxy for GPP. Examples of SIF observation operators include simple linear relationships with GPP (Bloom et al., 2020; MacBean et al., 2018) through to more complex operators based on the underlying photochemistry and radiative transfer in the canopy, either using empirical simplifications of those processes (Bacour et al., 2019) or using fully

mechanistic models for the operator (Norton et al., 2019). Another example is vegetation optical depth which has been used to constrain above-ground biomass and leaf area index (Scholze et al., 2019).

Scholze et al. (2016, 2019) also developed observation operators to map surface soil moisture (SSM) retrievals to simulated volumetric soil moisture of the surface layer of BETHY, which were also used by Wu et al. (2018, 2020, 2024). SSM is subject to large biases, which therefore necessitates this type of transformation. Numerous models employ methods to map SSM to the climatology of their model, for example through cumulative density function (CDF) matching. Another approach is to focus solely on dynamics (e.g., dry downs, Raoult et al., 2021). The dynamics approach is often used when assimilating vegetation indices, FAPAR or leaf area index (LAI) — retrievals are normalised to estimate the seasonality of phenology instead of the absolute values (MacBean et al., 2015). The optimisation then focuses on a reduced set of phenology-related parameters, rather than including those related to photosynthesis (Bacour et al., 2015).

Forward modelling of remote sensing data — i.e., the process of simulating remote sensing data directly from the LSM outputs rather than assimilating processed satellite products — like in the example of SIF, is the opposite approach to the assimilation of high-level satellite products such as LAI or GPP. A key argument for taking this approach is that assumptions in the retrieval process used in these products are likely inconsistent with the assumptions embedded in the land surface model they are being assimilated into. A clear example of this is the use of satellite GPP products which typically employ a production efficiency approach (e.g. the MODIS GPP product, Running et al., 2021) whereas land surface models often use limiting-rate enzyme kinetic schemes derived from those of Farquhar et al. (1980) and Collatz et al. (1992). Furthermore, satellite-derived GPP estimates typically use environmental drivers such as downwelling shortwave radiation which will almost certainly differ from those used to drive the land surface model they are being assimilated into. Finally, there are often substantial differences between the satelitte-derived estimates (e.g. of GPP or LAI) where the assimilation of any one product is likely biased with respect to the ‘truth’ (which is the primary reason for using the seasonal dynamics rather than the actual values of time series data, as discussed in the previous paragraph). Consequently, discrepancies between these high-level observations and the values of the same variables predicted by a LSM may differ due to these factors and be non-trivial to characterise.

It is appealing, therefore, to assimilate low-level products like SIF or canopy reflectance (Quaife et al., 2008). For canopy reflectance, this typically requires the use of radiative transfer models and is analogous to so-called “radiance assimilation” which is used extensively in numerical weather prediction. In that way, any systematic error between the model and the observations can be attributed to the land model (including the radiative transfer model) itself. For example, Shiklomanov et al. (2021) modified the existing canopy radiative transfer model in the Ecosystem Demography v2 model (ED2) to predict full hyperspectral waveforms, instead of just aggregate visible, near-infrared, and thermal bands, and then used this observation operator to calibrate ED2 against airborne AVIRIS imaging spectroscopy across the eastern temperate US.

Meunier et al. (2022) later used this observation operator in the development of a novel tropical liana PFT. However, low-level satellite products often exhibit variability across domains that are not inherently resolved by the land model, leading to some level of compromise between i) adding complexity to the land model, ii) having an observation operator that is not completely consistent with the underlying model or, iii) accepting that some of the variability in the observations themselves will not be resolved. In the examples of SIF and canopy reflectance, both vary with the relative geometry of the sun and sensor - correctly capturing that directional variability using an observation operator that is physically consistent with the description of the radiative transfer regime implemented in global land surface models (which typically only predict total fluxes, i.e. integrated across the viewing hemisphere) is not currently possible. Nevertheless, the selection and processing of observation data can help mitigate some of these issues. For example, space-time binning of space-borne SIF data across multiple observation geometries can limit the impact of directional effects and potentially increase the consistency between model assumptions and the observed variables.

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 machine learning may offer a unique opportunity within DA applications, as discussed in Sect. 4.3.

## 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” optimisations, grouping sites and performing a single optimisation 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 optimisations (Kuppel et al., 2012; Raoult et al., 2016). Another approach is to average the results of single-site optimisations. While usually less effective than multi-site optimisations, 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 optimised parameters improved simulated land-surface temperature performance.

Both these approaches can be thought of as end-members (all sites the same versus all sites different) in a continuum representing the statistical independence of calibrations across sites. While only just beginning to be utilised to calibrate ecosystem models (Dokoohaki et al., 2022; Fer, Shiklomanov, et al., 2021), hierarchical models have a long history of use in ecology as a way of capturing this continuum, allowing parameters to vary across space and through time,

but constraining that variability with multivariate statistical models that describe that variability. Since the across-site and within-site calibrations are fit simultaneously, this would allow LSM models to “borrow strength” across sites (e.g., reducing equifinality as described above) without forcing parameters to be the same everywhere. Hierarchical models also provide a formal framework for accounting for the fact that out-of-sample predictions are more uncertain (because their parameter vectors need to be predicted) than in-sample predictions at sites where parameter vectors are known. To date, existing hierarchical ecosystem model calibrations have assumed a simple “random effects” structure (i.e. different sites are drawn from the same across-site distribution), but there are important opportunities to explore hierarchical models with across-site spatiotemporal covariances (i.e., sites closer together should be more similar) and across-site covariates (i.e., parameters that explain, and help predict, parameter variability).

A further alternative is the use of intermediate complexity models (e.g., DALEC), which, due to their reduced computational complexity, can retrieve parameters at the pixel scale utilising 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, Smallman, 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 optimised parameters can be inserted into more complex models, enhancing their performance and helping to better understand their internal dynamics (Caen et al., 2022).

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* datasets.

## 3.5 Dealing with large and multiple observational datasets

Although EO instruments can provide global gridded datasets with which to calibrate the models, fully exploiting these opportunities is challenging. Running experiments at the same resolution as the satellite products (e.g., 500m MODIS resolution; Justice et al., 2002) requires a lot of computational power and time, and we do not always have access to matching meteorological forcing data. The resolution of products to be assimilated may also not be meaningful for the objectives of the experiment. Additionally, when assimilating more than one remote sensing data constraint, we must address multiple competing resolutions. This requires decisions about scaling (see Sect. 3.3), determining which products are to be upscaled (aggregated) versus downscaled (interpolated). Generally, satellite products are scaled to

match the chosen model grid, usually dictated by the resolution of the forcing data, although this scaling can result in an over-generalisation or loss of information.

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 optimisation (e.g., MacBean et al., 2015), although defining what is representative is a challenge in itself. Once selected, the representative pixel approach helps to i) reduce the dimensionality of the problem, allowing for efficient and multi-data-stream calibrations, ii) focus on points with close to homogenous coverage to be able to calibrate class-specific parameters (e.g., plant functional types), and iii) define a different evaluation set of pixels with which to assess the optimisations, especially sites with additional ground data. After selecting representative pixels, multi-pixel optimisations are performed (as described in Sect. 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 optimisation 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 optimisations are becoming paramount to provide adequate constraints since parameters are likely to impact different parts of the model. By selecting only one specific data stream in an optimisation, we risk degrading the model's overall predictive capacity if some of the optimised 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 optimisation, 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 optimisations 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 and the characterisation of error correlations between them (Bacour et al., 2023). Nevertheless,

it must be stressed 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 Sect. 3.2).

## 3.6 Including the spin-up and transient historical period in the assimilation to better constrain land carbon sink projections

Many LSM simulations include both a spin-up phase that brings the prognostic variables including vegetation state, soil carbon pools, and soil moisture content into equilibrium prior to the industrial revolution (c. 1750). This is followed by a transient historical simulation where the model is driven by changing climate forcing, rising $CO_2$ levels, nitrogen deposition, and prescribed land management and land cover change since the equilibrium time point up to the present day. Even with transient forcings, this historical period is likely not accurately simulated, in part due to the lack of accurate historical climate and land use forcing data, in part because “slow” carbon cycling parameters (e.g. carbon allocation or turnover rates) that control the magnitude of the equilibrium carbon stock are poorly constrained, and in part because the effects of key global change drivers on carbon storage (including recovery from disturbance) are often missing or not reliably represented in models. The result is a large spread in the magnitude and dynamics of various carbon pools and fluxes which underpin the current and future projections of the land carbon sink (Arora et al., 2020; Friedlingstein et al., 2023).

To obtain reliable estimates of the current or future land carbon sink and trend in atmospheric $CO_2$ we need accurate simulations of global carbon stock trajectories (i.e., *changes* in carbon stocks). The trend in carbon stocks depends on the magnitude of carbon stocks post spin-up, which in turn is strongly controlled by soil carbon pool turnover rates (Exbrayat, Bloom, et al., 2018) (in addition to other parameters involved in soil carbon decomposition that moderate that turnover rate). This is because for the CENTURY type model (Parton et al., 1987) used in many LSMs, heterotrophic respiration is partly dependent on the size of carbon stocks. Global sensitivity analyses (Sect. 3.1) of soil carbon cycle models performed for multiple different biomes worldwide have rarely been performed (though see Huang et al., 2018) due to the computational expense of running long-timescale simulations needed to model carbon stock trajectories. For the same reason, relatively few past parameter DA studies with computationally expensive LSMs at multi-site or global scale have included these slow-acting carbon cycle parameters in their assimilation experiments. However, we know from past DA studies that optimising “fast” carbon cycle flux related parameters related to photosynthesis, phenology, and ecosystem respiration has limited impact on regional to global scale carbon stocks (MacBean, Bacour, et al., 2022), as expected, while “slow” carbon cycle process parameters (such as those related to carbon allocation to different biomass pools, or biomass and soil carbon pool turnover times) are important for constraining long-term carbon stock trajectories (Thum et al., 2017).

To optimise the “slow” acting carbon cycle parameters involved in carbon allocation, biomass turnover and soil carbon cycling, LSM assimilation experiments would need to include the spin-up and transient runs in the assimilation, which would be prohibitively costly given the computational cost of LSM runs. Therefore, neither the spin-up or transient period (prior to the assimilation window) are usually included in LSM assimilations (Peylin et al., 2016; Raoult et al., 2016; Schürmann et al., 2016). This presents challenges for obtaining accurate model estimates of carbon fluxes and stocks because an incorrect initial carbon stock will likely result in biased parameter retrievals that are accounting for the model errors contributing to the incorrect initial carbon stock. Note this is not the case for carbon cycle and ecosystem models that have much faster run times and who have therefore been able to include biomass and soil carbon turnover rates and other related “slow” carbon cycling parameters in their optimisations (e.g., CARDAMOM-DALEC – Bloom et al., 2016).

To make up for incorrect carbon pool magnitudes and the fact that including spin-up and transient in the assimilation is not yet feasible, most past carbon cycle parameter DA studies have included scalars on the initial C pools in the optimisation, resulting in an improved fit to NEE and atmospheric $CO_2$ data (e.g., $\eta$, Carvalhais et al. (2008, 2010); $K_{soilC}$ in ORCHIDEE PDA studies, e.g., Peylin et al. (2016); $f_{slow}$ in CCDAS studies, Castro-Morales et al. (2019); Schürmann et al., (2016)). These scalars alter the initial carbon pool size to account for model and forcing errors mentioned above that contribute to incorrect soil carbon stock sizes. Studies differ in how many such scalars to include, both in terms of which carbon pools to relax (all C pools as in Santaren et al. (2007) versus slow and/or passive as in Peylin et al. (2016), whether to scale aboveground biomass or not (Carvalhais et al., 2010), and to how many to use spatially in global simulations (1 in CCDAS, Castro-Morales et al. (2019), Schürmann et al. (2016), versus 30 regional factors used in ORCHIDEE studies, Bacour et al. (2023), Peylin et al. (2016)). Other options for avoiding spin-up include directly initialising models with carbon stock observations, and including parameter calibration within iterative state DA approaches. 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.

Adjusting initial carbon stocks without optimising the “slow” carbon cycle parameters to which the equilibrium carbon stock 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 is insufficient. In longer runs (up to 2100 or 2300) those “slow” carbon cycling parameters that resulted in the original incorrect carbon stock magnitude will start to push the model back to that original (inaccurate) equilibrium, resulting in an artificial trend in the modelled carbon pools (and resultant biases in carbon fluxes and land carbon sink estimates). Thus, for long term projections of carbon-climate feedbacks, all parameters that are important for carbon pool trajectories need to be included in the assimilations. This means that longer time windows (lasting several hundreds to thousands of years) governing the periods over which these “slower” carbon cycle parameters operate will

need to be included in the assimilation experiments (Raiho et al., 2021; Thum et al., 2017). This will materially increase the computational cost of an experiment enough to be prohibitive for computationally expensive LSMs with current simulation protocols and assimilation algorithms. Methods for increasing the simulation speed (e.g., model emulation - see Sect. 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 data assimilation (Hararuk et al., 2014; Huang et al., 2018; Luo et al., 2022; Tao et al., 2020, 2024). Intermediate complexity ecosystem models may be able to assist by providing much constrained priors of soil carbon pool turnover times (and other parameters to which equilibrium/initial carbon stock magnitude are sensitive) (Bloom et al., 2016).

This problem is specific to long-term, slowly changing carbon (and other nutrient like nitrogen and phosphorus) stocks: e.g., for water storage (e.g., soil moisture), usually only a few years are required either for spin-up or to adjust to a given perturbation. Therefore, for hydrology-focused simulations both the spin-up and historical period spanning the perturbation from equilibrium can be included in the experiment. In fact, by including this shorter spinup, the assimilation also gives an estimate of the initial state (e.g., soil moisture, Pinnington et al., (2021); snow albedo, Raoult et al. (2023)). While carbon cycling is interlinked with water and energy cycles, long-term carbon stock trajectories are insensitive to short-term fluctuations in soil moisture.

In addition to longer assimilation time windows, assimilating measurements of aboveground biomass or soil C stocks in conjunction with carbon fluxes provides 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, and 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 datasets 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) in addition to nighttime and soil respiration data that so far have typically not been utilised 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 (Raczka et al., 2018). 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 minimisation algorithms

To perform optimisations effectively, careful consideration must be given to the choice of algorithm and its implementation. As discussed in Sect. 2, various algorithms are available, each with distinct characteristics, such as local versus global optimisation, each having different computational demands. 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 (6 options), crossover (5 options), and mutation (3 options), resulting in 90 possible combinations. Users can also adjust crossover and mutation probabilities. The success of the optimisation process greatly depends on how the optimisation 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 Sect. 4.1) rather than an LSM to find an optimal experimental design (Dagon et al., 2020); once the design has been identified, the optimisation 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, the tangent linear and adjoint models can take years to develop, even when using automatic differentiation software, since the code first needs to be cleaned and structural adjustments need to be made to ensure the code is differentiable without changing the fundamental physics. For example, this may require replacing look-up tables with their continuous formulations and reformulating minimum and maximum calculations to allow a smooth transition at the edge (Schürmann et al., 2016). The years taken to derive the tangent linear/adjoint models mean they quickly become outdated, especially with big community models like JULES and ORCHIDEE, where new processes are added approximately every six months. For JULES, the adjoint was developed for v2.2 of the model (Raoult et al., 2016), whereas JULES is currently at v7.3 at the time of writing. Similarly, while the tangent linear exists for ORCHIDEE, it exists for an old version of the model (AR5) that predates the addition of a multi-layered soil hydrology scheme and nitrogen cycle. To address this issue, the ORCHIDEE DA team has been developing a tool to do the required preprocessing of any version of ORCHIDEE so the tangent linear version of the model can be easily derived using Transformation of Algorithms in Fortran (Giering, 2010). On the other hand, BETHY's lower complexity has allowed it to be kept compliant with automatic differentiation software for decades, which provided efficient derivative code of the up-to-date version of the model. This is also the case for its successor D&B (Knorr et al., 2024), which is the model component of the European Space Agency supported TCASS system, and for the Nanjing University Carbon Assimilation System (NUCAS, Zhu et al., 2023). Alternatively, models written directly in an auto-differentiable language (Julia or python-JAX; see Sect. 5.4) alleviate this issue (Gelbrecht et al., 2023; C. Shen et al., 2023). Although these languages have slower computational performance than Fortran, these new languages often also facilitate the use of graphic processing units (GPU), e.g., through packages like pyTorch (Paszke et al., 2019).

As discussed in Sect. 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 (Pinnington et al., 2020) or the use of emulators, which can be used to either avoid gradient-based approaches in favour of Monte Carlo 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 land surface models and rigorously diagnose their uncertainties. We now have access to large observational datasets at high spatio-temporal resolutions and increased computational capacity and efficiency. These factors, combined with recent advances in machine learning (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 help us address the current challenges and limitations in land PDA outlined above, as well as areas where ML has the potential to improve the DA workflow (Fig. 1). With the large number of studies currently being published in the field of machine learning, we only provide a short overview of the relevant literature. In the context of ML for PDA, we can broadly group the existing studies and applications into four categories: i) the use of ML to emulate the relationship between LSM parameters and its outputs or performance (Sect. 4.1), ii) the creation of ‘hybrid models’ in which ML replaces or complements a component of a larger LSM (Sect. 4.2), iii) the use of ML to improve or pre-process observation datasets prior to their use in PDA (Sect. 4.3), and iv) the use of ML to optimise the parameter estimation process itself (Sect. 4.4).

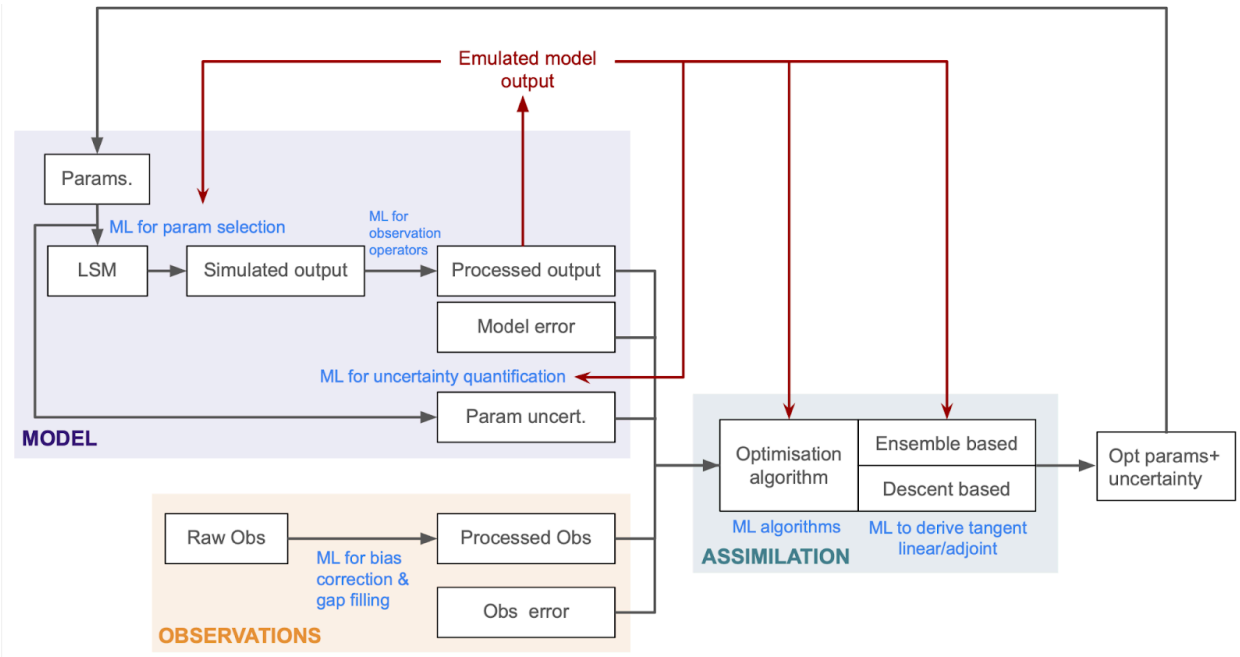


**Figure 1**: Examples of where ML can facilitate each part of the land surface model PDA workflow

## 4.1 Parameter perturbation emulators

The computational cost of high-complexity LSMs hinders the use of the more computationally demanding PDA techniques such as MCMC. However, machine learning methods can mitigate a portion of these computational burdens. By building a 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).

**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, see McNeall et al. (2024) and Kennedy et al. (2024) for PPEs constructed for JULES and CLM, respectively). 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, it can be more cost effective to use a non-random and targeted sampling strategy, such as active learning which tries to optimise the selection of the next sample (e.g., 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 optimise 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.

**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 itself (i.e., the response surface of model error as a function of parameter value) which is much lower dimensional and often much smoother that the model output itself (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 dataset, they are less feasible for larger datasets. One option is to develop sparse Gaussian processes, as demonstrated by Baker et al. (2022). Running JULES at a 1km resolution over Great Britain, they exploit the fact that LSMs typically do not exchange information laterally between grid cells (river routing is generally done as a separate step) to select a subset of coordinates representative of different parameter settings and forcing data regimes.

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 datasets without performance degradation (Shwartz-Ziv & Armon, 2022). (D. Lu & Ricciuto, 2019) used singular value decomposition with Bayesian optimisation to create a reduced number of surrogate models for carbon modelling parameter perturbation. Their approach showed minimal accuracy loss, making it effective for extensive parameter space exploration and uncertainty quantification. Other examples of NNs used to emulate LSMs include, Dagon et al. (2020), where a series of artificial feed-forward NNs were trained to emulate CLM5 output given important biophysical parameter values and Meyer et al. (2022), where an NN was trained to emulate the ensemble mean of several urban LSMs combining the strengths of the different into one ML model. While artificial NNs do not provide a probabilistic prediction, new methods are emerging such as neural processes (e.g., (Garnelo et al., 2018) or randomised 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 (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 (LSTM) methods, which for example have been applied to ECLand (Boussetta et al., 2021), 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 for example, (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 (e.g., Datta & Faroughi, 2023; Guo et al., 2021; Wesselkamp et al., 2024).

**Computational Cost Reduction**

Once an emulator is trained it becomes computationally feasible to apply PDA techniques that require a large number of samples from a prior parameter distribution, e.g., MCMC. Fer et al., (2018) showed how emulators sped up an MCMC optimisation for the relatively simple SIPNET model by over two orders of magnitude (>100x). Further applying their method to the more complex Ecosystem Demography model v2 (ED2), whose complexity precluded it from a direct application of the MCMC methodology for parameter tuning, they found that emulators helped achieve a >20,000x increase in speed (27 hr versus a predicted 74 years by traditional MCMC). Similarly, Sawada (2020) and Cleary et al. (2021) both used emulators to perform Bayesian inversion using the otherwise costly MCMC approach to sample the approximate posterior parameter distribution after calibration. Torres-Rojas et al. (2022) combine surrogate modelling with a multi-objective Pareto efficiency analysis to infer LSM's optimal subgrid parameters at 1% of the computational cost. The emulators were trained on forward model runs used to initially calibrate the model using Ensemble Kalman sampling - a derivative-free optimisation method. 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).

**History Matching**

Emulators are commonly used in the field of uncertainty quantification, and one key method from this field that is gaining traction in land surface modelling is the so-called history matching (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). 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), it is arguably less dangerous to get them wrong here than in the DA case - if the errors are overestimated, HM gives a clear diagnostic of this being the case, for example, 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, in other optimisation approaches, we would still get a solution even if one does not exist. HM also allows the user to

test many different metrics to see if parameters can capture specific features, similar to multi-objective optimisations, giving a clear diagnosis of model structure error. HM 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 optimisation.

## 4.2 Hybrid modelling

ML can also be used in a hybrid modelling approach to substitute components of the physical model with an ML approximation (Eyring et al., 2024). 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). ML and process models can be combined in a number of different ways, including i) substituting a specific model parameterisation with an ML approximation, ii) deriving spatial parameterisations that better capture observed physical behaviour, iii) training on model-data residuals to predict process-model biases and characterise structural errors, and iv) replace computationally costly parts of the model. Hybrid modelling 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 modelling (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). As with all parameter estimation methods, hybrid modelling can be subject to parameters compensating for model structural errors or errors in parameters outside the calibration set (see also Sect. 4.2). This can be counteracted through the use of multivariate independent observation constraints in the calibration.

**Substitution of Uncertain or Missing Parameterisations and Processes**

In the context of land DA, hybrid modelling has been used to improve the representation of complex processes, such as the representation of human processes and their impact, which are often not represented in their full complexity or missing completely in traditional LSMs. 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 can effectively account for human processes by mapping observations into the model climatology (thus removing global biases). At the same time, they can retain the independent information on human processes that is inherent in the observations but typically removed in traditional bias correction approaches (e.g.,

Kumar et al., 2012). Kolassa et al. (2017) used an artificial NN observation operator trained on brightness temperature observations from the Soil Moisture Active Passive (SMAP) mission and GEOS land model outputs to assimilate soil moisture information, which introduced the impact of irrigation and tile drainage in a model that does not normally represent these processes. Assem et al. (2017) developed a Deep Convolutional NN, trained on historic water flow and water level observations, to predict water flow in urban areas from runoff estimates generated by a physical LSM. Hybrid modelling 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.

**Improved Spatial Parameterisations**

Hybrid modelling techniques have also been used successfully to generate model parameterisations that better capture the parameter spatial distribution and thus the observed physical behaviour (Tao et al., 2020, 2024). Process-model parameterisations 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 modelling can address these issues by mapping environmental variables into model parameters or using high-resolution, high-fidelity model simulations to derive new parameterisations for coarse-resolution models (e.g., Gentine et al., 2018). Bao et al. (2023) replaced the traditional PFT-based parameterisation of a light use efficiency model with an ecosystem-property-based parameterisation 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 (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.

**Model Error Identification/Characterisation**

Additionally, hybrid modelling 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 characterise 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.

**Computational Cost Reduction**

Finally, hybrid modelling can be used to replace computationally costly parts of the model. For example, emulating the spinup, which can account for up to 98% of computational time in complex LSMs, would greatly alleviate challenges linked to this bottleneck (see Sect. 3.6). A successful undertaking by Sun et al. (2023) showed how bagging decision trees (an ensemble ML method based on (Breiman, 1996) could be used to emulate the spin-up of the ORCHIDEE LSM. Koppa et al. (2022) developed a deep learning-based hybrid model combining a process-based land surface model with remotely-sensed observations to estimate global evaporation. They showed how hybrid models can significantly improve predictive accuracy while reducing the computational cost.

**Data Requirements**

Hybrid modelling has the potential to be very powerful, but it is also susceptible to issues linked to equifinality (Kraft et al., 2022; Sawada, 2020). 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.

## 4.3 Observation Processing

There are many examples of using ML to improve or pre-process the observational datasets that can be assimilated into LSMs, especially from the field of remote sensing. Many of these novel datasets have yet to be exploited in the LSM parameter estimation studies, presenting exciting new opportunities.

**Observation Operators**

One such application is the use of ML-generated observation operators to translate satellite-observed radiances into model states or parameters (see challenges raised in Sect. 3.3). The use of ML techniques in this context has several advantages: i) ML-based observation operators are relatively simple to implement compared to physically-based approaches, which often involve the inversion of radiative transfer models, ii) they are able to easily accommodate the simultaneous assimilation of multiple observation types, iii) they can inherently correct climatological biases between model and observations, and iv) they facilitate the assimilation of radiance observations rather than retrieval products, thus reducing errors stemming from possible inconsistencies between retrieval algorithm assumptions and models. Due to these advantages, ML-based observation operators have been applied in several land data assimilation studies, including for soil moisture (Kolassa et al., 2017; Rodríguez-Fernández et al., 2019), leaf area index (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).

**Retrieval Algorithms**

Similarly, ML approaches have been used to develop data-driven retrieval algorithms in cases where physical retrieval algorithms are very complex. For example, Chen et al. (2022), Gentine

& Alemohammad (2018), Shen et al. (2022) and Zhang et al. (2018) each used ML to estimate SIF from MODIS radiances, OCO-2, and TROPOMI observations, respectively. Alemohammad et al. (2017) developed an ML approach to retrieve global, monthly GPP estimates from GOME-2 SIF observations only.

**Gap-Filling**

ML approaches can also be used to improve observation datasets by making them more suitable for data assimilation applications. One approach is to use ML to generate gap-filled observations or generate higher temporal resolution datasets. For example, Yatheendradas & Kumar (2022) used an ML approach to create a gap-filled, high-resolution dataset of observed snow cover fraction and Fang et al. (2019) used a deep learning Long Short-Term Memory framework to predict daily "SMAP Level-3 like" soil moisture estimates from atmospheric forcing data and static physiographic attributes. Vekuri et al. (2023) used extreme graident boosting to gap-fill eddy covariance data reducing the northern biases in the data found after using more traditional gap-filling methods. Nevertheless, one must exert caution when using gap-filled data (or other model-derived data, such as retrieval products) for parameter estimation, since they are dependent on the assumptions of the selected gap-filling method. Furthermore, gap-filled data can artificially inflate sample size, which leads to falsely precise parameter estimates.

**Upscaling**

Another approach is to use ML to map local observations to the global scale to mitigate representativeness issues that can arise from the assimilation of local observations. For example, studies by Beer et al. (2010), Joiner et al. (2018), Jung et al. (2011) and Tramontana et al. (2016) all have used ML approaches in combination with remote sensing observations to generate global estimates of carbon and energy fluxes from local flux-tower observations. Vergopolan et al. (2021) used a high-resolution LSM and an ML Bayesian merging scheme trained on in-situ soil moisture data to learn LSM and SMAP satellite biases and obtain 30m satellite-based soil moisture estimates over the contiguous United States. One caveat to using ML to upscale point observations is that large discrepancies can exist between different data products based on the same observations, highlighting the need for thorough evaluation and uncertainty assessment of ML-based products.

**Derived Quantities**

Finally, ML can be used to improve the algorithms used to generate observation datasets. For example, Tramontana et al. (2020) used a combined neural network approach that accounts for the influence of soil property and micrometeorological drivers to generate improved estimates of the partitioning of observed NEE into GPP and ecosystem respiration (RECO), while Zeng et al. (2022) used an ML approach to separate the natural and anthropogenic contributions to satellite-estimated evapotranspiration.

## 4.4 Optimisation process

Since optimisation is a key component to both ML and DA, there are many algorithms common to both fields including gradient-based and evolutionary algorithms (Sect. 2). Indeed, the strong

mathematical similarities between ML and DA mean that both fields can learn from each other and share methodologies (Geer, 2021). ML approaches can be used to improve optimisation 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 optimisation algorithms. For example, clustering methods can be used to set the population size, crossover probability and mutation probability parameters in genetic algorithms (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 optimise ML structure and hyperparameters (Akiba et al., 2019). ML techniques can also be used to choose the best-performing algorithm for a particular optimisation 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 optimising the parameter estimation algorithm hyperparameters themselves.

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 Sect. 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 neural networks to emulate the tangent linear and adjoint models since neural networks can be differentiated trivially (Hatfield et al., 2021).

Table 1: Summary of challenges outlined in Sect. 3 and their ML opportunities

| **PDA challenge** | **ML opportunity** |
|---|---|
| Selecting parameters and their prior distributions (Sect. 3.1) | |
| - Identifying which model parameters to optimise is challenging, due to high dimensionality and strong parameter covariances.<br>- Choosing prior distributions for parameters is crucial yet difficult, requiring detailed structural insights and data. | - Emulators can reduce the computational demand of running models with many different parameter settings needed for sensitivity analyses (Sect. 4.1).<br>- Emulators can be used to facilitate uncertainty quantification, for example, through history matching (Sect. 4.1). |
| Characterisation of model and data/observation errors (Sect. 3.2) | |
| - Model errors are difficult to quantify due to uncertainties in process representation, missing processes, and the challenge of specifying an informative prior.<br>- Quantifying data errors is tricky because of sampling variability, instrument inaccuracies, and complex error correlations that are often ignored. | - Hybrid modelling can be used to replace model processes that are missing or poorly understood, helping to diagnose model structural errors (Sect. 4.2).<br>- ML methods can be used to generate improved estimates of derived quantities, thus reducing observation errors (Sect. 4.3). |

| | |
|---|---|
| Developing observation operators (Sect. 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 (Sect. 4.3). |
| Tackling spatial and temporal heterogeneity (Sect. 3.4) | |
| - Variability in surface properties, driven by diverse climates, soils, and ecosystems, complicates parameter estimation across locations.<br>- High computational demands make it difficult to calibrate LSMs across large spatial domains.<br>- Temporal variability and short data series hinder the capture of both seasonal cycles and long-term trends. | - Hybrid modelling can be used to improve spatial parameterisations (Sect. 4.2).<br>- Emulators can help reduce the computational demand of running the model over large domains (Sect. 4.1).<br>- Long Short-Term Memory encoder-decoder networks consider long-term dependencies and therefore may help capture seasonal and interannual trends (Sect. 4.1). |
| Dealing with large and multiple observational datasets (Sect. 3.5) | |
| - Scaling satellite products to match model grids can lead to information loss.<br>- Products may be subject to regional biases due to varying data quality and processing methods.<br>- Assimilating multiple data streams in model calibrations presents challenges in consistency, error characterisation, and balancing different data sources. | - ML methods can be used to upscale sparse observational data (e.g., flux tower observations) or map satellite observations to a model grid (Sect. 4.3).<br>- ML can be applied to improve the algorithms used to produce observational datasets (Sect. 4.3).<br>- ML-based observation operators are able to easily accommodate multiple observation types and adjust their respective impacts in the assimilation (Sect. 4.3). |
| Including the historical period in the assimilation window (Sect. 3.6) | |
| - Spin-up and transient parts of model runs can be computationally demanding. | - Hybrid modelling can be used to replace computationally costly parts of the model (Sect. 4.2). |
| Choice and implementation of minimisation algorithms (Sect. 3.7) | |
| - Algorithms requiring a large number of model runs are computationally costly and therefore rarely applied to complex LSMs.<br>- For different algorithms, there can be a large number of configuration options and tuneable hyperparameters. | - ML can enhance computational efficiency, enabling the use of algorithms that require numerous model runs (Sect. 4.1).<br>- ML can help find the best configurations and hyperparameters to use when optimising (Sect. 4.4). |

| | |
|---|---|
| - Maintaining tangent linear/adjoint models for gradient-based optimisation 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 (Sect. 4.4). |

# 5. Future priorities

Moving beyond the ML avenues outlined in the previous section and summarised in Table 1, 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 datasets and experimental configurations

In addition to the traditional datasets used to optimise LSM parameters, our data-rich world offers access to a wide array of 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 fertilisation effect at nitrogen-limited sites (Thomas et al., 2017; Jiang et al., 2020; Mahmud et al., 2018; Raoult, Edouard-Rambaut, et al., 2024).
- **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 (Lawrence et al., 2020) can help constrain rates of soil carbon cycling (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 datasets:** New datasets, such as full-waveform lidar data from the GEDI (Global Ecosystem Dynamics Investigation) 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). 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 utilised to improve the representation of methane production processes.

By combining these data and implementing novel DA approaches described in this paper, we can aspire to assess how this information influences both short-term and long-term forecasts and reduces model discrepancies. The focus should be on refining core processes driving ecosystem-scale carbon and water fluxes and testing their responses to global change, beyond just fitting historical data.

As with all past carbon cycle DA studies, before novel datasets 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 observational characteristics (temporal sampling interval, record length, observation uncertainty, choice of minimisation algorithm and its configuration, etc. – Sect. 3.7) are required to retrieve the correct parameter values with the strong assumption that there is no modelling 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; Fox et al., 2009). As these data are model outputs, the “true” value of the parameters is known. Synthetic DA experiments can also be used prior to data collection, where they can help optimise sampling over space, time, and sampling design. Indeed, calibration has yet to be adequately integrated into the broader literature on model-driven observing system simulation experiments. To improve this, advocating for standardised community benchmark protocols and datasets could address different challenges, such as assessing resistance to noise and evaluating forcing variability. Results from such community-driven experimental setups could reveal common challenges and development opportunities, enhancing the robustness and effectiveness of DA methods across the field (see Sect. 5.4).

Additional tests of DA experimental configuration that are rarely performed (or rarely reported in the literature) should include testing i) how parameters retrieved at individual sites compare to parameters retrieved 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 Sect. 3.4), ii) the utility of PFT dependent parameters versus alternative approaches for grouping parameters (e.g., regionally dependent PFTs - e.g. Dahlin et al., 2017; Bao et al., 2023), iii) how retrieved parameters vary with the forcing dataset used in the simulations, iv) how retrieved values depend on which parameters and/or PFTs are optimised or which terms to include in the cost function, and v) how retrieved parameters vary in space and time within PFTs and what this tells us about missing processes, among other factors. A critical test of any parameterisation 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 optimised or a novel dataset is used in the assimilation. Building DA frameworks to include this technical testing will give confidence in using retrieved parameter values in operational versions of the models.

## 5.2 Moving towards land surface–atmospheric transport and full Earth system model coupling in data assimilation

Atmospheric $CO_2$ mole fraction measurements collected at tall towers around the world have proven valuable in improving NEE predictions at regional to global scales within a carbon cycle DA framework (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 (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. One benefit of assimilating atmospheric $CO_2$ concentration data is that it is one of the only datasets that can provide a large spatial scale constraint (albeit indirect) on RECO because it is heavily influenced by soil carbon stocks; thus, assimilating atmospheric $CO_2$ data presents an opportunity to improve the representation of both soil carbon flux and stock trajectories in LSMs, which is crucial for future predictions regarding the carbon sink capacity of terrestrial ecosystems.

However, the assimilation of atmospheric $CO_2$ data requires coupling LSMs with atmospheric transport models in order to scale 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 modelling errors associated with the atmospheric model itself (physics and spatial/vertical discretisation) 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) and Kaminski & Mathieu (2017). Recent advances in the utilisation of graphics accelerators (Chevallier et al., 2023) offer hope for a significant reduction in computational times and the development of 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 optimisations. To date, there has been limited assessment of whether posterior parameter values from offline DA experiments compare to retrieved values from fully coupled runs (nor how retrieved values vary when different offline climate reanalysis forcing products are used). To achieve this goal, LSM DA groups should learn from advances made in the NWP 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 (Sect. 4.1) (Watson-Parris et al., 2021) and automatic differentiation of model code (Gelbrecht et al., 2023) should help considerably in alleviating this problem (see Sect. 3.7 for remaining challenges).

## 5.3 Identifying and improving structural errors and model representation

The best estimates of different parameters are very dependent on the experimental setup and so few of the optimised 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). Instead, the main strength of parameter estimation for LSMs and, therefore, its main purpose thus far, has been to identify structural errors. If we cannot match observations within the bounds of their known uncertainties by simply changing the parameter values, this suggests that a process is poorly represented or missing from the model. This critical information is then fed back to the model developers to ensure changes are made to the model, before restarting the cycle of model calibration. Although this exchange is key in developing any LSM, it is rarely published. Nevertheless, a few documented examples from the ORCHIDEE land surface model workflow exist. MacBean et al. (2015) demonstrated that temperate broadleaved temperature thresholds for senescence in the ORCHIDEE LSM were too low. The newly optimised parameters have since been included in ORCHIDEE trunk versions. Salmon et al. (2022) found that when constraining parameters of the ORCHIDEE LSM against methane emissions in northern peatlands, the process providing enough active carbon for methanogenesis was missing. Raoult et al. (2023) found by assimilating MODIS snow albedo over Greenland that a three-layered ice sheet model was insufficient to simulate accurately both the snow albedo and runoff rates, leading to further discretisation of the model.

However, careful consideration is needed to avoid equating the status quo of making changes to models—often involving increased complexity—with progress in model development. While identifying and addressing structural errors is crucial, introducing new processes or refining existing ones can sometimes lead to models that are more complex without necessarily improving their predictive power. It is important to strike a balance between enhancing model accuracy and maintaining model parsimony. Overly complex models can become difficult to validate and manage, potentially obscuring rather than clarifying underlying processes. Therefore, the goal should be to make thoughtful adjustments that improve model performance while ensuring that the added complexity is justified by significant improvements in accuracy or functionality. This approach ensures that models remain robust and efficient and that any modifications contribute meaningfully to their overall effectiveness.

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 datasets 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 optimisation theories) and facilitate collaboration across different modelling groups.

## 5.4 International collaboration: intercomparison studies and shared toolboxes

Efforts by AIMES and ILMF to build a Land Data Assimilation 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 modelling 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 27th 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 utilising 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 land surface model benchmarking tools (iLAMB: Collier et al., 2018; Seiler et al., 2022). 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 datasets, 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).

## 5.5 Propagation of error reductions to constrain climate predictions

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 to enable observationally constrained probabilistic statements to be made about the future of the land biosphere. 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 trivial, but 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). However, with increased computational power, we are acquiring the capability to run LSMs as ensembles rather than relying on a single realisation, thereby enabling us to better capture the uncertainty of model predictions (Arora et al., 2023).

By sampling from the posterior distributions after a PDA experiment, we can generate ensemble simulations which can be used to explore future scenarios and idealised 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, which likely requires the use of emulators (see Sect. 4.1). The second is how to intelligently select parameter vectors from those distributions. Currently, simple models might propagate uncertainty by using 100-1000 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 towards 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 (IPCC) 8th Assessment Report, the annual Global Carbon Budget (GCB) and other emerging frameworks quantifying land carbon trajectories. Therefore, we must develop methods to maximise 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 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.

# 6. Summary and conclusion

Improving the accuracy of land surface models (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 data assimilation 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 Earth System Models) are complex in nature leading to many challenges when performing global scale optimisations. Nevertheless, advances in computational capability, novel datasets 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 land surface model data assimilation, helping to streamline the assimilation process, manage large datasets and speed up otherwise computationally demanding processes. International collaboration is crucial in this endeavour, as shared knowledge and resources can significantly accelerate the advancement of LSM calibration and data assimilation. Efforts to build a Land Data Assimilation Community, such as those by the AIMES Land Data Assimilation 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 standardised methodologies, ultimately leading to more accurate LSMs.

# Acknowledgements

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 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 Award DE-SC0022070, the National Science Foundation (NSF) under IA 1947282, and the National Center for Atmospheric Research (NCAR) through 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 number 337552) and Horizon Europe, HORIZON-MISS-2022-SOIL-01-05. 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).

# Open research

**Data Availability statement:** This article discusses the challenges and priorities in the field of parameter estimation for land data assimilation, and the opportunities offered by machine learning—it does not include the specific use of any particular software or results involving specific data products.

# Appendix

The process-based models mentioned through the paper are listed in Table A1. This list cover wide spectrum of land models ranging in complexity and computational demand, including 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 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.

**Table A1**. References for the process-based models mentioned in this article.

| **Acronym** | **Full name** | **Model reference** |
|---|---|---|
| BETHY | Biosphere Energy Transfer Hydrology | Knorr (2000) |
| 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 | Lawrence et al. (2019) |
| D&B | DALEC & BETHY | Knorr et al. (2024) |
| DALEC | Data Assimilation Linked Ecosystem Carbon | Williams et al. (2005) |
| ED | Ecosystem Demography | Ma et al. (2022); Moorcroft et al. (2001) |
| ECLand | European Centre 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 Biosphera-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 | Organising 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 & Heimann (1995) |
| SIPNET | Simplified Photosynthesis and Evapotranspiration | Braswell et al. (2005) |
| TECOS | terrestrial ecosystem | Xu et al., (2006) |

1

# References


Abadie, C., Maignan, F., Remaud, M., Kohonen, K.-M., Sun, W., Kooijmans, L., Vesala, T., Seibt, U., Raoult, N., Bastrikov, V., Belviso, S., & Peylin, P. (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., Finzi, A., Moorhead, D., Schimel, J., Torn, M., & Mayes, M. A. (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

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1907.10902

Alemohammad, S. H., Fang, B., Konings, A. G., Aires, F., Green, J. K., Kolassa, J., Miralles, D., Prigent, C., & Gentine, P. (2017). Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically-based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. *Biogeosciences*, *14*(18), 4101–4124. https://doi.org/10.5194/bg-14-4101-2017

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

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., … Ziehn, T. (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., Beaudoing, H. K., Getirana, A., Navari, M., Li, B., Jacob, J., Wegiel, J., & Peters-Lidard, C. D. (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., Fennel, K., Beck, H., De Lannoy, G., Pauwels, V. R. N., Raiho, A., Montzka, C., Williams, M., Mishra, U., Poppe, C., Zacharias, S., Lausch, A., Samaniego, L., Van Looy, K., …

Vereecken, H. (2021). Reanalysis in earth system science: Toward terrestrial ecosystem reanalysis. *Reviews of Geophysics (Washington, D.C.: 1985)*, *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., Peylin, P., Chevallier, F., Vuichard, N., & Bastrikov, V. (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., Weiss, M., Demarty, J., Santaren, D., Baret, F., Berveiller, D., Dufrêne, E., & Prunet, P. (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., Dormann, C. F., & Hartig, F. (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., Blyth, E., Bousserez, N., Boussetta, S., Brown, A., Buizza, R., Buontempo, C., Chevallier, F., Choulga, M., Cloke, H., Cronin, M. F., Dahoui, M., De Rosnay, P., Dirmeyer, P. A., … Zeng, X. (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). 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://gmd.copernicus.org/articles/11/4739/2018/

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., … Papale, D. (2010). Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science (New York, N.Y.)*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R., Ménard, C. B., & Harding R. (2011). The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://gmd.copernicus.org/articles/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., Meyer, V., Reager, J. T., Worden, H. M., Jiang, Z., Quetin, G. R., Smallman, T. L., Exbrayat, J.-F., Yin, Y., Saatchi, S. S., Williams, M., & Schimel, D. S. (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., Melton, J. R., Pongratz, J., Turton, R. H., Yoshimura, K., & Yuan, H. (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 (New York, N.Y.)*, *359*(6375). https://doi.org/10.1126/science.aam8328

Booth, B. B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., Huntingford, C., Betts, R. A., Harris, G. R., & Lloyd, J. (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., Agustí-Panareda,

A., Beljaars, A., Wedi, N., Munõz-Sabater, J., de Rosnay, P., Sandu, I., Hadade, I., Carver, G., Mazzetti, C., Prudhomme, C., Yamazaki, D., & Zsoter, E. (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., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., & Zhang, Q. (2018). *JAX: composable transformations of Python+NumPy programs* (Version 0.3.13). 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., & Kueppers L. (2021). Capturing functional strategies and compositional dynamics in vegetation demographic models. *Biogeosciences*, *18*(14), 4473–4490. https://bg.copernicus.org/articles/18/4473/2021/bg-18-4473-2021.html

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). 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., Papale, D., Rambal, S., & Seixas, J. (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., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M. J., & Valentini, R. (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), 13,218–13,235. https://doi.org/10.1002/2016JD024821

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

Chen, X., Huang, Y., Nie, C., Zhang, S., Wang, G., Chen, S., & Chen, Z. (2022). A long-term reconstructed TROPOMI solar-induced fluorescence dataset using machine learning algorithms. *Scientific Data*, *9*(1), 427. https://doi.org/10.1038/s41597-022-01520-1

Chevallier, F. (2007). Impact of correlated observation errors on inverted CO2surface fluxes from OCO measurements. *Geophysical Research Letters*, *34*(24). 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). https://doi.org/10.1029/2022gl102135

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., & Cox P. (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://gmd.copernicus.org/articles/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., Mu, M., & Randerson, J. T. (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., Rio, C., Audouin, O., Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M.-P., Madeleine, J.-B., Rodier, Q., & Xu, W. (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). 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., Valdayskikh, V., Radchenko, T., Nekrasova, O., Zakharov, V., & Jouzel, J. (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), 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. *Atmosphere-Ocean*, *0*(0), 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., Crow, W. T., de Rosnay, P., Durand, M., Forman, B., Geppert, G., Girotto, M., Hendricks Franssen, H.-J., Jonas, T., Kumar, S., Lievens, H., Lu, Y., Massari, C., Pauwels, V. R. N., … Steele-Dunne, S. (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*. https://doi.org/10.3389/frwa.2022.981745

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., Schepers, D., Weston, P., Zuo, H., Alonso-Balmaseda, M., Balsamo, G., Bonavita, M., Borman, N., Brown, A., Chrust, M., Dahoui, M., Chiara, G., English, S., Geer, A., … Rabier, F. (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

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., Kooper, R., LeBauer,

D., Mantooth, J., McHenry, K., & Wang, D. (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. (n.d.). *xgboost-distribution: Probabilistic prediction with XGBoost*. Github. Retrieved August 15, 2024, 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

Duan, T., Anand, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., & Schuler, A. (13--18 Jul 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. https://proceedings.mlr.press/v119/duan20a.html

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (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., Gayno, G., & Tarpley, J. D. (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). 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., Bloom, A. A., Falloon, P., Ito, A., Smallman, T. L., & Williams, M. (2018). Reliability ensemble averaging of 21st century projections of terrestrial net primary productivity reduces global and regional uncertainties. *Earth System Dynamics*, *9*(1), 153–165. https://doi.org/10.5194/esd-9-153-2018

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., Bocquet, M., Bretherton, C. S., Christensen, H. M., Dagon, K., Gagne, D. J., Hall, D., Hammerling, D., Hoyer, S., Iglesias-Suarez, F., Lopez-Gomez, I., McGraw, M. C., Meehl, G. A., Molina, M. J., … Zanna, L. (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., Parazoo, N. C., Stettz, S. G., Yang, Y., Bonal, D., Bloom, A. A., Williams, M., & Konings, A. G. (2021). Optimal model complexity for terrestrial carbon cycle prediction. *Biogeosciences* , *18*(8), 2727–2754. https://doi.org/10.5194/bg-18-2727-2021

Fang, K., Pan, M., & Shen, C. (2019). The value of SMAP for long-term soil moisture estimation with the help of deep learning. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *57*(4), 2221–2233. https://doi.org/10.1109/tgrs.2018.2872131

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. Royal Meteorological Society (Great Britain)*, *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 CO2 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., Desai, A., Duveneck, M. J., Fisher, J. B., Haynes, K. D., Hoffman, F. M., Johnston, M. R., Kooper, R., LeBauer, D. S., Mantooth, J., Parton, W. J., Poulter, B., Quaife, T., Raiho, A., … Dietze, M. C. (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., Murphy, B., Desai, A. R., & Dietze, M. C. (2021). Capturing site-to-site variability through Hierarchical Bayesian calibration of a process-based dynamic vegetation model. In *bioRxiv* (p. 2021.04.28.441243). 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). https://doi.org/10.1029/2018ms001453

Fisher, R. A., Muszala, S., Verteinstein, M., Lawrence, P., Xu, C., McDowell, N. G., Knox, R. G., Koven, C., Holm, J., Rogers, B. M., Spessa, A., Lawrence, D., & Bonan, G. (2015). Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes, CLM4.5(ED). *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., Fisher, J. B., Shi, M., Walker, A. P., & Lawrence, D. M. (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., v. 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., von Bloh, W., & Carvalhais, N. (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., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., & Van Wijk, M. T. (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

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., … Zheng, B. (2023). Global Carbon Budget 2023. *Earth System Science Data*, *15*(12), 5301–5369. https://doi.org/10.5194/essd-15-5301-202310.5194/essd-15-5301-2023-supplement

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Ali Eslami, S. M., & Teh, Y. W. (2018). Neural Processes. In *arXiv [cs.LG]*. arXiv. 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., & Alemohammad, S. H. (2018). Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. *Geophysical Research Letters*, *45*(7), 3136–3146.

https://doi.org/10.1002/2017GL076294
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. In *arXiv [physics.ao-ph]*. arXiv. https://doi.org/10.5194/egusphere-2024-277
Giering, R. (2010). *Transformation of algorithms in Fortran Version 1.15 (TAF Version 1.9. 70)*.
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. (1997). *On Automatic Differentiation*. 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? In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2207.08815
Groenendijk, M., Dolman, A. J., Ammann, C., Arneth, A., Cescatti, A., Dragoni, D., Gash, J. H. C., Gianelle, D., Gioli, B., Kiely, G., Knohl, A., Law, B. E., Lund, M., Marcolla, B., van der Molen, M. K., Montagnani, L., Moors, E., Richardson, A. D., Roupsard, O., … Wohlfahrt, G. (2011). Seasonal variation of photosynthetic model parameters and leaf area index from global Fluxnet eddy covariance data. *Journal of Geophysical Research: Biogeosciences*, *116*(G4). 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., Loisel, J., Malhotra, A., Jackson, R. B., Ogle, S., Phillips, C., Ryals, R., Todd-Brown, K., Vargas, R., Vergara, S. E., Cotrufo, M. F., Keiluweit, M., Heckman, K. A., Crow, S. E., … Nave, L. E. (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). https://doi.org/10.1029/2021ms002521
Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms* (2nd ed.) [PDF]. 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., Campanharo, W. A., House, J. I., Sitch, S., Hales, T. C., Adami, M., Anderson, L. O., & Aragão, L. E. O. C. (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., Cassol, H. L. G., Achard, F., Jucker, T., Silva, C. A., House, J., Sitch, S., Hales, T. C., & Aragão, L. E. O. C. (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., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., … Zuo, H. (2018). Operational global reanalysis: progress, future directions and synergies with NWP | ERA Report Series. *ECMWF*. https://www.ecmwf.int/en/elibrary/80922-operational-global-reanalysis-progress-future-directions-and-synergies-nwp
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., Guimberteau, M., Jornet-Puig, A., Lu, X., & Luo, Y. (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
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., Evans, M. N., Babst, F., & Luyssaert, S. (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-14-5891-202110.5194/gmd-14-5891-2021-supplement
Jiang, M., Medlyn, B. E., Drake, J. E., Duursma, R. A., Anderson, I. C., Barton, C. V. M., Boer, M. M., Carrillo, Y., Castañeda-Gómez, L., Collins, L., Crous, K. Y., De Kauwe, M. G., Dos Santos, B. M., Emmerson, K. M., Facey, S. L., Gherlenda, A. N., Gimeno, T. E., Hasegawa,

S., Johnson, S. N., … Ellsworth, D. S. (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

Jian, J., Vargas, R., Anderson-Teixeira, K., Stell, E., Herrmann, V., Horn, M., Kholod, N., Manzon, J., Marchesi, R., Paredes, D., & Bond-Lamberty, B. (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

Joiner, J., Yoshida, Y., Zhang, Y., Duveiller, G., Jung, M., Lyapustin, A., Wang, Y., & Tucker, C. J. (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., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., … Williams, C. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *116*(G00J07). https://doi.org/10.1029/2010jg001566

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., … Reichstein, M. (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., Roy, D., & Morisette, J. (2002). An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, *83*, 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., Vestin, P., & Weslien, P. (2024). Assimilating multi-site eddy-covariance data to calibrate the $CH_4$ wetland emission module in a terrestrial ecosystem model. In *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 Biogeochem. 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., Blessing, S.,

Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J. P., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., … Ziehn, T. (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., 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

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., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar C, C., Aleixo, I., Ali, H., … Wirth, C. (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., Carbone, M. S., Reichstein, M., & Richardson, A. D. (2011). The model-data fusion pitfall: assuming certainty in an uncertain world. *Oecologia*, *167*(3), 587–597. https://doi.org/10.1007/s00442-011-2106-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

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

Kennedy, D., Dagon, K., Lawrence, D. M., Fisher, R. A., Sanderson, B. M., Collier, N., Hoffman, F. M., Koven, C. D., Levis, S., Oleson, K. W., Zarakas, C. M., Cheng, Y., Foster, A., Lu, X., Fowler, M., Hawkins, L., Kavoo, T., Kluzek, E. B., Kumar, S., … Luo, Y. (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

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*. https://www.jstor.org/stable/1913641?casa_token=1EgxpT-njfMAAAAA:EPhIlE24zFcuOWFrDS_G8r0LMGIuWx60_Ft7yeBlx9XViRf0sI27b3uD1MmKewRbJkO2Lcgb7G3kDQS-sQBi1vItqRAyTEHOyQfEkAYw1W6DtaIc-PGm

Knorr, W. (2000). Annual and interannual $CO_2$ exchanges of the terrestrial biosphere: process-based simulations and uncertainties: CO2 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 CO2 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., 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). 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., Quaife, T., Smallmann, L., Steele-Dunne, S., Vreugdenhil, M., Green, T., Zähle, S., Aurela, M., Bouvet, A., Bueechi, E., Dorigo, W., El-Madany, T., Migliavacca, M., Honkanen, M., … Drusch, M. (2024). A comprehensive land surface vegetation model for multi-stream data assimilation, D&B v1.0. In *EGUsphere*. https://doi.org/10.5194/egusphere-2024-1534

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). https://doi.org/10.1029/2010gb003900

Kolassa, J., Reichle, R. H., Liu, Q., Cosh, M., Bosch, D. D., Caldwell, T. G., Colliander, A., Collins, C. H., Jackson, T. J., Livingston, S. J., Moghaddam, M., & Starks, P. J. (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., Lewis, J., Lindsay, K., Mathesius, S., Meinshausen, M., Mills, M., Nicholls, Z., Sanderson, B. M., Séférian, R., Swart, N. C., Wieder, W. R., & Zickfeld, K. (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.,

Detto, M., Dietze, M. C., Faybishenko, B., Holm, J., Huang, M., Kovenock, M., Kueppers, L. M., Lemieux, G., Massoud, E., McDowell, N. G., Muller-Landau, H. C., Needham, J. F., Norby, R. J., … Xu, C. (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*. Aspendale, Vic., 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., Ciais, P., Sitch, S., & Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system: DVGM FOR COUPLED CLIMATE STUDIES. *Global Biogeochemical Cycles*, *19*(1). https://doi.org/10.1029/2003gb002199

Kumar, S., Kolassa, J., Reichle, R., Crow, W., de Lannoy, G., de Rosnay, P., MacBean, N., Girotto, M., Fox, A., Quaife, T., Draper, C., Forman, B., Balsamo, G., Steele-Dunne, S., Albergel, C., Bonan, B., Calvet, J.-C., Dong, J., Liddy, H., & Ruston, B. (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). 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). 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://gmd.copernicus.org/articles/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://gmd.copernicus.org/articles/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., Jacobs, E. M., McNellis, R., Rady, J. M., Smith, N. G., Thomas, R. Q., Wieder, W. R., & Grandy, A. S. (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., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., & Battaglia, P. (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., Heckman, K., Blankinship, J. C., Crow, S. E., McNicol, G., Trumbore, S., Levine, P. A., Vindušková, O., Todd-Brown, K., Rasmussen, C., Hicks Pries, C. E., Schädel, C., McFarlane, K., Doetterl, S., … Wagai, R. (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., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., … Zeng, X. (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

Li, L., Fang, Y., Zheng, Z., Shi, M., Longo, M., Koven, C. D., Holm, J. A., Fisher, R. A., McDowell, N. G., Chambers, J., & Leung, L. R. (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-16-4017-202310.5194/gmd-16-4017-2023-supplement

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., Wang, H., & Zhou, G. (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

Lu, D., Ricciuto, D., Walker, A., Safta, C., & Munger, W. (2017). Bayesian calibration of

terrestrial ecosystem models: a study of advanced Markov chain Monte Carlo methods. *Biogeosciences*, *14*, 4295–4314. https://doi.org/10.5194/BG-14-4295-2017

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., Hararuk, O., Hou, E., Jiang, L., Liao, C., Lu, X., Shi, Z., Smith, B., Tao, F., & Wang, Y.-P. (2022). Matrix approach to land carbon cycle modeling. *Journal of Advances in Modeling Earth Systems*, *14*(7). 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

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

MacBean, N., Bacour, C., Raoult, N., Bastrikov, V., Koffi, E. N., Kuppel, S., Maignan, F., Ottlé, C., Peaucelle, M., Santaren, D., & Peylin, P. (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., Köhler, P., Gómez-Dans, J., & Disney, M. (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., Oinonen, M., Straková, P., & Viskari, T. (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., Lohila, A., Mammarella, I., Margolis, H., Markkanen, T., Susiluoto, J., Thum, T., Viskari, T., Zaehle, S., & Aalto, T. (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

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

Massoud, E. C., Xu, C., Fisher, R. A., Knox, R. G., Walker, A. P., Serbin, S. P., Christoffersen, B. O., Holm, J. A., Kueppers, L. M., Ricciuto, D. M., Wei, L., Johnson, D. J., Chambers, J. Q., Koven, C. D., McDowell, N. G., & Vrugt, J. A. (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., Williams, S., Motew, M., Campbell, E. E., Brummitt, C. D., Paustian, K., & Kumar, A. A. (2023). Validating DayCent-CR for cropland soil carbon offset reporting at a national scale. *Geoderma*, *438*(116647), 116647. https://doi.org/10.1016/j.geoderma.2023.116647

Mauritsen, T., Bader, J., & Becker, T. (2019). *Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2*. https://doi.org/10.1029/2018ms001400

McGovern, A., Elmore, K. L., Gagne, D. J., II, Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., & Williams, J. K. (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: Biogeosciences*, *114*(G1). 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., De Deurwaerder, H. P. T., Krishna Moorthy, S. M., Schnitzer, S. A., Marvin, D. C.,

Longo, M., Liu, C., Broadbent, E. N., Almeyda Zambrano, A. M., Muller-Landau, H. C., Detto, M., & Verbeeck, H. (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

Migliavacca, M., Meroni, M., Busetto, L., Colombo, R., Zenone, T., Matteucci, G., Manca, G., & Seufert, G. (2009). Modeling gross primary production of agro-forestry ecosystems by assimilation of satellite-derived information in a process-based model. *Sensors (Basel, Switzerland)*, *9*(2), 922–942. https://doi.org/10.3390/s90200922

Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological Monographs*. 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 CO2 and H2O fluxes. *Agricultural and Forest Meteorology*, *148*(10), 1467–1477. https://doi.org/10.1016/j.agrformet.2008.04.013

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., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Zhang, W., Hamdi, Z. M., … Jung, M. (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 CH4 and N2O flux measurements in terrestrial ecosystems. *Atmospheric Environment (Oxford, England: 1994)*, *81*, 311–319. https://doi.org/10.1016/j.atmosenv.2013.09.030

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., MacBean, N., Raoult, N., Scott, R. L., Shi, Y., Sulman, B. N., & Wood, J. D. (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., & Hartig. (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://gmd.copernicus.org/articles/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*. *16*, 19–22. 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

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Neural Information Processing Systems*, *abs/1912.01703*. https://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library

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., Koffi, E., Kane, A., Maignan, F., Chevallier, F., Ciais, P., & Prunet, P. (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., Robinson, E., Morrison, R., Osborne, S., & Quaife, T. (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. M., Casella, E., Dance, S. L., Lawless, A. S., Morison, J. I. L., Nichols, N. K., Wilkinson, M., & Quaife, T. L. (2017). Understanding the effect of disturbance from selective felling on the carbon dynamics of a managed woodland by combining observations with model predictions. *J. Geophys. Res. Biogeosci.*, *122*, 886–902. https://doi.org/10.1002/2017JG003760

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://hess.copernicus.org/articles/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

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., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*(112165), 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., Lomas, M., D'Alessandro, M. M., Paillou, P., Papathanassiou, K., Rocca, F., Saatchi, S., Scipal, K., Shugart, H., Smallman, T. L., Soja, M. J., Tebaldini, S., Ulander, L., Villard, L., & Williams, M. (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

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

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., Charbit, S., Dumas, C., Maignan, F., Ottlé, C., & Bastrikov, V. (2023). Improving

modelled albedo over the Greenland ice sheet through parameter optimisation and MODIS snow albedo retrievals. *The Cryosphere*, *17*(7), 2705–2724. https://doi.org/10.5194/tc-17-2705-2023

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. 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

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

Rayner, P. J. (2010). *The current state of carbon-cycle data assimilation*. *2*(4), 289–296. https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=H3up71wAAAAJ:isC4tDSrTZIC

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). https://doi.org/10.1029/2004gb002254

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (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). 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 CO2 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). https://doi.org/10.1029/2010jg001400

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., Ponton, S., Bonal, D., Meir, P., & Williams, M. (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 CO2 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., Guimbaud, C., Zhu, D., Ciais, P., Peylin, P., Gogo, S., Laggoun-Défarge, F., Aurela, M., Bret-Harte, M. S., Chen, J., Chojnicki, B. H., Chu, H., Edgar, C. W., Euskirchen, E. S., … Ziemblińska, K. (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., Saisana, M., & Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Wiley & Sons, Limited, John. 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., Ingram, W. J., Piani, C., Stainforth, D. A., Stone, D. A., & Allen, M. R. (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., Bacour, C., Ciais, P., & Longdoz, B. (2014). Ecosystem model optimization using in situ flux observations: benefit of Monte Carlo versus variational schemes and analyses of the year-to-year model performances. *Biogeosciences*, *11*(24), 7137–7158. https://doi.org/10.5194/bg-11-7137-2014

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). 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). 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., Kerr, Y., Mialon, A., Richaume, P., Rodríguez-Fernández, N., Vittucci, C., Wigneron, J.-P., Mecklenburg, S., & Drusch, M. (2019). Mean European carbon sink over 2010–2015 estimated by simultaneous assimilation of atmospheric $CO_2$, 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. *Https://agupubs.onlinelibrary.wiley.com › 2007JD008642https://agupubs.onlinelibrary.wiley.com › 2007JD008642*. https://doi.org/10.1029/2007JD008642

Schürmann, G. J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., Giering, R., Rödenbeck, C., Heimann, M., & Zaehle, S. (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., Goll, D., Jain, A. K., Joetzjer, E., Lienert, S., Lombardozzi, D., Luyssaert, S., Nabel, J. E. M. S., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., & Zaehle, S. (2022). Are terrestrial biosphere models fit for simulating the global land carbon sink? *Journal of Advances in Modeling Earth Systems*, *14*(5). https://doi.org/10.1029/2021ms002946

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., Albergel, C., Calvet, J.-C., Ku, O., & Georgievska, S. (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), 113116. https://doi.org/10.1016/j.rse.2022.113116

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., … Lawson, K. (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

Shen, H., Wang, Y., Guan, X., Huang, W., Chen, J., Lin, D., & Gan, W. (2022). A spatiotemporal constrained machine learning method for OCO-2 solar-induced chlorophyll fluorescence (SIF) reconstruction. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *60*, 1–17. https://doi.org/10.1109/tgrs.2022.3204885

Shiklomanov, A. N., Cowdery, E. M., Bahn, M., Byun, C., Jansen, S., Kramer, K., Minden, V., Niinemets, Ü., Onoda, Y., Soudzilovskaia, N. A., & Dietze, M. C. (2018). Does the leaf economic spectrum hold within plant functional types? A Bayesian multivariate trait

meta-analysis. In *bioRxiv* (p. 475038). 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
Shi, Z., Allison, S. D., He, Y., Levine, P. A., Hoyt, A. M., Beem-Miller, J., Zhu, Q., Wieder, W. R., Trumbore, S., & Randerson, J. T. (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
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., Arneth, A., Arora, V. K., Bastos, A., Bastrikov, V., Bellouin, N., Canadell, J. G., Chini, L., Ciais, P., Falk, S., Harris, I., Hurtt, G., Ito, A., Jain, A. K., … Zaehle, S. (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., Jenkins, S., Leach, N., Mathison, C., & Partanen, A.-I. (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), 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., Lu, L., & Denning, A. S. (2008). Remote sensing data assimilation for a prognostic phenology model: DATA ASSIMILATION AND PHENOLOGY MODELING. *Journal of Geophysical Research*, *113*(G4). 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., Mishra, U., Jiang, L., Huang, X., & Luo, Y. (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., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Doughty, R., Ren, Z., & Luo, Y. (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 Netherlands. 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., Dinon-Aldridge, H., Burkhart, H. E., Domec, J.-C., Fox, T. R., Gonzalez-Benecke, C. A., Martin, T. A., Noormets, A., Sampson, D. A., & Teskey, R. O. (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., Loustau, D., & Ciais, P. (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). https://doi.org/10.1029/2022wr032233

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S.,

& Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Tramontana, G., Migliavacca, M., Jung, M., Reichstein, M., Keenan, T. F., Camps-Valls, G., Ogee, J., Verrelst, J., & Papale, D. (2020). Partitioning net carbon dioxide fluxes into photosynthesis and respiration using neural networks. *Global Change Biology*, *26*(9), 5235–5253. https://doi.org/10.1111/gcb.15203

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., & Nikolova, S. (2007). OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *Journal of Geophysical Research: Biogeosciences*, *112*(G2). 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

Vekuri, H., Tuovinen, J.-P., Kulmala, L., Papale, D., Kolari, P., Aurela, M., Laurila, T., Liski, J., & Lohila, A. (2023). A widely-used eddy covariance gap-filling method creates systematic bias in carbon balance estimates. *Scientific Reports*, *13*(1), 1720. https://doi.org/10.1038/s41598-023-28827-2

Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., & Ciais, P. (2011). Seasonal patterns of CO2fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. *Journal of Geophysical Research*, *116*(G2). 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 the Environment*, *242*(111740), 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., Torres-Rojas, L., Sadri, S., & Wood, E. F. (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 Simulation*, *10*(3). https://doi.org/10.1515/IJNSNS.2009.10.3.273

Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., Bastrikov, V., & Peylin, P. (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). https://doi.org/10.1029/2023jd038556

Wang, J., Jiang, F., Wang, H., Qiu, B., Wu, M., He, W., Ju, W., Zhang, Y., Chen, J. M., & Zhou, Y. (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), 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), 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., Boedecker, J., Dormann, C. F., Pappenberger, F., & Balsamo, G. (2024). Advances in Land Surface Model-based Forecasting: A comparative study of LSTM, Gradient Boosting, and Feedforward Neural Network Models as prognostic state emulators. In *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., Kooijmans, L. M. J., Hilton, T. W., Belviso, S., Peylin, P., Commane, R., Sun, W., Chen, H., Kuai, L., Mammarella, I., Maseyk, K., Berkelhammer, M., Li, K.-F., Yakir, D., … Campbell, J. E. (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. In *arXiv [physics.comp-ph]*. arXiv. http://arxiv.org/abs/2401.00035

Wu, M., Jiang, F., Scholze, M., Chen, D., Ju, W., Wang, S., Kaminski, T., Lu, Z., Vossbeck, M., & Zheng, M. (2024). Regional responses of vegetation productivity to the two phases of ENSO. *Geophysical Research Letters*, *51*(8). https://doi.org/10.1029/2024gl108176

Wu, M., Scholze, M., Kaminski, T., Voßbeck, M., & Tagesson, T. (2020). Using SMOS soil moisture data combining CO2 flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS). *Remote Sensing of Environment*, *240*(111719), 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, K., Liu, P., Zhang, J., Han, D., Wang, G., & Shen, C. (2021). Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *Journal of Hydrology*, *603*(127043), 127043. https://doi.org/10.1016/j.jhydrol.2021.127043

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., Du, M., & Liu, C. (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

Yatheendradas, S., & Kumar, S. (2022). A novel machine learning–based gap-filling of

fine-resolution remotely sensed snow cover fraction data by combining downscaling and regression. *Journal of Hydrometeorology*, *23*(5), 637–658. https://doi.org/10.1175/jhm-d-20-0111.1

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). https://doi.org/10.1029/2009gb003522

Zeng, H., Elnashar, A., Wu, B., Zhang, M., Zhu, W., Tian, F., & Ma, Z. (2022). A framework for separating natural and anthropogenic contributions to evapotranspiration of human-managed land covers in watersheds based on machine learning. *The Science of the Total Environment*, *823*(153726), 153726. https://doi.org/10.1016/j.scitotenv.2022.153726

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

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., & Gentine, P. (2018). A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences*, *15*(19), 5779–5800. https://bg.copernicus.org/articles/15/5779/2018/bg-15-5779-2018.html

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., Lin, C., Li, X., & Qiu, G. Y. (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. In *arXiv [cs.DC]*. arXiv. http://arxiv.org/abs/2405.00018

Zhu, H., Wu, M., Jiang, F., Vossbeck, M., Kaminski, T., Xing, X., Wang, J., Ju, W., & Chen, J. M. (2023). Assimilation of Carbonyl Sulfide (COS) fluxes within the adjoint-based data assimilation system–Nanjing University Carbon Assimilation System (NUCAS v1.0). In *EGUsphere* (pp. 1–35). https://doi.org/10.5194/egusphere-2023-1955

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

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

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

[1]Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK
[2]National Centre for Earth Observation, Department of Meteorology, University of Reading, Reading, UK
[3]Departments of Geography and Environment & Biology, Western University, Canada
[4]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[5]Science Systems and Applications, Inc., Lanham, MD, USA
[6]Computing and Data Sciences, Boston University, Boston, MA 02215 USA
[7]CICERO, Norway
[8]Finnish Meteorological Institute, Helsinki, Finland
[9]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Gif-sur-Yvette, France
[10]NSF National Center for Atmospheric Research, Boulder, CO, USA
[11]Earth and Environmental Engineering Department, Columbia University, New York, NY 10027, USA
[12]Max Planck Institute for Biogeochemistry, Hans Knöll Strasse 10, 07745 Jena, Germany;
[13]Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[14]ELLIS Unit Jena, Jena, Germany
[15]UK Centre for Ecology and Hydrology, Wallingford, UK
[16]Department of Earth & Environment, Boston University, Boston, MA 02215 USA
[17]Earth and Environmental Engineering Department, Columbia University, New York, NY 10027, USA
[18]The Inversion Lab, Hamburg, Germany
[19]NSF National Center for Atmospheric Research, Boulder, CO, USA
[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, 85721, USA
[23]European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK
[24]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden
[25]School of Environmental Studies, Queen's University, Kingston, ON, K7L3N6, Canada
[26]School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh, EH9 3FF
[27]Earth Environment and Planetary Sciences, Rice University, Houston, TX 77006, USA
[28]European Commission Joint Research Center, Ispra, 21027 VA, Italy
[29]Department of Mathematics, Statistics, and Computer Science, Augsburg University, Minneapolis, MN 55454 USA
*now at European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK

## Key points (max 140 characters each)

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

## Abstract (max 250 words)

Accurately predicting terrestrial ecosystem responses to climate change is crucial for addressing global challenges. This relies on mechanistic modelling 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 optimisation, 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, characterising 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 datasets, and including the spin-up and historical period in the assimilation window. We then 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 datasets, harness machine learning advances, and enhance the predictive capabilities of LSMs.

## Plain language summary (max 200 words)

Improving the accuracy of land surface models (LSMs) is crucial for reducing uncertainties in climate change projections. Parameter data assimilation, 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 optimisation. Advances in computational power, novel datasets, and machine learning (ML) offer promising solutions to improve these models. ML can streamline the data assimilation process, handling large datasets 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 machine learning 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 Data Assimilation Working Group and the International Land Model Forum, is essential for accelerating progress, facilitating knowledge exchange, and developing standardised methods for more accurate climate modelling.

# 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. This strongly relies on our ability to accurately model the underlying processes at the global scale. Such global-scale, mechanistic or process-based models of the terrestrial biosphere, often embedded in Earth system models (wherein they are called Land Surface Models – LSMs; Blyth et al., 2021), 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 (including emissions from land use change), 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 also indispensable tools in assessing climate change mitigation strategies, for example, to assess how effective nature-based solutions such as reforestation will be in curbing rising $CO_2$ emissions.

Representing all the requisite processes corresponding to interacting vegetation, biogeochemistry, water and energy cycles mechanistically (and accurately) in LSMs over a wide range of timescales, from sub-daily flux exchanges with the atmosphere to decadal-century timescales representative of changes in biomass and soil carbon pools required for carbon-climate feedbacks, is critical for robust and reliable projections (Watson-Parris, 2021). However, LSMs are highly complex and subject to large uncertainties, both in terms of missing processes, inadequate representation of processes, and poorly constrained parameters. Furthermore, when trying to address model structural uncertainty, implementing new processes tends to introduce additional parameters and, therefore, more parameter uncertainty. As a result, LSMs often diverge significantly in their representation of many terrestrial processes (Gier et al., 2024; Green et al., 2024; Varney et al., 2024). Consequently, their predictions of important ecosystem responses under future climate change scenarios often vary widely. For example, LSMs disagree on the magnitude of the land carbon sink (Koven et al., 2022; Shi et al., 2024), and the potential constraints on $CO_2$ fertilisation due to water (Green et al., 2019) and nutrient (Davies-Barnard et al., 2022) limitations.

Parametric uncertainty is one of the largest sources of uncertainty in all types of land models ( simple, intermediate and full complexity models), particularly for 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). Indeed, it has been shown for one LSM that even perturbing a single carbon flux related parameter within its range of uncertainty can result in a projection spread in atmospheric $CO_2$ by 2100 that is larger than running the model under different emissions scenarios (Booth et al., 2012). We urgently need to reduce this uncertainty to ensure we can utilise the full potential of LSMs—parameter optimisation is one way to achieve this.

Many processes in LSMs (as well as processes in ecosystem models, see Table A1 for all process-based models mentioned in the paper) are controlled by parameters that represent the

functioning of individual elements of the system. While some of these parameters can be directly observed (e.g. photosynthetic capacity, wood density, rooting depth, 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. Even those parameters that can be directly measured can often only be observed at scales that differ from the grid resolution of most global-scale LSM simulations (typically 0.5 degrees or greater). As a result, LSM predictions – particularly for vegetation and carbon cycle related processes – can be highly sensitive to parameter choices (in addition to model parameterisation or structural uncertainties) (Booth et al., 2012; Buotte et al., 2021; Exbrayat et al., 2014; Fisher et al., 2019; Oberpriller et al., 2022; Smallman et al., 2021; Zaehle, Friedlingstein, et al., 2010).

Historically, LSM parameters have simply been manually tuned (adjusted by hand to produce more realistic model behaviour or to better fit a given important model variable to a given dataset). Manual tuning of LSM parameters was often the only option given the required rapid pace of LSM development, the lack of available data at the correct scales for LSM parameter optimisation, or the computational demand of optimising the large number of parameters (typically >200) in LSMs with many complex, interacting processes. However, in the last two decades, the hurdles associated with performing rigorous LSM parameter optimisation (as opposed to tuning) have diminished to the point that it has become feasible: many datasets have become available at LSM-relevant scales, and the computational cost of running LSMs has decreased (although it remains a challenge – see Sect. 3). LSM groups have therefore started to optimise a selection of parameters using statistically robust data assimilation (DA) methods.

DA methods are powerful as they allow observational data to be combined with numerical methods to optimise estimates of chosen variables at the time of observations, either to update the state (state estimation) or to optimise internal parameters (parameter estimation) while accounting for uncertainties in both the model and the data (Rayner et al., 2019). However, the distinct requirements of LSMs compared to the atmospheric and ocean components of ESMs result in subtle but important differences in how DA techniques are applied. The atmospheric and ocean components of ESMs rely on fluid dynamic models, where the underlying fundamental laws are relatively well understood, even if complex to simulate, and many of the model parameters are known physical quantities that can be observed. Therefore, DA activities using atmospheric or ocean components of ESMs have thus far been heavily focused on numerical weather forecasting (NWP) 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, parametric and structural uncertainties dominate their spread (Bonan & Doney, 2018; Draper, 2021; Luo et al., 2015). LSM parameters are often linked to biological processes and organismal traits and are dependent on plant functional type (PFT). Therefore, these parameters have a wide range of possible values where they have been measured (in addition to a lack of data on parameters for some PFTs and the role of "effective" parameters as discussed above). Characterising and

simplifying the diversity of life into relatively few parameters is thus a challenge faced in LSM development that is less of an issue for atmospheric and ocean modeling.

Early efforts in global model calibration in the 1990s and 2000s focused on optimising vegetation and carbon cycle parameters of simplified or intermediate complexity land, carbon, or ecosystem models. These studies, such as Knorr & Heimann's (1995) work optimising parameters of the Simple Diagnostic Biosphere Model (SDBM) using site $CO_2$ measurements, laid the groundwork for DA-focused land model parameter optimisation. Knorr & Heimann's (1995) study was followed by further studies constrain carbon flux related processes in simple and intermediate complexity ecosystem models like BETHY (Rayner et al., 2005; Scholze et al., 2007), using frameworks referred to carbon cycle data assimilation systems (CCDASs). Parallel to this, there was significant progress in using local eddy-covariance flux tower measurements to optimise parameters related to photosynthesis, respiration, and energy flux in ecosystem models at the site level (e.g., Moore et al., 2008; Sacks et al., 2006; Y.-P. Wang et al., 2001; Williams et al., 2005). Two key intercomparison projects, OptIC and REFLEX, played a pivotal role in assessing various data assimilation techniques for parameter estimation in simple and intermediate complexity land, carbon cycle or ecosystem models (Fox et al., 2009; Trudinger et al., 2007).

Parameter optimisation of computational expensive land models using DA started in the late 2000s (Medvigy et al., 2009; Rayner, 2010; Santaren et al., 2007). These studies used similar data (*in situ* fluxes and biomass) and similar experimental configurations (site scale optimisations) as past studies with simple and intermediate complexity models but often with different DA methods due to the increase in computational expense of running much more complex models (Sect. 2). Building on the formative DA work with the SDBM (Kaminski et al., 2002) and BETHY models (Rayner et al., 2005), other LSM groups also started using global networks of *in situ* atmospheric $CO_2$ mole concentration data for constraining regional to global scale surface net $CO_2$ exchange (Kaminski et al., 2013; Peylin et al., 2016; Schürmann et al., 2016). Testing of DA configuration at site scale (data type, sampling interval, record length, and combinations of data - e.g., carbon fluxes and stocks or carbon fluxes) continued with all types of land models (Bastrikov et al., 2018; Bloom et al., 2016; Bloom & Williams, 2015; Braswell et al., 2005; Dietze et al., 2014; Keenan et al., 2013; Medvigy et al., 2009; Moore et al., 2008; Ricciuto et al., 2008, 2011; Santaren et al., 2014; Thum et al., 2017; Weng et al., 2011; Weng & Luo, 2011; Wutzler & Carvalhais, 2014; Xu et al., 2006). One example was the emergence of "multi-site" experiments – parameter estimation studies in which data from multiple sites (often grouped by PFT) were included simultaneously in the assimilation, with the retrieved parameters then compared to those from assimilations with only individual site data (see Sect. 3.4 for further discussion). These were initially performed against data from the global FLUXNET network for a range of intermediate and full complexity LSMs, including many LSMs used within ESMs (e.g., Carvalhais et al., 2008, 2010; Groenendijk et al., 2011; Kato et al., 2013; Knorr et al., 2010; Wu et al., 2018; Xiao et al., 2014, ORCHIDEE: Kuppel et al., 2012, 2014; JULES: Alton, 2013; Raoult et al., 2016; Noah: Chaney et al., 2016; CLM: Post et al., 2017). With the advent of satellite products, remote sensing indicators of vegetation dynamics (phenology and photosynthetic uptake) began to be employed to constrain model parameters at various spatial scales, including reflectance (Shiklomanov et al., 2021); vegetation indices

(Migliavacca et al., 2009; NDVI – MacBean et al., 2015), FAPAR (Bacour et al., 2015; Forkel et al., 2014, 2019; Kaminski et al., 2012; Knorr et al., 2010; Stöckli et al., 2008; Zobitz et al., 2014), solar-induced fluorescence (SIF; (Bacour et al., 2019; Forkel et al., 2019; Knorr et al., 2024; MacBean et al., 2018; Norton et al., 2018, 2019; J. Wang et al., 2021), aboveground biomass and burned area (Forkel et al., 2019). Over the past decade, parameter estimation has advanced to constrain the terrestrial carbon, water, and energy cycles simultaneously, driven by new remote sensing data on total column-integrated $CO_2$ fluxes ($XCO_2$), satellite-derived vegetation optical depth, soil moisture, snow cover, and river flow measurements, which have been successfully integrated, for example, into BETHY (Scholze et al., 2016), the new community D&B model developed by the European Space Agency (ESA)'s Carbon Cluster (Knorr et al., 2024); JULES (Pinnington et al., 2018, 2021), and ORCHIDEE (Raoult et al., 2021). Further details on the history of parameter optimisation in all types of land models are provided in Rayner (2010), Kaminski et al. (2013), Scholze et al. (2017), Rayner et al. (2019), Baatz et al. (2021), and MacBean, Bacour, et al. (2022).

While substantial progress in complex LSM parameter optimisation has been made (particularly for constraining parameters of short timescale vegetation dynamics and carbon fluxes, as described above), a number of challenges hindering objective calibration of the full high-dimensional LSM parameter space remain. Despite advances in the use of analytical techniques to dramatically reduce the time for LSM simulations (Luo et al., 2022; Sun et al., 2023), these highly complex models still have computational requirements – even for one global scale simulation – that are too high for efficient multi-site to global DA experiments. This is true even for "offline" simulations (i.e., LSM simulations forced with climate reanalysis data, as opposed to "online" cases when LSMs are run within the whole ESM). High dimensionality and computational cost make it difficult to calibrate LSMs using conventional statistical approaches like Markov Chain Monte Carlo. Methods used with simpler models often fail with LSMs due to their complexity. These challenges have also meant that LSMs currently struggle to fully leverage the large amount of data from ground networks and Earth observation platforms for calibration.

As an LSM community, thus far, we have no overall strategy for how to proceed towards a system that allows for objective parameter estimation. However, this field is rapidly expanding and we are in a unique position to learn from each other, especially in relation to the technical challenges we face with computational expensive LSM parameter DA. Efforts to build a Land Data Assimilation Community (https://land-da-community.github.io/) by the Analysis, Integration and Modeling of the Earth System (AIMES) Land Data Assimilation 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. Capitalising 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 modelling with the help of DA.

In this paper, we summarise the current state of parameter estimation in land surface modelling,

starting with DA methods, before outlining the different challenges and opportunities our community faces. We then highlight how some of these challenges can be potentially addressed by capitalising on emerging ML techniques and increasing computational capabilities. Finally, we propose future priorities for advancing the field given the urgent need for more accurate and precise LSM projections. We focus on the techniques and challenges related to optimising carbon-vegetation-water interactions in full complexity LSMs but also discuss parameter DA and ML methods applied to intermediate complexity land, carbon cycle and ecosystem models.

This paper complements Kumar et al. (2022) which addresses land surface model data assimilation in the context of state estimation, with a focus on vegetation and hydrology processes. A water cycle-focused perspective, tackling both state and parameter estimation, is offered by De Lannoy et al. (2022).

# 2. Data assimilation methods for parameter estimation in land surface models

LSMs have many parameters that need to be calibrated to accurately reflect the real world (ideally based on observations) and to increase confidence in their future projections. Expert knowledge and empirical measurements of some LSM parameters provide approximate values or their respective ranges. However, due to uncertainties in observations and processes, and the conceptual nature of most parameters, the exact values of LSM parameters are inherently difficult to determine. Instead, we make use of the abundance of observational data indirectly related to the parameters via the processes they are related to, and thus the problem of parameter estimation 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* $\mathbf{y} \cong G(\boldsymbol{\theta})$. In the context of parameter estimation, 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 minimisation effort that locates the argument of a cost function that minimises the discrepancy between $\mathbf{y}$ and $G(\boldsymbol{\theta})$. 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 regularisation, is the Bayesian approach. Bayesian methods include information on the prior distribution of the parameters p($\boldsymbol{\theta}$) to define an entire posterior distribution:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (1)$$

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 (MAP) estimate - the argument that maximises the posterior distribution (i.e., its mode) - provides a point estimate for $\boldsymbol{\theta}$ and is equivalent to a loss minimisation estimate regularised with prior parameter information under Gaussian assumptions. Under such assumptions, maximising the posterior distribution corresponds to minimising the so-called variational cost function:

$$J(\boldsymbol{\theta}) = ½[ (G(\boldsymbol{\theta}) - \mathbf{y})^{T}\mathbf{R}^{-1}(G(\boldsymbol{\theta}) - \mathbf{y}) + (\boldsymbol{\theta} - \boldsymbol{\theta}_{b})^{T}\mathbf{B}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_{b})], \quad (2)$$

where $\mathbf{R}$ and $\mathbf{B}$ are the model/data and prior error covariance matrix, respectively, and $\boldsymbol{\theta}_{b}$ are the prior parameter values.

With the emergence of novel ground and satellite observation sets came the advent and development of techniques to implement them in a field of mathematics originally coined Data Assimilation (DA) (Talagrand & Courtier, 1987). Along with the differences in the aforementioned approaches to solving the inverse problem, these methods also 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). There is some confusion in the community regarding the terminology used when describing DA methods, for example, the false dichotomies sometimes used between “variational and sequential” and “optimisation-based versus Bayesian” - these dichotomies have been marred over time with hybridisation and the continual development of the techniques. Rayner et al. (2019) have made a significant effort to harmonise the notation and clarify overlapping terminology within the community.

Although DA is primarily used in numerical weather forecasting to correct the model state, in LSMs, DA is often employed to reduce parametric uncertainty, a process referred to as parameter data assimilation (PDA). Techniques used in numerical weather forecasting can be adapted for parameter estimation in LSMs. One of the key methods is 4DVar, which involves minimising Eq. 2 (called 4DVar to contrast with 3DVar, where the observations are instead compared to a single model output at a time). The next part of this section looks a little deeper into methods used to reduce this cost function, as well as outlining alternative DA methods that extract the full posterior distribution.

**Methods for reducing cost functions:**

Methods commonly used to minimise the cost (e.g., Eq. 2) require numerical optimisation due to their complex structure and these can usually be grouped into local gradient-descent or global random search techniques. Although more computationally efficient, gradient-descent methods require the gradient of the cost function (either exact, which requires differentiating the entire LSM - see Sect. 3.7, or approximated when exact is not possible or desirable) and they can

result in the location of a local minimum. A common gradient-based minimisation method used in LSM parameter estimation 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. These gradients can be obtained by hand or using automatic differentiation software (Gelbrecht et al., 2023; Griewank, 1997). 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 Sect. 3.7)—often necessitate alternatives. To address this, approximate gradient methods can be employed. One approach is to estimate gradients using finite difference, 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. In cases where gradient information 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 towards the minimum of a cost function, eliminating the need for direct gradient calculations. Additionally, more advanced approaches, such as the ensemble-based 4DVar (4DEnVar) algorithm proposed by Liu et al. (2008) use an ensemble of model trajectories to approximate gradient information via a control variable transform.

Alternatively, global search methods can be used to minimise the cost function. These methods use techniques that try to scan the entire parameter space in some defined way to avoid this pitfall but often require heavy computational power to do so. These global search methods can be cateogrised as Monte Carlo (MC), since they are methods that make use of repeated trials (or sampling) generated using random numbers (Owen, 2013). An example of such a 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.

Although these gradient-descent and global search methods are very efficient in finding an optimal point-estimate of the parameters that minimise the given cost function, usually they do not directly offer information about the posterior error statistics. Nevertheless, it is possible to exploit information about the curvature of the cost function (via the Hessian) at the optimum to obtain such information, but this is typically more complicated than deriving gradient information and more costly in the case of global search.

**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($\boldsymbol{\theta}$|**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 optimisation, as opposed to gradient-descent methods, these techniques are often Monte Carlo in nature and hence also derivative-free (black-box).

Techniques include importance sampling (Kloek & Van Dijk, 1978), a relatively simple approach that samples random values from the prior and accumulates accepted parameterisations 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 Monte Carlo (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 (i.e., 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 parallelisation.

Particle filters provide an alternative to MCMC for sampling from the posterior distribution, particularly in time-evolving systems. They represent the posterior using a set of particles, updating them with each new data point. While computationally intensive and prone to particle degeneracy, particle filters are useful for real-time tracking of system states and time-varying parameters. However, many of the parameters in land surface models are linked to biological processes and thus are subject to change over time due to acclimation, phenotypic plasticity, adaptation and evolution. While some attempts have been made to explore the seasonal variability in parameters (Rowland et al., 2014; Verbeeck et al., 2011), the majority of the literature in land model parameter estimation so far operates on the assumption that parameters are fixed in time. As such, particle filters are rarely used in PDA (Speich et al., 2021) (unless part of joint state-parameter DA, for example, Zhang et al., 2017).

**Applications in LSMs**:
Due to the high number of required model evaluations, MCMC methods have primarily been applied to computationally inexpensive land, carbon cycle, and ecosystem models, or to calibrate isolated processes such as fitting parameters of a two-pool model of substrate dependence in plant respiration (Jones et al., 2024) or parameters of the wetlands CH4 emissions module in the second generation dynamic global vegetation model LPJ-GUESS (Kallingal et al., 2024). For example, MCMC methods have been used to estimate parameters of the Simplified PnET (SIPNET) ecosystem model (Fer et al., 2018; M. Liu et al., 2015; Sacks et al., 2006), TECOS (Xu et al., 2006), FöBAAR forest carbon cycle model (Keenan et al., 2012), BETHY (Knorr & Kattge, 2005) and the DALEC suite of intermediate complexity ecosystem models (Famiglietti et al., 2021; Keenan et al., 2011; D. Lu et al., 2017). DALEC is also at the heart of the cutting-edge CARbon DAta MOdel fraMework (CARDAMOM) where the full potential of MCMC-based carbon parameter estimation is performed (Bloom et al., 2016; Exbrayat, Smallman, et al., 2018; Smallman et al., 2021).

While computationally expensive LSMs build on this foundation, their complexity and parameter volume have made MCMC methods computationally prohibitive. Consequently, 4DVar has been the preferred approach for these models. When the tangent linear or adjoint models have been available (e.g., Bacour et al., 2015; Knorr et al., 2024; Kuppel et al., 2012; Raoult et al., 2016; Schürmann et al., 2016), these have been directly used to minimise the cost function and calculate the Hessian. Alternatively, the Nelder-Mead simplex algorithm (Pinnington et al., 2018), finite differences (Bacour et al., 2019; Bastrikov et al., 2018; MacBean et al., 2015) and 4DEnVar (Pinnington et al., 2020) have all been used to circumvent the need of such models. While some Monte Carlo approaches have been used to calibrate complex LSM parameters—either for global search methods to minimise the cost function or to extract the full posterior distribution—these are typically applied at the site scale and fall short of full global calibrations. Examples include the adaptive population importance sampler used to calibrate the JSBACH model (Mäkelä et al., 2019), the genetic algorithm used to calibrate ORCHIDEE (Bastrikov et al., 2018), and multichain MCMC method DiffeRential Evolution Adaptive Metropolis (DREAM(zs)) (Vrugt et al., 2009) used with CLM (Post et al., 2017) and LPJ-GUESS (Bagnara et al., 2019).

# 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 and the prior distributions over which they are allowed to vary. A common first step is to select from the (potentially quite large) number of model parameters, a subset that is deemed the most influential in some sense. The excluded parameters are then fixed at their nominal values, yielding a parameter space of reduced dimension. 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 i) PDA techniques are computationally demanding, scaling with the number of parameters used in the optimisation, and ii) due to the high degree of equifinality in most parameter spaces (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 at equifinality.

Which model output and metric is tested fundamentally affects the crucial parameter selection if relying primarily on sensitivity analysis. Furthermore, parameter sensitivity is often a function of the parameter prior distributions, about which for many parameters we may have poor knowledge. Indeed, a key distinction between a traditional sensitivity analysis, which may vary all parameters by the same arbitrary amount (e.g. +/- 10%), and an uncertainty partitioning analysis is whether the prior distributions accurately represent our knowledge about model parameters prior to calibration (direct data constraints, formal expert elicitation, etc.) (Dietze et al., 2014; LeBauer et al., 2013; Raczka et al., 2018).

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 solution to combat this is to use spatial pattern correlations as a metric for parameter selection to ensure that the parameters selected are not highly correlated (Dagon et al., 2020). More sophisticated methods include using the adjoint model to determine local sensitivities and global sensitivity methods such as Morris (Morris, 1991) and the variance-based Sobol (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 CABLE (Lu et al., 2013), CLASSIC (Deepak et al., 2024), CLM4.5(FATES) (Massoud et al., 2019), JULES (Pianosi et al., 2017), Noah-MP (Wang et al., 2023) and ORCHIDEE (Dantec-Nédélec et al., 2017; Novick et al., 2022). However, these methods can be hard to implement (see Sect. 3.7 for the discussion about adjoint models) or require a large number of model runs (e.g., O(10,000) for Sobol). Nevertheless, once the adjoint or ensemble exists, it is relatively easy to test the sensitivity of different model outputs.

In complex LSMs, even after selecting the most influential parameters, the large number of vegetation (e.g., 15 plant functional types 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 independently. One way to tackle this issue is to assume that the parameter differences among different groups vary proportionally and, therefore, optimise 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 localising parameters rather than using PFT-specific parameterisations may be more appropriate. As such, methods that allow for independent tuning of parameters within each PFT, or even localisation of parameters, may be necessary. Scaling factors can also be used to target processes without needing to deeply explore detailed parameterisations (e.g., Raoult et al., 2021) used a factor to scale the bare soil resistance to evapotranspiration parameterisation in ORCHIDEE).

Selecting parameters is only one part of the problem - choosing the prior distributions is equally important. In the existing LSM calibration literature, it is very common to assume uniform prior

distributions, either explicitly within Bayesian calibrations or implicitly when selecting uniform range restrictions within parameter estimation using a naive objective function (unlike, for example, classic variational DA techniques such as 4DVar which use an explicit Gaussian prior). In these cases, uniform ranges are often based on informal “expert judgment” or ad hoc trial and error. In some cases, parameter uncertainty ranges can be obtained from *in situ* measurements, such as the TRY database (Kattge et al., 2020). Alternatively, the range can be set based on the operational value of the parameter (e.g., ±20%) - although this should only be done as a last resort. When selecting ranges, extra considerations are needed to ensure that the ranges make physical sense (e.g., not sampling negative values if the parameter needs to be positive), that parameter dependencies are maintained (e.g., two parameters whose ratio should not surpass a given threshold, or multiple parameters that must sum to one) and that plausible relationships are retained (e.g., longevity of wood should be longer than that of foliage).

While uniform distributions are frequently chosen due to the lack of a more specific prior distribution, and often to ensure the range is broad enough to cover edge cases, this approach has significant drawbacks. Uniform priors rarely represent our actual prior knowledge of a system, as they imply that all values within a range are equally likely, but values even a little bit outside that range are impossible. In practice, parameter values in certain parts of parameter space are often known a priori to be more plausible than others. An alternative to assuming uniform prior distributions is to select from any of a plethora of other distributions, with such choices usually driven by a combination of structural constraints (e.g., using zero-bound distributions for non-negative parameters), formal syntheses and meta-analyses of trait data, and structured expert-elicitation exercises (Dietze, 2017; Dietze et al., 2014; LeBauer et al., 2013). However, selecting an inappropriate distribution can be as problematic as using a uniform distribution, especially given that the true prior distribution is often not well known at the start of the calibration process. This highlights the importance of conducting formal prior predictive checks to validate assumptions before proceeding.

Priors constructed from trait data, where available, can often be quite well constrained, acting as a form of data fusion (i.e. combining multiple constraints) and helping to constrain subsequent calibrations to biologically-plausible parts of parameters space. Indeed, accounting for prior trait knowledge can lead to very different conclusions about what parameters need to be included in a calibration, as there are cases where very sensitive parameters may be well constrained a priori (e.g., the parameter controlling the maximum rate of carboxylation - Vcmax) while in other cases much less sensitive, but unconstrained, parameters may plausibly span multiple orders of magnitude and thus contribute more to overall model predictive uncertainty (Dietze, 2017; LeBauer et al., 2013).

Informative non-uniform priors do not have to assume parameter independence; multivariate priors can be constructed to capture known correlation structures and trait trade-offs, both within- and across-PFTs (Shiklomanov et al., 2018). However, quantifying these correlations can be a challenge, and so error covariances are often omitted in PDA, neglecting natural parameter relationships. This simplification can result in an ill-posed inversion problem.

Finally, adopting informative non-uniform priors makes it easier to take advantage of the iterative nature of Bayesian inference, where the posteriors from one round of model calibration can be used as priors in the next round without requiring the recalibration of models to earlier data constraints. Not only does this greatly simplify the updating of model calibrations as new data becomes available, but it offers considerable computational advantages.

It is important to stress that no matter the method used for parameter estimation, solutions only exist in the parameter space defined by the parameter selection and authorised prior ranges (Williamson et al., 2013). Changing the number of parameters, their prior distributions, and/or the model process representation will require new calibrations since the solution may differ due to new parameter interactions and the equifinality of solutions.

## 3.2 Characterisation 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 characterising 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 (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 modelling step from the raw data measurement to any given physical quantity (see Sect. 3.3). Due to the difficulty in separating model and observation errors, they have often been combined in past studies. In fact, the mathematical formalisation commonly used in PDA assumes that observation errors include model errors, thereby treating model discrepancy as part of the observational error.

Although common, combining model error with data 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.

There are a number of arguments for keeping process and observation error distinct. Model process error propagates in space and time when making predictions, while observation error

does not. Additionally, addressing a large process error requires improving the model structure, while addressing a large observation error calls for improving data quality. Furthermore, calibrating models using cost functions that rely solely on fixed a priori observation errors can distort parameter uncertainty estimates as well as the relative weight assigned to different data constraints, as there's often no inherent reason to assume that model skill at predicting a variable is proportional to the accuracy of its measurement. Indeed, it is easy to point to examples where the uncertainty in our ability to model something differs in rank order from our ability to measure that same thing (e.g., at local scale, model predictions of net ecosystem exchange (NEE) are more uncertain than gross primary productivity (GPP: the flux of carbon absorbed into the land surface due to photosynthesis), but observations of GPP are more uncertain than 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 data assimilation 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 modelling 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 data assimilation (Chevallier, 2007; Kuppel et al., 2013). 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 datasets, random errors tend to average out, leaving systematic errors to dominate. These errors have long been recognised 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 appreciated. 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 (Kennedy & O'Hagan, 2001; Oberpriller et al., 2021). Additionally, hybrid models that integrate machine learning with process-based models are being explored as a means to address these challenges (see Sect. 4.2).

Ultimately, interconnected efforts, such as the characterisation of 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 in providing ways forward for discerning errors in data from those in model structure.

## 3.3 Developing observation operators

The term “observation operator” refers to any transformation of the modelled quantity used to allow comparison against observations (Kaminski & Mathieu, 2017). Note that what are often called observations are themselves complex transformations of raw data measurements used to estimate physical quantities comparable to the LSM output. For example, radiances observed by a satellite at the top of the atmosphere can be translated into any number of land surface data products, such as leaf area index. This processing can also be seen as a complex model, such as the inversion of a radiative transfer scheme. Furthermore, these data are usually prepared in such a way that they are available on the model grid.

In some cases, it is possible to assume a one-to-one relationship between the model output and assimilated data, in which case the observation operator is the identity matrix. However, in all other cases, an observation operator is required for DA, and the choice of observation operator can significantly impact the results (Cooper et al., 2019). A common use of an observation operator is to bridge the spatial scale between model and observations, either by aggregating the gridded observations to the resolution of the model or vice-versa (Pinnington et al., 2021). More complex examples of spatial scaling operators utilise a weighted averaging process to match a more detailed description of the observation, such as modelling the point spread function of satellite data, or the footprint of an eddy-covariance flux measurement. For example, Vergopolan et al. (2020) introduced a cluster-based observation operator that maps the Gaussian footprint of satellite observations to the sub-grid scale of high-resolution LSMs. This enables efficiently assimilating coarse soil moisture observations while bridging the spatial scale mismatch with fine-scale LSMs and ground observations (Vergopolan et al., 2021). In an application with flux tower data, Pinnington et al. (2017) partitioned the fluxes to observe different parts of the forest and run separate assimilation experiments for logged and unlogged forest stands.

In another example, atmospheric transport is used to map surface fluxes of gas species, such as $CO_2$, into atmospheric concentrations of that species at sampling points. In this way, flask measurements of $CO_2$ have been used to constrain parameters in models of the terrestrial biosphere (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) and to evaluate simulated net $CO_2$ fluxes after optimising against eddy-covariance data (Kuppel et al., 2014). For non-reactive species, it is sufficient to have data on winds 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.

Observation operators are also used to predict observed quantities that are not directly computed by the model itself. A recent example is the assimilation of SIF data, which is typically assumed to be a proxy for GPP. Examples of SIF observation operators include simple linear relationships with GPP (Bloom et al., 2020; MacBean et al., 2018) through to more complex operators based on the underlying photochemistry and radiative transfer in the canopy, either using empirical simplifications of those processes (Bacour et al., 2019) or using fully

mechanistic models for the operator (Norton et al., 2019). Another example is vegetation optical depth which has been used to constrain above-ground biomass and leaf area index (Scholze et al., 2019).

Scholze et al. (2016, 2019) also developed observation operators to map surface soil moisture (SSM) retrievals to simulated volumetric soil moisture of the surface layer of BETHY, which were also used by Wu et al. (2018, 2020, 2024). SSM is subject to large biases, which therefore necessitates this type of transformation. Numerous models employ methods to map SSM to the climatology of their model, for example through cumulative density function (CDF) matching. Another approach is to focus solely on dynamics (e.g., dry downs, Raoult et al., 2021). The dynamics approach is often used when assimilating vegetation indices, FAPAR or leaf area index (LAI) — retrievals are normalised to estimate the seasonality of phenology instead of the absolute values (MacBean et al., 2015). The optimisation then focuses on a reduced set of phenology-related parameters, rather than including those related to photosynthesis (Bacour et al., 2015).

Forward modelling of remote sensing data — i.e., the process of simulating remote sensing data directly from the LSM outputs rather than assimilating processed satellite products — like in the example of SIF, is the opposite approach to the assimilation of high-level satellite products such as LAI or GPP. A key argument for taking this approach is that assumptions in the retrieval process used in these products are likely inconsistent with the assumptions embedded in the land surface model they are being assimilated into. A clear example of this is the use of satellite GPP products which typically employ a production efficiency approach (e.g. the MODIS GPP product, Running et al., 2021) whereas land surface models often use limiting-rate enzyme kinetic schemes derived from those of Farquhar et al. (1980) and Collatz et al. (1992). Furthermore, satellite-derived GPP estimates typically use environmental drivers such as downwelling shortwave radiation which will almost certainly differ from those used to drive the land surface model they are being assimilated into. Finally, there are often substantial differences between the satelitte-derived estimates (e.g. of GPP or LAI) where the assimilation of any one product is likely biased with respect to the ‘truth’ (which is the primary reason for using the seasonal dynamics rather than the actual values of time series data, as discussed in the previous paragraph). Consequently, discrepancies between these high-level observations and the values of the same variables predicted by a LSM may differ due to these factors and be non-trivial to characterise.

It is appealing, therefore, to assimilate low-level products like SIF or canopy reflectance (Quaife et al., 2008). For canopy reflectance, this typically requires the use of radiative transfer models and is analogous to so-called “radiance assimilation” which is used extensively in numerical weather prediction. In that way, any systematic error between the model and the observations can be attributed to the land model (including the radiative transfer model) itself. For example, Shiklomanov et al. (2021) modified the existing canopy radiative transfer model in the Ecosystem Demography v2 model (ED2) to predict full hyperspectral waveforms, instead of just aggregate visible, near-infrared, and thermal bands, and then used this observation operator to calibrate ED2 against airborne AVIRIS imaging spectroscopy across the eastern temperate US.

Meunier et al. (2022) later used this observation operator in the development of a novel tropical liana PFT. However, low-level satellite products often exhibit variability across domains that are not inherently resolved by the land model, leading to some level of compromise between i) adding complexity to the land model, ii) having an observation operator that is not completely consistent with the underlying model or, iii) accepting that some of the variability in the observations themselves will not be resolved. In the examples of SIF and canopy reflectance, both vary with the relative geometry of the sun and sensor - correctly capturing that directional variability using an observation operator that is physically consistent with the description of the radiative transfer regime implemented in global land surface models (which typically only predict total fluxes, i.e. integrated across the viewing hemisphere) is not currently possible. Nevertheless, the selection and processing of observation data can help mitigate some of these issues. For example, space-time binning of space-borne SIF data across multiple observation geometries can limit the impact of directional effects and potentially increase the consistency between model assumptions and the observed variables.

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 machine learning may offer a unique opportunity within DA applications, as discussed in Sect. 4.3.

## 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” optimisations, grouping sites and performing a single optimisation 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 optimisations (Kuppel et al., 2012; Raoult et al., 2016). Another approach is to average the results of single-site optimisations. While usually less effective than multi-site optimisations, 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 optimised parameters improved simulated land-surface temperature performance.

Both these approaches can be thought of as end-members (all sites the same versus all sites different) in a continuum representing the statistical independence of calibrations across sites. While only just beginning to be utilised to calibrate ecosystem models (Dokoohaki et al., 2022; Fer, Shiklomanov, et al., 2021), hierarchical models have a long history of use in ecology as a way of capturing this continuum, allowing parameters to vary across space and through time,

but constraining that variability with multivariate statistical models that describe that variability. Since the across-site and within-site calibrations are fit simultaneously, this would allow LSM models to “borrow strength” across sites (e.g., reducing equifinality as described above) without forcing parameters to be the same everywhere. Hierarchical models also provide a formal framework for accounting for the fact that out-of-sample predictions are more uncertain (because their parameter vectors need to be predicted) than in-sample predictions at sites where parameter vectors are known. To date, existing hierarchical ecosystem model calibrations have assumed a simple “random effects” structure (i.e. different sites are drawn from the same across-site distribution), but there are important opportunities to explore hierarchical models with across-site spatiotemporal covariances (i.e., sites closer together should be more similar) and across-site covariates (i.e., parameters that explain, and help predict, parameter variability).

A further alternative is the use of intermediate complexity models (e.g., DALEC), which, due to their reduced computational complexity, can retrieve parameters at the pixel scale utilising 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, Smallman, 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 optimised parameters can be inserted into more complex models, enhancing their performance and helping to better understand their internal dynamics (Caen et al., 2022).

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* datasets.

## 3.5 Dealing with large and multiple observational datasets

Although EO instruments can provide global gridded datasets with which to calibrate the models, fully exploiting these opportunities is challenging. Running experiments at the same resolution as the satellite products (e.g., 500m MODIS resolution; Justice et al., 2002) requires a lot of computational power and time, and we do not always have access to matching meteorological forcing data. The resolution of products to be assimilated may also not be meaningful for the objectives of the experiment. Additionally, when assimilating more than one remote sensing data constraint, we must address multiple competing resolutions. This requires decisions about scaling (see Sect. 3.3), determining which products are to be upscaled (aggregated) versus downscaled (interpolated). Generally, satellite products are scaled to

match the chosen model grid, usually dictated by the resolution of the forcing data, although this scaling can result in an over-generalisation or loss of information.

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 optimisation (e.g., MacBean et al., 2015), although defining what is representative is a challenge in itself. Once selected, the representative pixel approach helps to i) reduce the dimensionality of the problem, allowing for efficient and multi-data-stream calibrations, ii) focus on points with close to homogenous coverage to be able to calibrate class-specific parameters (e.g., plant functional types), and iii) define a different evaluation set of pixels with which to assess the optimisations, especially sites with additional ground data. After selecting representative pixels, multi-pixel optimisations are performed (as described in Sect. 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 optimisation 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 optimisations are becoming paramount to provide adequate constraints since parameters are likely to impact different parts of the model. By selecting only one specific data stream in an optimisation, we risk degrading the model's overall predictive capacity if some of the optimised 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 optimisation, 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 optimisations 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 and the characterisation of error correlations between them (Bacour et al., 2023). Nevertheless,

it must be stressed 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 Sect. 3.2).

## 3.6 Including the spin-up and transient historical period in the assimilation to better constrain land carbon sink projections

Many LSM simulations include both a spin-up phase that brings the prognostic variables including vegetation state, soil carbon pools, and soil moisture content into equilibrium prior to the industrial revolution (c. 1750). This is followed by a transient historical simulation where the model is driven by changing climate forcing, rising $CO_2$ levels, nitrogen deposition, and prescribed land management and land cover change since the equilibrium time point up to the present day. Even with transient forcings, this historical period is likely not accurately simulated, in part due to the lack of accurate historical climate and land use forcing data, in part because “slow” carbon cycling parameters (e.g. carbon allocation or turnover rates) that control the magnitude of the equilibrium carbon stock are poorly constrained, and in part because the effects of key global change drivers on carbon storage (including recovery from disturbance) are often missing or not reliably represented in models. The result is a large spread in the magnitude and dynamics of various carbon pools and fluxes which underpin the current and future projections of the land carbon sink (Arora et al., 2020; Friedlingstein et al., 2023).

To obtain reliable estimates of the current or future land carbon sink and trend in atmospheric $CO_2$ we need accurate simulations of global carbon stock trajectories (i.e., *changes* in carbon stocks). The trend in carbon stocks depends on the magnitude of carbon stocks post spin-up, which in turn is strongly controlled by soil carbon pool turnover rates (Exbrayat, Bloom, et al., 2018) (in addition to other parameters involved in soil carbon decomposition that moderate that turnover rate). This is because for the CENTURY type model (Parton et al., 1987) used in many LSMs, heterotrophic respiration is partly dependent on the size of carbon stocks. Global sensitivity analyses (Sect. 3.1) of soil carbon cycle models performed for multiple different biomes worldwide have rarely been performed (though see Huang et al., 2018) due to the computational expense of running long-timescale simulations needed to model carbon stock trajectories. For the same reason, relatively few past parameter DA studies with computationally expensive LSMs at multi-site or global scale have included these slow-acting carbon cycle parameters in their assimilation experiments. However, we know from past DA studies that optimising “fast” carbon cycle flux related parameters related to photosynthesis, phenology, and ecosystem respiration has limited impact on regional to global scale carbon stocks (MacBean, Bacour, et al., 2022), as expected, while “slow” carbon cycle process parameters (such as those related to carbon allocation to different biomass pools, or biomass and soil carbon pool turnover times) are important for constraining long-term carbon stock trajectories (Thum et al., 2017).

To optimise the “slow” acting carbon cycle parameters involved in carbon allocation, biomass turnover and soil carbon cycling, LSM assimilation experiments would need to include the spin-up and transient runs in the assimilation, which would be prohibitively costly given the computational cost of LSM runs. Therefore, neither the spin-up or transient period (prior to the assimilation window) are usually included in LSM assimilations (Peylin et al., 2016; Raoult et al., 2016; Schürmann et al., 2016). This presents challenges for obtaining accurate model estimates of carbon fluxes and stocks because an incorrect initial carbon stock will likely result in biased parameter retrievals that are accounting for the model errors contributing to the incorrect initial carbon stock. Note this is not the case for carbon cycle and ecosystem models that have much faster run times and who have therefore been able to include biomass and soil carbon turnover rates and other related “slow” carbon cycling parameters in their optimisations (e.g., CARDAMOM-DALEC – Bloom et al., 2016).

To make up for incorrect carbon pool magnitudes and the fact that including spin-up and transient in the assimilation is not yet feasible, most past carbon cycle parameter DA studies have included scalars on the initial C pools in the optimisation, resulting in an improved fit to NEE and atmospheric $CO_2$ data (e.g., $\eta$, Carvalhais et al. (2008, 2010); $K_{soilC}$ in ORCHIDEE PDA studies, e.g., Peylin et al. (2016); $f_{slow}$ in CCDAS studies, Castro-Morales et al. (2019); Schürmann et al., (2016)). These scalars alter the initial carbon pool size to account for model and forcing errors mentioned above that contribute to incorrect soil carbon stock sizes. Studies differ in how many such scalars to include, both in terms of which carbon pools to relax (all C pools as in Santaren et al. (2007) versus slow and/or passive as in Peylin et al. (2016), whether to scale aboveground biomass or not (Carvalhais et al., 2010), and to how many to use spatially in global simulations (1 in CCDAS, Castro-Morales et al. (2019), Schürmann et al. (2016), versus 30 regional factors used in ORCHIDEE studies, Bacour et al. (2023), Peylin et al. (2016)). Other options for avoiding spin-up include directly initialising models with carbon stock observations, and including parameter calibration within iterative state DA approaches. 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.

Adjusting initial carbon stocks without optimising the “slow” carbon cycle parameters to which the equilibrium carbon stock 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 is insufficient. In longer runs (up to 2100 or 2300) those “slow” carbon cycling parameters that resulted in the original incorrect carbon stock magnitude will start to push the model back to that original (inaccurate) equilibrium, resulting in an artificial trend in the modelled carbon pools (and resultant biases in carbon fluxes and land carbon sink estimates). Thus, for long term projections of carbon-climate feedbacks, all parameters that are important for carbon pool trajectories need to be included in the assimilations. This means that longer time windows (lasting several hundreds to thousands of years) governing the periods over which these “slower” carbon cycle parameters operate will

need to be included in the assimilation experiments (Raiho et al., 2021; Thum et al., 2017). This will materially increase the computational cost of an experiment enough to be prohibitive for computationally expensive LSMs with current simulation protocols and assimilation algorithms. Methods for increasing the simulation speed (e.g., model emulation - see Sect. 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 data assimilation (Hararuk et al., 2014; Huang et al., 2018; Luo et al., 2022; Tao et al., 2020, 2024). Intermediate complexity ecosystem models may be able to assist by providing much constrained priors of soil carbon pool turnover times (and other parameters to which equilibrium/initial carbon stock magnitude are sensitive) (Bloom et al., 2016).

This problem is specific to long-term, slowly changing carbon (and other nutrient like nitrogen and phosphorus) stocks: e.g., for water storage (e.g., soil moisture), usually only a few years are required either for spin-up or to adjust to a given perturbation. Therefore, for hydrology-focused simulations both the spin-up and historical period spanning the perturbation from equilibrium can be included in the experiment. In fact, by including this shorter spinup, the assimilation also gives an estimate of the initial state (e.g., soil moisture, Pinnington et al., (2021); snow albedo, Raoult et al. (2023)). While carbon cycling is interlinked with water and energy cycles, long-term carbon stock trajectories are insensitive to short-term fluctuations in soil moisture.

In addition to longer assimilation time windows, assimilating measurements of aboveground biomass or soil C stocks in conjunction with carbon fluxes provides 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, and 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 datasets 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) in addition to nighttime and soil respiration data that so far have typically not been utilised 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 (Raczka et al., 2018). 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 minimisation algorithms

To perform optimisations effectively, careful consideration must be given to the choice of algorithm and its implementation. As discussed in Sect. 2, various algorithms are available, each with distinct characteristics, such as local versus global optimisation, each having different computational demands. 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 (6 options), crossover (5 options), and mutation (3 options), resulting in 90 possible combinations. Users can also adjust crossover and mutation probabilities. The success of the optimisation process greatly depends on how the optimisation 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 Sect. 4.1) rather than an LSM to find an optimal experimental design (Dagon et al., 2020); once the design has been identified, the optimisation 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, the tangent linear and adjoint models can take years to develop, even when using automatic differentiation software, since the code first needs to be cleaned and structural adjustments need to be made to ensure the code is differentiable without changing the fundamental physics. For example, this may require replacing look-up tables with their continuous formulations and reformulating minimum and maximum calculations to allow a smooth transition at the edge (Schürmann et al., 2016). The years taken to derive the tangent linear/adjoint models mean they quickly become outdated, especially with big community models like JULES and ORCHIDEE, where new processes are added approximately every six months. For JULES, the adjoint was developed for v2.2 of the model (Raoult et al., 2016), whereas JULES is currently at v7.3 at the time of writing. Similarly, while the tangent linear exists for ORCHIDEE, it exists for an old version of the model (AR5) that predates the addition of a multi-layered soil hydrology scheme and nitrogen cycle. To address this issue, the ORCHIDEE DA team has been developing a tool to do the required preprocessing of any version of ORCHIDEE so the tangent linear version of the model can be easily derived using Transformation of Algorithms in Fortran (Giering, 2010). On the other hand, BETHY's lower complexity has allowed it to be kept compliant with automatic differentiation software for decades, which provided efficient derivative code of the up-to-date version of the model. This is also the case for its successor D&B (Knorr et al., 2024), which is the model component of the European Space Agency supported TCASS system, and for the Nanjing University Carbon Assimilation System (NUCAS, Zhu et al., 2023). Alternatively, models written directly in an auto-differentiable language (Julia or python-JAX; see Sect. 5.4) alleviate this issue (Gelbrecht et al., 2023; C. Shen et al., 2023). Although these languages have slower computational performance than Fortran, these new languages often also facilitate the use of graphic processing units (GPU), e.g., through packages like pyTorch (Paszke et al., 2019).

As discussed in Sect. 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 (Pinnington et al., 2020) or the use of emulators, which can be used to either avoid gradient-based approaches in favour of Monte Carlo 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 land surface models and rigorously diagnose their uncertainties. We now have access to large observational datasets at high spatio-temporal resolutions and increased computational capacity and efficiency. These factors, combined with recent advances in machine learning (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 help us address the current challenges and limitations in land PDA outlined above, as well as areas where ML has the potential to improve the DA workflow (Fig. 1). With the large number of studies currently being published in the field of machine learning, we only provide a short overview of the relevant literature. In the context of ML for PDA, we can broadly group the existing studies and applications into four categories: i) the use of ML to emulate the relationship between LSM parameters and its outputs or performance (Sect. 4.1), ii) the creation of ‘hybrid models’ in which ML replaces or complements a component of a larger LSM (Sect. 4.2), iii) the use of ML to improve or pre-process observation datasets prior to their use in PDA (Sect. 4.3), and iv) the use of ML to optimise the parameter estimation process itself (Sect. 4.4).

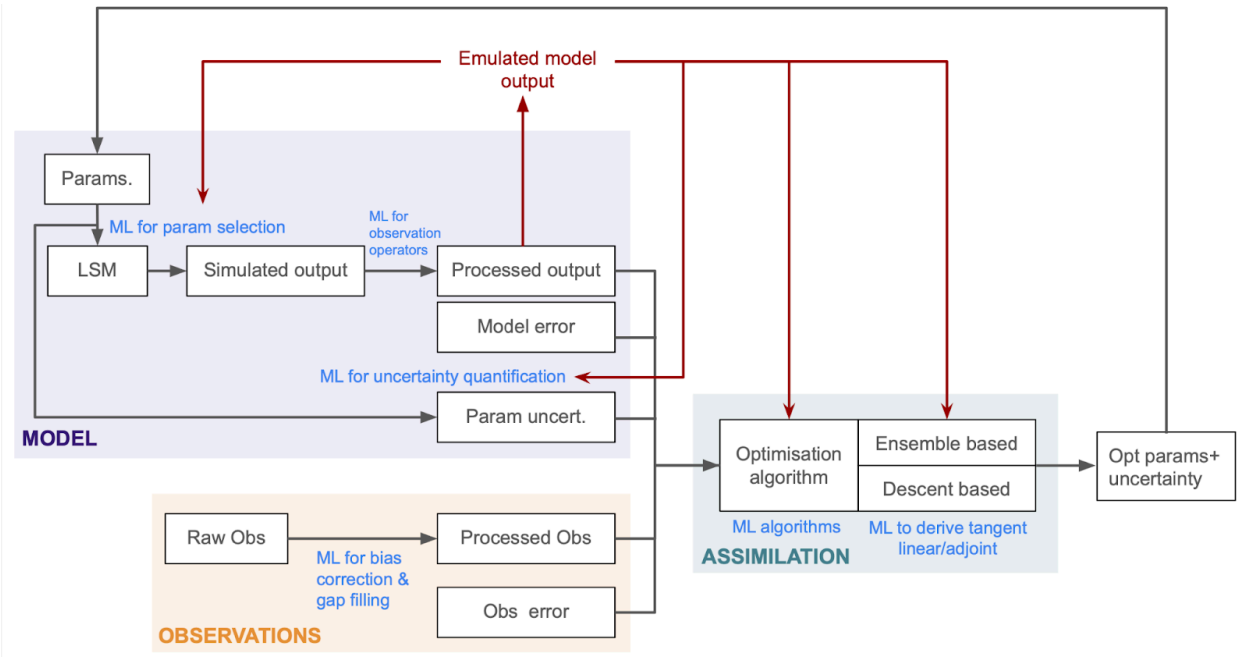


**Figure 1**: Examples of where ML can facilitate each part of the land surface model PDA workflow

## 4.1 Parameter perturbation emulators

The computational cost of high-complexity LSMs hinders the use of the more computationally demanding PDA techniques such as MCMC. However, machine learning methods can mitigate a portion of these computational burdens. By building a 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).

**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, see McNeall et al. (2024) and Kennedy et al. (2024) for PPEs constructed for JULES and CLM, respectively). 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, it can be more cost effective to use a non-random and targeted sampling strategy, such as active learning which tries to optimise the selection of the next sample (e.g., 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 optimise 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.

**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 itself (i.e., the response surface of model error as a function of parameter value) which is much lower dimensional and often much smoother that the model output itself (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 dataset, they are less feasible for larger datasets. One option is to develop sparse Gaussian processes, as demonstrated by Baker et al. (2022). Running JULES at a 1km resolution over Great Britain, they exploit the fact that LSMs typically do not exchange information laterally between grid cells (river routing is generally done as a separate step) to select a subset of coordinates representative of different parameter settings and forcing data regimes.

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 datasets without performance degradation (Shwartz-Ziv & Armon, 2022). (D. Lu & Ricciuto, 2019) used singular value decomposition with Bayesian optimisation to create a reduced number of surrogate models for carbon modelling parameter perturbation. Their approach showed minimal accuracy loss, making it effective for extensive parameter space exploration and uncertainty quantification. Other examples of NNs used to emulate LSMs include, Dagon et al. (2020), where a series of artificial feed-forward NNs were trained to emulate CLM5 output given important biophysical parameter values and Meyer et al. (2022), where an NN was trained to emulate the ensemble mean of several urban LSMs combining the strengths of the different into one ML model. While artificial NNs do not provide a probabilistic prediction, new methods are emerging such as neural processes (e.g., (Garnelo et al., 2018) or randomised 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 (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 (LSTM) methods, which for example have been applied to ECLand (Boussetta et al., 2021), 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 for example, (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 (e.g., Datta & Faroughi, 2023; Guo et al., 2021; Wesselkamp et al., 2024).

**Computational Cost Reduction**

Once an emulator is trained it becomes computationally feasible to apply PDA techniques that require a large number of samples from a prior parameter distribution, e.g., MCMC. Fer et al., (2018) showed how emulators sped up an MCMC optimisation for the relatively simple SIPNET model by over two orders of magnitude (>100x). Further applying their method to the more complex Ecosystem Demography model v2 (ED2), whose complexity precluded it from a direct application of the MCMC methodology for parameter tuning, they found that emulators helped achieve a >20,000x increase in speed (27 hr versus a predicted 74 years by traditional MCMC). Similarly, Sawada (2020) and Cleary et al. (2021) both used emulators to perform Bayesian inversion using the otherwise costly MCMC approach to sample the approximate posterior parameter distribution after calibration. Torres-Rojas et al. (2022) combine surrogate modelling with a multi-objective Pareto efficiency analysis to infer LSM's optimal subgrid parameters at 1% of the computational cost. The emulators were trained on forward model runs used to initially calibrate the model using Ensemble Kalman sampling - a derivative-free optimisation method. 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).

**History Matching**

Emulators are commonly used in the field of uncertainty quantification, and one key method from this field that is gaining traction in land surface modelling is the so-called history matching (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). 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), it is arguably less dangerous to get them wrong here than in the DA case - if the errors are overestimated, HM gives a clear diagnostic of this being the case, for example, 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, in other optimisation approaches, we would still get a solution even if one does not exist. HM also allows the user to

test many different metrics to see if parameters can capture specific features, similar to multi-objective optimisations, giving a clear diagnosis of model structure error. HM 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 optimisation.

## 4.2 Hybrid modelling

ML can also be used in a hybrid modelling approach to substitute components of the physical model with an ML approximation (Eyring et al., 2024). 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). ML and process models can be combined in a number of different ways, including i) substituting a specific model parameterisation with an ML approximation, ii) deriving spatial parameterisations that better capture observed physical behaviour, iii) training on model-data residuals to predict process-model biases and characterise structural errors, and iv) replace computationally costly parts of the model. Hybrid modelling 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 modelling (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). As with all parameter estimation methods, hybrid modelling can be subject to parameters compensating for model structural errors or errors in parameters outside the calibration set (see also Sect. 4.2). This can be counteracted through the use of multivariate independent observation constraints in the calibration.

**Substitution of Uncertain or Missing Parameterisations and Processes**

In the context of land DA, hybrid modelling has been used to improve the representation of complex processes, such as the representation of human processes and their impact, which are often not represented in their full complexity or missing completely in traditional LSMs. 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 can effectively account for human processes by mapping observations into the model climatology (thus removing global biases). At the same time, they can retain the independent information on human processes that is inherent in the observations but typically removed in traditional bias correction approaches (e.g.,

Kumar et al., 2012). Kolassa et al. (2017) used an artificial NN observation operator trained on brightness temperature observations from the Soil Moisture Active Passive (SMAP) mission and GEOS land model outputs to assimilate soil moisture information, which introduced the impact of irrigation and tile drainage in a model that does not normally represent these processes. Assem et al. (2017) developed a Deep Convolutional NN, trained on historic water flow and water level observations, to predict water flow in urban areas from runoff estimates generated by a physical LSM. Hybrid modelling 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.

**Improved Spatial Parameterisations**

Hybrid modelling techniques have also been used successfully to generate model parameterisations that better capture the parameter spatial distribution and thus the observed physical behaviour (Tao et al., 2020, 2024). Process-model parameterisations 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 modelling can address these issues by mapping environmental variables into model parameters or using high-resolution, high-fidelity model simulations to derive new parameterisations for coarse-resolution models (e.g., Gentine et al., 2018). Bao et al. (2023) replaced the traditional PFT-based parameterisation of a light use efficiency model with an ecosystem-property-based parameterisation 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 (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.

**Model Error Identification/Characterisation**

Additionally, hybrid modelling 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 characterise 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.

**Computational Cost Reduction**

Finally, hybrid modelling can be used to replace computationally costly parts of the model. For example, emulating the spinup, which can account for up to 98% of computational time in complex LSMs, would greatly alleviate challenges linked to this bottleneck (see Sect. 3.6). A successful undertaking by Sun et al. (2023) showed how bagging decision trees (an ensemble ML method based on (Breiman, 1996) could be used to emulate the spin-up of the ORCHIDEE LSM. Koppa et al. (2022) developed a deep learning-based hybrid model combining a process-based land surface model with remotely-sensed observations to estimate global evaporation. They showed how hybrid models can significantly improve predictive accuracy while reducing the computational cost.

**Data Requirements**

Hybrid modelling has the potential to be very powerful, but it is also susceptible to issues linked to equifinality (Kraft et al., 2022; Sawada, 2020). 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.

## 4.3 Observation Processing

There are many examples of using ML to improve or pre-process the observational datasets that can be assimilated into LSMs, especially from the field of remote sensing. Many of these novel datasets have yet to be exploited in the LSM parameter estimation studies, presenting exciting new opportunities.

**Observation Operators**

One such application is the use of ML-generated observation operators to translate satellite-observed radiances into model states or parameters (see challenges raised in Sect. 3.3). The use of ML techniques in this context has several advantages: i) ML-based observation operators are relatively simple to implement compared to physically-based approaches, which often involve the inversion of radiative transfer models, ii) they are able to easily accommodate the simultaneous assimilation of multiple observation types, iii) they can inherently correct climatological biases between model and observations, and iv) they facilitate the assimilation of radiance observations rather than retrieval products, thus reducing errors stemming from possible inconsistencies between retrieval algorithm assumptions and models. Due to these advantages, ML-based observation operators have been applied in several land data assimilation studies, including for soil moisture (Kolassa et al., 2017; Rodríguez-Fernández et al., 2019), leaf area index (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).

**Retrieval Algorithms**

Similarly, ML approaches have been used to develop data-driven retrieval algorithms in cases where physical retrieval algorithms are very complex. For example, Chen et al. (2022), Gentine

& Alemohammad (2018), Shen et al. (2022) and Zhang et al. (2018) each used ML to estimate SIF from MODIS radiances, OCO-2, and TROPOMI observations, respectively. Alemohammad et al. (2017) developed an ML approach to retrieve global, monthly GPP estimates from GOME-2 SIF observations only.

**Gap-Filling**

ML approaches can also be used to improve observation datasets by making them more suitable for data assimilation applications. One approach is to use ML to generate gap-filled observations or generate higher temporal resolution datasets. For example, Yatheendradas & Kumar (2022) used an ML approach to create a gap-filled, high-resolution dataset of observed snow cover fraction and Fang et al. (2019) used a deep learning Long Short-Term Memory framework to predict daily "SMAP Level-3 like" soil moisture estimates from atmospheric forcing data and static physiographic attributes. Vekuri et al. (2023) used extreme graident boosting to gap-fill eddy covariance data reducing the northern biases in the data found after using more traditional gap-filling methods. Nevertheless, one must exert caution when using gap-filled data (or other model-derived data, such as retrieval products) for parameter estimation, since they are dependent on the assumptions of the selected gap-filling method. Furthermore, gap-filled data can artificially inflate sample size, which leads to falsely precise parameter estimates.

**Upscaling**

Another approach is to use ML to map local observations to the global scale to mitigate representativeness issues that can arise from the assimilation of local observations. For example, studies by Beer et al. (2010), Joiner et al. (2018), Jung et al. (2011) and Tramontana et al. (2016) all have used ML approaches in combination with remote sensing observations to generate global estimates of carbon and energy fluxes from local flux-tower observations. Vergopolan et al. (2021) used a high-resolution LSM and an ML Bayesian merging scheme trained on in-situ soil moisture data to learn LSM and SMAP satellite biases and obtain 30m satellite-based soil moisture estimates over the contiguous United States. One caveat to using ML to upscale point observations is that large discrepancies can exist between different data products based on the same observations, highlighting the need for thorough evaluation and uncertainty assessment of ML-based products.

**Derived Quantities**

Finally, ML can be used to improve the algorithms used to generate observation datasets. For example, Tramontana et al. (2020) used a combined neural network approach that accounts for the influence of soil property and micrometeorological drivers to generate improved estimates of the partitioning of observed NEE into GPP and ecosystem respiration (RECO), while Zeng et al. (2022) used an ML approach to separate the natural and anthropogenic contributions to satellite-estimated evapotranspiration.

## 4.4 Optimisation process

Since optimisation is a key component to both ML and DA, there are many algorithms common to both fields including gradient-based and evolutionary algorithms (Sect. 2). Indeed, the strong

mathematical similarities between ML and DA mean that both fields can learn from each other and share methodologies (Geer, 2021). ML approaches can be used to improve optimisation 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 optimisation algorithms. For example, clustering methods can be used to set the population size, crossover probability and mutation probability parameters in genetic algorithms (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 optimise ML structure and hyperparameters (Akiba et al., 2019). ML techniques can also be used to choose the best-performing algorithm for a particular optimisation 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 optimising the parameter estimation algorithm hyperparameters themselves.

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 Sect. 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 neural networks to emulate the tangent linear and adjoint models since neural networks can be differentiated trivially (Hatfield et al., 2021).

Table 1: Summary of challenges outlined in Sect. 3 and their ML opportunities

| **PDA challenge** | **ML opportunity** |
|---|---|
| Selecting parameters and their prior distributions (Sect. 3.1) | |
| - Identifying which model parameters to optimise is challenging, due to high dimensionality and strong parameter covariances.<br>- Choosing prior distributions for parameters is crucial yet difficult, requiring detailed structural insights and data. | - Emulators can reduce the computational demand of running models with many different parameter settings needed for sensitivity analyses (Sect. 4.1).<br>- Emulators can be used to facilitate uncertainty quantification, for example, through history matching (Sect. 4.1). |
| Characterisation of model and data/observation errors (Sect. 3.2) | |
| - Model errors are difficult to quantify due to uncertainties in process representation, missing processes, and the challenge of specifying an informative prior.<br>- Quantifying data errors is tricky because of sampling variability, instrument inaccuracies, and complex error correlations that are often ignored. | - Hybrid modelling can be used to replace model processes that are missing or poorly understood, helping to diagnose model structural errors (Sect. 4.2).<br>- ML methods can be used to generate improved estimates of derived quantities, thus reducing observation errors (Sect. 4.3). |

| | |
|---|---|
| Developing observation operators (Sect. 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 (Sect. 4.3). |
| Tackling spatial and temporal heterogeneity (Sect. 3.4) | |
| - Variability in surface properties, driven by diverse climates, soils, and ecosystems, complicates parameter estimation across locations.<br>- High computational demands make it difficult to calibrate LSMs across large spatial domains.<br>- Temporal variability and short data series hinder the capture of both seasonal cycles and long-term trends. | - Hybrid modelling can be used to improve spatial parameterisations (Sect. 4.2).<br>- Emulators can help reduce the computational demand of running the model over large domains (Sect. 4.1).<br>- Long Short-Term Memory encoder-decoder networks consider long-term dependencies and therefore may help capture seasonal and interannual trends (Sect. 4.1). |
| Dealing with large and multiple observational datasets (Sect. 3.5) | |
| - Scaling satellite products to match model grids can lead to information loss.<br>- Products may be subject to regional biases due to varying data quality and processing methods.<br>- Assimilating multiple data streams in model calibrations presents challenges in consistency, error characterisation, and balancing different data sources. | - ML methods can be used to upscale sparse observational data (e.g., flux tower observations) or map satellite observations to a model grid (Sect. 4.3).<br>- ML can be applied to improve the algorithms used to produce observational datasets (Sect. 4.3).<br>- ML-based observation operators are able to easily accommodate multiple observation types and adjust their respective impacts in the assimilation (Sect. 4.3). |
| Including the historical period in the assimilation window (Sect. 3.6) | |
| - Spin-up and transient parts of model runs can be computationally demanding. | - Hybrid modelling can be used to replace computationally costly parts of the model (Sect. 4.2). |
| Choice and implementation of minimisation algorithms (Sect. 3.7) | |
| - Algorithms requiring a large number of model runs are computationally costly and therefore rarely applied to complex LSMs.<br>- For different algorithms, there can be a large number of configuration options and tuneable hyperparameters. | - ML can enhance computational efficiency, enabling the use of algorithms that require numerous model runs (Sect. 4.1).<br>- ML can help find the best configurations and hyperparameters to use when optimising (Sect. 4.4). |

| | |
|---|---|
| - Maintaining tangent linear/adjoint models for gradient-based optimisation 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 (Sect. 4.4). |

# 5. Future priorities

Moving beyond the ML avenues outlined in the previous section and summarised in Table 1, 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 datasets and experimental configurations

In addition to the traditional datasets used to optimise LSM parameters, our data-rich world offers access to a wide array of 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 fertilisation effect at nitrogen-limited sites (Thomas et al., 2017; Jiang et al., 2020; Mahmud et al., 2018; Raoult, Edouard-Rambaut, et al., 2024).
- **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 (Lawrence et al., 2020) can help constrain rates of soil carbon cycling (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 datasets:** New datasets, such as full-waveform lidar data from the GEDI (Global Ecosystem Dynamics Investigation) 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). 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 utilised to improve the representation of methane production processes.

By combining these data and implementing novel DA approaches described in this paper, we can aspire to assess how this information influences both short-term and long-term forecasts and reduces model discrepancies. The focus should be on refining core processes driving ecosystem-scale carbon and water fluxes and testing their responses to global change, beyond just fitting historical data.

As with all past carbon cycle DA studies, before novel datasets 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 observational characteristics (temporal sampling interval, record length, observation uncertainty, choice of minimisation algorithm and its configuration, etc. – Sect. 3.7) are required to retrieve the correct parameter values with the strong assumption that there is no modelling 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; Fox et al., 2009). As these data are model outputs, the “true” value of the parameters is known. Synthetic DA experiments can also be used prior to data collection, where they can help optimise sampling over space, time, and sampling design. Indeed, calibration has yet to be adequately integrated into the broader literature on model-driven observing system simulation experiments. To improve this, advocating for standardised community benchmark protocols and datasets could address different challenges, such as assessing resistance to noise and evaluating forcing variability. Results from such community-driven experimental setups could reveal common challenges and development opportunities, enhancing the robustness and effectiveness of DA methods across the field (see Sect. 5.4).

Additional tests of DA experimental configuration that are rarely performed (or rarely reported in the literature) should include testing i) how parameters retrieved at individual sites compare to parameters retrieved 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 Sect. 3.4), ii) the utility of PFT dependent parameters versus alternative approaches for grouping parameters (e.g., regionally dependent PFTs - e.g. Dahlin et al., 2017; Bao et al., 2023), iii) how retrieved parameters vary with the forcing dataset used in the simulations, iv) how retrieved values depend on which parameters and/or PFTs are optimised or which terms to include in the cost function, and v) how retrieved parameters vary in space and time within PFTs and what this tells us about missing processes, among other factors. A critical test of any parameterisation 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 optimised or a novel dataset is used in the assimilation. Building DA frameworks to include this technical testing will give confidence in using retrieved parameter values in operational versions of the models.

## 5.2 Moving towards land surface–atmospheric transport and full Earth system model coupling in data assimilation

Atmospheric $CO_2$ mole fraction measurements collected at tall towers around the world have proven valuable in improving NEE predictions at regional to global scales within a carbon cycle DA framework (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 (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. One benefit of assimilating atmospheric $CO_2$ concentration data is that it is one of the only datasets that can provide a large spatial scale constraint (albeit indirect) on RECO because it is heavily influenced by soil carbon stocks; thus, assimilating atmospheric $CO_2$ data presents an opportunity to improve the representation of both soil carbon flux and stock trajectories in LSMs, which is crucial for future predictions regarding the carbon sink capacity of terrestrial ecosystems.

However, the assimilation of atmospheric $CO_2$ data requires coupling LSMs with atmospheric transport models in order to scale 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 modelling errors associated with the atmospheric model itself (physics and spatial/vertical discretisation) 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) and Kaminski & Mathieu (2017). Recent advances in the utilisation of graphics accelerators (Chevallier et al., 2023) offer hope for a significant reduction in computational times and the development of 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 optimisations. To date, there has been limited assessment of whether posterior parameter values from offline DA experiments compare to retrieved values from fully coupled runs (nor how retrieved values vary when different offline climate reanalysis forcing products are used). To achieve this goal, LSM DA groups should learn from advances made in the NWP 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 (Sect. 4.1) (Watson-Parris et al., 2021) and automatic differentiation of model code (Gelbrecht et al., 2023) should help considerably in alleviating this problem (see Sect. 3.7 for remaining challenges).

## 5.3 Identifying and improving structural errors and model representation

The best estimates of different parameters are very dependent on the experimental setup and so few of the optimised 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). Instead, the main strength of parameter estimation for LSMs and, therefore, its main purpose thus far, has been to identify structural errors. If we cannot match observations within the bounds of their known uncertainties by simply changing the parameter values, this suggests that a process is poorly represented or missing from the model. This critical information is then fed back to the model developers to ensure changes are made to the model, before restarting the cycle of model calibration. Although this exchange is key in developing any LSM, it is rarely published. Nevertheless, a few documented examples from the ORCHIDEE land surface model workflow exist. MacBean et al. (2015) demonstrated that temperate broadleaved temperature thresholds for senescence in the ORCHIDEE LSM were too low. The newly optimised parameters have since been included in ORCHIDEE trunk versions. Salmon et al. (2022) found that when constraining parameters of the ORCHIDEE LSM against methane emissions in northern peatlands, the process providing enough active carbon for methanogenesis was missing. Raoult et al. (2023) found by assimilating MODIS snow albedo over Greenland that a three-layered ice sheet model was insufficient to simulate accurately both the snow albedo and runoff rates, leading to further discretisation of the model.

However, careful consideration is needed to avoid equating the status quo of making changes to models—often involving increased complexity—with progress in model development. While identifying and addressing structural errors is crucial, introducing new processes or refining existing ones can sometimes lead to models that are more complex without necessarily improving their predictive power. It is important to strike a balance between enhancing model accuracy and maintaining model parsimony. Overly complex models can become difficult to validate and manage, potentially obscuring rather than clarifying underlying processes. Therefore, the goal should be to make thoughtful adjustments that improve model performance while ensuring that the added complexity is justified by significant improvements in accuracy or functionality. This approach ensures that models remain robust and efficient and that any modifications contribute meaningfully to their overall effectiveness.

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 datasets 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 optimisation theories) and facilitate collaboration across different modelling groups.

## 5.4 International collaboration: intercomparison studies and shared toolboxes

Efforts by AIMES and ILMF to build a Land Data Assimilation 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 modelling 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 27th 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 utilising 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 land surface model benchmarking tools (iLAMB: Collier et al., 2018; Seiler et al., 2022). 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 datasets, 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).

## 5.5 Propagation of error reductions to constrain climate predictions

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 to enable observationally constrained probabilistic statements to be made about the future of the land biosphere. 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 trivial, but 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). However, with increased computational power, we are acquiring the capability to run LSMs as ensembles rather than relying on a single realisation, thereby enabling us to better capture the uncertainty of model predictions (Arora et al., 2023).

By sampling from the posterior distributions after a PDA experiment, we can generate ensemble simulations which can be used to explore future scenarios and idealised 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, which likely requires the use of emulators (see Sect. 4.1). The second is how to intelligently select parameter vectors from those distributions. Currently, simple models might propagate uncertainty by using 100-1000 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 towards 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 (IPCC) 8th Assessment Report, the annual Global Carbon Budget (GCB) and other emerging frameworks quantifying land carbon trajectories. Therefore, we must develop methods to maximise 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 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.

# 6. Summary and conclusion

Improving the accuracy of land surface models (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 data assimilation 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 Earth System Models) are complex in nature leading to many challenges when performing global scale optimisations. Nevertheless, advances in computational capability, novel datasets 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 land surface model data assimilation, helping to streamline the assimilation process, manage large datasets and speed up otherwise computationally demanding processes. International collaboration is crucial in this endeavour, as shared knowledge and resources can significantly accelerate the advancement of LSM calibration and data assimilation. Efforts to build a Land Data Assimilation Community, such as those by the AIMES Land Data Assimilation 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 standardised methodologies, ultimately leading to more accurate LSMs.

# Acknowledgements

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 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 Award DE-SC0022070, the National Science Foundation (NSF) under IA 1947282, and the National Center for Atmospheric Research (NCAR) through 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 number 337552) and Horizon Europe, HORIZON-MISS-2022-SOIL-01-05. 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).

# Open research

**Data Availability statement:** This article discusses the challenges and priorities in the field of parameter estimation for land data assimilation, and the opportunities offered by machine learning—it does not include the specific use of any particular software or results involving specific data products.

# Appendix

The process-based models mentioned through the paper are listed in Table A1. This list cover wide spectrum of land models ranging in complexity and computational demand, including 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 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.

**Table A1**. References for the process-based models mentioned in this article.

| **Acronym** | **Full name** | **Model reference** |
|---|---|---|
| BETHY | Biosphere Energy Transfer Hydrology | Knorr (2000) |
| 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 | Lawrence et al. (2019) |
| D&B | DALEC & BETHY | Knorr et al. (2024) |
| DALEC | Data Assimilation Linked Ecosystem Carbon | Williams et al. (2005) |
| ED | Ecosystem Demography | Ma et al. (2022); Moorcroft et al. (2001) |
| ECLand | European Centre 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 Biosphera-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 | Organising 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 & Heimann (1995) |
| SIPNET | Simplified Photosynthesis and Evapotranspiration | Braswell et al. (2005) |
| TECOS | terrestrial ecosystem | Xu et al., (2006) |

1

# References


Abadie, C., Maignan, F., Remaud, M., Kohonen, K.-M., Sun, W., Kooijmans, L., Vesala, T., Seibt, U., Raoult, N., Bastrikov, V., Belviso, S., & Peylin, P. (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., Finzi, A., Moorhead, D., Schimel, J., Torn, M., & Mayes, M. A. (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

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1907.10902

Alemohammad, S. H., Fang, B., Konings, A. G., Aires, F., Green, J. K., Kolassa, J., Miralles, D., Prigent, C., & Gentine, P. (2017). Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically-based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. *Biogeosciences*, *14*(18), 4101–4124. https://doi.org/10.5194/bg-14-4101-2017

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

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., … Ziehn, T. (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., Beaudoing, H. K., Getirana, A., Navari, M., Li, B., Jacob, J., Wegiel, J., & Peters-Lidard, C. D. (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., Fennel, K., Beck, H., De Lannoy, G., Pauwels, V. R. N., Raiho, A., Montzka, C., Williams, M., Mishra, U., Poppe, C., Zacharias, S., Lausch, A., Samaniego, L., Van Looy, K., …

Vereecken, H. (2021). Reanalysis in earth system science: Toward terrestrial ecosystem reanalysis. *Reviews of Geophysics (Washington, D.C.: 1985)*, *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., Peylin, P., Chevallier, F., Vuichard, N., & Bastrikov, V. (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., Weiss, M., Demarty, J., Santaren, D., Baret, F., Berveiller, D., Dufrêne, E., & Prunet, P. (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., Dormann, C. F., & Hartig, F. (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., Blyth, E., Bousserez, N., Boussetta, S., Brown, A., Buizza, R., Buontempo, C., Chevallier, F., Choulga, M., Cloke, H., Cronin, M. F., Dahoui, M., De Rosnay, P., Dirmeyer, P. A., … Zeng, X. (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). 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://gmd.copernicus.org/articles/11/4739/2018/

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., … Papale, D. (2010). Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science (New York, N.Y.)*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R., Ménard, C. B., & Harding R. (2011). The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://gmd.copernicus.org/articles/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., Meyer, V., Reager, J. T., Worden, H. M., Jiang, Z., Quetin, G. R., Smallman, T. L., Exbrayat, J.-F., Yin, Y., Saatchi, S. S., Williams, M., & Schimel, D. S. (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., Melton, J. R., Pongratz, J., Turton, R. H., Yoshimura, K., & Yuan, H. (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 (New York, N.Y.)*, *359*(6375). https://doi.org/10.1126/science.aam8328

Booth, B. B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., Huntingford, C., Betts, R. A., Harris, G. R., & Lloyd, J. (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., Agustí-Panareda,

A., Beljaars, A., Wedi, N., Munõz-Sabater, J., de Rosnay, P., Sandu, I., Hadade, I., Carver, G., Mazzetti, C., Prudhomme, C., Yamazaki, D., & Zsoter, E. (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., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., & Zhang, Q. (2018). *JAX: composable transformations of Python+NumPy programs* (Version 0.3.13). 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., & Kueppers L. (2021). Capturing functional strategies and compositional dynamics in vegetation demographic models. *Biogeosciences*, *18*(14), 4473–4490. https://bg.copernicus.org/articles/18/4473/2021/bg-18-4473-2021.html

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). 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., Papale, D., Rambal, S., & Seixas, J. (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., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M. J., & Valentini, R. (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), 13,218–13,235. https://doi.org/10.1002/2016JD024821

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

Chen, X., Huang, Y., Nie, C., Zhang, S., Wang, G., Chen, S., & Chen, Z. (2022). A long-term reconstructed TROPOMI solar-induced fluorescence dataset using machine learning algorithms. *Scientific Data*, *9*(1), 427. https://doi.org/10.1038/s41597-022-01520-1

Chevallier, F. (2007). Impact of correlated observation errors on inverted CO2surface fluxes from OCO measurements. *Geophysical Research Letters*, *34*(24). 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). https://doi.org/10.1029/2022gl102135

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., & Cox P. (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://gmd.copernicus.org/articles/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., Mu, M., & Randerson, J. T. (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., Rio, C., Audouin, O., Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M.-P., Madeleine, J.-B., Rodier, Q., & Xu, W. (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). 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., Valdayskikh, V., Radchenko, T., Nekrasova, O., Zakharov, V., & Jouzel, J. (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), 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. *Atmosphere-Ocean*, *0*(0), 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., Crow, W. T., de Rosnay, P., Durand, M., Forman, B., Geppert, G., Girotto, M., Hendricks Franssen, H.-J., Jonas, T., Kumar, S., Lievens, H., Lu, Y., Massari, C., Pauwels, V. R. N., … Steele-Dunne, S. (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*. https://doi.org/10.3389/frwa.2022.981745

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., Schepers, D., Weston, P., Zuo, H., Alonso-Balmaseda, M., Balsamo, G., Bonavita, M., Borman, N., Brown, A., Chrust, M., Dahoui, M., Chiara, G., English, S., Geer, A., … Rabier, F. (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

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., Kooper, R., LeBauer,

D., Mantooth, J., McHenry, K., & Wang, D. (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. (n.d.). *xgboost-distribution: Probabilistic prediction with XGBoost*. Github. Retrieved August 15, 2024, 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

Duan, T., Anand, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., & Schuler, A. (13--18 Jul 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. https://proceedings.mlr.press/v119/duan20a.html

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (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., Gayno, G., & Tarpley, J. D. (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). 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., Bloom, A. A., Falloon, P., Ito, A., Smallman, T. L., & Williams, M. (2018). Reliability ensemble averaging of 21st century projections of terrestrial net primary productivity reduces global and regional uncertainties. *Earth System Dynamics*, *9*(1), 153–165. https://doi.org/10.5194/esd-9-153-2018

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., Bocquet, M., Bretherton, C. S., Christensen, H. M., Dagon, K., Gagne, D. J., Hall, D., Hammerling, D., Hoyer, S., Iglesias-Suarez, F., Lopez-Gomez, I., McGraw, M. C., Meehl, G. A., Molina, M. J., … Zanna, L. (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., Parazoo, N. C., Stettz, S. G., Yang, Y., Bonal, D., Bloom, A. A., Williams, M., & Konings, A. G. (2021). Optimal model complexity for terrestrial carbon cycle prediction. *Biogeosciences* , *18*(8), 2727–2754. https://doi.org/10.5194/bg-18-2727-2021

Fang, K., Pan, M., & Shen, C. (2019). The value of SMAP for long-term soil moisture estimation with the help of deep learning. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *57*(4), 2221–2233. https://doi.org/10.1109/tgrs.2018.2872131

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. Royal Meteorological Society (Great Britain)*, *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 CO2 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., Desai, A., Duveneck, M. J., Fisher, J. B., Haynes, K. D., Hoffman, F. M., Johnston, M. R., Kooper, R., LeBauer, D. S., Mantooth, J., Parton, W. J., Poulter, B., Quaife, T., Raiho, A., … Dietze, M. C. (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., Murphy, B., Desai, A. R., & Dietze, M. C. (2021). Capturing site-to-site variability through Hierarchical Bayesian calibration of a process-based dynamic vegetation model. In *bioRxiv* (p. 2021.04.28.441243). 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). https://doi.org/10.1029/2018ms001453

Fisher, R. A., Muszala, S., Verteinstein, M., Lawrence, P., Xu, C., McDowell, N. G., Knox, R. G., Koven, C., Holm, J., Rogers, B. M., Spessa, A., Lawrence, D., & Bonan, G. (2015). Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes, CLM4.5(ED). *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., Fisher, J. B., Shi, M., Walker, A. P., & Lawrence, D. M. (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., v. 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., von Bloh, W., & Carvalhais, N. (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., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., & Van Wijk, M. T. (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

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., … Zheng, B. (2023). Global Carbon Budget 2023. *Earth System Science Data*, *15*(12), 5301–5369. https://doi.org/10.5194/essd-15-5301-202310.5194/essd-15-5301-2023-supplement

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Ali Eslami, S. M., & Teh, Y. W. (2018). Neural Processes. In *arXiv [cs.LG]*. arXiv. 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., & Alemohammad, S. H. (2018). Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. *Geophysical Research Letters*, *45*(7), 3136–3146.

https://doi.org/10.1002/2017GL076294
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. In *arXiv [physics.ao-ph]*. arXiv. https://doi.org/10.5194/egusphere-2024-277
Giering, R. (2010). *Transformation of algorithms in Fortran Version 1.15 (TAF Version 1.9. 70)*.
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. (1997). *On Automatic Differentiation*. 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? In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2207.08815
Groenendijk, M., Dolman, A. J., Ammann, C., Arneth, A., Cescatti, A., Dragoni, D., Gash, J. H. C., Gianelle, D., Gioli, B., Kiely, G., Knohl, A., Law, B. E., Lund, M., Marcolla, B., van der Molen, M. K., Montagnani, L., Moors, E., Richardson, A. D., Roupsard, O., … Wohlfahrt, G. (2011). Seasonal variation of photosynthetic model parameters and leaf area index from global Fluxnet eddy covariance data. *Journal of Geophysical Research: Biogeosciences*, *116*(G4). 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., Loisel, J., Malhotra, A., Jackson, R. B., Ogle, S., Phillips, C., Ryals, R., Todd-Brown, K., Vargas, R., Vergara, S. E., Cotrufo, M. F., Keiluweit, M., Heckman, K. A., Crow, S. E., … Nave, L. E. (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). https://doi.org/10.1029/2021ms002521
Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms* (2nd ed.) [PDF]. 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., Campanharo, W. A., House, J. I., Sitch, S., Hales, T. C., Adami, M., Anderson, L. O., & Aragão, L. E. O. C. (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., Cassol, H. L. G., Achard, F., Jucker, T., Silva, C. A., House, J., Sitch, S., Hales, T. C., & Aragão, L. E. O. C. (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., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., … Zuo, H. (2018). Operational global reanalysis: progress, future directions and synergies with NWP | ERA Report Series. *ECMWF*. https://www.ecmwf.int/en/elibrary/80922-operational-global-reanalysis-progress-future-directions-and-synergies-nwp
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., Guimberteau, M., Jornet-Puig, A., Lu, X., & Luo, Y. (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
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., Evans, M. N., Babst, F., & Luyssaert, S. (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-14-5891-202110.5194/gmd-14-5891-2021-supplement
Jiang, M., Medlyn, B. E., Drake, J. E., Duursma, R. A., Anderson, I. C., Barton, C. V. M., Boer, M. M., Carrillo, Y., Castañeda-Gómez, L., Collins, L., Crous, K. Y., De Kauwe, M. G., Dos Santos, B. M., Emmerson, K. M., Facey, S. L., Gherlenda, A. N., Gimeno, T. E., Hasegawa,

S., Johnson, S. N., … Ellsworth, D. S. (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

Jian, J., Vargas, R., Anderson-Teixeira, K., Stell, E., Herrmann, V., Horn, M., Kholod, N., Manzon, J., Marchesi, R., Paredes, D., & Bond-Lamberty, B. (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

Joiner, J., Yoshida, Y., Zhang, Y., Duveiller, G., Jung, M., Lyapustin, A., Wang, Y., & Tucker, C. J. (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., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., … Williams, C. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *116*(G00J07). https://doi.org/10.1029/2010jg001566

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., … Reichstein, M. (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., Roy, D., & Morisette, J. (2002). An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, *83*, 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., Vestin, P., & Weslien, P. (2024). Assimilating multi-site eddy-covariance data to calibrate the $CH_4$ wetland emission module in a terrestrial ecosystem model. In *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 Biogeochem. 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., Blessing, S.,

Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J. P., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., … Ziehn, T. (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., 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

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., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar C, C., Aleixo, I., Ali, H., … Wirth, C. (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., Carbone, M. S., Reichstein, M., & Richardson, A. D. (2011). The model-data fusion pitfall: assuming certainty in an uncertain world. *Oecologia*, *167*(3), 587–597. https://doi.org/10.1007/s00442-011-2106-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

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

Kennedy, D., Dagon, K., Lawrence, D. M., Fisher, R. A., Sanderson, B. M., Collier, N., Hoffman, F. M., Koven, C. D., Levis, S., Oleson, K. W., Zarakas, C. M., Cheng, Y., Foster, A., Lu, X., Fowler, M., Hawkins, L., Kavoo, T., Kluzek, E. B., Kumar, S., … Luo, Y. (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

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*. https://www.jstor.org/stable/1913641?casa_token=1EgxpT-njfMAAAAA:EPhIlE24zFcuOWFrDS_G8r0LMGIuWx60_Ft7yeBlx9XViRf0sI27b3uD1MmKewRbJkO2Lcgb7G3kDQS-sQBi1vItqRAyTEHOyQfEkAYw1W6DtaIc-PGm

Knorr, W. (2000). Annual and interannual $CO_2$ exchanges of the terrestrial biosphere: process-based simulations and uncertainties: CO2 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 CO2 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., 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). 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., Quaife, T., Smallmann, L., Steele-Dunne, S., Vreugdenhil, M., Green, T., Zähle, S., Aurela, M., Bouvet, A., Bueechi, E., Dorigo, W., El-Madany, T., Migliavacca, M., Honkanen, M., … Drusch, M. (2024). A comprehensive land surface vegetation model for multi-stream data assimilation, D&B v1.0. In *EGUsphere*. https://doi.org/10.5194/egusphere-2024-1534

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). https://doi.org/10.1029/2010gb003900

Kolassa, J., Reichle, R. H., Liu, Q., Cosh, M., Bosch, D. D., Caldwell, T. G., Colliander, A., Collins, C. H., Jackson, T. J., Livingston, S. J., Moghaddam, M., & Starks, P. J. (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., Lewis, J., Lindsay, K., Mathesius, S., Meinshausen, M., Mills, M., Nicholls, Z., Sanderson, B. M., Séférian, R., Swart, N. C., Wieder, W. R., & Zickfeld, K. (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.,

Detto, M., Dietze, M. C., Faybishenko, B., Holm, J., Huang, M., Kovenock, M., Kueppers, L. M., Lemieux, G., Massoud, E., McDowell, N. G., Muller-Landau, H. C., Needham, J. F., Norby, R. J., … Xu, C. (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*. Aspendale, Vic., 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., Ciais, P., Sitch, S., & Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system: DVGM FOR COUPLED CLIMATE STUDIES. *Global Biogeochemical Cycles*, *19*(1). https://doi.org/10.1029/2003gb002199

Kumar, S., Kolassa, J., Reichle, R., Crow, W., de Lannoy, G., de Rosnay, P., MacBean, N., Girotto, M., Fox, A., Quaife, T., Draper, C., Forman, B., Balsamo, G., Steele-Dunne, S., Albergel, C., Bonan, B., Calvet, J.-C., Dong, J., Liddy, H., & Ruston, B. (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). 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). 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://gmd.copernicus.org/articles/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://gmd.copernicus.org/articles/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., Jacobs, E. M., McNellis, R., Rady, J. M., Smith, N. G., Thomas, R. Q., Wieder, W. R., & Grandy, A. S. (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., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., & Battaglia, P. (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., Heckman, K., Blankinship, J. C., Crow, S. E., McNicol, G., Trumbore, S., Levine, P. A., Vindušková, O., Todd-Brown, K., Rasmussen, C., Hicks Pries, C. E., Schädel, C., McFarlane, K., Doetterl, S., … Wagai, R. (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., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., … Zeng, X. (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

Li, L., Fang, Y., Zheng, Z., Shi, M., Longo, M., Koven, C. D., Holm, J. A., Fisher, R. A., McDowell, N. G., Chambers, J., & Leung, L. R. (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-16-4017-202310.5194/gmd-16-4017-2023-supplement

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., Wang, H., & Zhou, G. (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

Lu, D., Ricciuto, D., Walker, A., Safta, C., & Munger, W. (2017). Bayesian calibration of

terrestrial ecosystem models: a study of advanced Markov chain Monte Carlo methods. *Biogeosciences*, *14*, 4295–4314. https://doi.org/10.5194/BG-14-4295-2017

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., Hararuk, O., Hou, E., Jiang, L., Liao, C., Lu, X., Shi, Z., Smith, B., Tao, F., & Wang, Y.-P. (2022). Matrix approach to land carbon cycle modeling. *Journal of Advances in Modeling Earth Systems*, *14*(7). 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

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

MacBean, N., Bacour, C., Raoult, N., Bastrikov, V., Koffi, E. N., Kuppel, S., Maignan, F., Ottlé, C., Peaucelle, M., Santaren, D., & Peylin, P. (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., Köhler, P., Gómez-Dans, J., & Disney, M. (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., Oinonen, M., Straková, P., & Viskari, T. (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., Lohila, A., Mammarella, I., Margolis, H., Markkanen, T., Susiluoto, J., Thum, T., Viskari, T., Zaehle, S., & Aalto, T. (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

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

Massoud, E. C., Xu, C., Fisher, R. A., Knox, R. G., Walker, A. P., Serbin, S. P., Christoffersen, B. O., Holm, J. A., Kueppers, L. M., Ricciuto, D. M., Wei, L., Johnson, D. J., Chambers, J. Q., Koven, C. D., McDowell, N. G., & Vrugt, J. A. (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., Williams, S., Motew, M., Campbell, E. E., Brummitt, C. D., Paustian, K., & Kumar, A. A. (2023). Validating DayCent-CR for cropland soil carbon offset reporting at a national scale. *Geoderma*, *438*(116647), 116647. https://doi.org/10.1016/j.geoderma.2023.116647

Mauritsen, T., Bader, J., & Becker, T. (2019). *Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2*. https://doi.org/10.1029/2018ms001400

McGovern, A., Elmore, K. L., Gagne, D. J., II, Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., & Williams, J. K. (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: Biogeosciences*, *114*(G1). 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., De Deurwaerder, H. P. T., Krishna Moorthy, S. M., Schnitzer, S. A., Marvin, D. C.,

Longo, M., Liu, C., Broadbent, E. N., Almeyda Zambrano, A. M., Muller-Landau, H. C., Detto, M., & Verbeeck, H. (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

Migliavacca, M., Meroni, M., Busetto, L., Colombo, R., Zenone, T., Matteucci, G., Manca, G., & Seufert, G. (2009). Modeling gross primary production of agro-forestry ecosystems by assimilation of satellite-derived information in a process-based model. *Sensors (Basel, Switzerland)*, *9*(2), 922–942. https://doi.org/10.3390/s90200922

Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological Monographs*. 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 CO2 and H2O fluxes. *Agricultural and Forest Meteorology*, *148*(10), 1467–1477. https://doi.org/10.1016/j.agrformet.2008.04.013

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., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Zhang, W., Hamdi, Z. M., … Jung, M. (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 CH4 and N2O flux measurements in terrestrial ecosystems. *Atmospheric Environment (Oxford, England: 1994)*, *81*, 311–319. https://doi.org/10.1016/j.atmosenv.2013.09.030

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., MacBean, N., Raoult, N., Scott, R. L., Shi, Y., Sulman, B. N., & Wood, J. D. (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., & Hartig. (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://gmd.copernicus.org/articles/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*. *16*, 19–22. 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

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Neural Information Processing Systems*, *abs/1912.01703*. https://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library

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., Koffi, E., Kane, A., Maignan, F., Chevallier, F., Ciais, P., & Prunet, P. (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., Robinson, E., Morrison, R., Osborne, S., & Quaife, T. (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. M., Casella, E., Dance, S. L., Lawless, A. S., Morison, J. I. L., Nichols, N. K., Wilkinson, M., & Quaife, T. L. (2017). Understanding the effect of disturbance from selective felling on the carbon dynamics of a managed woodland by combining observations with model predictions. *J. Geophys. Res. Biogeosci.*, *122*, 886–902. https://doi.org/10.1002/2017JG003760

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://hess.copernicus.org/articles/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

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., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*(112165), 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., Lomas, M., D'Alessandro, M. M., Paillou, P., Papathanassiou, K., Rocca, F., Saatchi, S., Scipal, K., Shugart, H., Smallman, T. L., Soja, M. J., Tebaldini, S., Ulander, L., Villard, L., & Williams, M. (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

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

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., Charbit, S., Dumas, C., Maignan, F., Ottlé, C., & Bastrikov, V. (2023). Improving

modelled albedo over the Greenland ice sheet through parameter optimisation and MODIS snow albedo retrievals. *The Cryosphere*, *17*(7), 2705–2724. https://doi.org/10.5194/tc-17-2705-2023

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. 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

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

Rayner, P. J. (2010). *The current state of carbon-cycle data assimilation*. *2*(4), 289–296. https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=H3up71wAAAAJ:isC4tDSrTZIC

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). https://doi.org/10.1029/2004gb002254

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (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). 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 CO2 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). https://doi.org/10.1029/2010jg001400

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., Ponton, S., Bonal, D., Meir, P., & Williams, M. (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 CO2 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., Guimbaud, C., Zhu, D., Ciais, P., Peylin, P., Gogo, S., Laggoun-Défarge, F., Aurela, M., Bret-Harte, M. S., Chen, J., Chojnicki, B. H., Chu, H., Edgar, C. W., Euskirchen, E. S., … Ziemblińska, K. (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., Saisana, M., & Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Wiley & Sons, Limited, John. 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., Ingram, W. J., Piani, C., Stainforth, D. A., Stone, D. A., & Allen, M. R. (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., Bacour, C., Ciais, P., & Longdoz, B. (2014). Ecosystem model optimization using in situ flux observations: benefit of Monte Carlo versus variational schemes and analyses of the year-to-year model performances. *Biogeosciences*, *11*(24), 7137–7158. https://doi.org/10.5194/bg-11-7137-2014

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). 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). 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., Kerr, Y., Mialon, A., Richaume, P., Rodríguez-Fernández, N., Vittucci, C., Wigneron, J.-P., Mecklenburg, S., & Drusch, M. (2019). Mean European carbon sink over 2010–2015 estimated by simultaneous assimilation of atmospheric $CO_2$, 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. *Https://agupubs.onlinelibrary.wiley.com › 2007JD008642https://agupubs.onlinelibrary.wiley.com › 2007JD008642*. https://doi.org/10.1029/2007JD008642

Schürmann, G. J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., Giering, R., Rödenbeck, C., Heimann, M., & Zaehle, S. (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., Goll, D., Jain, A. K., Joetzjer, E., Lienert, S., Lombardozzi, D., Luyssaert, S., Nabel, J. E. M. S., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., & Zaehle, S. (2022). Are terrestrial biosphere models fit for simulating the global land carbon sink? *Journal of Advances in Modeling Earth Systems*, *14*(5). https://doi.org/10.1029/2021ms002946

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., Albergel, C., Calvet, J.-C., Ku, O., & Georgievska, S. (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), 113116. https://doi.org/10.1016/j.rse.2022.113116

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., … Lawson, K. (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

Shen, H., Wang, Y., Guan, X., Huang, W., Chen, J., Lin, D., & Gan, W. (2022). A spatiotemporal constrained machine learning method for OCO-2 solar-induced chlorophyll fluorescence (SIF) reconstruction. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *60*, 1–17. https://doi.org/10.1109/tgrs.2022.3204885

Shiklomanov, A. N., Cowdery, E. M., Bahn, M., Byun, C., Jansen, S., Kramer, K., Minden, V., Niinemets, Ü., Onoda, Y., Soudzilovskaia, N. A., & Dietze, M. C. (2018). Does the leaf economic spectrum hold within plant functional types? A Bayesian multivariate trait

meta-analysis. In *bioRxiv* (p. 475038). 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
Shi, Z., Allison, S. D., He, Y., Levine, P. A., Hoyt, A. M., Beem-Miller, J., Zhu, Q., Wieder, W. R., Trumbore, S., & Randerson, J. T. (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
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., Arneth, A., Arora, V. K., Bastos, A., Bastrikov, V., Bellouin, N., Canadell, J. G., Chini, L., Ciais, P., Falk, S., Harris, I., Hurtt, G., Ito, A., Jain, A. K., … Zaehle, S. (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., Jenkins, S., Leach, N., Mathison, C., & Partanen, A.-I. (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), 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., Lu, L., & Denning, A. S. (2008). Remote sensing data assimilation for a prognostic phenology model: DATA ASSIMILATION AND PHENOLOGY MODELING. *Journal of Geophysical Research*, *113*(G4). 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., Mishra, U., Jiang, L., Huang, X., & Luo, Y. (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., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Doughty, R., Ren, Z., & Luo, Y. (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 Netherlands. 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., Dinon-Aldridge, H., Burkhart, H. E., Domec, J.-C., Fox, T. R., Gonzalez-Benecke, C. A., Martin, T. A., Noormets, A., Sampson, D. A., & Teskey, R. O. (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., Loustau, D., & Ciais, P. (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). https://doi.org/10.1029/2022wr032233

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S.,

& Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Tramontana, G., Migliavacca, M., Jung, M., Reichstein, M., Keenan, T. F., Camps-Valls, G., Ogee, J., Verrelst, J., & Papale, D. (2020). Partitioning net carbon dioxide fluxes into photosynthesis and respiration using neural networks. *Global Change Biology*, *26*(9), 5235–5253. https://doi.org/10.1111/gcb.15203

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., & Nikolova, S. (2007). OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *Journal of Geophysical Research: Biogeosciences*, *112*(G2). 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

Vekuri, H., Tuovinen, J.-P., Kulmala, L., Papale, D., Kolari, P., Aurela, M., Laurila, T., Liski, J., & Lohila, A. (2023). A widely-used eddy covariance gap-filling method creates systematic bias in carbon balance estimates. *Scientific Reports*, *13*(1), 1720. https://doi.org/10.1038/s41598-023-28827-2

Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., & Ciais, P. (2011). Seasonal patterns of CO2fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. *Journal of Geophysical Research*, *116*(G2). 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 the Environment*, *242*(111740), 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., Torres-Rojas, L., Sadri, S., & Wood, E. F. (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 Simulation*, *10*(3). https://doi.org/10.1515/IJNSNS.2009.10.3.273

Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., Bastrikov, V., & Peylin, P. (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). https://doi.org/10.1029/2023jd038556

Wang, J., Jiang, F., Wang, H., Qiu, B., Wu, M., He, W., Ju, W., Zhang, Y., Chen, J. M., & Zhou, Y. (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), 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), 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., Boedecker, J., Dormann, C. F., Pappenberger, F., & Balsamo, G. (2024). Advances in Land Surface Model-based Forecasting: A comparative study of LSTM, Gradient Boosting, and Feedforward Neural Network Models as prognostic state emulators. In *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., Kooijmans, L. M. J., Hilton, T. W., Belviso, S., Peylin, P., Commane, R., Sun, W., Chen, H., Kuai, L., Mammarella, I., Maseyk, K., Berkelhammer, M., Li, K.-F., Yakir, D., … Campbell, J. E. (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. In *arXiv [physics.comp-ph]*. arXiv. http://arxiv.org/abs/2401.00035

Wu, M., Jiang, F., Scholze, M., Chen, D., Ju, W., Wang, S., Kaminski, T., Lu, Z., Vossbeck, M., & Zheng, M. (2024). Regional responses of vegetation productivity to the two phases of ENSO. *Geophysical Research Letters*, *51*(8). https://doi.org/10.1029/2024gl108176

Wu, M., Scholze, M., Kaminski, T., Voßbeck, M., & Tagesson, T. (2020). Using SMOS soil moisture data combining CO2 flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS). *Remote Sensing of Environment*, *240*(111719), 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, K., Liu, P., Zhang, J., Han, D., Wang, G., & Shen, C. (2021). Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *Journal of Hydrology*, *603*(127043), 127043. https://doi.org/10.1016/j.jhydrol.2021.127043

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., Du, M., & Liu, C. (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

Yatheendradas, S., & Kumar, S. (2022). A novel machine learning–based gap-filling of

fine-resolution remotely sensed snow cover fraction data by combining downscaling and regression. *Journal of Hydrometeorology*, *23*(5), 637–658. https://doi.org/10.1175/jhm-d-20-0111.1

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). https://doi.org/10.1029/2009gb003522

Zeng, H., Elnashar, A., Wu, B., Zhang, M., Zhu, W., Tian, F., & Ma, Z. (2022). A framework for separating natural and anthropogenic contributions to evapotranspiration of human-managed land covers in watersheds based on machine learning. *The Science of the Total Environment*, *823*(153726), 153726. https://doi.org/10.1016/j.scitotenv.2022.153726

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

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., & Gentine, P. (2018). A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences*, *15*(19), 5779–5800. https://bg.copernicus.org/articles/15/5779/2018/bg-15-5779-2018.html

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., Lin, C., Li, X., & Qiu, G. Y. (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. In *arXiv [cs.DC]*. arXiv. http://arxiv.org/abs/2405.00018

Zhu, H., Wu, M., Jiang, F., Vossbeck, M., Kaminski, T., Xing, X., Wang, J., Ju, W., & Chen, J. M. (2023). Assimilation of Carbonyl Sulfide (COS) fluxes within the adjoint-based data assimilation system–Nanjing University Carbon Assimilation System (NUCAS v1.0). In *EGUsphere* (pp. 1–35). https://doi.org/10.5194/egusphere-2023-1955

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