JAMES Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2024MS004715

Key Points:

- We constructed a parameter perturbation ensemble of the Community Land Model, v5.1, perturbing 211 parameters across six forcing scenarios
- Parameter effects can exceed scenario effects and parameter effect rankings differ by biome and based on the forcing scenario
- The software infrastructure developed in our experiment has greatly reduced the human and computer time needed for constructing future PPEs

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

D. Kennedy,
djk2120@ucar.edu

Citation:

Kennedy, D., Dagon, K., Lawrence, D. M., Fisher, R. A., Sanderson, B. M., Collier, N., et al. (2025). One-at-a-time parameter perturbation ensemble of the Community Land Model, version 5.1. *Journal of Advances in Modeling Earth Systems*, *17*, e2024MS004715. https://doi.org/10.1029/2024MS004715

Received 16 SEP 2024
Accepted 23 MAY 2025

One-at-a-Time Parameter Perturbation Ensemble of the Community Land Model, Version 5.1

D. Kennedy[1,2], K. Dagon[1], D. M. Lawrence[1], R. A. Fisher[3], B. M. Sanderson[3], N. Collier[4], F. M. Hoffman[4], C. D. Koven[5], E. Kluzek[1], S. Levis[1], X. Lu[6], K. W. Oleson[1], C. M. Zarakas[7], Y. Cheng[8], A. C. Foster[1], M. D. Fowler[1], L. R. Hawkins[9], T. Kavoo[10], S. Kumar[10], A. J. Newman[8], P. J. Lawrence[1], F. Li[11], D. L. Lombardozzi[1,12], Y. Luo[13], J. K. Shuman[14], A. L. S. Swann[7,15], S. C. Swenson[1], G. Tang[1], W. R. Wieder[1,16], and A. W. Wood[1]

[1]Climate and Global Dynamics Laboratory, NSF NCAR, Boulder, CO, USA, [2]Earth Research Institute, University of California, Santa Barbara, CA, USA, [3]CICERO Centre for International Climate and Environmental Research, Oslo, Norway, [4]Oak Ridge National Laboratory, Oak Ridge, TN, USA, [5]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA, USA, [6]School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, China, [7]Department of Atmospheric and Climate Science, University of Washington, Seattle, WA, USA, [8]Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [9]Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA, [10]College of Forestry, Wildlife and Environment, Auburn University, Alabama, AL, USA, [11]International Center for Climate and Environment Sciences, Chinese Academy of Sciences, Institute of Atmospheric Physics, Beijing, China, [12]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA, [13]School of Integrative Plant Science, Cornell University, Ithaca, NY, USA, [14]Earth Science Division, National Aeronautics and Space Administration Ames Research Center, Moffett Field, CA, USA, [15]Department of Biology, University of Washington, Seattle, WA, USA, [16]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA

Abstract Comprehensive land models are subject to significant parametric uncertainty, which can be hard to quantify due to the large number of parameters and high model computational costs. We constructed a large parameter perturbation ensemble (PPE) for the Community Land Model version 5.1 with biogeochemistry configuration (CLM5.1-BGC). We performed more than 2,000 simulations perturbing 211 parameters across six forcing scenarios. This provides an expansive data set, which can be used to identify the most influential parameters on a wide range of output variables globally, by biome, or by plant functional type. We found that parameter effects can exceed scenario effects and that a small number of parameters explains a large fraction of variance across our ensemble. The most important parameters can differ regionally and also based on the forcing scenario. The software infrastructure developed for this experiment has greatly reduced the human and computer time needed for CLM PPEs, which can facilitate routine investigation of parameter sensitivity and uncertainty, as well as automated calibration.

Plain Language Summary The Community Land Model includes a large set of numerical settings that help describe attributes of the various components of the land system. Each setting has a default value, but we know that other values may also be reasonable within a certain range. We ran a large set of simulations, increasing and decreasing each setting independently to better understand its influence on model outputs, such as plant productivity and the water cycle. We repeated these experiments across a range of scenarios, including present-day conditions and introducing (or removing) various aspects of climate change. We found that changing certain model settings could influence our results as much as the influence of climate change, itself. We also found that the most influential settings varied by geographic region. Understanding the influence of all of these settings can help us improve our model and also help us gauge our confidence in model predictions.

1. Introduction

Water availability, land temperature extremes, fire risk, and crop productivity will all see impacts from climate change, and are among the many processes represented within the terrestrial components of Earth System Models. Understanding how these processes respond to rising CO_2 and even influence CO_2 concentration trajectories is a critical facet of climate change research. Land processes substantially influence climate directly through, for example, evapotranspiration (ET) (Zarakas et al., 2024), and indirectly through carbon-climate feedbacks (Cox et al., 2000). Uncertainty in climate model projections varies by domain and generally increases with extended time horizons (Koven et al., 2022). Making centennial-scale projections of the cumulative terrestrial carbon sink has been especially challenging, with high uncertainty persisting across model generations (Arora et al., 2020; Friedlingstein et al., 2006, 2014), and limited efficacy of emergent constraints (L. Liu et al., 2024). Emissions trajectories and land management scenarios contribute some uncertainty, but the majority of this uncertainty is inherent to the challenge of predicting vegetation dynamics in a novel climate (Lovenduski & Bonan, 2017). Still, our expectation is that some of this uncertainty is indeed reducible, if we can more effectively utilize the ongoing expansion of observational data sources from remote sensing, meteorological stations, flux towers, and field campaigns (Baldocchi et al., 2024; Worden et al., 2021). The capability to efficiently ingest these data and improve simulation performance with improved parameters is especially valuable for actionable science (Cheng et al., 2023). However, given increasingly comprehensive land modeling systems, and the wide array of observational products, several technical hurdles exist that hinder effective model development and calibration.

Model inter-comparison projects (MIPs) are a major component of the model development cycle, and have been the primary means by which model projection uncertainty is assessed (Eyring et al., 2016; Friedlingstein et al., 2022; Henderson-Sellers et al., 1995; Pitman et al., 1999; Schlosser et al., 2000; Wood et al., 1998). While MIPs have had tremendous utility in capturing and assessing wide ranges of model assumptions, it can nonetheless be difficult to interpret the differences between models, or even between subsequent versions of the same model, due to the multiplicity of structural and parametric variations (McNeall et al., 2016). In most MIPs, each model is typically allowed only a single parameterization (despite the existence of many plausible parameter sets) due to the high cost of each simulation, for example, the TRENDY land model intercomparison (Sitch et al., 2024) or the Coupled Model Intercomparison Project (Eyring et al., 2016). Thus, MIPs typically conflate parametric and structural uncertainty, and de-emphasize the consequences of uncertain model calibration in estimates of future land state trajectories.

Understanding the range of outcomes arising from the many plausible parameter combinations is a critical step in robust uncertainty quantification. Parameter sensitivity tests can be used to gauge parametric uncertainty, and a collection of systematic parameter sensitivity tests across multiple parameters is often termed a Perturbed Parameter Ensemble (PPE, also referred to as Perturbed Physics Ensemble), with examples in the coupled (Murphy et al., 2004) and land-only contexts (Baker et al., 2022; Dagon et al., 2020; McNeall et al., 2024). One important application of PPEs has been assessing uncertainty in climate model projections (Booth et al., 2012; Hawkins et al., 2019; Murphy et al., 2004; Peatier et al., 2022; Sanderson et al., 2008; Tett et al., 2022; Yamazaki et al., 2021). Beyond uncertainty quantification, PPEs also form a basis for automated model calibration through various approaches like history matching, which does not require assumptions regarding parameter distributions (D. Williamson et al., 2013; D. B. Williamson et al., 2017; Hourdin et al., 2020; Couvreux et al., 2021; McNeall et al., 2024), or Bayesian methods (Cleary et al., 2021; Fer et al., 2018; Ziehn et al., 2012), or using genetic algorithms (Bastrikov et al., 2018). Several promising avenues for model calibration are in development (Alonso-González et al., 2022; Cleary et al., 2021; Pinnington et al., 2020), many of which require the construction of large PPEs (Qian et al., 2018). In a more general sense, PPEs yield a robust knowledge basis for understanding and working with a given climate model. This can aid in introducing new users to the model, or to help steer and facilitate model development.

The process of parameter estimation for the complex land models typically embedded within Earth System Models is challenging on account of the intrinsic complexity of the heterogeneous land surface, the diversity and plasticity of plant and microbial life, and the multiplicity of domains that comprise the terrestrial biosphere (e.g.., hydrology, snow, biogeochemistry, physiology, land management, etc.). The number of model parameters required to represent all the relevant processes is large and model simulations are relatively expensive, presenting a major barrier to robust calibration, model interpretation, and development (Dagon et al., 2020; Fisher & Koven, 2020). Thorough assessment of the parametric sensitivity of land models in a manner that is repeatable, open, and integrated with ongoing code development is desirable but challenging (Balaji et al., 2022; Hourdin et al., 2017). It typically requires extensive computational time, software engineering, and domain-specific expert scientific knowledge. Our work here tries to alleviate each of these challenges in turn, at least within the context of one land model. While we are not yet at the point of automated calibration, we present herein several innovations that will serve to make an automated calibration system more technically feasible.

In this project we create a large PPE with the Community Land Model 5.1 with biogeochemistry configuration (CLM5.1-BGC, hereafter CLM) (Lawrence et al., 2019). We extend the work of Dagon et al. (2020) which utilized a PPE to calibrate a subset of CLM parameters. Here we include a broader suite of parameters and continue to enhance the software tools for parameter perturbation and optimization. The goal of this project is to systematize the parameter perturbation process within the CLM modeling framework, and develop the necessary tools and data sets to efficiently test parameter effects across a wide range of land model processes. In doing so, we have generated a large PPE, comprising thousands of parameter sensitivity tests. In this paper we present that data set, describe how it was produced, and survey some potential applications. The data set has already demonstrated utility for diagnosing parameter effects (Cheng et al., 2023; Yan et al., 2023a, 2023b; Zarakas et al., 2024), while the software and modeling infrastructure has greatly expanded our capability to generate insights about parameter effects and sensitivities within CLM.

2. Experiment Description

2.1. Model Description

This experiment utilizes the Community Land Model configuration (version 5.1, i.e. CLM5.1) of the Community Terrestrial Systems Model. The model source code and documentation are available online (https://github.com/ESCOMP/CTSM), as is a full model description (Lawrence et al., 2019).

Relative to CLM5.0, version 5.1 includes minor bug fixes, parameter adjustments (Birch et al., 2021), and the implementation of biomass heat storage (Swenson et al., 2019). The PPE experiment required additional code modifications to enable systematic variation of the full suite of model parameters (many of which were previously "hard-coded", i.e., written directly as numerical values in the model source code). These modifications were incorporated into the main branch of the CLM codebase for future use (see Open Research Section). We utilized the biogeochemistry version of CLM, which includes active carbon and nitrogen cycles. This is as opposed to the satellite phenology mode, which is driven by observed canopy properties and thus has fewer internal feedbacks, and was explored in Dagon et al. (2020). We used the model in land-only mode driven by a data atmosphere (see Section 2.5 for details), with the crop model turned off.

2.2. Model Spin-Up

Model spin-up for the equilibration of carbon and nitrogen pools within biogeochemistry-enabled land models can consume up to 98% of computational time needed for a simulation (Sun et al., 2023). Depending on the evaluation criteria and model configuration, CLM5 requires between 800 and 2,000 years (or more) to reach steady-state conditions (Lawrence et al., 2019). In the absence of equilibrium, the drift toward steady state can obscure important model dynamics or features (Séférian et al., 2016). Because each member of the PPE can have a unique steady state, we performed an independent spin-up for each ensemble member.

To manage computational cost we leveraged the Semi-Analytic spin-up mode (SASU) recently implemented within CLM5 (Liao et al., 2023; Lu et al., 2020). This new module utilizes a matrix representation on CLM's biogeochemistry to reduce spin-up time by one order of magnitude (Liao et al., 2023; Luo et al., 2022), and had been previously leveraged for a PPE with the Organizing Carbon and Hydrology in Dynamic Ecosystems model (Huang et al., 2018). Our spin-up protocol featured 20 years in Accelerated Decomposition spin-up mode (see Lawrence et al. (2019) and Thornton and Rosenbloom (2005) for details), followed by 80 years of SASU, followed by 40 years of Native Dynamic spin-up mode (details on forcing data in Section 2.5). This protocol was designed to achieve sufficiently equilibrated carbon and nitrogen states, while minimizing computational time. This spin-up methodology did not always reach full equilibrium of soil carbon (Figure S1 in Supporting Information S1). However we believe that this spinup methodology is sufficient to robustly infer parameter rankings for carbon cycle processes, if not the exact absolute values of parameter effects.

2.3. Sparse Grid

Another control on model cost is spatial resolution. Most global CLM simulations utilize nominal 1° resolution, which equates to about 20,000 land grid cells. To reduce computational costs parameter perturbation experiments often use lower resolution, such as $4° \times 5°$ (Dagon et al., 2020). As opposed to selecting a coarse grid resolution, here we apply an alternative sparse grid approach by using a clustering algorithm to achieve an alternative low

Table 1
Clustering Inputs Categorized Into Three Groups, With the CTSM Variable Name in Parentheses

Climate variables	Ecosystem state variables	Ecosystem flux variables
2 m air temperature (TSA)	Leaf area index (TLAI)	Gross primary production (GPP)
Atmospheric rain (RAIN)	Ecosystem carbon (TOTECOSYSC)	Heterotrophic respiration (HR)
Atmospheric snow (SNOW)	Ecosystem nitrogen (TOTECOSYSN)	Autotrophic respiration (AR)
2 m specific humidity (Q2M)	Soil ice (TOTSOILICE)	Net biome production (NBP)
Solar radiation (FSDS)	Soil liquid water (TOTSOILLIQ)	Total liquid runoff (QRUNOFF)
	Snow cover fraction (FSNO)	Sensible heat (FSH)
		Latent heat (EFLX_LH_TOT)

resolution configuration. The goal is to more strategically distribute grid cells, opting for more grid cells in areas that are especially distinct or variable, and fewer in areas that can be easily approximated based on output elsewhere (e.g., neighboring grid cells). Multivariate spatio-temporal clustering (MVSC) has been utilized to extract patterns of climatological significance from climate model output (Hoffman et al., 2005) and applied to design a representativeness-based sampling network (Hoffman et al., 2013). Instead of lowering resolution by coarsening a rectilinear grid, we used MVSC to strategically remove effectively redundant grid cells. We used k-means clustering to identify groups of grid cells with similar dynamics based on a 2° transient simulation (1850–2014) using the CLM-PPE codebase. We selected one representative grid cell from each cluster to stand in for the entire cluster. The representative grid cell is whichever is located nearest the cluster centroid in climate space. The set of representative grid cells comprise a "sparse grid", which are used in lieu of a "coarse" grid. To recompose mapped output and compute global means, the output from the representative grid cell is substituted for all members of the cluster cohort. This naive interpolation will introduce errors that grow in relative magnitude as the domain of interest shrinks.

Clustering was based on a subset of 18 CLM variables (Table 1). The MVSC algorithm delineates clusters based on the mean and variability of each variable at each grid cell computed for six 30-year climatology windows (1865–1894, 1895–1924, …, 1985–2014). Clusters were delineated to equalize the multi-dimensional variance across the user-specified number of groups, k. We tested 15 values of k, ranging from 10 to 800. Utilizing the ILAMB2.5 benchmarking software (Collier et al., 2018), we calculated skill scores to quantify how well each candidate sparsegrid mirrored the full grid output (Figure S2 in Supporting Information S1). We opted for a 400-cluster sparsegrid, to balance computational cost against model fidelity. Because our emphasis is on vegetated regions, we masked out Antarctica within the clustering algorithm, whereby we do not provide any output below 60°S.

2.4. Parameter Identification and Ranges

Identification of the complete parameter set of a land surface model is in itself a non-trivial exercise, as in practice, many empirical constants are hard-coded and not amenable to systematic perturbation (Cuntz et al., 2016; Mendoza et al., 2015). For the purposes of this activity, we identified a broad set of 211 CLM-BGC parameters, and enabled their modification via the model parameter file. Several domains were not perturbed, including crops, methane emissions, emissions of biogenic volatile organic compounds and dust, urban properties, and river flow parameters. The crop model was not utilized as our initial focus is on unmanaged land. Of the 211 parameters, 145 are scalar-valued, with a single parameter value acting globally. The remaining 66 parameters include a plant functional type (PFT) dimension, with the potential for different parameter values for each unique PFT. For any given PFT parameter, all PFTs were perturbed in tandem (i.e., all 16 PFTs to their maximum values and then to their minimum values), but with potentially different ranges for each PFT. Explicit parameter ranges for each parameter (and each PFT) are provided in a supplemental spreadsheet.

Given the large parameter space (Table 2), to conduct an initial analysis of the response of the model to parametric uncertainty, we decided to vary each parameter independently, exploring the impact of low and high values. To define parameter ranges we created an online spreadsheet and solicited domain-area experts to provide a minimum and maximum value for each parameter. In some cases literature values were directly utilized, but in many

Table 2
In Total, 211 Parameters Were Perturbed in This Experiment

Parameter	Description	Model domain	Varies by PFT?
cv	Turbulent transfer coefficient	Sensible, latent heat and momentum fluxes	
dleaf	Leaf characteristic length	Sensible, latent heat and momentum fluxes	yes
d_max	Dry surface layer (DSL) parameter	Sensible, latent heat and momentum fluxes	
frac_sat_soil_dsl_init	Fraction of saturated soil at which DSL initiates	Sensible, latent heat and momentum fluxes	
fff	Decay factor for fractional saturated area	Hydrology	
liq_canopy_storage_scalar	Canopy-storage-of-liquid-water parameter	Hydrology	
maximum_leaf_wetted_fraction	Maximum leaf wetted fraction	Hydrology	
sand_pf	Perturbation factor for sand content of soil column	Hydrology	
sucsat_sf	Scale factor for minimum soil water potential	Hydrology	
medlynintercept	Medlyn intercept of conductance-photosynthesis relationship	Stomatal resistance and photosynthesis	yes
medlynslope	Medlyn slope of conductance-photosynthesis relationship	Stomatal resistance and photosynthesis	yes
tpu25ratio	Ratio of tpu25top to vcmax25top	Stomatal resistance and photosynthesis	
jmaxb0	Baseline proportion of nitrogen allocated for electron transport	Photosynthetic capacity	
jmaxb1	Response of electron transport rate to light	Photosynthetic capacity	
slatop	Specific leaf area at top of canopy	Photosynthetic capacity	yes
theta_cj	Empirical curvature parameter for photosynthesis colimitation	Photosynthetic capacity	yes
wc2wjb0	Baseline ratio of wc:wj	Photosynthetic capacity	
FUN_fracfixers	fraction of carbon available for N fixation	Fixation and uptake of Nitrogen	yes
KCN	Nitrogen acquisition parameter group	Fixation and uptake of Nitrogen	yes
kmax	Plant segment max conductance	Plant hydraulics	yes
krmax	Root segment max conductance	Plant hydraulics	yes
psi50	Water potential at 50% loss of conductance	Plant hydraulics	yes
nstem	Stem number	Biomass heat storage	yes
lmr_intercept_atkin	Intercept in the calculation of leaf maintenance respiration	Plant respiration	yes
froot_leaf	Allocation parameter: new fine root C per new leaf C	Carbon and nitrogen allocation	yes
leafcn	Leaf C:N	Carbon and nitrogen allocation	yes
leaf_long	Leaf longevity	Vegetation phenology and turnover	yes
cpha	Activation energy for cp	Acclimation parameters	
jmaxhd	Deactivation energy for jmax	Acclimation parameters	
kcha	Activation energy for kc	Acclimation parameters	
lmrha	Activation energy for lmr	Acclimation parameters	
lmrhd	Deactivation energy for lmr	Acclimation parameters	
tpuha	Activation energy for tpu	Acclimation parameters	
tpuse_sf	Scale factor for tpu entropy term	Acclimation parameters	
vcmaxha	Activation energy for vcmax	Acclimation parameters	
vcmaxhd	Deactivation energy for vcmax	Acclimation parameters	

Note. These are the parameters mentioned within the main text. A complete set of parameters, with ranges, is included in a spreadsheet in Supporting Information S1.

cases, expert judgment was used or else a default perturbation of ±20%. The parameter range spreadsheet is provided within Supporting Information S1.

In some cases parameters could not or should not be perturbed independently, because they feature inherent co-dependence. In the case of nitrogen fixation costs, we opted to perturb the parameters independently, but also as a group (“KCN”), to enforce a change in overall nitrogen limitation in lieu of switching from one uptake pathway to another. For the grouped simulations, the given set of parameters were perturbed in tandem, either to their

Table 3
Forcing Scenarios

Name	Meteorology	CO_2 (ppmv)	N addition	Description
CTL2010	2005–2014	367	–	Control experiment
C285	2005–2014	285	–	Low CO_2
C867	2005–2014	867	–	High CO_2
AF1855	1851–1860	367	–	Pre-industrial climate
AF2095	2091–2100	367	–	Late century climate (SSP3-7.0)
NDEP	2005–2014	367	5g N m^{-2} y^{-1}	Enhanced nitrogen deposition

minimum or maximum values. We likewise opted to perturb soil hydraulics directly in two distinct ways, first by directly perturbing the hydraulic properties (e.g., saturated hydraulic conductivity) and also in a parallel set of simulations perturbing the soil texture (percent sand, silt, and clay) and allowing the soil hydraulic properties to respond through the pedotransfer functions. Future experiments could utilize a multifactor perturbation sampling strategy to resolve parameter interdependencies and tradeoffs, such as along major axes of ecosystem function (Migliavacca et al., 2021).

For a small percentage of parameters, the perturbation ranges were severe enough to eradicate a given PFT across all or part of its geographical range (Figure S3 in Supporting Information S1). In the case that one or more PFTs did not survive a parameter perturbation, we iteratively re-ran the simulation with slightly less extreme parameter values until the PFT survived. A PFT was classified as surviving in a given gridcell if it achieves an leaf area index (LAI) of at least 0.1 m2/m2 at any point after the spinup period. A PFT was classified as surviving globally if at least 50% of the area that was alive with the default parameters is alive in the perturbed simulation. For the analyses in this paper, we constrained the ranges of several parameters based on survivability in the low CO_2, pre-industrial climate, and control climate ensembles to improve the chances that all PFTs would be able to survive in a transient 1850 to present-day simulation. We did not constrain parameter ranges based on future climate, high CO_2, or increased nitrogen deposition. In total, 16 of the 211 parameters had the parameter range constrained to some extent. Original and constrained ranges are available in Supporting Information S1 spreadsheet.

In one case, a significant CLM bug was identified after the completion of our primary experiment (https://github.com/ESCOMP/CTSM/issues/2120). Namely that the values for two nitrogen uptake parameters had been transposed in Brzostek et al. (2014). Resolving the issue resulted in an approximately 5% decrease in gross primary production (GPP) of (Figure S4 in Supporting Information S1) in the control simulations, and also significantly altered the size of the various parameter effects. Any PPE should be seen as intrinsically connected to its underlying codebase (bugs and all), and, reflecting this, we did not rerun the rest of the PPE. Therefore the majority of simulations contain the erroneous parameter values, and may slightly overestimate GPP. The interacting effect of this bug fix with all of the other parameter perturbations was not explored. The bug has since been resolved in later model versions and will be tested in future PPEs.

2.5. Experimental Design

For each of the 211 parameters, we ran 12 simulations, one for each of the minimum and maximum parameter values, across six different forcing scenarios. Each simulation ran for 150 years, with the first 140 for spin-up (as described above in Section 2.2), followed by a 10-year period for analysis. It is important to note that these are equilibrium-style simulations, where the land has been allowed to reach steady state conditions, relative to the climate forcing. This is achieved by repeating a single 10-year forcing data set 15 times during the spin-up procedure and production phase of our simulations. As such, these simulations are not directly comparable to transient simulations or satellite observations, where present-day vegetation has not yet reached equilibrium with present-day climate and CO_2 concentrations.

We opted for six different forcing scenarios to understand the intersection of parameter effects with different forcing agents associated with climate change (Table 3). The GSWP3v1 reanalysis product (http://hydro.iis.u-tokyo.ac.jp/GSWP3/) served as our atmospheric forcing, as it is the default forcing data for CLM5 (Lawrence et al., 2019). Our six forcing scenarios were: a 2010-era control climate, high and low CO_2, future and pre-

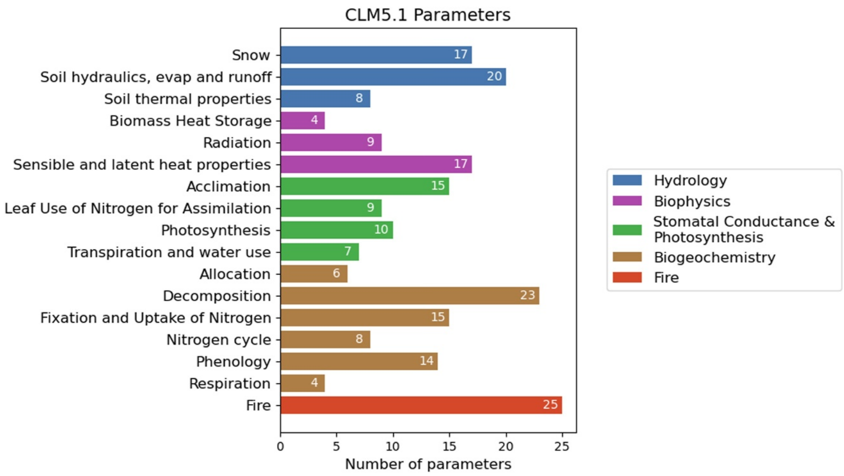

Figure 1. 211 parameters were perturbed across the various domains of the land model.

industrial climate anomaly forcing, and enhanced nitrogen deposition (see Table 3). We applied climate and CO_2 anomalies independently, in order to disentangle their effects on parameter rankings. Future and pre-industrial climate anomaly forcing data sets were prepared by adding GSWP3v1 anomalies from 2005 to 2014 to a mean climate change signal. We inferred the mean climate change signal using the CESM2 large ensemble experiment (Rodgers et al., 2021), computed as the average of the difference between the period of interest (either 1850s or 2090s) and present day for six atmospheric forcing variables (temperature, humidity, precipitation, wind, longwave and shortwave radiation; see Figure S5 in Supporting Information S1 for climatologies). The future climate scenario utilizes the end-of-century climate from the SSP3-7.0, a high-emissions scenario, which was chosen to align with the existing CESM2 large ensemble (Rodgers et al., 2021). Since completing the experiment we uncovered a software bug affecting the AF1855 and AF2095 simulations. Namely that climate anomalies were not applied to certain coastal gridcells (Figure S6 in Supporting Information S1). In all this affected less than 3% of land area, and did not significantly alter parameter rankings (Figure S7 in Supporting Information S1).

2.6. Biome Definitions

In addition to analyzing parameter effects on various output variables globally, we were also interested in understanding parameter effects on different biomes. As such, we categorized each of the 400 sparse grid cells according to their Whittaker biome (Whittaker, 1970), which delineates biomes based on average temperature and precipitation. Biomes were delineated based on the atmospheric forcing temperature and precipitation data averaged across 2005 to 2014 (Figure S8 in Supporting Information S1). Overall there are nine vegetated biomes and an additional ice sheet biome, where no vegetation is present (Figures S9 and 10 in Supporting Information S1). These biomes are somewhat arbitrary, and the biomes would vary slightly based on the climatology period, but they provide some coarse delineation that we feel has meaningful ecological significance and have found useful for analyzing this ensemble.

3. Results

3.1. Computational Cost

Given the very large size of the CLM5-BGC parameter space (Figure 1), and the relatively expensive computational cost of standard model configurations, a major goal of this project was to develop a fast configuration of CLM5-BGC that would allow for a large number of simulations. Combining the SASU spinup approach with the sparse grid formulation (see Section 2.3 for details) yielded a configuration approximately 500 times less costly than the 1-degree configuration most often used for CLM simulations (Figure 2). There are 22,648, 5,666, and 1,764 land grid cells in standard 1°, 2°, and $4° \times 5°$ CLM5.1 simulations, respectively, as compared to just 400 grid cells in the sparsegrid. Likewise, whereas previous spin-up methodologies required 1,500 years or longer to satisfy equilibrium criteria, the SASU approach enabled by matrix representation yielded satisfactory spin-up efficiency for this experiment within 140 years (Figure S1 in Supporting Information S1).

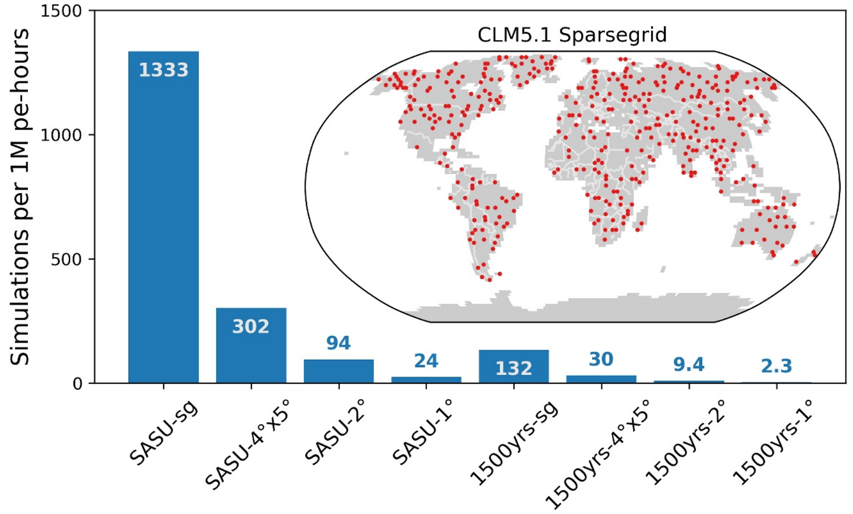

Figure 2. The approximate number of simulations afforded by 1 million core-hours on the Cheyenne supercomputer for a range of CLM configurations. Configurations are labeled according to spin-up procedure (Semi-Analytic spin-up mode or the standard 1,500-year spinup) and horizontal resolution ("sg" signifies sparsegrid). The inset map shows the locations of the 400 sparse grid cells. See Section 2 for spin-up and sparsegrid details.

Choosing the number of clusters to generate the sparsegrid involved balancing the computational savings against representational fidelity. Given more clusters, the sparsegrid will generally provide a better approximation of output from the full grid. As one example, accurate reconstruction of global photosynthesis could be achieved with a relatively small number of clusters, with $R^2 > 0.95$ achieved with only 200 clusters (Figure 3). Performance

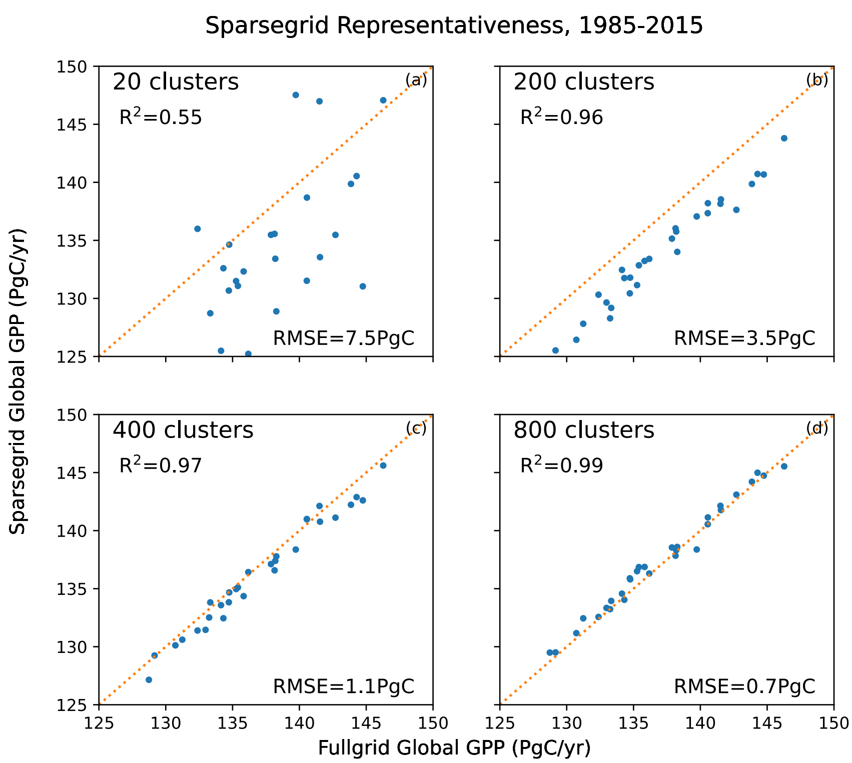

Figure 3. Sparsegrid versus fullgrid (2° resolution) global annual gross primary production across the last 40 years of a transient CLM5.1 simulation with different cluster number settings (a–d). We opted for 400 clusters to balance computational cost against representativeness.

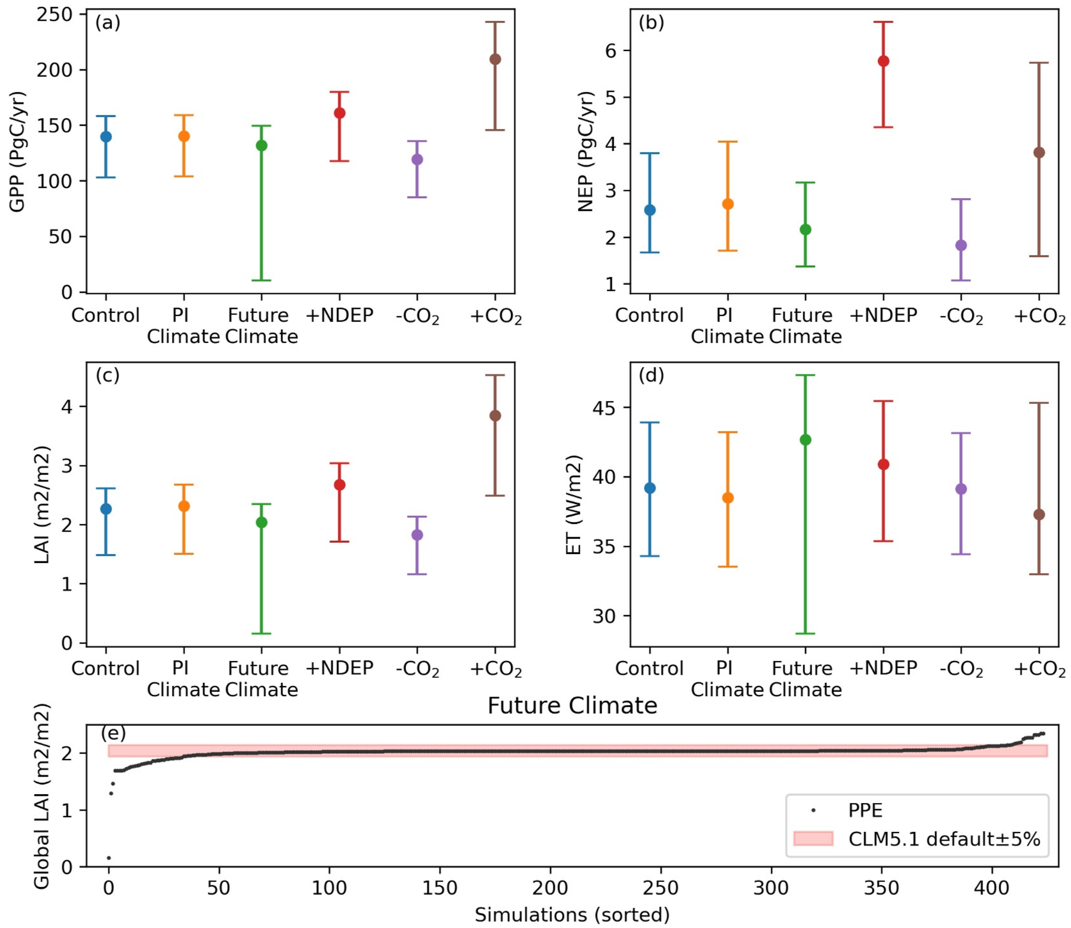

Figure 4. Global annual mean values for the default simulations (circles) and ensemble ranges (bars) for (a) gross primary production, (b) net ecosystem production, (c) leaf area index (LAI), and (d) evapotranspiration across six forcing scenarios. (e) Global LAI across the full set of parameter perturbation simulations under the future climate scenario. Most simulations fall within 5% of the default LAI, with a small number of parameters generating larger effects. One perturbation, reducing sand heat capacity by 20% caused a very low LAI.

continues to improve with added clusters, but with marginal returns (Figure S2 in Supporting Information S1). We had hoped to see a clear point of diminishing returns across a wide range of variables as we increase the number of clusters. Instead, for most variables, we observe a quick reduction of bias approaching 200 clusters, but after this point, improvement slows but does not saturate (Figure S2 in Supporting Information S1). We chose 400 clusters, because it provides satisfactory performance across a range of important metrics, while affording a sufficient number of simulations to perform the full experiment within our computational budget. The main goal of the sparsegrid is to obtain a coarse understanding of parameter effects at low computational cost, which could later be improved by full grid simulations or alternative sparsegrids with more clusters.

This model configuration allowed for over 2,500 simulations across our six forcing scenarios. Overall, we found substantial impacts of parametric sensitivity on model behavior, in some cases exceeding the magnitude of scenario effects (Figure 4). In the control scenario, the global sum of GPP ranged from 103 to 158 PgC per year across all ensemble members, and net ecosystem production (NEP) ranged from 1.7 to 3.8 PgC per year. NEP was especially variable in the high CO_2 scenario, varying from 1.6 to 5.7 PgC per year. One perturbation, reducing the heat capacity of sand by 20%, proved destructive in the future climate scenario, resulting in inhospitably hot soil conditions and widespread plant death. A perturbation of ±20% may exceed the reasonable range for this parameter, but this simulation was instructive for exposing the model's response to hot soils. Carrying out an extensive PPE increases the possibility of exposing unexpected model behavior, including unforeseen tipping points, brittle parameterizations, and/or bugs.

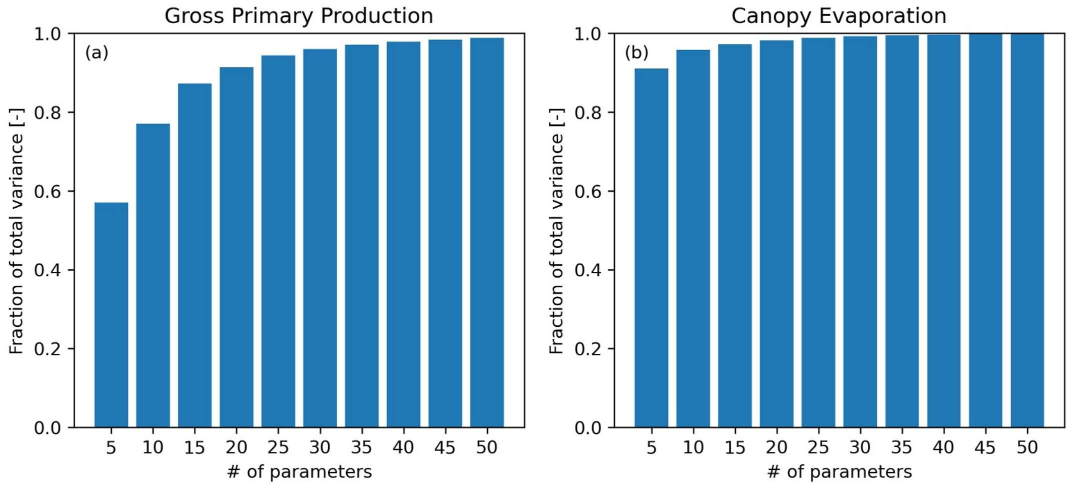

Figure 5. Cumulative fraction of variance explained by the most influential parameters on gross primary production (GPP) and canopy evaporation. Approximately 96% of canopy evaporation variance can be explained by the 10 most influential parameters, whereas 30 parameters are required to explain 96% of GPP variance.

3.2. Main Results

A small fraction of parameters tends to explain a large amount of the ensemble variance for any given variable or metric (Figure 5). GPP, which is a complicated process, requires 20 parameters to explain 90% of the control ensemble variance. Conversely, canopy evaporation is governed by fewer parameters. Just four parameters (maximum leaf wetted fraction, liquid canopy storage, the turbulent transfer coefficient, and leaf characteristic length) explain upwards of 90% of the canopy evaporation variance across all the forcing scenarios (Figure S11 in Supporting Information S1). In general, most of the parameter perturbations had a relatively small impact on any given output variable. Though this experiment encompasses 211 parameters, no single variable responds to all 211 parameters. In fact, the effective parameter space dimensionality for any given output variable tends to be much smaller than 211. While the principle of land models is that these are all interlinked via complex feedback processes, many parameters nonetheless have impacts that are mostly limited to their own domain.

3.3. Comparing Scenarios and Biomes

We opted to rank parameters by absolute parameter effect, given that perturbations while unequal in relative size, were defined with the same uncertainty range in mind. We repeated the full set of parameter perturbations across the six forcing scenarios in Table 3 (Figure 4). This ensures that we can identify parameters that are important not just under present-day conditions, but also parameters that control the response to forcing changes (CO_2, nitrogen deposition and climate). For example, the parameters that control present-day NEP differ from the parameters that control NEP in the high CO_2 scenario (Figures 6a and 6b). NEP increases by approximately 50% with the increase of CO_2 from 367 to 867 ppm when using the CLM5.1 default parameters, but can actually decrease with certain parameter settings (Figure 6c). One key parameter is *tpuse_sf*, which is a scalar perturbation factor influencing the triose phosphate limitations on photosynthesis. Triose phosphate limitation is difficult to quantify and has not been observed to significantly hinder photosynthesis under present-day CO_2 concentrations, but may play a larger role in the future (Kumarathunge et al., 2019; Lombardozzi et al., 2018). This parameter is not particularly influential on NEP at 367 ppm (ranked 39, not shown), but is third most influential when we increased CO_2 to 867 ppm.

The future climate experiment is 4.3 K warmer averaged over our study domain (land-only, Antarctica excluded, Figure S5 in Supporting Information S1). This contributed to an increase in ET of 8.8% relative to the control simulation when using the default CLM5.1 parameters (Figure 6f). While the parameters that most strongly affect global average ET in both the control and future climates are generally the same (Figures 6d and 6e), a distinct set of parameters govern the change in ET due to warming (Figure 6f). While the control simulation is primarily influenced by hydrology and stomatal conductance parameters, the response of ET to warming is largely controlled by photosynthesis acclimation parameters (i.e., *tpuse_sf*, *kcha*, *vcmaxha*, *vcmaxhd*, *lmrhd*, and *cpha*).

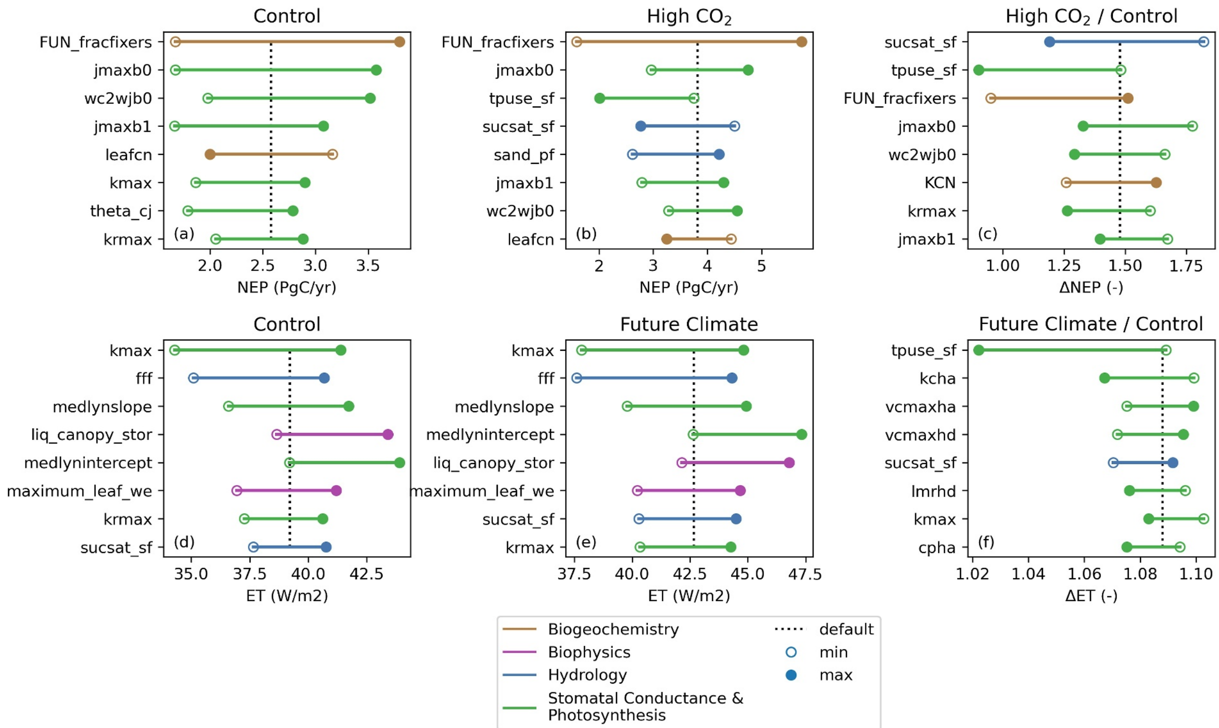

Figure 6. The eight most influential parameters (ranked by absolute parameter effect) on net ecosystem production (top row) in the control simulations (a), high CO_2 simulations (b), and the relative response to high CO_2 (c), as well as on evapotranspiration (ET, bottom row) in the control simulations (d), future climate simulations (e) and the relative response to future climate (f). The solid lines span the range between the two simulations for each parameter, with an open circle indicating the simulation where a parameter is set to its minimum value and a filled circle for its maximum. The lines are color-coded according to scientific domain. The dashed lines indicate the value in the simulation with default CLM5.1 parameters. The parameters that are most influential on present-day climate (a), (d) can differ significantly from the parameters that control the response to forcing perturbations (c), (f).

In a coupled ESM, land-atmosphere interactions would modulate these parameter effects, which could dampen parameter effects. While we feel relatively confident identifying influential parameters using land-only simulations, projecting their exact impacts in a coupled framework will require coupled model sensitivity tests.

We repeated parameter ranking analyses in each of nine Whittaker biomes (see Section 2.6). We found that the parameters controlling LAI, for example, vary significantly by biome (Figure 7). Plant hydraulics parameters were the most important in the tropical rain forest, photosynthetic capacity in the boreal forest, and runoff and soil evaporation in the temperate grassland/desert biome. The most influential parameters for LAI globally included parameters that were important in each of these three biomes.

In this paper, we focus primarily on global and biome-level parameter rankings. However it is also possible to inspect the geographic footprint of parameter perturbations by projecting the sparsegrid output to standard lat/lon coordinates (see Section 2.3 for details). For example, perturbing the *medlynslope* parameter (which controls stomatal conductance) has a large effect on global runoff, but primarily via its effects in vegetated areas (Figure 8). Because the number of potential variables, parameters, and geographical ranges of interest to the wider CLM community is larger than we can document here, we provide a tool that can be used to explore an extended diagnostics set which summarizes the >2 TB of output data via approximately 2,000 plots (https://webext.cgd.ucar.edu/I2000/PPEn11_OAAT). The diagnostics website includes ranking plots (as in Figure 7) and maps of parameter effects (as in Figure 8). These plots are repeated across a combination of model output variables and model parameters. Likewise figures are repeated for each of the various forcing scenarios.

4. Discussion

In this project, we identified and perturbed 211 CLM parameters to create a large one-at-a-time PPE. There were several barriers to perturbing the full set of CLM-BGC parameters. First, many parameters had not been officially

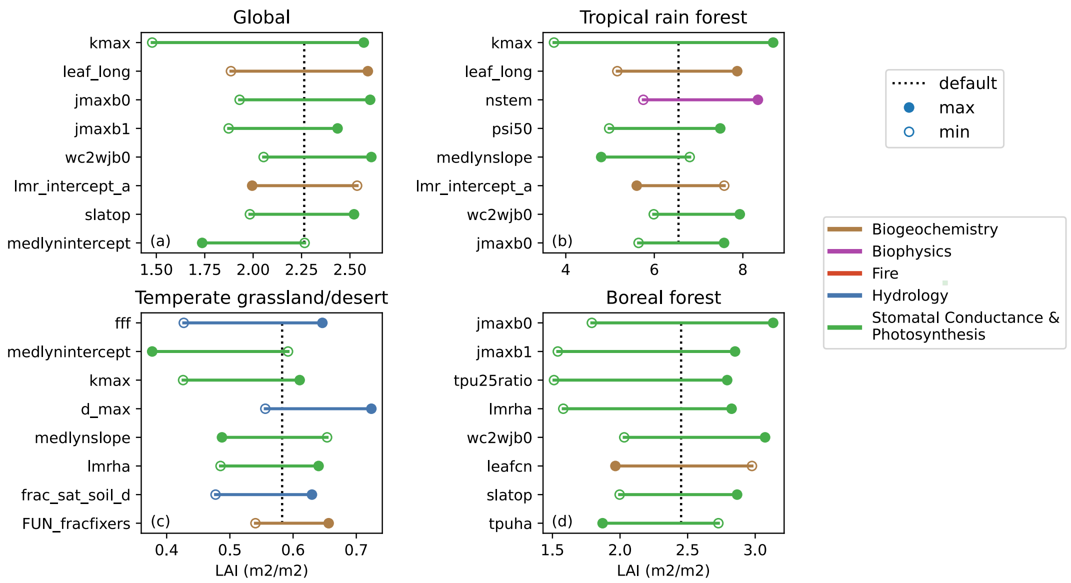

Figure 7. The eight most influential parameters (ranked by absolute parameter effect) on leaf area index within the control ensemble, globally and within three biomes. Lines and circles are as described in Figure 6. Parameter rankings vary by biome, with the global rankings seeming to reflect contributions from each of these three biomes.

identified as parameters. In such cases, we identified hard-coded values, established an appropriate parameter name, and extracted that parameter to the CLM parameter file for easier manipulation. Though we perturbed a large number of parameters across a variety of CLM processes (Figure 1), this still does not cover the full set of CLM parameters, as some processes were not included, such as crop phenology and management. In carrying out this process of parameter identification, we likewise unearthed many nuances in the epistemology of what "parameters" are in the context of Earth system models. For instance, is soil texture a forcing variable or a parameter? Defining the parameters within a given model structure can be somewhat subjective, such that it would be difficult to collate a comprehensive or definitive set of parameters for a model like CLM. e.g., for several empirical regressions, we scaled the slope and intercept terms in tandem, because we lacked sufficient information to define two independent perturbation ranges.

The second challenge involved defining a perturbation range for each parameter. We solicited expert judgment to set a minimum and maximum reasonable value for each parameter. In some cases, literature values were explicitly used (e.g., for the slope parameter of the stomatal conductance model, "medlynslope", Lin et al. (2015)), but the most common range was ±20%. It is exceedingly difficult to set parameter ranges that sample comparable probability density, even in a univariate experiment, like this one. The number of parameters is large, with many

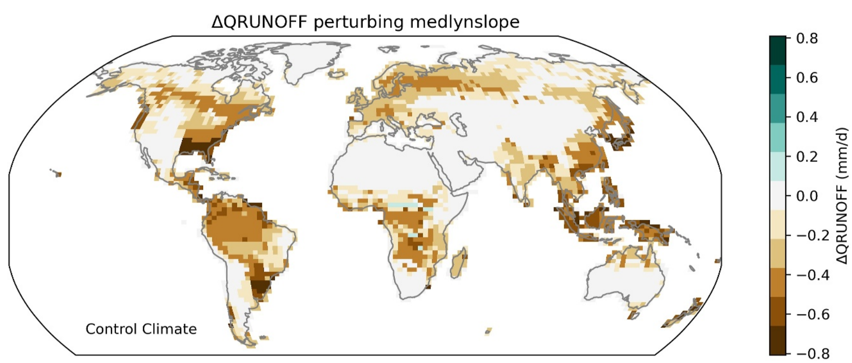

Figure 8. Map of the effect of perturbing the medlynslope stomatal conductance parameter on runoff within the control ensemble. Increasing medlynslope tends to reduce runoff, but only in regions with sufficient vegetation activity. This is one example of many plots available online (https://webext.cgd.ucar.edu/I2000/PPEn11_OAAT) in a broader diagnostics set.

lacking sufficient empirical backing for robust range evaluation. Defining appropriate ranges for the full suite of land model parameters is a major activity of the scientific subdomains that comprise land models (Kattge et al., 2020). Even with parameters that have an empirical basis from field studies, uncertainty quantification is difficult due to the scale mismatch between site-level observations and land model grid cells (Y. Liu et al., 2021). As such, we cannot claim that the parameters are equivalently sampled, whereby parameter effects and parameter rankings could be subject to sampling asymmetries.

The third challenge involved managing computational cost. Quantifying parameter effects is a necessary prerequisite to automated calibration and uncertainty quantification. Because the parameter space of CLM is quite large and the model response is potentially non-linear, large ensembles are required to adequately resolve response surfaces. With standard CLM configurations, such ensembles would far exceed our computing resources. As such, the computational cost of inferring parameter effects is a major constraint. We were able to reduce ensemble generation computational cost 500x by strategically reducing the number of model grid cells (Figure 3) and by leveraging a Semi-Analytic Spin-Up (SASU) offered by the matrix approach to land biogeochemistry models (Liao et al., 2023; Lu et al., 2020; Luo et al., 2022) (see Sections 2.2 and 2.3 for details). As long as computational resources remain constrained, designing faster model configurations and efficient sampling strategies will be important for effective model calibration and uncertainty quantification.

Our sparsegrid formulation allows us to run far fewer gridcells than a standard rectilinear grid, but at the cost of model fidelity. We visualized interannual variations in global GPP for different sparsegrids (Figure 3) to demonstrate the tradeoff between number of clusters and representational accuracy. A wider array of scoring metrics (Figure S3 in Supporting Information S1, using the ILAMB2.5 scoring methodologies, Collier et al., 2018) shows that scores continue to improve beyond 400 clusters (our choice for these experiments). For example, bias scores for total carbon stocks are consistently lower than other variables and may require a larger number of clusters or an alternative clustering strategy that accounts for the outsized influence of high latitude grid cells. The sparsegrid configuration will generate a coarse understanding of parameter effects at low computational cost, that can be honed by iteratively increasing the number of clusters or by supplementing with full grid simulations.

For this experiment we opted for a one-at-a-time perturbation strategy, testing a minimum and maximum value for each parameter. We found that parameter effects could be quite large, in some cases exceeding the effects of our forcing scenarios (Figure 4). That said, the majority of parameter perturbations had small effects for any specific model variable or metric, such that a majority of simulations were clustered around the default simulation (Figure 4e). In the case of canopy evaporation, for example, more than 95% of the ensemble variance could be explained by the 10 most influential parameters (Figure 5). This indicates that there may be tractable parameter estimation sub-problems if global calibration of all 211 parameters proves infeasible.

We opted for six forcing scenarios to identify parameter effects not just in present-day but in response to pre-industrial or future climate conditions. The parameters that were most influential under present-day conditions did not necessarily match the set of parameters controlling the response to future forcing (Figure 6). We found that many acclimation parameters, which were not as important for determining present-day ET, were among the most influential on the response of ET to future climate. This emphasizes the value of testing models not only under mean-state conditions, but also under experimentally perturbed conditions under global change analogs (Wieder et al., 2019). These acclimation parameters likewise would not be constrained by observations of mean ET, indicating that alternative metrics may be required or that we may lack capacity to reduce projection uncertainties associated with these parameters.

Furthermore, because parameter values in CLM are time-invariant, plants must be able to survive pre-industrial conditions in order to be present for the later stages of a transient simulation, such as those performed for many model intercomparison experiments. Several perturbations did indeed cause survivability issues (Figure S2 in Supporting Information S1), in which case we opted to constrain the parameter ranges until all PFTs were able to survive in the low CO_2, control, and pre-industrial climate simulations. In future experiments we may instead opt for fully transient simulations from 1850 to 2100, as the computational cost is comparable given the overhead of independent spinups for each forcing scenario.

Just as the most influential parameters varied by forcing scenario, parameter rankings varied significantly by biome (Figure 7). Different PFTs present varying trait strategies, and different climates impose unique challenges to vegetation activity, such as light availability, extreme temperatures, or water stress. As such, perturbations that serve to ease or exacerbate the most substantial environmental stress for the set of PFTs in a given biome will have greater effect. For example, modifying photosynthetic capacity (via jmaxb0, which sets baseline proportion of nitrogen allocated for electron transport) has a large effect in boreal forests, as light is relatively limiting at high latitudes. Understanding the regional signatures of parameter perturbations is important for model calibration. Previous efforts have experienced difficulty resolving regional biases due to a limited amount of parameter flexibility. For example (Dagon et al., 2020), could not simultaneously optimize plant productivity in the Amazon and Sahel regions. In some cases, perturbing PFT-specific parameters independently could aid in providing regional levers for bias reduction. In our experiments, PFT-specific parameters were perturbed in unison, but their effects can be analyzed by PFT for many variables that are modeled at the PFT-level, such as photosynthesis and transpiration. To first order, we expect that PFT-specific parameter effects will be independent, but CLM does feature some implicit competition among natural vegetation PFTs, such as through a shared soil moisture reservoir, which could impose some cross PFT interaction.

A one-at-a-time PPE cannot capture parameter interactions, and our min/max sampling protocol precludes diagnosing non-linearities. As such, this data set will be insufficient for most calibration activities or for estimating overall parametric uncertainty (Raoult et al., 2024). A primary utility of our data set is that we can diagnose parameter effects without the uncertainty contributed by an emulator or a regression model. As such, it is easy to diagnose which parameters are most influential on a given process. We have published a large set of ensemble diagnostic plots online (https://webext.cgd.ucar.edu/I2000/PPEn11_OAAT/), which serves as a valuable enhancement to our model technical documentation. Now, in addition to seeing the definition of a given parameter, and the relevant equations, a model user can easily investigate the magnitude and spatial patterns of its effects (e.g., Figure 8). This could be useful for investigating obvious model deficiencies, such as regional LAI biases, or to understand potential model responses under different forcing scenarios (e.g., plant survival under low CO_2 conditions).

Another caveat is that our PPE utilizes land-only simulations, with prescribed atmospheric forcing. This is likely to be adequate for estimating many land surface fluxes and pools, in particular of carbon, but precludes the influence of land-atmosphere feedbacks. This could lead to biased estimates of water and energy fluxes, as atmospheric feedbacks can modify the state of the land surface itself (Laguë et al., 2019). Likewise, fire parameters, which were not among the most important parameters for global ET, NEP, or LAI, may still influence the coupled model when coupled fire emissions are allowed to modify aerosol loads. Additional work is ongoing to extend this work in a coupled model framework to quantify how much, where, and for which processes atmospheric feedbacks are likely to modulate land parameter effects substantially. Furthermore our experiments are conducted at nominal 2-degree resolution, certain processes subject to land heterogeneity may respond differently at higher resolution, such as permafrost depth in the Arctic (Schickhoff et al., 2024). Our expectation is that relative parameter rankings are likely robust, but that the absolute value of parameter effects may be significantly biased relative to their effects in the coupled model.

5. Conclusion

The primary deliverable of this project is the PPE data set that captures the effects of 211 parameters on a wide range of CLM variables across six forcing scenarios. Several spin-off projects have leveraged this data set, including some that have already reached publication (Cheng et al., 2023; Yan et al., 2023a, 2023b). These projects utilized our ensemble to filter for parameters that influence the relevant study domains and variables, and performed follow-on experiments using the parameter ranges collated in Section 2.5. Maintaining, improving, and interpreting complex land models benefits from thoughtful investment in software to automate and routinize important components of the development process, such as the International Land Model Benchmarking system (Collier et al., 2018). Therefore we also see the infrastructure underpinning our experiment as a major deliverable. Our project has accelerated parameter exploration work within our collaborator network by providing:

- Parameter ranges for more than 200 parameters
- Purpose-built unstructured grid (sparsegrid)
- Accelerated spinup procedure
- Ensemble generation scripting toolchain
- Global, PFT, and biome-level parameter sensitivity diagnostics across six forcing scenarios

The software, data sets, and analysis tools generated to enable and utilize this experiment will greatly reduce the burden of generating future PPEs and parameter sensitivity experiments. Due to the large parameter space of comprehensive land models like CLM, reducing model computational costs can expand the scope of feasible parameter estimation problems. All of the code to generate this ensemble is available via github. We have also made available the collated parameter ranges and the output from this ensemble (see the Open Research Section for details).

We have already started a follow-on activity that perturbs a subset of important parameters with a more exhaustive sampling approach to work toward calibration of LAI. By investing in the infrastructure that we introduce here, all of our subsequent parameter perturbation experiments have required much less time and effort. We continue to extend our efforts in this domain toward model calibration and uncertainty quantification, and we anticipate repeating this foundational one-at-a-time experiment with subsequent model releases. Reducing both the human and computer time required for these experiments is allowing us to repeat them more frequently through the model development cycle.

Data Availability Statement

The model code for this experiment is contained in a development tag of the CTSM (https://github.com/ESCOMP/CTSM/tree/branch_tags/PPE.n11_ctsm5.1.dev030). The CTSM component set longname is: 2000_DATM%GSWP3v1_CLM51%BGC_SICE_SOCN_SROF_SGLC_SWAV_SIAC_SESP. A relatively small (∼700 MB), post-processed data set is publicly available online (Kennedy et al., 2025). All of the figures on our diagnostics website (https://webext.cgd.ucar.edu/I2000/PPEn11_OAAT) can be made from this one file. The raw data set will be available on the NSF-NCAR computing system, and can be made available online as well, if use cases exist beyond this smaller file. Ensemble generation scripts are available via: https://github.com/djk2120/ppe_tools. All of the code to make the figures in this manuscript is available via: https://github.com/djk2120/oaat_clm5_ppe.

Acknowledgments
This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement Number 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX) were provided by the Climate Simulation Laboratory at NSF-NCAR's Computational and Information Systems Laboratory. DK and DML acknowledge funding from the National Oceanic and Atmospheric Administration Modeling, Analysis, Predictions and Projections program (award number NA20OA R4310413) and funding from the NSF National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. KD is supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research (BER), Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program under Award Number DE-SC0022070 and National Science Foundation (NSF) IA 1947282. ALSS recognizes support from the National Science Foundation (NSF) (AGS-1553715, AGS-2330096). FL is supported by the National Key Research and Development Program of China (2022YFE0106500). CDK acknowledges support by the Director, Office of Science, Office of Biological and Environmental Research of the US Department of Energy under contract DE-AC02-05CH11231 through the Regional and Global Model Analysis Program (RUBISCO SFA). RF and BS acknowledge funding by the European Union's Horizon 2020 (H2020) research and innovation program under Grant Agreement Number 101003536 (ESM2025—Earth System Models for the Future) and 821003 (4C, Climate-Carbon Interactions in the Coming Century).

References

Alonso-González, E., Aalstad, K., Baba, M. W., Revuelto, J., López-Moreno, J. I., Fiddes, J., et al. (2022). The multiple snow data assimilation system (MuSA v1.0). *Geoscientific Model Development*, *15*(24), 9127–9155. https://doi.org/10.5194/gmd-15-9127-2022

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2020). Carbon–concentration and carbon–climate feedbacks in cmip6 models and their comparison to cmip5 models. *Biogeosciences*, *17*(16), 4173–4222. https://doi.org/10.5194/bg-17-4173-2020

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

Balaji, V., Couvreux, F., Deshayes, J., Gautrais, J., Hourdin, F., & Rio, C. (2022). Are general circulation models obsolete? *Proceedings of the National Academy of Sciences*, *119*(47), e2202075119. https://doi.org/10.1073/pnas.2202075119

Baldocchi, D., Novick, K., Keenan, T., & Torn, M. (2024). Ameriflux: Its impact on our understanding of the "breathing of the biosphere", after 25 years. *Agricultural and Forest Meteorology*, *348*, 109929. https://doi.org/10.1016/j.agrformet.2024.109929

Bastrikov, V., MacBean, N., Bacour, C., Santaren, D., Kuppel, S., & Peylin, P. (2018). Land surface model parameter optimisation using in situ flux data: Comparison of gradient-based versus random search algorithms, a case study using ORCHIDEE v1.9.5.2. *Geoscientific Model Development*, *11*(12), 4739–4754. https://doi.org/10.5194/gmd-11-4739-2018

Birch, L., Schwalm, C. R., Natali, S., Lombardozzi, D., Keppel-Aleks, G., Watts, J., et al. (2021). Addressing biases in Arctic–boreal carbon cycling in the community land model version 5. *Geoscientific Model Development*, *14*(6), 3361–3382. https://doi.org/10.5194/gmd-14-3361-2021

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

Brzostek, E. R., Fisher, J. B., & Phillips, R. P. (2014). Modeling the carbon cost of plant nitrogen acquisition: Mycorrhizal trade-offs and multipath resistance uptake improve predictions of retranslocation. *Journal of Geophysical Research: Biogeosciences*, *119*(8), 1684–1697. https://doi.org/10.1002/2014JG002660

Cheng, Y., Musselman, K. N., Swenson, S., Lawrence, D., Hamman, J., Dagon, K., et al. (2023). Moving land models toward more actionable science: A novel application of the community terrestrial systems model across Alaska and the Yukon River basin. *Water Resources Research*, *59*(1), e2022WR032204. https://doi.org/10.1029/2022WR032204

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

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The international land model benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018MS001354

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., et al. (2021). Process-based climate model development harnessing machine learning: I. A calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, *13*(3), e2020MS002217. https://doi.org/10.1029/2020MS002217

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*(6809), 184–187. https://doi.org/10.1038/35041539

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., et al. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the noah-mp land surface model. *Journal of Geophysical Research: Atmospheres*, *121*(18), 10676–10700. https://doi.org/10.1002/2016JD025097

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

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

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

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2018MS001453. https://doi.org/10.1029/2018MS001453

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., et al. (2006). Climate–carbon cycle feedback analysis: Results from the c4mip model intercomparison. *Journal of Climate*, *19*(14), 3337–3353. https://doi.org/10.1175/JCLI3800.1

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*(2), 511–526. https://doi.org/10.1175/jcli-d-12-00579.1

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. https://doi.org/10.5194/essd-14-4811-2022

Hawkins, L. R., Rupp, D. E., McNeall, D. J., Li, S., Betts, R. A., Mote, P. W., et al. (2019). Parametric sensitivity of vegetation dynamics in the triffid model and the associated uncertainty in projected climate change impacts on western u.s. forests. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2787–2813. https://doi.org/10.1029/2018MS001577

Henderson-Sellers, A., Pitman, A. J., Love, P. K., Irannejad, P., & Chen, T. H. (1995). The project for intercomparison of land surface parameterization schemes (PILPS): Phases 2 and 3*. *Bulletin of the American Meteorological Society*, *76*(4), 489–504. https://doi.org/10.1175/1520-0477(1995)076⟨0489:TPFIOL⟩2.0.CO;2

Hoffman, F. M., Hargrove, W. W., Erickson, D. J., & Oglesby, R. J. (2005). Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models. *Earth Interactions*, *9*(10), 1–27. https://doi.org/10.1175/EI110.1

Hoffman, F. M., Kumar, J., Mills, R. T., & Hargrove, W. W. (2013). Representativeness-based sampling network design for the state of Alaska. *Landscape Ecology*, *28*(8), 1567–1586. https://doi.org/10.1007/s10980-013-9902-0

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. https://doi.org/10.1175/BAMS-D-15-00135.1

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., et al. (2020). LMDZ6A: The atmospheric component of the IPSL climate model with improved and better tuned physics. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS001892. https://doi.org/10.1029/2019MS001892

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

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

Kennedy, D., Dagon, K., & Lawrence, D. M. (2025). One at a time parameter perturbation ensemble of the community land model, version 5.1. *Dryad*. https://doi.org/10.5061/dryad.j6q573nsn

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

Kumarathunge, D. P., Medlyn, B. E., Drake, J. E., Rogers, A., & Tjoelker, M. G. (2019). No evidence for triose phosphate limitation of light-saturated leaf photosynthesis under current atmospheric co2 concentration. *Plant, Cell and Environment*, *42*(12), 3241–3252. https://doi.org/10.1111/pce.13639

Laguë, M. M., Bonan, G. B., & Swann, A. L. S. (2019). Separating the impact of individual land surface properties on the terrestrial surface energy budget in both the coupled and uncoupled land–atmosphere system. *Journal of Climate*, *32*(18), 5725–5744. https://doi.org/10.1175/JCLI-D-18-0812.1

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. in press. https://doi.org/10.1029/2018MS001583

Liao, C., Lu, X., Huang, Y., Tao, F., Lawrence, D. M., Koven, C. D., et al. (2023). Matrix approach to accelerate spin-up of clm5. *Journal of Advances in Modeling Earth Systems*, *15*(8), e2023MS003625. https://doi.org/10.1029/2023MS003625

Lin, Y.-S., Medlyn, B. E., Duursma, R. A., Prentice, I. C., Wang, H., Baig, S., et al. (2015). Optimal stomatal behaviour around the world. *Nature Climate Change*, *5*(5), 459–464. https://doi.org/10.1038/nclimate2550

Liu, L., Fisher, R. A., Douville, H., Padrón, R. S., Berg, A., Mao, J., et al. (2024). No constraint on long-term tropical land carbon-climate feedback uncertainties from interannual variability. *Communications Earth and Environment*, *5*(1), 348. https://doi.org/10.1038/s43247-024-01504-6

Liu, Y., Konings, A. G., Kennedy, D., & Gentine, P. (2021). Global coordination in plant physiological and rooting strategies in response to water stress. *Global Biogeochemical Cycles*, *35*(7), 1–23. https://doi.org/10.1029/2020GB006758

Lombardozzi, D. L., Smith, S. J., andCheng, N. G., Dukes, J. S., Sharkey, T. D., Fisher, R. A., & Bonan, G. B. (2018). Triose phosphate limitation in photosynthesis models reduces leaf photosynthesis and global terrestrial carbon storage. *Environmental Research Letters*, *13*(7), 074025. https://doi.org/10.1088/1748-9326/aacf68

Lovenduski, N. S., & Bonan, G. B. (2017). Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters*, *12*(4), 044020. https://doi.org/10.1088/1748-9326/aa66b8

Lu, X., Du, Z., Huang, Y., Lawrence, D., Kluzek, E., Collier, N., et al. (2020). Full implementation of matrix approach to biogeochemistry module of CLM5. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002105. https://doi.org/10.1029/2020MS002105

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., et al. (2022). Matrix approach to land carbon cycle modeling. *Journal of Advances in Modeling Earth Systems*, *14*(7), e2022MS003008. https://doi.org/10.1029/2022MS003008

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

McNeall, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wiltshire, A., & Sexton, D. (2016). The impact of structural error on parameter constraint in a climate model. *Earth System Dynamics*, *7*(4), 917–935. https://doi.org/10.5194/esd-7-917-2016

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., & Gupta, H. (2015). Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, *51*(1), 716–728. https://doi.org/10.1002/2014WR015820

Migliavacca, M., Musavi, T., Mahecha, M. D., Nelson, J. A., Knauer, J., Baldocchi, D. D., et al. (2021). The three major axes of terrestrial ecosystem function. *Nature*, *598*(7881), 468–472. https://doi.org/10.1038/s41586-021-03939-9

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, *430*(7001), 768–772. https://doi.org/10.1038/nature02771

Peatier, S., Sanderson, B. M., Terray, L., & Roehrig, R. (2022). Investigating parametric dependence of climate feedbacks in the atmospheric component of cnrm-cm6-1. *Geophysical Research Letters*, *49*(9), e2021GL095084. https://doi.org/10.1029/2021GL095084

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

Pitman, A. J., Henderson-Sellers, A., Desborough, C. E., Yang, Z.-L., Abramopoulos, F., Boone, A., et al. (1999). Key results and implications from phase 1(c) of the project for intercomparison of Land-surface parametrization schemes. *Climate Dynamics*, *15*(9), 673–684. https://doi.org/10.1007/s003820050309

Qian, Y., Wan, H., Yang, B., Golaz, J.-C., Harrop, B., Hou, Z., et al. (2018). Parametric sensitivity and uncertainty quantification in the version 1 of e3sm atmosphere model based on short perturbed parameter ensemble simulations. *Journal of Geophysical Research: Atmospheres*, *123*(23), 13046–13073. https://doi.org/10.1029/2018JD028927

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

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., et al. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, *12*(4), 1393–1411. https://doi.org/10.5194/esd-12-1393-2021

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

Schickhoff, M., de Vrese, P., Bartsch, A., Widhalm, B., & Brovkin, V. (2024). Effects of land surface model resolution on fluxes and soil state in the arctic. *Environmental Research Letters*, *19*(10), 104032. https://doi.org/10.1088/1748-9326/ad6019

Schlosser, C. A., Slater, A. G., Robock, A., Pitman, A. J., Vinnikov, K. Y., Henderson-Sellers, A., et al., Contributors, T. P. D. (2000). Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(d). *Monthly Weather Review*, *128*(2), 301–321. https://doi.org/10.1175/1520-0493(2000)128⟨0301:SOABGH⟩2.0.CO;2

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., et al. (2016). Inconsistent strategies to spin up models in cmip5: Implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, *9*(5), 1827–1851. https://doi.org/10.5194/gmd-9-1827-2016

Sitch, S., O'Sullivan, M., Robertson, E., Friedlingstein, P., Albergel, C., Anthoni, P., et al. (2024). Trends and drivers of terrestrial sources and sinks of carbon dioxide: An overview of the trendy project. *Global Biogeochemical Cycles*, *38*(7), e2024GB008102. https://doi.org/10.1029/2024GB008102

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

Swenson, S. C., Burns, S. P., & Lawrence, D. M. (2019). The impact of biomass heat storage on the canopy energy balance and atmospheric stability in the community land model. *Journal of Advances in Modeling Earth Systems*, *11*(1), 83–98. https://doi.org/10.1029/2018MS001476

Tett, S. F. B., Gregory, J. M., Freychet, N., Cartis, C., Mineter, M. J., & Roberts, L. (2022). Does model calibration reduce uncertainty in climate projections? *Journal of Climate*, *35*(8), 2585–2602. https://doi.org/10.1175/JCLI-D-21-0434.1

Thornton, P. E., & Rosenbloom, N. A. (2005). Ecosystem model spin-up: Estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. *Ecological Modelling*, *189*(1), 25–48. https://doi.org/10.1016/j.ecolmodel.2005.04.008

Whittaker, R. H. (1970). *Communities and ecosystems*. Macmillan.

Wieder, W. R., Lawrence, D. M., Fisher, R. A., Bonan, G. B., Cheng, S. J., Goodale, C. L., et al. (2019). Beyond static benchmarking: Using experimental manipulations to evaluate land model assumptions. *Global Biogeochemical Cycles*, *33*(10), 1289–1309. https://doi.org/10.1029/2018GB006141

Williamson, D. B., Blaker, A. T., & Sinha, B. (2017). Tuning without over-tuning: Parametric uncertainty quantification for the NEMO ocean model. *Geoscientific Model Development*, *10*(4), 1789–1816. https://doi.org/10.5194/gmd-10-1789-2017

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*(7–8), 1703–1729. https://doi.org/10.1007/s00382-013-1896-4

Wood, E. F., Lettenmaier, D. P., Liang, X., Lohmann, D., Boone, A., Chang, S., et al. (1998). The project for intercomparison of Land-surface parameterization schemes (PILPS) phase 2(c) red–arkansas river basin experiment:: 1. Experiment description and summary intercomparisons. *Global and Planetary Change*, *19*(1), 115–135. https://doi.org/10.1016/S0921-8181(98)00044-7

Worden, J., Saatchi, S., Keller, M., Bloom, A. A., Liu, J., Parazoo, N., et al. (2021). Satellite observations of the tropical terrestrial carbon balance and interactions with the water cycle during the 21st century. *Reviews of Geophysics*, *59*(1), e2020RG000711. https://doi.org/10.1029/2020RG000711

Yamazaki, K., Sexton, D. M. H., Rostron, J. W., McSweeney, C. F., Murphy, J. M., & Harris, G. R. (2021). A perturbed parameter ensemble of hadgem3-gc3.05 coupled model projections: Part 2: Global performance and future changes. *Climate Dynamics*, *56*(11), 3437–3471. https://doi.org/10.1007/s00382-020-05608-5

Yan, H., Sun, N., Eldardiry, H., Thurber, T. B., Reed, P. M., Malek, K., et al. (2023b). Large ensemble diagnostic evaluation of hydrologic parameter uncertainty in the community land model version 5 (clm5). *Journal of Advances in Modeling Earth Systems*, *15*(5), e2022MS003312. https://doi.org/10.1029/2022MS003312

Yan, H., Sun, N., Eldardiry, H., Thurber, T. B., Reed, P. M., Malek, K., et al. (2023a). Characterizing uncertainty in community land model version 5 hydrological applications in the United States. *Scientific Data*, *10*(1), 187. https://doi.org/10.1038/s41597-023-02049-7

Zarakas, C. M., Kennedy, D., Dagon, K., Swann, A. L. S., Liu, A., Bonan, G., et al. (2024). Land processes can substantially impact the mean climate state. *Geophysical Research Letters*, *51*(21), e2024GL108372. https://doi.org/10.1029/2024GL108372

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. *Global Biogeochemical Cycles*, *26*(3), GB3025. https://doi.org/10.1029/2011GB004185