The Community Land Model, version 5.1 One-at-a-time Parameter Perturbation Ensemble

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], and A.W. Wood[1]

[1]Climate and Global Dynamics Laboratory, 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, Guangdong, China
[7]Department of Atmospheric Sciences, 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, AL, USA
[11]International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[12]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA
[13]School of Integrative Plant Science, Cornell University, 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

Key Points:

- enter point 1 here

Corresponding author: Daniel Kennedy, djk2120@ucar.edu

Abstract

- 211 parameters perturbed across six forcing scenarios
- expansive dataset available 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
- parameter effects can exceed scenario effects
- small number of parameters explains a large fraction of variance
- most important parameters can vary regionally and also based on climate forcing
- software infrastructure developed to facilitate routine investigation of parameter sensitivity and uncertainty, as well as automated calibration

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 and influence CO_2 concentrations is a critical facet of climate change research. Uncertainty in climate model projections varies by domain and generally increases with extended time horizons (Koven et al., 2022). Land processes substantially influence climate directly through, for example, evapotranspiration (Zarakas et al., 2024), and indirectly through carbon-climate feedbacks (?, ?). Making centennial-scale projections of the cumulative terrestrial carbon sink gas been especially challenging, with high uncertainty persisting across model generations (Friedlingstein et al., 2014; Arora et al., 2020), and limited efficacy of emergent constraints. A portion of this uncertainty is irreducible, 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. The capability to efficiently ingest these data and improve simulation performance 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 (Henderson-Sellers et al., 1995; Pitman et al., 1999; Wood et al., 1998; Schlosser et al., 2000; Eyring et al., 2016; Friedlingstein et al., 2022). 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, e.g. 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 Param-

eter Ensemble (PPE, also referred to as Perturbed Physics Ensemble), with examples in the coupled (Murphy et al., 2004) and land-only contexts (Dagon et al., 2020; McNeall et al., 2024). One important application of PPEs has been assessing uncertainty in climate model projections (Murphy et al., 2004; Sanderson et al., 2008; Booth et al., 2012; Hawkins et al., 2019; Yamazaki et al., 2021; Peatier et al., 2022; Tett et al., 2022). Beyond uncertainty quantification, PPEs also form a basis for automated model calibration through, for example, history matching (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). Several promising avenues for model calibration are in development (Pinnington et al., 2020; Cleary et al., 2021; Alonso-González et al., 2022), 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 impact the terrestrial biosphere (e.g, hydrology, snow and ice sheets, 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 objective calibration, model interpretation, and development (Fisher & Koven, 2020; Dagon et al., 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 (Hourdin et al., 2017; Balaji et al., 2022). 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 Terrestrial Systems Model (CTSM) (Lawrence et al., 2019), in the Community Land Model with biogeochemistry configuration (CLM-BGC). 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 CTSM modeling framework, and develop the necessary tools and datasets 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 dataset, describe how it was produced, and survey some potential applications. The dataset 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 uncertainties within CTSM/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 CTSM codebase for future use (see Open Research Section). We utilized the biogeochemical cycling 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 reanalysis atmosphere 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 2000 years (or more) to reach steady-state conditions (Lawrence et al., 2019). In the absence of equilibrium, the drift towards 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 member.

To manage computational cost we leveraged the Semi-Analytic spin-up mode (SASU) recently implemented within CLM5 (Lu et al., 2020; Liao et al., 2023). This new module utilizes a matrix representation on CLM's biogeochemistry to reduce spin-up time by one order of magnitude (Luo et al., 2022; Liao et al., 2023), and had been previously leveraged for a PPE with the Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) 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 equilibration of deep soil carbon (beyond 1 meter depth) in very few points at high latitude. Certain inferences about deep soil carbon from this ensemble may therefore subject to uncertainty due to spin-up concerns.

2.3 Sparse grid

Another control on model cost is spatial resolution. Most global CTSM simulations utilize nominal 1° resolution, which equates to about 20,000 land grid cells. In order to manage computational cost, parameter perturbation experiments often use lower resolution, such as 4°x5° (Dagon et al., 2020). Here, we use a clustering algorithm to achieve an alternative low 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 here 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 CTSM-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 'sparsegrid', 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.

Clustering was based on a subset of 18 CTSM 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 (Supp Figure 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.

Table 1. Clustering inputs categorized into three groups, with the CTSM variable name in parentheses

Climate variables	Ecosystem state variables	Ecosystem flux variables
2m 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)
2m 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)

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 (Mendoza et al., 2015; Cuntz et al., 2016). For the purposes of this activity, we identified a broad set of 211 CLM-BGC parameters, and enabled their modification via the model input files. Several domains were not perturbed, including crops, CH_4, VOC, and dust emissions, 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 plant functional type. Not all parameters are currently utilizing this functionality, such as *medlynintercept* where the default value is 100 μmol H_2O m^{-2} s^{-1} for all sixteen of CLM's natural vegetation PFTs.

Given the large parameter space, 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 cases, expert judgment was used or else a default perturbation of ±20%. The parameter range spreadsheet, with literature references and parameter descriptions

is available online and in appendix zqz. 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. We likewise opted to perturb soil hydraulic parameters directly (e.g. saturated hydraulic conductivity) and also indirectly, in a parallel set of simulations via perturbations to the soil texture (percent sand, clay, and silt).

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 (Supp Figure S2). 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 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 a supplemental materials spreadsheet.

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 dataset 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 CO_2.

We opted for six different forcing scenarios to understand the intersection of parameter effects with different forcing agents associated with climate change (Table 2). 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). We applied climate and CO_2 anomalies independently, in order to disentangle their effects on parameter rankings. Future and pre-industrial climate forcing datasets were prepared by adding GSWP3v1 anomalies from 2005-2014 to the corresponding mean climate change signal (see Table 2; AF2095 and AF1855). 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 and present day for six atmospheric forcing variables (temperature, humidity, precipitation, wind, longwave and shortwave radiation; see Supp Figure S3 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 anomalies were not applied to certain coastal gridcells (Supp Fig S4). In all this affected less than 3% of land area, and did not significantly alter parameter rankings (Supp Fig S5).

Table 2. 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

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, which delineates biomes based on average temperature and precipitation (Whittaker, 1970). Biomes were delineated based on the atmospheric forcing temperature and precipitation data averaged across 2005 to 2014 (Supp Figure S6). Overall there are nine vegetated biomes and an additional ice sheet biome, where no vegetation is present (Supp Figures S7,8).

3 Results

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 sparsegrid 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 22648, 5666, and 1764 land grid cells in standard 1°, 2°, and 4°x5° CLM5 simulations, respectively, as compared to just 400 grid cells in the sparsegrid. Likewise, whereas previous spin-up methodologies required 1500 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.

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 continues to improve with added clusters, but with marginal returns (Supp Figure S2). 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, convergence slows and presents some small amplitude oscillations. We chose 400 clusters, because it provides satisfactory performance across a range of important metrics, while affording us sufficient simulations to perform the full experiment within our computational budget.

This model configuration allowed for over 2000 simulations across six forcing scenarios. Overall, we found substantial impacts of parametric uncertainty on model behavior, in some cases exceeding the magnitude of climate scenario effects (Figure 4). In the control scenario, gross primary productivity (GPP) ranged from 103 to 158 PgC, and

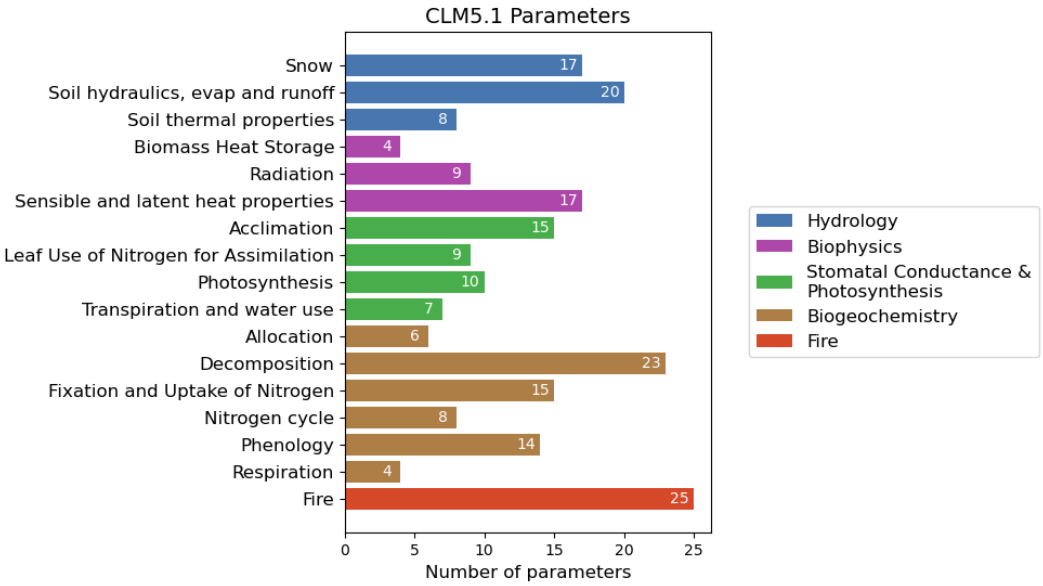

Figure 1. 211 parameters were perturbed across the various domains of the land model.

net ecosystem productivity (NEP) ranged from 1.7 to 3.8 PgC. NEP was especially variable in the high CO_2 scenario, varying from 1.6 to 5.7 PgC. 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.

A small fraction of parameters tends to explain a large amount of the ensemble variance for any given variable or metric (Figure 5). For example, 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 (Supp Fig S3). Conversely, GPP, which is a significantly more complicated process, requires 20 parameters to explain 90% of the control ensemble variance. 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.

We repeated the full set of parameter perturbations across the six forcing scenarios in Table 2. 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 (CO2, 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 (Figure 6a,b). 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

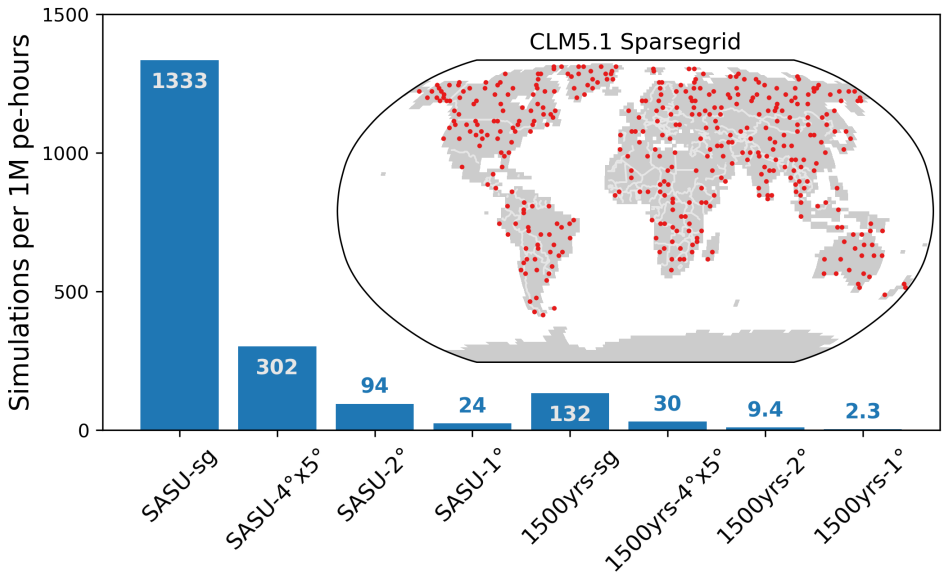

Figure 2. The approximate number of simulations afforded by 1 million core-hours on the Cheyenne supercomputer for a range of CLM configurations. Configurations are labeled according to spin-up procedure (SASU or the standard 1500-year spinup) and horizontal resolution ('sg' signifies sparsegrid). The inset map shows the locations of the 400 sparse grid cells. See Section 2 for spin-up and sparsegrid details.

phosphate limitation is difficult to quantify and has not been observed to significantly hinder photosynthesis under present-day CO_2 concentrations (Kumarathunge et al., 2019). In CLM5.1, this parameter is not particularly influential on NEP at 367 ppm (ranked 39, not shown), but is third most influential with high CO_2.

The future climate experiment is 4.3K warmer averaged over our study domain (land-only, Antarctica excluded, Supp Figure S4). This led to an increase in evapotranspiration 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 (Figure 6d,e), 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*). 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 leaf area index, 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 leaf area index globally included parameters that were important in each of these three biomes.

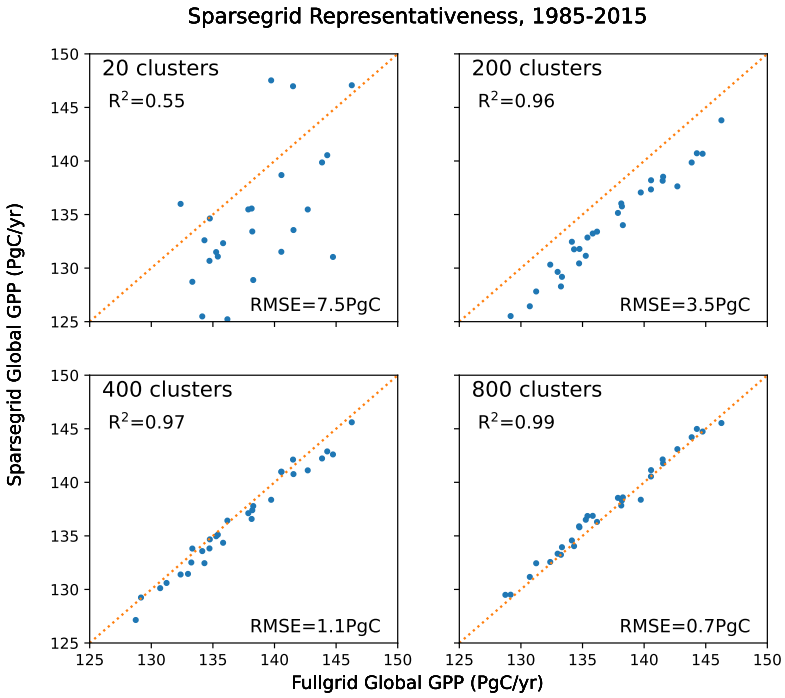

Figure 3. Sparsegrid vs fullgrid (2° resolution) global annual gross primary production (GPP) across the last forty years of a transient CLM5.1 simulation. We opted for 400 clusters to balance computational cost against representativeness.

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 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 summarises the >2TB of output data via approximately 2000 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 CLM5-BGC parameters. First, many parameters had not been officially 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. Although in this experiment 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

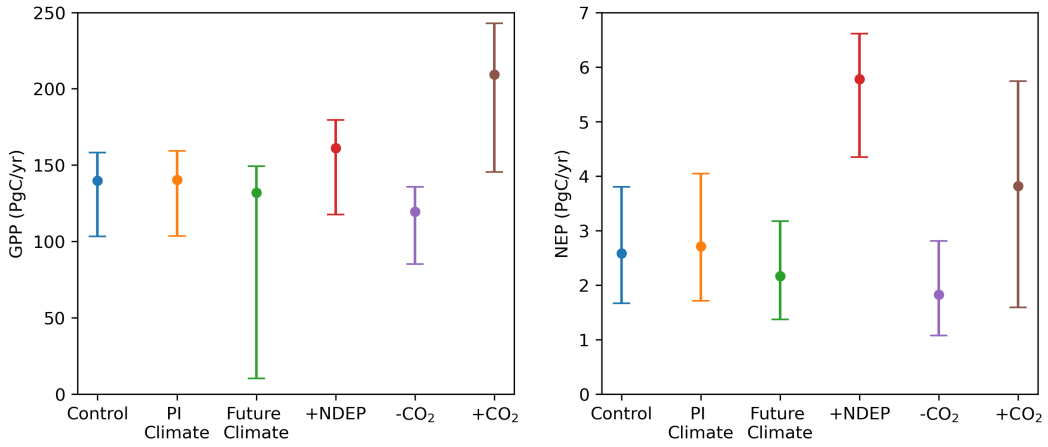

Figure 4. Global annual gross primary production (GPP) and net ecosystem production (NEP) across six forcing scenarios. Circles mark the CLM5.1 default simulation, and the bars span the ensemble ranges generated by individual parameter perturbations.

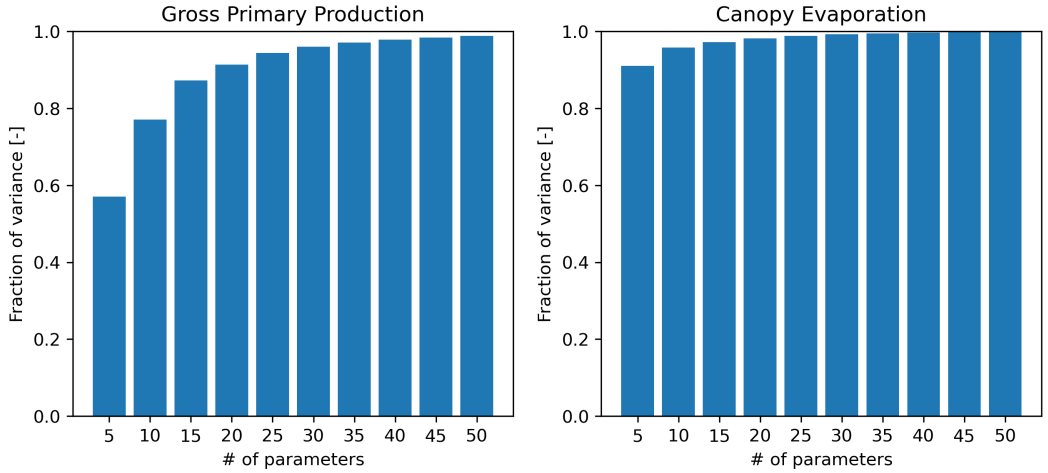

Figure 5. Cumulative fraction of variance explained by the most influential parameters on GPP and canopy evaporation. Approximately 96% of canopy evaporation variance can be explained by the ten most influential parameters, whereas 30 parameters are required to explain 96% of GPP variance.

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. For example, 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

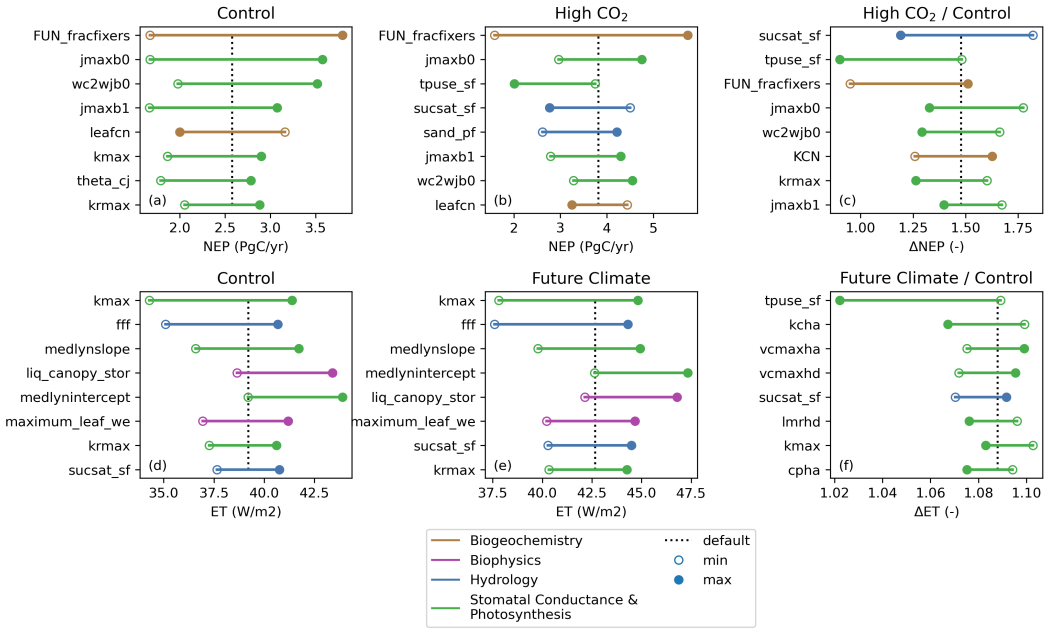

Figure 6. The eight most influential parameters on NEP in the control simulations (a), high CO_2 simulations (b), and the relative response to high CO_2 (c), as well as on ET in the control simulations (d), future climate simulations (e) and the relative response to future climate (f). The parameters 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).

sample comparable probability density, even in a univariate experiment, like this one. The number of parameters is large, with many 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, they may behave differently at the coarse climate modeling spatial resolution (zqz), complicating the process of uncertainty quantification. 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 (Lu et al., 2020; Luo et al., 2022; Liao et al., 2023) (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.

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

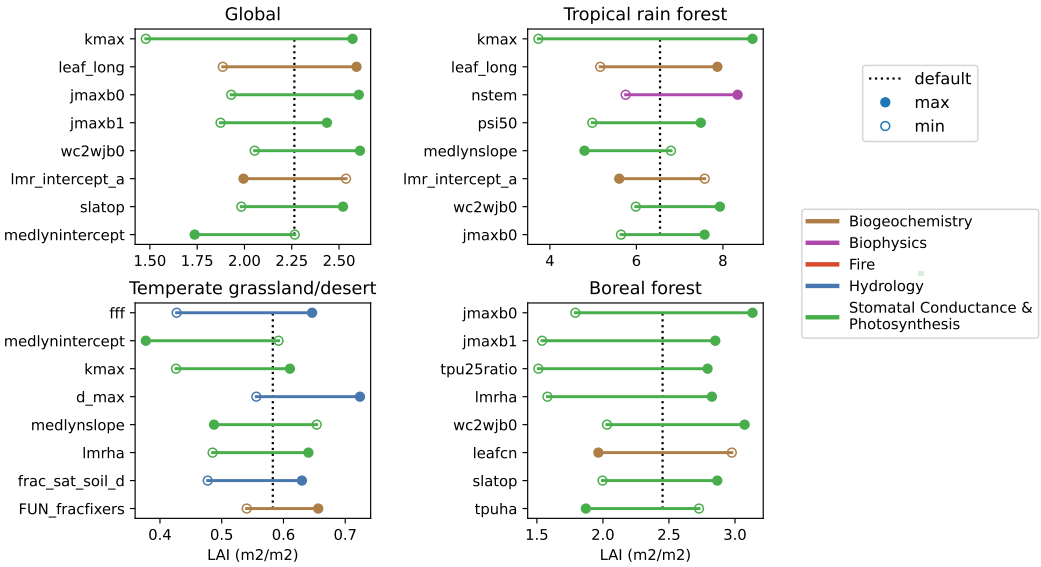

Figure 7. The eight most influential parameters on leaf area index within the control ensemble, globally and within three biomes. The solid lines span the range between the two simulations for each parameter, with an open circle indicating the simulation where a parameter is set to its minimum value and a filled circle for its maximum. The lines are color-coded according to scientific domain. The dashed lines indicate the LAI for the simulation with default CLM5.1 parameters. Parameter rankings vary by biome, with the global rankings seeming to reflect contributions from each of these three biomes.

be quite large, in some cases exceeding the effects of our various 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. In the case of canopy evaporation, for example, more than 95% of the ensemble variance could be explained by the ten 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 evapotranspiration (ET), were among the most influential on the response of ET to future climate. This emphasizes the importance of testing models not only under mean-state conditions, but also under experimentally perturbed conditions under global change analogues (Wieder et al., 2019). 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 (Supp Figure S2), 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 1850-2100, as the computational cost is comparable given the overhead of independent spinups for each forcing scenario.

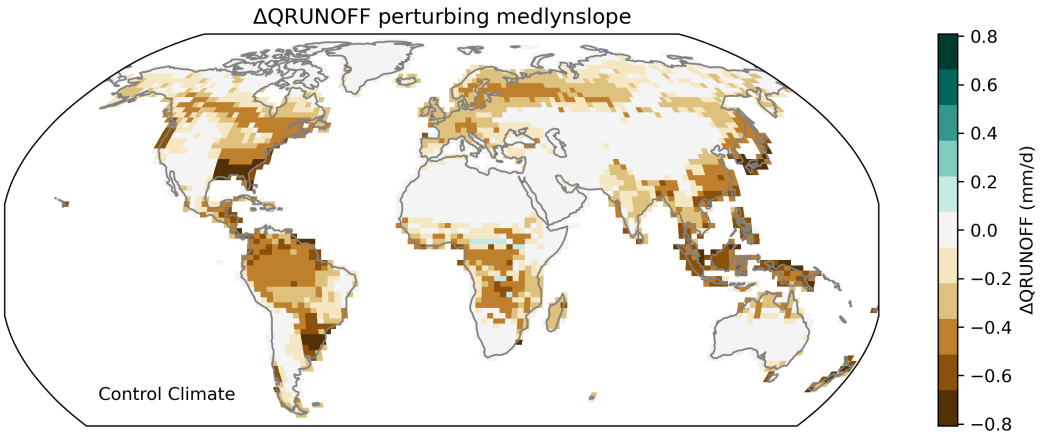

Figure 8. Map of the effect of perturbing medlynslope on runoff within the control ensemble. Increasing medlynslope tends to reduce runoff, but only in regions with sufficient vegetation activity. This is one example of many plots available online (`https://webext.cgd.ucar.edu/I2000/PPEn11_OAAT`) in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

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 inflict 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 stressor for the set of PFTs in a given biome will have greater effect. 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 dataset will be insufficient for most calibration activities or for estimating overall parametric uncertainty. A primary utility of our dataset 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, 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 CO2 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). 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. 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.

The primary deliverable of this project is the PPE dataset that captures the effects of 212 parameters on a wide range of CLM variables across six forcing scenarios. Several spin-off projects have leveraged this dataset, 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.

The infrastructure underpinning this experiment is another 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, datasets, and analysis tools generated to enable and utilize this experiment will greatly reduce the burden of generating future PPEs and parameter sensitivity experiments. 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 begun a follow-on activity that perturbs a subset of important parameters with a more exhaustive sampling approach to work towards calibration of leaf area index. 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 towards model calibration and uncertainty quantification, and we anticipate repeating the 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.

Open Research Section

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 (∼300MB), post-processed dataset is freely available online. All of the figures on our diagnostics website can be made from this one file. The raw dataset will be available on the NSF-NCAR computing system, and can be made available on-

line as well, but we are gauging interest before committing to finding a host for the 2TB of data.

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 National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX) were provided by the Climate Simulation Laboratory at NCAR's Computational and Information Systems Laboratory (CISL). DK is supported via the Drought Task Force IV organized by the NOAA Climate Program Office's Modeling, Analysis, Predictions, and Projections. 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 No. 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., ... Gascoin, S. (2022, December). The Multiple Snow Data Assimilation System (MuSA v1.0). *Geoscientific Model Development*, *15*(24), 9127–9155. Retrieved 2023-08-30, from `https://gmd.copernicus.org/articles/15/9127/2022/` doi: 10.5194/gmd-15-9127-2022

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

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. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.2202075119` doi: 10.1073/pnas.2202075119

Birch, L., Schwalm, C. R., Natali, S., Lombardozzi, D., Keppel-Aleks, G., Watts, J., ... Rogers, B. M. (2021). Addressing biases in Arctic–boreal carbon cycling in the Community Land Model Version 5. *Geoscientific Model Development*, *14*(6), 3361–3382. Retrieved 2023-08-31, from `https://gmd.copernicus.org/articles/14/3361/2021/` doi: 10.5194/gmd-14-3361-2021

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

Cheng, Y., Musselman, K. N., Swenson, S., Lawrence, D., Hamman, J., Dagon, K., ... Newman, A. J. (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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022WR032204 doi: 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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0021999120304903 doi: 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., ... 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001354 doi: https://doi.org/10.1029/2018MS001354

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., ... 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*, e2020MS002217. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002217 doi: 10.1029/2020MS002217

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., ... Thober, S. (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), 10,676-10,700. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025097 doi: 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. Retrieved from https://ascmo.copernicus.org/articles/6/223/2020/ doi: 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, May). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved 2023-08-30, from https://gmd.copernicus.org/articles/9/1937/2016/ doi: 10.5194/gmd-9-1937-2016

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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001453 doi: https://doi.org/10.1029/2018MS001453

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014, Jan 15). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*(2), 511-526. Retrieved from http://ezproxy.cul.columbia.edu/login?url=https://search.proquest.com/docview/1495899892?accountid=10226

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., ... Zheng, B. (2022, November). Global Carbon Budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. Retrieved 2023-08-29, from https://essd.copernicus.org/articles/14/4811/2022/ doi: 10.5194/essd-14-4811-2022

Hawkins, L. R., Rupp, D. E., McNeall, D. J., Li, S., Betts, R. A., Mote, P. W., ... Wallom, D. C. H. (2019). Parametric sensitivity of vegetation dy-

namics 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001577 doi: https://doi.org/10.1029/2018MS001577

Henderson-Sellers, A., Pitman, A. J., Love, P. K., Irannejad, P., & Chen, T. H. (1995, April). The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3*. *Bulletin of the American Meteorological Society*, *76*(4), 489–504. Retrieved 2023-08-29, from https://journals.ametsoc.org/view/journals/bams/76/4/1520-0477_1995_076_0489_tpfiol_2_0_co_2.xml doi: 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. Retrieved from https://journals.ametsoc.org/view/journals/eint/9/10/ei110.1.xml doi: 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. Retrieved from https://doi.org/10.1007/s10980-013-9902-0 doi: 10.1007/s10980-013-9902-0

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... Williamson, D. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589 - 602. Retrieved from https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-15-00135.1.xml doi: https://doi.org/10.1175/BAMS-D-15-00135.1

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., ... Ghattas, J. (2020). LMDZ6A: The Atmospheric Component of the IPSL Climate Model With Improved and Better Tuned Physics. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001892. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001892 doi: 10.1029/2019MS001892

Huang, Y., Zhu, D., Ciais, P., Guenet, B., Huang, Y., Goll, D. S., ... 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017MS001237 doi: https://doi.org/10.1029/2017MS001237

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., ... Wirth, C. (2020). Try plant trait database – enhanced coverage and open access. *Global Change Biology*, *26*(1), 119-188. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.14904 doi: https://doi.org/10.1111/gcb.14904

Koven, C. D., Arora, V. K., Cadule, P., Fisher, R. A., Jones, C. D., Lawrence, D. M., ... 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. Retrieved from https://esd.copernicus.org/articles/13/885/2022/ doi: 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 & Environment*, *42*(12), 3241-3252. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/pce.13639 doi: https://doi.org/10.1111/pce.13639

Laguë, M. M., Bonan, G. B., & Swann, A. L. S. (2019). Separating the im-

pact 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. Retrieved from https://journals.ametsoc.org/view/journals/clim/32/18/jcli-d-18-0812.1.xml doi: 10.1175/JCLI-D-18-0812.1

Lawrence, D. M., et al., & 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*, *in press*. doi: 10.1029/2018MS001583

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

Lin, Y.-S., Medlyn, B. E., Duursma, R. A., Prentice, I. C., Wang, H., Baig, S., ... Wingate, L. (2015, May). Optimal stomatal behaviour around the world. *Nature Climate Change*, *5*(5), 459–464. Retrieved 2023-08-02, from https://www.nature.com/articles/nclimate2550 doi: 10.1038/nclimate2550

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

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

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., ... Wang, Y.-P. (2022). Matrix approach to land carbon cycle modeling. *Journal of Advances in Modeling Earth Systems*, *14*(7), e2022MS003008. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003008 doi: 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. Retrieved from https://gmd.copernicus.org/articles/17/1059/2024/ doi: 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. Retrieved from https://esd.copernicus.org/articles/7/917/2016/ doi: 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015820 doi: https://doi.org/10.1002/2014WR015820

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*, 768–772. Retrieved from https://www.nature.com/articles/nature02771 doi: 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL095084 (e2021GL095084 2021GL095084) doi: https://doi.org/10.1029/

2021GL095084

Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., & Scoby, D. (2020, January). The Land Variational Ensemble Data Assimilation Framework: LAVENDAR v1.0.0. *Geoscientific Model Development*, *13*(1), 55–69. Retrieved 2023-08-30, from https://gmd.copernicus.org/articles/13/55/2020/ doi: 10.5194/gmd-13-55-2020

Pitman, A. J., Henderson-Sellers, A., Desborough, C. E., Yang, Z.-L., Abramopoulos, F., Boone, A., ... Xue, Y. (1999, September). Key results and implications from phase 1(c) of the Project for Intercomparison of Land-surface Parametrization Schemes. *Climate Dynamics*, *15*(9), 673–684. Retrieved 2023-08-29, from https://doi.org/10.1007/s003820050309 doi: 10.1007/s003820050309

Qian, Y., Wan, H., Yang, B., Golaz, J.-C., Harrop, B., Hou, Z., ... Zhang, K. (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), 13,046-13,073. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD028927 doi: https://doi.org/10.1029/2018JD028927

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

Sanderson, B. M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D. J., ... 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. Retrieved from https://journals.ametsoc.org/view/journals/clim/21/11/2008jcli1869.1.xml doi: https://doi.org/10.1175/2008JCLI1869.1

Schlosser, C. A., Slater, A. G., Robock, A., Pitman, A. J., Vinnikov, K. Y., Henderson-Sellers, A., ... Contributors, T. P. D. (2000, February). Simulations of a Boreal Grassland Hydrology at Valdai, Russia: PILPS Phase 2(d). *Monthly Weather Review*, *128*(2), 301–321. Retrieved 2023-08-30, from https://journals.ametsoc.org/view/journals/mwre/128/2/1520-0493_2000_128_0301_soabgh_2.0.co_2.xml doi: 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., ... Romanou, A. (2016). Inconsistent strategies to spin up models in cmip5: implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, *9*(5), 1827–1851. Retrieved from https://gmd.copernicus.org/articles/9/1827/2016/ doi: 10.5194/gmd-9-1827-2016

Sitch, S., O'Sullivan, M., Robertson, E., Friedlingstein, P., Albergel, C., Anthoni, P., ... 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GB008102 (e2024GB008102 2024GB008102) doi: 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*, *n/a*(n/a). doi: 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. doi: 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. Retrieved from https://journals.ametsoc.org/view/journals/clim/35/8/JCLI-D-21-0434.1.xml doi: 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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304380005001948 doi: 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., ... Thomas, R. Q. (2019). Beyond static benchmarking: Using experimental manipulations to evaluate land model assumptions. *Global Biogeochemical Cycles*, *33*(10), 1289-1309. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GB006141 doi: https://doi.org/10.1029/2018GB006141

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*, 1703–1729.

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*, 1789–1816. Retrieved from https://gmd.copernicus.org/articles/10/1789/2017/ doi: 10.5194/gmd-10-1789-2017

Wood, E. F., Lettenmaier, D. P., Liang, X., Lohmann, D., Boone, A., Chang, S., ... Zeng, Q.-c. (1998, December). 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. Retrieved 2023-08-30, from https://www.sciencedirect.com/science/article/pii/S0921818198000447 doi: 10.1016/S0921-8181(98)00044-7

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. Retrieved from https://doi.org/10.1007/s00382-020-05608-5 doi: 10.1007/s00382-020-05608-5

Yan, H., Sun, N., Eldardiry, H., Thurber, T. B., Reed, P. M., Malek, K., ... Rice, J. S. (2023a). Characterizing uncertainty in Community Land Model version 5 hydrological applications in the United States. *Scientific Data*, *10*(1), 187. Retrieved 2023-08-02, from https://www.nature.com/articles/s41597-023-02049-7 doi: 10.1038/s41597-023-02049-7

Yan, H., Sun, N., Eldardiry, H., Thurber, T. B., Reed, P. M., Malek, K., ... Rice, J. S. (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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003312 doi: https://doi.org/10.1029/2022MS003312

Zarakas, C. M., Kennedy, D., Dagon, K., & Swann, A. L. S. (2024). Land processes can substantially impact the mean climate state. *Geophysical Research Letters*, *tk*(tk). doi: tk

Appendix A Supplementary Figures

extra text: Maintaining, improving, and interpreting complex land models benefits from thoughtful investment in software to automate and routinize important components of the development process, e.g. Collier et al. (2018).

Will be deleted:

A1 Parameters

Table A1. Some key parameters

Parameter	Description	Model Domain
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
medlynintercept	Medlyn intercept of conductance-photosynthesis relationship	Stomatal resistance and photosynthesis
medlynslope	Medlyn slope of conductance-photosynthesis relationship	Stomatal resistance and photosynthesis
tpu25ratio	Ratio of tpu25top to vcmax25top	Stomatal resistance and photosynthesis
jmaxb0	Baseline proportion of nitrogen allocated for electron transport	Photosynthetic capacity (LUNA)
jmaxb1	Response of electron transport rate to light	Photosynthetic capacity (LUNA)
slatop	Specific leaf area at top of canopy	Photosynthetic capacity (LUNA)
wc2wjb0	Baseline ratio of wc:wj	Photosynthetic capacity (LUNA)
kmax	Plant segment max conductance	Plant hydraulics
krmax	Root segment max conductance	Plant hydraulics
psi50	Water potential at 50% loss of conductance	Plant hydraulics
nstem	Stem number	Biomass heat storage
lmr_intercept_atkin	Intercept in the calculation of leaf maintenance respiration	Plant respiration
froot_leaf	Allocation parameter: new fine root C per new leaf C	Carbon and nitrogen allocation
leafcn	Leaf C:N	Carbon and nitrogen allocation
leaf_long	Leaf longevity	Vegetation phenology and turnover
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

manuscript submitted to *Journal of Advances in Modeling Earth Systems (JAMES)*