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

D. Kennedy[1]**and ...**

[1]Climate and Global Dynamics Laboratory, NCAR, Boulder, CO, USA.

Key Points:

- enter point 1 here

Corresponding author: Daniel Kennedy, djk2120@ucar.edu

Abstract

[enter your Abstract here]

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 land components of Earth System Models. Understanding how these processes respond to CO_2 concentrations, and how they themselves impact CO_2 concentrations, is a critical facet of climate change research. Our certainty in climate model projections varies by domain and generally decreases with extended time horizons (Koven et al., 2022). Centennial-scale estimates of the cumulative terrestrial carbon sink have been especially challenging, with high uncertainty persisting across model generations (Friedlingstein et al., 2014; Arora et al., 2020). A portion of this uncertainty is irreducible, inherent to the challenge of predicting vegetation dynamics in a novel climate (Lovenduski & Bonan, 2017). Effectively synthesizing the ongoing expansion of observational data sources from remote sensing, meteorological stations, flux towers, and field campaigns can help to parameterize increasingly comprehensive land modeling platforms.

Uncertainty in the model projections of the carbon cycle is manifest in both uncertainty around the appropriate algorithms and model structures used, and also in the parameterizations of these models. Model inter-comparison projects (MIPs) are the primary means by which structural uncertainty is assessed (refs). While these 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). To allow for more tractable experimental design each model is typically allowed only a single parameterization of many plausible parameter sets. Thus, MIPs typically conflate parametric and structural uncertainty, and de-emphasize the consequences of uncertain model calibration in estimates of future land state trajectories.

The process of parameter selection for land 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, cryology, biogeochemistry, physiology, land management, micrometeorology, etc.). The number of model parameters required to represent all the relevant processes is large, and model speeds are relatively slow, presenting a major barrier to robust objective calibration, model interpretation and development (Fisher & Koven, 2020; Dagon et al., 2020). Thorough assessment of the parameter 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.

Several promising avenues for model calibration are in development (lavendar, musa, CES), many of which require the construction of large parameter perturbation ensembles (PPEs) (Qian et al., 2018). PPEs have played an important role in assessing projection uncertainty in climate models (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). 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) or Bayesian uncertainty quantification (Cleary et al., 2021). A major goal of this project is to enable automated calibration of the Community Land Model (CLM) (Lawrence et al., 2019). This extends 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.

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). Our goal in this project is to systematize the parameter perturbation process within the CLM modeling framework, and develop the necessary tools and datasets to efficiently test parameter effects across the full suite 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), while the software and modeling infrastructure has greatly expanded our capability to generate insights about parameter effects and uncertainties within CLM.

2 Experiment Description

2.1 Model description

The is developed by the CESM Land Model Working Group and maintained at the National Center for Atmospheric Research (NCAR). This experiment utilizes the Community Land Model configuration version 5.1 (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, 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'). 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 was explored in Dagon et al. (2020). We used the model in land-only mode driven by observed meteorological forcing from the GSWP3v1 reanalysis product (`http://hydro.iis.u-tokyo.ac.jp/GSWP3/`), 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. 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 Matrix-CN spin-up mode recently implemented within CLM (Lu et al., 2020). This new module utilizes a linearized simplification of CLM's biogeochemistry to significantly reduce spin-up time. Our spin-up protocol featured 20 years in accelerated decomposition mode (see Lawrence et al. (2019) for details), followed by 80 years of Matrix-CN, followed by 40 years of 'normal' mode, cycling over a ten-year forcing dataset (described below). This protocol was designed to achieve sufficiently equilibrated model states, while minimizing computational time. This spin-up methodology did not always reach full equilibration of deep soil carbon (beyond 1 meter depth). Certain inferences about deep soil carbon from this ensemble are therefore subject to uncertainty due to spin-up concerns.

2.3 Sparsegrid

Another control on model cost is resolution. Most CLM 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.

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, leaving only 400 grid cells that efficiently sample important model dynamics across the globe.

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 '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 meaningful CLM variables (Table 1). The clustering algorithm analyzed 12 observations of each variable per grid cell, namely the mean and interannual variability 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 A1). 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 CLM variable name in parentheses

Climate forcing 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 Experimental Design

Identification of the complete parameter set of a land surface model is in itself a non-trivial exercise, as in practice, many empirical constants previously considered as 'internal' are hard-coded and not amenable to systematic perturbation. For the purposes of this activity, we identified what we consider a very broad set of 193 CLM-BGC pa-

rameters, and enabled their modification via the model input files. We opted not to perturb crop parameters within this ensemble, as our initial focus is on unmanaged land.

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 the many cases, expert judgment was used. The 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 via perturbations to the soil texture (percent sand, clay, and silt).

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. 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, and 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). 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 the six atmospheric forcing variables. Our future climate scenario utilizes the SSP3-7.0 forcing in order to allow comparison with the existing CESM2 large ensemble. Our software infrastructure has greatly reduced the effort and computational cost associated with ensemble generation, such that it would be easy to add additional ensembles, to explore other emission scenarios or climate regimes (e.g. Last Glacial Maximum).

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	$5g/m^2$	enhanced nitrogen deposition

2.5 Analyses

Biome analysis

3 Results

A total of 193 parameters were targeted for sensitivity testing across a wide variety of CLM subcomponents (Figure 1).

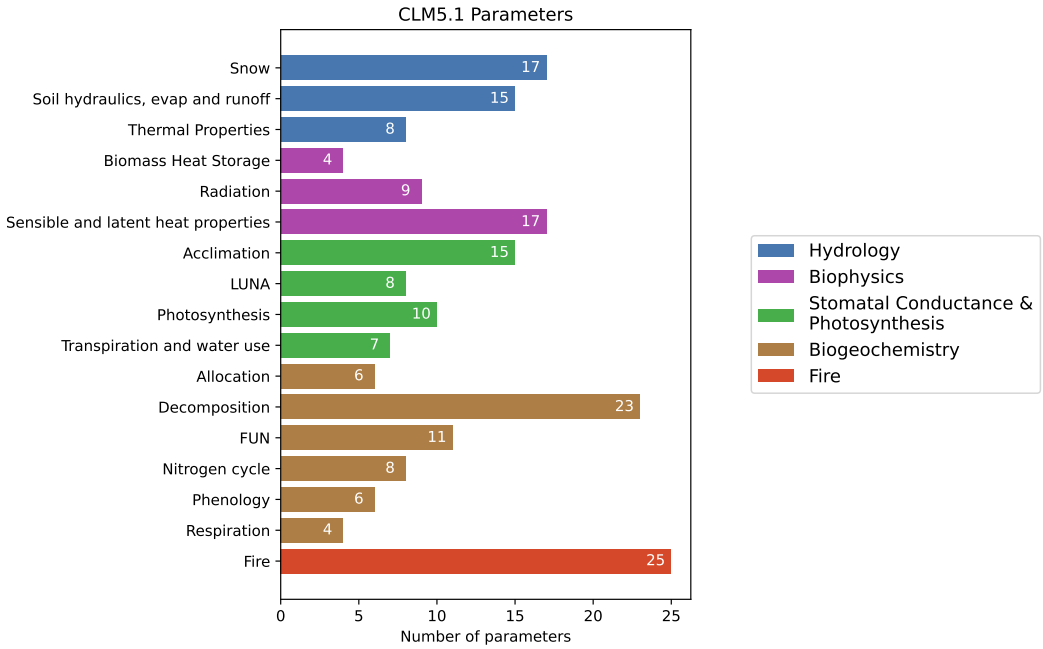

Figure 1. 193 parameters were perturbed across the various domains of the land model.

Given the challenging size of the CLM5-BGC parameter space, and the need to probe and document parametric uncertainties within operational land surface models, a major goal of this project was to develop a fast configuration of CLM5-BGC that would allow for a large number of simulations given the available computational resources. Combining the CN-Matrix spinup approach with our sparsegrid formulation (see Section 2.3 for details) yielded a configuration approximately 500 times cheaper 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° CLM 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 CN-matrix approach yielded satisfactory spin-up for our 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 provides a better approximation of output from the full grid (Figure 3). 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, and RMSE<1PgC requiring approximately 700 clusters (Supp Figure zqz). We chose 400 clusters, based on ILAMB scores of a wide variety of output variables (Supp Figure A1)

Overall, we found substantial impacts of parametric uncertainty on model behaviour, in some cases exceeding the magnitude of climate scenario effects (Figure 4). 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% likely exceeds 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, which may belie unforeseen tipping points, brittle parameterizations, and/or bugs.

A small fraction of parameters tend to explain a large amount of the ensemble variance (Figure 5) for any given variable or metric. For example the same four parameters

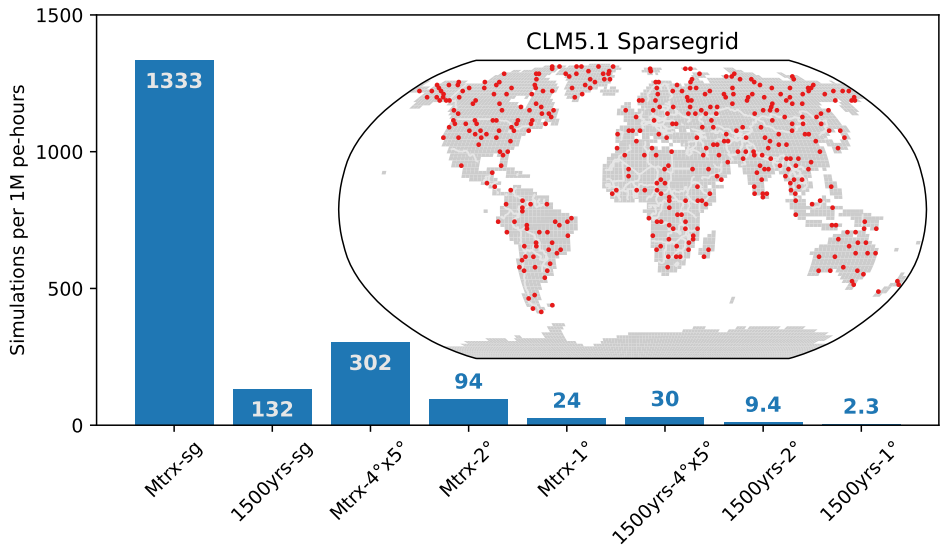

Figure 2. The approximate number of simulations afforded by 1 million core-hours on the Cheyenne supercomputer for a range of CLM configurations. Configurations are labeled according to spin-up procedure (CN-matrix or the standard 1500-year spinup) and horizontal resolution ('sg' signifies sparsegrid).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

explain up to 90% of the canopy evaporation variance across all the forcing scenarios (Supp Fig A2). Most of the parameter perturbations had a relatively small impact on any given output variable. As such, the effective parameter space dimensionality for any given CLM process is likely much smaller than would be expected from having 193 independent parameters, reflecting the fact that land surface models like CLM comprise representations of numerous interconnected domains. 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 responses to individual forcing variables (CO2, nitrogen deposition) and climate. In the case of evapotranspiration (ET), for example, the parameters that control the increase in ET due to warming tend to be different from the parameters that control present-day ET (Figure 6). ET response can differ significantly from the default case (i.e. stray from the orange line), even when ET is near the default under present-day conditions (i.e. within the shaded region). Besides *kmax*, all of the parameters that appear in the top eight parameters for ET response to future climate govern plant acclimation to temperature. These parameters, while not especially influential on present-day ET, would be among the most influential in governing the trend in ET going forward, illustrating the potential limitations of calibration to ambient conditions.

The PPE is likewise useful for determining which parameters influence which model processes and where. For example the parameters controlling leaf area index 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 param-

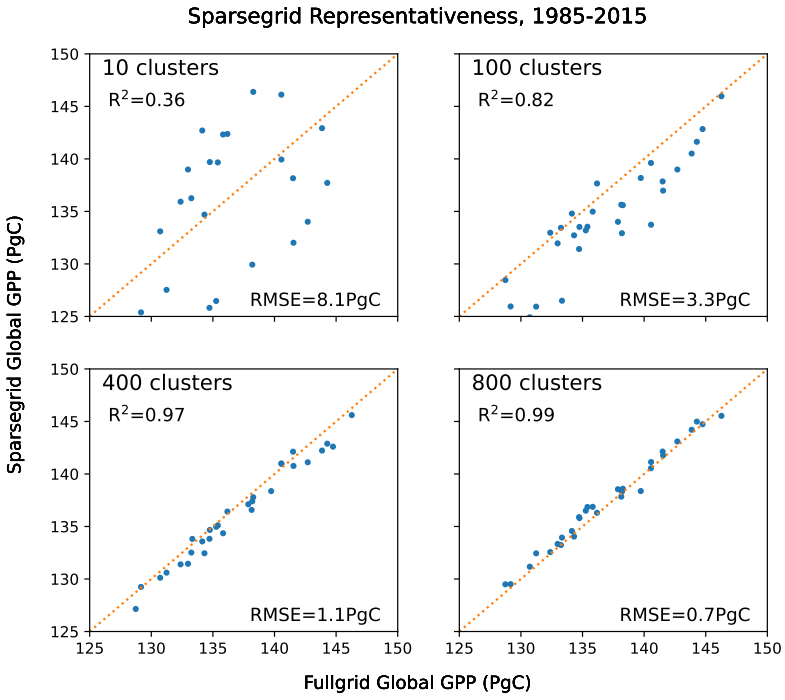

Figure 3. Sparsegrid vs fullgrid (2° resolution) global annual gross primary production (GPP) across the last forty years of a transient CLM5.1 simulation. We opted for 400 clusters to balance computational cost against representativeness.

eters globally include parameters that were important in each of these three biomes. 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 and contains 1720 figures (`https://webext.cgd.ucar.edu/I2000/PPEn11_OAAT`).

In many cases we found it easier to utilize interactive plotting utilities to toggle through the various parameters, output variables, and forcing scenarios (Figure 8). Interactive widgets are relatively easy to generate within a data analysis session, but more difficult to host online, hence the screenshot.

4 Discussion

In this project, we identified and perturbed 193 CLM parameters to create a large one-at-a-time PPE. This ensemble is useful for understanding parametric controls on CLM processes. The software and analysis tools generated to enable and utilize this experiment will greatly reduce the burden of generating future PPEs and parameter sensitivity experiments.

There were several barriers to perturbing the full set of CLM5-BGC parameters. Firstly, many parameters had not been officially identified as such. 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. In this experiment we perturbed 193 parameters across a variety of CLM processes (Figure 1). This is not the full set of CLM parameters, as some processes were not included, such as crops. In carry-

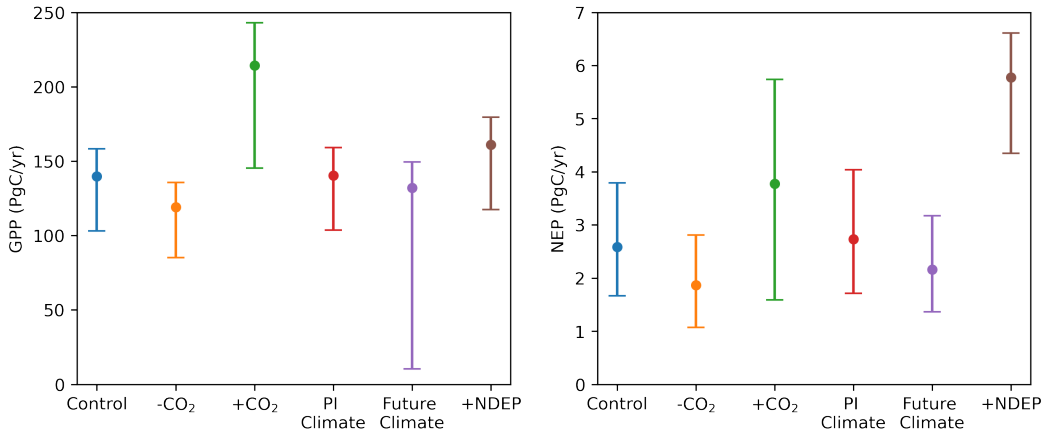

Figure 4. Global annual gross primary production (GPP) and net ecosystem production (NEP) across six forcing scenarios. Circles mark the CLM5.1 default simulation, and the bars span the ensemble ranges.

ing out this experiment, we have likewise realized that the epistemology of climate model 'parameters' is highly nuanced, and that 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.

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. 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 lacking sufficient empirical backing for robust range evaluation. Even with parameters that have an empirical basis, 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 potentially non-linear, large ensembles are required to adequately resolve response surfaces. With standard CLM configurations, such ensembles would far exceed our computing allocation. 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 linearized spin-up solver (see Sections 2.2 and 2.3 for details). As long as computational resources remain constrained, designing faster model configurations and more 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 be quite large, in some cases exceeding the effects of our various forcing scenarios (Figure 4). That being said, the majority of parameter perturbations had small effects for any specific model variable or metric, such that a majority of simulations were clustered

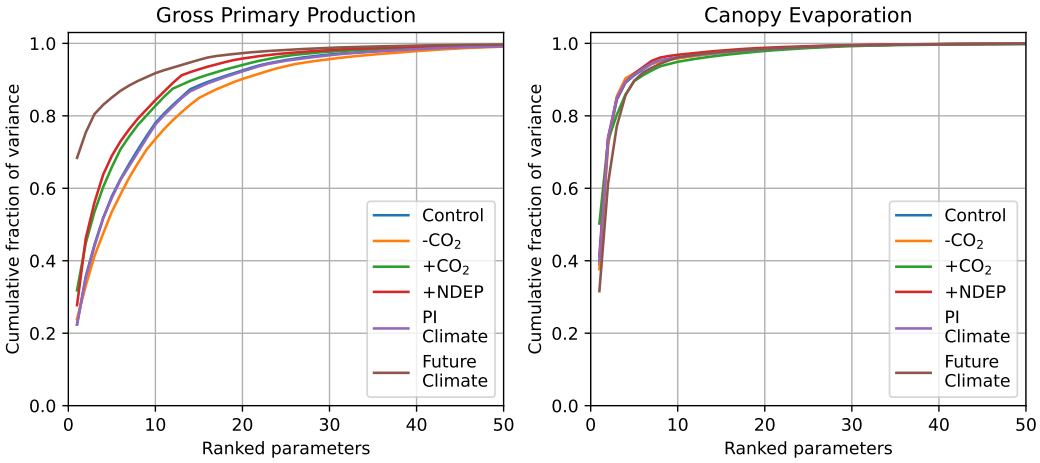

Figure 5. Cumulative fraction of variance explained by the most influential parameters on gross primary production and canopy evaporation. For example approximately 96% of canopy evaporation variance is explained by the first ten parameters regardless of forcing, whereas between 74 and 92% of GPP variance is explained by the first ten parameters. Only the top 50 parameters are shown (of 193) to improve the visualization. Note that parameter rankings differ by variable.

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 193 parameters proves unattainable.

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. Similarly, the parameters controlling leaf area index varied significantly depending on biome (Figure 7).

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).

Several spin-off projects have leveraged the work presented here, including some that have already reached publication status (Cheng et al., 2023; Yan et al., 2023a, 2023b). These projects utilized the output from this ensemble to filter for parameters that influence the relevant study domains and used the parameter ranges collated in Section

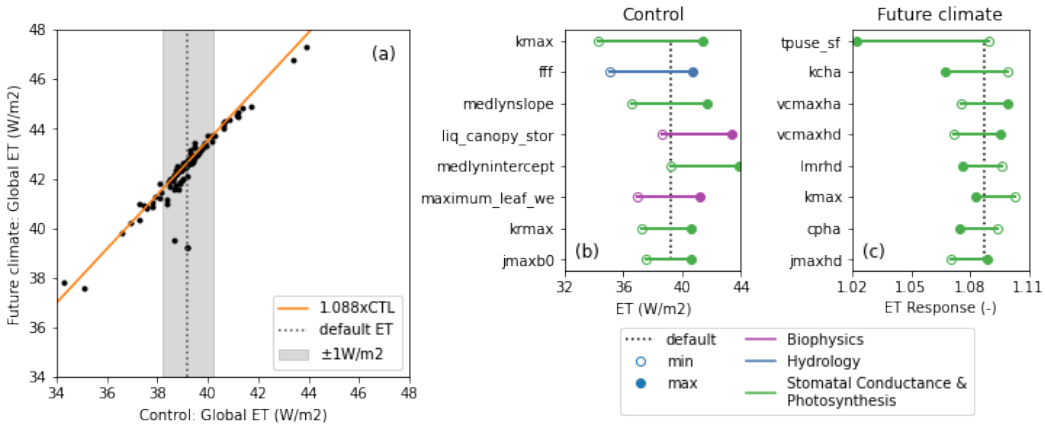

Figure 6. (a) Global annual average evapotranspiration (ET) in the future climate scenario vs. our control experiment. The shading spans the default ET in the control experiment, ±1 W/m2. The response of ET to future climate with default parameters is +8.8%. The orange line shows the expected future ET assuming a linear default ET response. (b) The top 8 parameters governing global ET in the control experiment. (c) The top 8 parameters governing ET response to future climate, i.e. the ratio of future ET to control ET. Parameters governing the change in ET expected in the future tend to differ from the parameters controlling present-day ET.

2.4. Our project has accelerated parameter exploration work within our collaborator network by providing:

- Parameter ranges for nearly 200 parameters
- Purpose-built unstructured grid (sparsegrid)
- Accelerated spinup procedure
- Ensemble generation scripting toolchain
- Parameter sensitivity diagnostics for nearly 200 parameters across 6 forcing scenarios

We have likewise begun a follow-on activity that perturbs a subset of important parameters with the Latin hypercube sampling strategy (?, ?) to work towards routine model calibration. Our first foray into exploring the potential for global calibration is with leaf area index due to controls that this variable has on biogeophysical and biogeochemical fluxes. By investing in the infrastructure that we introduce here, all of our subsequent parameter perturbation experiments have required much less time and effort. We expect to continue to extend our efforts in this domain towards model calibration, as well as repeat the foundational one-at-a-time experiments with subsequent model releases.

Open Research Section

The model code for this experiment is contained in a development tag (`https://github.com/ESCOMP/CTSM/tree/branch_tags/PPE.n11_ctsm5.1.dev030`).

(CTSM component set longname is:
`2000_DATM%GSWP3v1_CLM51%BGC_SICE_SOCN_SROF_SGLC_SWAV_SIAC_SESP`)

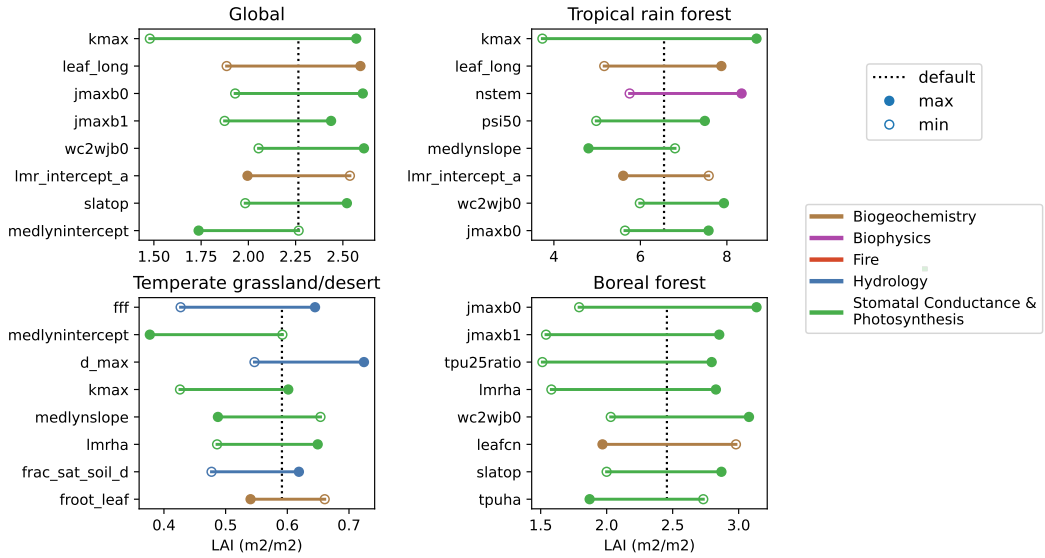

Figure 7. The eight most influential parameters on leaf area index within the control ensemble, globally and within three biomes.

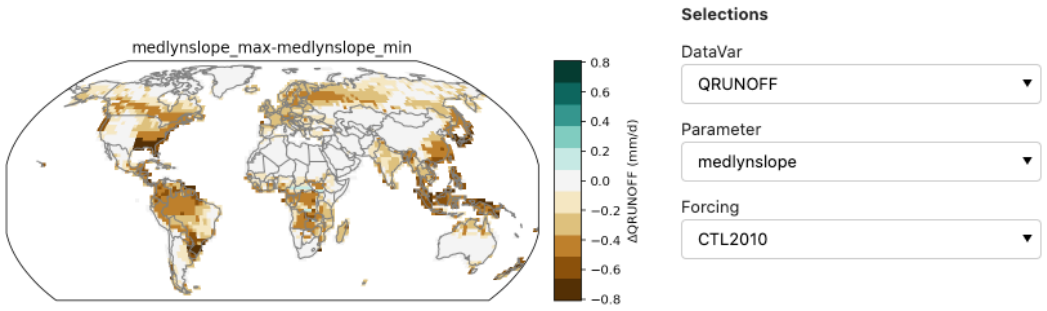

Figure 8. Screenshot of interactive diagnostic for exploring maps of parameter effects. In this case, the effect of medlynslope on runoff within the control ensemble. Increasing medlynslope tends to reduce runoff, but only in regions with sufficient vegetation activity.

This section MUST contain a statement that describes where the data supporting the conclusions can be obtained. Data cannot be listed as "Available from authors" or stored solely in supporting information. Citations to archived data should be included in your reference list. Wiley will publish it as a separate section on the paper's page. Examples and complete information are here: https://www.agu.org/Publish with AGU/Publish/Author Resources/Data for Authors

Acknowledgments

Enter acknowledgments here. This section is to acknowledge funding, thank colleagues, enter any secondary affiliations, and so on.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 parame-

ter 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

Appendix A Supplementary Figures

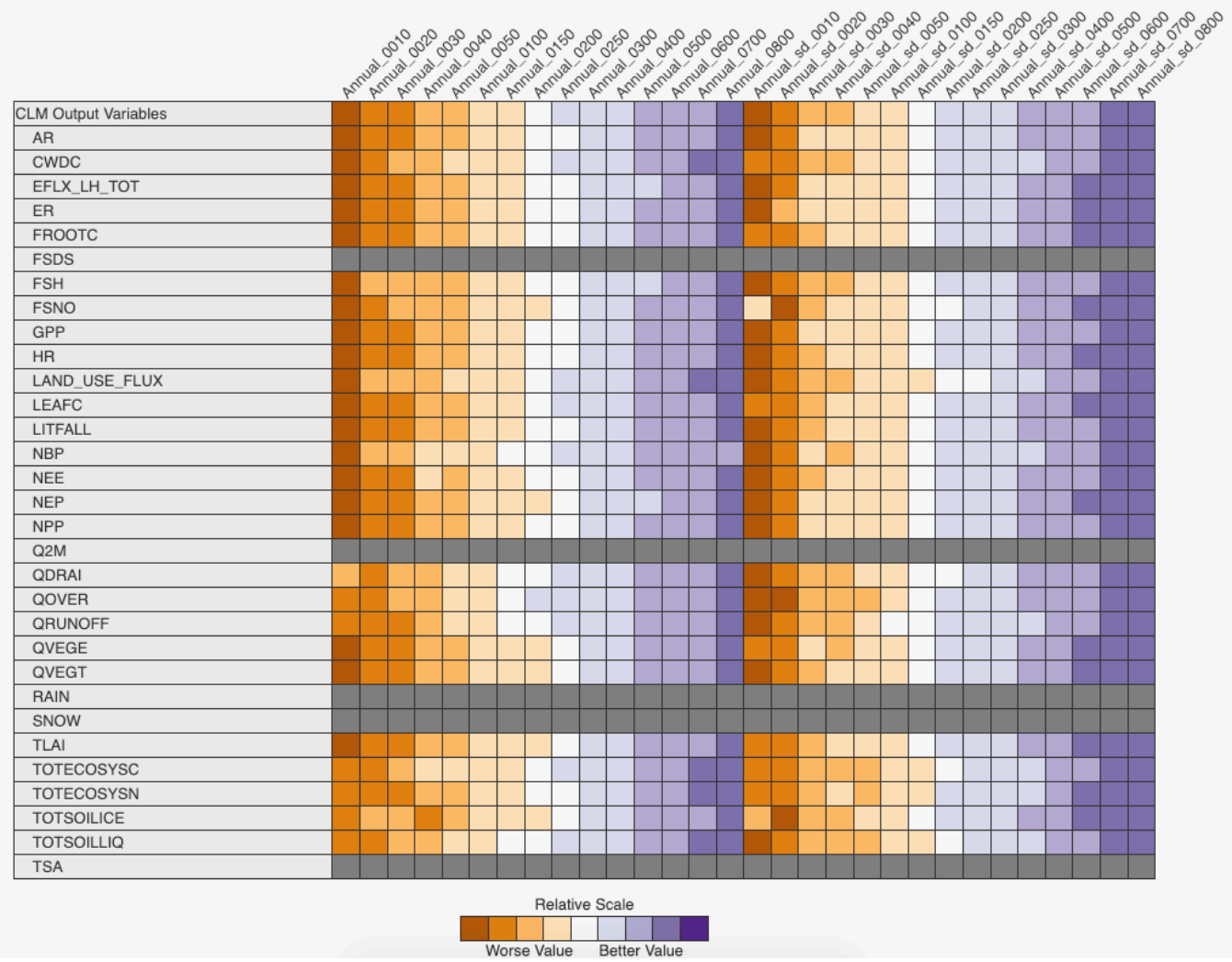

Figure A1. ILAMB 2.5 overall scores comparing remapped sparsegrid output to the full grid 2° model output. Column headings indicate the number of clusters, and whether annual means or annual means and standard deviations were used as input to the clustering algorithm. We should try to remove the gray bits (zqz). Full, interactive results are available at www.ilamb.org/PPE/CLM/2021-02. See main text Section 2.3 for clustering details.

Will be deleted:

A1 Parameters

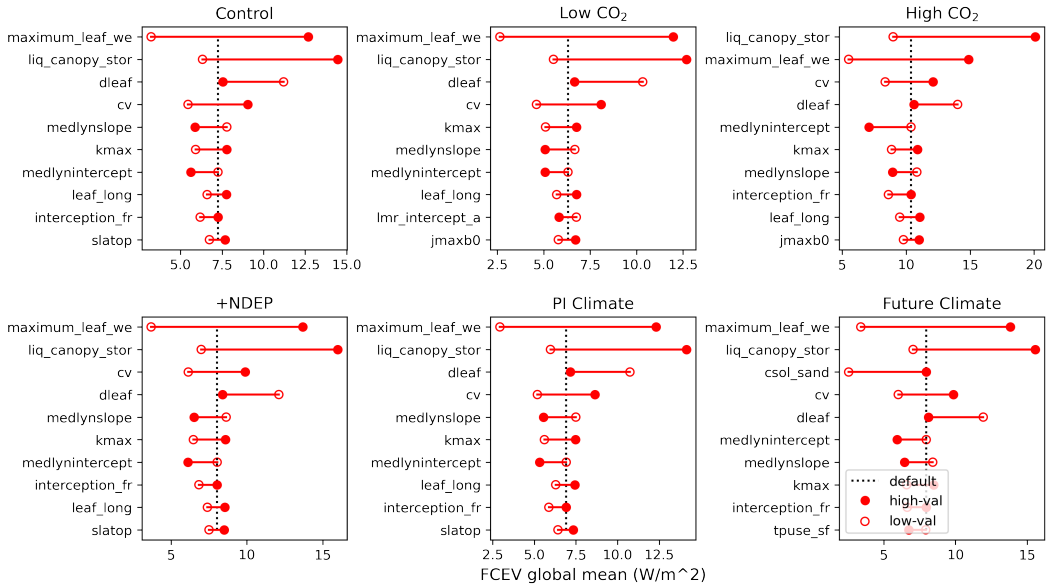

Figure A2. The top ten parameters controlling global annual canopy evaporation (CLM variable: FCEV) across the six forcing scenarios. Note that this plot is one of many from our more comprehensive diagnostics set `https://webext.cgd.ucar.edu/I2000/PPEn11_OAAT/`.

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)*