Understanding the Influence of Parameter Value Uncertainty on Climate Model Output: Developing an Interactive Dashboard

PROPOSERS

Sofia Ingersoll | Master of Environmental Data Science Candidate | Bren School of Environmental Science & Management | singersoll@bren.ucsb.edu | (805) 636 – 6859

Heather Childers | Master of Environmental Data Science Candidate | Bren School of Environmental Science & Management | hmchilders@bren.ucsb.edu | (805) 570 – 0365

CLIENT

Daniel Kennedy | Project Scientist | Climate & Global Dynamics Lab at the National Center for Atmospheric Research | djk2120@ucar.edu | (617) 922-8813

OBJECTIVE

Scientists at the National Center for Atmospheric Research (NCAR) have recently carried out several experiments to better understand the uncertainties associated with future climate projections. In particular, the Community Land Model (CLM) working group has completed a large parameter perturbation ensemble, testing the effects of over 200 model parameters. The CLM parameter perturbation experiment has been designed for uncertainty quantification, model calibration, and as an educational resource to document model behavior. **The objective of this project is to develop an improved data visualization dashboard for communicating the results of the CLM parameter perturbation experiment.**

SIGNIFICANCE

The current website for displaying model results has been useful for the scientists within the Climate & Global Dynamics Lab (CGDL) at NCAR. Still, the interface is not intuitive or informative to the broader scientific audience or the general public. The current website is structured for the user to click multiple links to find the desired output, and the only available outputs are static images with non-descriptive titles. Outputs from the many model simulations total over 4TB and span several scientific disciplines. An ideal website would allow users to specify their desired parameter or variable and run a function that graphs the output data with updated legends and titles. Providing a public interface for perusing the results of the experiment has been an essential but understaffed task within the larger project. The interactive dashboard would provide an interface where new or experienced users can query the experiment database to ask which environmental processes are affected by a given model parameter, or vice versa, and which parameters are the most influential on a given environmental process.

BACKGROUND

Around the world, unprecedented record-breaking climate extremes continue to occur more and more frequently. The major driving forces responsible for this are greenhouse gasses; however, they are not the only significant contributing factor to climate change. What we consider to be impactful to our climate is the key to deepening our understanding of the causation of climate change. A broad scope of environmental parameters collectively influence our planet's homeostasis capabilities. Approximately a year ago, the CGDL at NCAR joined the team of over 50 climate modeling centers worldwide participating in the sixth evolution of the World Coupled Model Intercomparison Project, CMIP6. The CMIP6 project aims to implement interdisciplinary environmental parameters to better understand their interactions and their subsequent effects on local climate systems through the application of the Community Land Model (CLM). The projections provided by the CLM directly influence the Intergovernmental Panel on Climate Change (IPCC) Assessment Reports.

The CLM model experiment is focused on limiting uncertainty around biogeophysical parameters that influence the balance of chemical cycling and sequestration variables. These variable outputs were analyzed for various conditions, including high/low carbon dioxide levels and present/future climate. Each variable is associated with a set of relevant parameters that were perturbed across a range of reasonable estimates for each condition.

EQUITY

The main equity component of this project relates to the broader research objectives of the CLM parameter perturbation experiment. The goal of that project is to significantly improve climate model projections through calibration and uncertainty quantification. Climate hazards are deeply intertwined with patterns of global inequality, such that more reliable future projections of climate change can aid in climate adaptation and decision-making. The dashboard in this project will substantially accelerate progress towards improved model calibration and efforts to improve its climate change projections. While there are components of equity, the primary objectives of this project are not rooted in environmental justice and equity.

DATA

The CLM parameter perturbation experiment encompasses two experiments where parameters were varied systematically to see the impacts on various climate variables. The full model output (4TB) is available on the NCAR supercomputer and can be downloaded to the Bren server if needed. A post-processed subset of the data (300MB) is hosted online here. Students can perform this project using NCAR's supercomputer, which would obviate data downloads and ease the web deployment aspect of this project.

The perturbation experiment looked at **32 variables associated with land-based climate change**.
The data includes simulations with LAI prescribed from satellite phenology (SP) and simulations with**:**

- prognostic vegetation state
- active biogeochemistry (BGC)
- sensitivities to land use
- sensitivities land cover change
- active crop management
- elevated concentrations of atmospheric CO2
- nitrogen enrichment

Some output variables include:

- annual maximum active layer depth [**Figure 4**]
- fire area burned [**Figure 5**]
- total projected leaf area index [**Figure 6**]
- total latent heat flux [**Figure 7**]
- total soil carbon [**Figure 8**].

Within each of these variables, a unique subset of parameters was perturbed. The outputs were then displayed categorically between global average, global impacts, adaptation and vulnerability, global amplitude, and biome-specific averages. The un-processed outputs from the parameter perturbation experiment are available as net Common Data Forms. These netCDFs are a community standard way of quickly accessing or sharing array-oriented scientific data. The files contain variable information and data that can interface easily with Python.

POSSIBLE APPROACHES

The project aims to refine the following target skills: data visualization of 'big' data in Python, statistical analysis, web integration and containerization, introductory machine learning, performance computing, and geospatial data analysis.

This project has the capacity to be scaled to the interests and time constraints of the students working on the project. The following points provide a general outline for the progression of this capstone project:

- Begin exploring the data and create documentation for the variable/parameter names, units, and significance
- Explore Python packages that can facilitate the development of a dashboard, including Pangeo, Dask, and Xarray
- Create an interactive data visualization dashboard to create maps or graphs with updated titles and variable names based on user input. **These visualizations would draw from the existing code used to make static images currently available on the website.**

If time allows, the following features can be added to the dashboard:

- Increase the versatility of user inputs to allow for searching by perturbed parameters or variables of interest (The current site requires users to select a variable and then select a parameter using point-and-click navigation)
- Create interactivity to perturbate multiple parameters at once
- Develop new plotting routines, e.g. user-defined regions of interest
- Use elementary machine learning to draw relevant conclusions about parameter interaction

COMPUTATIONAL TOOLS & NEEDS

Language: Python
Data: Current Web application , General CESM Data
Packages: Pangeo, Dask, Xarray, Holoviz
Server: supercomputer

DELIVERABLES

- **A user-friendly, interactive dashboard** that will allow for a broad scope of individuals and NCAR employees to access the land model data easily and efficiently. The interactive dashboard should provide an interface where new or experienced users can query the experiment database.

Potential expansion of the project scope, if time and abilities allow:

- Serve visualizations from the raw data output (2.4TB) in lieu of the post-processed subset (300Mb).
- Incorporate results from the second experiment where parameters are varied simultaneously, which would require some basic machine learning.
- Deploy the prototype online and/or within a docker container with a robust Python script that requires minimal maintenance. The upkeep of this web-deployed deliverable would be managed by the CGDL at NCAR.

AUDIENCE

The primary audience for this project is the climate scientists between NCAR and CMIP who are working with the CGDL data but were not involved in creating the original experiment. Improving the readability of the data will also increase the likelihood of the public accessing the data.

BUDGET

The production of the project is not expected to require funding.

ABBREVIATIONS

CMIP: Coupled Model Intercomparison Project
IPCC: Intergovernmental Panel on Climate Change
NCAR: National Center of Atmospheric Research
CGDL: Climate & Global Dynamics Lab

REFERENCES

1. Friedlingstein, Pierre, et al. "Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks." Journal of Climate, vol. 27, no. 2, 15 Jan. 2014, pp. 511–526, https://doi.org/10.1175/jcli-d-12-00579.1.

2. Lawrence, David M., et al. "The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty." Journal of Advances in Modeling Earth Systems, 12 Dec. 2019, agupubs.onlinelibrary.wiley.com/doi/10.1029/2018MS001583, https://doi.org/10.1029/2018ms001583.

3. "Diagnostics." Webext.cgd.ucar.edu, webext.cgd.ucar.edu/I2000/PPEn11_OAAT/. Accessed 18 Oct. 2023.

4. Michaut, Catherine. "CMIP." Wcrp-Climate.org, 26 Nov. 2013, www.wcrp-climate.org/wgcm-cmip.

5. IPCC. "About — IPCC." Ipcc.ch, IPCC, 2019, www.ipcc.ch/about/.

6. "Mission & Vision | Climate & Global Dynamics." Www.cgd.ucar.edu, www.cgd.ucar.edu/about/mission-vision. Accessed 18 Oct. 2023.

7. "Index of /I2000/PPEn11_OAAT/Data." Webext.cgd.ucar.edu, webext.cgd.ucar.edu/I2000/PPEn11_OAAT/data/. Accessed 18 Oct. 2023.

8. "Documentation | ARC NCAR." Arc.ucar.edu, arc.ucar.edu/knowledge_base/70549550. Accessed 18 Oct. 2023.

9. "About Pangeo — Pangeo Documentation." Pangeo.io, pangeo.io/about.html#about. Accessed 18 Oct. 2023.

10. "Dask — Dask Documentation." Docs.dask.org, docs.dask.org/en/stable/index.html. Accessed 18 Oct. 2023.

11. "Xarray Documentation." Xarray, docs.xarray.dev/en/stable/index.html. Accessed 18 Oct. 2023.

12. "Overview — HoloViz 0.16.2 Documentation." Holoviz.org, holoviz.org/tutorial/Overview.html. Accessed 18 Oct. 2023.

13. "Dataset: CLM Land-Only Release." Earthsystemgrid.org, 2023, www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.CLM_LAND_ONLY.html.

SUPPLEMENTAL FIGURES

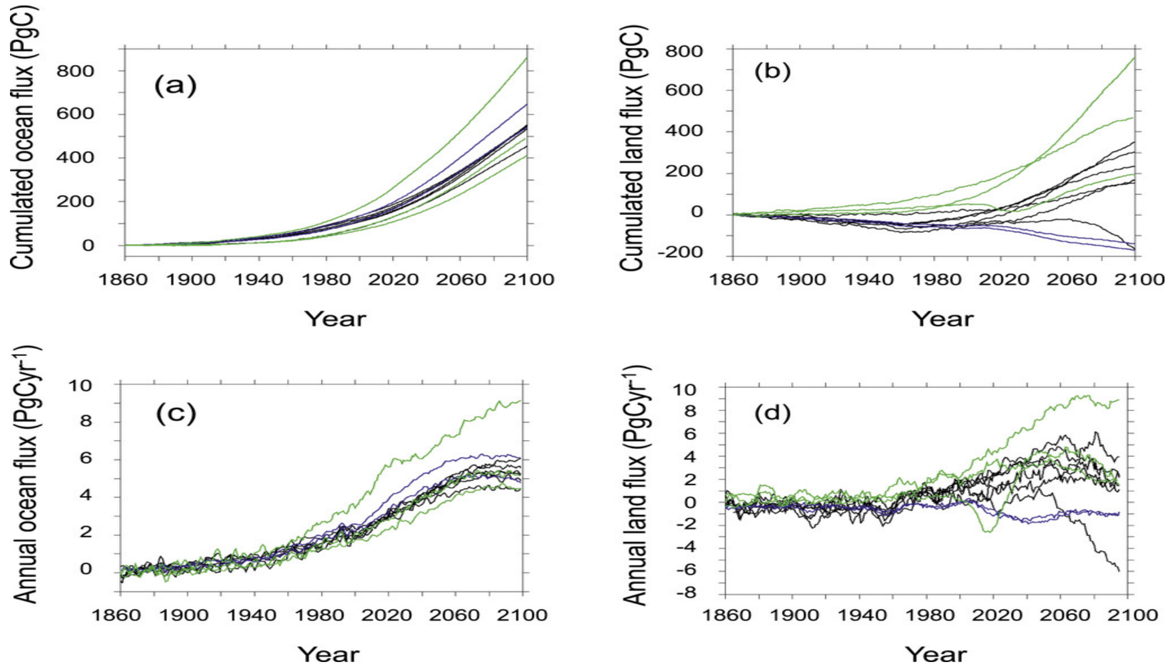


Figure 1: **Publication that utilized the multi-model outputs collected by the Global Carbon Project.** The images depict a range of (a) cumulative global air to ocean carbon flux (PgC), (b) cumulative global air to land carbon flux (PgC) from the 11 ESMs E-driven simulations, (c) the annual global air to ocean carbon flux, and (d) annual global air to land carbon flux (Friedlingstein, et. al).

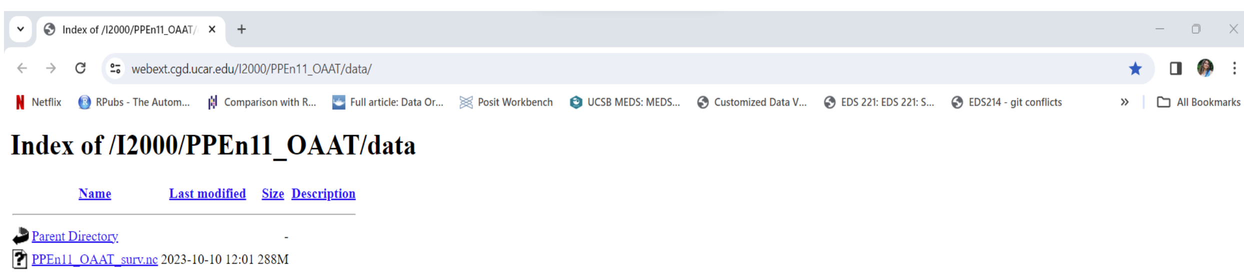
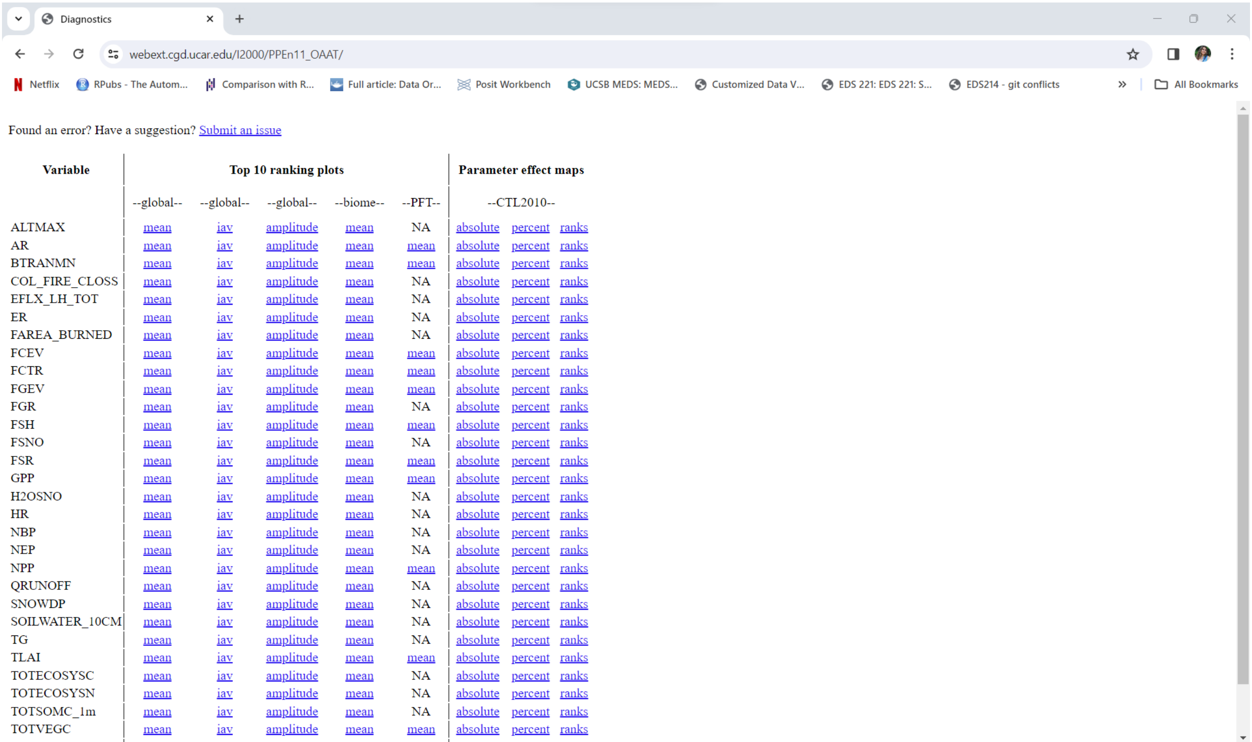


Figure 2: Home toggle page of metadata, includes a **link to** the **Parent Directory of visualizations** for the individual perturbation metadata and a **link to download the NETCDF file**: PPEn11_OAAT_surv.nc.



Variable	Top 10 ranking plots					Parameter effect maps		
	--global--	--global--	--global--	--biome--	--PFT--	--CTL2010--		
ALTMAX	mean	iav	amplitude	mean	NA	absolute	percent	ranks
AR	mean	iav	amplitude	mean	mean	absolute	percent	ranks
BTRANMN	mean	iav	amplitude	mean	mean	absolute	percent	ranks
COL_FIRE_CLOSS	mean	iav	amplitude	mean	NA	absolute	percent	ranks
EFLX_LH_TOT	mean	iav	amplitude	mean	NA	absolute	percent	ranks
ER	mean	iav	amplitude	mean	NA	absolute	percent	ranks
FAREA_BURNED	mean	iav	amplitude	mean	NA	absolute	percent	ranks
FCEV	mean	iav	amplitude	mean	mean	absolute	percent	ranks
FCTR	mean	iav	amplitude	mean	mean	absolute	percent	ranks
FGEV	mean	iav	amplitude	mean	mean	absolute	percent	ranks
FGR	mean	iav	amplitude	mean	NA	absolute	percent	ranks
FSH	mean	iav	amplitude	mean	mean	absolute	percent	ranks
FSNO	mean	iav	amplitude	mean	NA	absolute	percent	ranks
FSR	mean	iav	amplitude	mean	mean	absolute	percent	ranks
GPP	mean	iav	amplitude	mean	mean	absolute	percent	ranks
H2OSNO	mean	iav	amplitude	mean	NA	absolute	percent	ranks
HR	mean	iav	amplitude	mean	NA	absolute	percent	ranks
NBP	mean	iav	amplitude	mean	NA	absolute	percent	ranks
NEP	mean	iav	amplitude	mean	NA	absolute	percent	ranks
NPP	mean	iav	amplitude	mean	mean	absolute	percent	ranks
QRUNOFF	mean	iav	amplitude	mean	NA	absolute	percent	ranks
SNOWDP	mean	iav	amplitude	mean	NA	absolute	percent	ranks
SOILWATER_10CM	mean	iav	amplitude	mean	NA	absolute	percent	ranks
TG	mean	iav	amplitude	mean	NA	absolute	percent	ranks
TLAI	mean	iav	amplitude	mean	mean	absolute	percent	ranks
TOTECOSYSC	mean	iav	amplitude	mean	NA	absolute	percent	ranks
TOTECOSYSN	mean	iav	amplitude	mean	NA	absolute	percent	ranks
TOTSOMC_1m	mean	iav	amplitude	mean	NA	absolute	percent	ranks
TOTVEGC	mean	iav	amplitude	mean	mean	absolute	percent	ranks

Figure 3: Parent Directory home page containing perturbation data for 32 environmental variables. Organized into **two categories**:

- "Top 10 ranking plots"
- "Parameter effect maps".

The subcategories for "**Top 10 ranking plots**" are:

- –global–mean
- –global–iav
- –global–amplitude
- –biome–
- – PFT–

The "**Parameter effect maps**" subcategories under "**– CTL2010–**" are:

- absolute
- percent
- ranks

Each link sends the user to a new page showing a screenshot of the outcome variables after running the sample. Examples shown in *Figures 4:1*

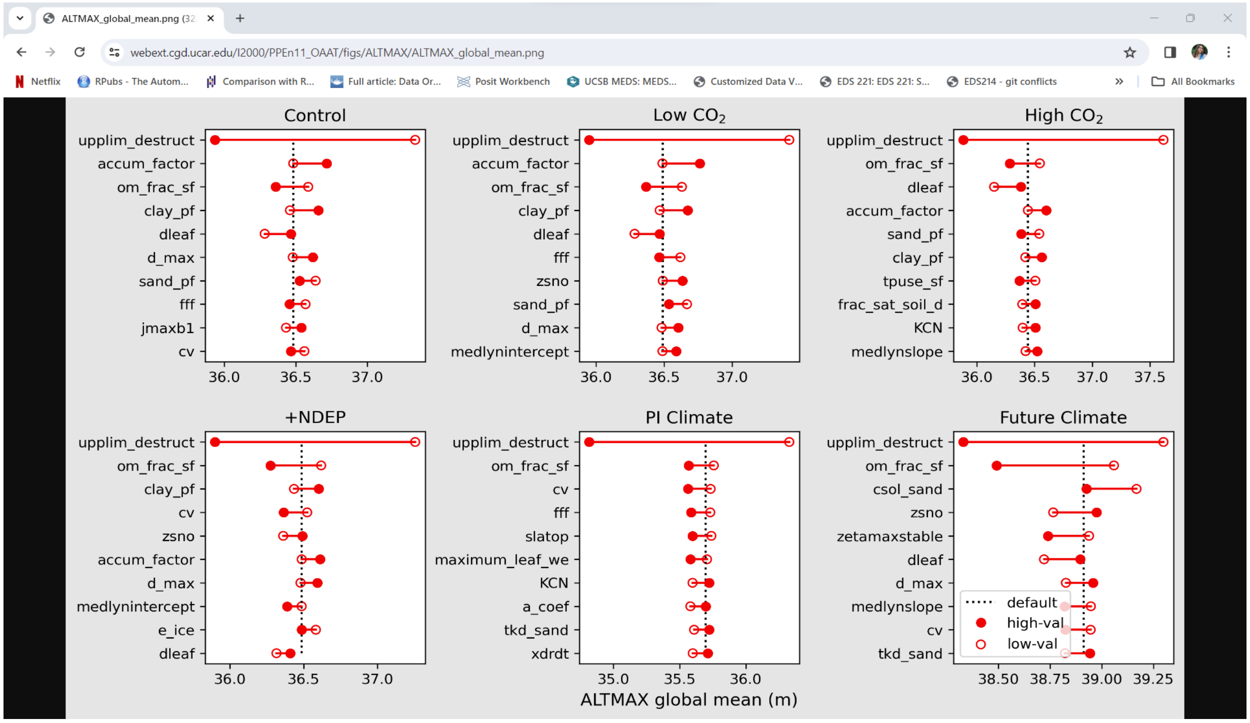


Figure 4: **Parent Directory—Top 10 Ranking plots–global–mean: ALTMAX** "Annual maximum active layer depth" also recognized as the average annual thaw depth in permafrost areas due to solar heating of the surface. **Visualization ATLMAX outcome under 6 various conditions with 10 influencing variables perturbed**. Data considered is a combination of **geophysical, vegetation, and chemical data**.

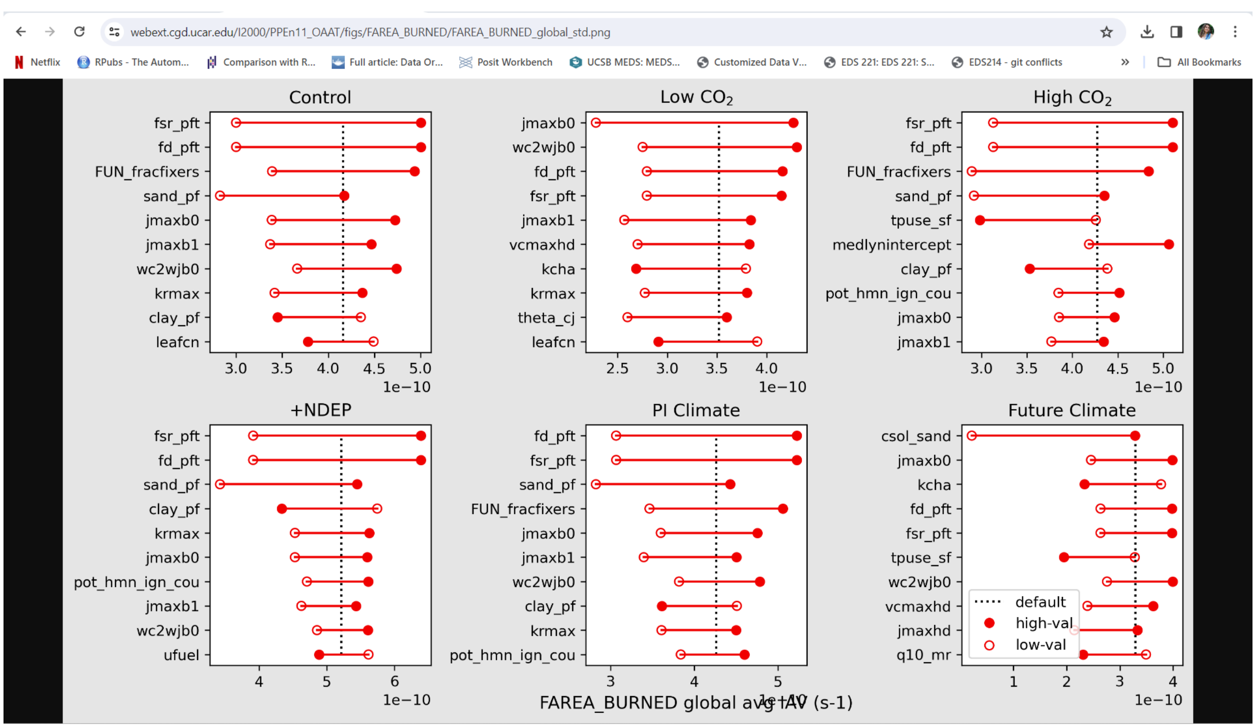


Figure 5: **Parent Directory—Top 10 Ranking plots–global–iav: FAREA_BURNED** "Fire area burned"

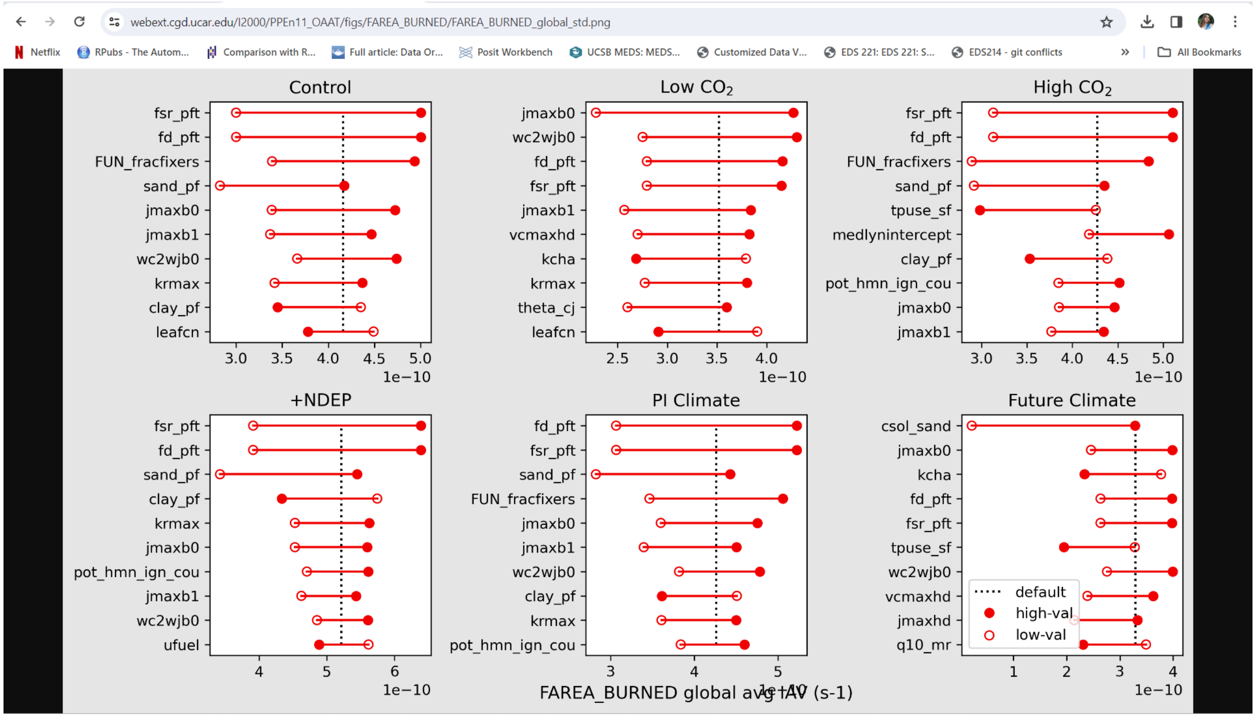


Figure 6: Parent Directory—Top 10 Ranking plots–global–amplitude:TLAI "Total projected leaf area index".

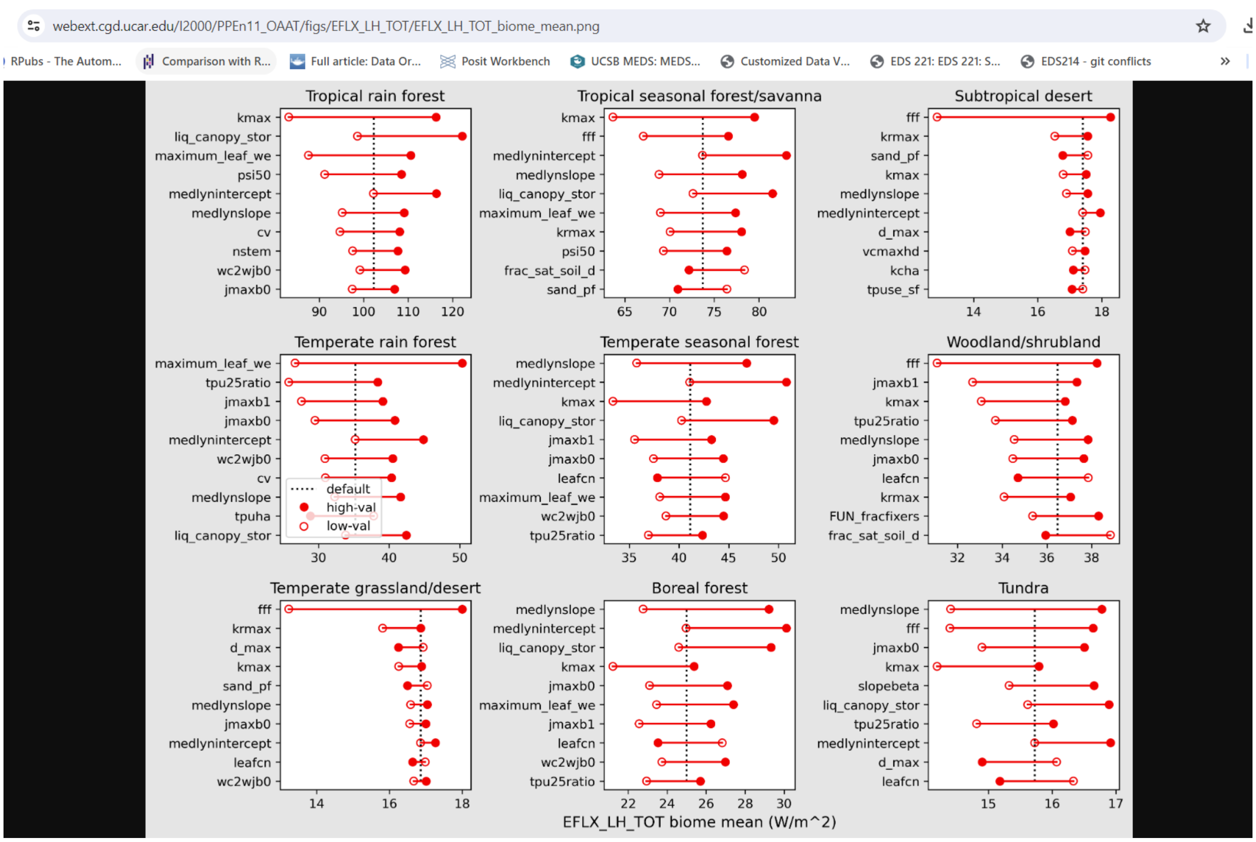
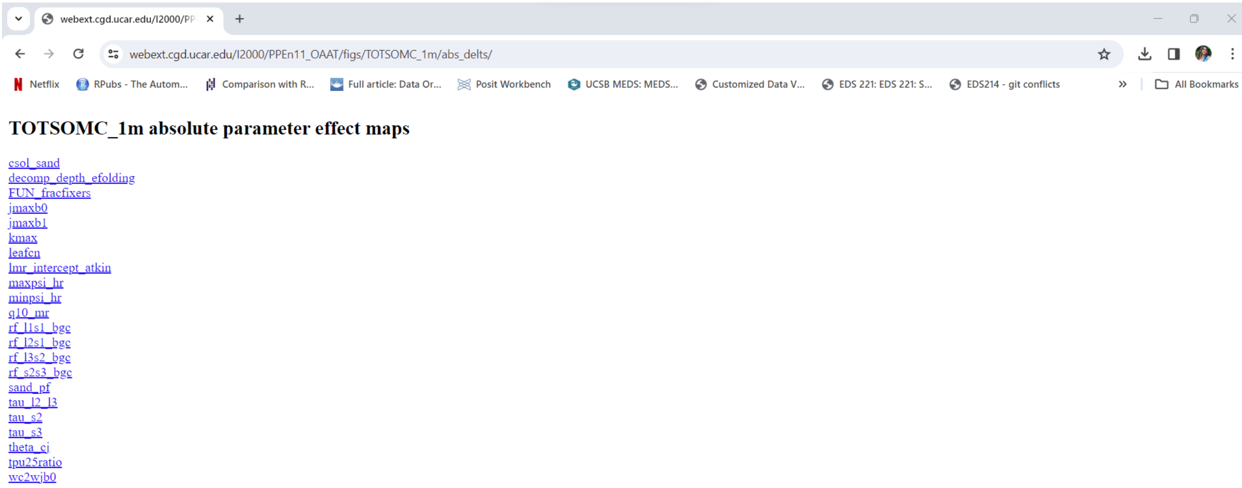


Figure 7: Parent Directory—— Top 10 Ranking plots–biome–mean:EFLX_LH_TOT "Total latent heat flux".


webext.cgd.ucar.edu/I2000/PPEn11_OAAT/figs/TOTSOMC_1m/abs_delts/

TOTSOMC_1m absolute parameter effect maps

csol_sand
decomp_depth_efolding
FUN_fracfixers
jmaxb0
jmaxb1
kmax
leafcn
lmr_intercept_atkin
maxpsi_hr
minpsi_hr
q10_mr
rf_l1s1_bgc
rf_l2s1_bgc
rf_l3s2_bgc
rf_s2s3_bgc
sand_pf
tau_l2_l3
tau_s2
tau_s3
theta_cj
tpu25ratio
wc2wjb0

Figure 8.A.: Parent Directory—Parameter Effects map–CLT2010–absolute:TOTSOMC_1m– “total soil carbon” directory to variable output maps.

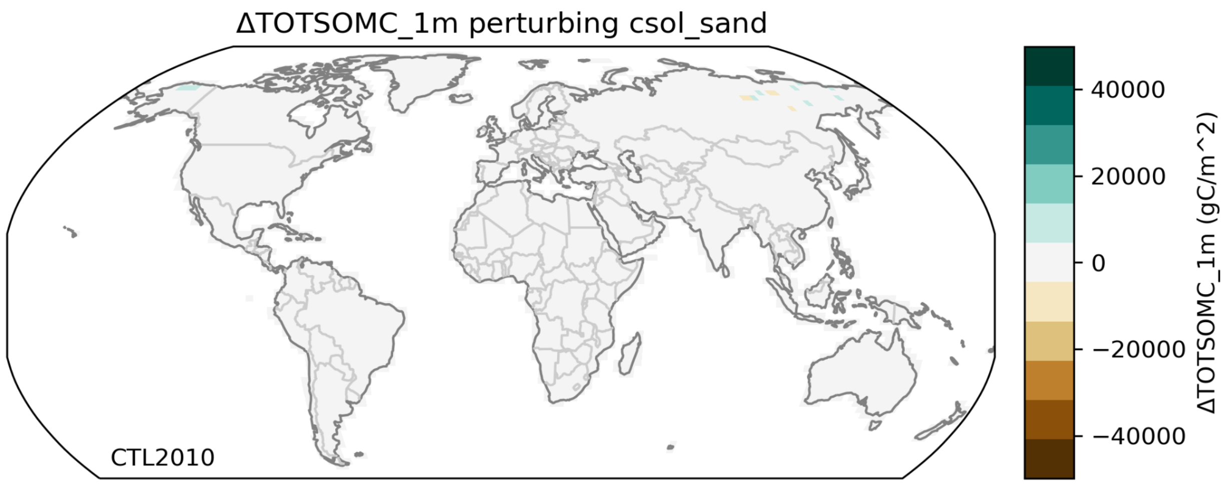
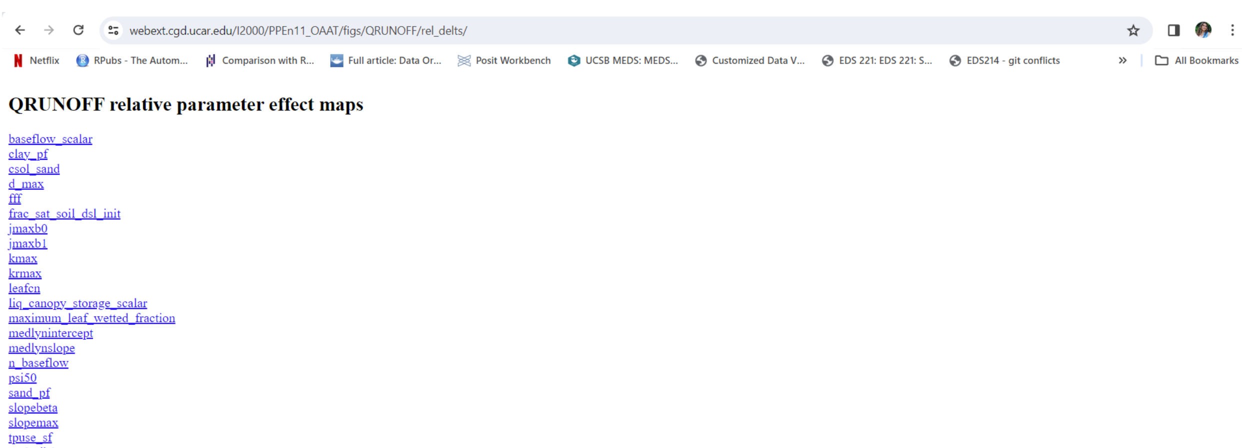


Figure 8.B.: Parent Directory—Parameter Effects map– CLT2010–absolute:TOTSOMC_1m: csol–sand. This figure in particular emphasizes the need for a scaling ability to better assess regions of interest.


webext.cgd.ucar.edu/I2000/PPEn11_OAAT/figs/QRUNOFF/rel_delts/

QRUNOFF relative parameter effect maps

baseflow_scalar
clay_pf
csol_sand
d_max
fff
frac_sat_soil_dsl_init
jmaxb0
jmaxb1
kmax
krmax
leafcn
liq_canopy_storage_scalar
maximum_leaf_wetted_fraction
medlynintercept
medlynslope
n_baseflow
psi50
sand_pf
slopebeta
slopemax
tpuse_sf
wc2wjb0

Figure 9.A.: Parent Directory—Parameter Effects map– CLT2010–percent:QRUNOFF

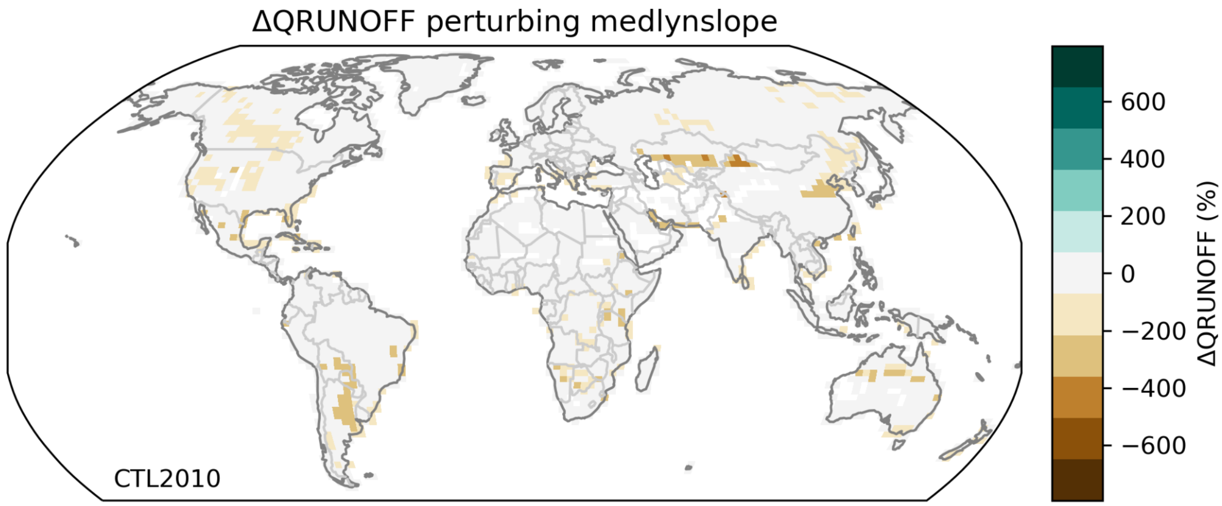


Figure 9.B.: Parent Directory—Parameter Effects map– CLT2010–percent:QRUNOFF: medlynslope "slope of the stomatal conductance–photosynthesis relationship"

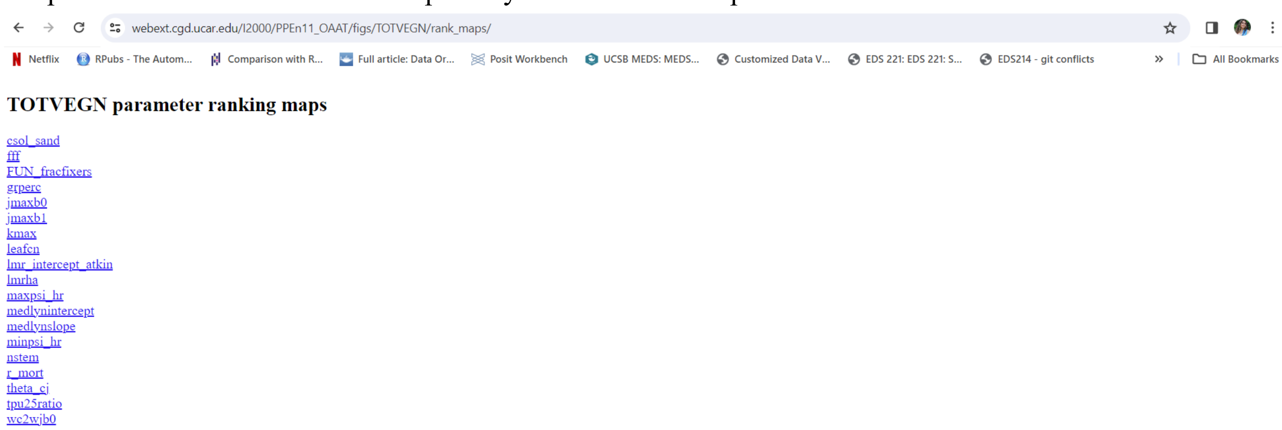


Figure 10.A.: Parent Directory—Parameter Effects map– CLT2010–rank:TOTVEGN "total vegetation nitrogen"

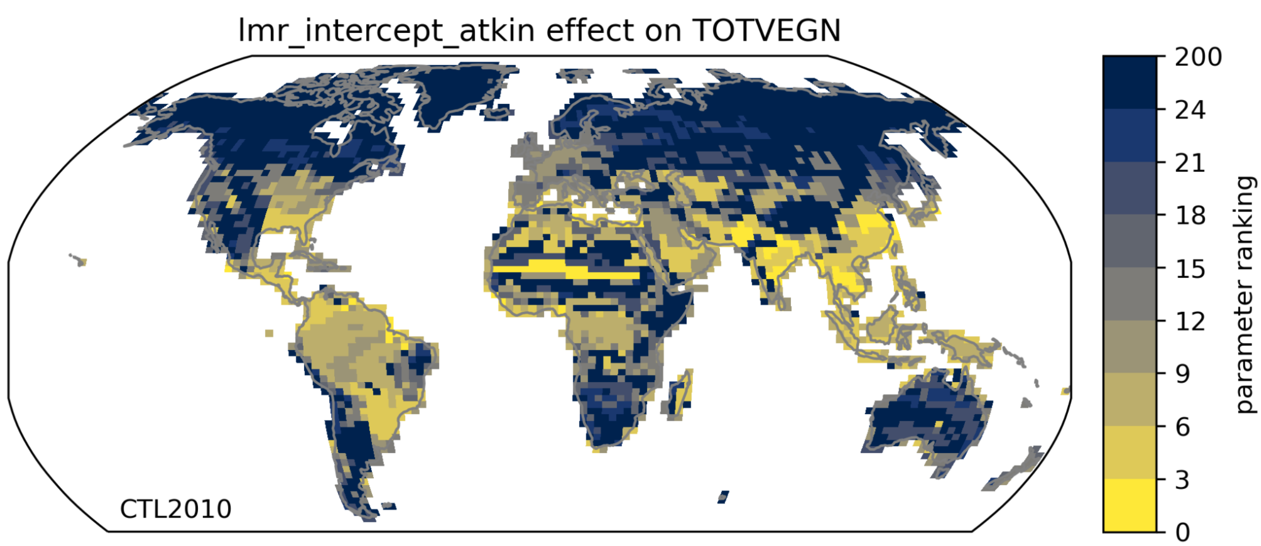


Figure 10.B.: Parent Directory—Parameter Effects map– CLT2010–rank:TOTVEGN: lmr_intercept_atkin "calculation of the top of canopy leaf maintenance respiration base rate µmolCO2 m−2 s−1".

CLIENT LETTER OF SUPPORT

MEDS capstone clienter letter of support

Understanding the Influence of Parameter Value Uncertainty on Climate Model Output: Developing an Interactive Dashboard

I have worked with Sofia Ingersoll and Heather Childers to help develop a potential capstone project for this year's MEDS cohort. We have proposed to develop an online dashboard to interface with a recent parameter sensitivity experiment carried out with the Community Land Model at NCAR. The experiment has generated
4TB of climate model output, which will be useful for uncertainty quantification, model calibration, and to supplement our technical documentation. The dashboard is meant to be outward-facing, whereby collaborators and new model users can peruse the results to better understand how the Community Land Model behaves and the science that is embedded within.

The experiments are complete and we have an existing dashboard, but would value something more interactive and/or flexible. I am hoping that the students can take existing plotting code that we have developed and implement it in a context that incorporates various toggles/widgets. This will likely involve using high-level python tools such as bokeh, plotly, or holoviews, in addition to more basic data analysis and plotting packages. There is also scope within this project for the students to develop new diagnostic plots if that is of interest, i.e. engaging specifically with the climate science side of this project. One or more of the students could also focus on containerizing the dashboard infrastructure if developing that skill was of particular interest.

I hope to engage students by offering the opportunity to gain data science experience working with a large dataset, while also offering some exposure to climate modeling and the general structure and intricacies of climate model outputs. This project is designed toward a specific deliverable, namely a data dashboard, but I am very flexible with regard to tailoring the project to students' interests and proficiencies. I personally enjoy working with beginner python users and trying to lower some of the barriers to entry for geospatial data analysis. Selfishly I am also very excited to work with the MEDS cohort in hopes that they will continue engaging with climate/land modeling. Furthermore, a successful version of this data dashboard requires that it be coherent and informative beyond my own research group. Therefore having the perspective of MEDS students while designing this dashboard would provide a lot of added value.

I have sufficient capacity to help advise with this project and look forward to working with a MEDS group. This could include weekly meetings throughout the project timeline. I have experience with most of the python libraries that would be required, and have worked with many other new python users during tutorials and in research collaborations. My preference is to set up the students on our HPC, because it is very well resourced and supported, but I can adapt if there is a specific reason to utilize Bren resources. The full dataset is 4TB, but we have a post-processed subset that students can use for prototyping that is only 300MB.

Daniel Kennedy
djk2120@ucar.edu

Project Scientist
Climate and Global Dynamics Laboratory
National Center for Atmospheric Research

Affiliate Research Associate
Earth Research Institute
University of California, Santa Barbara