The CLM5 Parameter Perturbation Experiment

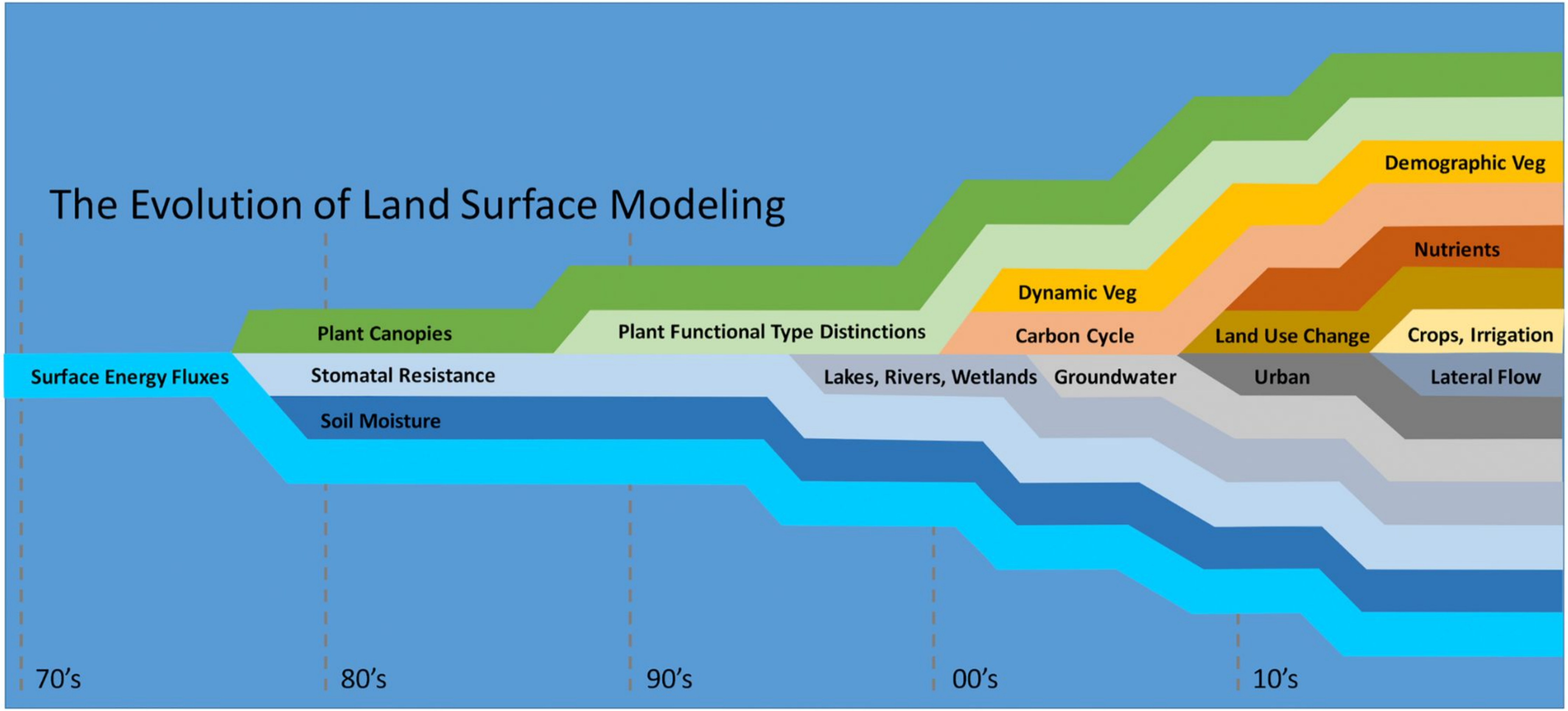
Quantifying parametric uncertainty and working towards automated calibration

Daniel Kennedy, Katie Dagon, Linnia Hawkins, Dave Lawrence
and the CLM5-PPE working group

CGD / LMWG Seminar
February 7, 2023

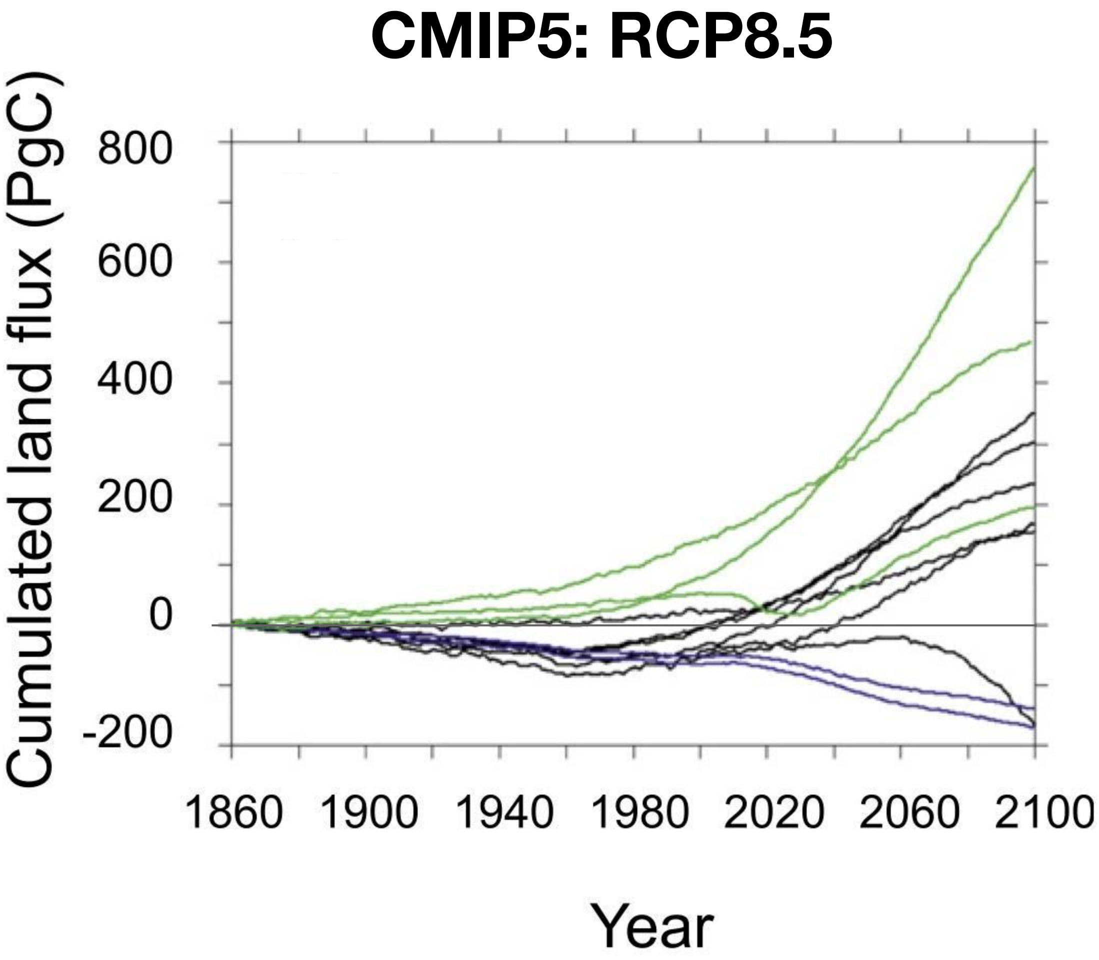
This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Increasing Complexity of Land Models



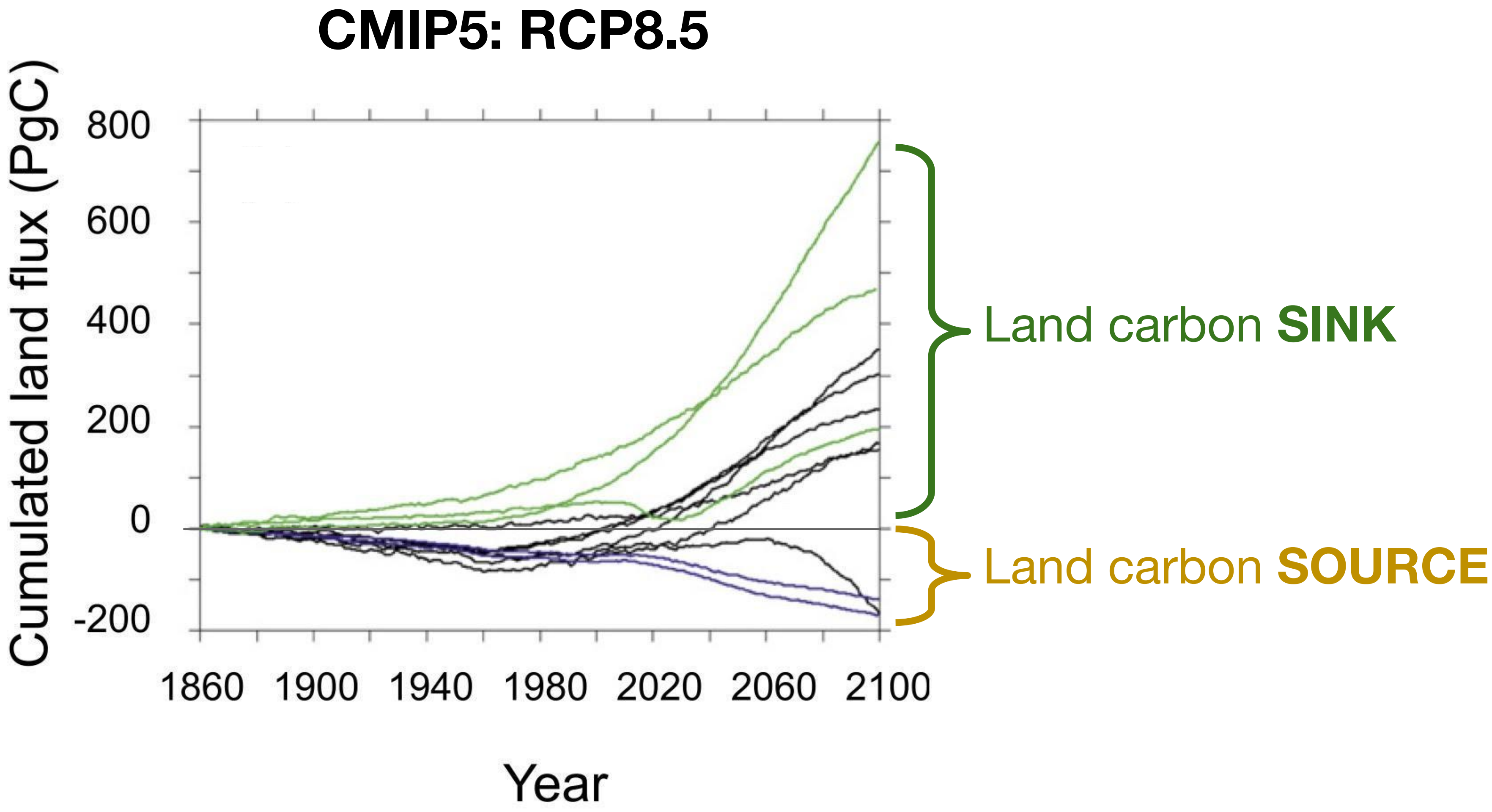
Fisher and Koven (2020)

Carbon Cycle Uncertainty in Land Model Projections



Friedlingstein et al. (2014)

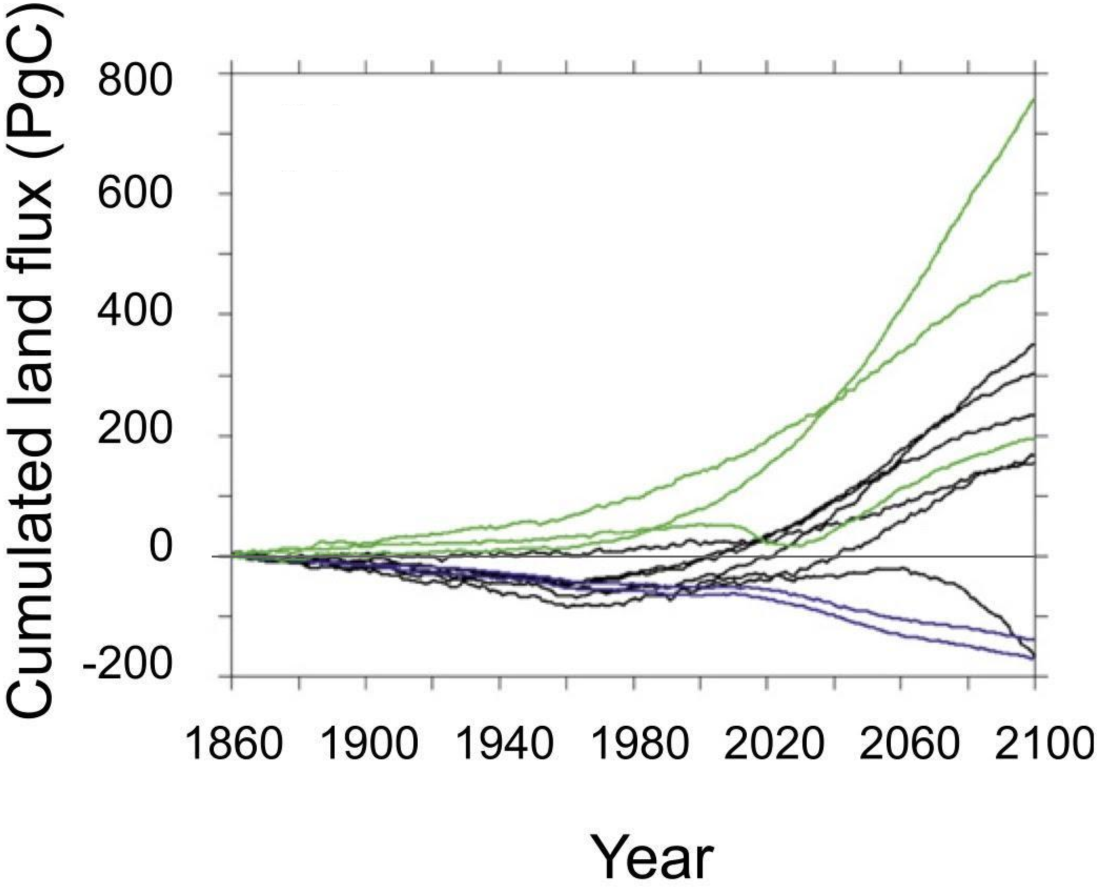
Carbon Cycle Uncertainty in Land Model Projections



Friedlingstein et al. (2014)

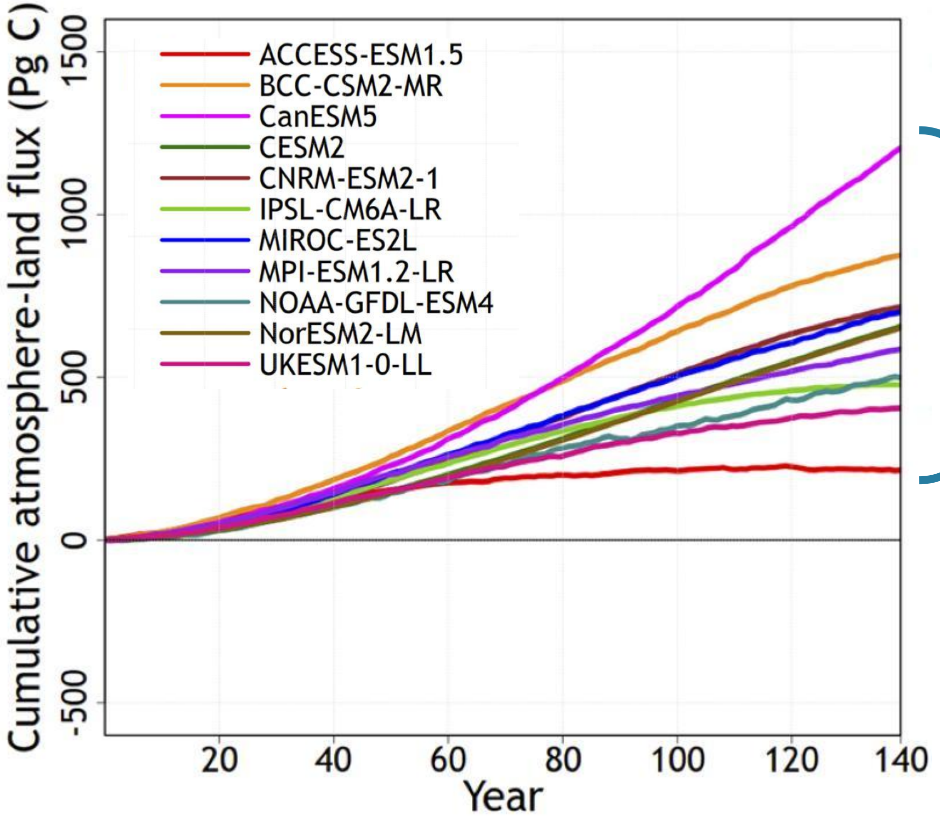
Carbon Cycle Uncertainty in Land Model Projections

CMIP5: RCP8.5



Friedlingstein et al. (2014)

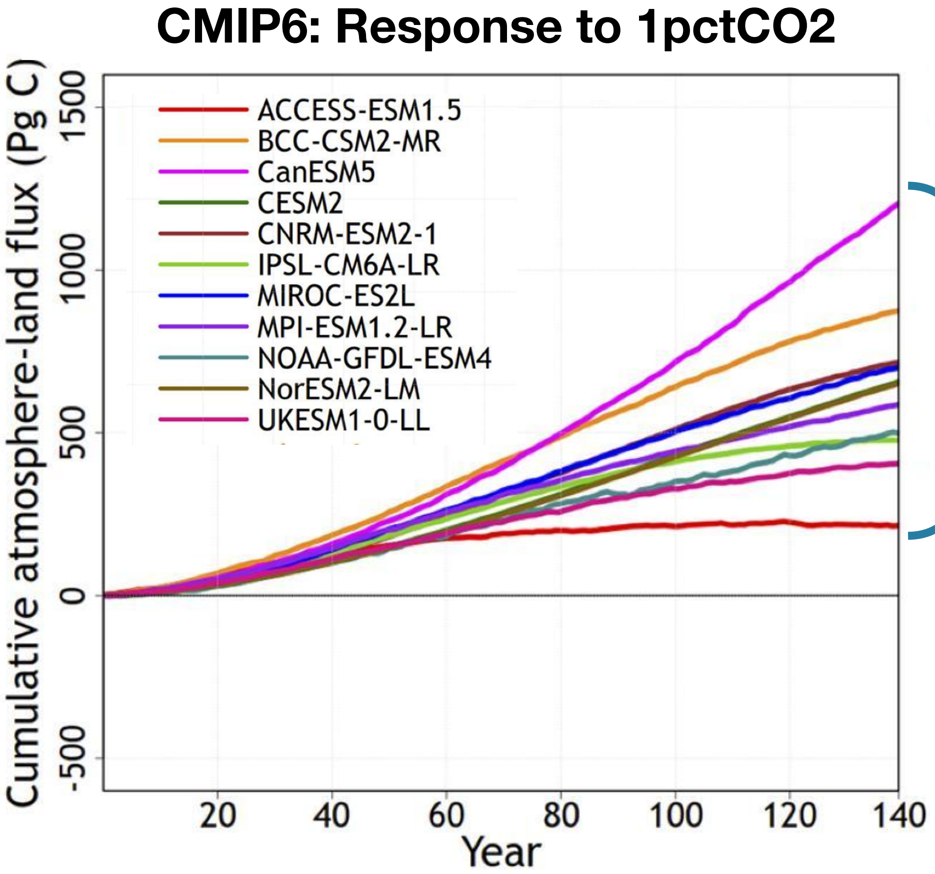
CMIP6: Response to 1pctCO2



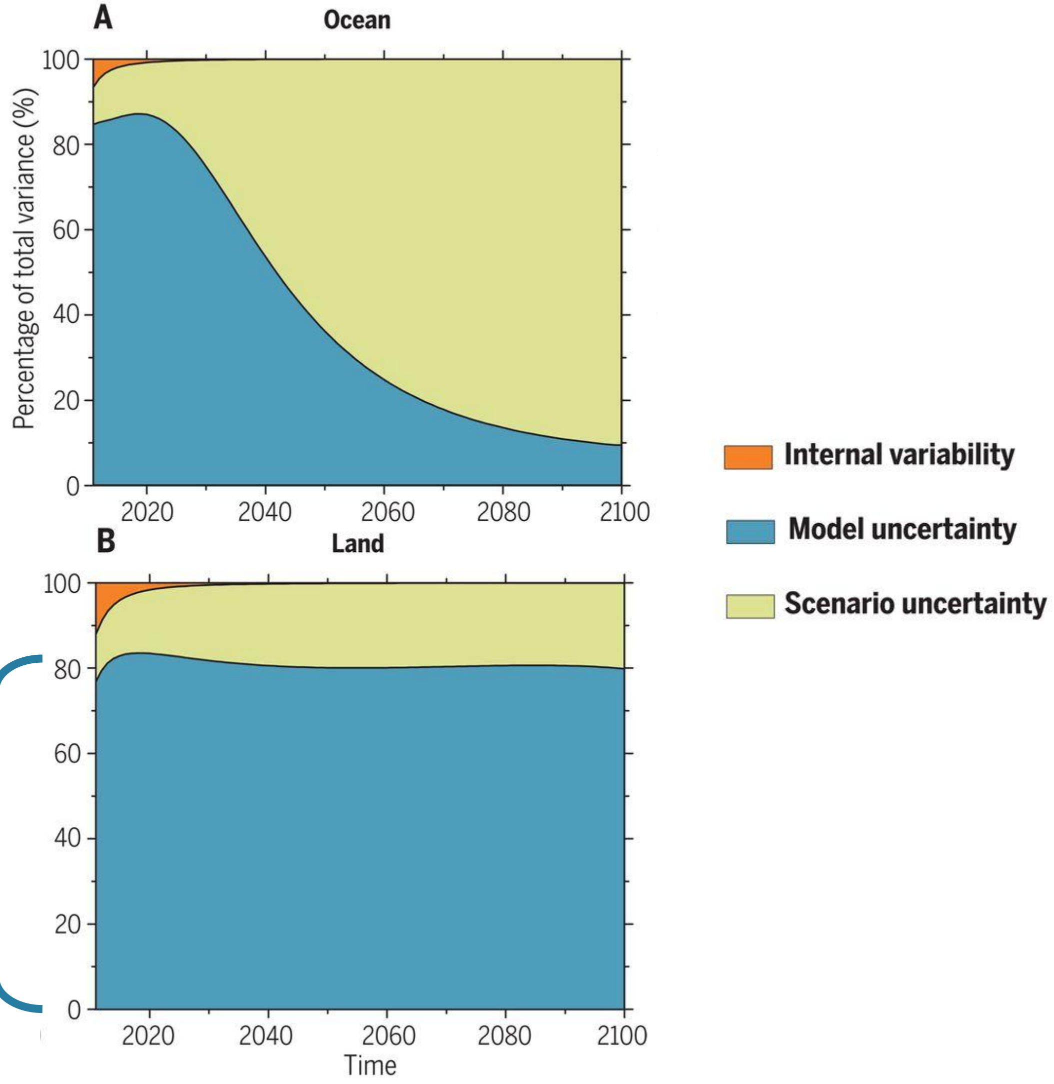
Arora et al. (2020)

Next generation models agree on the sign, but feature roughly **equivalent spread** (~1000 PgC)

Carbon Cycle Uncertainty in Land Model Projections

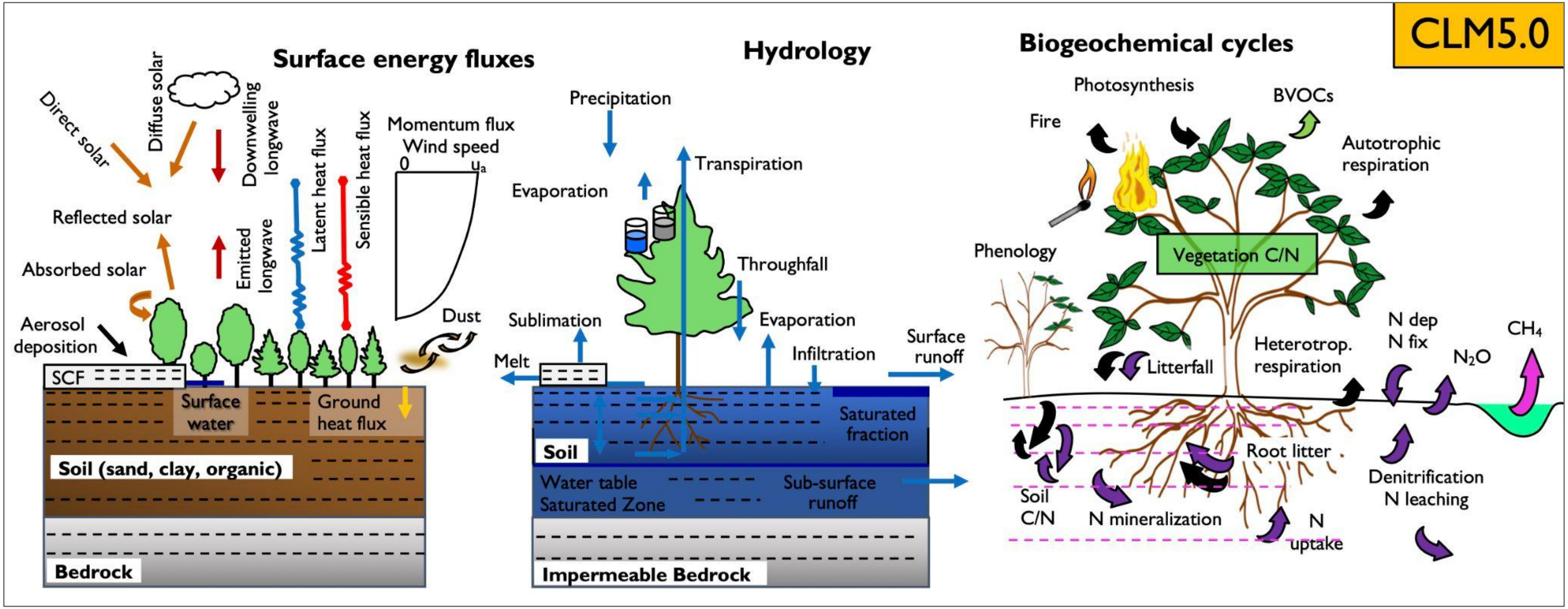


Uncertainty in land model structure and **parameters**



Bonan and Doney (2018), based on Lovenduski and Bonan (2017)

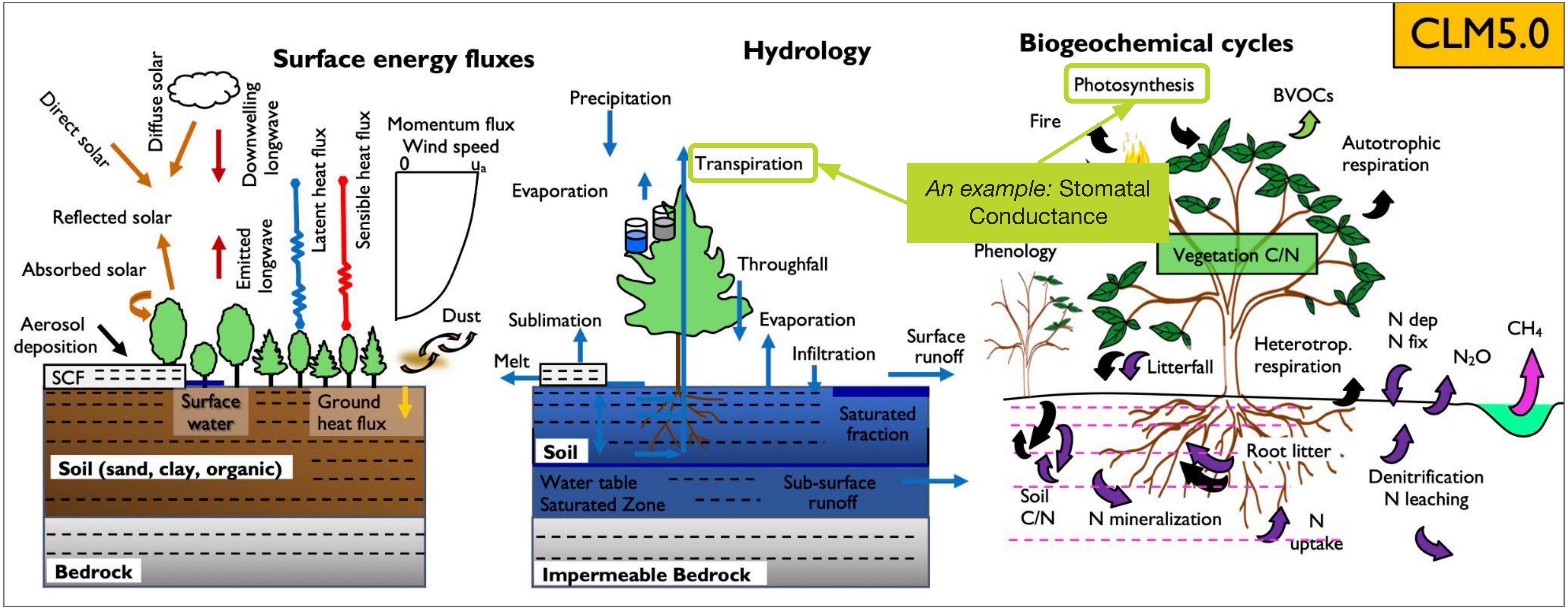
Uncertainty in Land Model Parameters



Schematic of the Community Land Model (CLM), version 5

Lawrence et al. (2019)

Uncertainty in Land Model Parameters



Schematic of the Community Land Model (CLM), version 5

Lawrence et al. (2019)

Example of Parameter Uncertainty: Stomatal Conductance

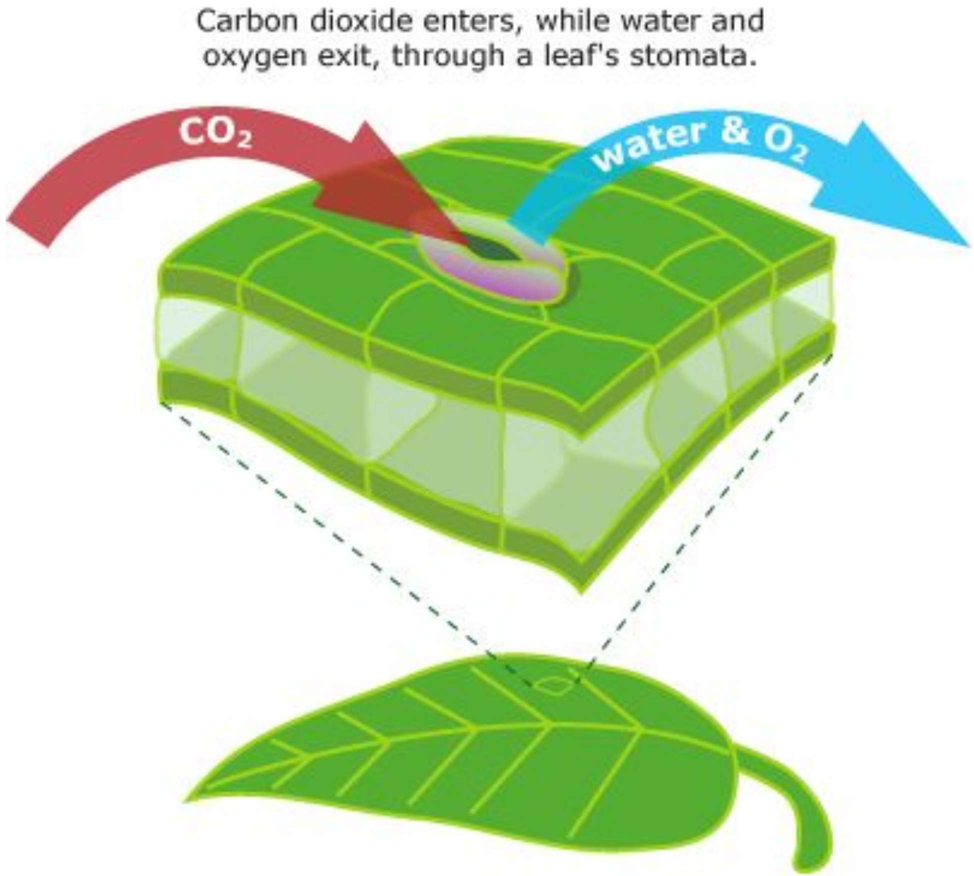
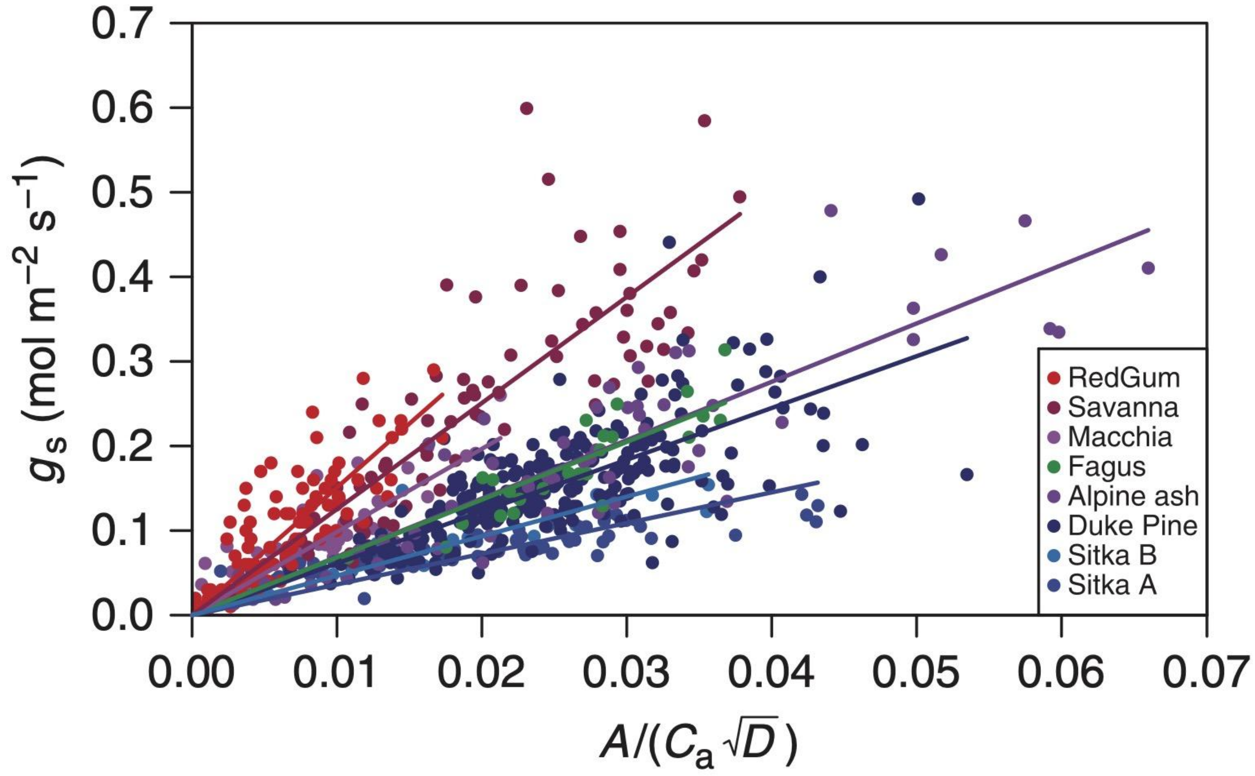


Image: evolution.berkeley.edu



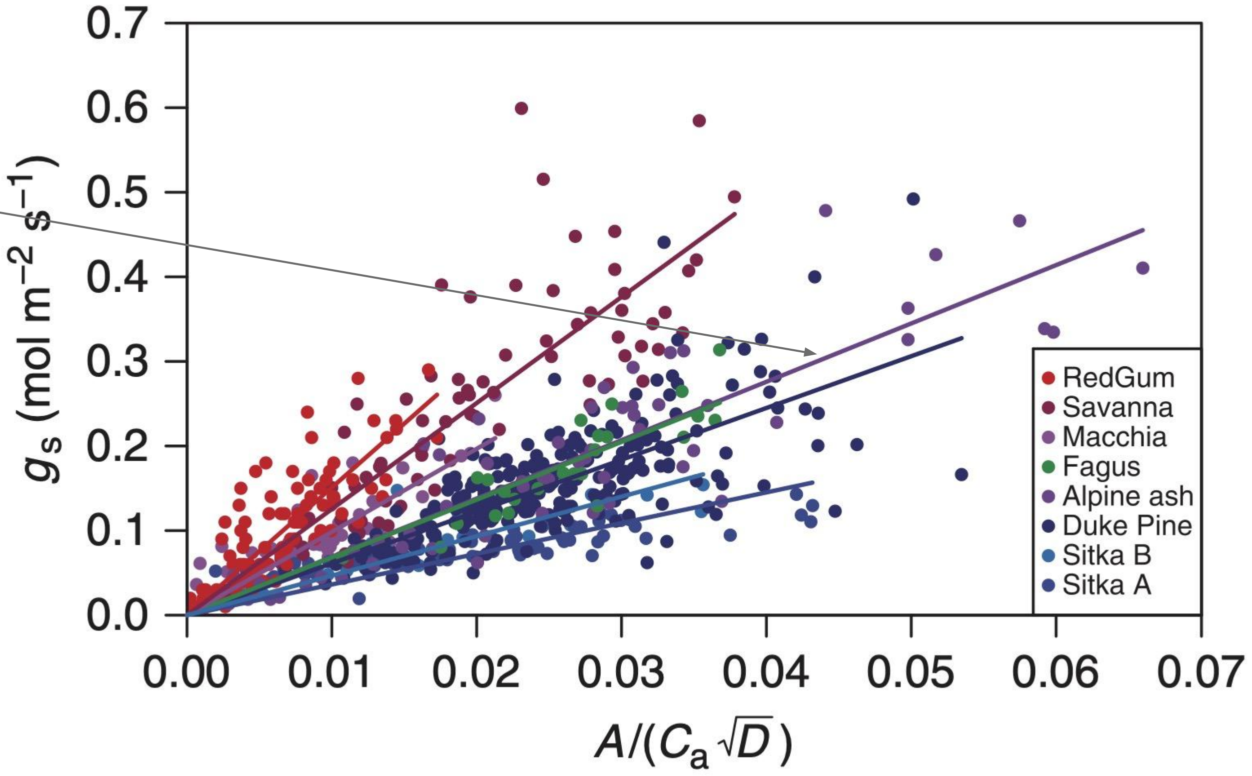
Medlyn et al. (2011)

Example of Parameter Uncertainty: Stomatal Conductance

Slope parameter represents **marginal water cost of carbon gain** and is an important model parameter.

$$g_s = g_o + 1.6\left(1 + \frac{g_1}{\sqrt{D}}\right)\frac{A_n}{c_s/P_{atm}}$$

g_1 = slope parameter
(mol H_2O/mol CO_2)



Medlyn et al. (2011)

Land Model Parameter Calibration

Hand-tuning parameter values takes a long time (many model runs, trial and error).

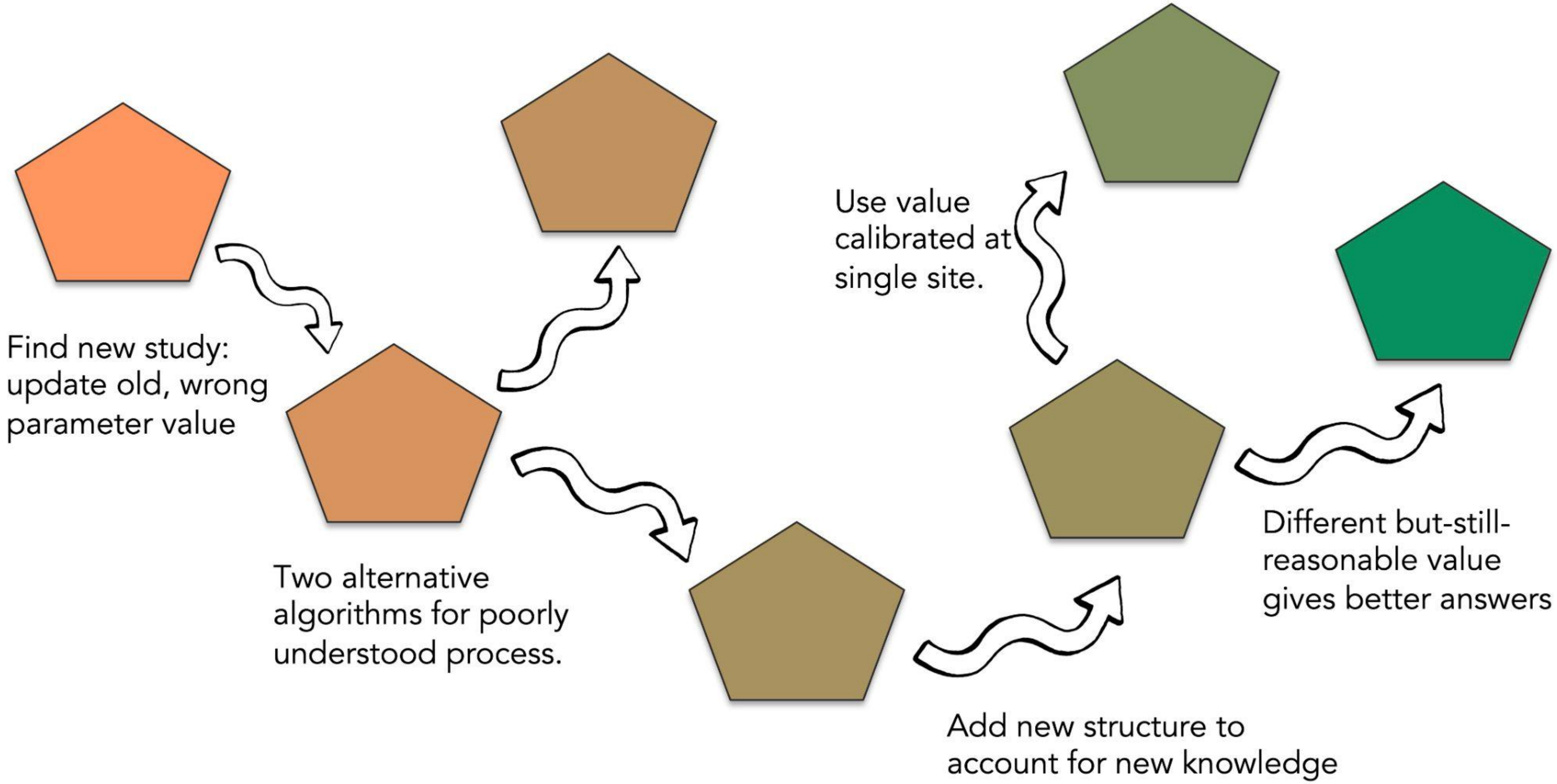


Figure from Rosie Fisher

Model Calibration, or Model Tuning

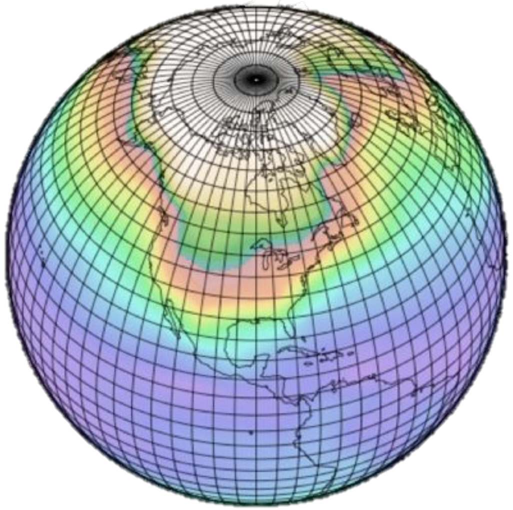
PNAS

PERSPECTIVE

Are general circulation models obsolete?

V. Balaji[a,b,1], Fleur Couvreux[c], Julie Deshayes[d], Jacques Gautrais[e], Frédéric Hourdin[f], and Catherine Rio[c]

*"Thus, model calibration is not a weakness of models; it, in fact, **holds the key to how model developers learn how their model behaves**, and, consequently, how the Earth system regulates itself."*
Balaji et al., 2022 [PNAS]



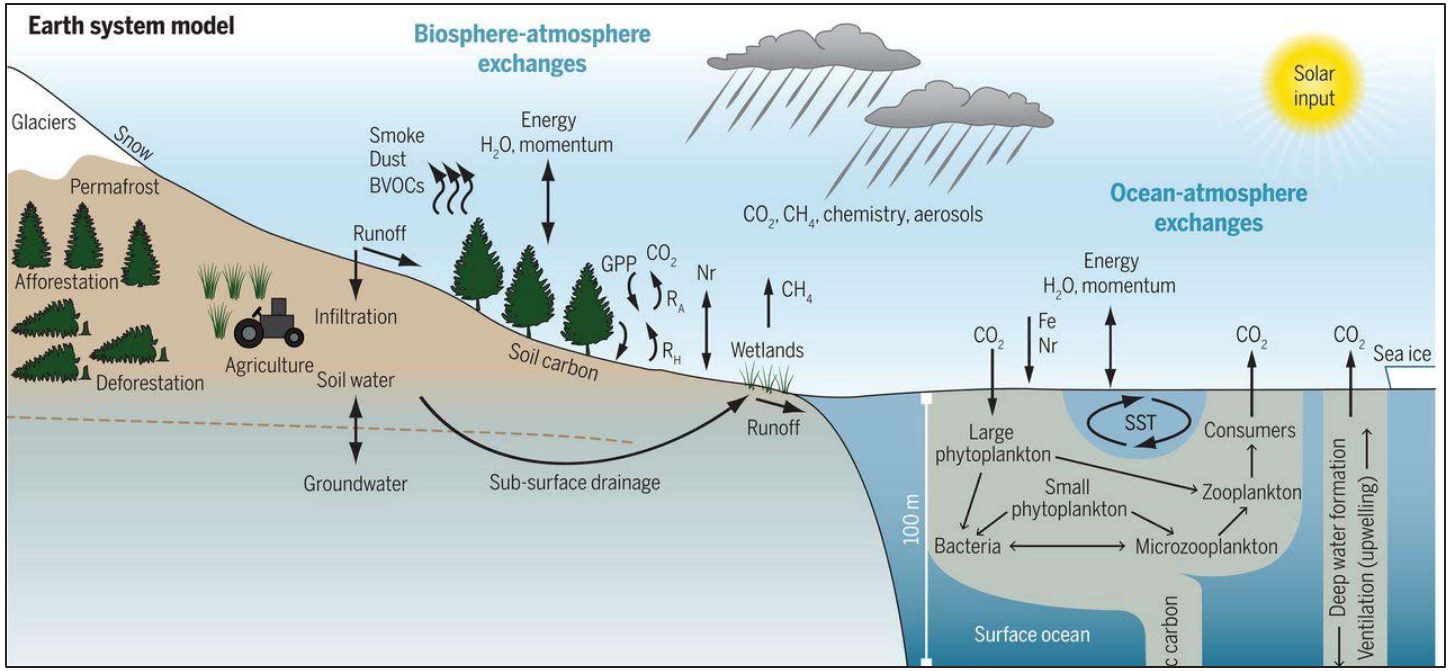
THE ART AND SCIENCE OF CLIMATE MODEL TUNING

Frédéric Hourdin, Thorsten Mauritsen, Andrew Gettelman, Jean-Christophe Golaz, Venkatramani Balaji, Qingyun Duan, Doris Folini, Duoying Ji, Daniel Klocke, Yun Qian, Florian Rauser, Catherine Rio, Lorenzo Tomassini, Masahiro Watanabe, and Daniel Williamson

*"As in art, there is also some **diversity and subjectivity in the tuning process** because of the complexity of the climate system and because of the choices made among the equally possible representations of the system."*
Hourdin et al., 2017 [BAMS]

Barriers to Systematic Model Calibration

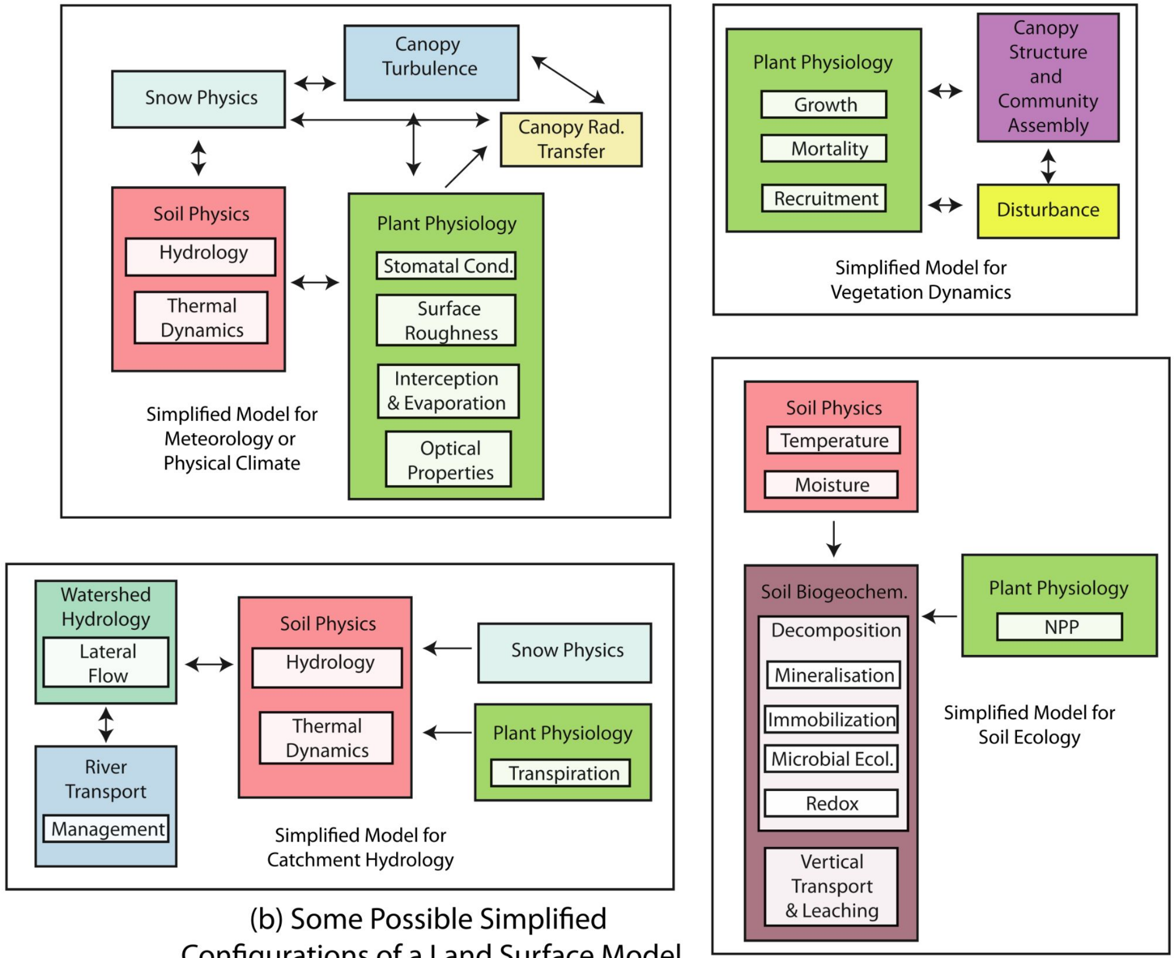
- *Interacting process complexity*
- *Large and poorly constrained parameter spaces*
- *Prohibitive computational costs (e.g., spin-up and model ensembles)*
- *Uncertainty in observations and model structure*
- *Challenges in removing subjectivity from model tuning*



Bonan and Doney (2018)

Benefits of Systematic Model Calibration

- ***Aiding model development:*** *testing new parameterizations, different resolutions, component coupling*
- ***Maintaining a hierarchy of model complexity*** *(e.g., CLM-SP, BGC, FATES)*
- ***Actionable science:*** *providing an easily deployable tool for different study domains*
- ***Increasing the model community*** *and user base*

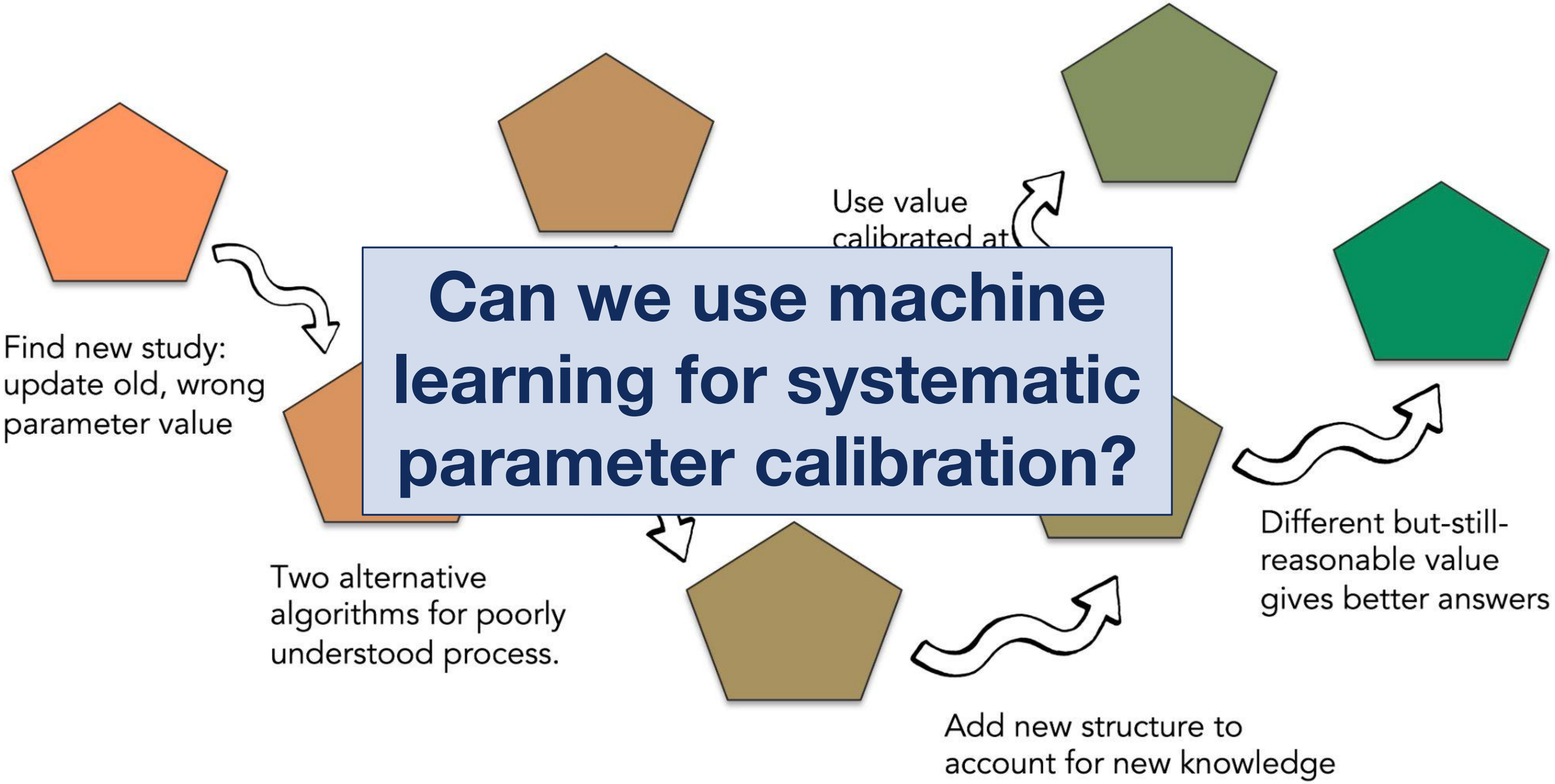


(b) Some Possible Simplified Configurations of a Land Surface Model

Fisher and Koven (2020)

Land Model Parameter Calibration

Hand-tuning parameter values takes a long time (many model runs, trial and error).



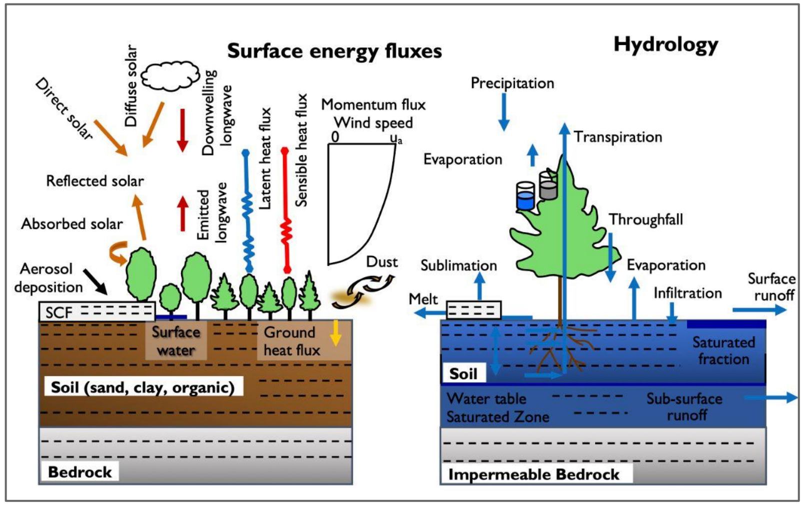
Can we use machine learning for systematic parameter calibration?

Figure from Rosie Fisher

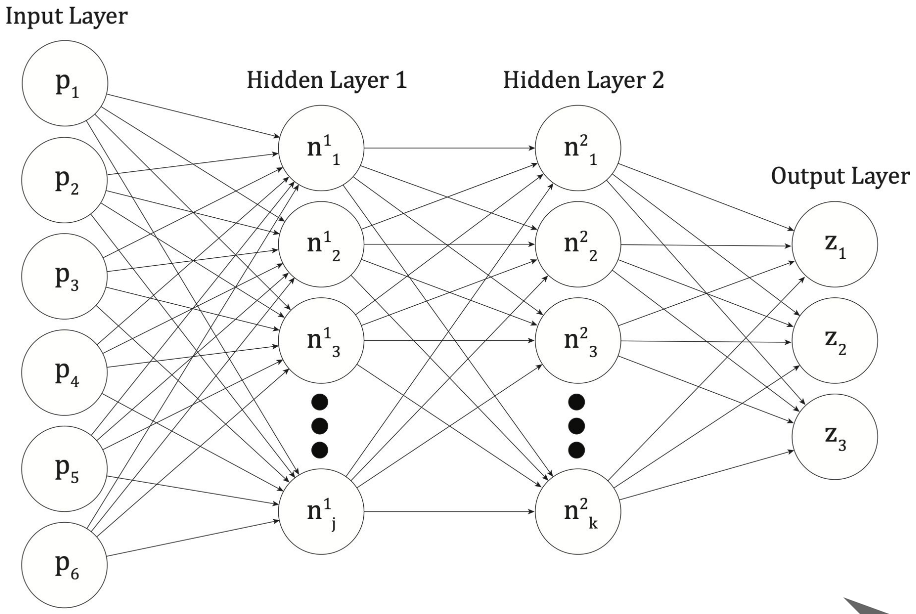
Machine Learning for Land Model Emulation

Training Mode

Input: **six** land model parameter values focused on **biophysical processes**

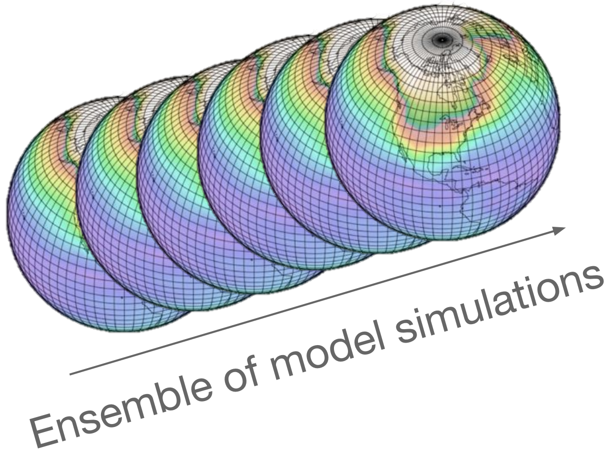


Neural network emulator



Learning the parameter response functions

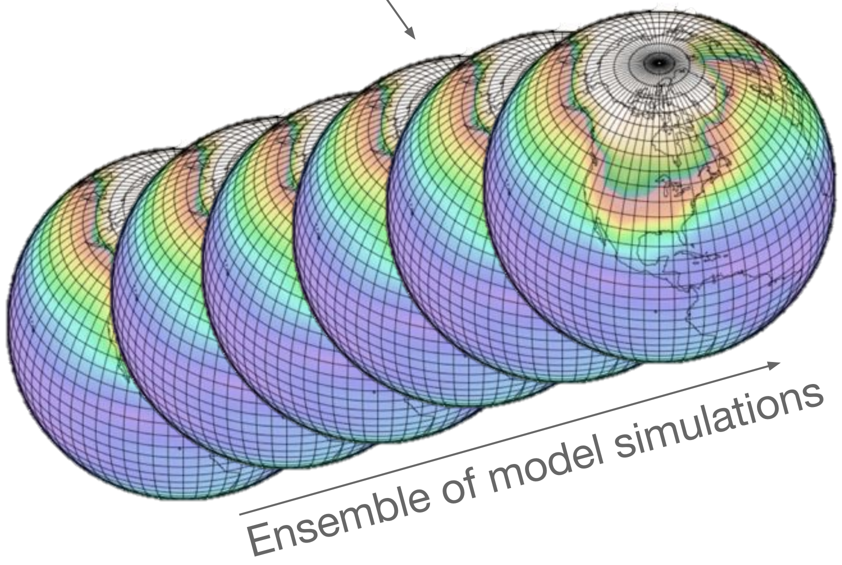
Output: variability in carbon and water fluxes using a **perturbed parameter ensemble**



Dagon et al. (2020)

Land Model Perturbed Parameter Ensemble

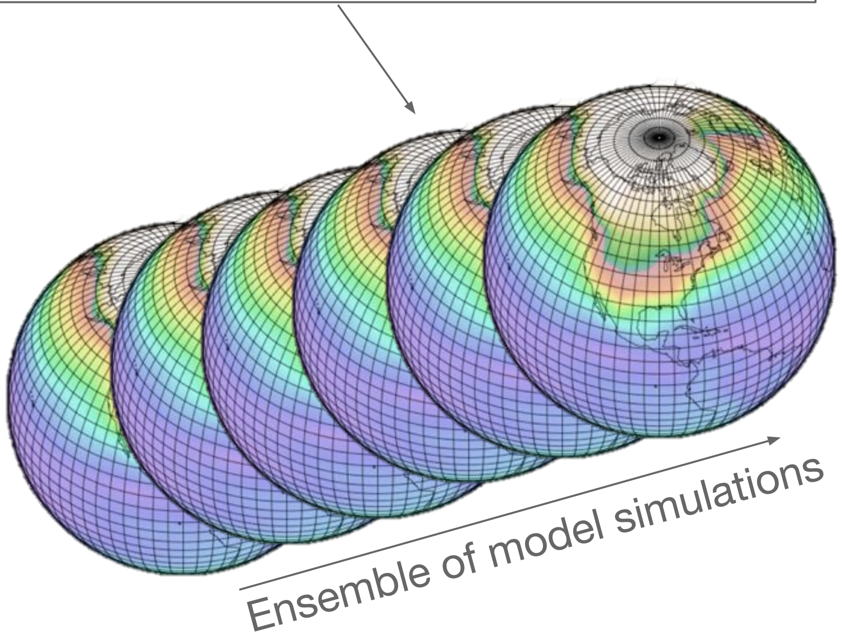
Land model* perturbed parameter ensemble (PPE) using 100 parameter combinations generated with Latin Hypercube sampling

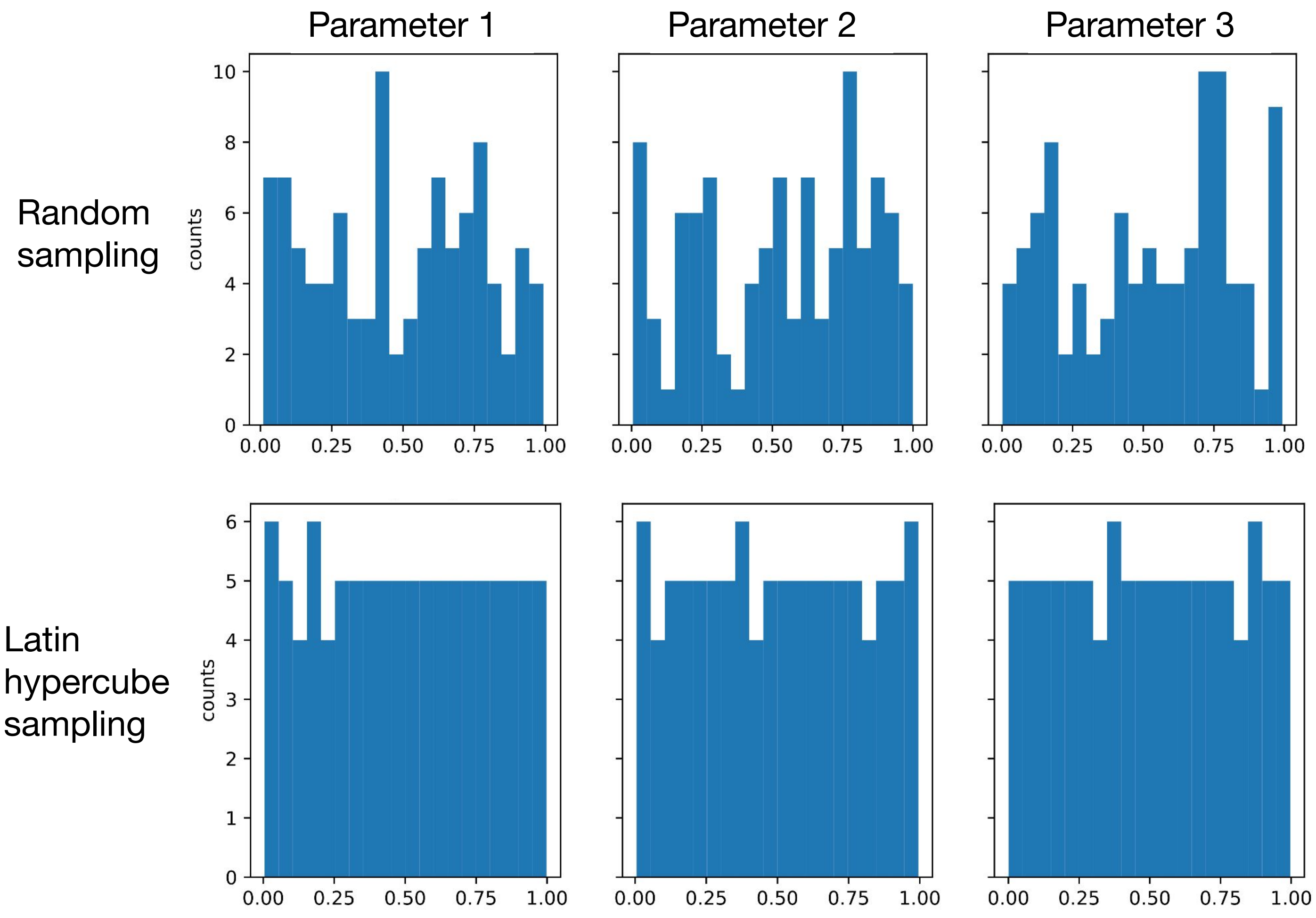


*Offline global land-only (CLM5SP) simulations forced by atmospheric reanalysis data

Land Model Perturbed Parameter Ensemble

Land model* perturbed parameter ensemble (PPE) using 100 parameter combinations generated with Latin Hypercube sampling

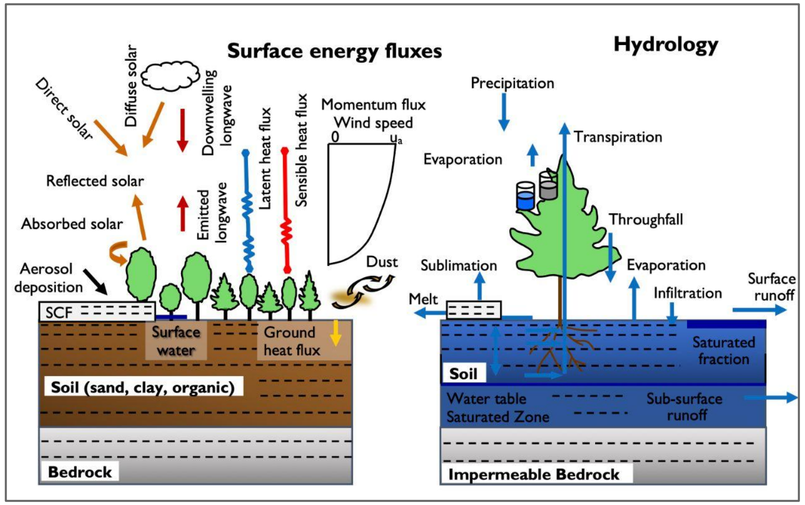




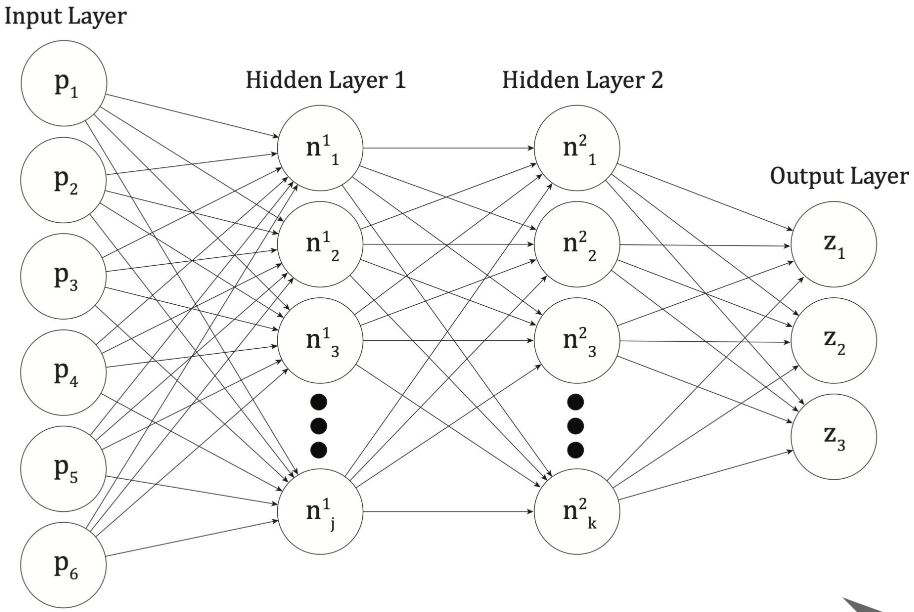
*Offline global land-only (CLM5SP) simulations forced by atmospheric reanalysis data

Inference Mode

Input: **NEW** land model parameter values unseen by the emulator

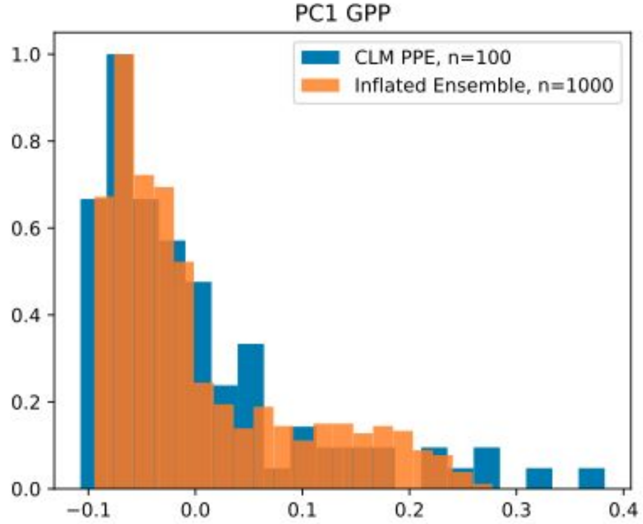


Trained neural network emulator



Inferring the parameter response

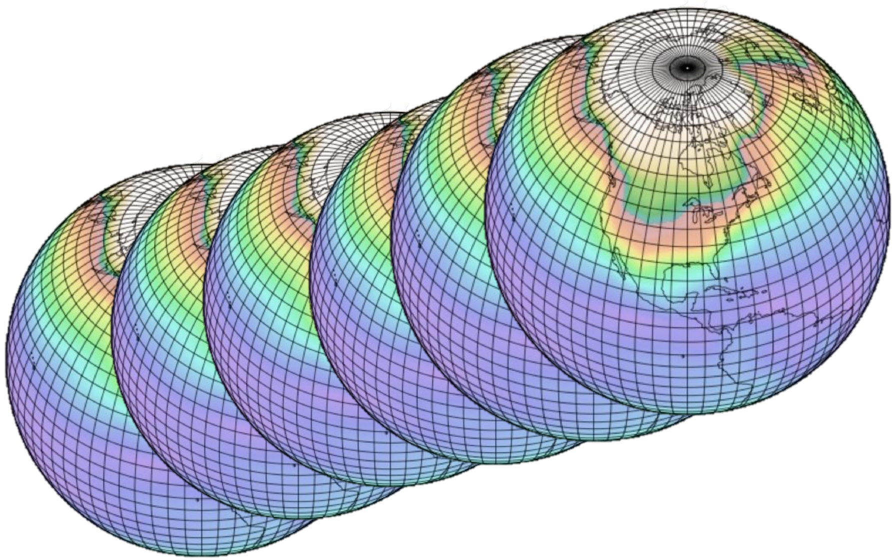
Output: **predictions** of variability in carbon and water fluxes



Dagon et al. (2020)

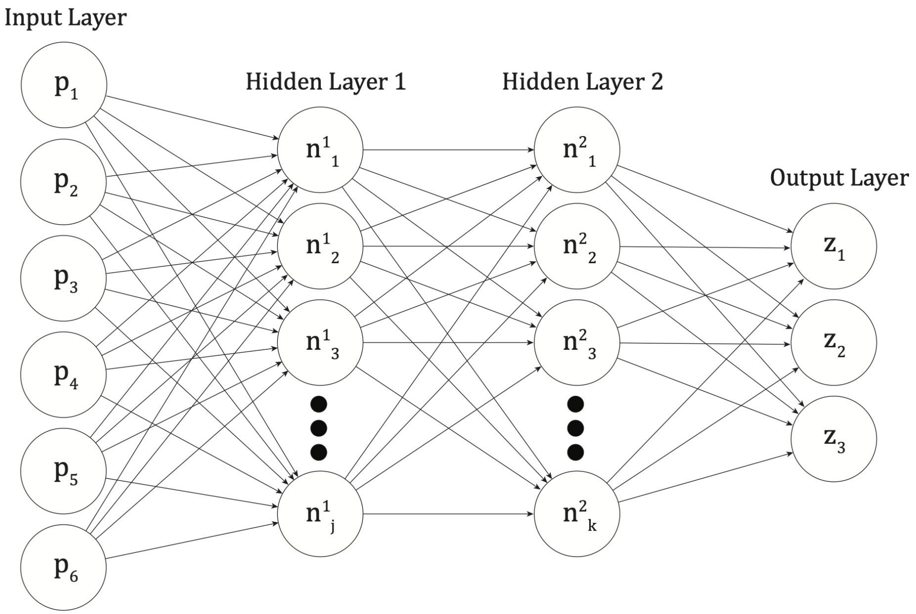
Increase in Computational Efficiency

Land model perturbed parameter ensemble



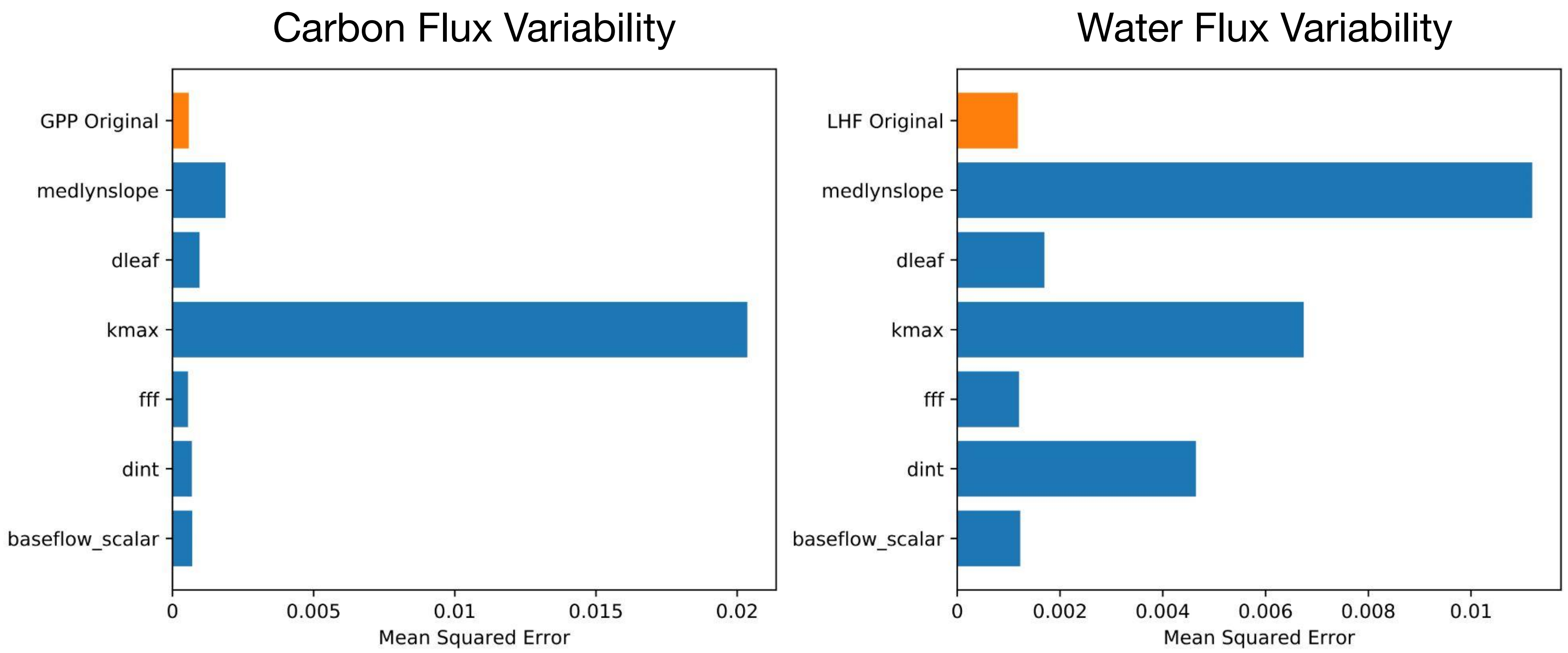
~2 hours per simulation

Neural network emulator



2.6 seconds to generate predictions!

Interpretability Sheds Light on Physics

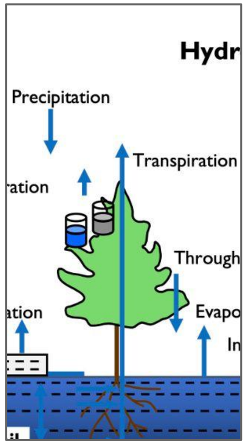


Variable Importance

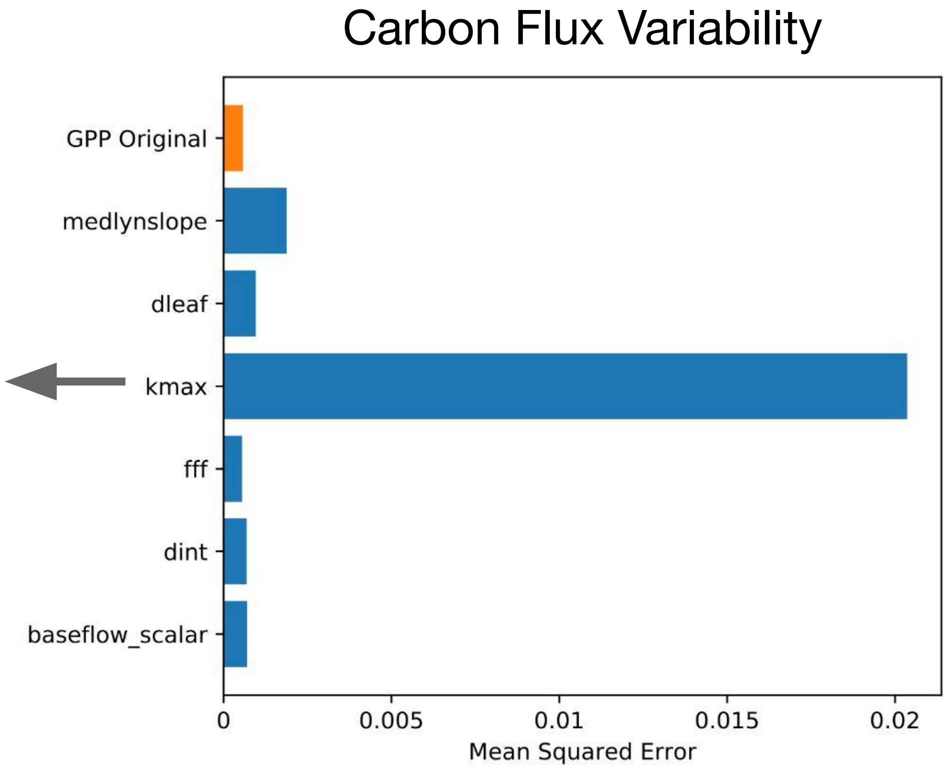
- Randomly shuffle values of one parameter (preserving others) and test performance of emulator.
- Skill metric is mean squared error between predictions and actual values.
- Larger bar means the parameter is **more important to the predictive skill** of the emulator.

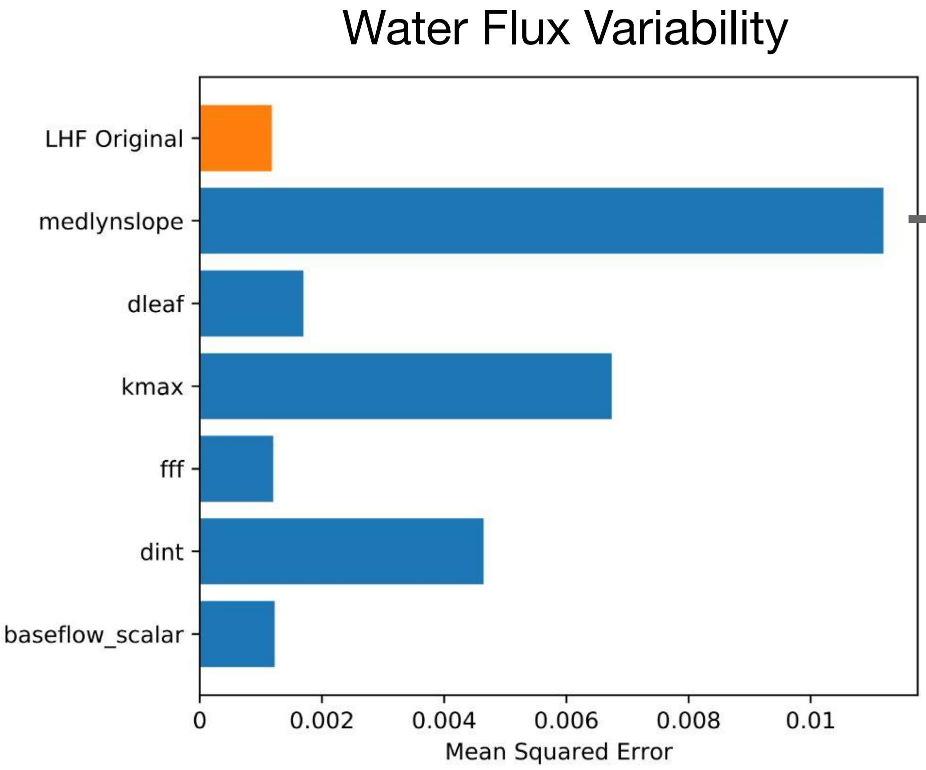
Dagon et al. (2020)

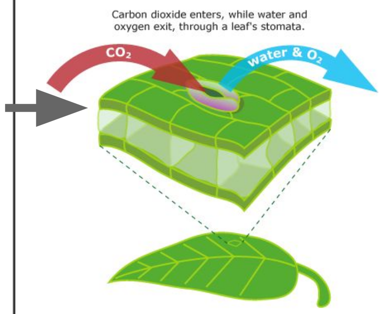
Interpretability Sheds Light on Physics



Kennedy et al. (2019)







Medlyn et al. (2011)

Variable Importance

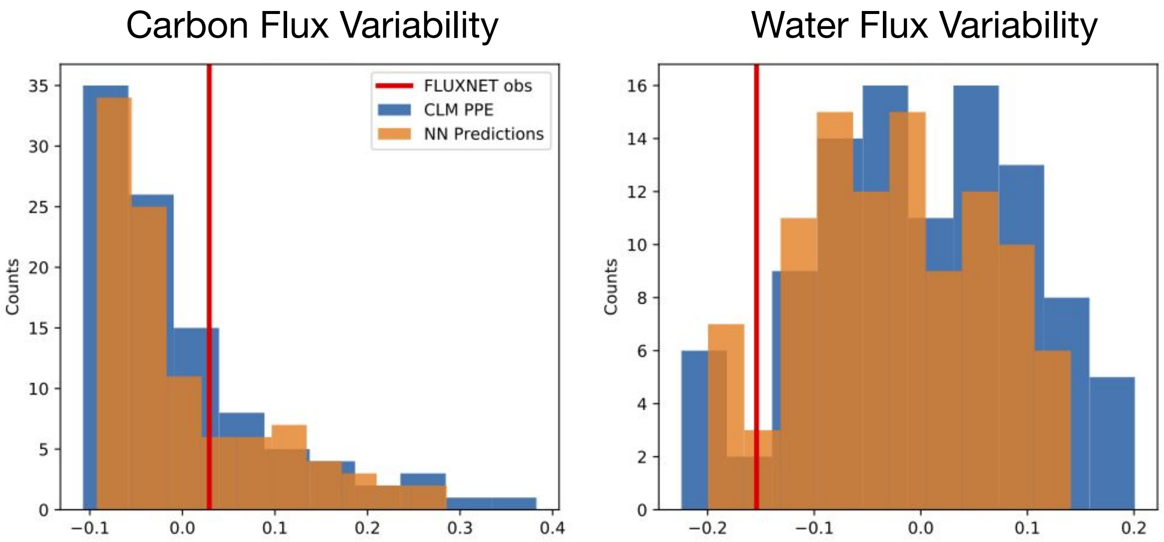
- Randomly shuffle values of one parameter (preserving others) and test performance of emulator.
- Skill metric is mean squared error between predictions and actual values.
- Larger bar means the parameter is **more important to the predictive skill** of the emulator.

Dagon et al. (2020)

Optimize Emulator Predictions for Calibration

Calibration

- Use a cost function to minimize error in emulator predictions relative to observations.
- Find some “best fit” parameter values.



$$J(p) = \sum_{v=1}^{2} \left[\sum_{m=1}^{3} \lambda_{v,m} \left(\frac{\hat{U}_{v,m}(p) - U_{obs,v,m}}{\sigma(U_{obs*,v,m})} \right)^2 \right]$$

Sum over output variables *v*

Sum over modes *m* for each term, weighting by % variance

Emulator predictions for parameters *p*

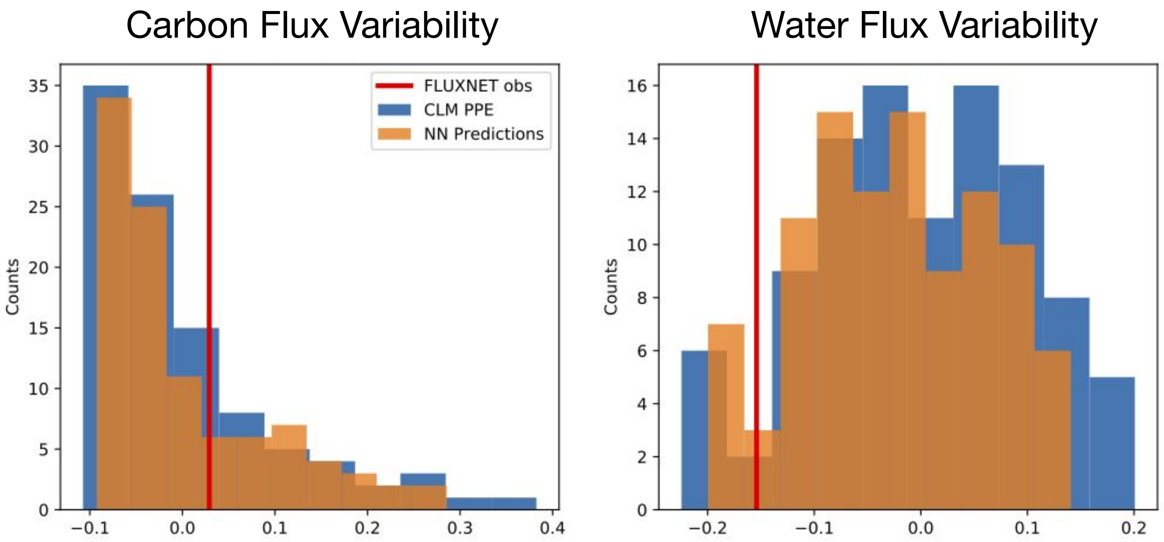
Normalize by standard deviation in observations

Observations

Dagon et al. (2020)

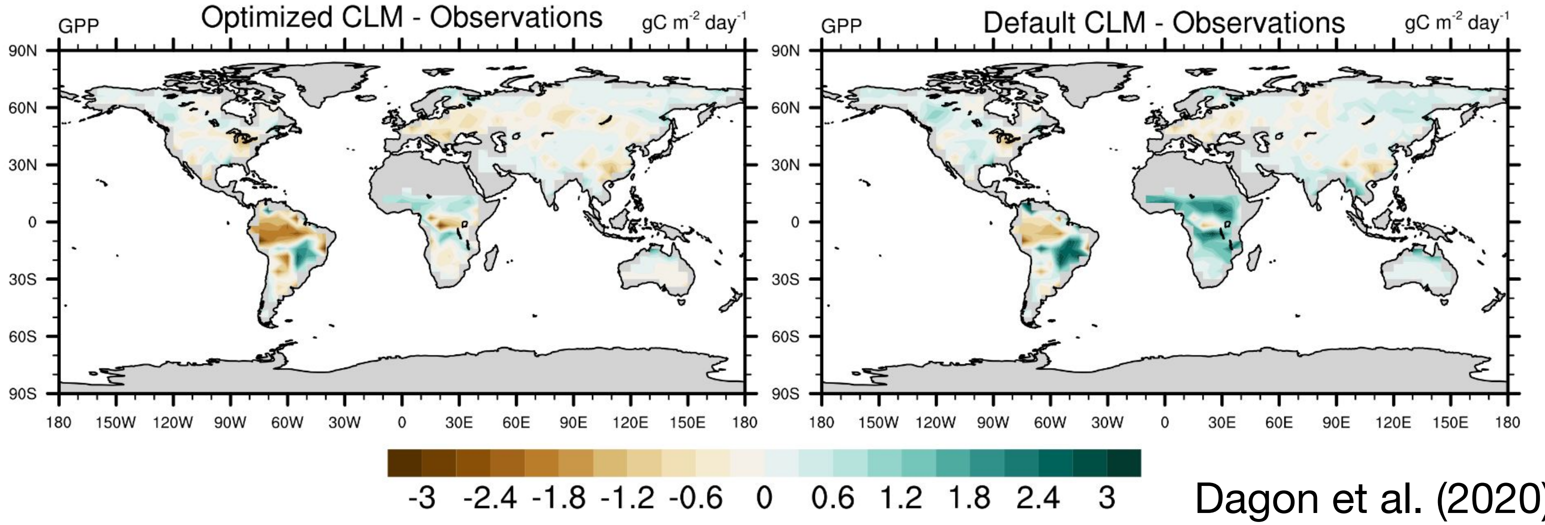
Calibration

- Use a cost function to minimize error in emulator predictions relative to observations.
- Find some "best fit" parameter values.



Testing

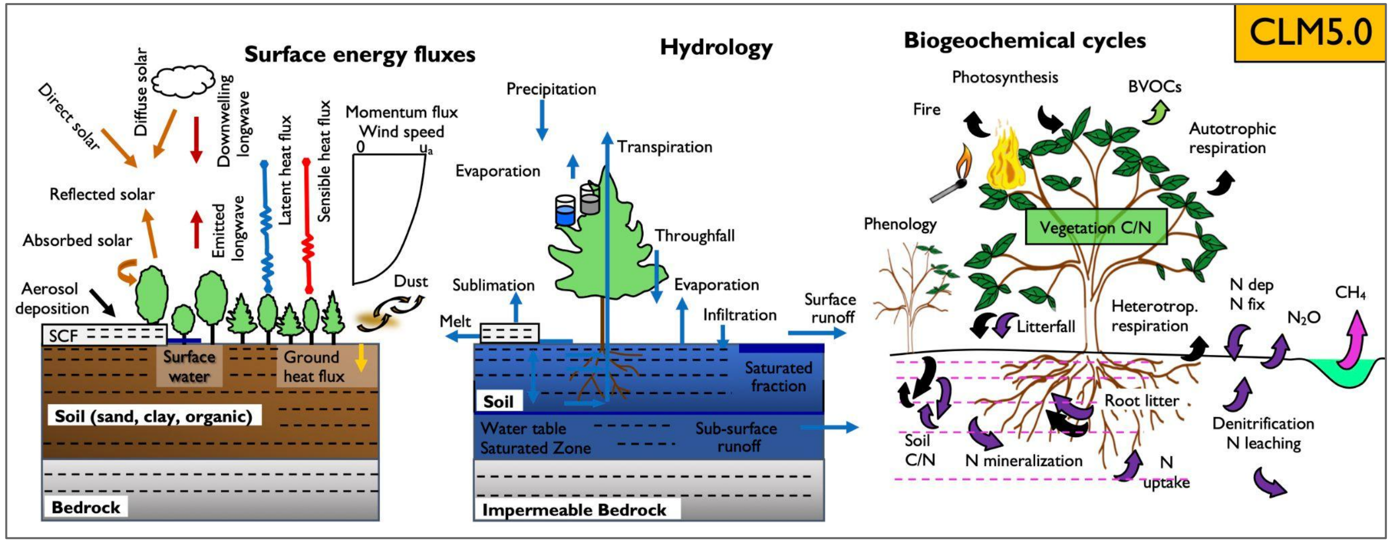
- Test optimized predictions in the full global land model.
- Improvement in global, annual mean biases; regional/seasonal results mixed- **what are we missing?**



Dagon et al. (2020)

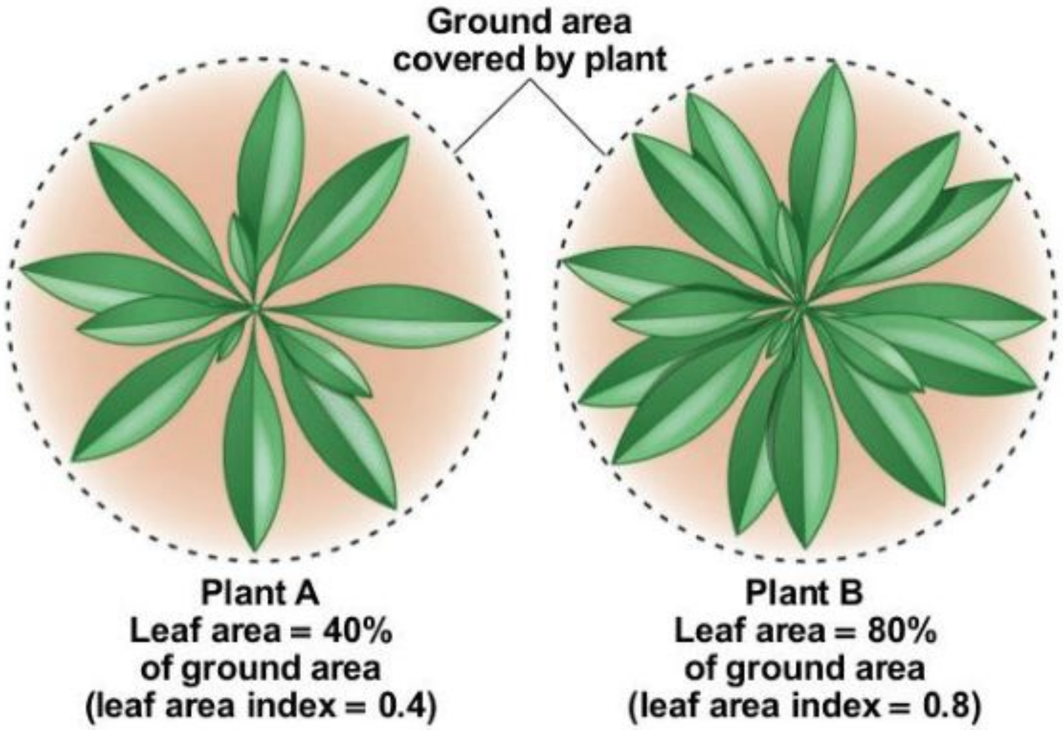
What are we missing?

- *Limited complexity:* CLM with satellite phenology (SP), not full biogeochemistry (BGC)
- *Minimal sampling:* 6 parameters, 100 ensemble members
- *Choice of metrics:* annual mean spatial variability
- *Choice of cost function:* accounting for additional sources of uncertainty
- *Choice of calibration targets:* carbon and water fluxes



Leaf Area Index (LAI)

The amount of leaf area per unit gridcell area



Pearson Education

Leaf Area Index (LAI)

The amount of leaf area per unit gridcell area

Closed canopy forest: LAI~4 m^2/m^2

USGS

Leaf Area Index (LAI)

The amount of leaf area per unit gridcell area

Closed canopy forest:	LAI~4 m^2/m^2
Grassland:	LAI~2 m^2/m^2
Global average:	LAI~1 m^2/m^2

USGS

Leaf area index correlated with many environmental variables

Photosynthesis
Carbon uptake

Fire emissions

Leaf area index correlated with many environmental variables

Photosynthesis
Carbon uptake

Fire emissions

Latent heat flux

Albedo

Leaf area index correlated with many environmental variables

Drought
Temperature
Human management
CO_2 concentration

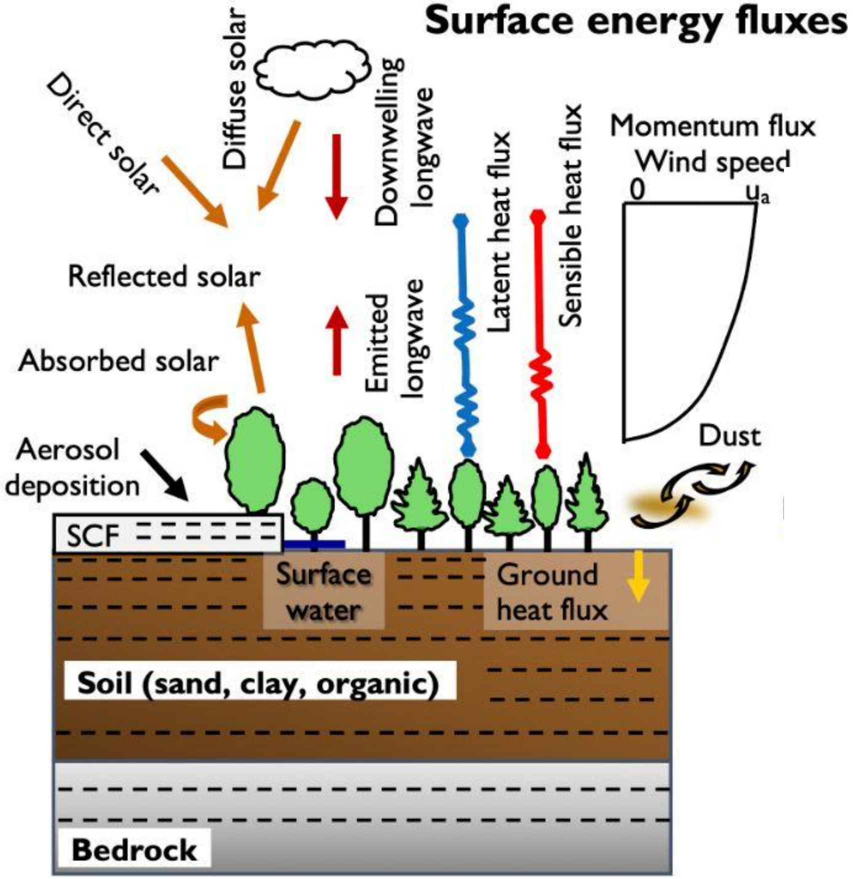
Photosynthesis
Carbon uptake

Fire emissions

Latent heat flux

Albedo

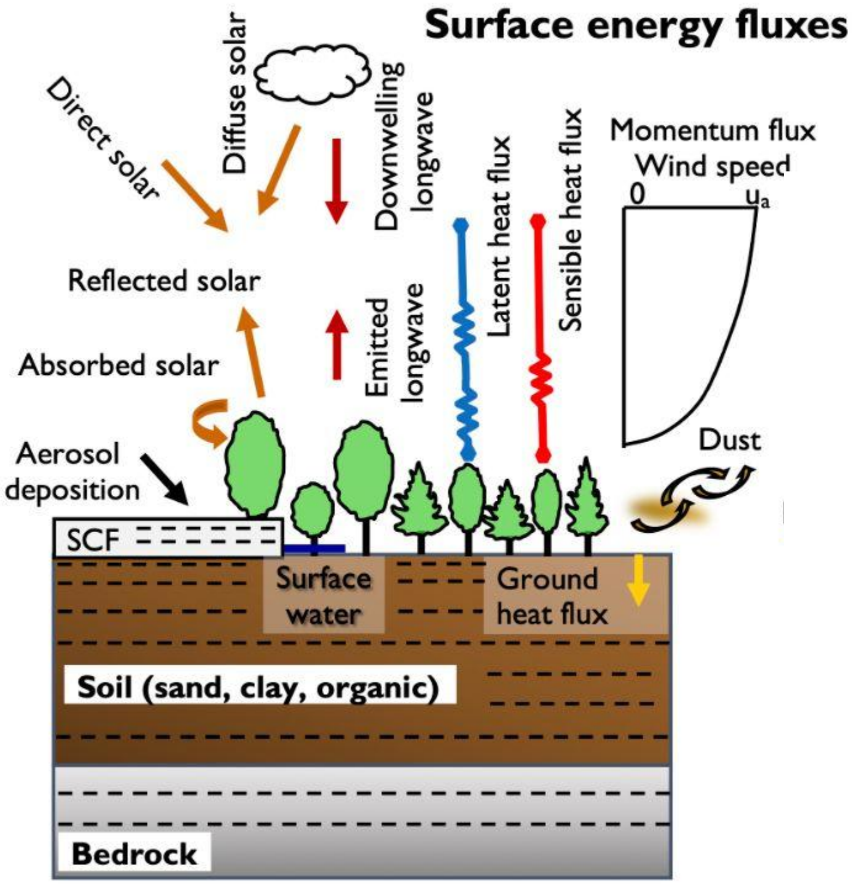
CLM can use prescribed or prognostic LAI



Satellite Phenology Mode

- leaf area prescribed from observations
- only worry about the biophysical modeling

CLM can use prescribed or prognostic LAI



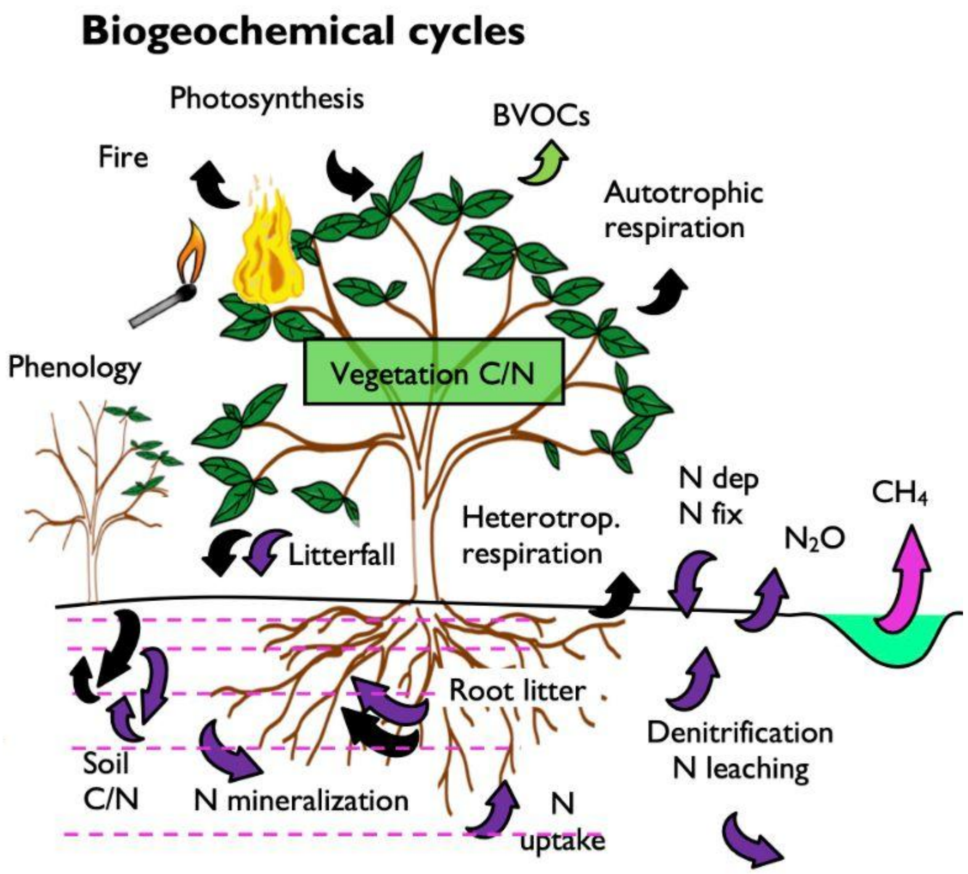
Satellite Phenology Mode

- leaf area prescribed from observations
- only worry about the biophysical modeling
- incomplete carbon cycle
- not suitable for SSP simulations

CLM can use prescribed or prognostic LAI

Biogeochemistry Mode

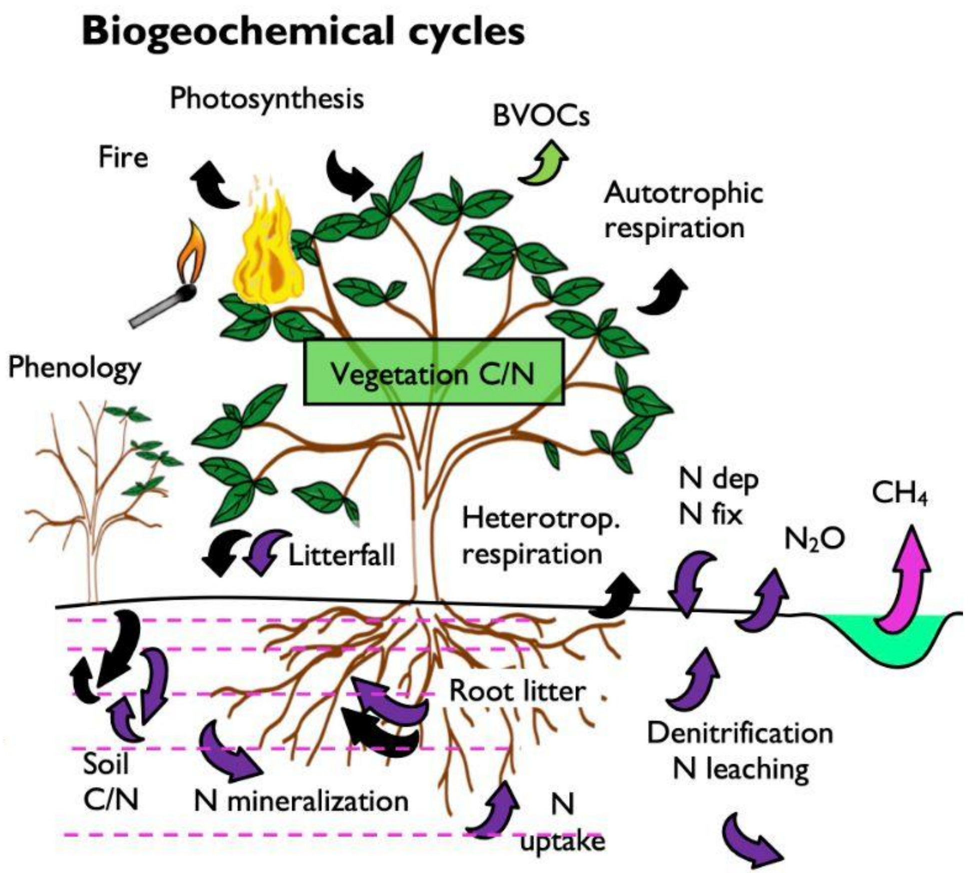
- prognostic leaf area
- full biogeochemistry



CLM can use prescribed or prognostic LAI

Biogeochemistry Mode

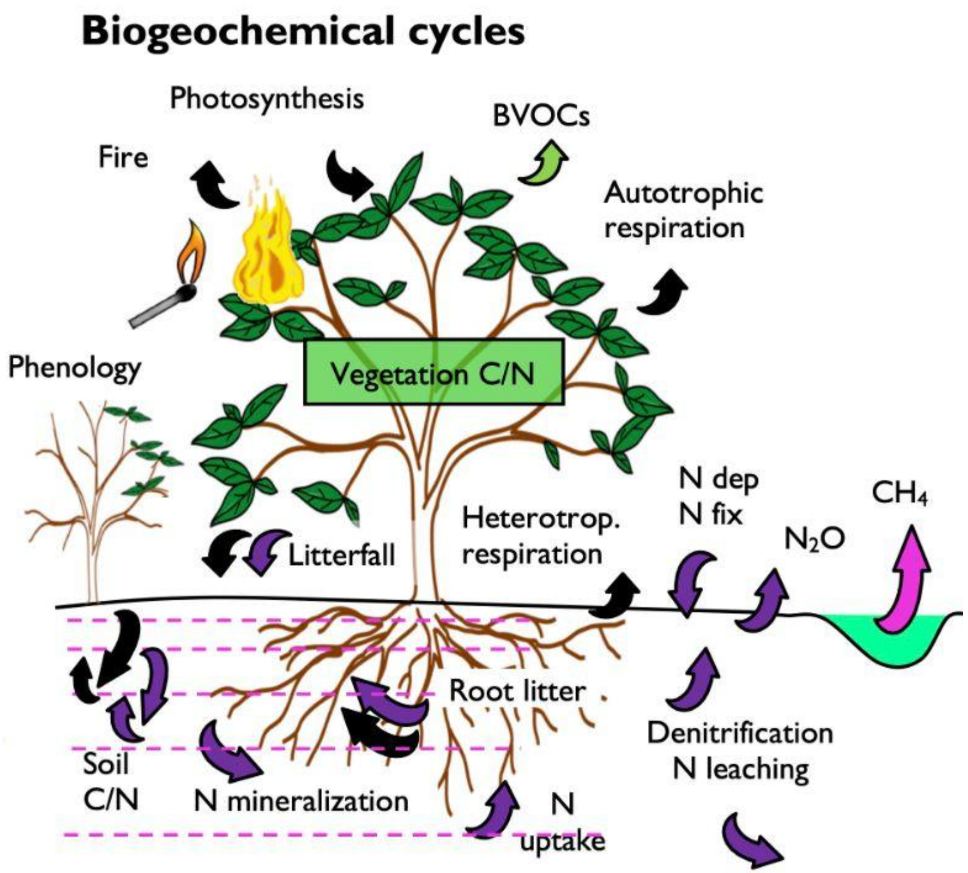
- prognostic leaf area
- full biogeochemistry
- 1.75x parameters (212 vs. 120)
- larger spinup burden
  - 1500 years vs. 20 years



CLM can use prescribed or prognostic LAI

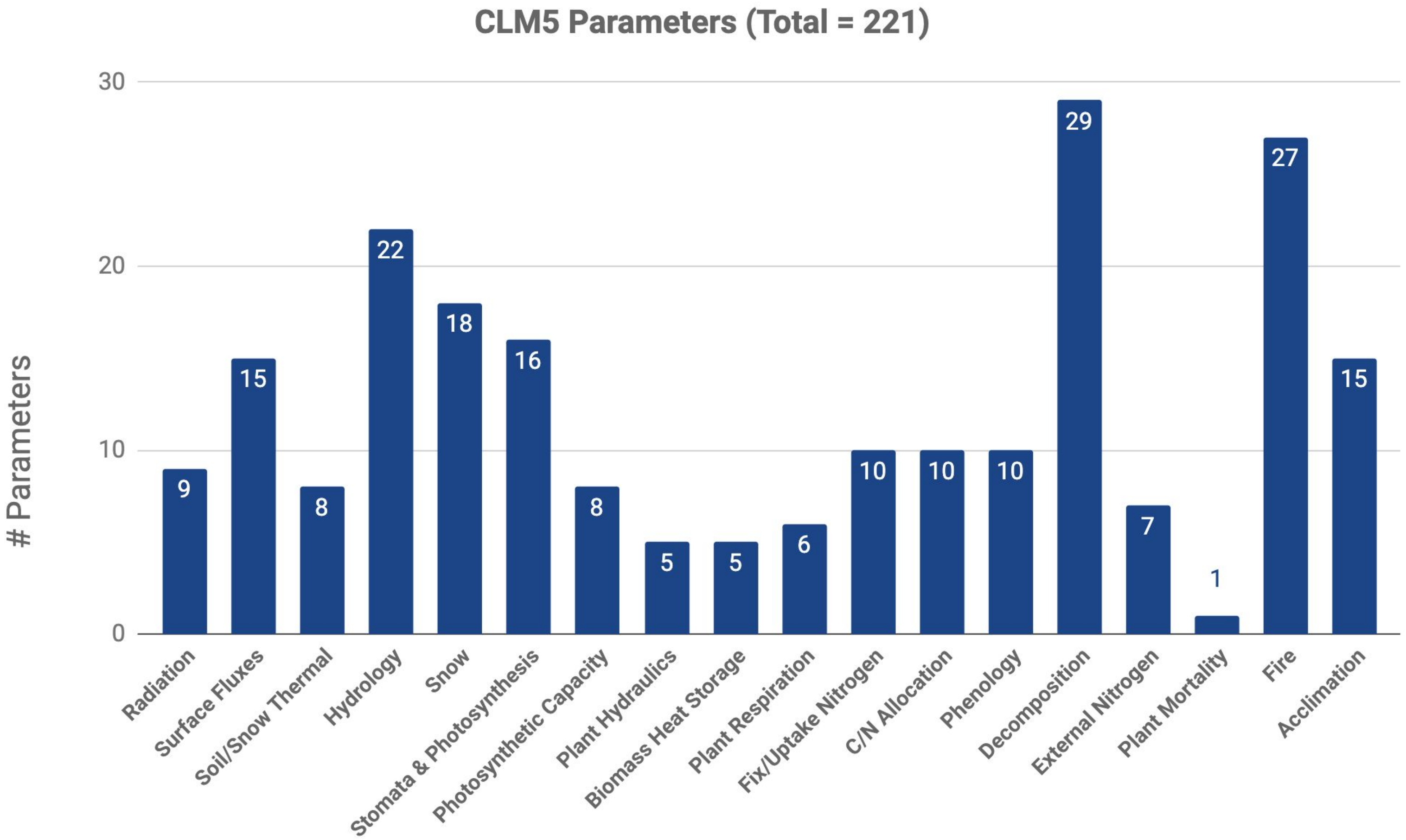
Biogeochemistry Mode

- prognostic leaf area
- full biogeochemistry
- 1.75x parameters (212 vs. 120)
- larger spinup burden
  - 1500 years vs. 20 years
- typically used within CESM



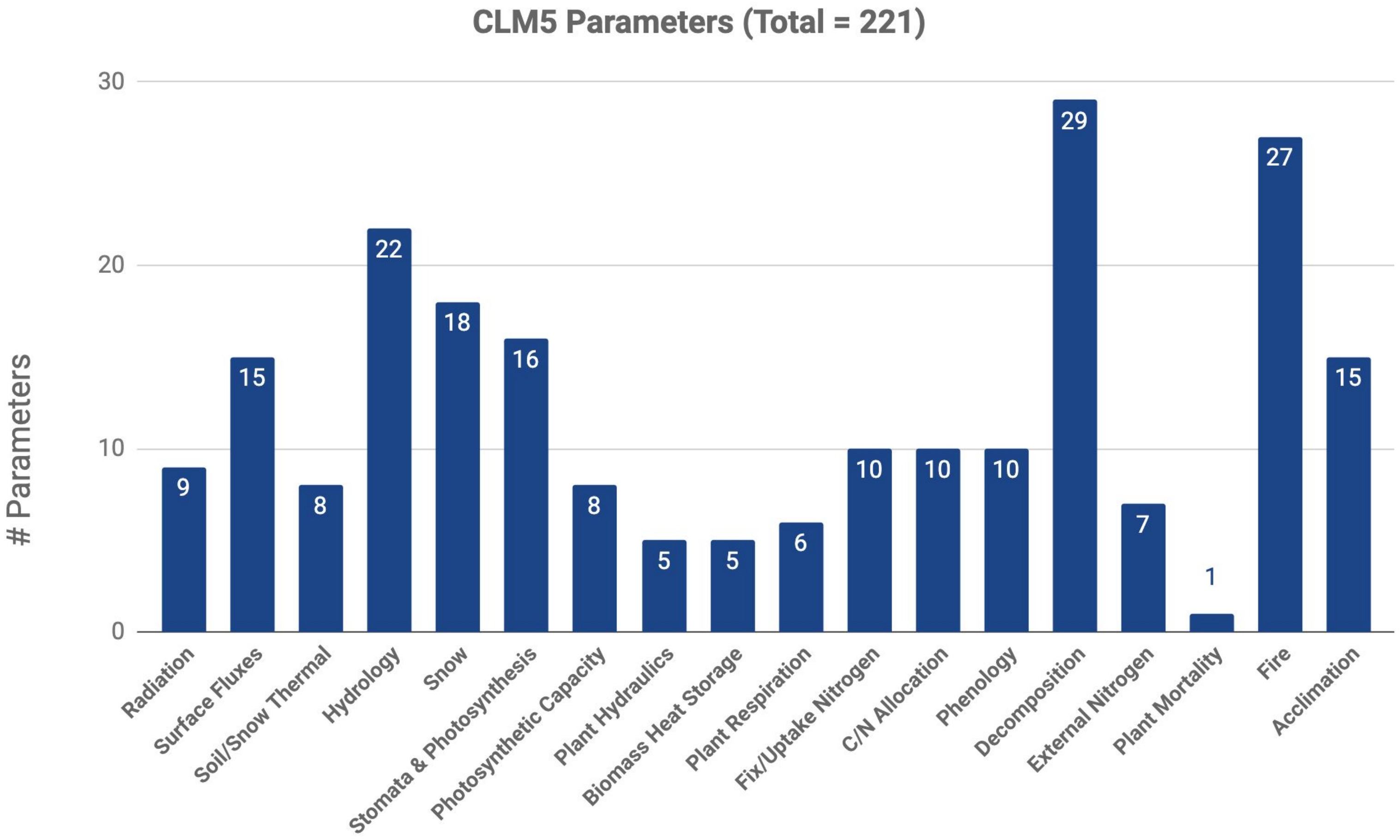
Land Model (Large) Perturbed Parameter Ensemble

Land model has over 200 parameters!
What if we sample all of them?



Land Model (Large) Perturbed Parameter Ensemble

Land model has over 200 parameters!
How do we sample all of them?



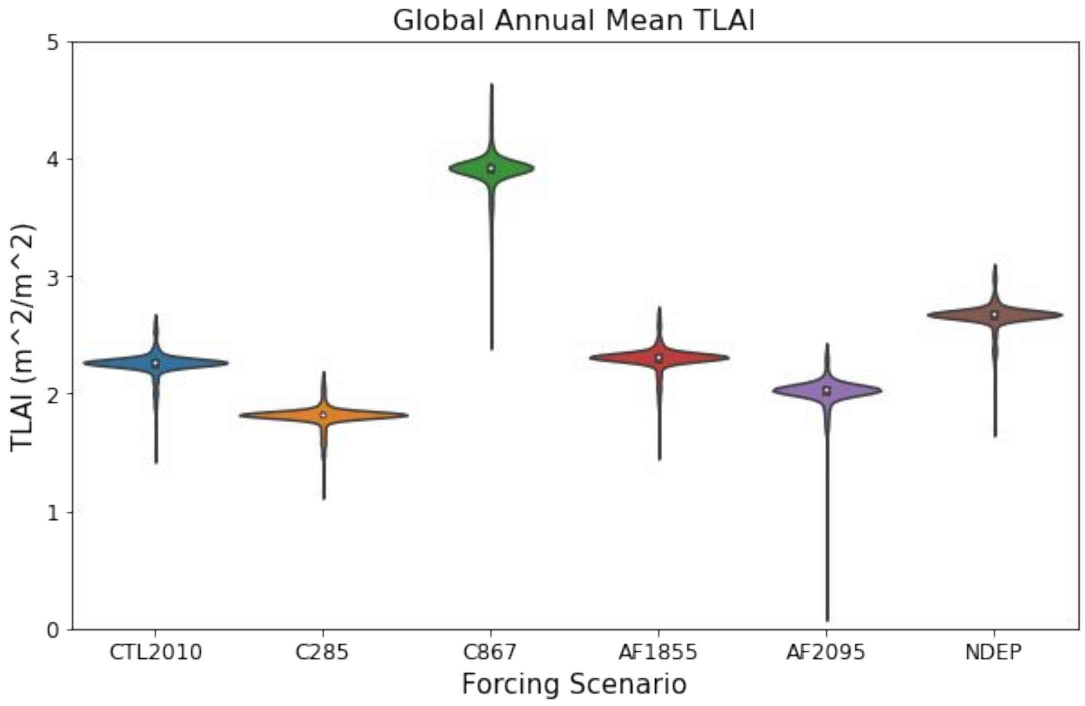
ROADMAP

1. **Identify and define parameter ranges** – what is a “parameter”?
2. **Infrastructure development:** parameter sampling, ensemble generation, computational efficiency

Land Model (Large) Perturbed Parameter Ensemble

Land model has over 200 parameters!
How do we sample different climate conditions?

Distributions of global, annual mean **total leaf area index (TLAI)** under various forcing scenarios.

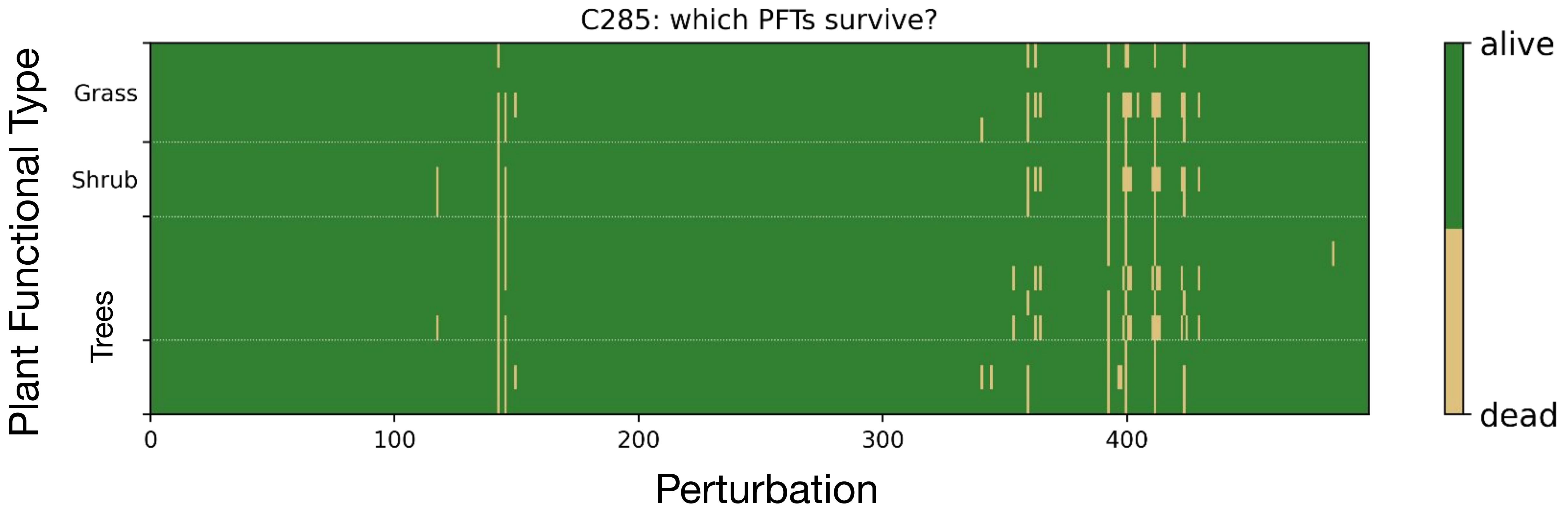


ROADMAP

1. **Identify and define parameter ranges** – what is a “parameter”?
2. **Infrastructure development:** parameter sampling, ensemble generation, computational efficiency
3. **Run one-at-a-time (OAAT) perturbations** with all 221 parameters, multiple forcing scenarios (e.g., low/high CO_2).

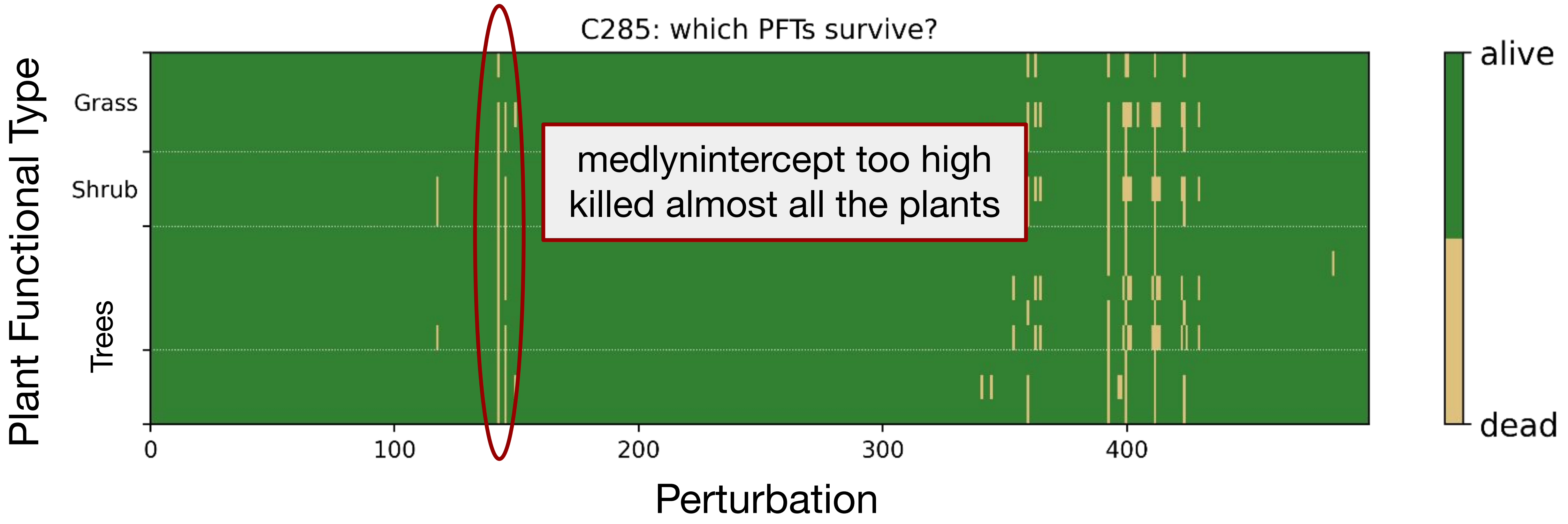
Can we avoid large-scale plant die off?

Parameter sets that look good in present-day
might not survive through pre-industrial conditions (low CO_2)



Can we avoid large-scale plant die off?

Parameter sets that look good in present-day
might not survive through pre-industrial conditions (low CO_2)



One-at-a-time ensemble (OAAT)

~200 parameters → **high value** / **low value** → **~400 simulations**

One-at-a-time ensemble (OAAT)

~200 parameters → high value / low value → ~2400 simulations

One-at-a-time ensemble (OAAT)

~200 parameters → high value / low value → ~2400 simulations

1° resolution

250 pe-hrs/yr * 10 yr simulation = 2,500 pe-hrs per simulation

One-at-a-time ensemble (OAAT)

~200 parameters → **high value / low value** → ~2400 simulations

1° resolution

250 pe-hrs/yr * (1500 yr spinup + 10 yr simulation) = 375,000 pe-hrs per simulation

One-at-a-time ensemble (OAAT)

allocate 2M pe-hours

~200 parameters → high value / low value → ~2400 simulations

1° resolution

250 pe-hrs/yr * 1500 yr spinup + 10 yr simulation → 5 simulations

One-at-a-time ensemble (OAAT)

allocate 2M pe-hours

~200 parameters → high value / low value → ~2400 simulations

2° resolution

70 pe-hrs/yr * 1500 yr spinup + 10 yr simulation → 20 simulations

One-at-a-time ensemble (OAAT)

allocate 2M pe-hours

~200 parameters → high value / low value → ~2400 simulations

sparsegrid

5 pe-hrs/yr * (1500 yr spinup + 10 yr simulation) → 250 simulations

Hoffman et al. 2013, *Landscape Ecology*

One-at-a-time ensemble (OAAT)

allocate 2M pe-hours

~200 parameters → high value / low value → ~2400 simulations

sparsegrid

matrix-CN

5 pe-hrs/yr * 140 yr spinup + 10 yr simulation → 2600 simulations

Hoffman et al. 2013, *Landscape Ecology*

Lu et al. 2020, *JAMES*

Infrastructure Developments

CLM5 Parameter List

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

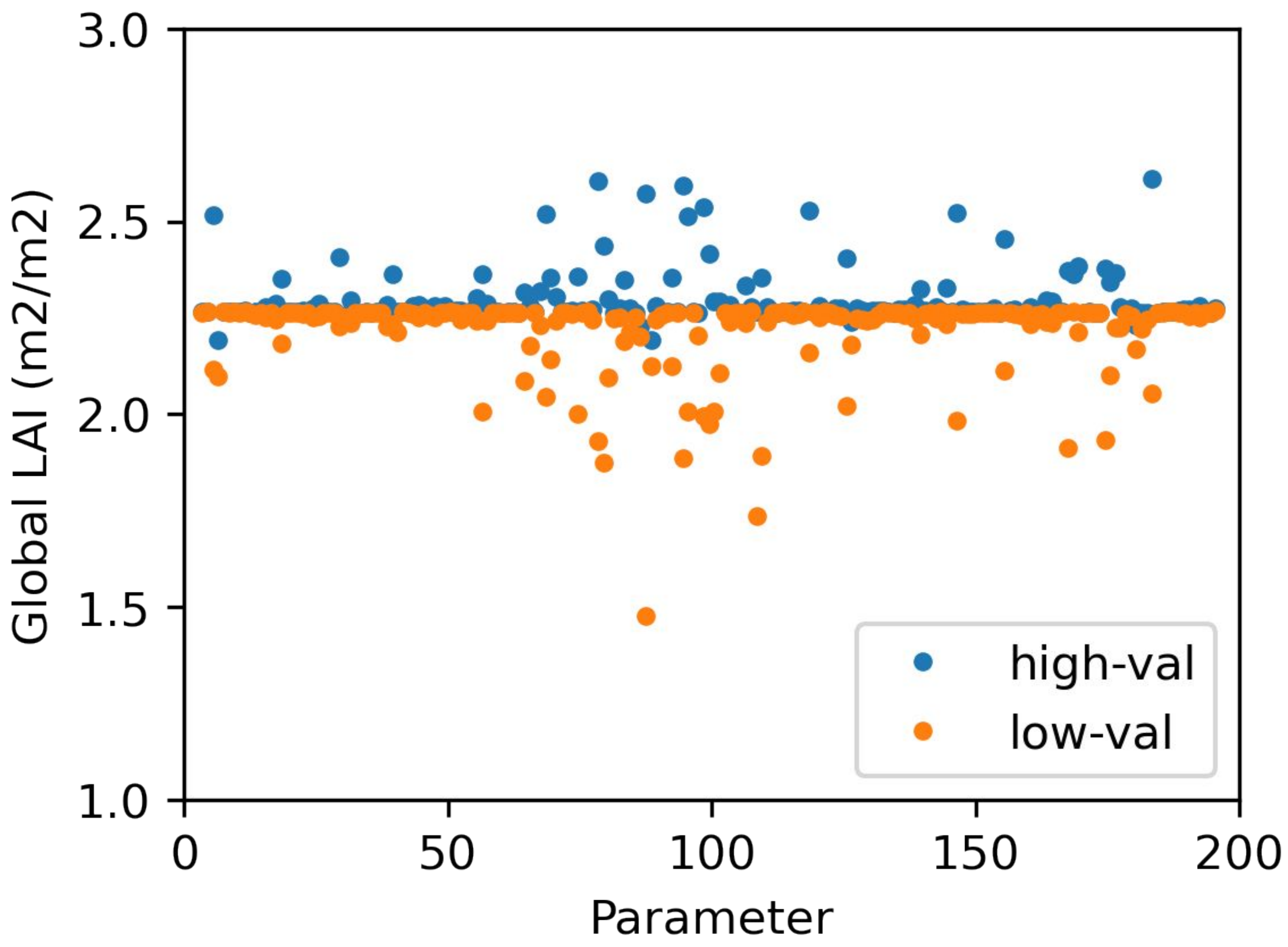
	D	E	F	G	S	
1	name	location	min	max	References for parameter ranges	Notes
97	**Photosynthetic capacity (LUNA)**					
98	slatop	P	pft	pft	Fisher et al. 2019, Kattge et al. 2011	
99	dsladlai	P	20percent	20percent	Thornton and Zimmerman 2007	
100	jmaxb0	P	0.01	0.05		
101	jmaxb1	N	0.05	0.25		
102	wc2wjb0	P	0.5	1.5		
103	enzyme_turnover_daily	P	0.05	0.15		
104	relhExp	P	4	8		
105	minrelh	P	0.2	0.3		
106	luna_theta_cj	P	0.8	0.99		

Define parameter ranges

One-at-a-time ensemble (OAAT)

why OAAT?

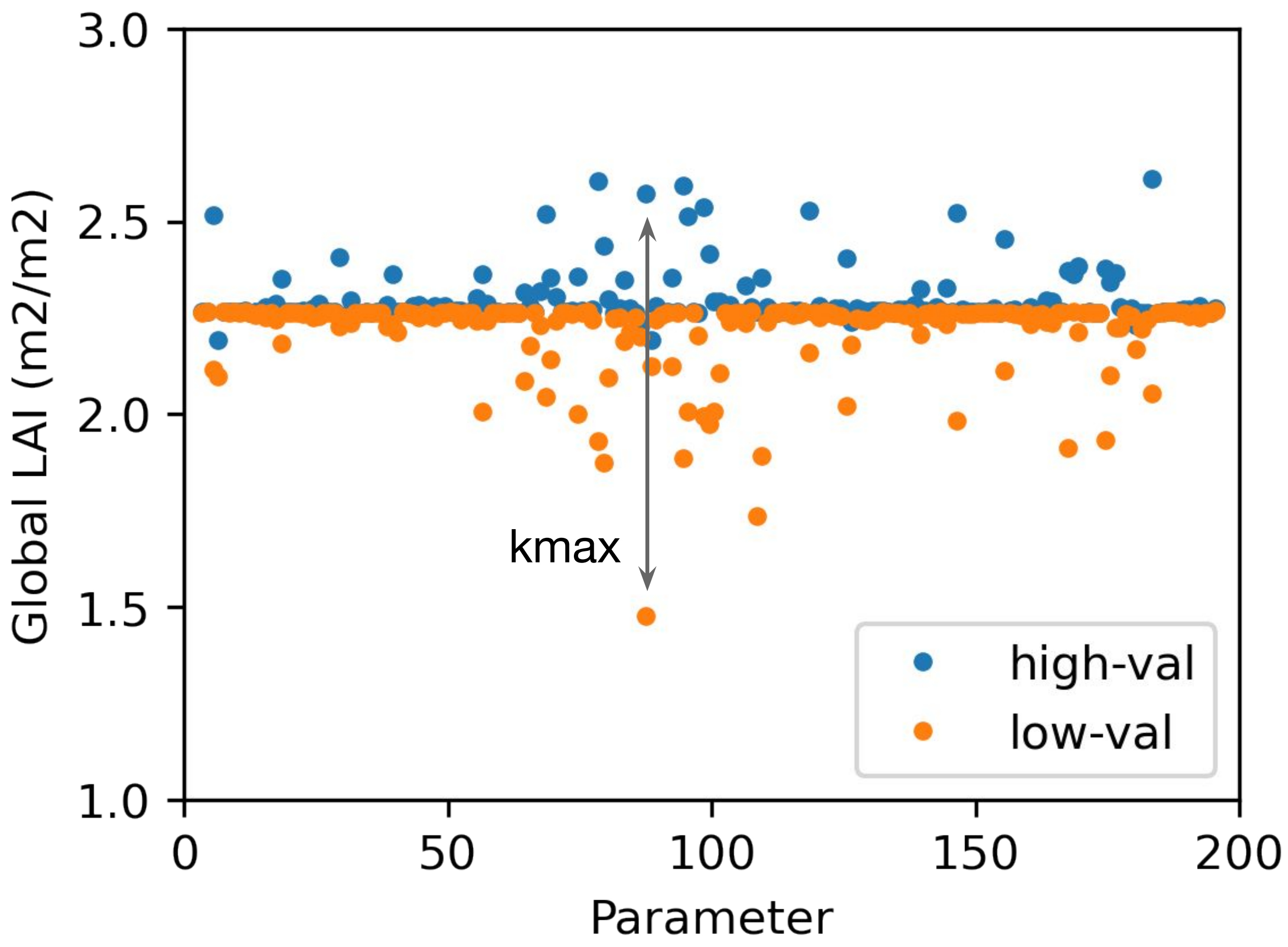
1. yields an easy-to-interpret dataset



One-at-a-time ensemble (OAAT)

why OAAT?

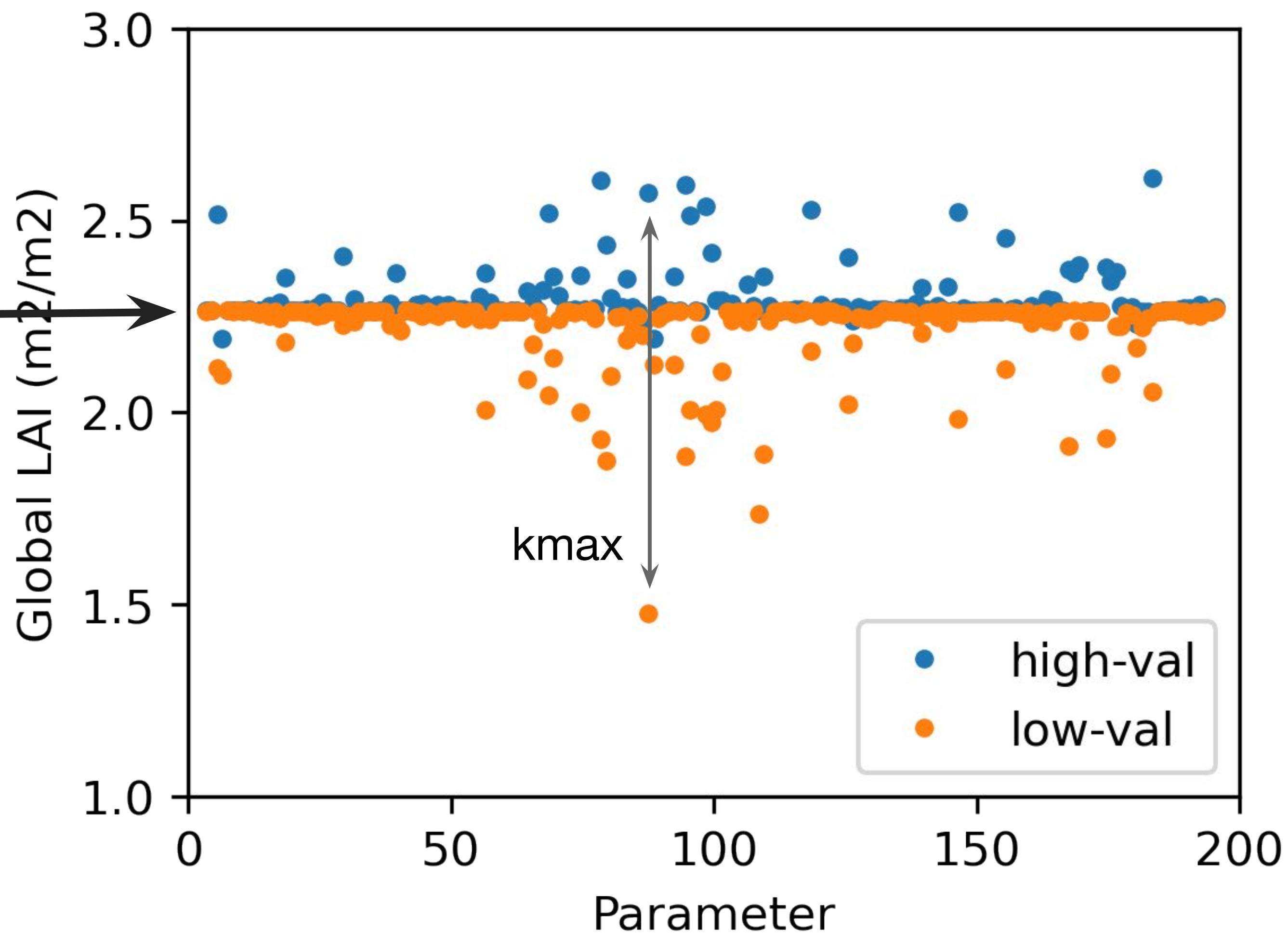
1. yields an easy-to-interpret dataset



One-at-a-time ensemble (OAAT)

why OAAT?

default CLM5

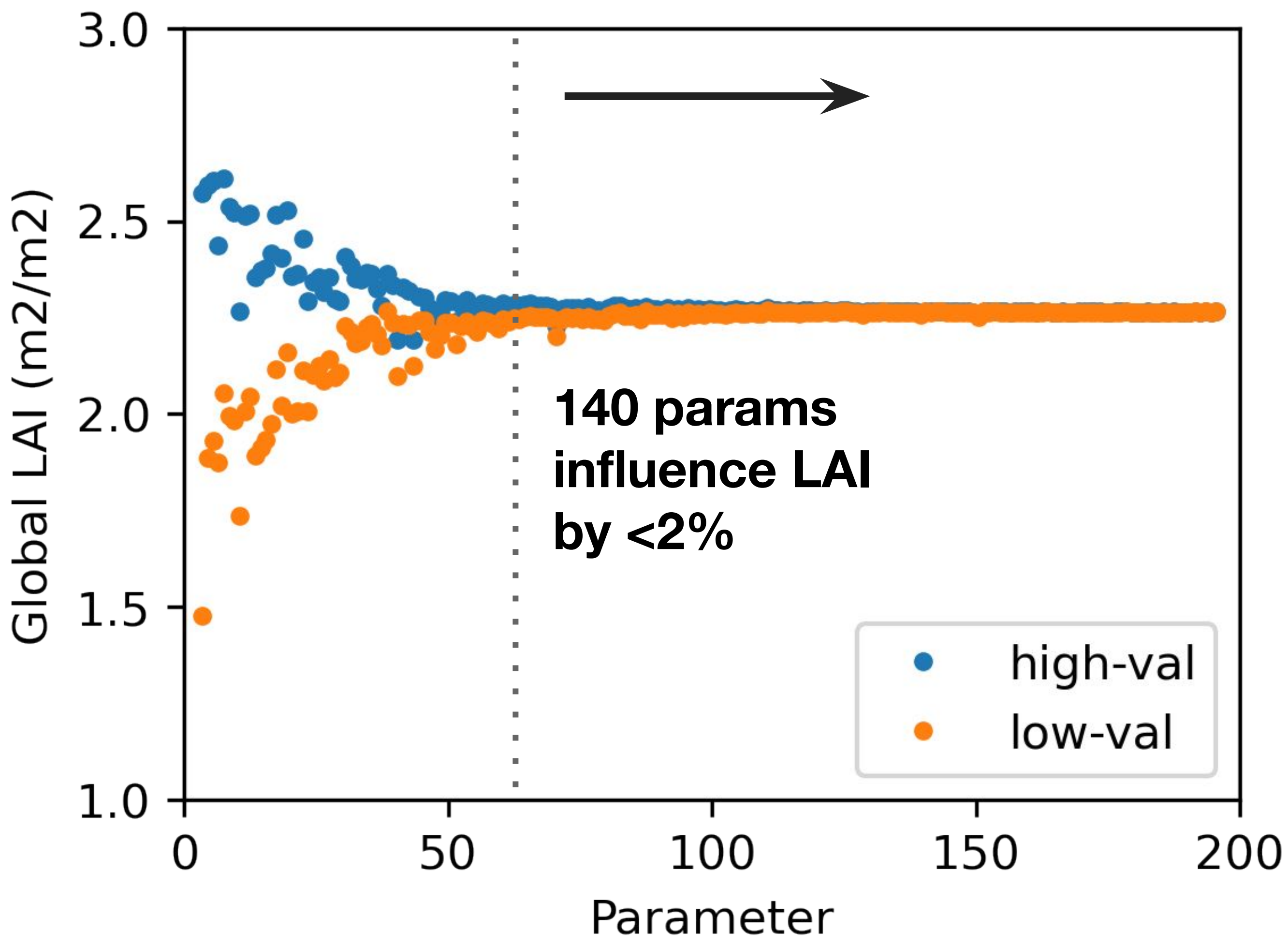


1. yields an easy-to-interpret dataset
2. many parameters have small effect

One-at-a-time ensemble (OAAT)

why OAAT?

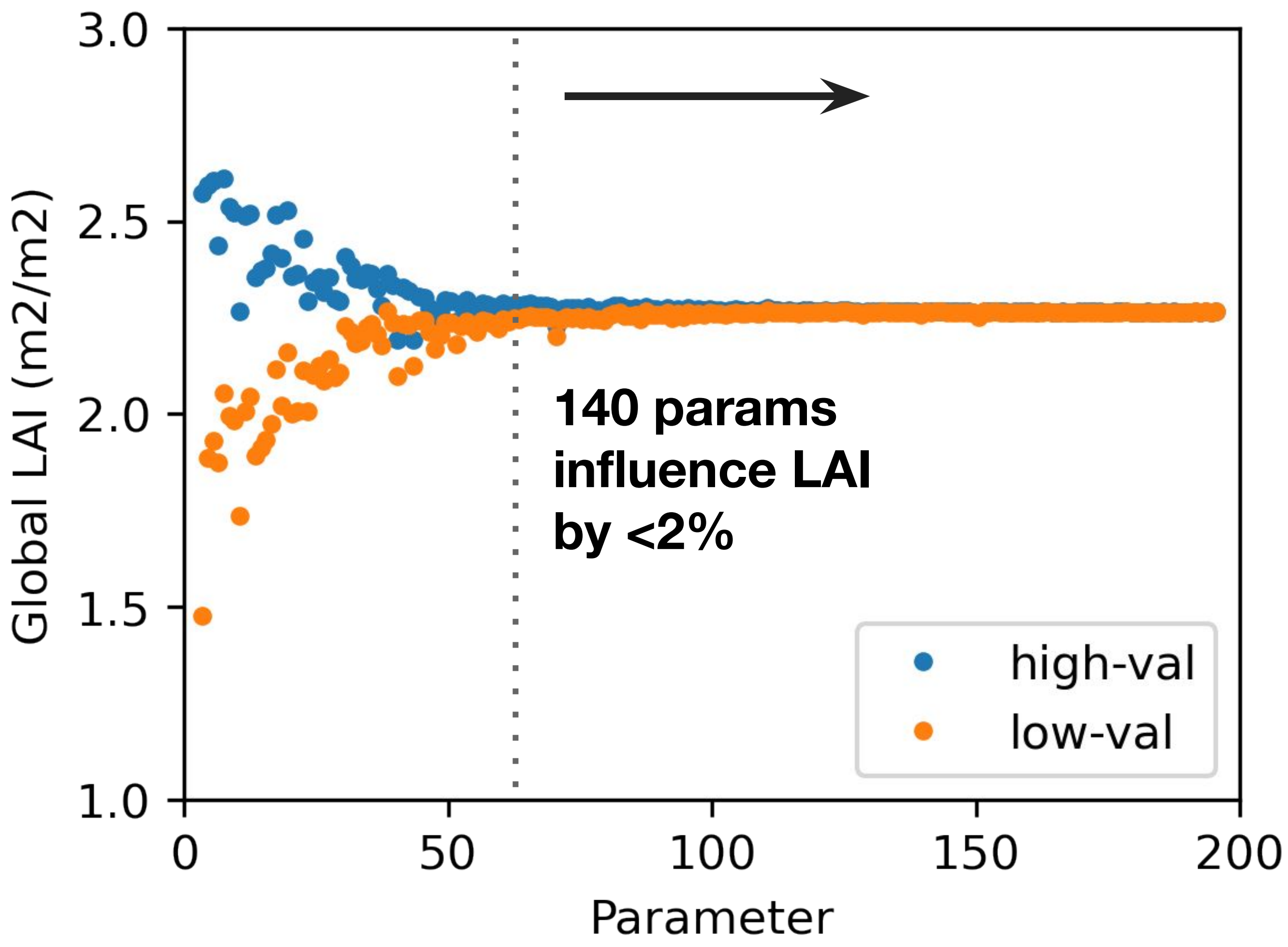
1. yields an easy-to-interpret dataset
2. many parameters have small effect



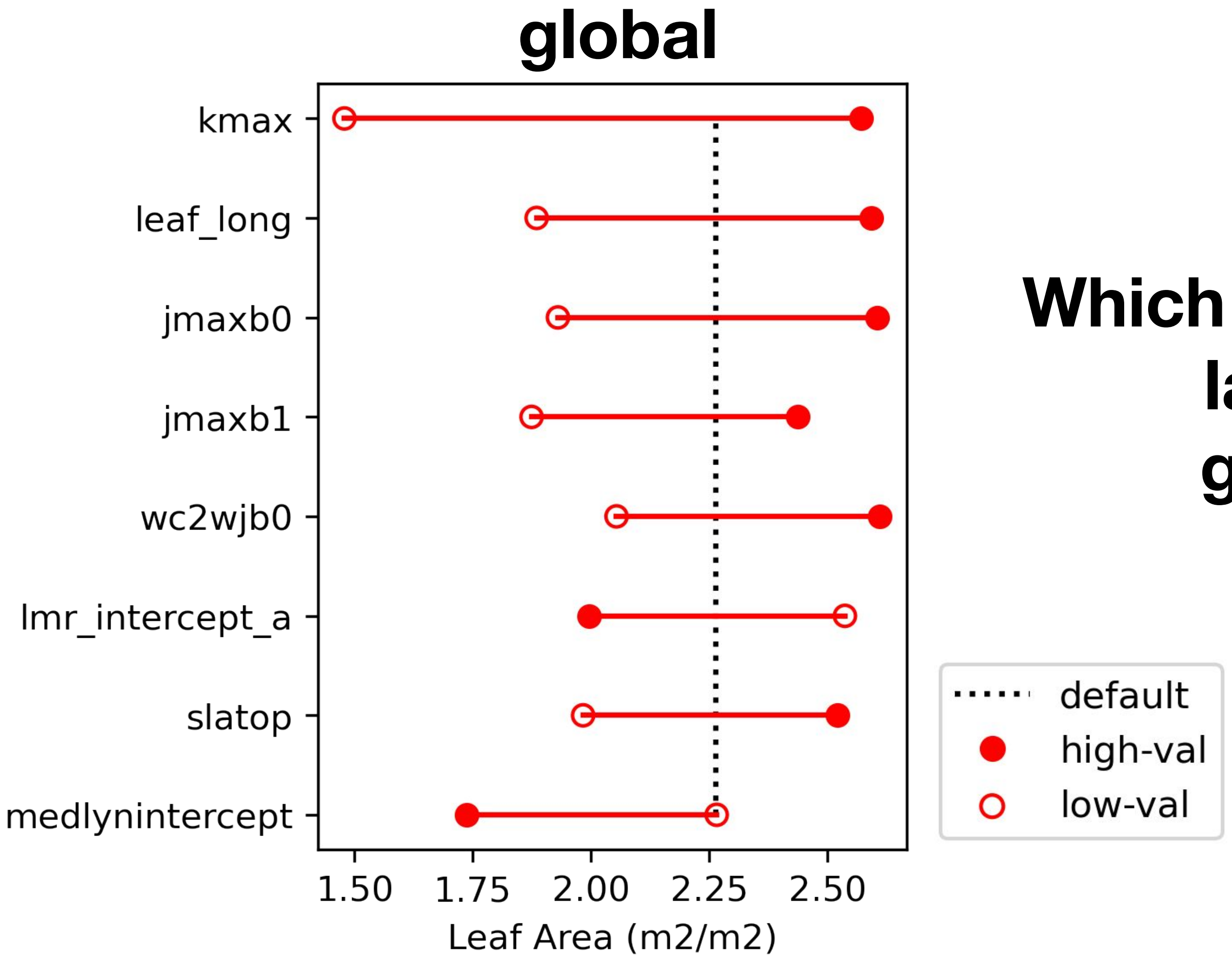
One-at-a-time ensemble (OAAT)

why OAAT?

1. yields an easy-to-interpret dataset
2. many parameters have small effect
3. latin hypercube of 200+ params seemed tenuous

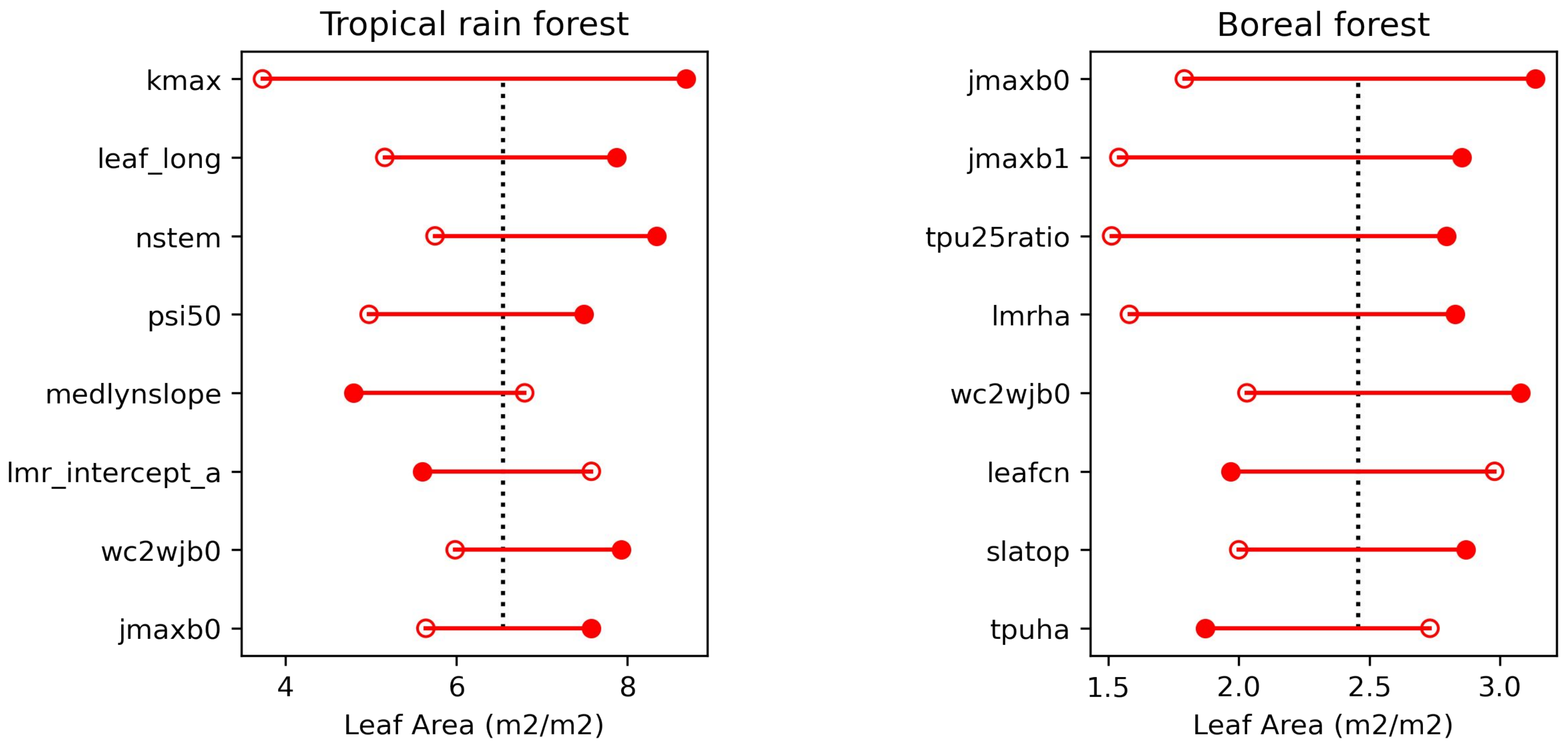


Parameter Ranking Diagnostics



Which parameters have the largest effect on global leaf area?

Parameter Ranking Diagnostics: by biome

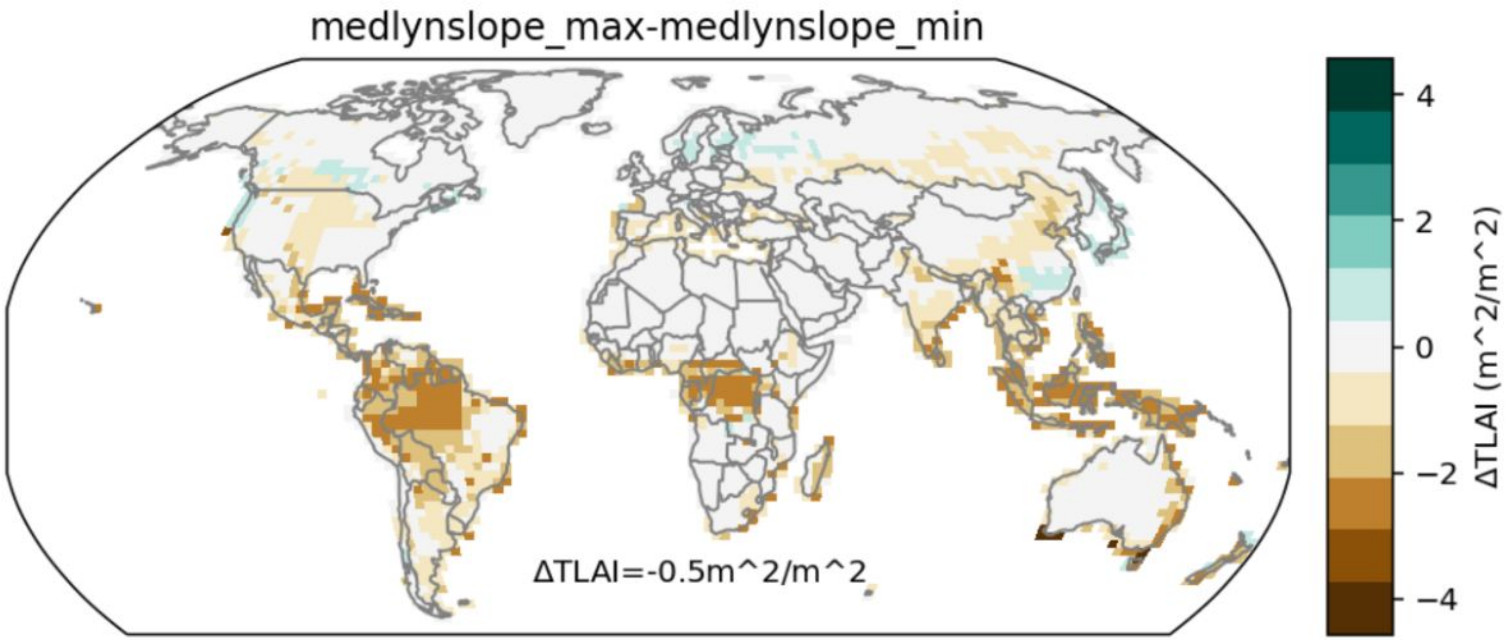


Mix and match parameter rankings

Variable	Domain	Metric	Forcing Scenario
Leaf area Photosynthesis Fire Energy fluxes Soil moisture Albedo …	Global Biome Plant type	Mean Interannual variability Seasonal amplitude …	Control High/Low CO_2 Future/PI Climate +Nitrogen

great for parameter screening

Interactive Visualization



DataVar

TLAI

Parameter

medlynslope

helpful model "smoke test"

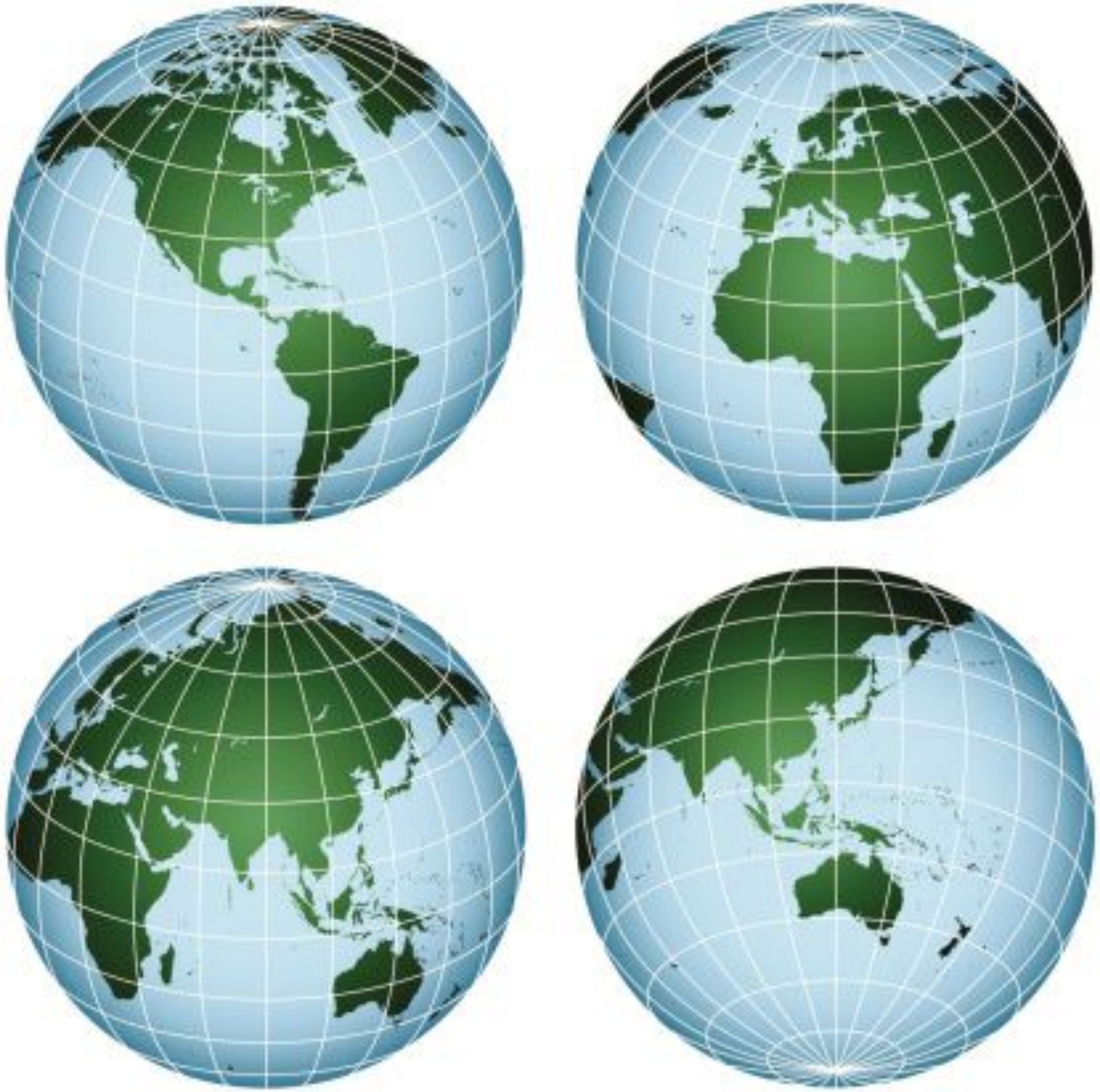
OAAT Deliverables

1. Featherweight CLM
2. Collated parameter ranges
3. Ensemble generation workflow
4. Python analysis library
5. Extensive diagnostics of CLM5 parameter sensitivity

github.com/djk2120/clm5ppe

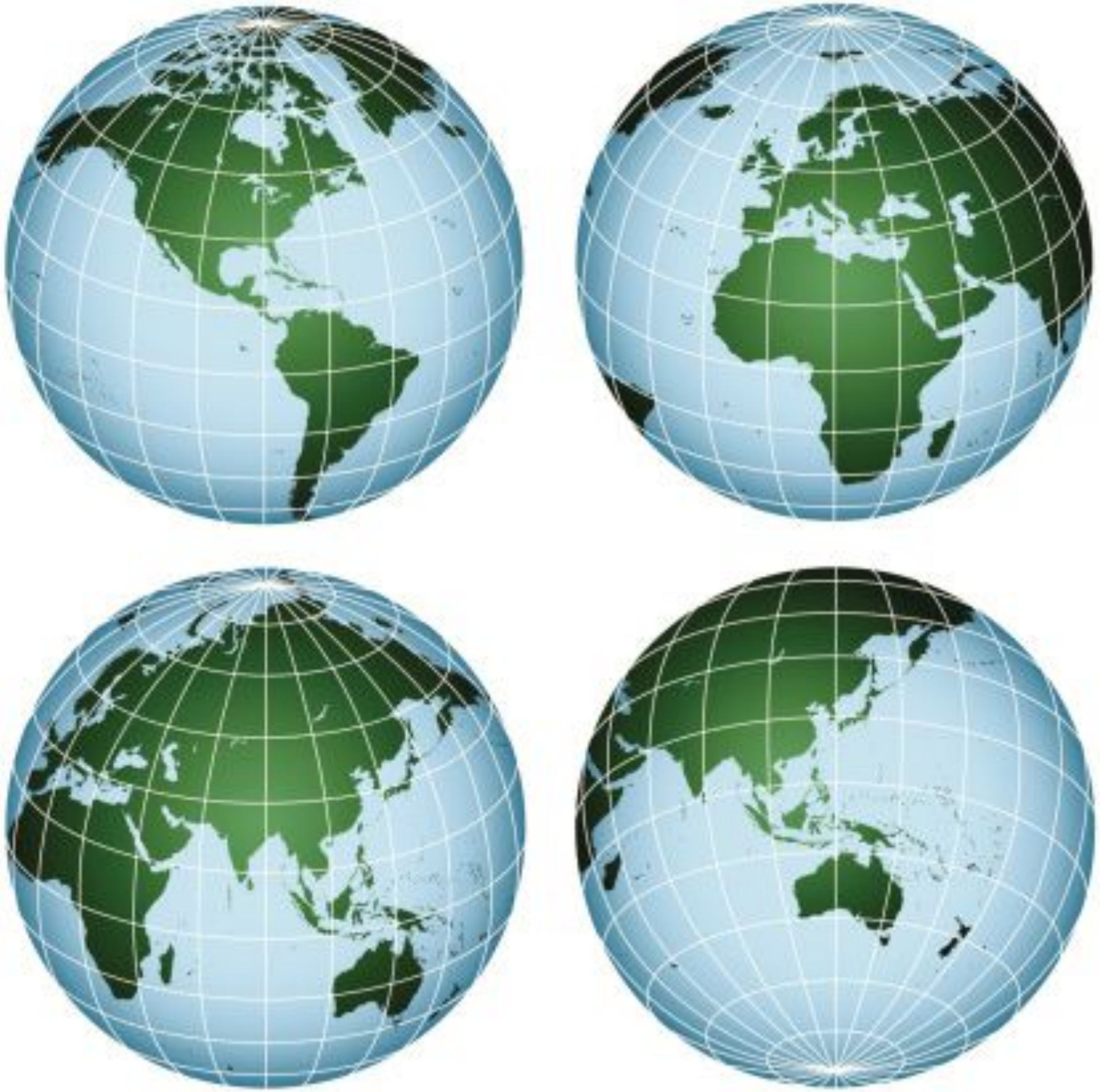
CLM PPE Coordinated Projects

- Arctic hydrology (NCAR-RAL)
  - Yifan Cheng: ***LMWG talk yesterday***
- Large-sample watershed modeling (NCAR-CGD)
  - Guoqiang Tang: ***talk yesterday***
- Runoff sensitivity (Michigan State)
  - Ahmed Elkouk: ***talk yesterday***
- Land-atmosphere interactions (Univ Washington)
  - Claire Zarakas: ***talk today at 1pm MT***
- NEON site calibration (Auburn Univ)
  - Thomas Kavoo: ***talk today at 1:15pm MT***
- LAI calibration (NCAR/Columbia)
  - Linnia Hawkins: ***talk today at 2:15pm MT***



CLM PPE Coordinated Projects

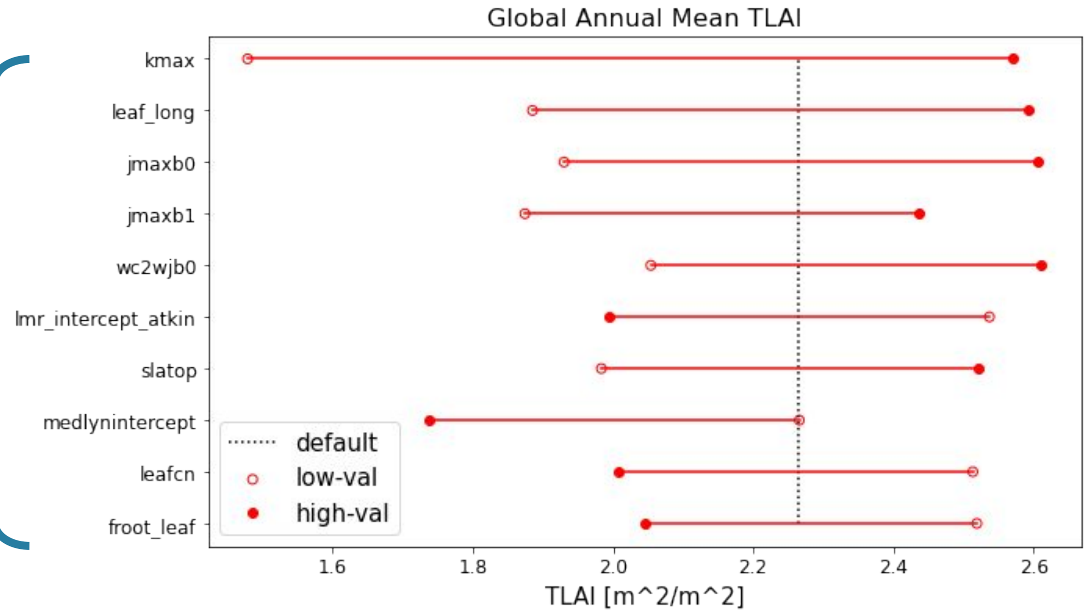
- Arctic hydrology (NCAR-RAL)
- Large-sample watershed modeling (NCAR-CGD)
- Runoff sensitivity (Michigan State)
- Land-atmosphere interactions (Univ Washington)
- NEON site calibration (Auburn Univ)
- LAI calibration (NCAR/Columbia)
- ET recession timescales (Oregon State)
- FATES PPE (NCAR)
- CONUS streamflow (PNNL)
- Land influence on drought (NCAR)
- Hydrologic sensitivity (Cornell Univ)
- Tropical carbon cycle interannual variability (JPL)
- GPP response to permafrost thaw (NAU)



Land Model (Large) Perturbed Parameter Ensemble

Land model has over 200 parameters!
How do we rank and select important parameters?

"Top 10" parameters for global, annual mean **total leaf area index (TLAI)** under control climate.



ROADMAP

1. **Identify and define parameter ranges** – what is a "parameter"?
2. **Infrastructure development:** parameter sampling, ensemble generation, computational efficiency (e.g., fast spin-up, sparse grid).
3. **Run one-at-a-time perturbations** with all 221 parameters, multiple forcing scenarios (e.g., low/high CO_2).
4. **Parameter selection:** variables of interest, metrics, biomes, environmental perturbations.
5. **All-at-once ensemble** with a subset of parameters focused on a particular calibration target.

Multifactor Ensemble

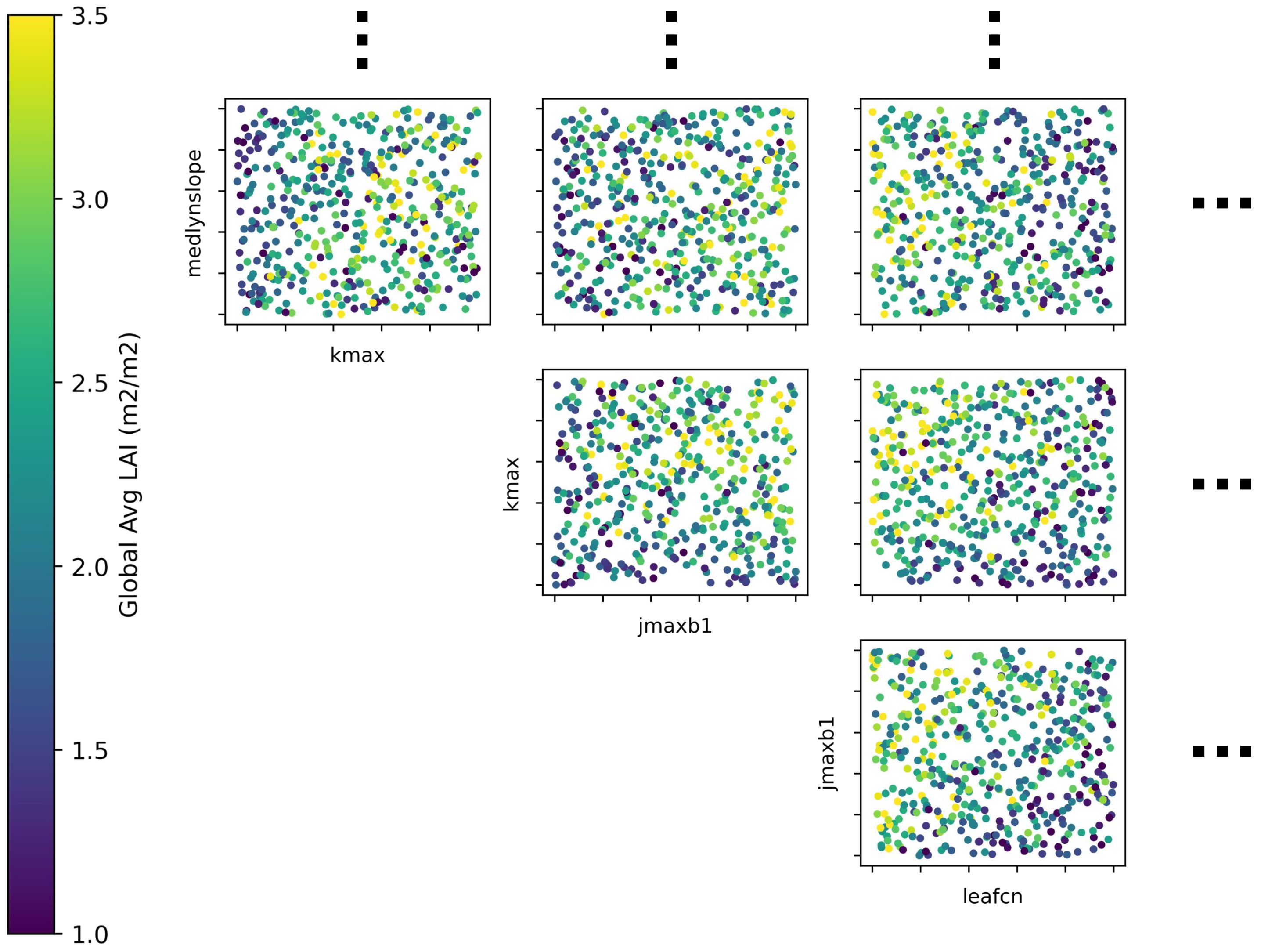
Focused on LAI calibration:

- challenging, but tractable
- foundational variable within CLM5-bgc
- observational constraints

Experimental Design:

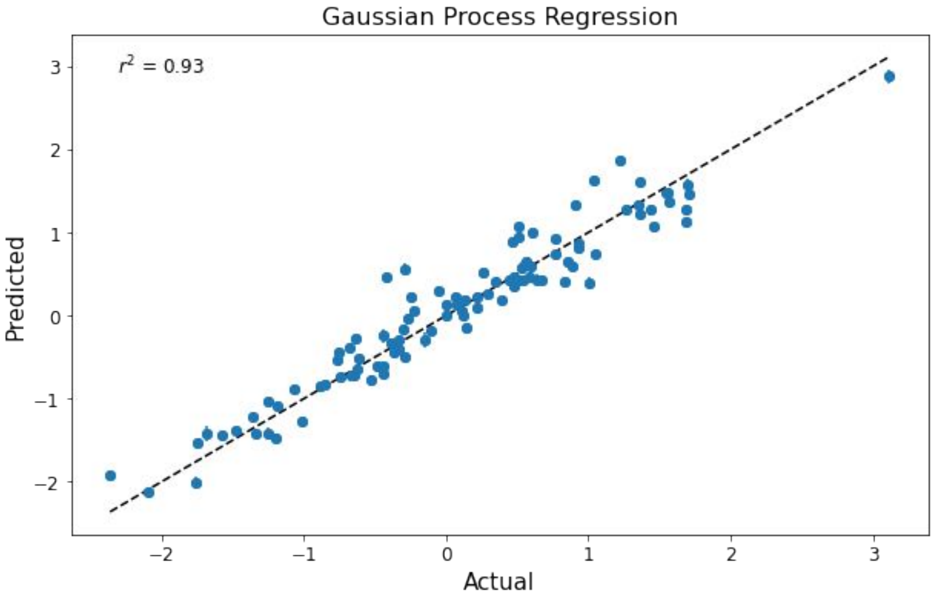
- subset 32 relevant parameters
- 500 simulations
- fully transient, 1850-2014
- Latin hypercube (LHC) sampling

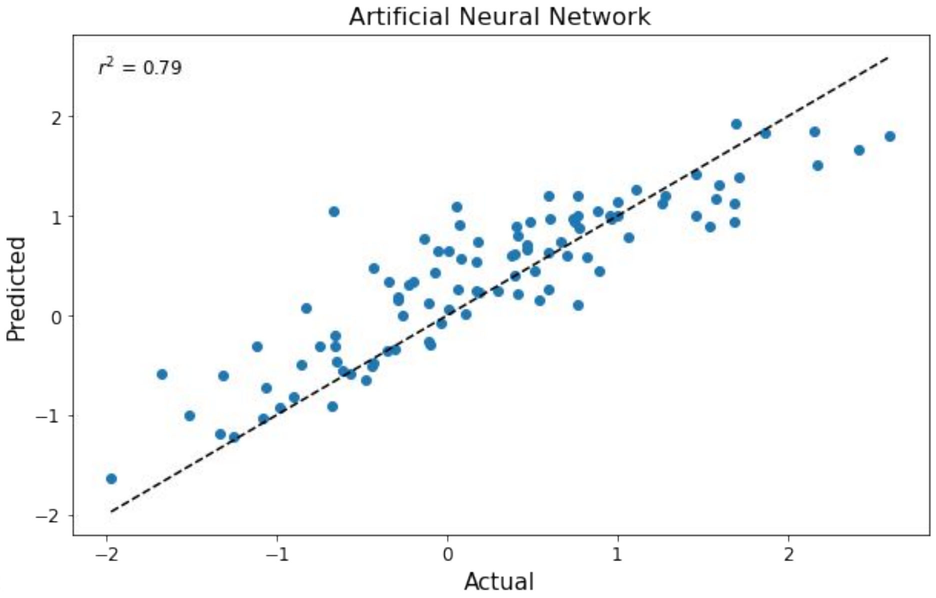
Hypercube results are difficult to parse directly

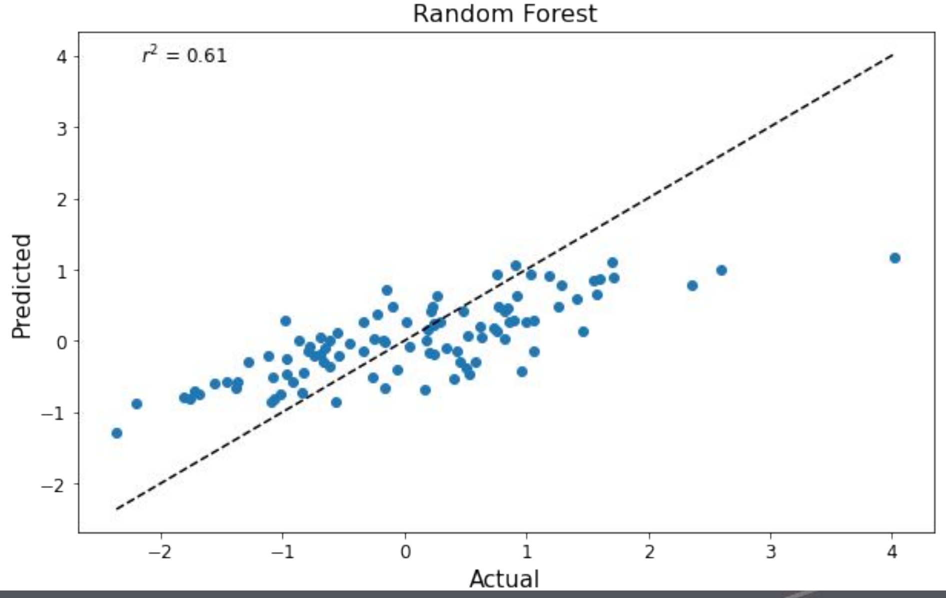


32 parameters varying all at once

Comparing Emulation Algorithms





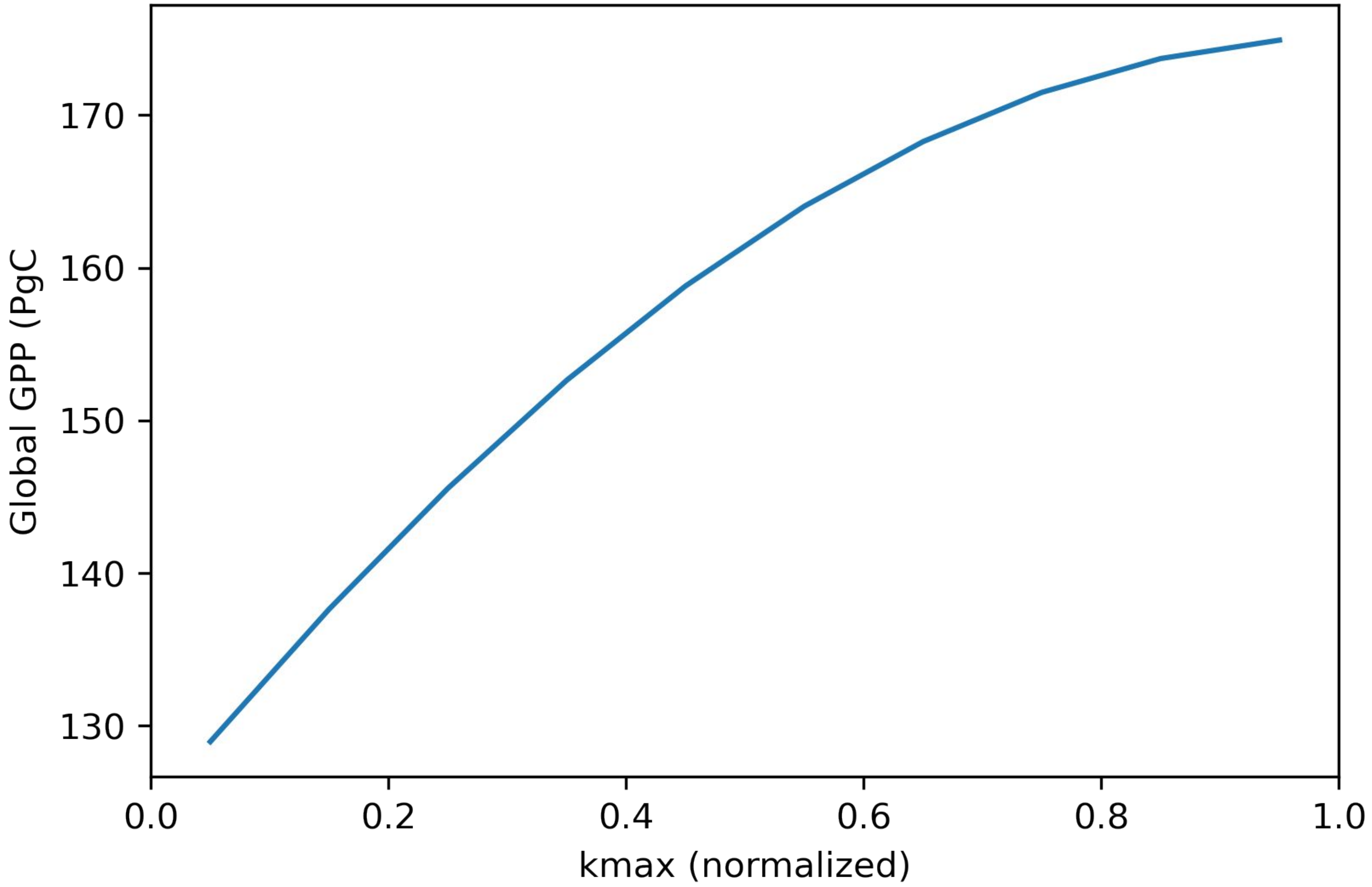


Emulating global annual mean leaf area index.

***Preliminary** hyperparameter tuning in progress!*

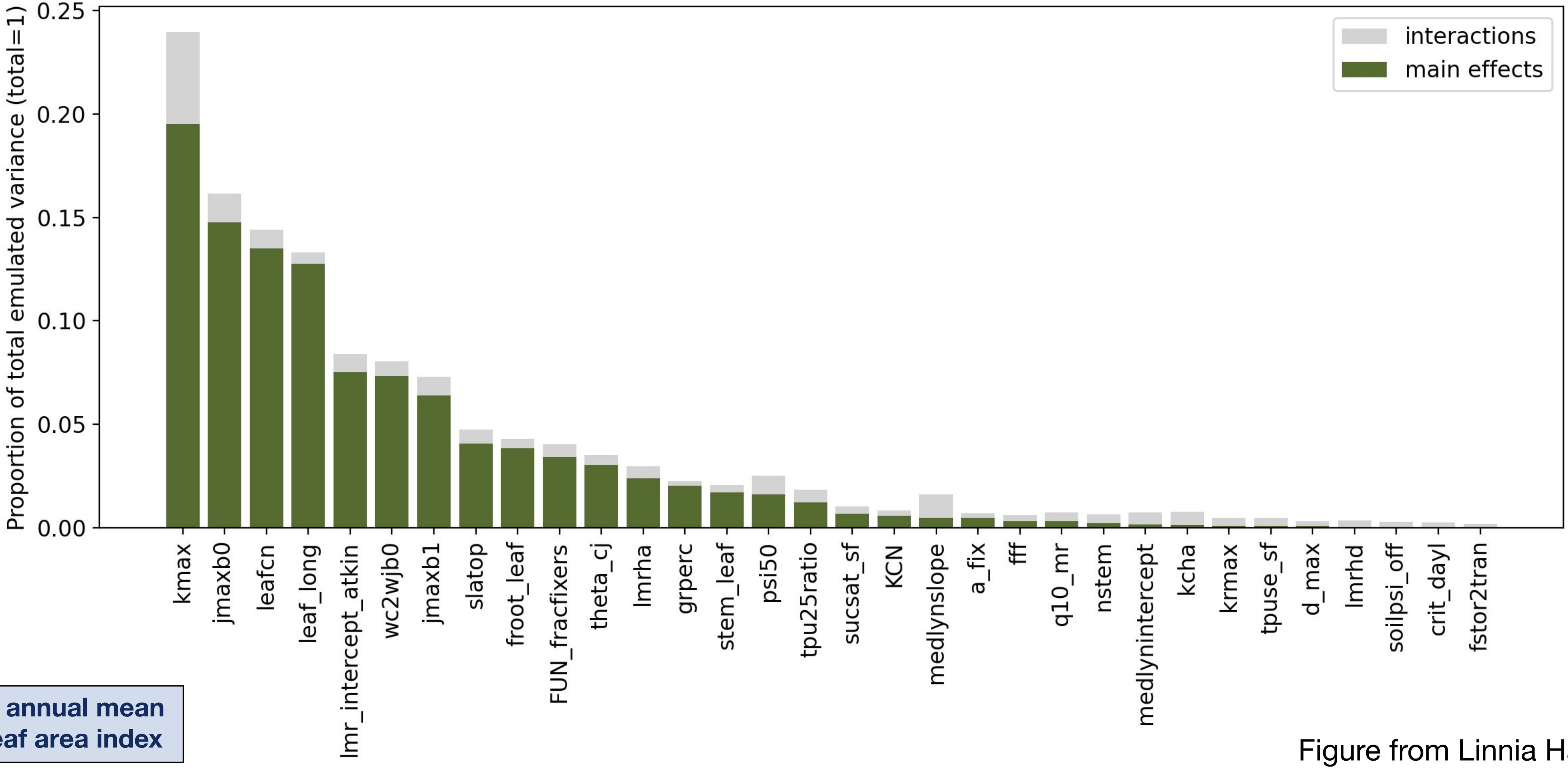
Thanks to the ESEm Python package:
https://github.com/duncanwp/ESEm
Watson-Parris et al. 2021

Identify nonlinear responses



increased kmax: saturating effect on photosynthesis

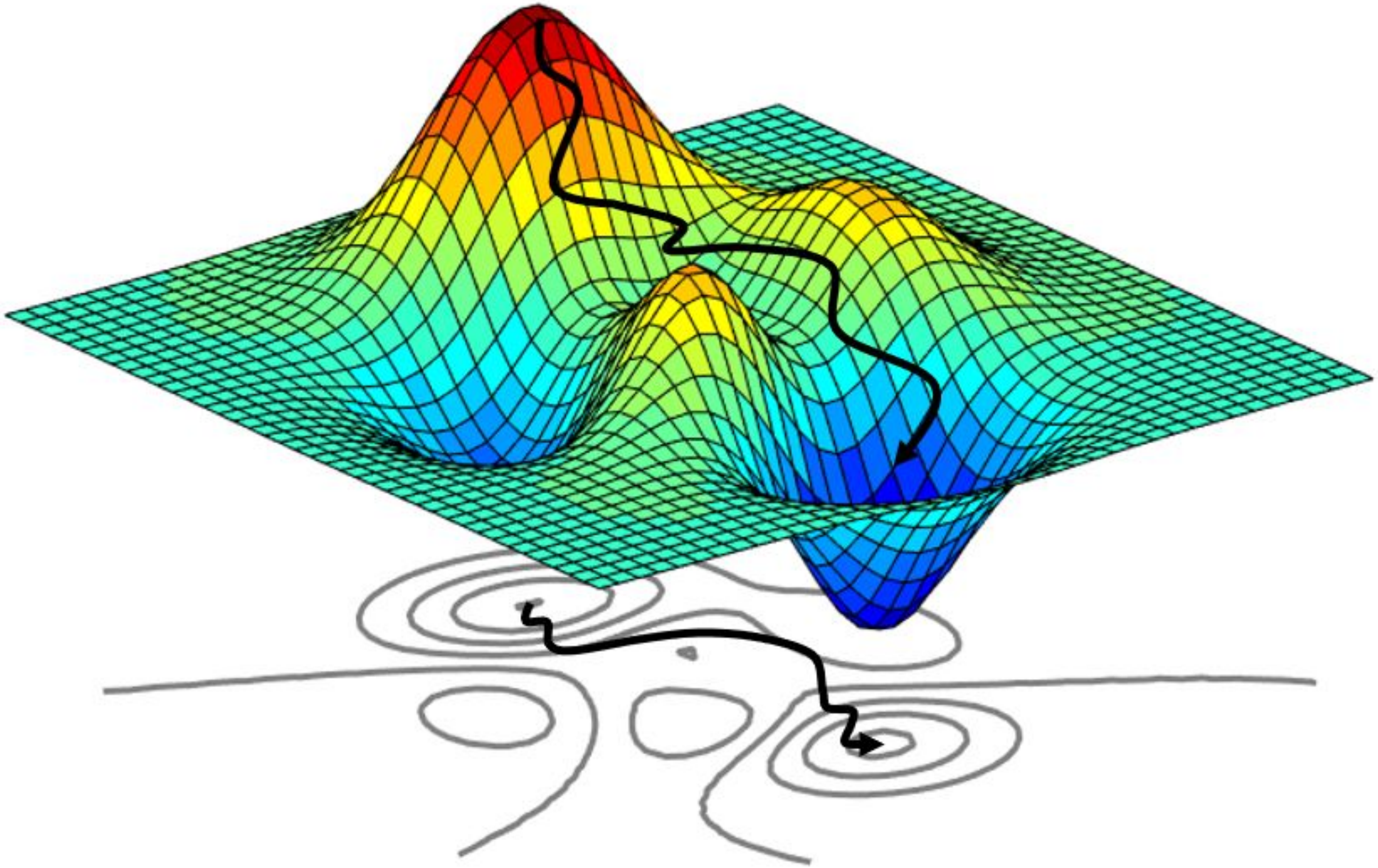
Global Sensitivity Analysis



Global annual mean total leaf area index

Figure from Linnia Hawkins

Model Calibration

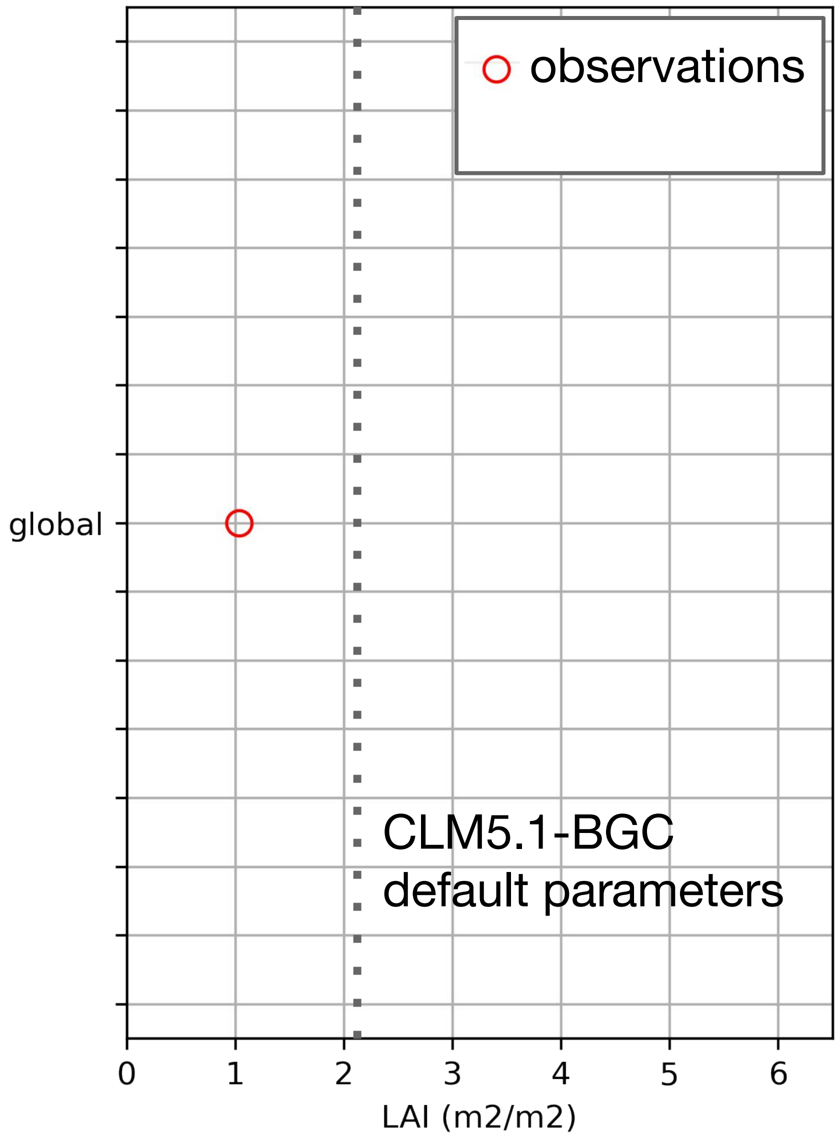


Main challenges:

- dimensionality
- equifinality
- structural errors

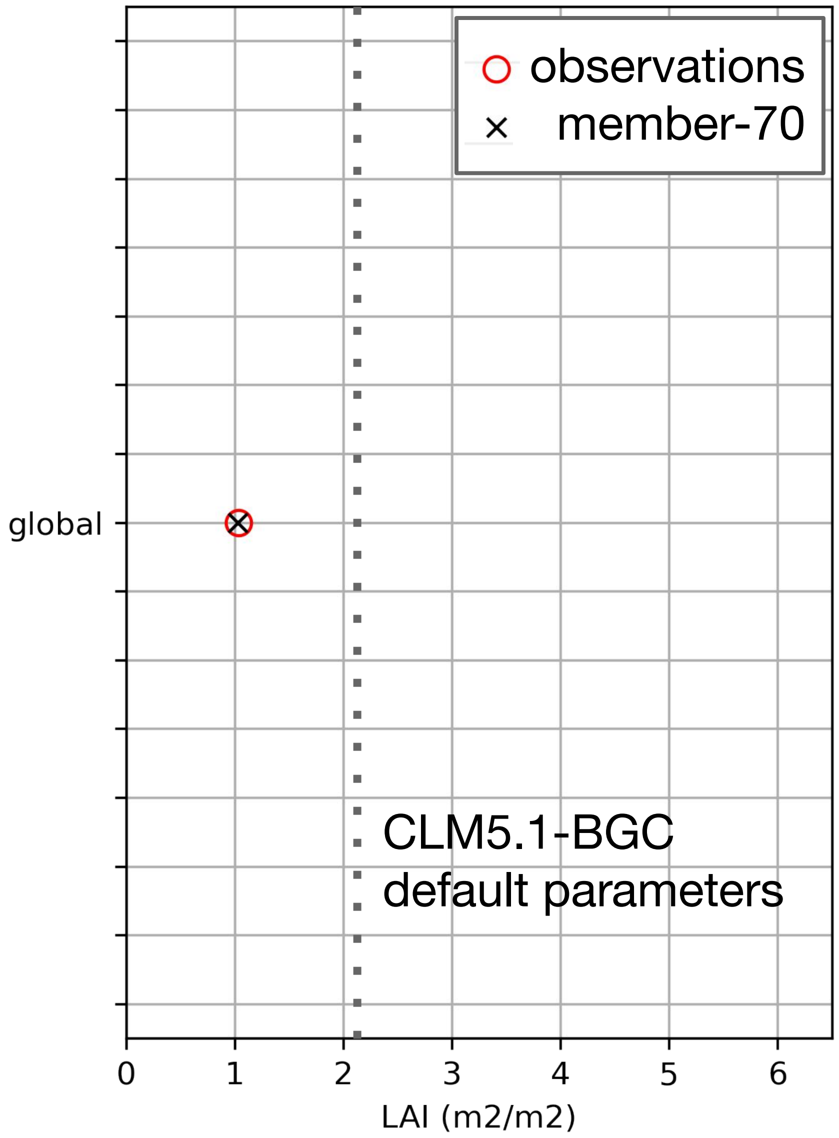
Global leaf area index

can we resolve the global LAI bias?

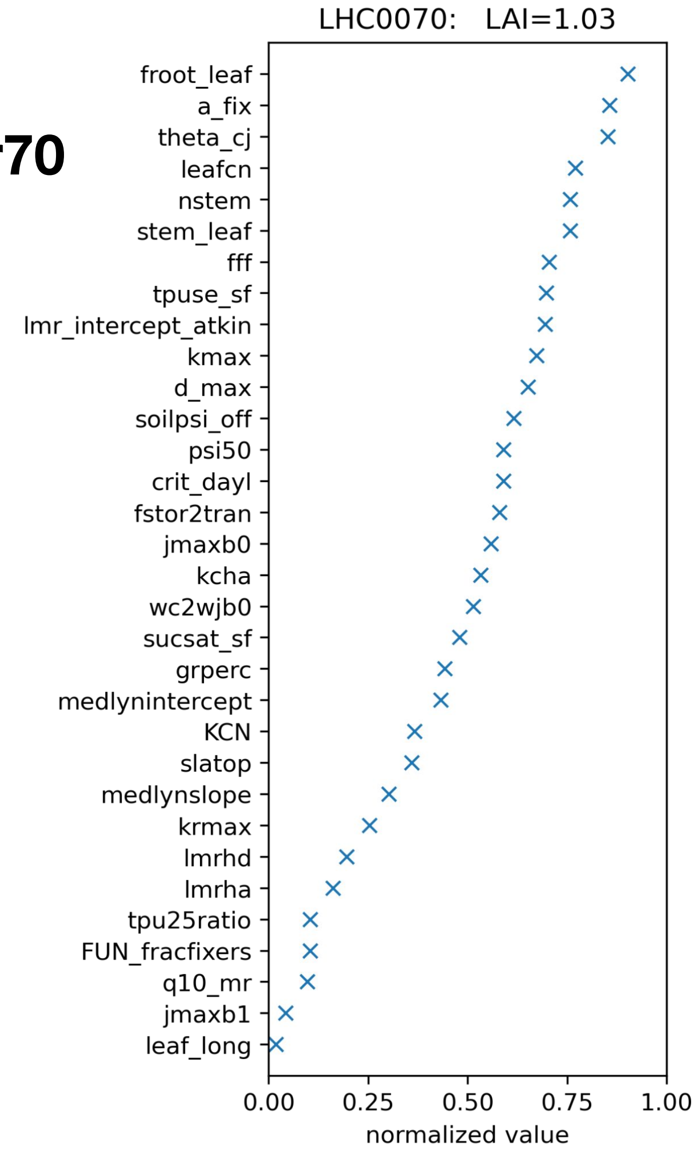


Global leaf area index

can we resolve the global LAI bias?

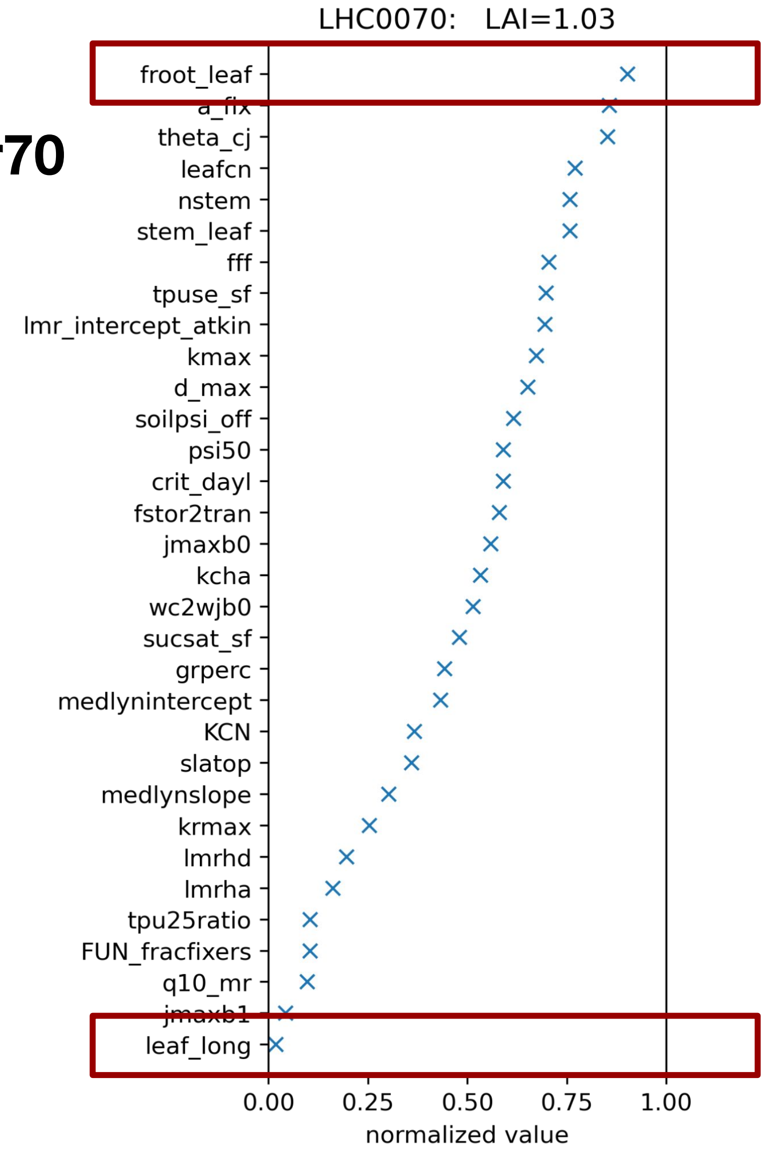


member70



what does the parameter set look like?

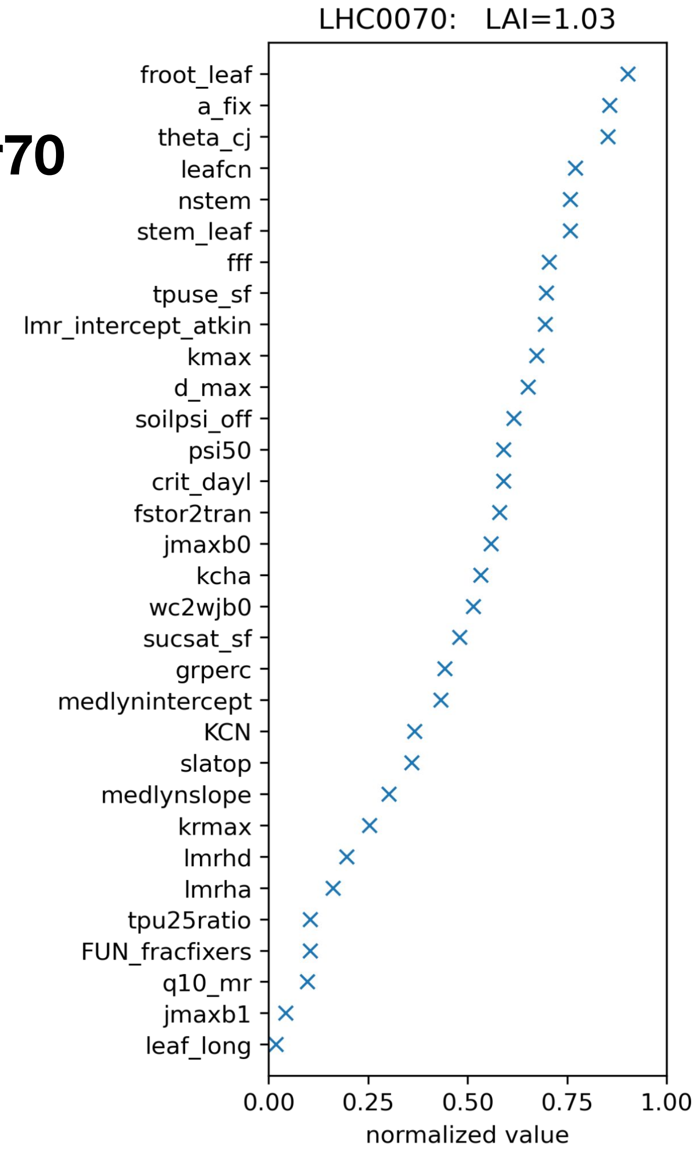
member70

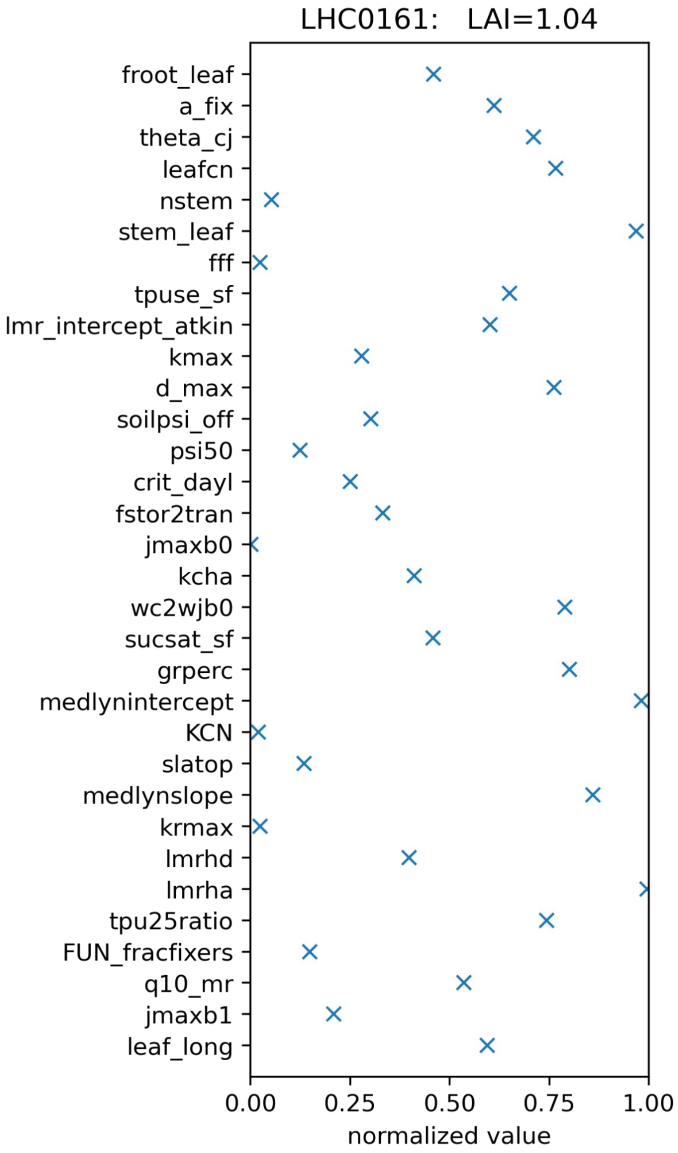


what does the parameter set look like?

Equifinality

member70

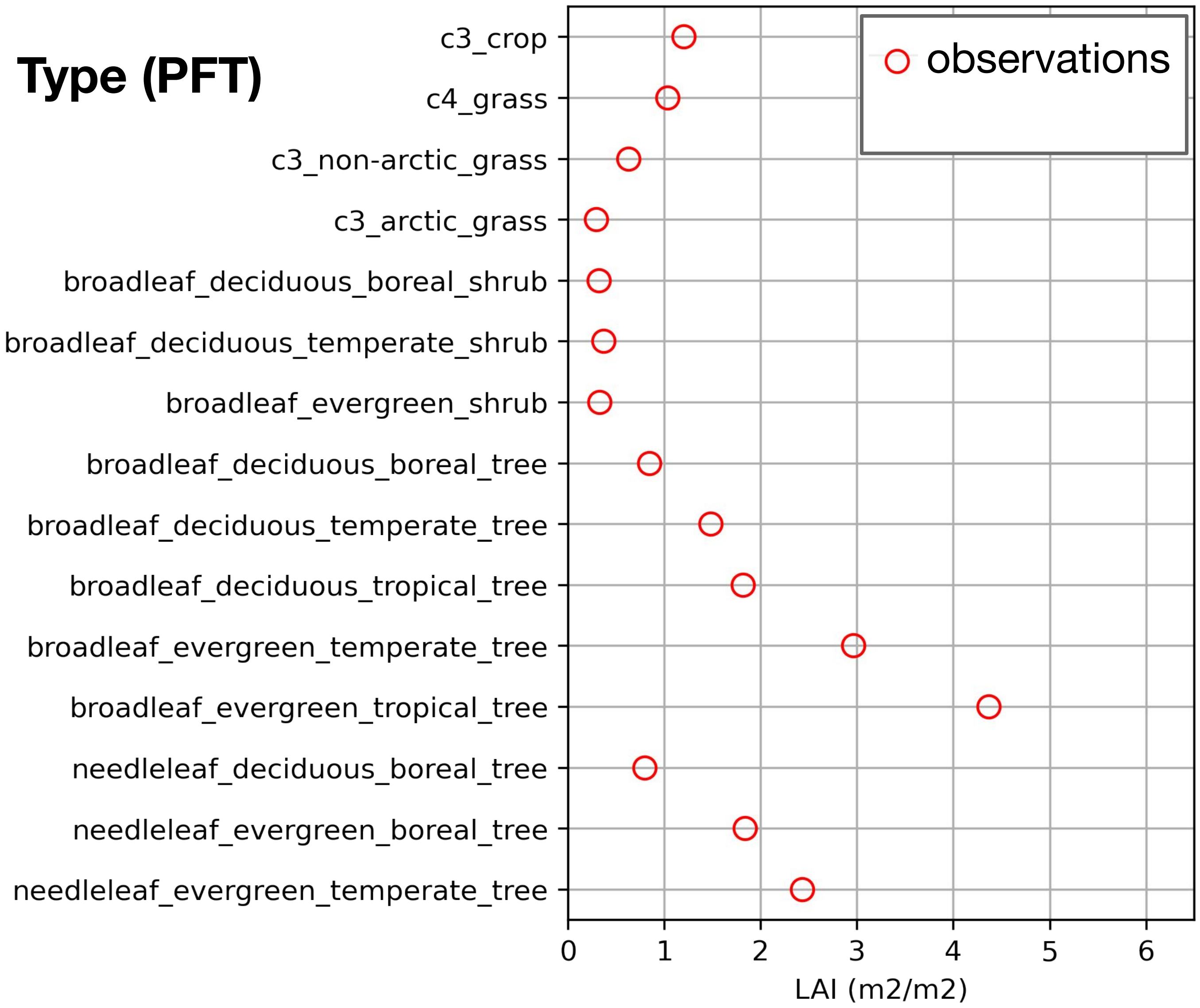




member161

PFT-Level Leaf Area Index

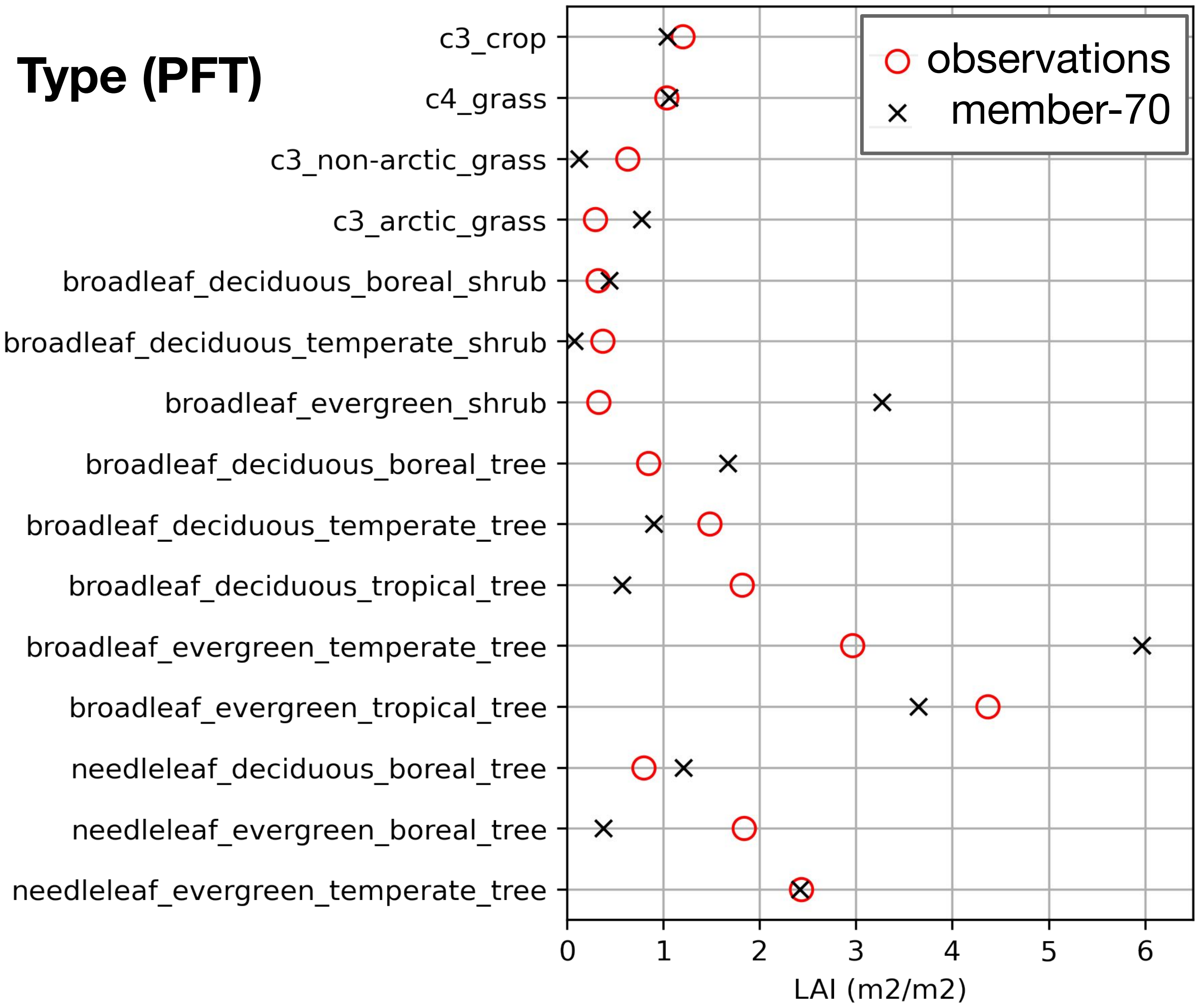
Plant Functional Type (PFT)



PFT-Level Leaf Area Index

Plant Functional Type (PFT)

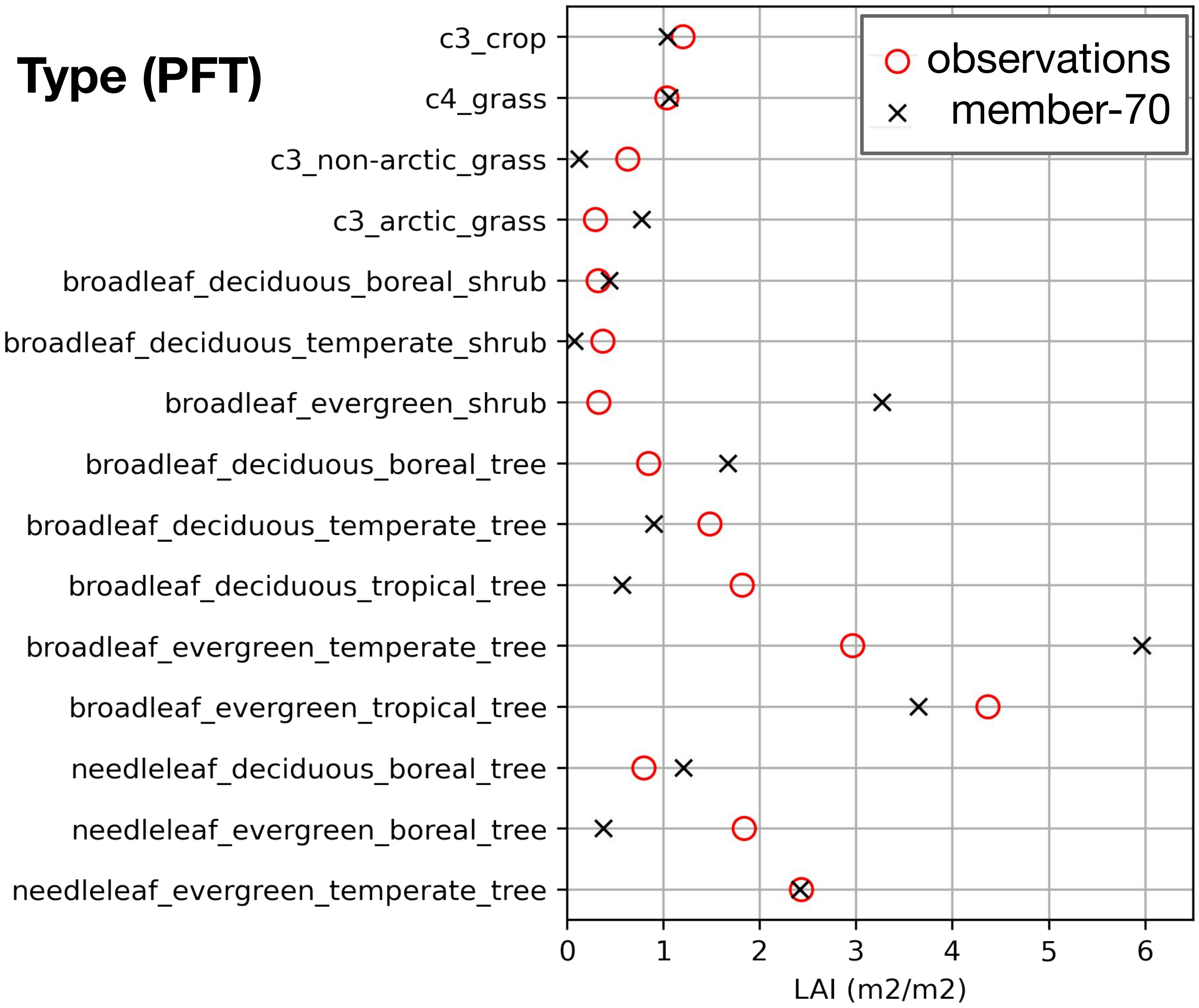
Beware of PFT tradeoffs



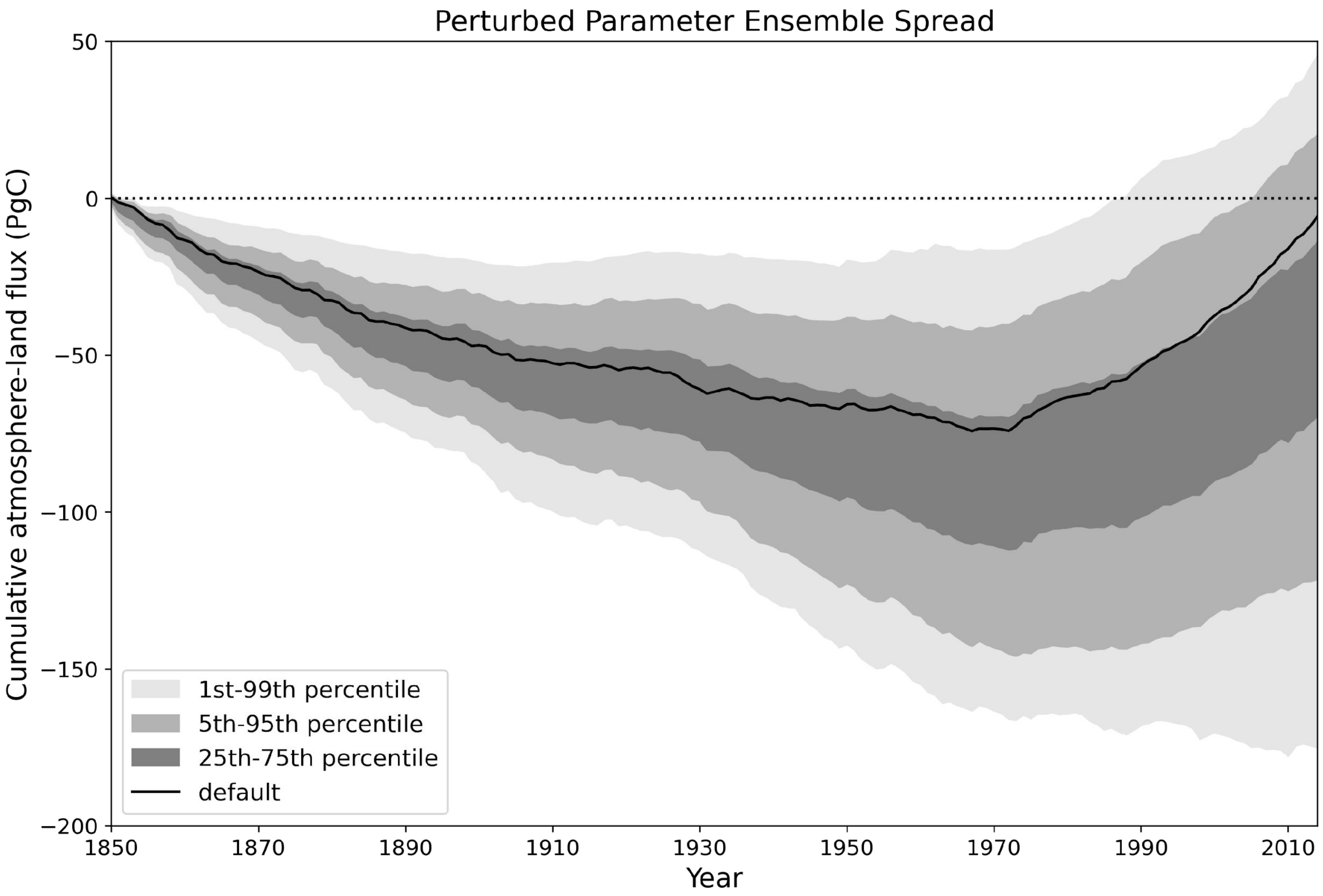
PFT-Level Leaf Area Index

Plant Functional Type (PFT)

leverage historical data to eliminate implausible paramsets



PPE Spread in Carbon Fluxes



Parameter sampling generates **~200 PgC** spread!

Repeatable Insight Generation

quick
repeatable
transparent

fast model

observations

sound
numerical
methods

parameter
priors

automated
workflows

parameter
insights

Summary

- **Systematic model calibration** is important to climate model development yet hindered by a variety of challenges.
- Land models in particular have **large parametric and structural uncertainties**, which impacts assessment of emergent climate features such as the **land carbon sink**.
- Machine learning emulation can help **optimize resources** and **reduce subjectivity** in model calibration.
- CLM5 PPE **community datasets** resulting in many offshoot projects and will be publicly available.
- We're making progress towards **open source** model calibration tools.

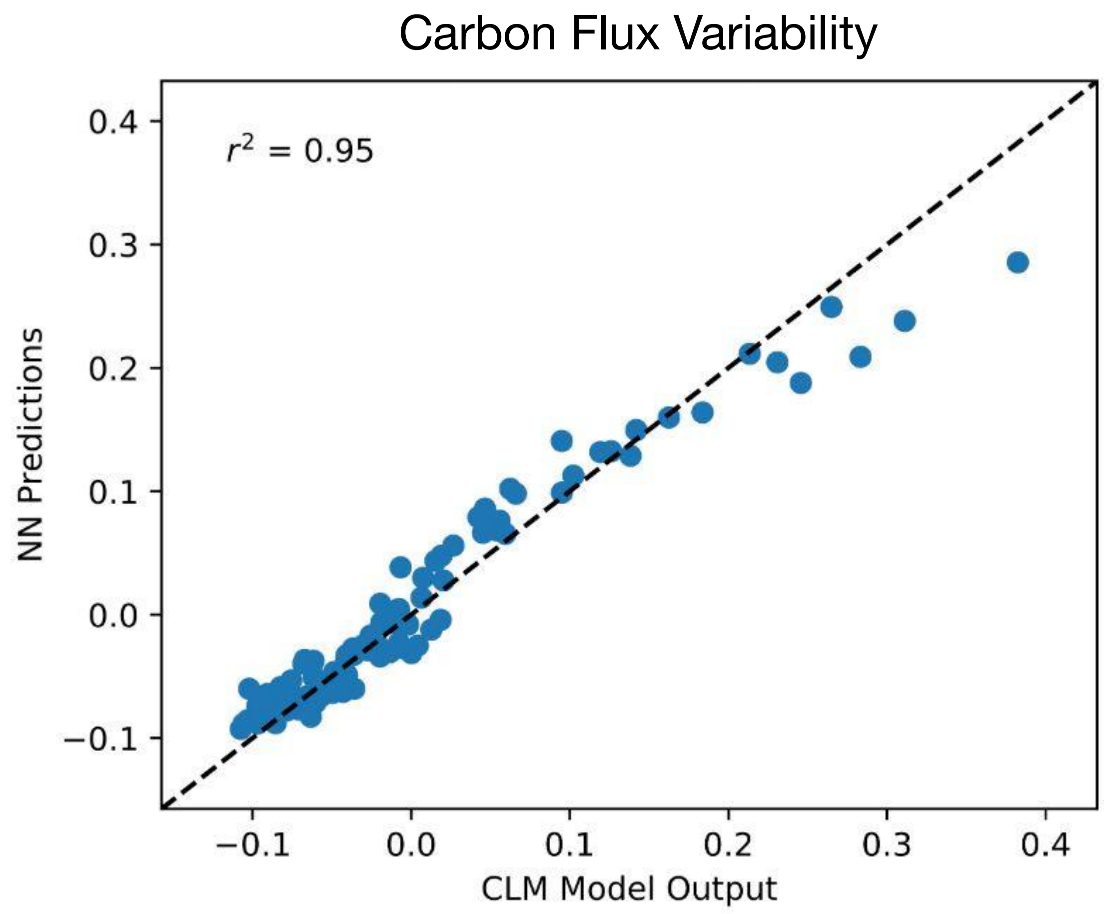
github.com/djk2120/ppe_tools
github.com/djk2120/CLM5PPE
github.com/katiedagon/CLM5_ParameterUncertainty

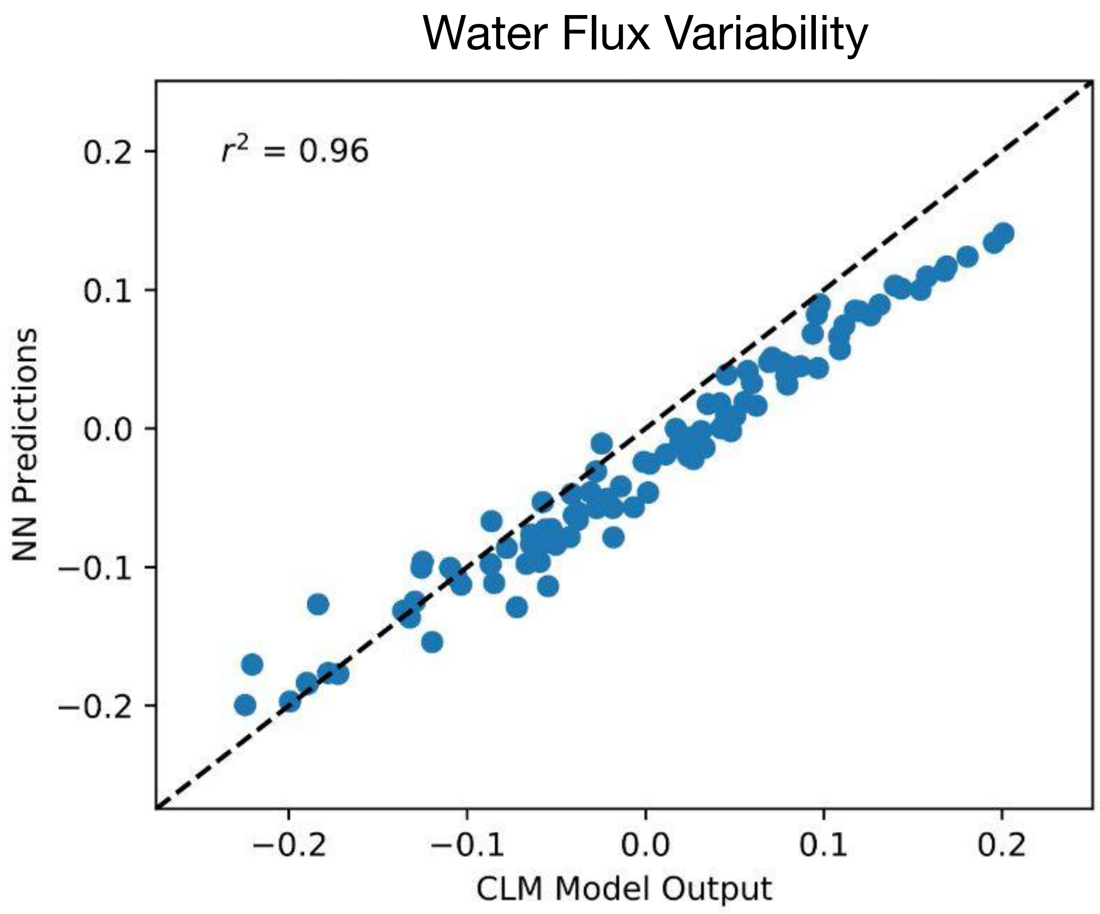
Thanks!
Questions?

djk2120@ucar.edu
kdagon@ucar.edu

BACKUP

Assessing Emulator Performance





"Best" emulator trained on original parameter values and model output.

Dagon et al. (2020)

Global Sensitivity Analysis

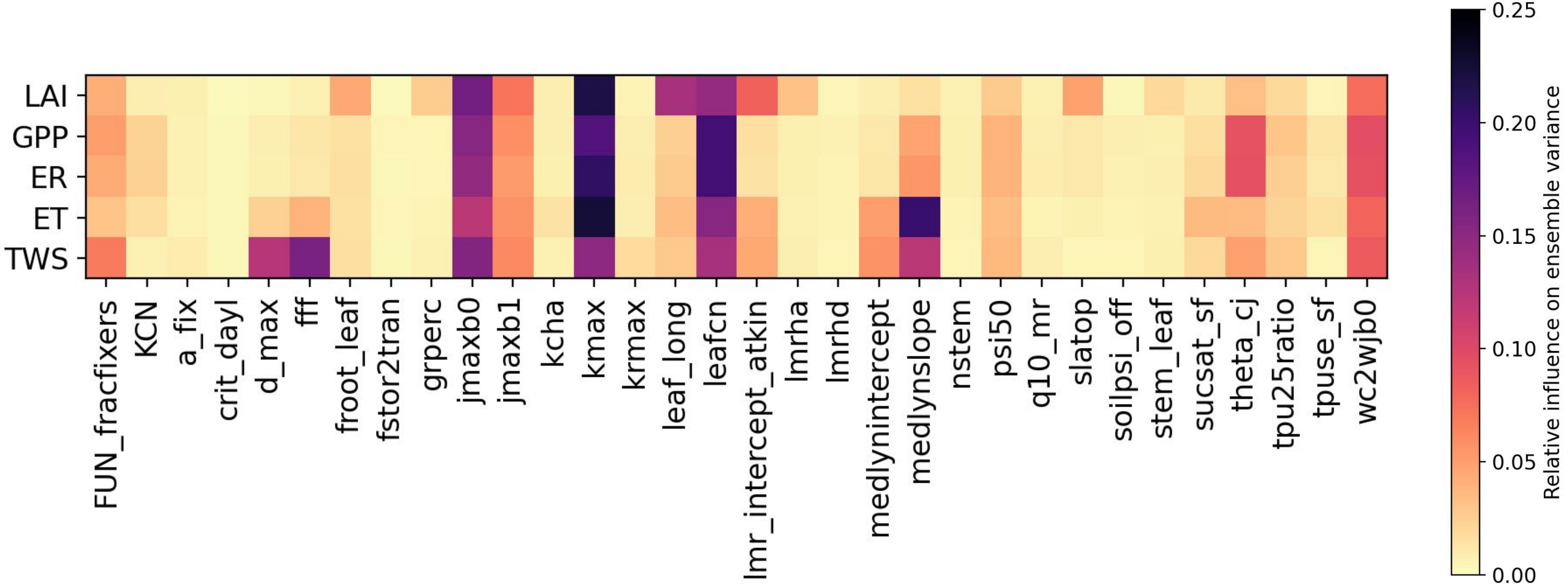


Figure from Linnia Hawkins

Global Sensitivity Analysis

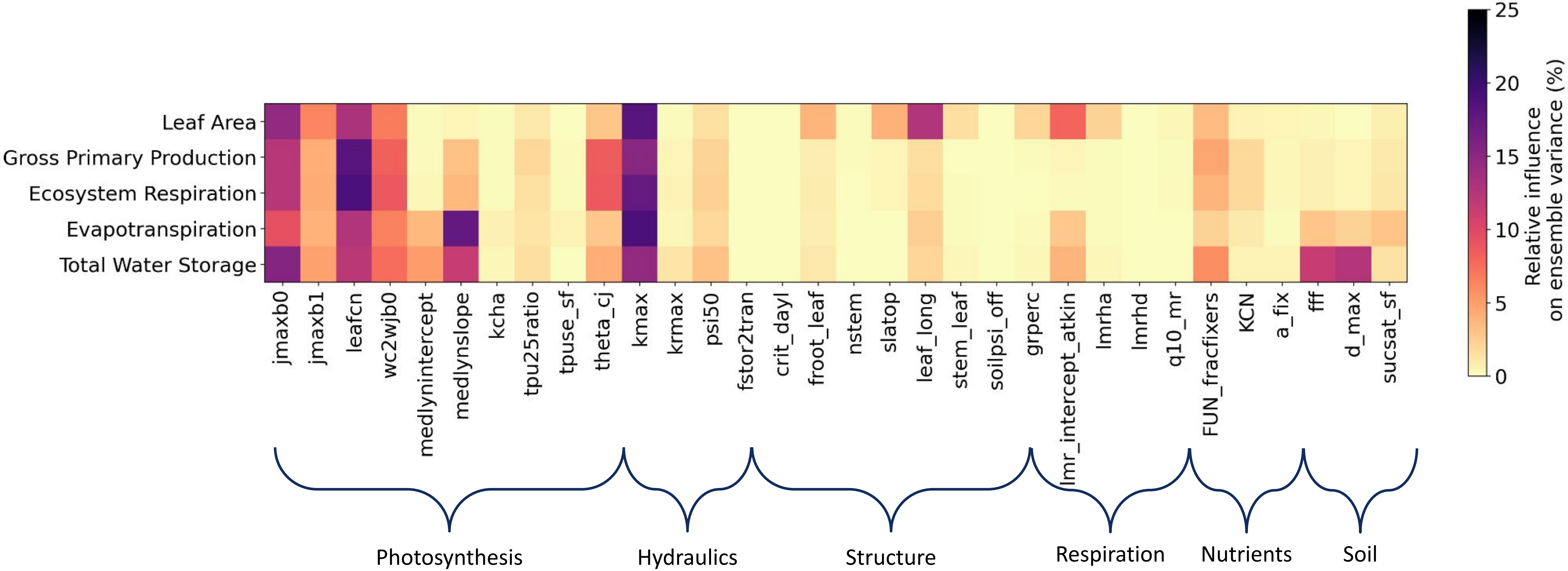
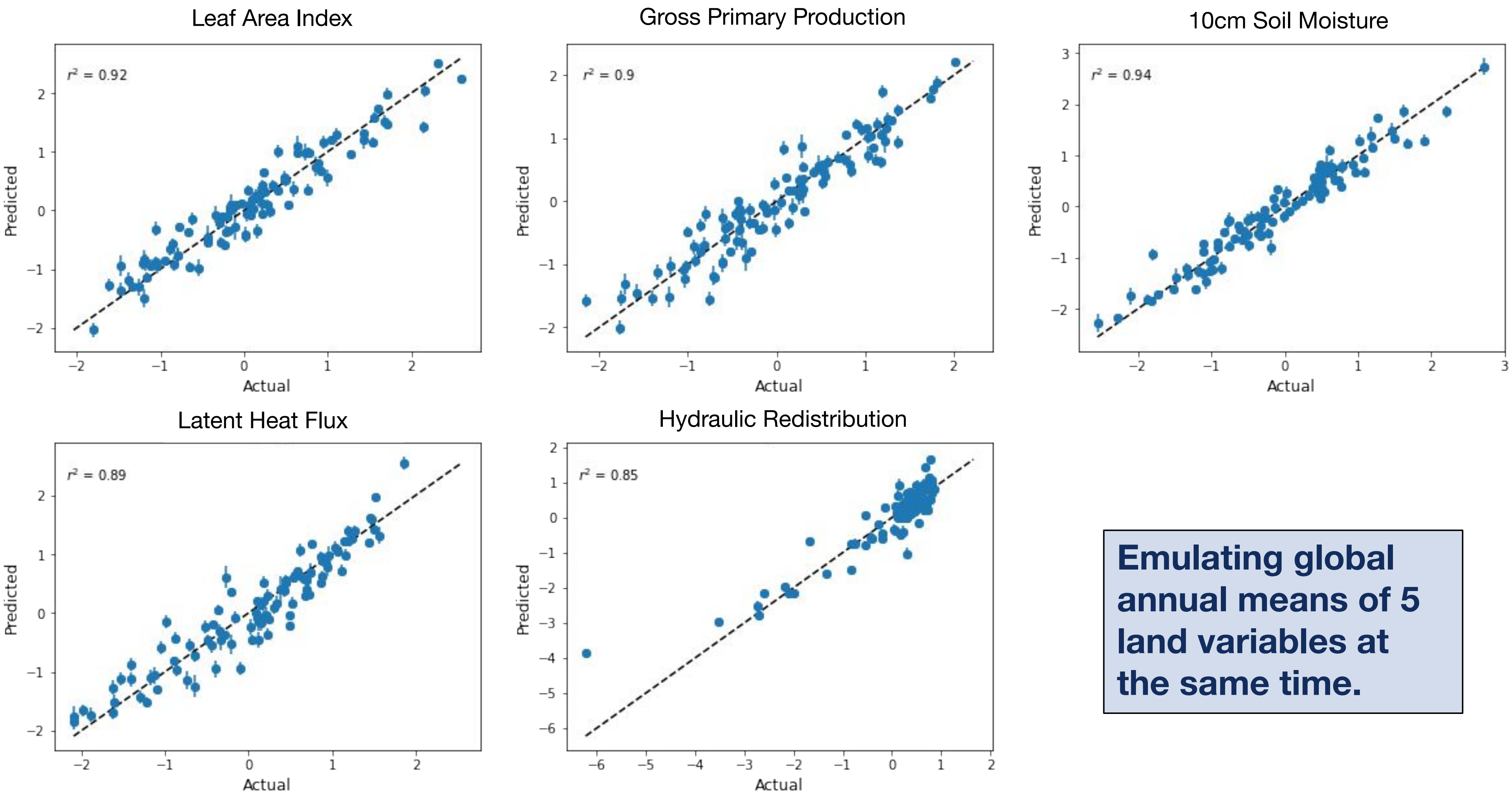


Figure from Linnia Hawkins

Gaussian Process Emulation



Emulating global annual means of 5 land variables at the same time.

PFT-Level Leaf Area Index