


LEAP

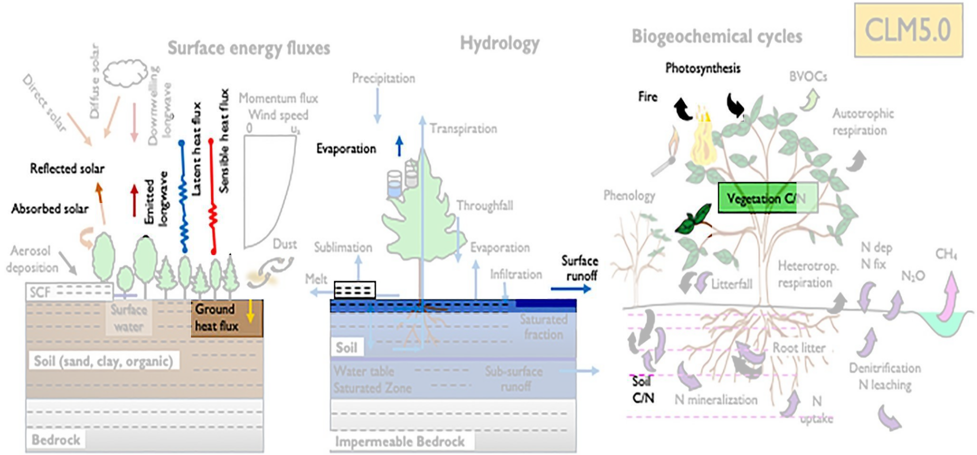
Tuning CLM6

Linnia Hawkins
Daniel Kennedy, Katie Dagon,
Dave Lawrence, Pierre Gentine
& CLM community

NSF | NCAR

Tuning CLM6 for CESM3

1) Align model output with observational constraints across diverse ecosystems.
2) Don't mess anything else up.



We only directly observe a small fraction of the processes simulated in CLM.

Adapted from Lawrence et al., (2019)

Constraining parametric uncertainty

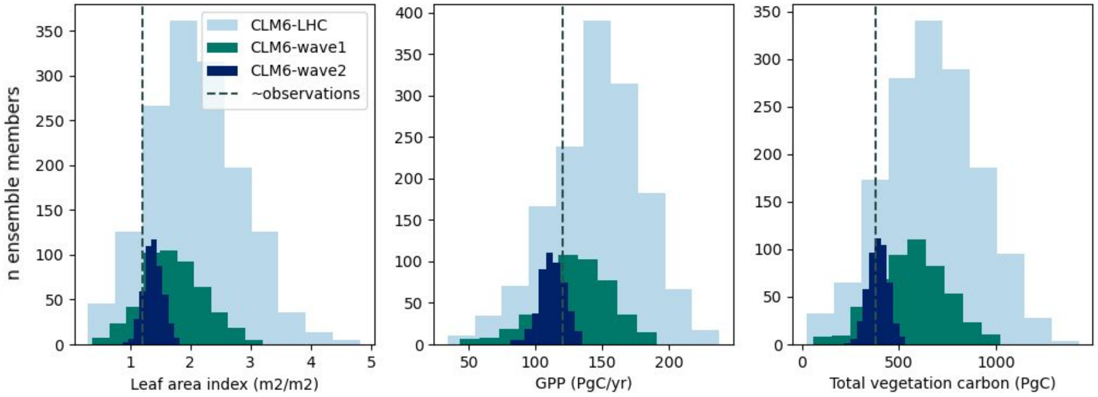
History Matching
- CLM6, 56 parameters

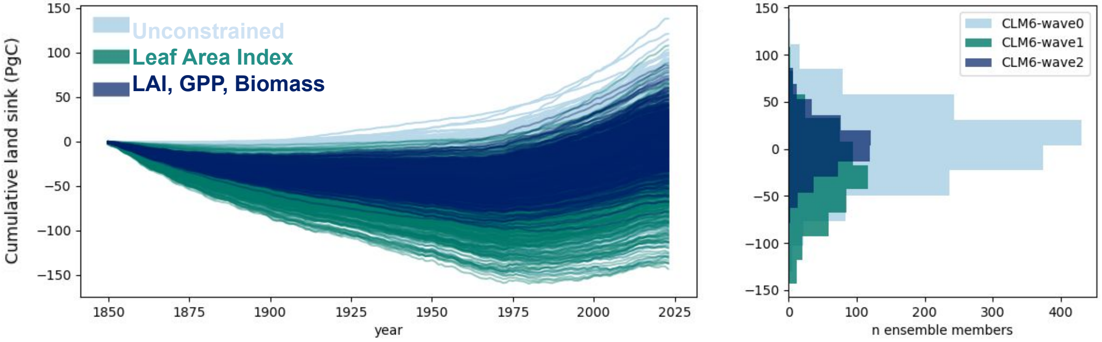
Wave 0: 1500 ensemble members
- Latin Hypercube sample
- Scaled PFT parameters

Wave 1: 500 ensemble members
- Independent PFT parameters
- Leaf Area Index

Wave 2: 500 ensemble members
- Independent PFT parameters
- LAI, GPP, VegC

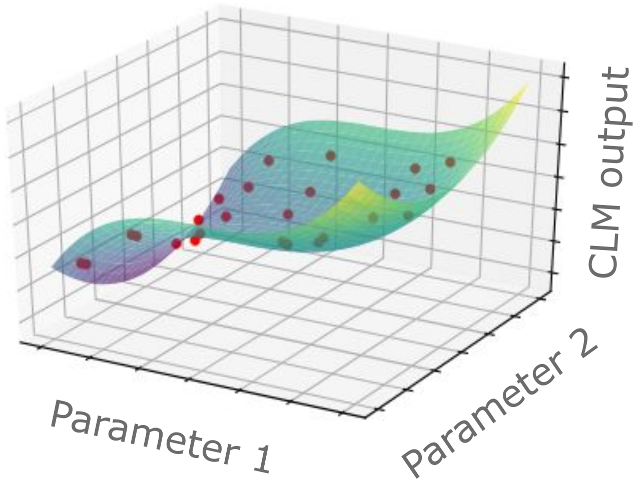
Wave 3: 500 ensemble members
- LAI, GPP, VegC, LH, SH





Emulators for Calibration

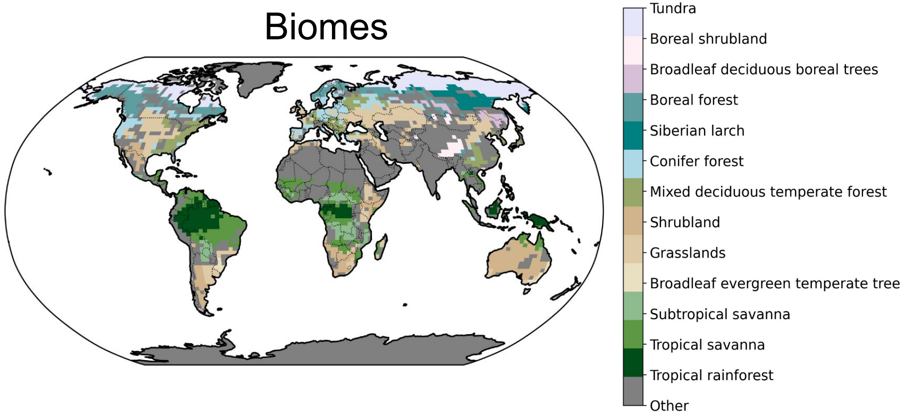
ML emulators
- Computationally efficient surrogate for CLM.
- Maps input parameters to output variables
- Differentiable

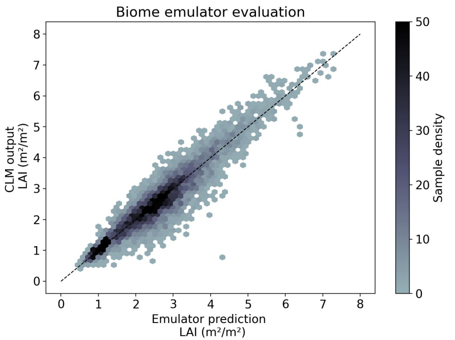


Emulators for Calibration

ML emulators
- Computationally efficient surrogate for CLM.
- Maps input parameters to output variables
- Differentiable

CLM emulators
- Training: 2500 ensemble members
- Independent PFT parameters
- Gaussian process regression
- Biome & variable specific





R^2 score

LAI	0.98	0.97	0.95	0.98	0.96	0.94	0.97	0.97	0.97	0.97	0.96	0.96	0.96
GPP	0.94	0.92	0.96	0.96	0.95	0.93	0.97	0.96	0.97	0.97	0.95	0.97	0.97
Veg C	0.96	0.97	0.94	0.95	0.93	0.94	0.97	0.97	0.96	0.96	0.95	0.85	0.93

Biomes

Calibration methods

	Objective	Output	Uncertainty quantification?	Uses gradients?	Scales well?
MCMC	Sampling proportional to likelihood × prior	Ensemble of samples from posterior	Yes	No (unless using HMC)	No
Numerical optimization	Minimizing a loss or cost function	Point estimate (best-fit parameters)	No	Depends on method	Depends on method
Gradient based optimization (ADAM)	Minimizing a loss or cost function	Point estimate (best-fit parameters)	No	Yes	Yes

$$\theta \rightarrow Emulator(\theta) \rightarrow variable \rightarrow \mathcal{L}oss \qquad \nabla_\theta \mathcal{L}$$

Calibrating CLM6 for CESM3

Gradient-based optimization (ADAM)

- Differentiable emulator and loss function constructed in TensorFlow enables end-to-end gradient computation.

Observational data (ILAMB)

- LAI, GPP, Biomass, Latent heat, Sensible heat

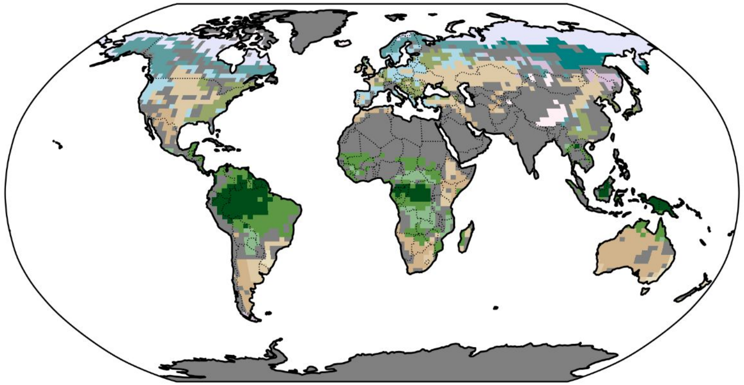
195 parameters

- Independent PFT parameters

Optimization

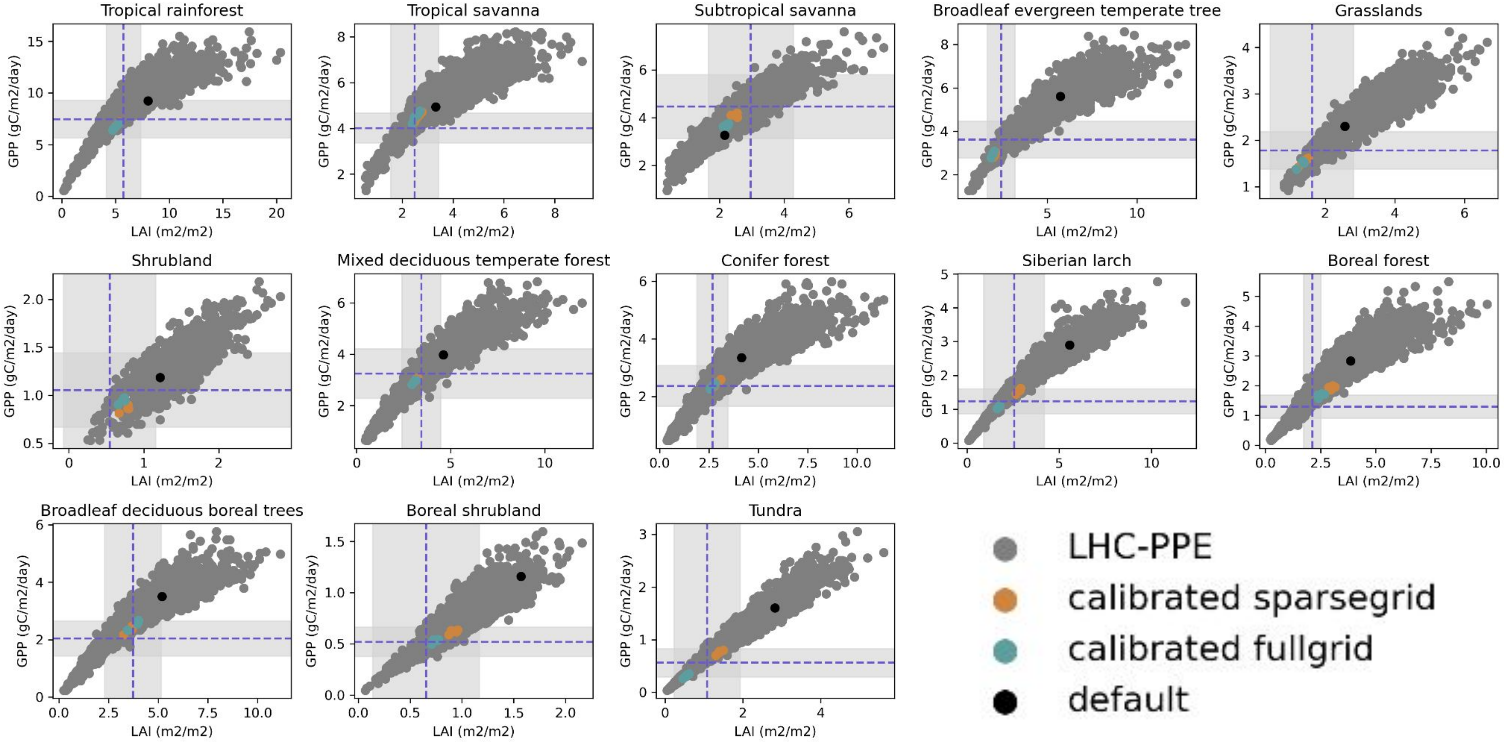
- Initialize from random sample close to default.
- Penalty for moving away from default parameters
- Terms stop contributing to loss once z-score < 1

Biomes

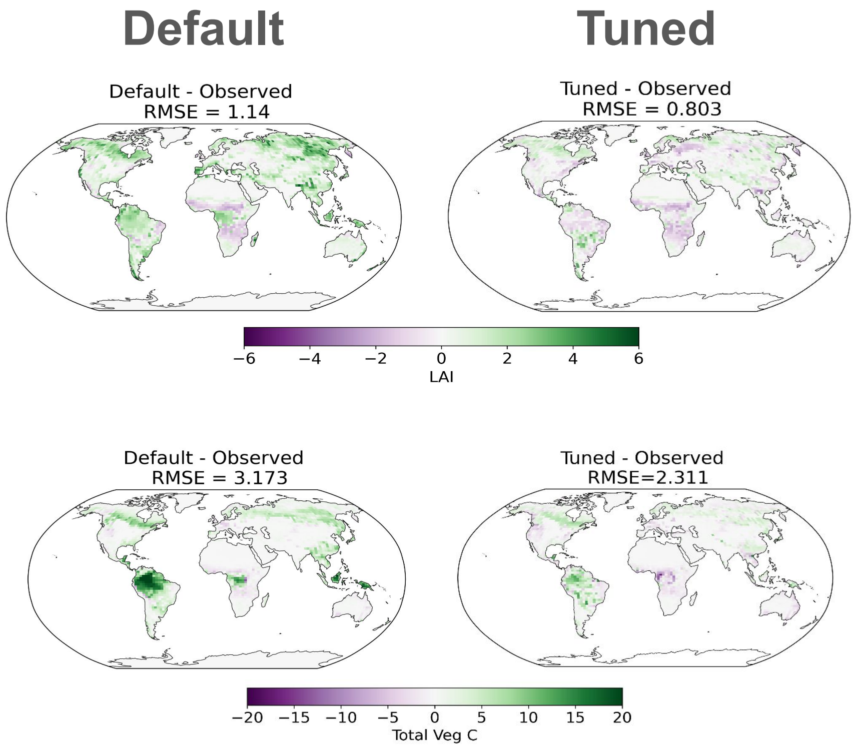


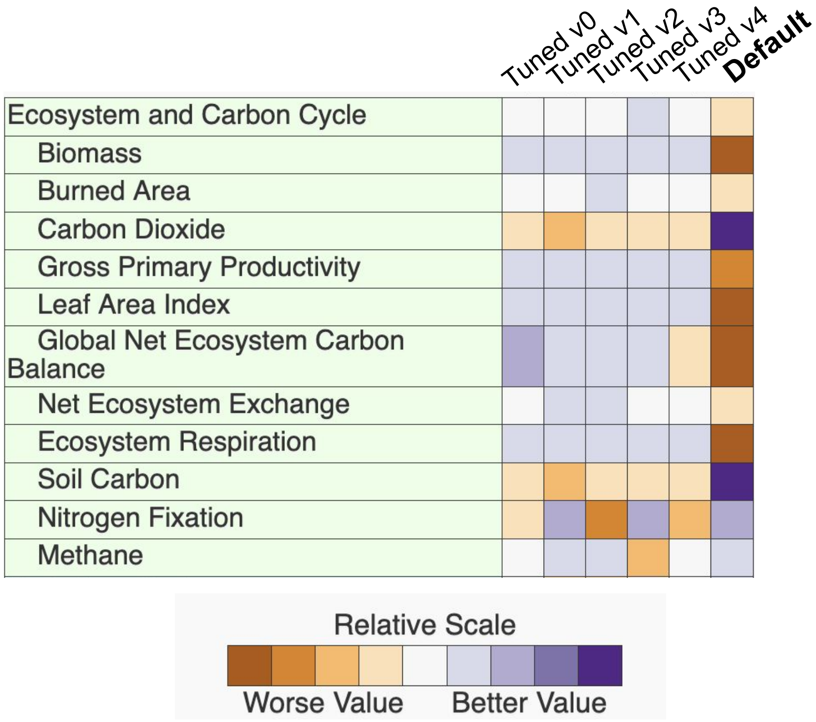
$$loss = \overset{variables}{\sum_{v}} \overset{biomes}{\sum_{b}} \left(Z_{v,b}^{2}\right) + penalty$$

Preliminary results



Preliminary results





Challenges

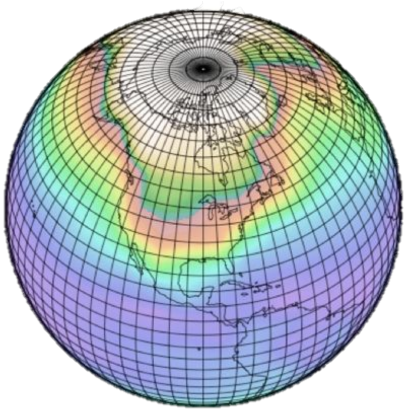
- Metric alignment
- Evolving code base
- Different applications: site, regional, global, resolution, science questions, etc.
- Feedbacks when coupled to atmosphere

Idealized future system

CLMulator

Training data:

- Reanalysis
- PI, Historical, Future
- Coupled
- Variable sparse grids



Application:

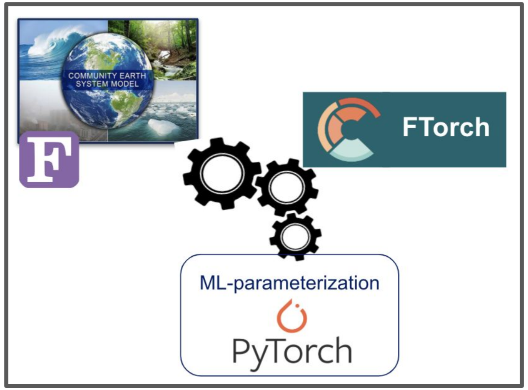
User: parameters, domain, metrics

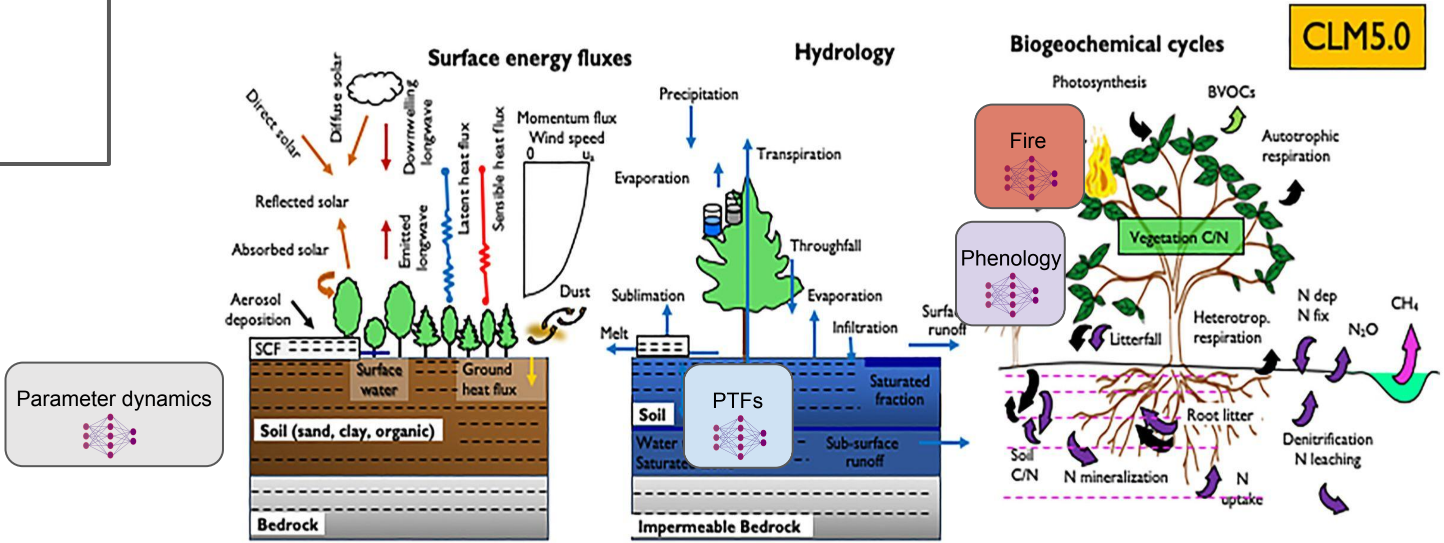
Returns: predictions, sensitivity,
calibrated parameter sets

New parameterization:

- Small ensemble
- Fine tune

Hybrid CLM: *ML parameterizations in development*





Adapted from Lawrence et al., (2019)

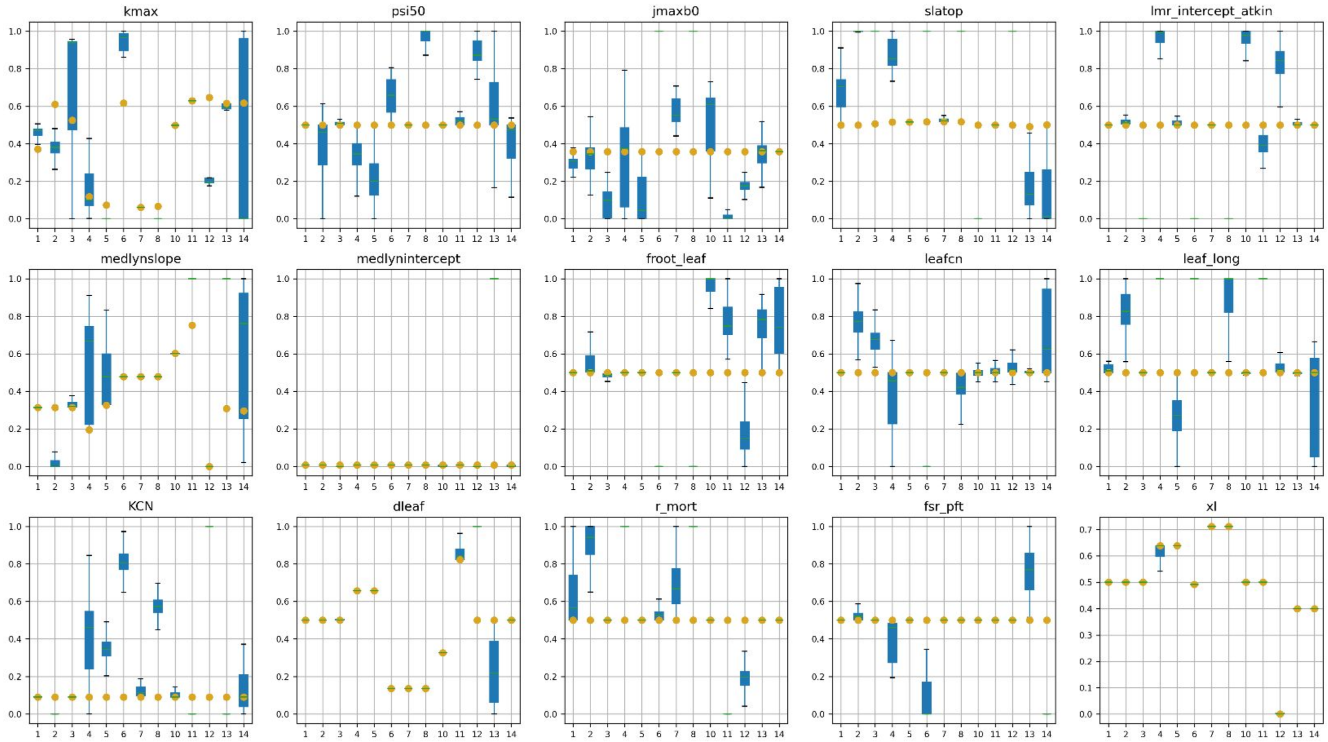
Summary

- **Constraining parametric uncertainty** in the strength of the land carbon sink to reduce overall uncertainty in climate change projections.

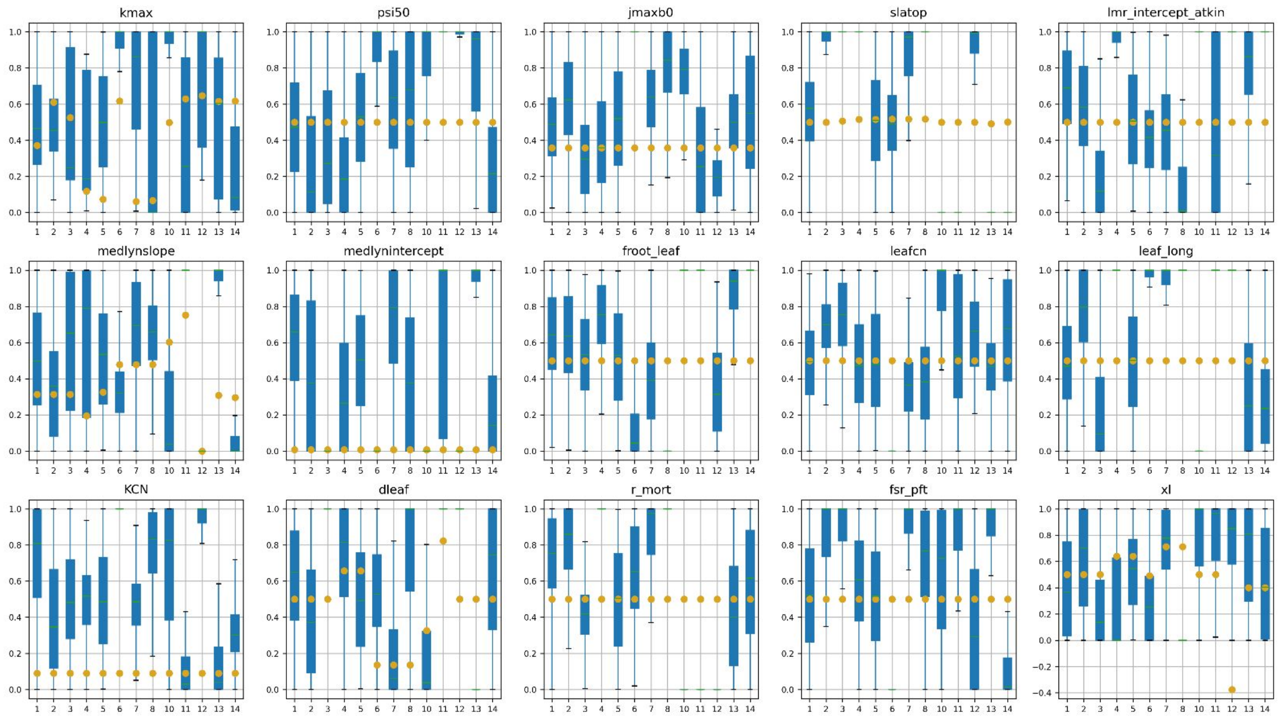
- **Calibration**: Emulator based calibration can be successful, and there are opportunities for increased flexibility.

- **CESM3-MLe:** CLM has capacity to support the implementation of ML-parameterizations.

Parameter posteriors: *stay close to default*



Parameter posteriors: *Best possible optimization*



LEAP

High equifinality

Recover a known parameter set.

Target: LAI
Error < 0.001