Parameter Uncertainty and Estimation in CLM5

Katie Dagon

ASP Postdoc, NCAR

CTSM Tutorial

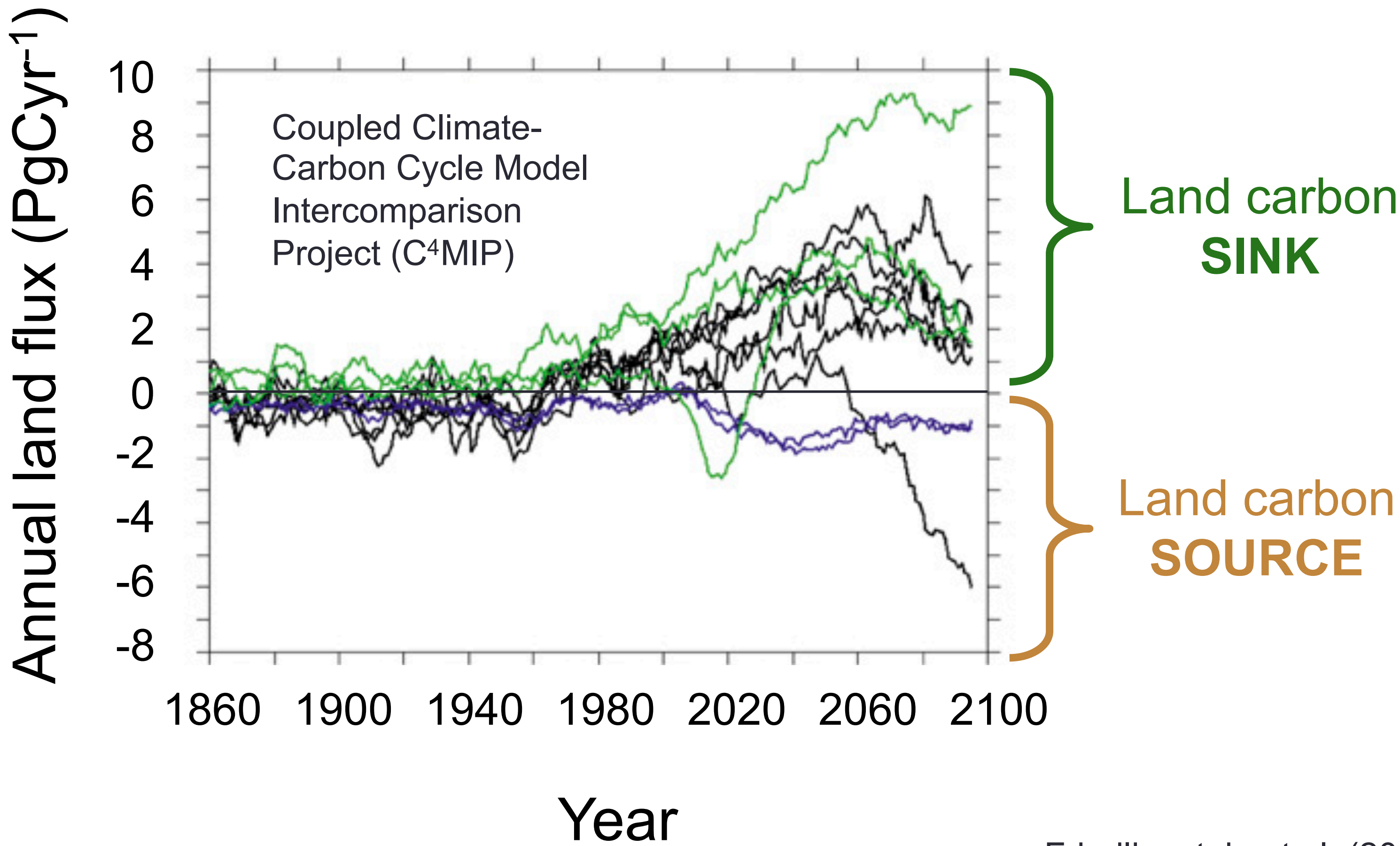
February 7, 2019

NCAR UCAR | ASP Advanced Study Program

NSF

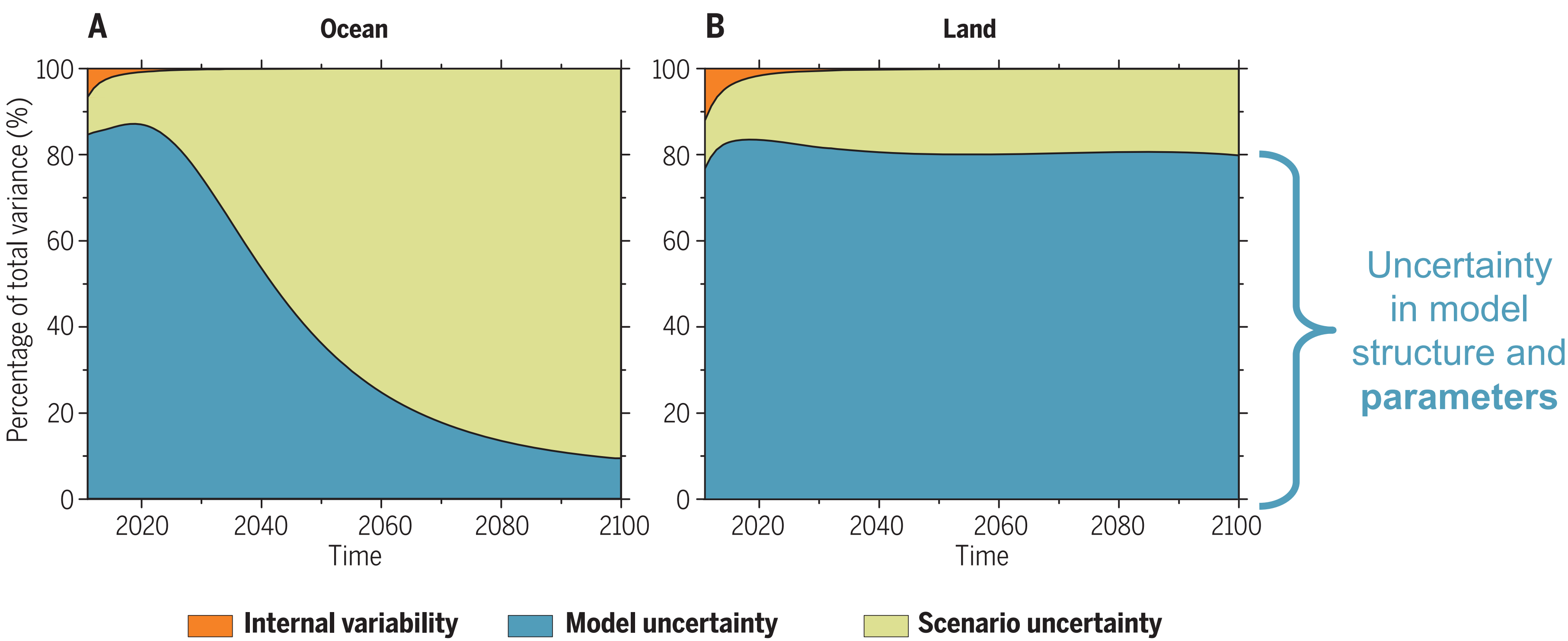
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

What is driving uncertainty in land surface model projections of climate change?



Friedlingstein et al. (2014)

What are the sources of uncertainty in land surface models?



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

What role do **parameter choices** play in overall **land model uncertainty**?

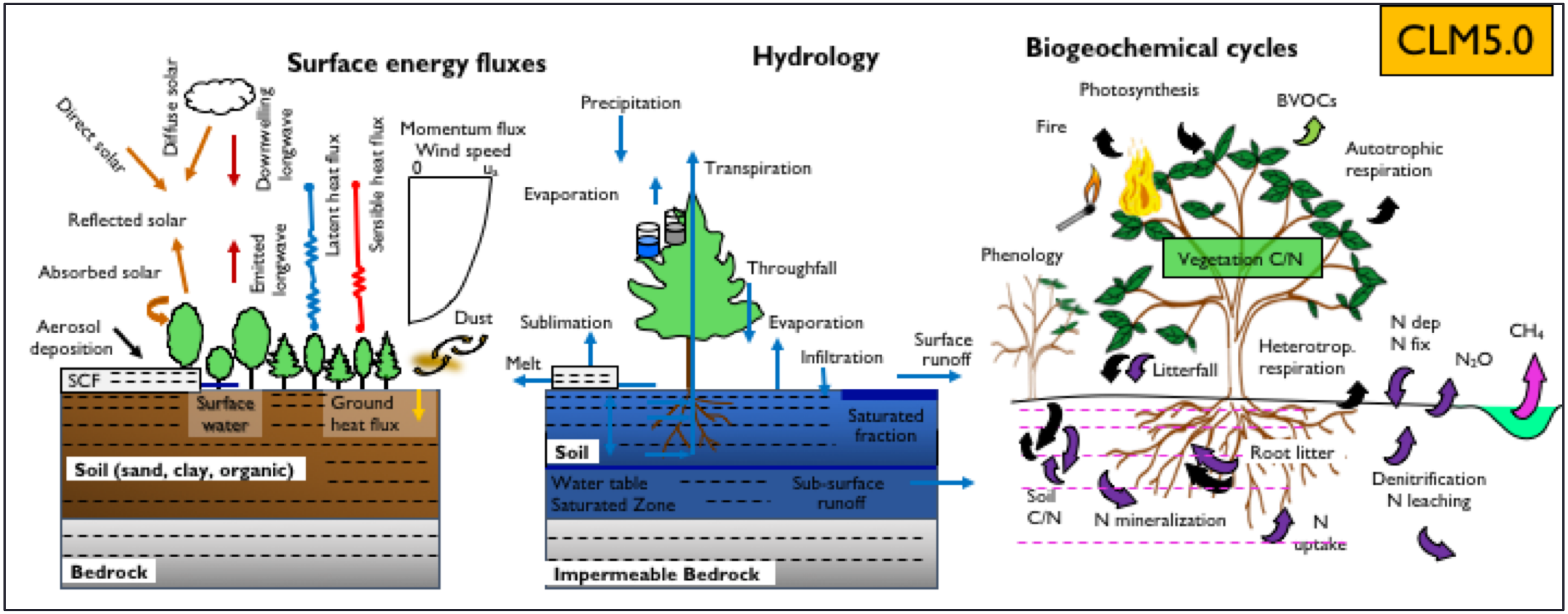
What role do **parameter choices** play in overall **land model uncertainty**?

1. Assessing parameter sensitivity through one-at-a-time changes in parameter values.

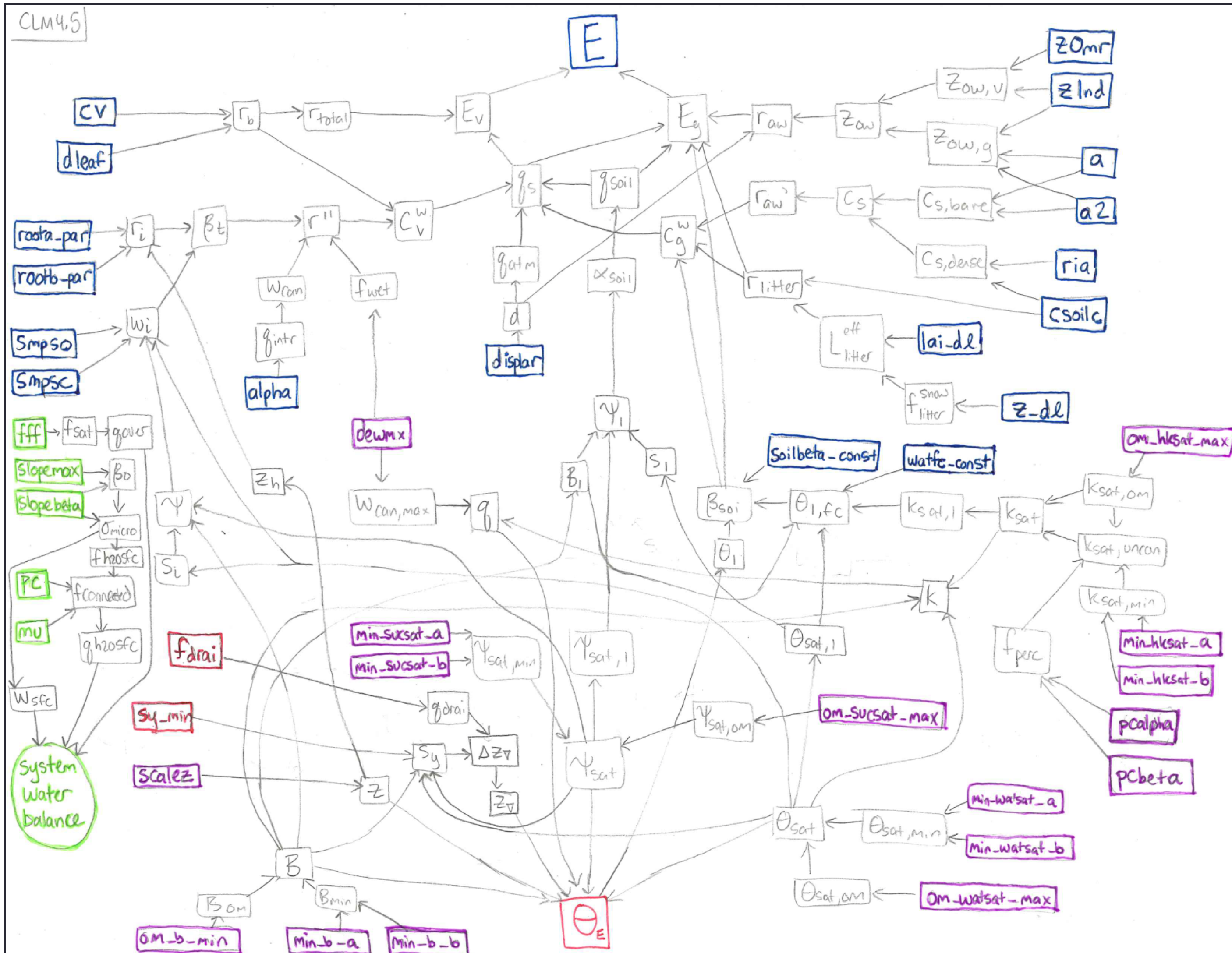
What role do **parameter choices** play in overall **land model uncertainty**?

1. Assessing parameter sensitivity through one-at-a-time changes in parameter values.
2. Using machine learning to emulate CLM and estimate parameter values.

Selecting land model parameters

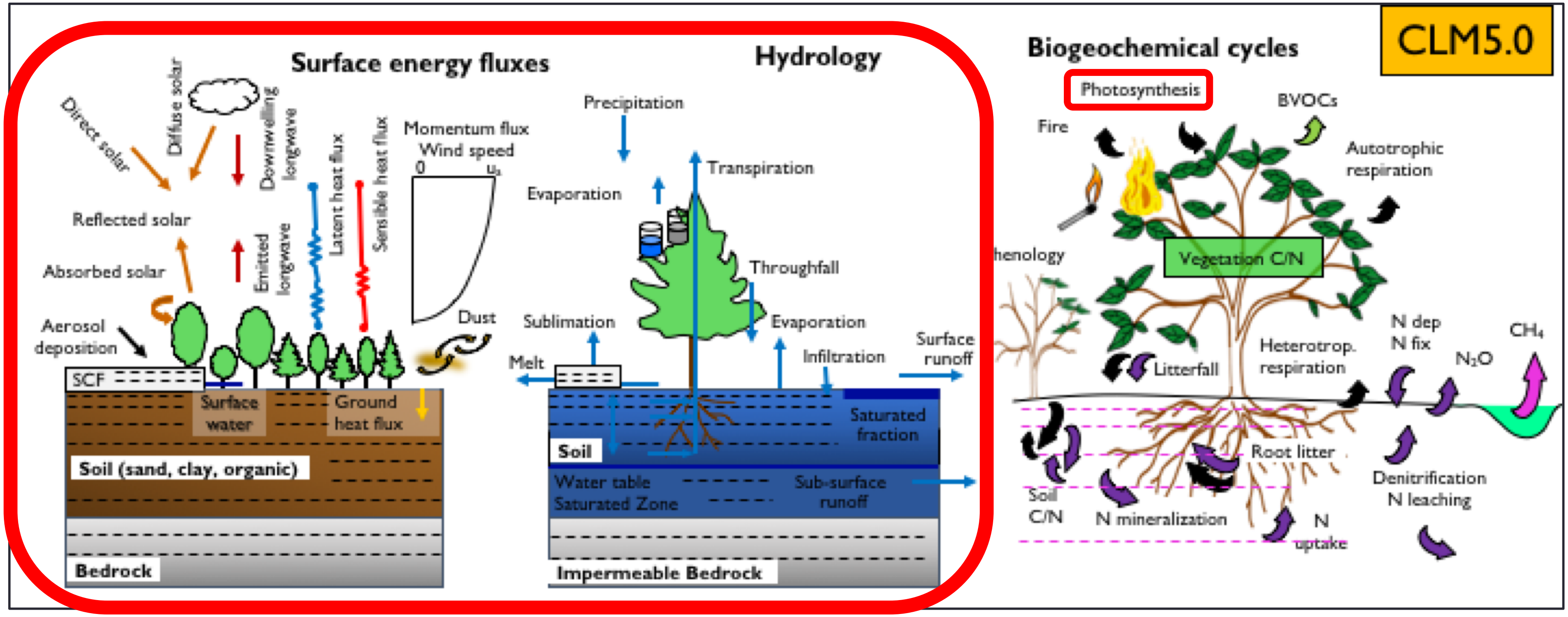


Evapotranspiration



Soil Moisture

Focus on CLM5 biogeophysical processes



Example parameter: **medlynslope**

This parameter represents the slope of the stomatal conductance – photosynthesis relationship.

From the CLM5 Documentation:

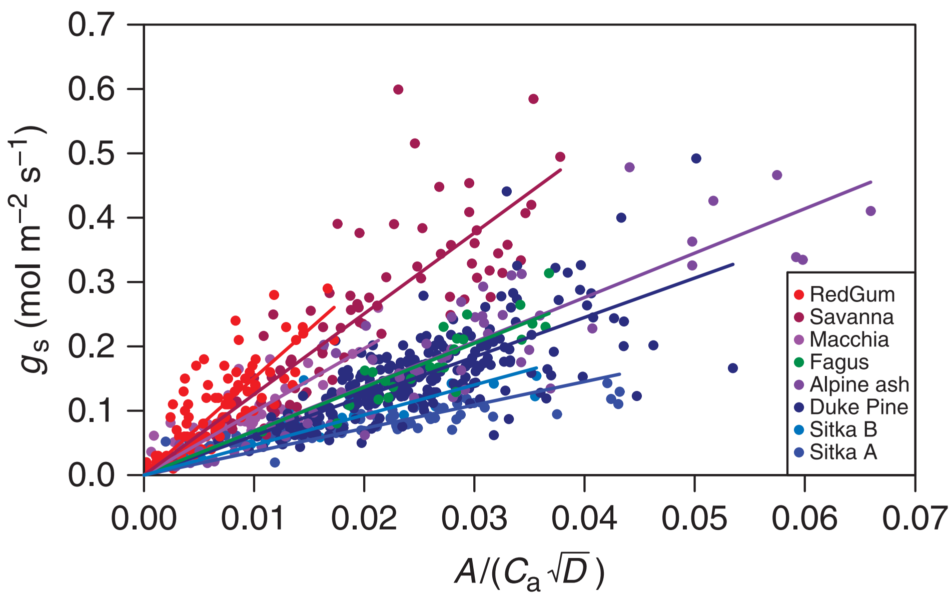
$$g_s = g_o + 1.6\left(1 + \frac{g_1}{\sqrt{D}}\right)\frac{A_n}{c_s/P_{atm}}$$

stomatal conductance

medlynslope

photosynthesis

- During day 2 practical, we experimented with a 10% decrease in this parameter.

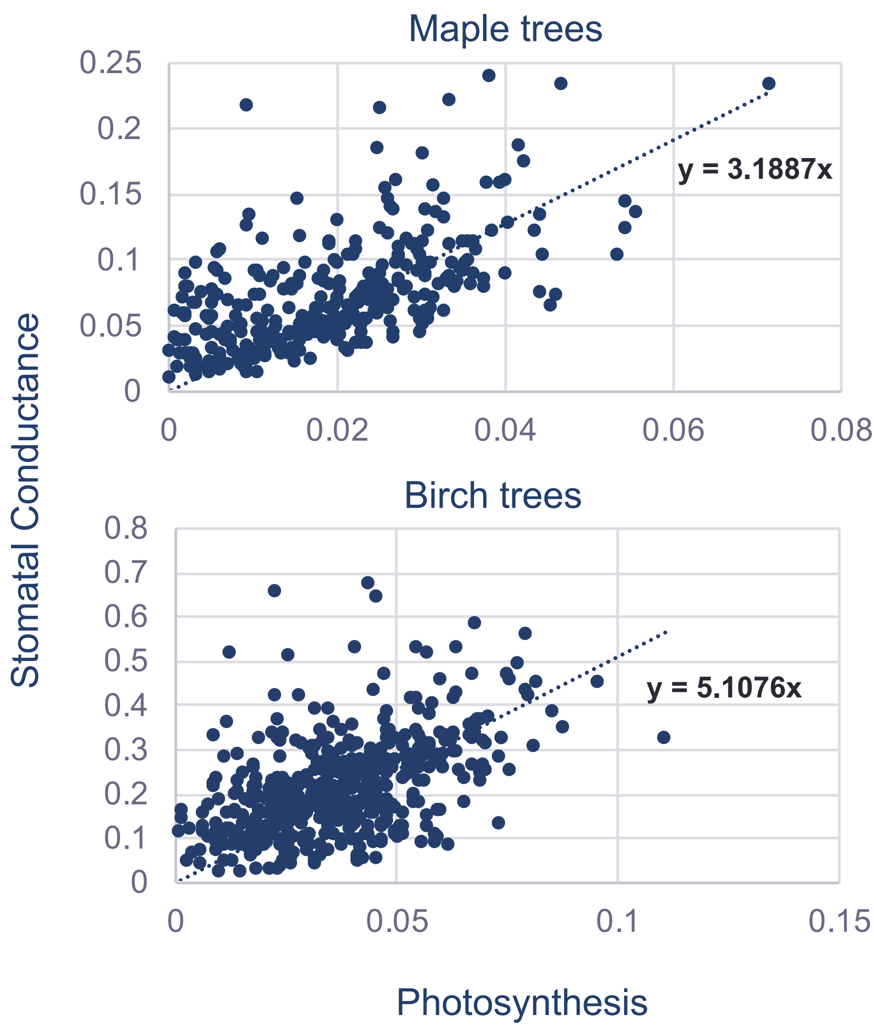


Medlyn et al. (2011)

Where do uncertainty ranges come from?

Observations of photosynthesis and stomatal conductance contribute to PFT-dependent **uncertainty range** for *medlynslope* parameter.

<ins>PFT = broadleaf deciduous trees</ins>
CLM default value = 4.45
Minimum = 3.1887
Maximum = 5.1076



Data from Lin et al. (2015)

Parameter Sensitivity Simulation Setup

- Which compset?
  - CLM5 with year 2000 forcing and satellite phenology (SP) mode
- What resolution?
  - 4x5 to speed up simulation time
- Simulation length?
  - 20 years total, sample last 5 years after 15 years of spin-up

Parameter Sensitivity Simulation Setup

- Which compset?
  - CLM5 with year 2000 forcing and satellite phenology (SP) mode
- What resolution?
  - 4x5 to speed up simulation time
- Simulation length?
  - 20 years total, sample last 5 years after 15 years of spin-up

```
./create_newcase --case $CASENAME --compset I2000Clm50Sp --res f45_f45

./xmlchange STOP_N=20
./xmlchange STOP_OPTION=nyears
```

Parameter Sensitivity Simulation Setup

- One-at-a-time parameter changes for 34 parameters, testing the min/max of their uncertainty ranges
  - 10 PFT-dependent parameters
  - 3 namelist parameters
  - 21 hard-coded parameters

Parameter Sensitivity Simulation Setup

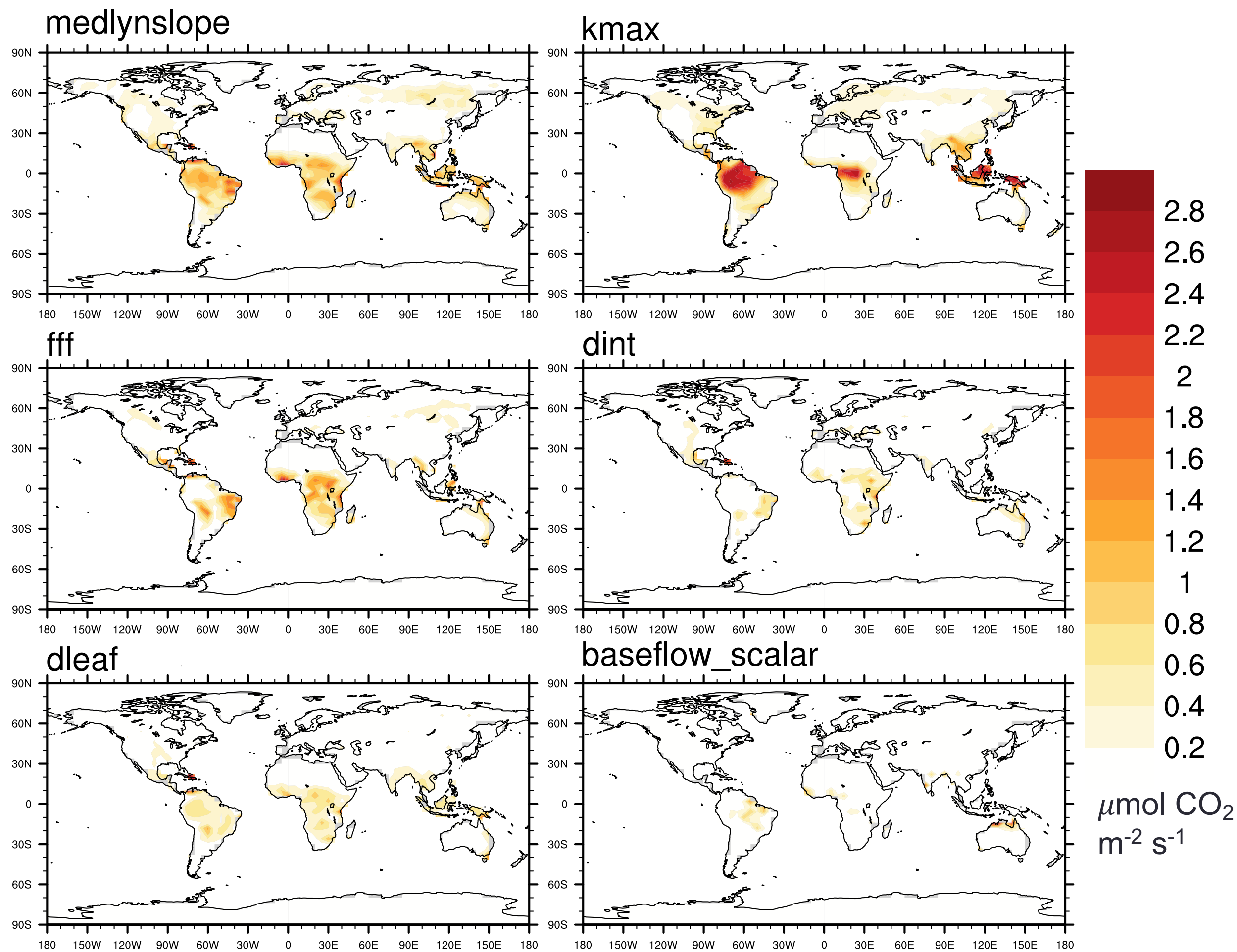
- One-at-a-time parameter changes for 34 parameters, testing the min/max of their uncertainty ranges
  - 10 PFT-dependent parameters `[modify params file]`
  - 3 namelist parameters `[make changes in user_nl_clm]`
  - 21 hard-coded parameters `[SourceMods]`

Parameter Sensitivity Simulation Setup

- 7 model outputs to assess sensitivity
  1. Gross Primary Productivity (GPP) → `FPSN`
  2. Evapotranspiration (ET) → `QFLX_EVAP_TOT`
  3. Transpiration Fraction = Transpiration/ET → `QVEGT/QFLX_EVAP_TOT`
  4. Sensible Heat Flux → `FSH`
  5. 10cm Soil Moisture → `SOILWATER_10CM`
  6. Total Column Soil Moisture → `SOILLIQ + SOILICE`
  7. Water Table Depth → `ZWT`

Assessing parameter sensitivity

Sensitivity* of gross primary productivity (GPP) to parameter perturbations



*Sensitivity = $|GPP_{max} - GPP_{min}|$

Ranked sensitivity to 7 outputs
Average rank across outputs

Using machine learning to emulate CLM and estimate parameter values

Using machine learning to emulate CLM and estimate parameter values

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

How can we speed this process up?

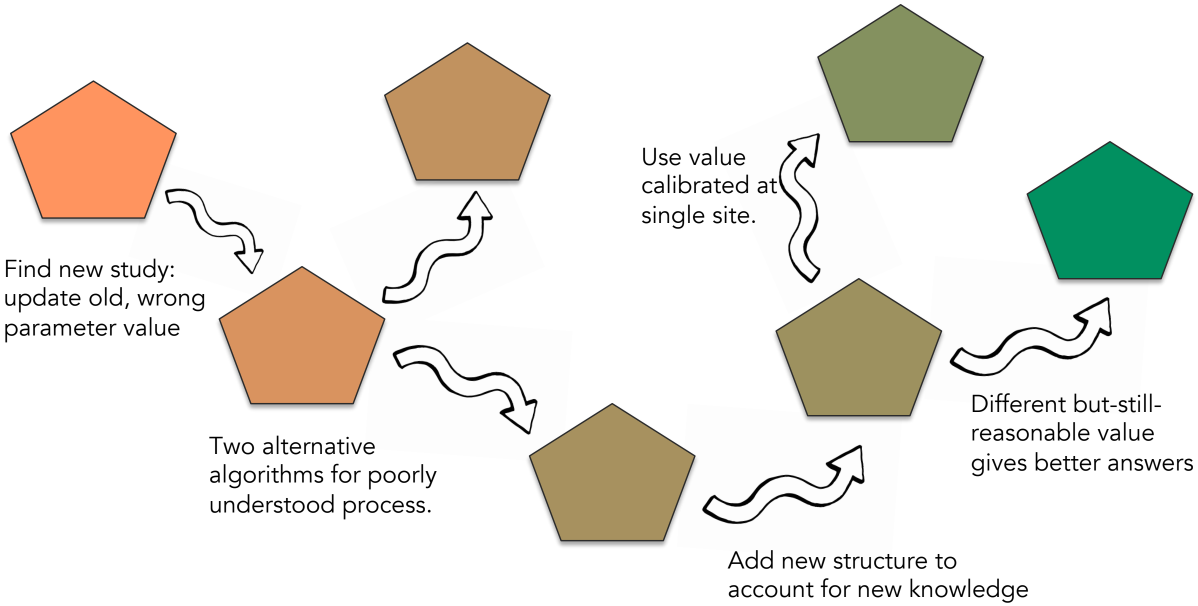


Figure from Rosie Fisher

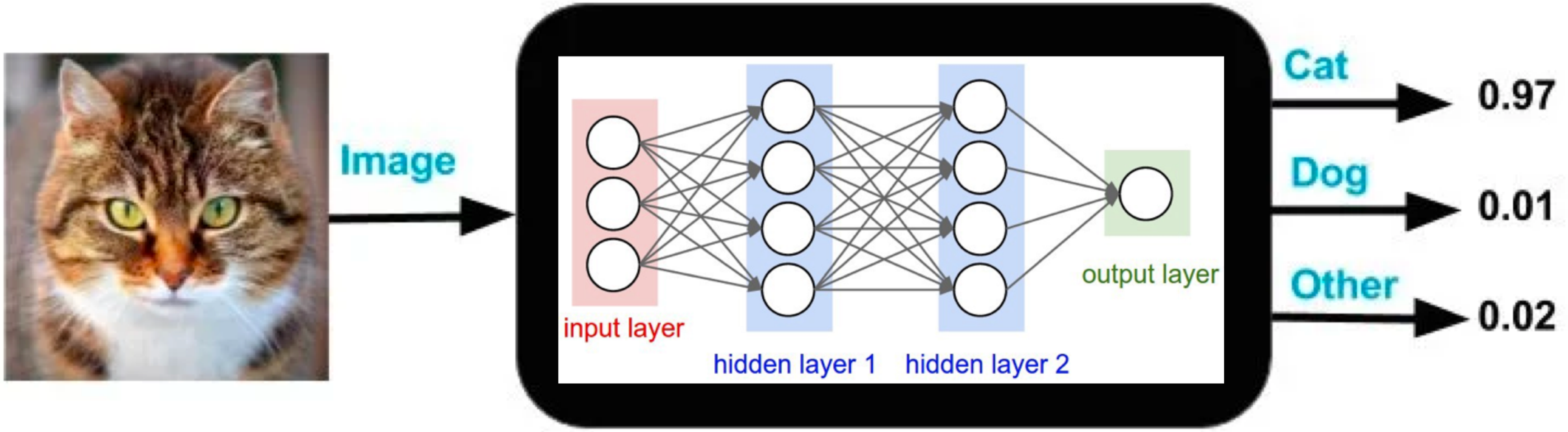
Using machine learning to emulate CLM and estimate parameter values

- Machine learning goals:
  1. Build and train a series of neural networks to predict CLM output, given parameter values as input.
  2. Inflate ensemble size of possible parameter combinations using trained networks.
  3. Compare network predictions with observations to **estimate parameter values**.

Machine learning by neural networks

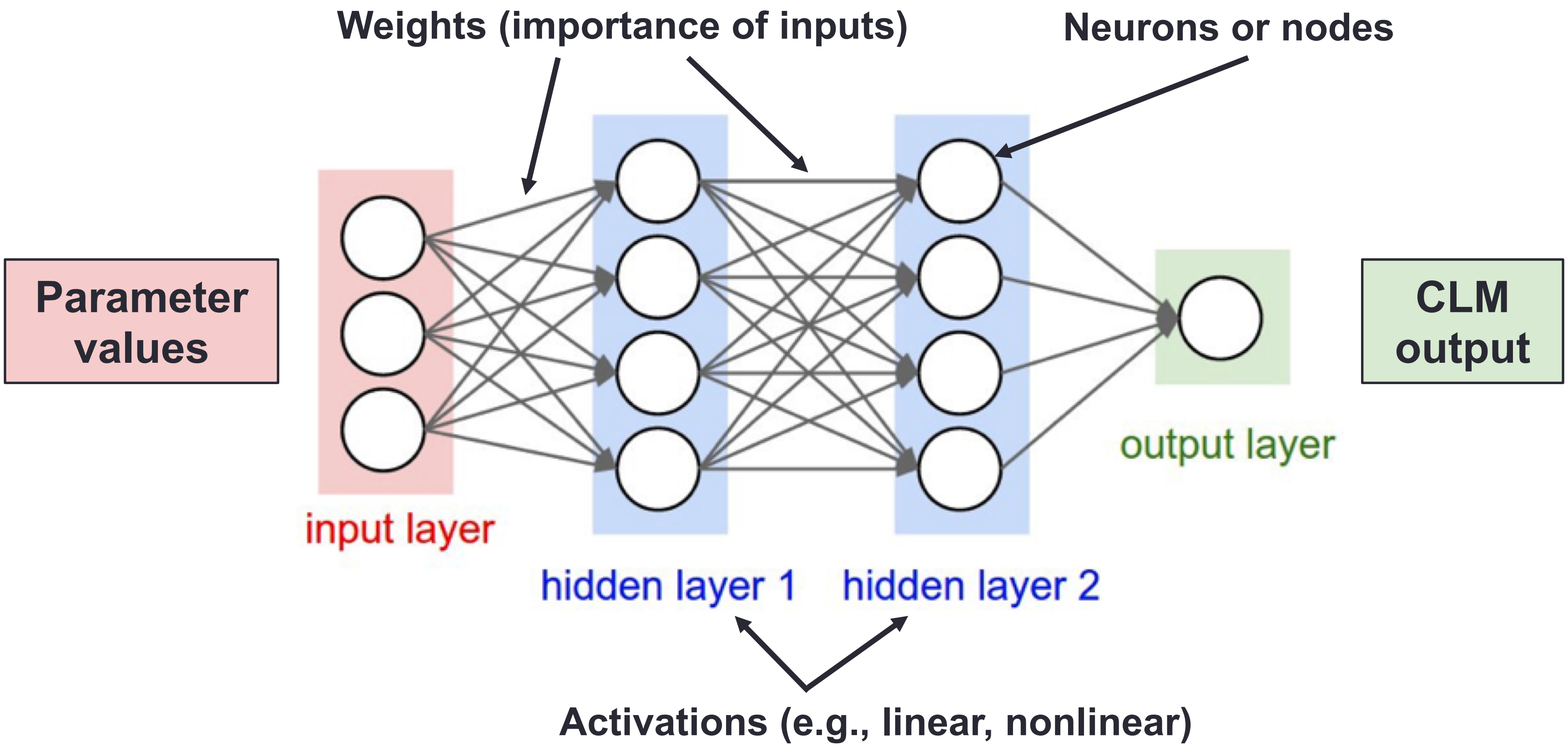
Network image: https://www.learnopencv.com/neural-networks-a-30000-feet-view-for-beginners/

Machine learning by neural networks



Network image: http://cs231n.github.io/neural-networks-1/

Machine learning by neural networks



Perturbed Parameter Ensemble Setup

- 100 randomly sampled parameter values from uncertainty ranges for 6 candidate parameters

Parameter values

Parameters

Simulations

	P1	P2	P3	P4	P5	P6
S1	x1,1	x1,2	x1,3	x1,4	x1,5	x1,6
S2	x2,1	x2,2	x2,3	x2,4	x2,5	x2,6
S3	x3,1	x3,2	x3,3	x3,4	x3,5	x3,6
…	…	…	…	…	…	…
S100	x100,1	x100,2	x100,3	x100,4	x100,5	x100,6

Perturbed Parameter Ensemble Setup

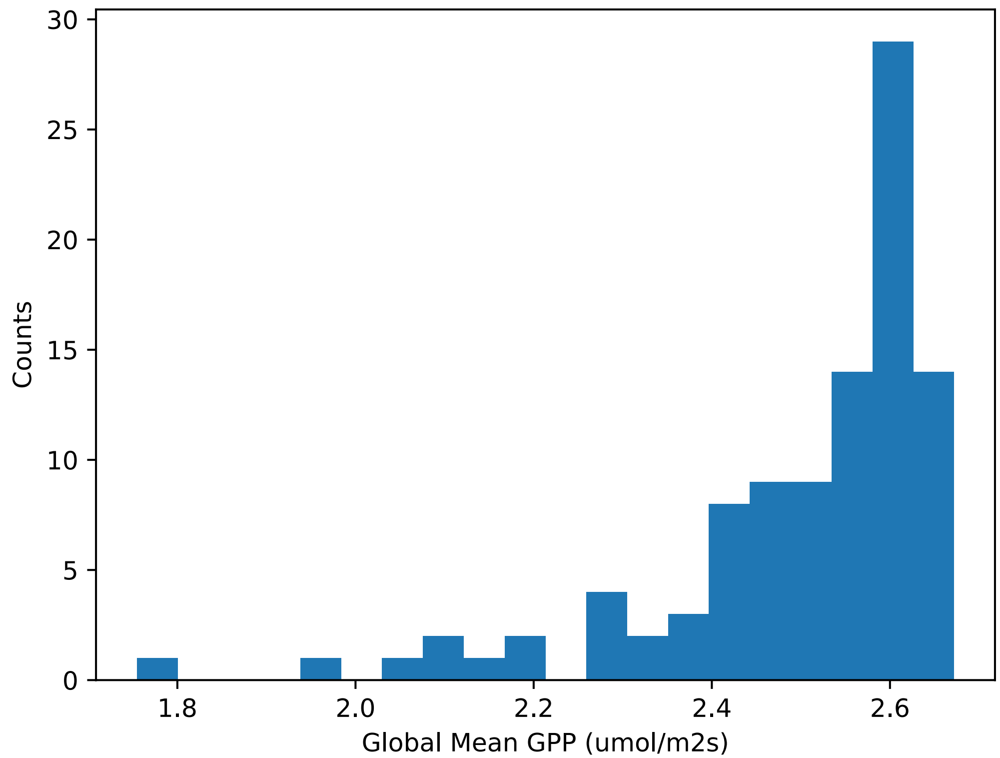
- 100 randomly sampled parameter values from uncertainty ranges for 6 candidate parameters
- CLM5 with year 2000 forcing and satellite phenology mode, 4x5 resolution, 20 years total, sample last 5 years

Perturbed Parameter Ensemble Setup

- 100 randomly sampled parameter values from uncertainty ranges for 6 candidate parameters
- CLM5 with year 2000 forcing and satellite phenology mode, 4x5 resolution, 20 years total, sample last 5 years
- Begin training neural network on output from 100 PPE simulations

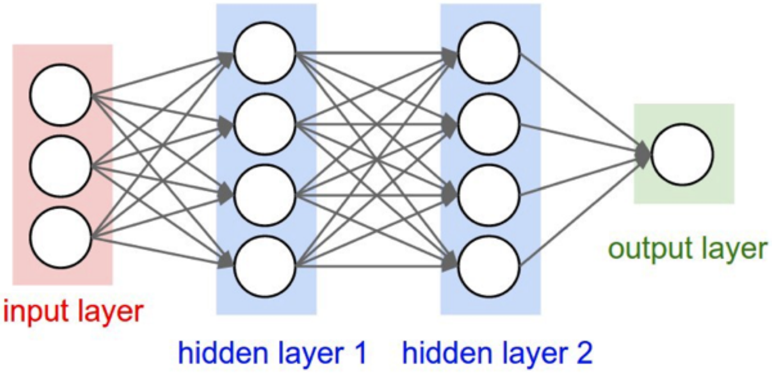
Begin training on simple global mean metric

Distribution of global mean gross primary productivity (GPP, μmol $m^{-2}s^{-1}$)



CLM output

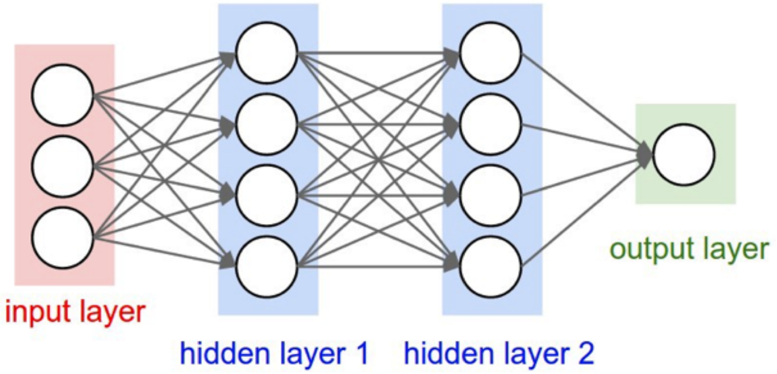
Build and train a neural network to predict land model output based on parameter values



Build and train a neural network to predict land model output based on parameter values

Input (100 parameter sets; 6 parameters)

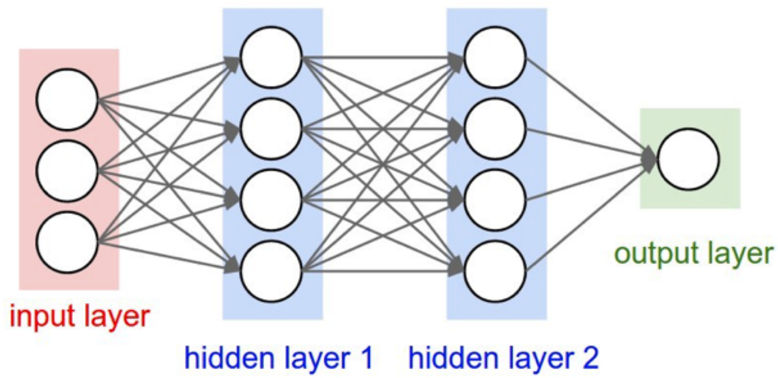
	P1	P2	P3	P4	P5	P6
S1	x1,1	x1,2	x1,3	x1,4	x1,5	x1,6
S2	x2,1	x2,2	x2,3	x2,4	x2,5	x2,6
S3	x3,1	x3,2	x3,3	x3,4	x3,5	x3,6
…	…	…	…	…	…	…
S100	x100,1	x100,2	x100,3	x100,4	x100,5	x100,6



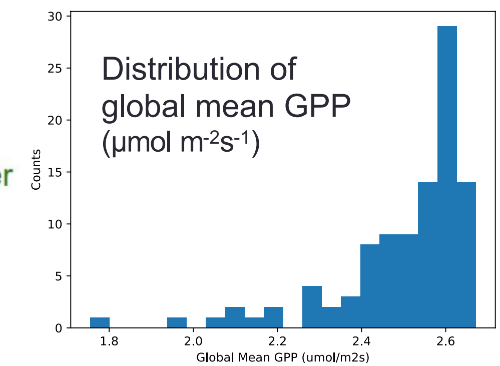
Build and train a neural network to predict land model output based on parameter values

Input (100 parameter sets; 6 parameters)

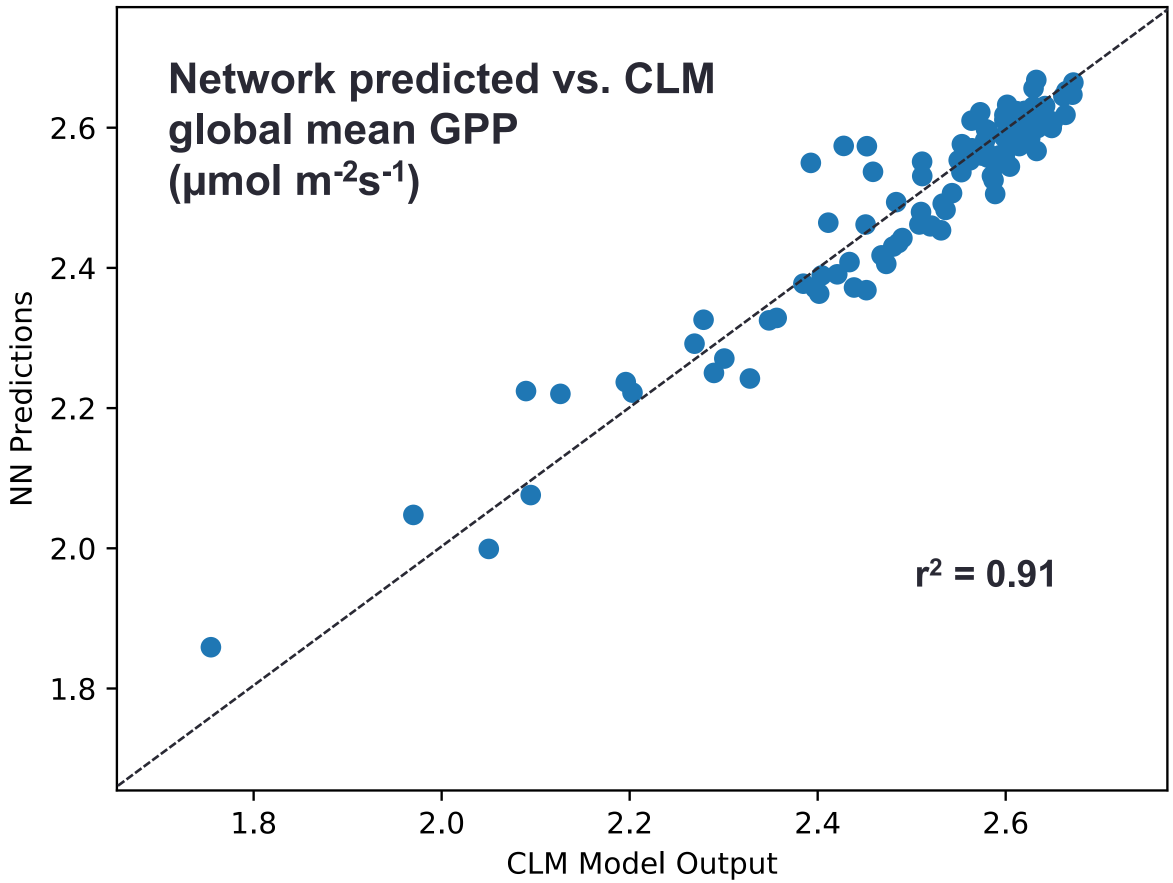
	P1	P2	P3	P4	P5	P6
S1	x1,1	x1,2	x1,3	x1,4	x1,5	x1,6
S2	x2,1	x2,2	x2,3	x2,4	x2,5	x2,6
S3	x3,1	x3,2	x3,3	x3,4	x3,5	x3,6
…	…	…	…	…	…	…
S100	x100,1	x100,2	x100,3	x100,4	x100,5	x100,6



Output (100 CLM simulations)



Assessing network performance



Climate model emulation

Input (1000 parameter sets; 6 parameters)

	P1	P2	P3	P4	P5	P6
S1	x1,1	x1,2	x1,3	x1,4	x1,5	x1,6
S2	x2,1	x2,2	x2,3	x2,4	x2,5	x2,6
S3	x3,1	x3,2	x3,3	x3,4	x3,5	x3,6
…	…	…	…	…	…	…
…	…	…	…	…	…	…
S1000	x1000,1	x1000,2	x1000,3	x1000,4	x1000,5	x1000,6

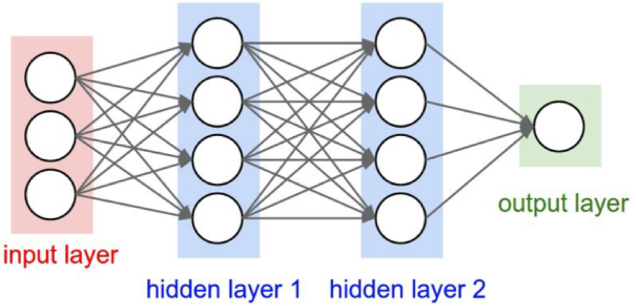
Increase the ensemble size from 100 to 1000 parameter values.

Climate model emulation

Input (1000 parameter sets; 6 parameters)

Run through trained neural network

	P1	P2	P3	P4	P5	P6
S1	x1,1	x1,2	x1,3	x1,4	x1,5	x1,6
S2	x2,1	x2,2	x2,3	x2,4	x2,5	x2,6
S3	x3,1	x3,2	x3,3	x3,4	x3,5	x3,6
…	…	…	…	…	…	…
…	…	…	…	…	…	…
S1000	x1000,1	x1000,2	x1000,3	x1000,4	x1000,5	x1000,6



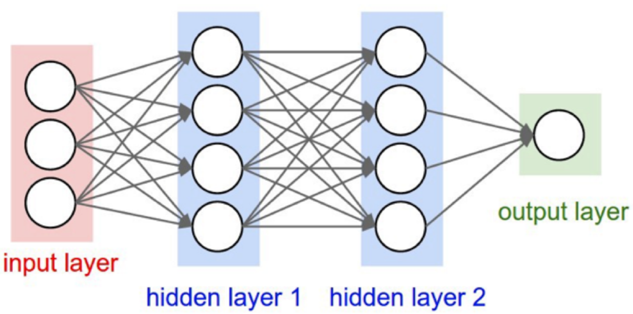
Climate model emulation

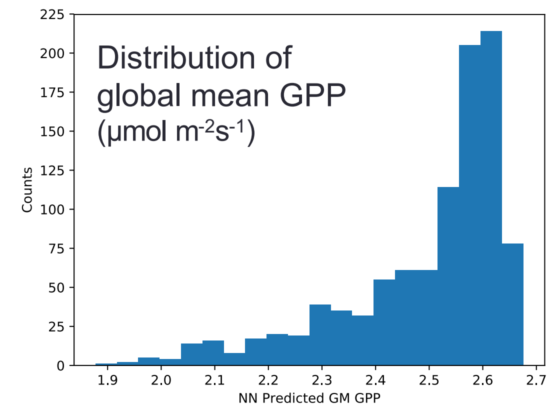
Input (1000 parameter sets; 6 parameters)

Run through trained neural network

Output (1000 neural network predictions)

	P1	P2	P3	P4	P5	P6
S1	x1,1	x1,2	x1,3	x1,4	x1,5	x1,6
S2	x2,1	x2,2	x2,3	x2,4	x2,5	x2,6
S3	x3,1	x3,2	x3,3	x3,4	x3,5	x3,6
…	…	…	…	…	…	…
…	…	…	…	…	…	…
S1000	x1000,1	x1000,2	x1000,3	x1000,4	x1000,5	x1000,6





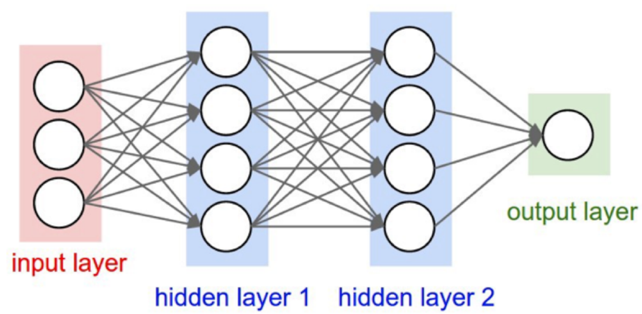
Climate model emulation

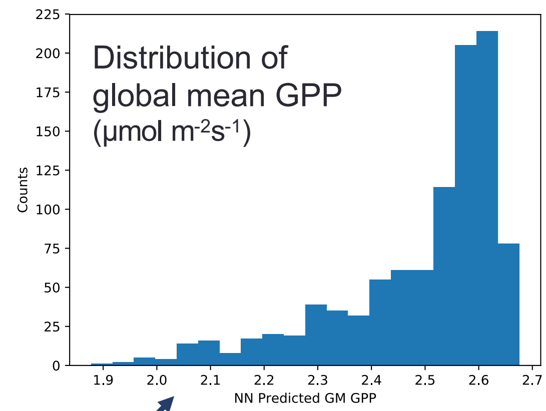
Input (1000 parameter sets; 6 parameters)

Run through trained neural network

Output (1000 neural network predictions)

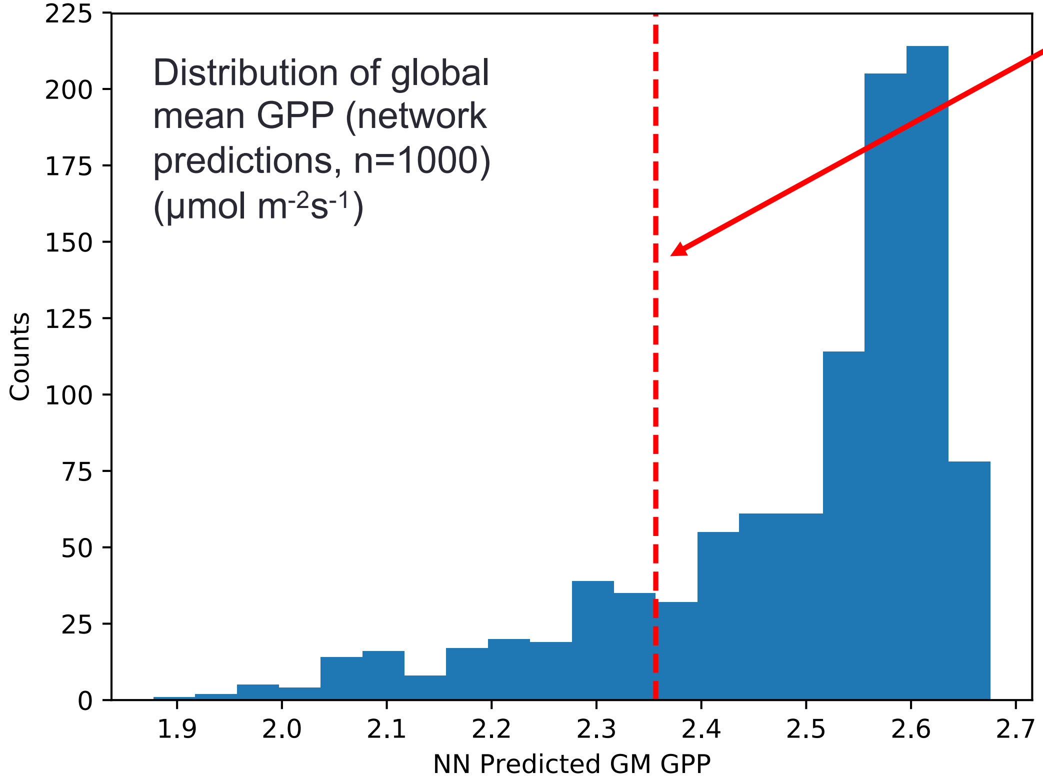
	P1	P2	P3	P4	P5	P6
S1	x1,1	x1,2	x1,3	x1,4	x1,5	x1,6
S2	x2,1	x2,2	x2,3	x2,4	x2,5	x2,6
S3	x3,1	x3,2	x3,3	x3,4	x3,5	x3,6
…	…	…	…	…	…	…
…	…	…	…	…	…	…
S1000	x1000,1	x1000,2	x1000,3	x1000,4	x1000,5	x1000,6





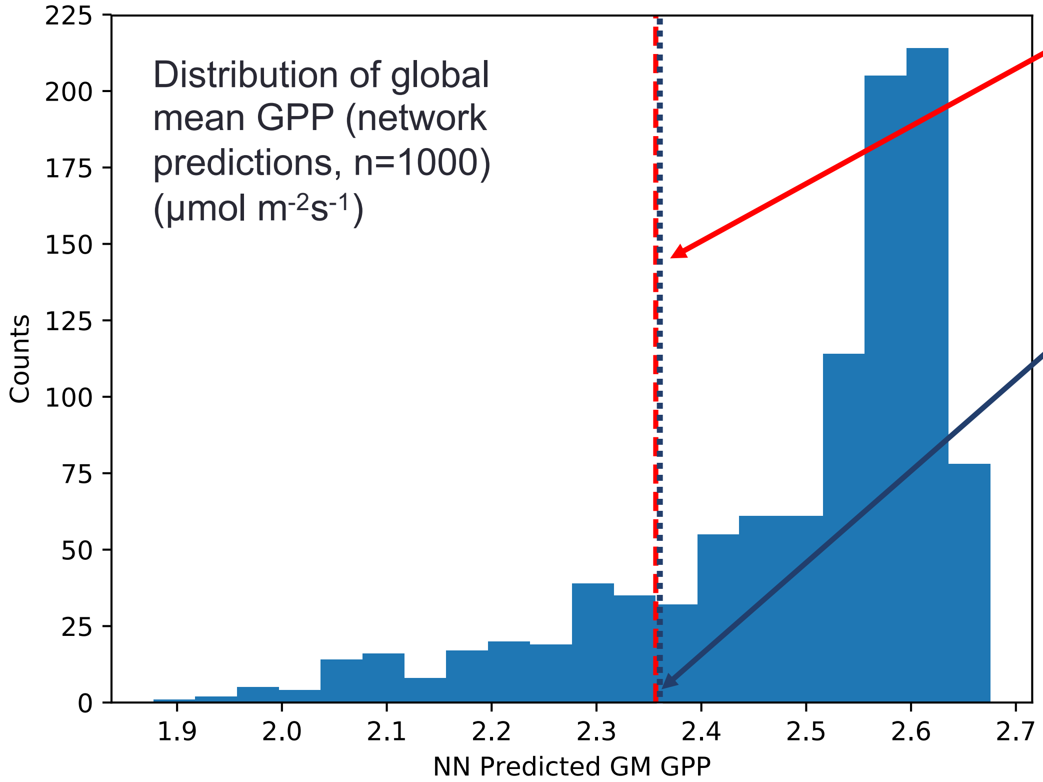
Emulated land model output! How good are these predictions? Can we use them to constrain parameter values?

Comparing with observations



Observational estimate of global mean GPP

Parameter estimation!



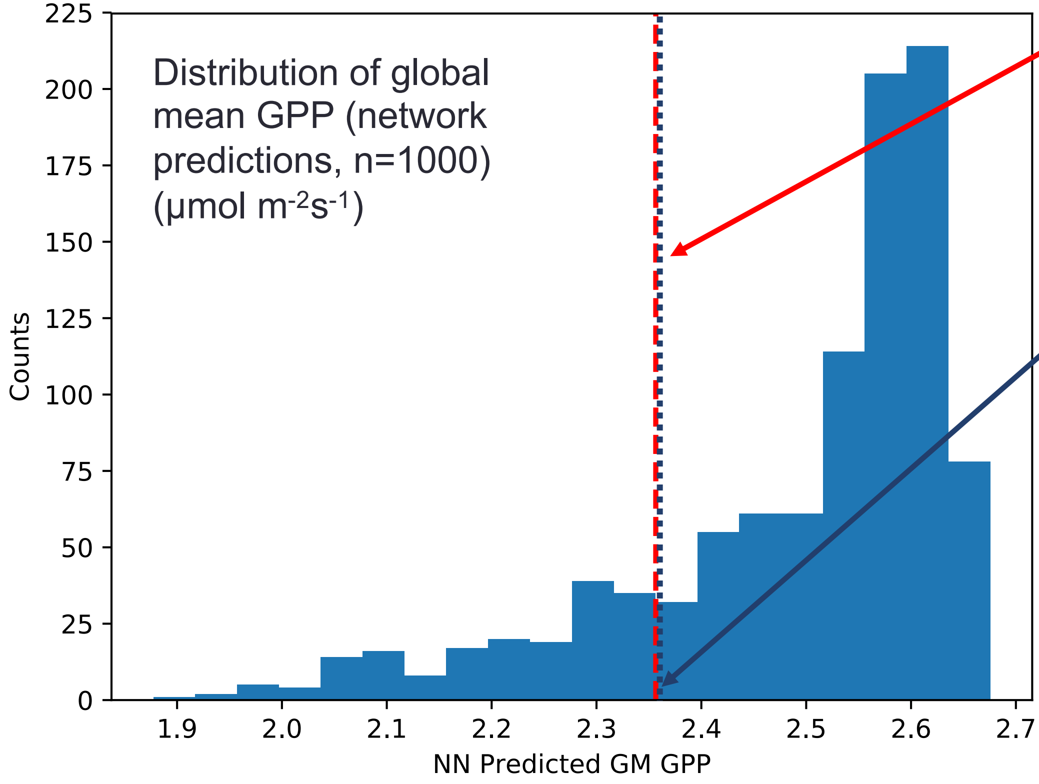
Observational estimate of global mean GPP

Isolate the emulator prediction closest to observations.

↓

This gives an estimate of "best" parameter values.

Parameter estimation!



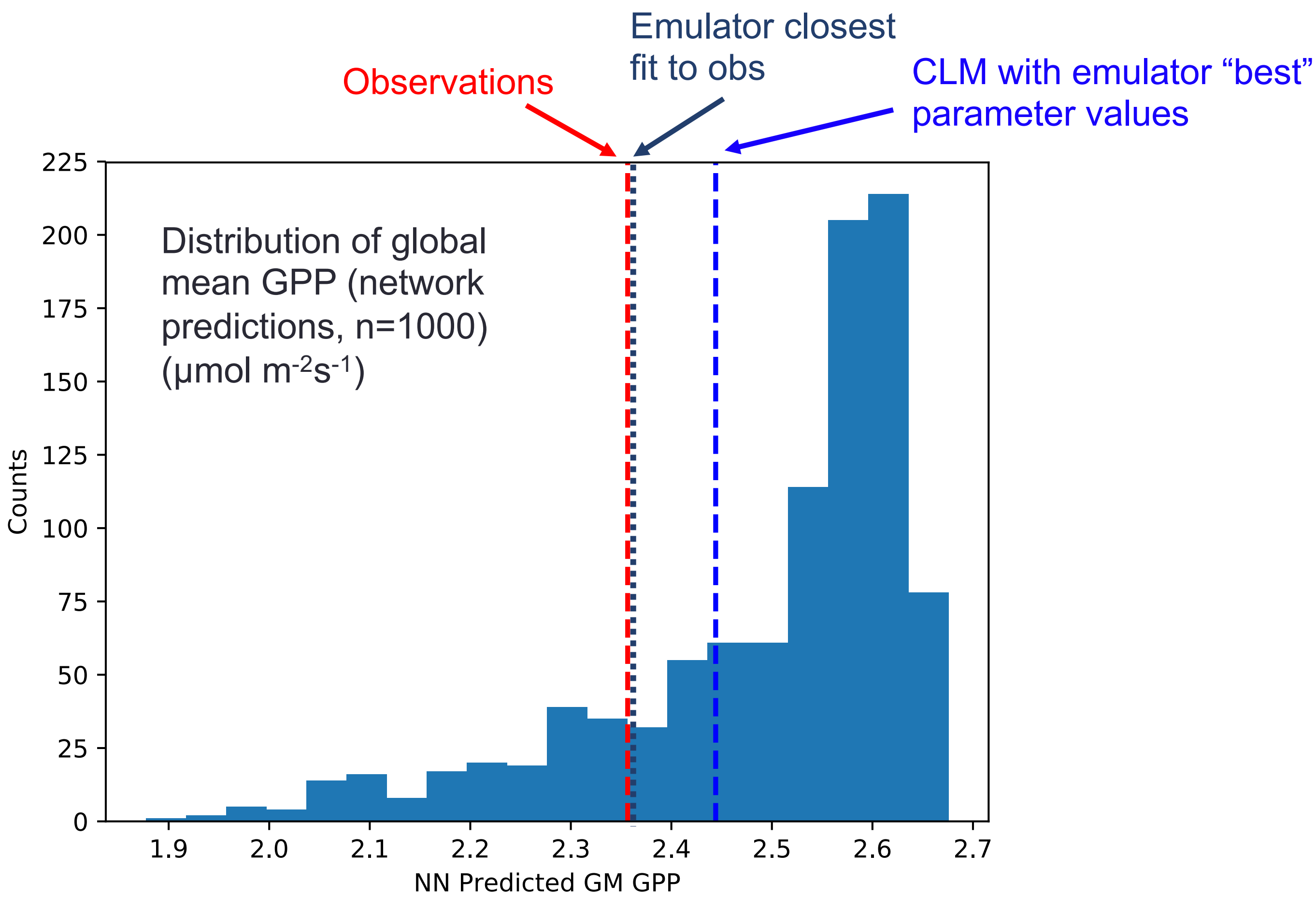
Observational estimate of global mean GPP

Isolate the emulator prediction closest to observations.

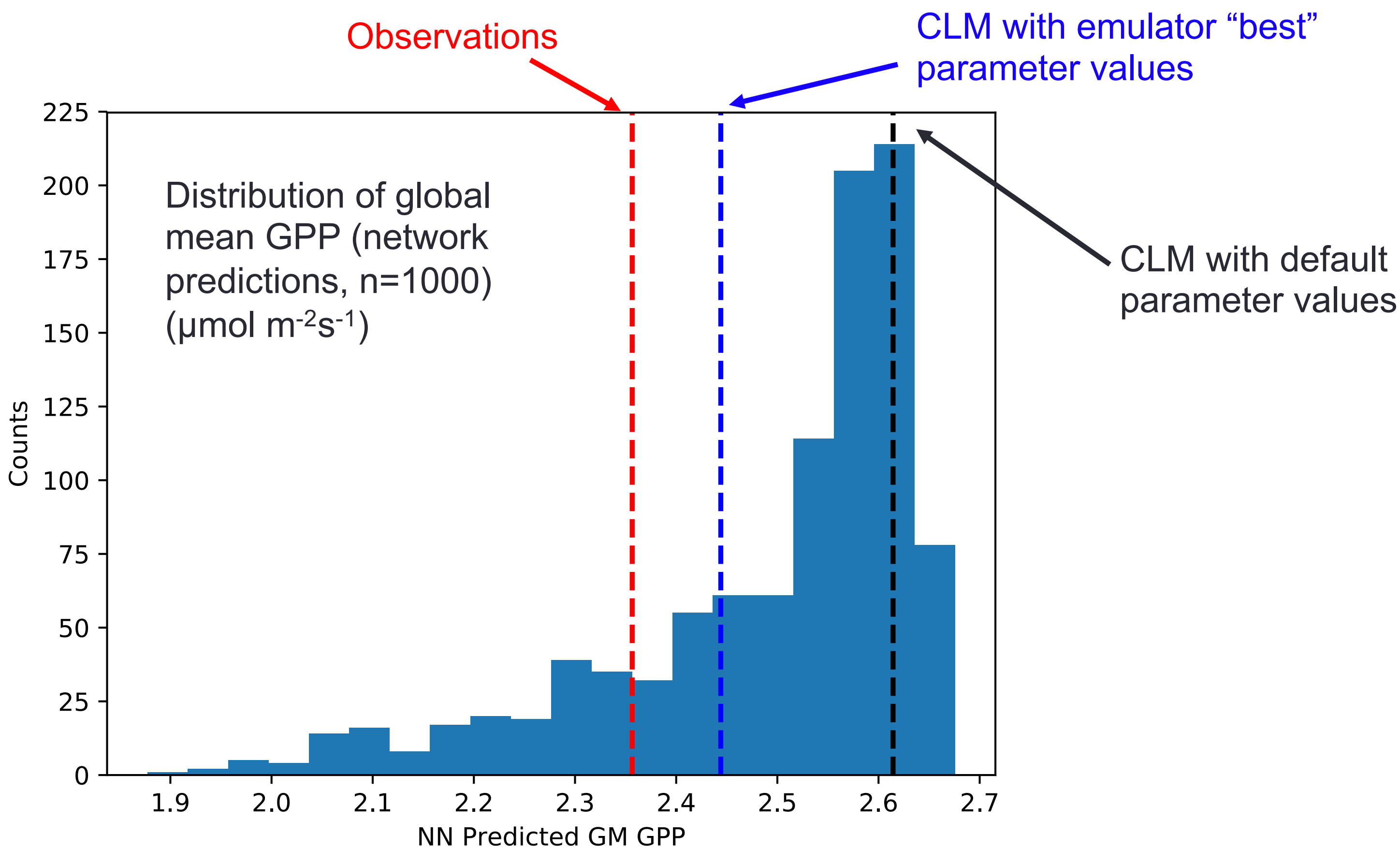
This gives an estimate of "best" parameter values.

- ➢ What happens if we run CLM with these parameter values?

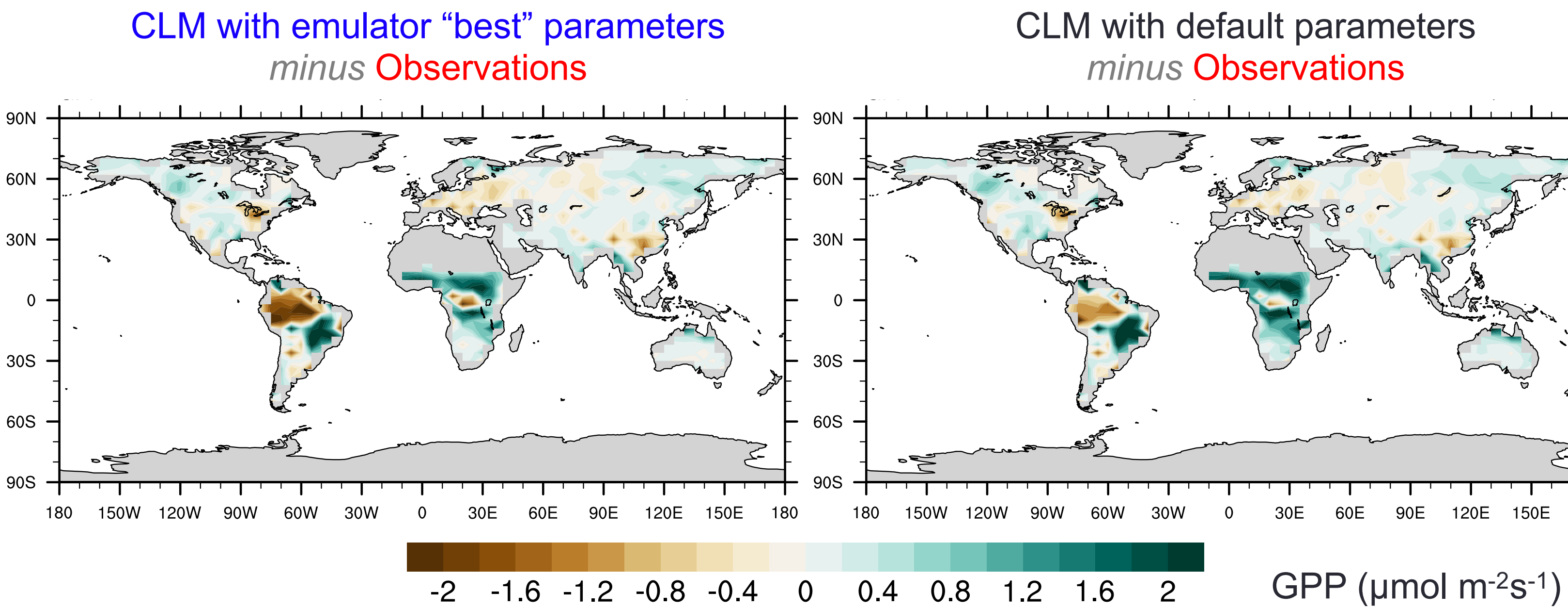
Testing the emulator



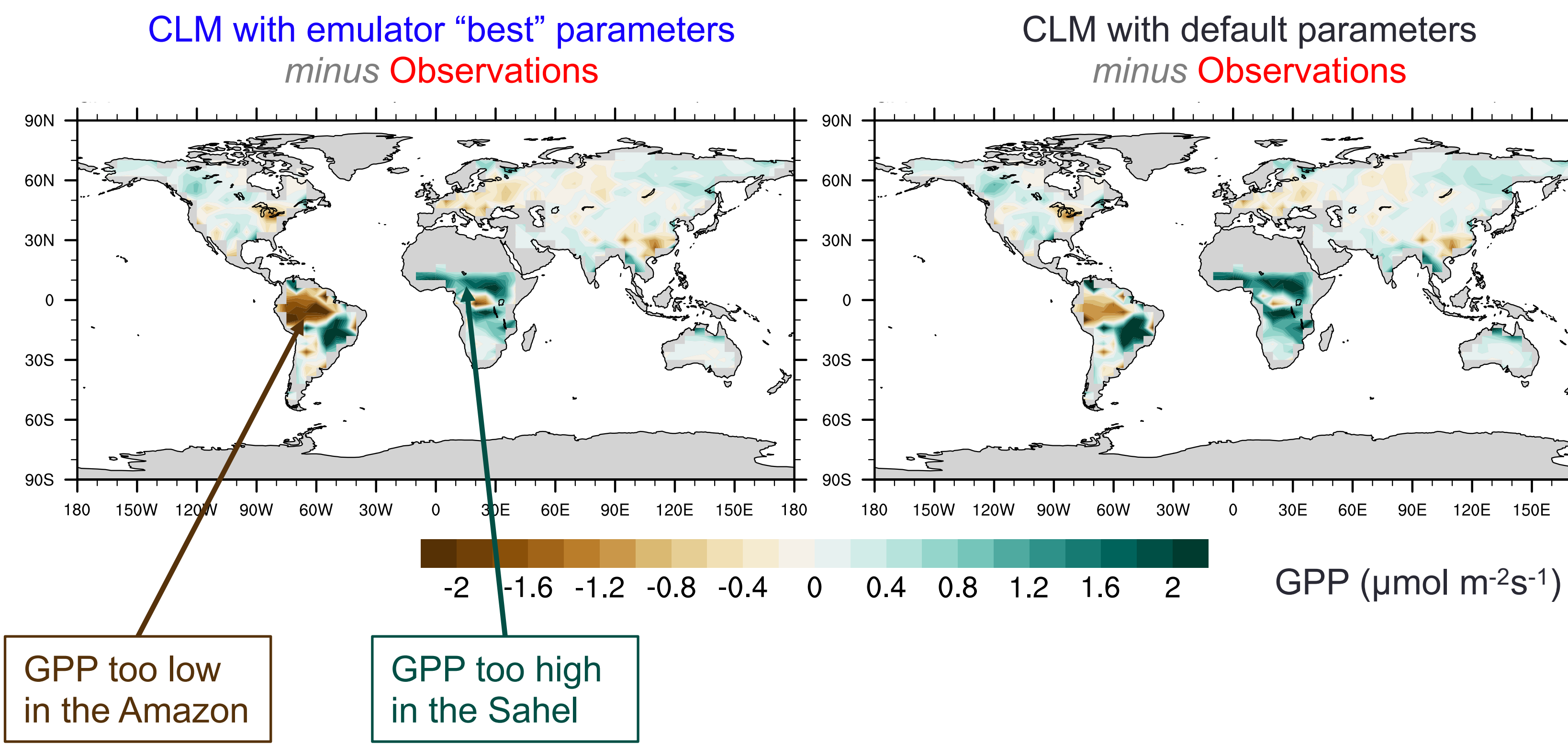
Making progress…



Regional performance needs improvement



Regional performance needs improvement

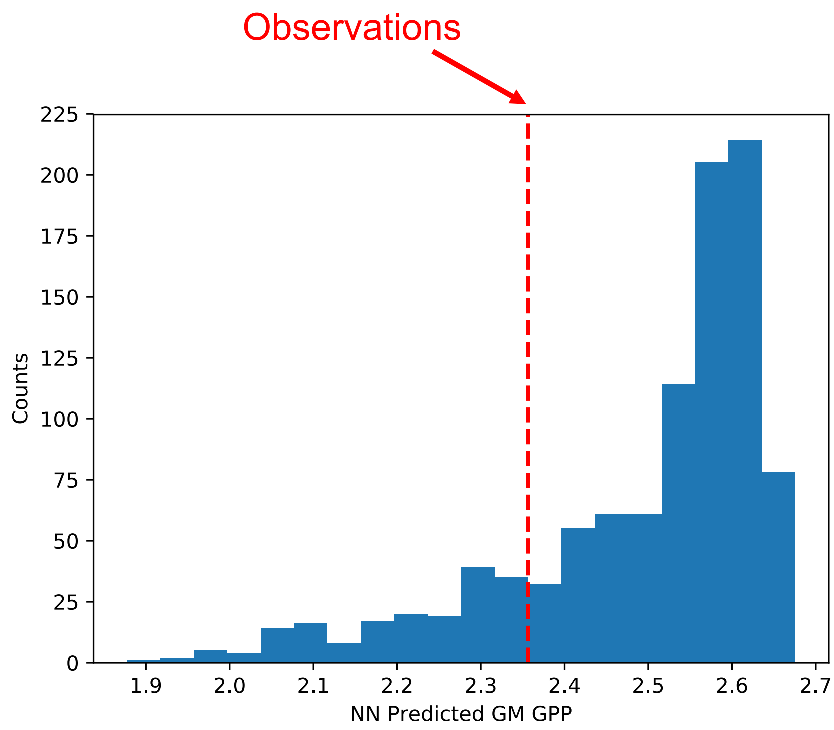


Next steps

- Assess ***multiple metrics*** (e.g., mean and ***spatial variability***)

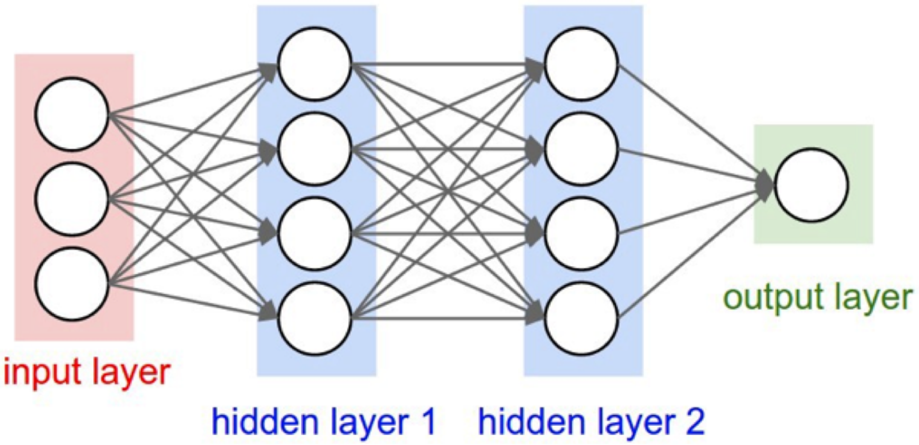
Next steps

- Assess multiple metrics (e.g., mean and spatial variability)
- Use multiple observational datasets to span ***observational uncertainty***



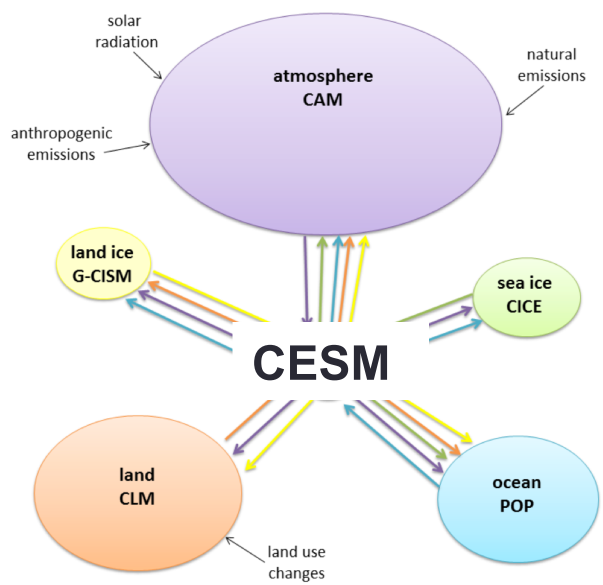
Next steps

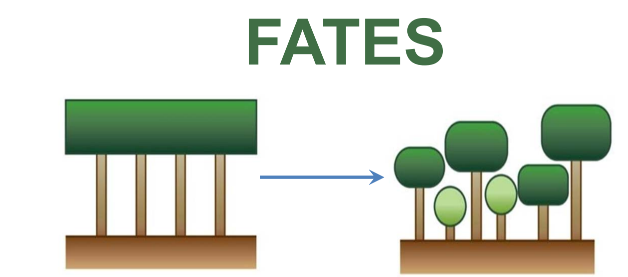
- Assess multiple metrics (e.g., mean and spatial variability)
- Use multiple observational datasets to span observational uncertainty
- Use multiple neural network configurations to span ***prediction uncertainty***



Next steps

- Assess multiple metrics (e.g., mean and spatial variability)
- Use multiple observational datasets to span observational uncertainty
- Use multiple neural network configurations to span prediction uncertainty
- Apply uncertainty framework to ***different models/model configurations***





Summary

- We can reduce uncertainty in land surface models by **studying parameters**.
- Machine learning can help in **climate model emulation**
- Land model emulator used to **estimate parameter values** by comparing with observations.

Summary

- We can reduce uncertainty in land surface models by **studying parameters**.
- Machine learning can help in **climate model emulation**
- Land model emulator used to **estimate parameter values** by comparing with observations.

github.com/ESCOMP/ctsm

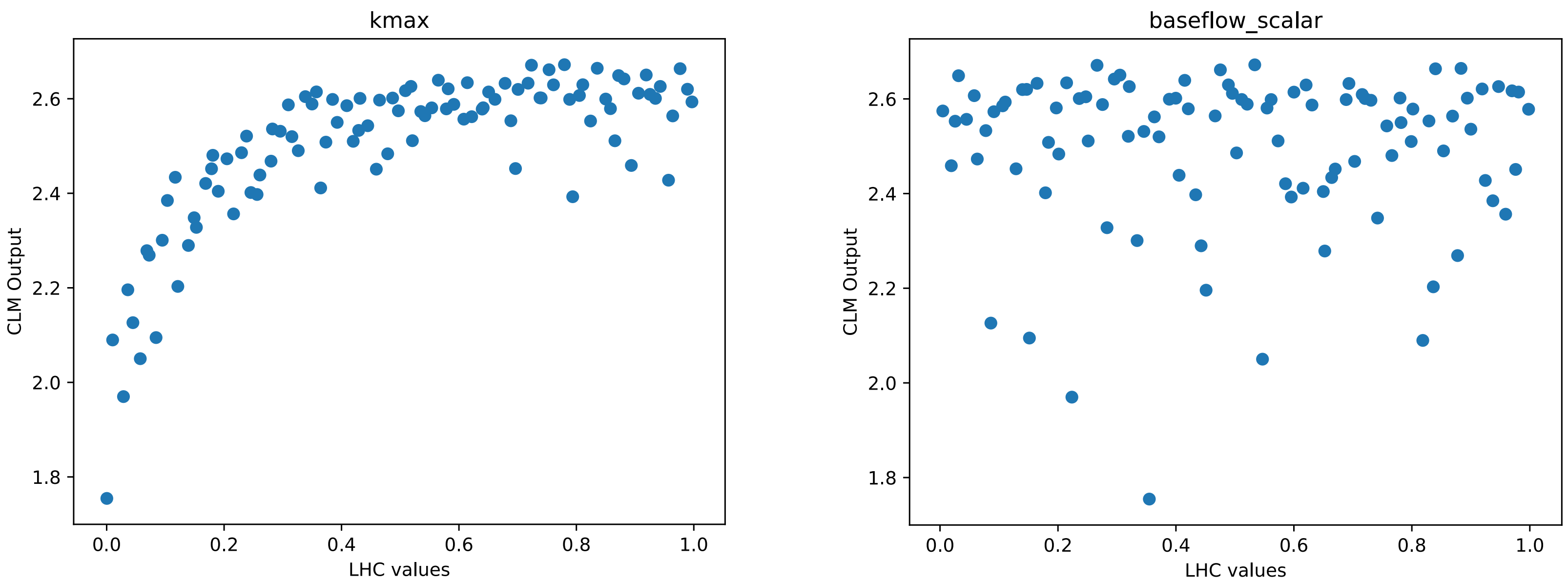
Questions?

Contact: *kdagon@ucar.edu*

@CLM_science

Backup Slides

Diagnosing skewness in global mean GPP



Train on spatial variability

- Use the spatial variability of GPP from 100 ensemble members (as condensed by singular value decomposition) to train the neural network