Parameter Uncertainty and Estimation in CLM5

Katie Dagon
ASP Postdoc, NCAR

CTSM Tutorial
February 7, 2019
What is driving uncertainty in land surface model projections of climate change?

K. Dagon
What are the sources of uncertainty in land surface models?

A. Ocean

B. Land

- Internal variability
- Model uncertainty
- Scenario uncertainty

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
Focus on CLM5 biogeophysical processes
Example parameter: medlynslope

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

From the CLM5 Documentation:

\[ g_s = g_o + 1.6(1 + \frac{g_1}{D}) \frac{A_n}{c_s/P_{atm}} \]

- **stomatal conductance**
- **medlynslope**
- **photosynthesis**

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

Medlyn et al. (2011)
Observations of photosynthesis and stomatal conductance contribute to PFT-dependent uncertainty range for medlynslope parameter.

PFT = broadleaf deciduous trees
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) → \texttt{FPSN}
2. Evapotranspiration (ET) → \texttt{QFLX\_EVAP\_TOT}
3. Transpiration Fraction = Transpiration/ET → \texttt{QVEGT/QFLX\_EVAP\_TOT}
4. Sensible Heat Flux → \texttt{FSH}
5. 10cm Soil Moisture → \texttt{SOILWATER\_10CM}
6. Total Column Soil Moisture → \texttt{SOILLIQ + SOILICE}
7. Water Table Depth → \texttt{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?

Find new study: update old, wrong parameter value

Two alternative algorithms for poorly understood process.

Use value calibrated at single site.

Different but-still-reasonable value gives better answers

Add new structure to account for new knowledge

Figure from Rosie Fisher

K. Dagon
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. Inflated ensemble size of possible parameter combinations using trained networks.
  3. Compare network predictions with observations to estimate parameter values.
Machine learning by neural networks

Machine learning by neural networks

Machine learning by neural networks

- Weights (importance of inputs)
- Neurons or nodes
- Parameter values
- CLM output
- Activations (e.g., linear, nonlinear)

input layer → hidden layer 1 → hidden layer 2 → output layer
Perturbed Parameter Ensemble Setup

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

<table>
<thead>
<tr>
<th>Simulations</th>
<th>Parameters</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>...</th>
<th>S100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
<td>x1,1</td>
<td>x2,1</td>
<td>x3,1</td>
<td>...</td>
<td>x100,1</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>x1,2</td>
<td>x2,2</td>
<td>x3,2</td>
<td>...</td>
<td>x100,2</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>x1,3</td>
<td>x2,3</td>
<td>x3,3</td>
<td>...</td>
<td>x100,3</td>
</tr>
<tr>
<td></td>
<td>P4</td>
<td>x1,4</td>
<td>x2,4</td>
<td>x3,4</td>
<td>...</td>
<td>x100,4</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>x1,5</td>
<td>x2,5</td>
<td>x3,5</td>
<td>...</td>
<td>x100,5</td>
</tr>
<tr>
<td></td>
<td>P6</td>
<td>x1,6</td>
<td>x2,6</td>
<td>x3,6</td>
<td>...</td>
<td>x100,6</td>
</tr>
</tbody>
</table>

- **Parameters**
  - S1
  - S2
  - S3
  - ...
  - S100
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)

<table>
<thead>
<tr>
<th>S1</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x1,1</td>
<td>x1,2</td>
<td>x1,3</td>
<td>x1,4</td>
<td>x1,5</td>
<td>x1,6</td>
</tr>
<tr>
<td>S2</td>
<td>x2,1</td>
<td>x2,2</td>
<td>x2,3</td>
<td>x2,4</td>
<td>x2,5</td>
<td>x2,6</td>
</tr>
<tr>
<td>S3</td>
<td>x3,1</td>
<td>x3,2</td>
<td>x3,3</td>
<td>x3,4</td>
<td>x3,5</td>
<td>x3,6</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S100</td>
<td>x100,1</td>
<td>x100,2</td>
<td>x100,3</td>
<td>x100,4</td>
<td>x100,5</td>
<td>x100,6</td>
</tr>
</tbody>
</table>
Build and train a neural network to predict land model output based on parameter values

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>x1,1</td>
<td>x1,2</td>
<td>x1,3</td>
<td>x1,4</td>
<td>x1,5</td>
<td>x1,6</td>
</tr>
<tr>
<td>S2</td>
<td>x2,1</td>
<td>x2,2</td>
<td>x2,3</td>
<td>x2,4</td>
<td>x2,5</td>
<td>x2,6</td>
</tr>
<tr>
<td>S3</td>
<td>x3,1</td>
<td>x3,2</td>
<td>x3,3</td>
<td>x3,4</td>
<td>x3,5</td>
<td>x3,6</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>S100</td>
<td>x100,1</td>
<td>x100,2</td>
<td>x100,3</td>
<td>x100,4</td>
<td>x100,5</td>
<td>x100,6</td>
</tr>
</tbody>
</table>

Output (100 CLM simulations)

Distribution of global mean GPP ($\mu$mol m$^{-2}$s$^{-1}$)

![Neural Network Diagram]

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

Network predicted vs. CLM global mean GPP (µmol m$^{-2}$s$^{-1}$)

$r^2 = 0.91$
Climate model emulation

Input (1000 parameter sets; 6 parameters)

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>x1,1</td>
<td>x1,2</td>
<td>x1,3</td>
<td>x1,4</td>
<td>x1,5</td>
<td>x1,6</td>
</tr>
<tr>
<td>S2</td>
<td>x2,1</td>
<td>x2,2</td>
<td>x2,3</td>
<td>x2,4</td>
<td>x2,5</td>
<td>x2,6</td>
</tr>
<tr>
<td>S3</td>
<td>x3,1</td>
<td>x3,2</td>
<td>x3,3</td>
<td>x3,4</td>
<td>x3,5</td>
<td>x3,6</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>S1000</td>
<td>x1000,1</td>
<td>x1000,2</td>
<td>x1000,3</td>
<td>x1000,4</td>
<td>x1000,5</td>
<td>x1000,6</td>
</tr>
</tbody>
</table>

Increase the ensemble size from 100 to 1000 parameter values.
Climate model emulation

Input (1000 parameter sets; 6 parameters)

Run through trained neural network

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>x1,1</td>
<td>x1,2</td>
<td>x1,3</td>
<td>x1,4</td>
<td>x1,5</td>
<td>x1,6</td>
</tr>
<tr>
<td>S2</td>
<td>x2,1</td>
<td>x2,2</td>
<td>x2,3</td>
<td>x2,4</td>
<td>x2,5</td>
<td>x2,6</td>
</tr>
<tr>
<td>S3</td>
<td>x3,1</td>
<td>x3,2</td>
<td>x3,3</td>
<td>x3,4</td>
<td>x3,5</td>
<td>x3,6</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>S1000</td>
<td>x1000,1</td>
<td>x1000,2</td>
<td>x1000,3</td>
<td>x1000,4</td>
<td>x1000,5</td>
<td>x1000,6</td>
</tr>
</tbody>
</table>
Climate model emulation

Input (1000 parameter sets; 6 parameters)

Run through trained neural network

Output (1000 neural network predictions)

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>x1,1</td>
<td>x1,2</td>
<td>x1,3</td>
<td>x1,4</td>
<td>x1,5</td>
<td>x1,6</td>
</tr>
<tr>
<td>S2</td>
<td>x2,1</td>
<td>x2,2</td>
<td>x2,3</td>
<td>x2,4</td>
<td>x2,5</td>
<td>x2,6</td>
</tr>
<tr>
<td>S3</td>
<td>x3,1</td>
<td>x3,2</td>
<td>x3,3</td>
<td>x3,4</td>
<td>x3,5</td>
<td>x3,6</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>S1000</td>
<td>x1000,1</td>
<td>x1000,2</td>
<td>x1000,3</td>
<td>x1000,4</td>
<td>x1000,5</td>
<td>x1000,6</td>
</tr>
</tbody>
</table>

Distribution of global mean GPP (µmol m⁻² s⁻¹)
Climate model emulation

Run through trained neural network

Input (1000 parameter sets; 6 parameters)

Output (1000 neural network predictions)

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

Distribution of global mean GPP (network predictions, n=1000) ($\mu$mol m$^{-2}$s$^{-1}$)

Observational estimate of global mean GPP
Parameter estimation!

Distribution of global mean GPP (network predictions, n=1000) (µmol m\(^{-2}\)s\(^{-1}\))

Observational estimate of global mean GPP

Isolate the emulator prediction closest to observations.

This gives an estimate of “best” parameter values.
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

Observations fit to obs

CLM with emulator “best” parameter values

Distribution of global mean GPP (network predictions, n=1000) (µmol m^{-2}s^{-1})
Making progress…

Distribution of global mean GPP (network predictions, n=1000) (µmol m\(^{-2}\)s\(^{-1}\))

Observations

CLM with emulator “best” parameter values

CLM with default parameter values
Regional performance needs improvement

CLM with emulator “best” parameters minus Observations

CLM with default parameters minus Observations

GPP (µmol m⁻²s⁻¹)
Regional performance needs improvement

CLM with emulator “best” parameters
minus Observations

CLM with default parameters
minus Observations

GPP too low in the Amazon

GPP too high in the Sahel

GPP (µmol m⁻²s⁻¹)
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.

Questions?

Contact: kdagon@ucar.edu

@CLM_science
Diagnosing skewness in global mean GPP

![Graphs showing relationship between CLM output and LHC values for kmax and baseflow_scalar.](image)
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.