# CESM Working Group Meetings 2024

## Computationally efficient method for predicting evapotranspiration using a Gaussian Process Regression Emulator

Thomas **Kavoo**
Climate, Water and Society Lab - CFWE, Auburn University
2/29/2024

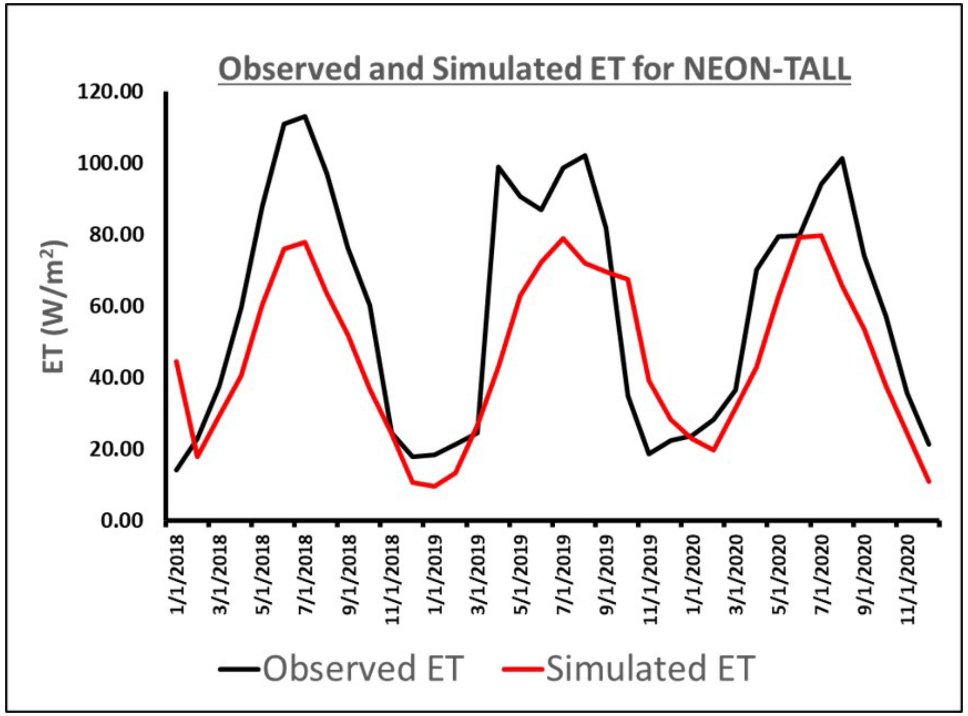
Advisor: Sanjiv **Kumar**

# Background

- Process-based Land surface models (LSMs) are pivotal in understanding the changing climate system
- Understanding Evapotranspiration (ET)
  - ET is the linkage between carbon, water, and energy cycles.
  - Sixty-five percent of annual precipitation returns back to the atmosphere as ET

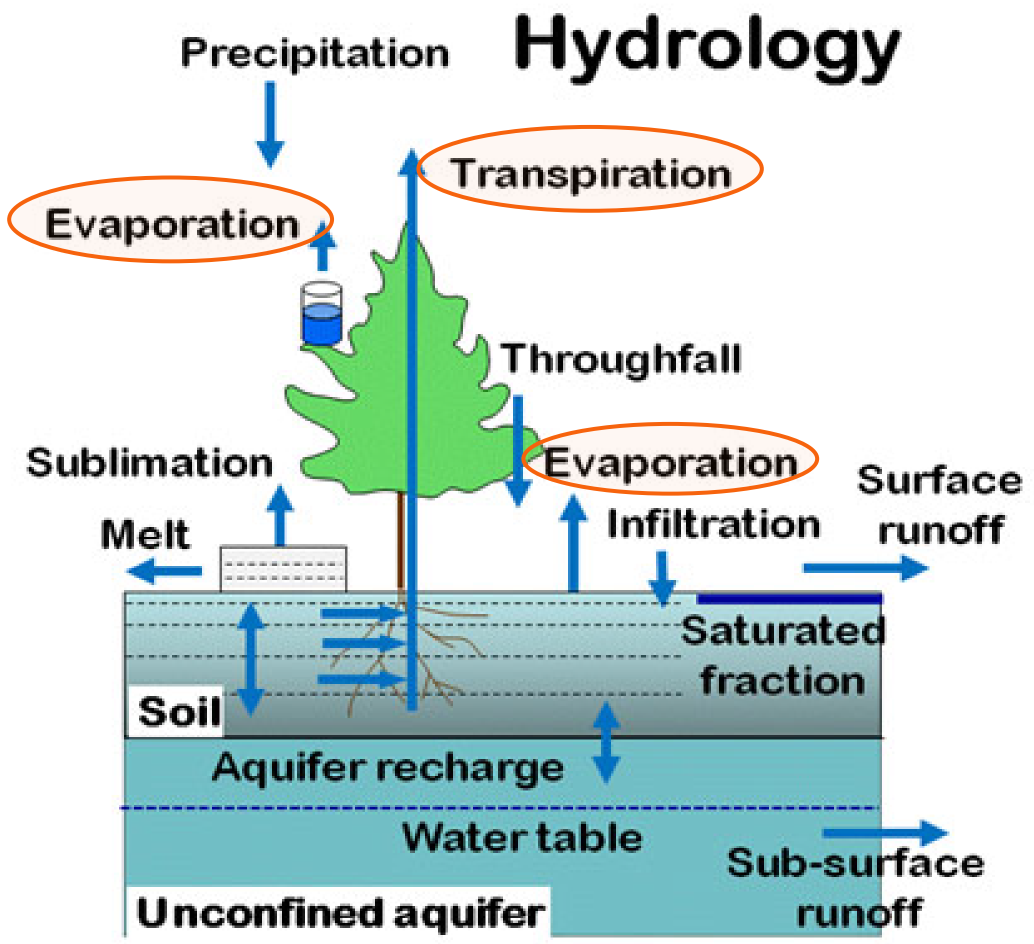
# Motivation and Overview

- Current representations of land surface processes in LSMs exhibit uncertainties.
- Need for more accurate climate projections and understanding climate-carbon feedback.
- Better representation -> More details -> Increase computational demands.
- Robust and computationally efficient models **(emulators)** are essential for scalability and practicality.



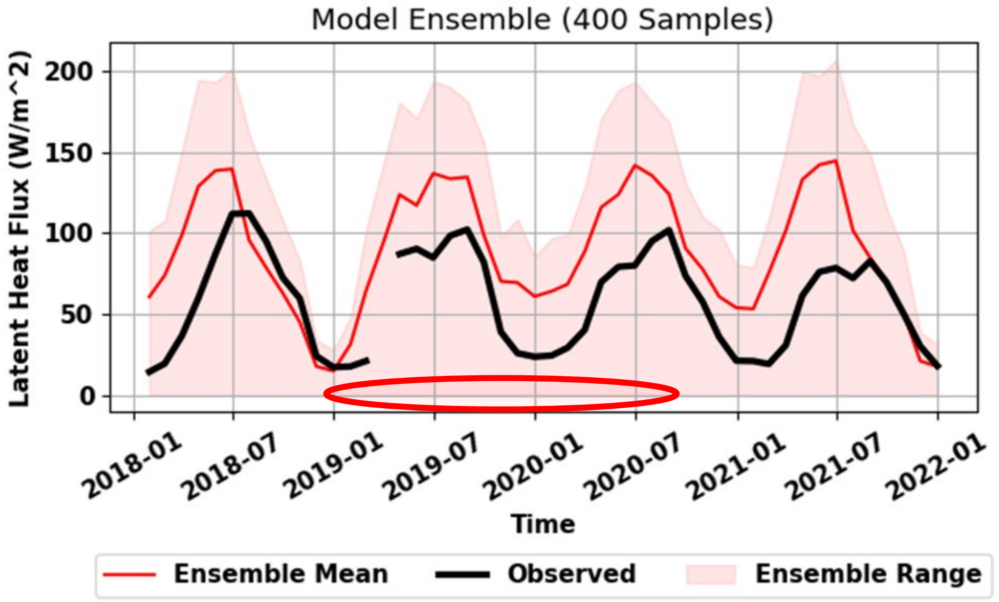
**Computational challenge:**

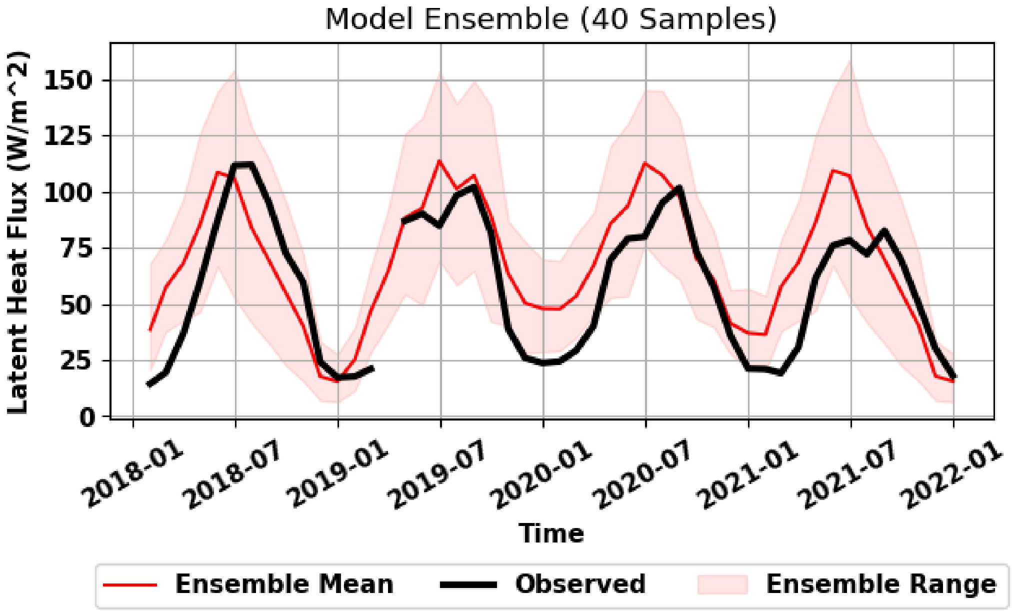
- LSMs strive for comprehensive representations of land surface processes.
- Increased complexity often leads to computational bottlenecks.

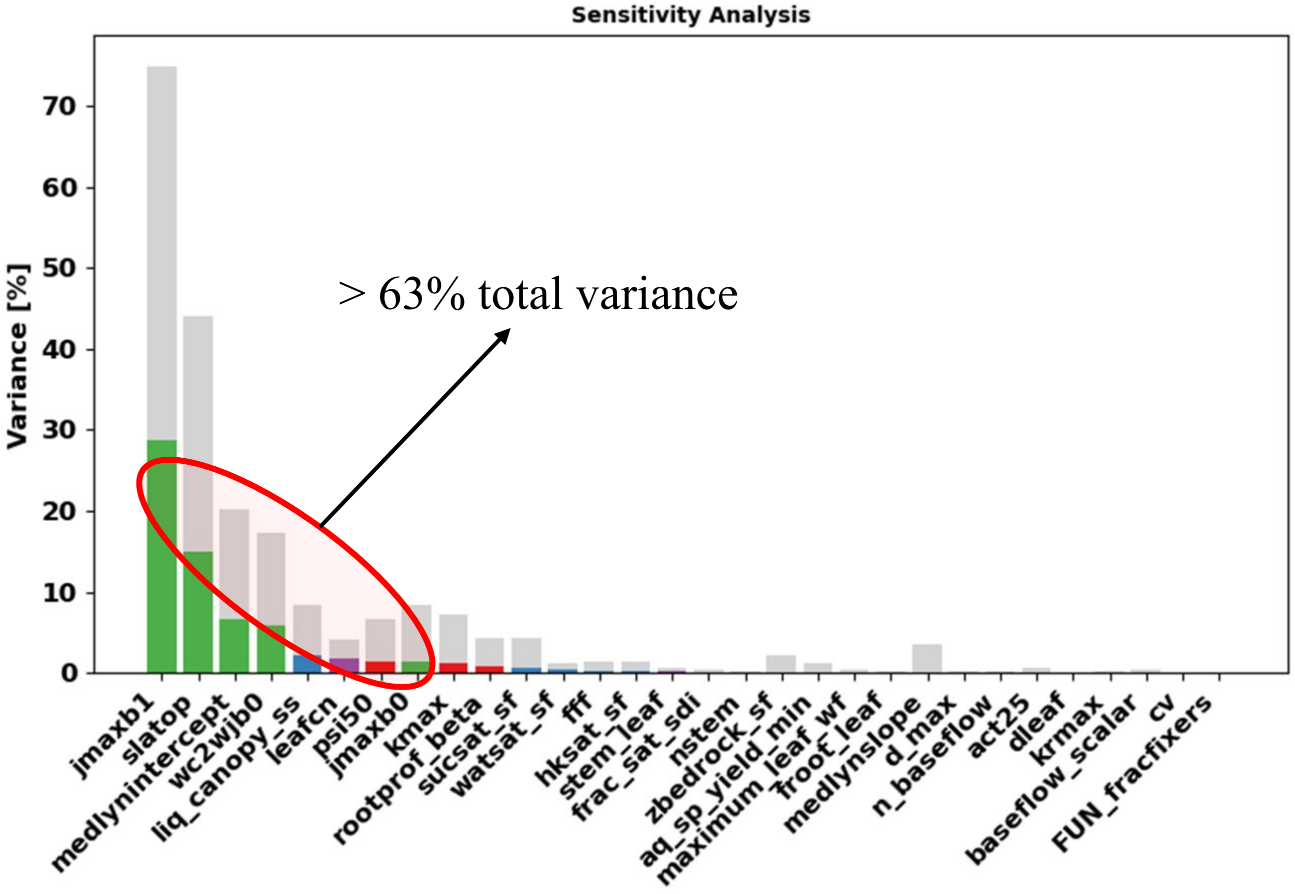



*Source: CLM5 Technical note (Figure 2.7.1)*

# Data Availability



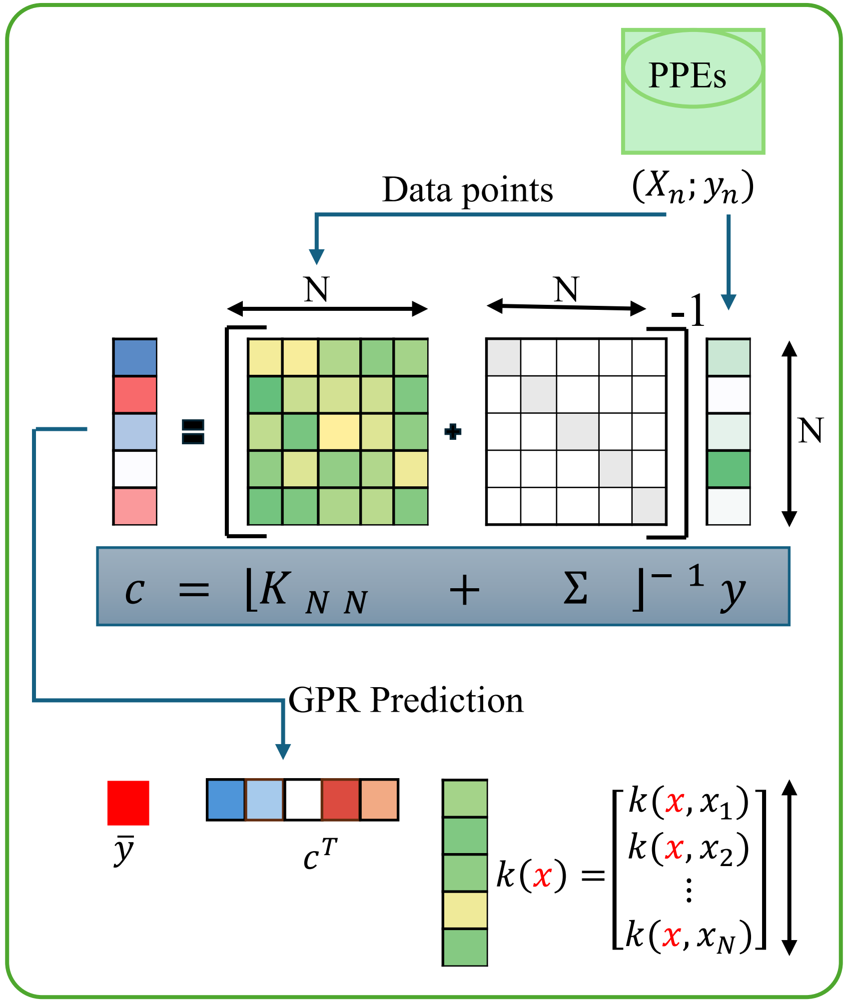
# Gaussian Process Regression

- GP is defined by a mean function $m(x)$ and a covariance function $k(x, x')$
- Given a set of training data X and corresponding function values y, the GP defines a prior distribution over functions $f$ such that:

$$f(X) \sim GP\big(m(X), k(X, X')\big)$$

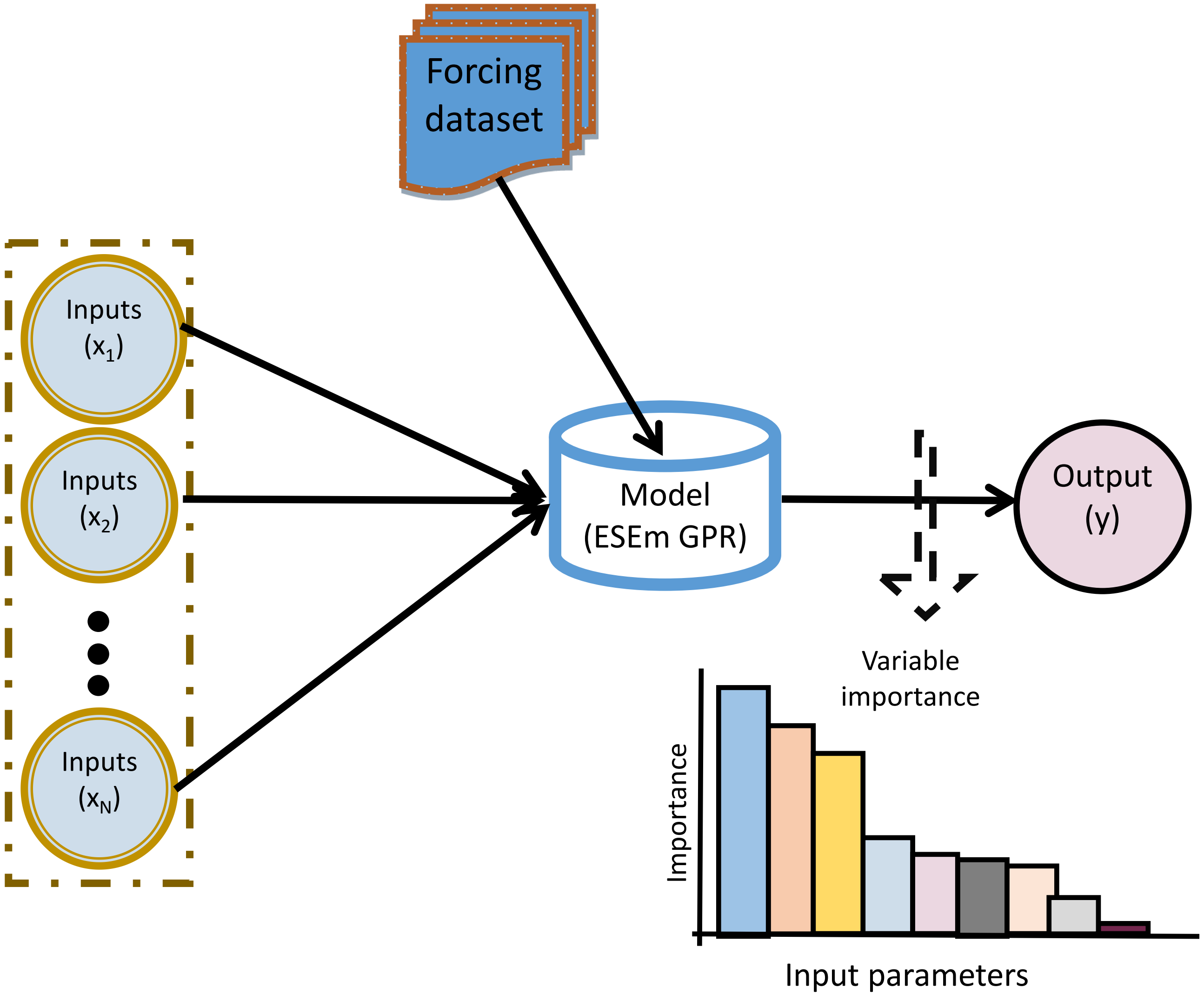
- The predictive distribution at point $x_*$ given training data $X$ and $y$ is:

$$f_* | X, y, x_* \sim GP(m_* k_*)$$




Source: Adapted from Deringer et. al., 2021

# Emulating Evapotranspiration




Output (ET)

Forcing datasets (7 – variables)

Perturbed parameters (7 – params)

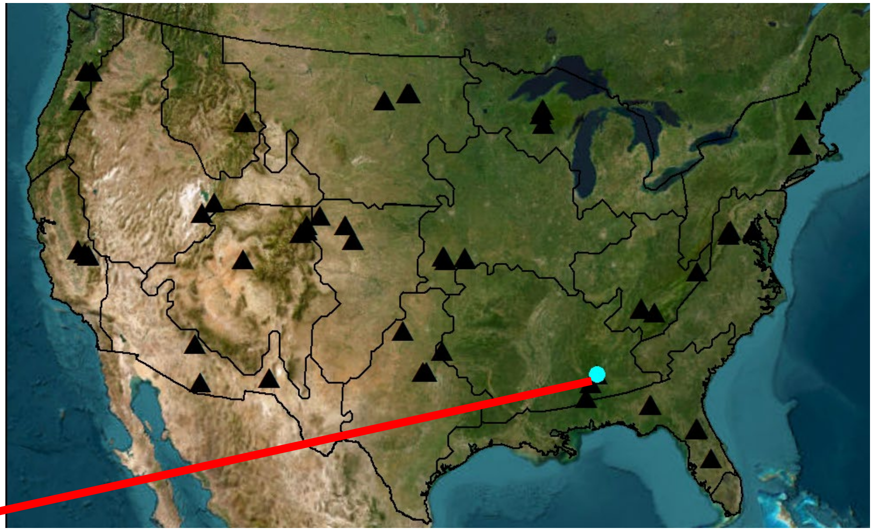
$$y_t(X_{\theta W}) \sim GP(\underbrace{m(X_{\theta W})}_{\text{Mean function}}, \overbrace{k(x, x')}^{\text{Covariance matrix}})$$

# ESEm Emulator Calibration

Bayes Theorem

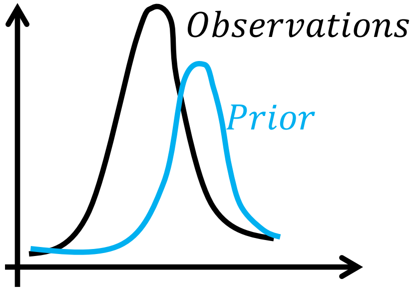
$$p(\theta|Y^0) \propto p(Y^0|\theta)p(\theta)$$

Observations

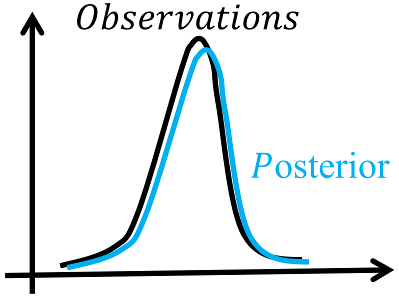


CLM5
Point-scale
simulations

Emulator
$y_t(X_{\theta W}) \sim GP(\mathrm{m}(X_{\theta W}), \mathrm{k}(\mathrm{x}, x'))$




Sampling the posterior distribution using MCMC



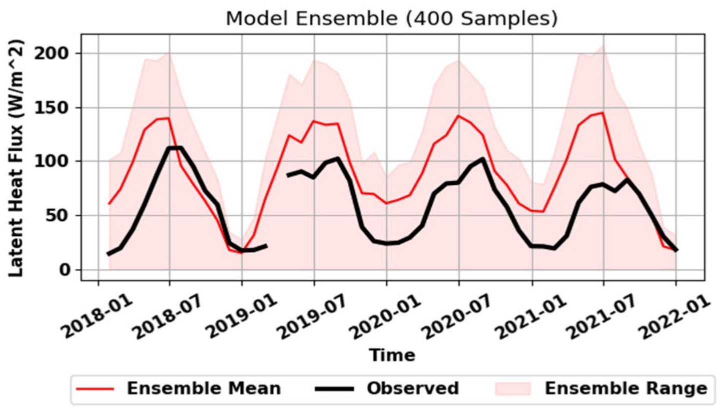
# Emulator Designs

**Mean**

- Mean ET from CLM5
- Observations 400 x 7
- No forcing dataset used

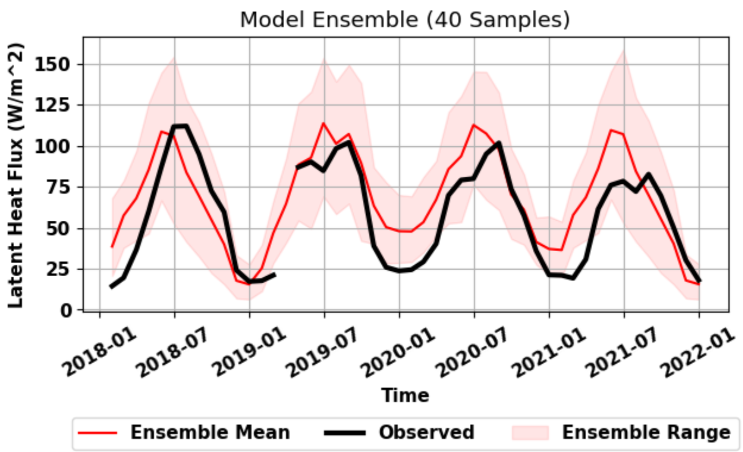
**PPE Guided**

- Monthly ET from CLM5
- Observations 1920 x 14
- Forcing dataset included



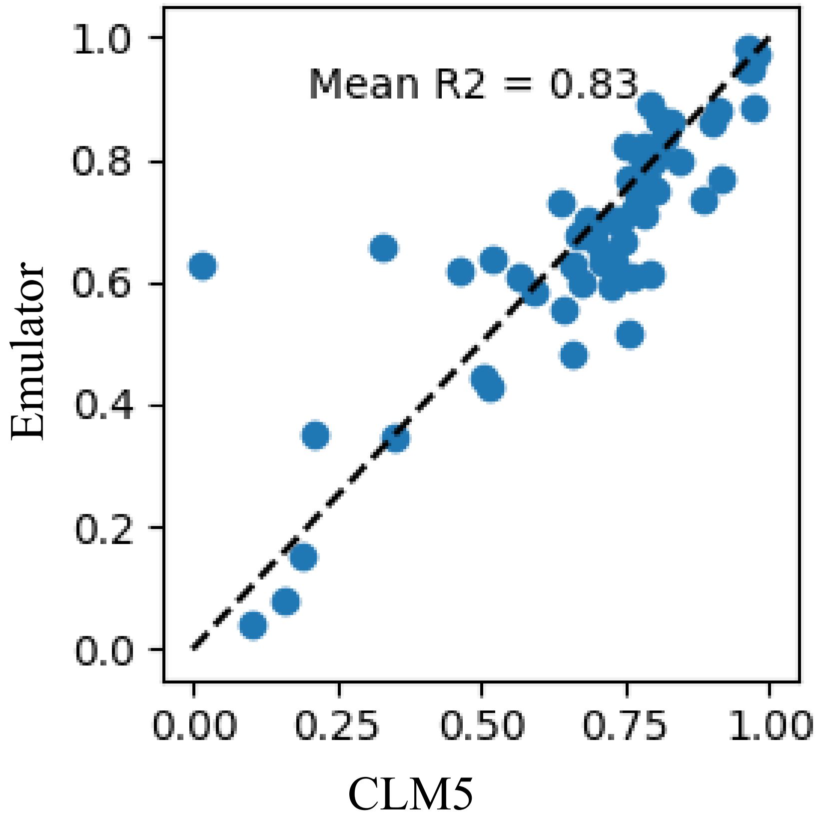

**Metric**

- Composite metric
- Observations 400 x 7
- No forcing dataset used

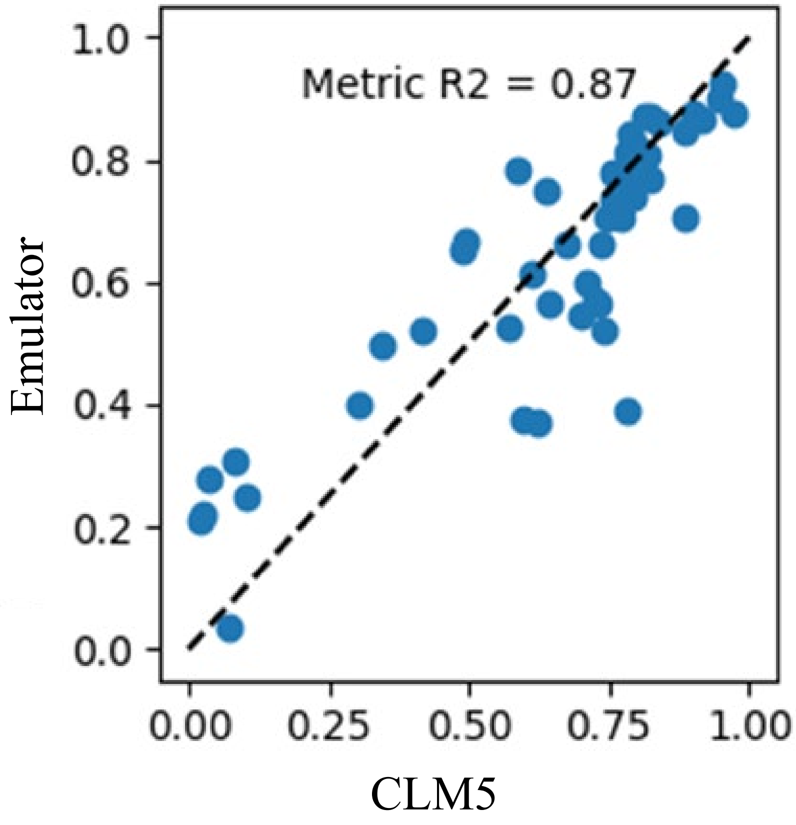


# Emulator Performance

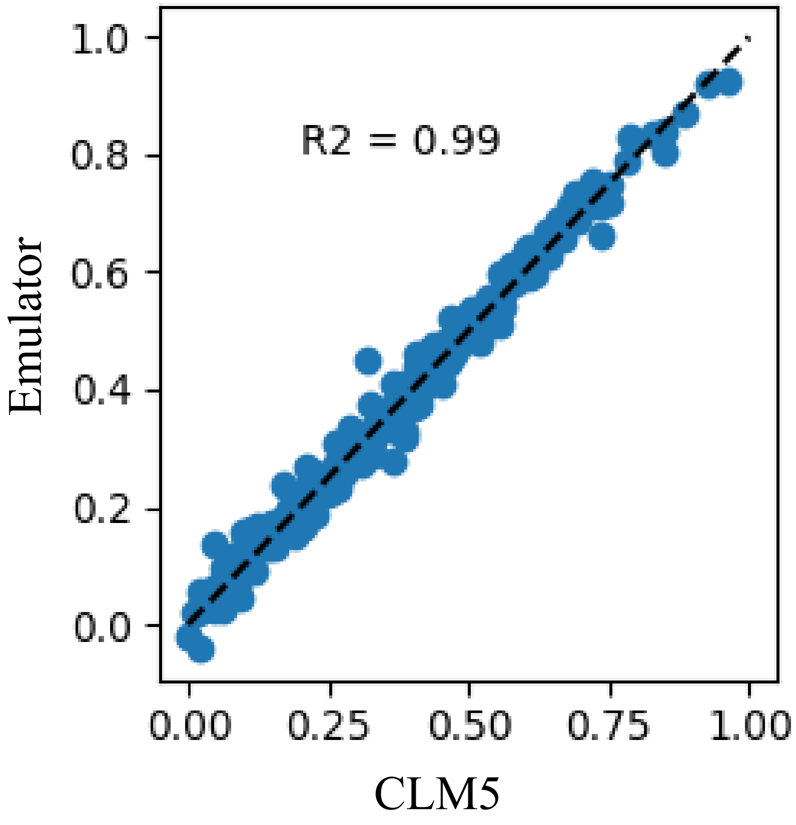
Mean



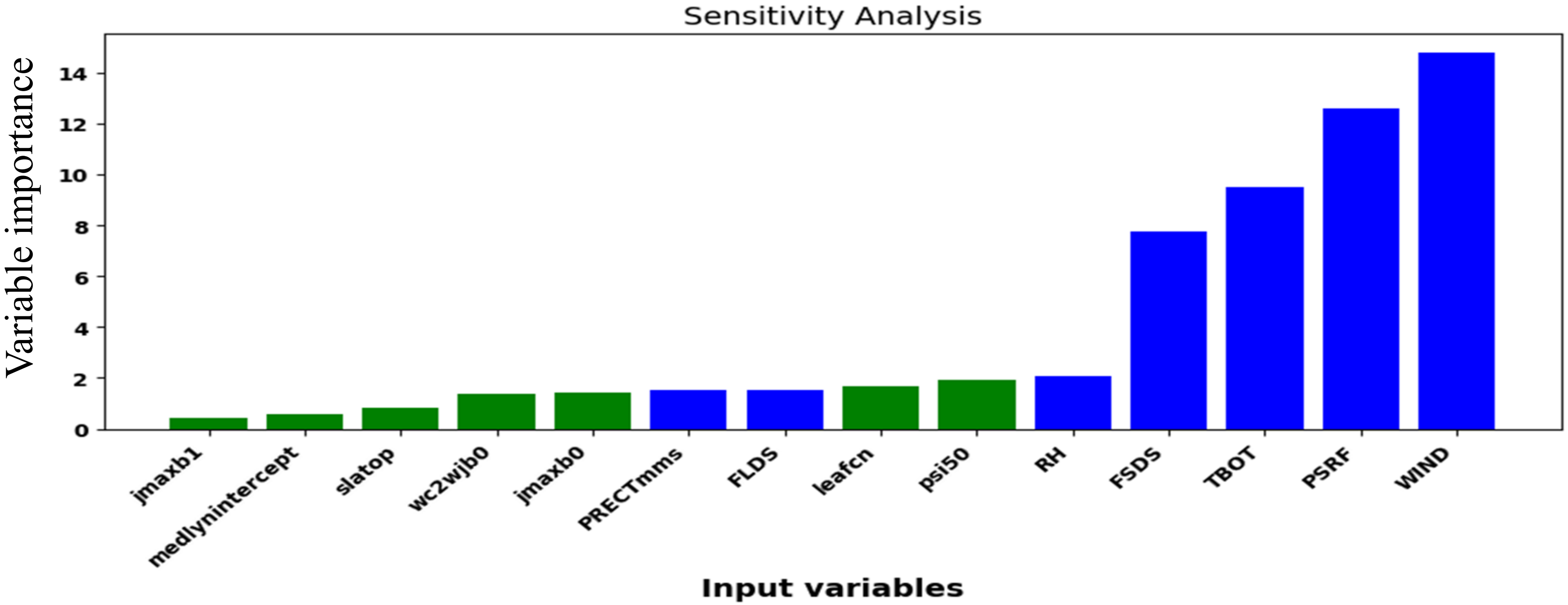

Metric



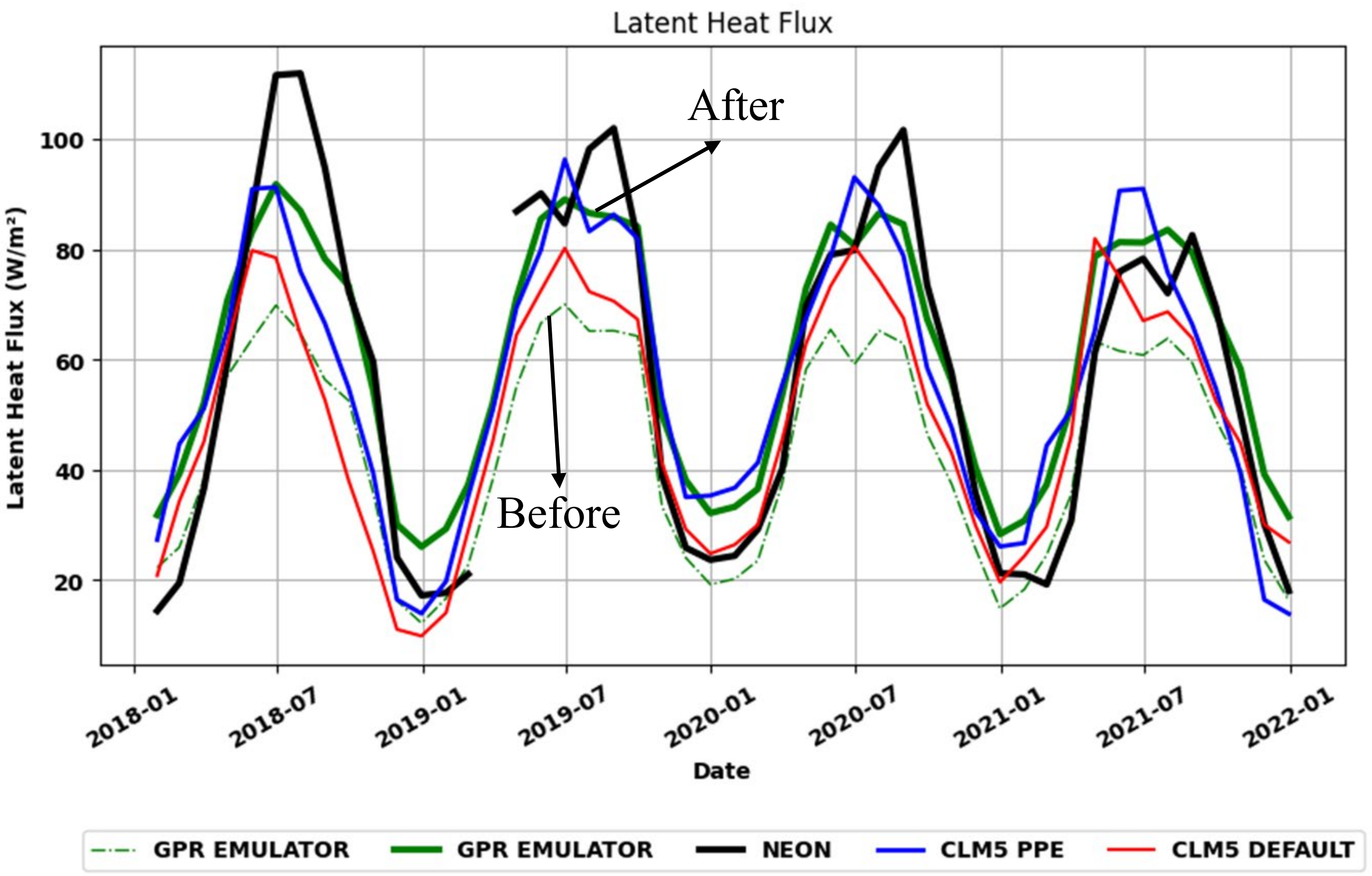

PPE Guided

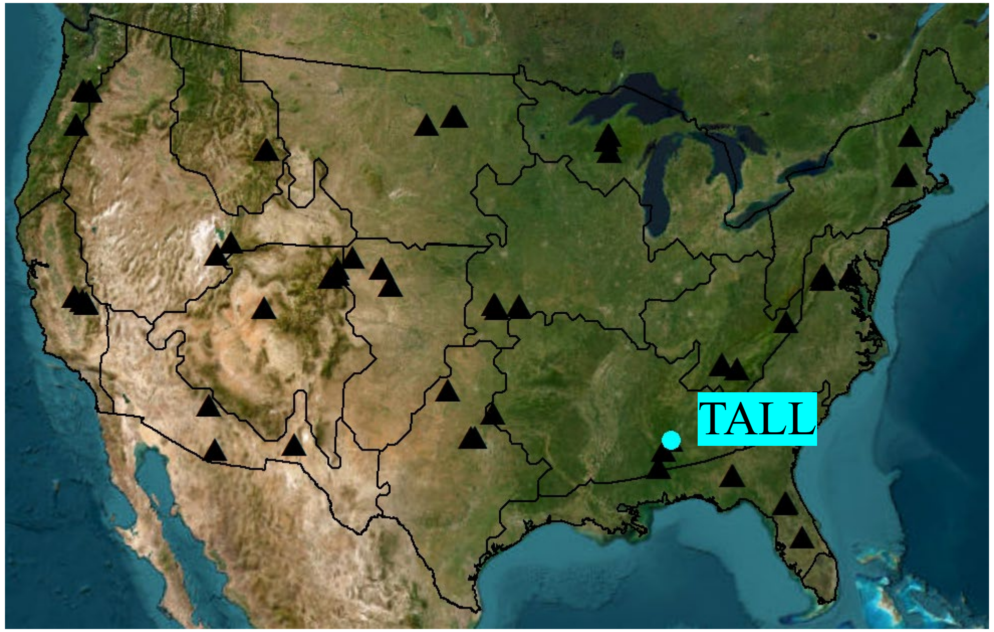


# Input variable importance



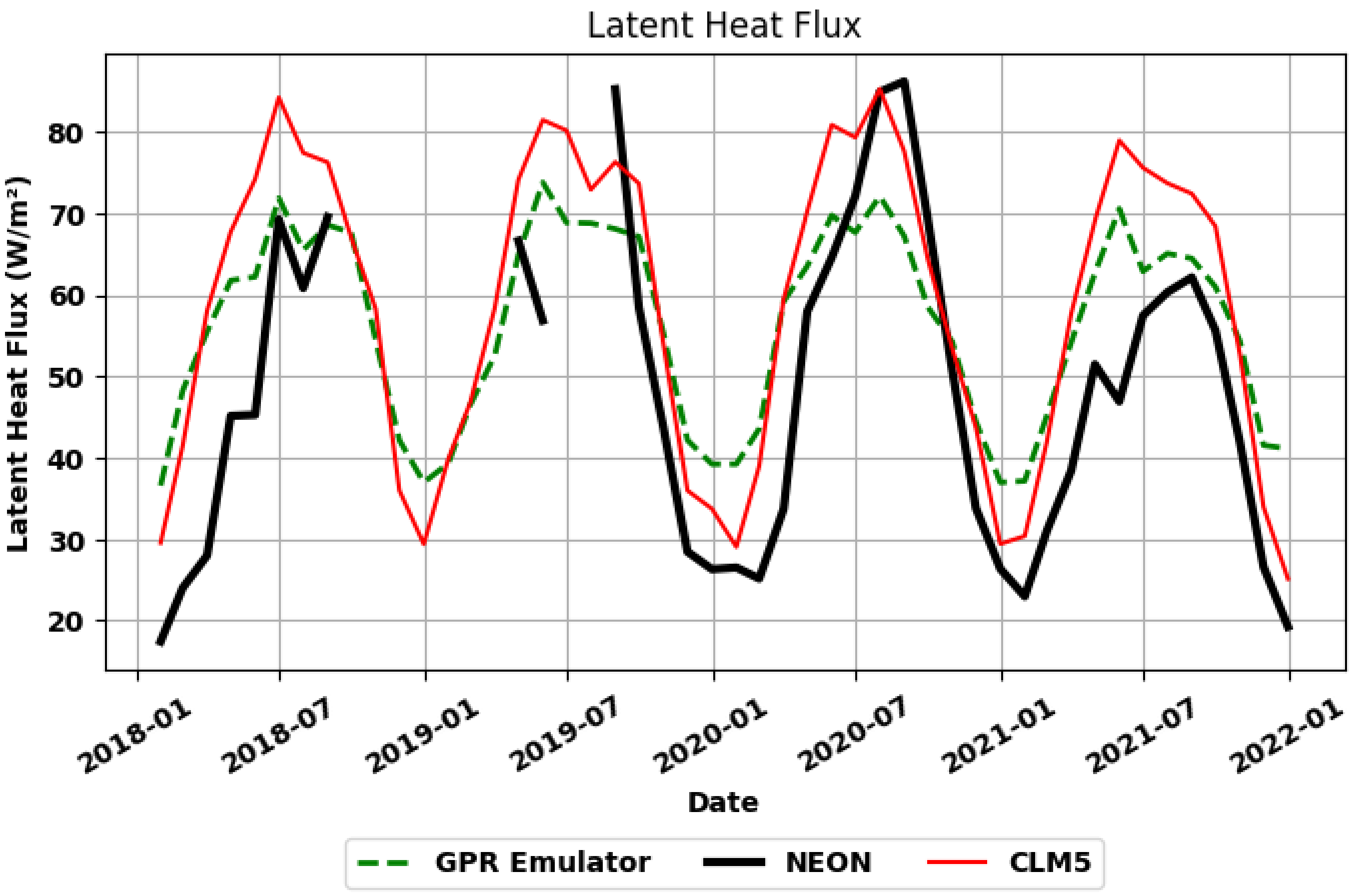
# Calibration improves emulator model performance

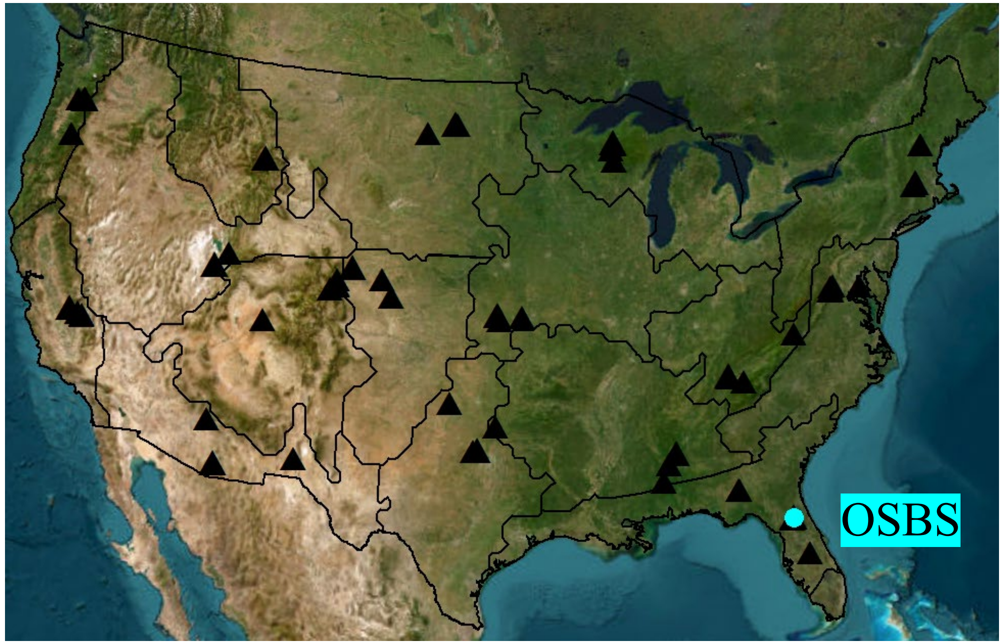



| | RMSE | MBE | R^2 |
|---|---|---|---|
| Default | 26 | -8.5 | 0.59 |
| PPE | 17.95 | -2.28 | 0.74 |
| Emulator | 19 | 5.6 | 0.81 |

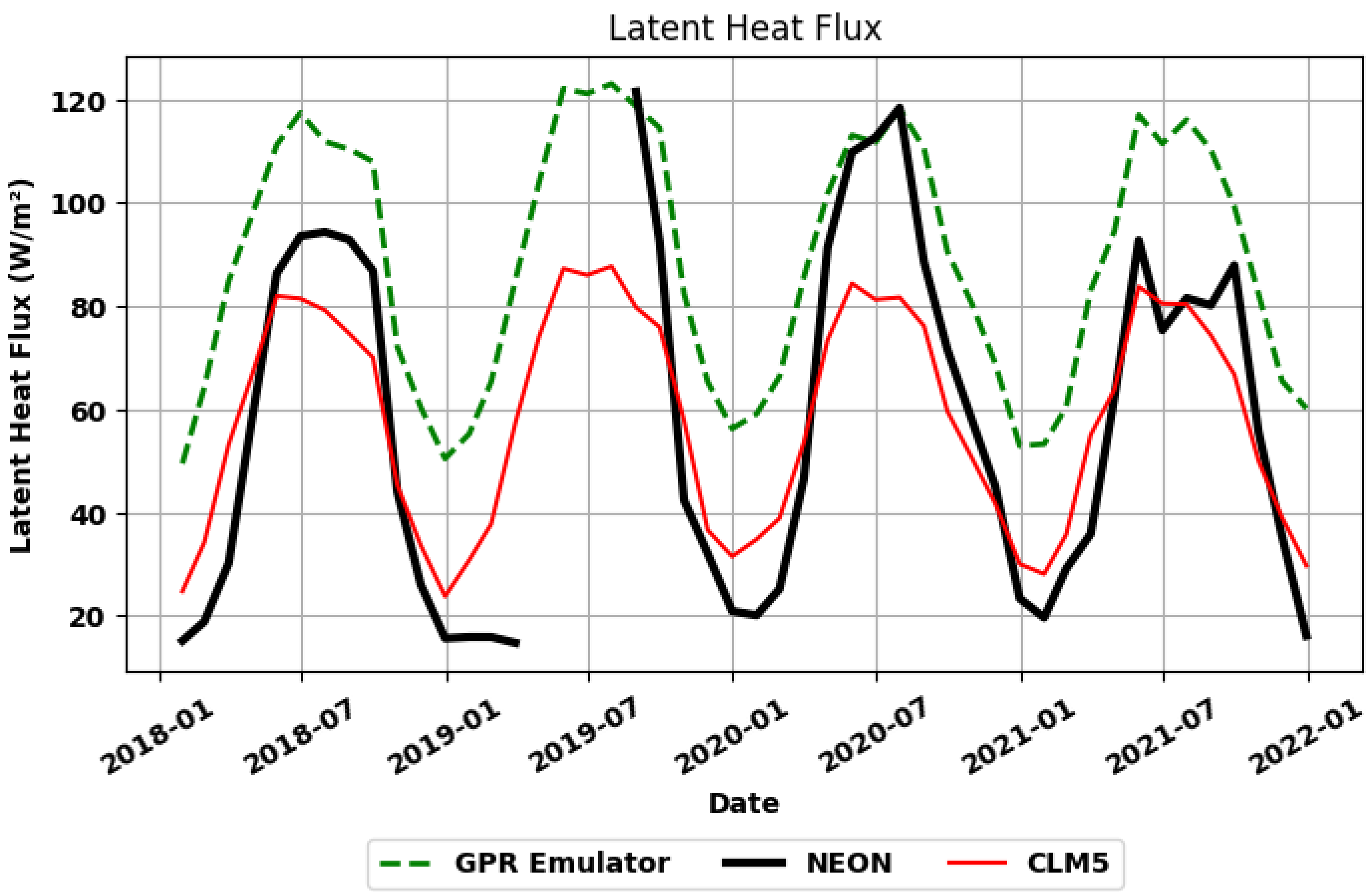
# The TALL calibrated emulator was applied to other NEON sites with similar characteristics.

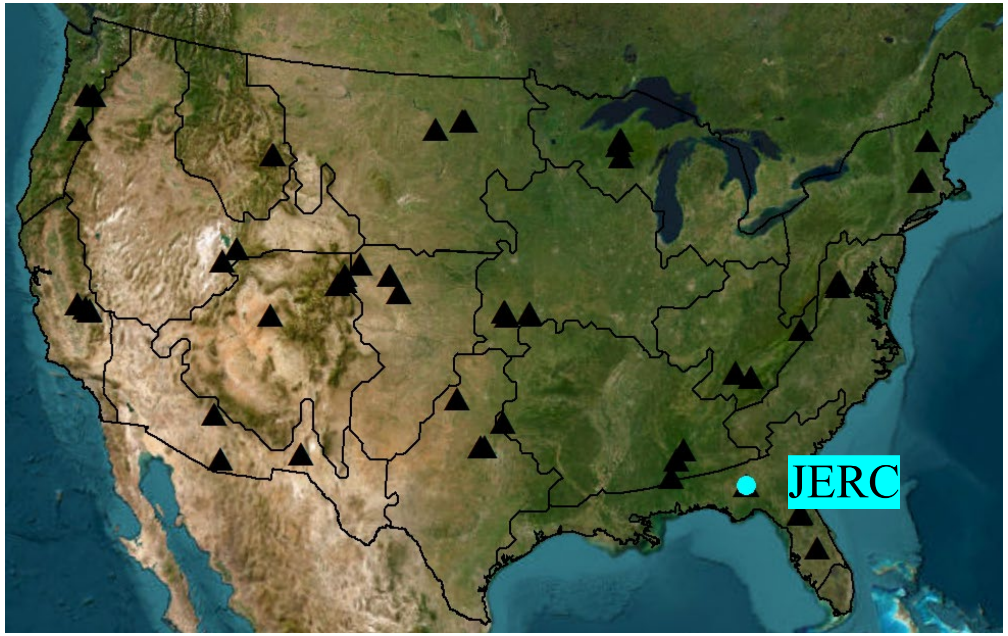



| | RMSE | MBE | $R^2$ |
|---|---|---|---|
| Default | 20 | 12 | 0.58 |
| Emulator | 18 | 11 | 0.67 |

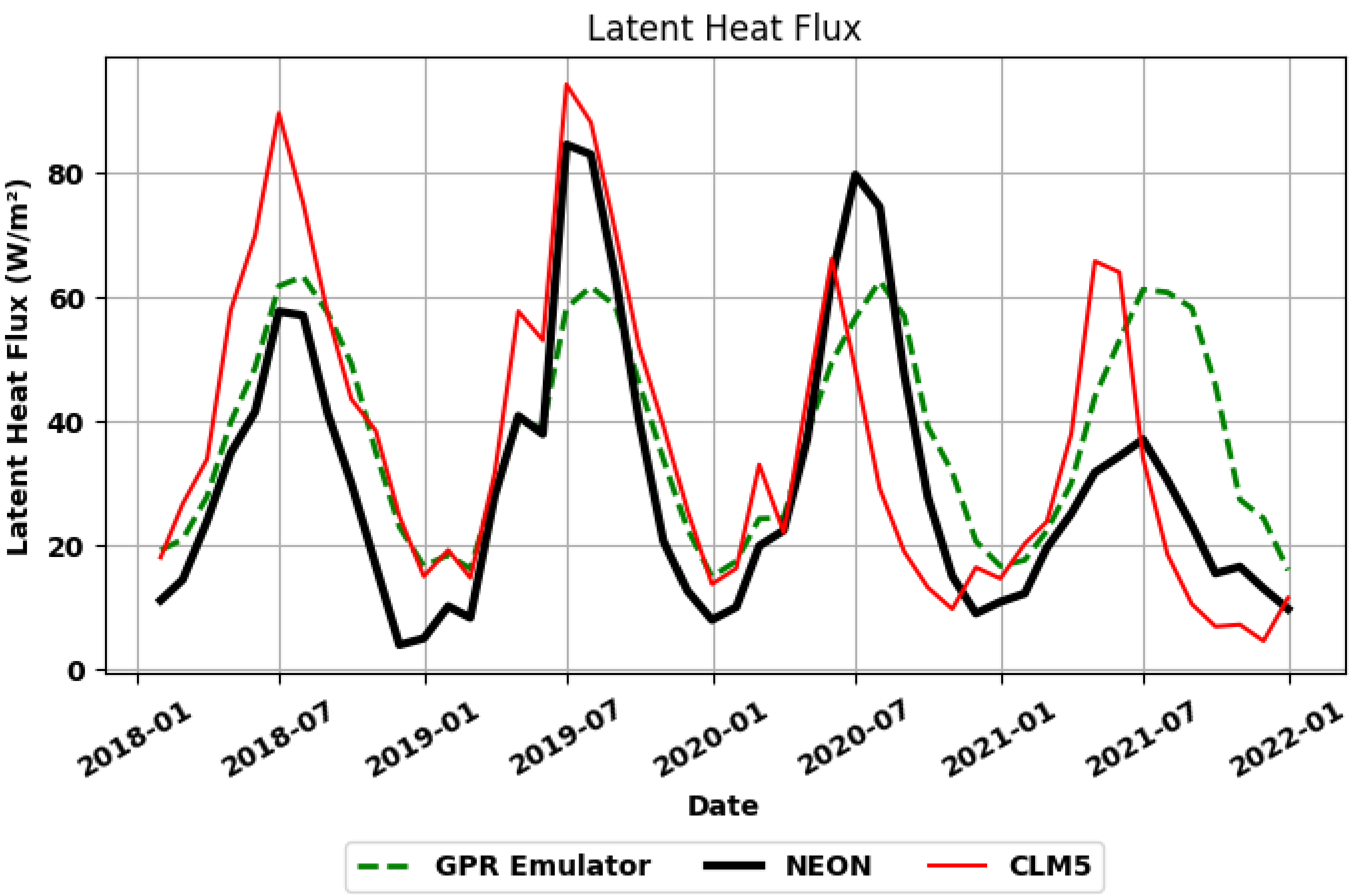
The TALL calibrated emulator was applied to other NEON sites with similar characteristics.

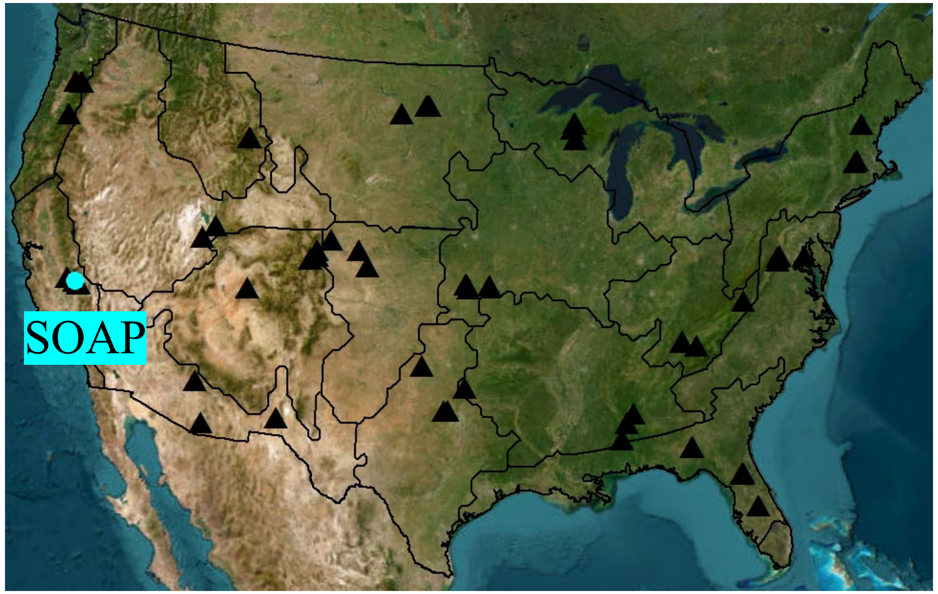



| | RMSE | MBE | R^2 |
|---|---|---|---|
| Default | 24 | -1.5 | 0.68 |
| Emulator | 34 | 28 | 0.80 |

The TALL calibrated emulator was applied to other NEON sites with similar characteristics.




| | RMSE | MBE | $R^2$ |
|---|---|---|---|
| Default | 17 | 7.1 | 0.51 |
| Emulator | 20 | 5 | 0.61 |

# Computation Cost: Comparing CLM-NEON run versus emulator

- CLM-NEON (point scale simulations - ~9 minutes
- CLM-NEON With initialization ~ 5.5 hrs. for 200 years
- Emulator - ~ 2 seconds

# Summary

- Emulator - fast prediction, parameter tuning, and sensitivity analysis
- Emulating CLM5 ET highlights the adaptability of the approach for various land surface model processes
- Emulator provides more flexibility – parameter sensitivity