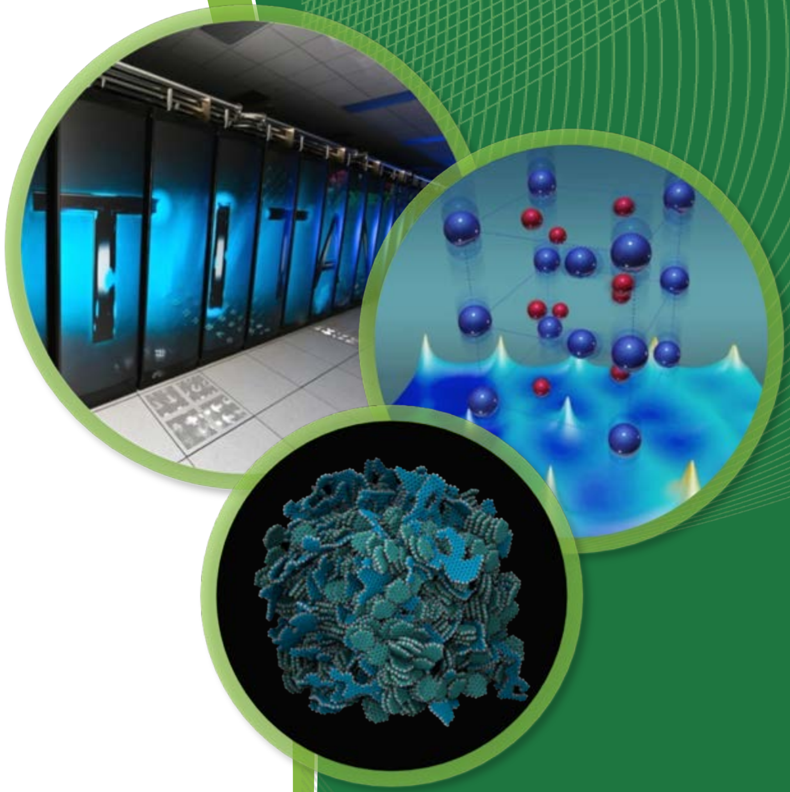
Optimization of Sensor Networks for Climate Models (OSCM): Overview

July 24th, 2018

Daniel Ricciuto[1], Khachik Sargsyan[2], Youssef Marzouk[3],

Dan Lu[1], Cosmin Safta[2], Kate Evans[1], Miroslav Stoyanov[1], Jayanth Jagalur Mohan[3], Fengyi Li[3]

[1]Oak Ridge National Laboratory
[2]Sandia National Laboratories
[3]Massachusetts Institute of Techology



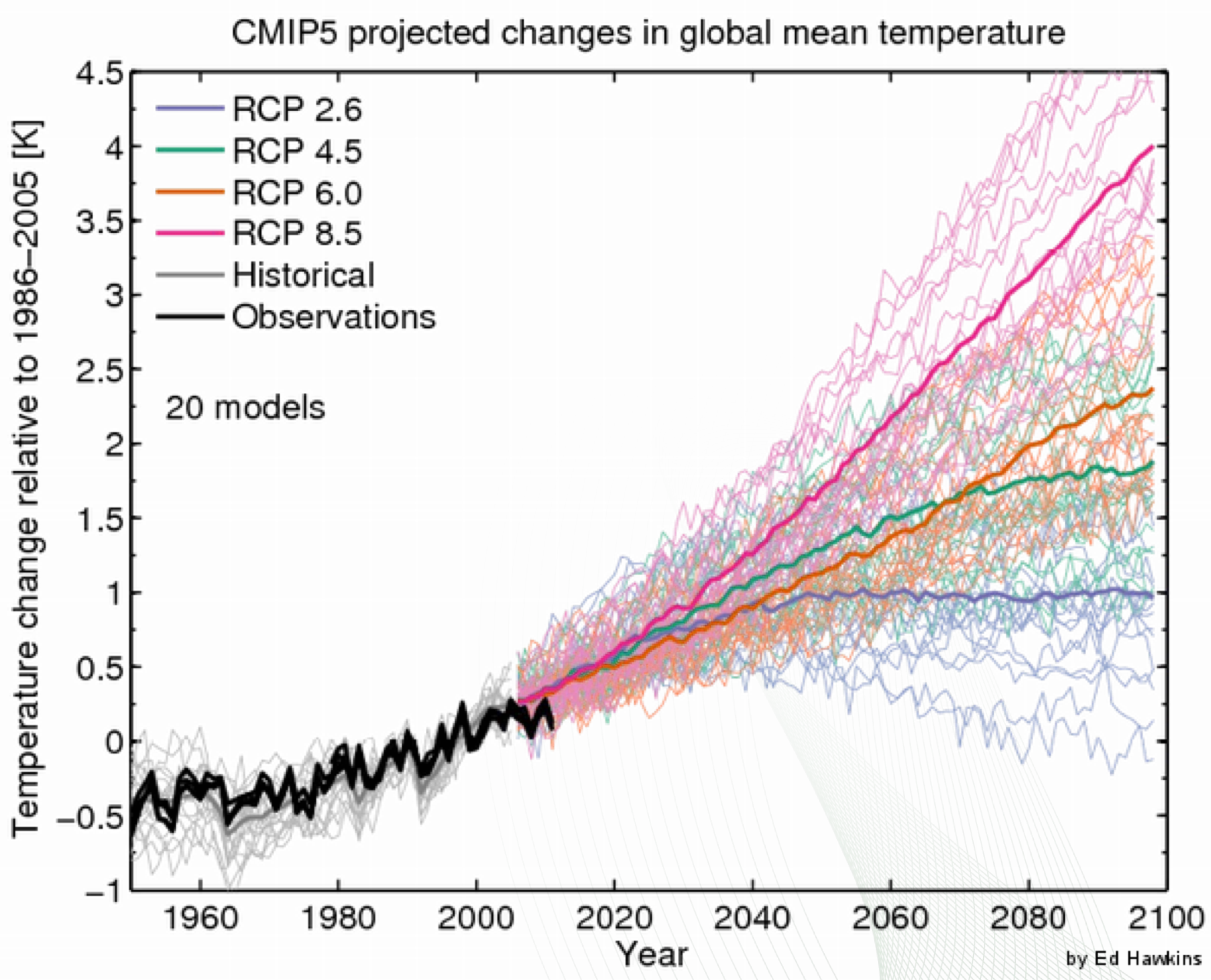
ORNL is managed by UT-Battelle
for the US Department of Energy

Sandia National Laboratories

OAK RIDGE National Laboratory

Key Challenges

- Earth System models like E3SM are computationally expensive to do even a single forward simulation, but contain many uncertain parameters/processes.
- Uncertainty quantification in or calibration of an ESM requires ensembles, which can become quite large given uncertain parameters
- Many of the relevant land and atmospheric observations are at “point” scale (1 model grid cell or smaller) – how can we best use these?



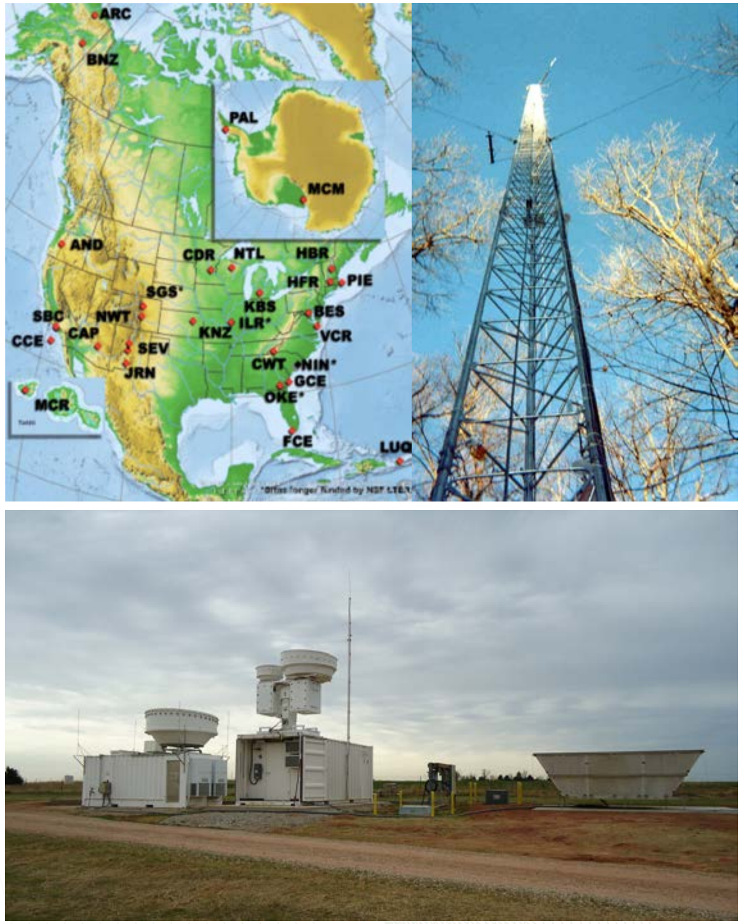
OAK RIDGE National Laboratory

Project objectives

- A new framework enabled by new UQ algorithms and high performance computing to guide placement of new observations to maximize uncertainty reduction in climate model predictions.

- A network of single-column coupled land-atmosphere models co-located with existing and proposed measurement networks, calibrated with existing or synthetic observations.

- Propagation of uncertainty with mutli-fidelity approach: Multi-level Monte Carlo (MLMC) can be used to propagate uncertainties *in fully coupled mode* over a range of fidelity and resolution.



Land/atmosphere measurement networks

Major project tasks

UQ methods development:

- Improve sensitivity analysis and surrogate modeling methods for E3SM
- Develop model calibration techniques, including representation of structural error (embedded approach)
- Implement methods to reconstruct global maps from multi-point simulations
- Develop algorithms for Network Optimization using E3SM single column models
- Develop and implement Multi-level Monte Carlo for uncertainty propagation in E3SM

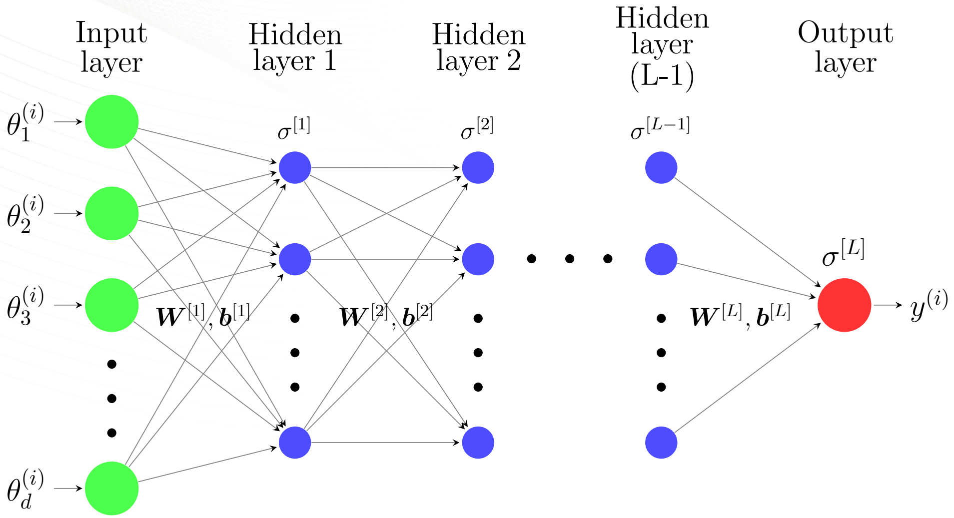
Earth System Model Infrastructure:

- Single column ensemble infrastructure and computational architecture
- Data synthesis for network calibration, process identification and validation

Oak Ridge National Laboratory

Strategies for building E3SM surrogate models

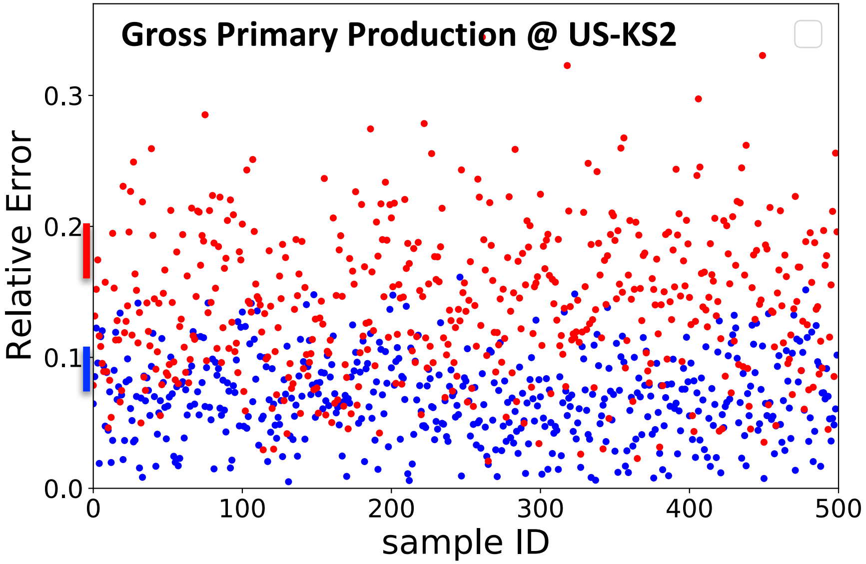
- Goal: Develop accurate surrogate models for E3SM-Land Model (65 input parameters)



3K simulations (approx. 50K CPU hours/grid cell)

- 2.5K samples for training using cross-validation to determine regularization knobs
  - 0.5K samples for testing
  - reshuffle and re-process data

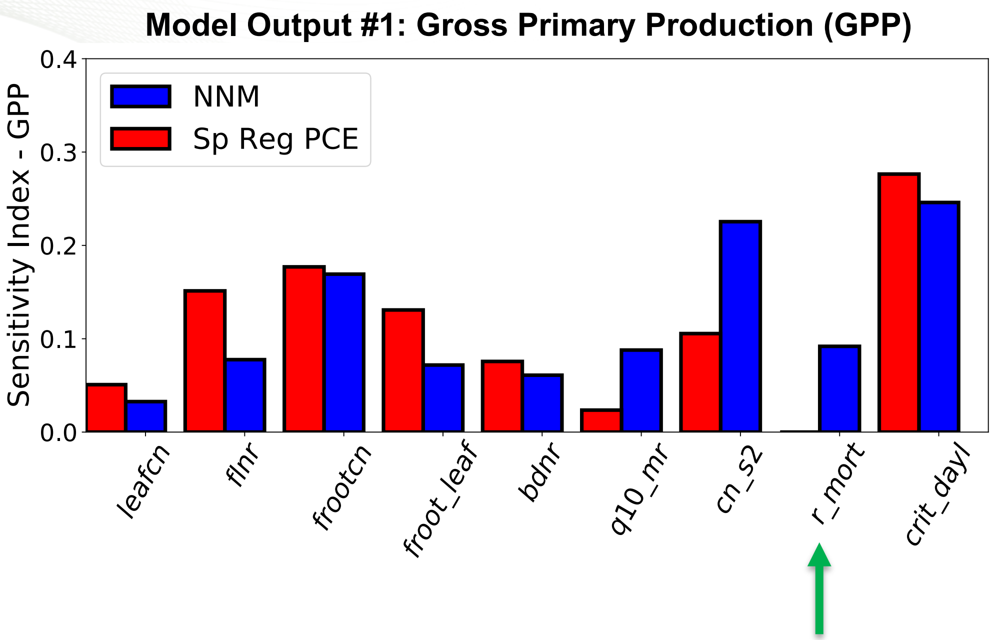
- The Neural Network Model (NNM) shows *increased accuracy* compared to Sparse Regression Polynomial Chaos Expansions for the ELM
  - **Sparse Regression PCE Model Error: 0.15…0.19**
  - **Neural Network Model Error: 0.06…0.09**

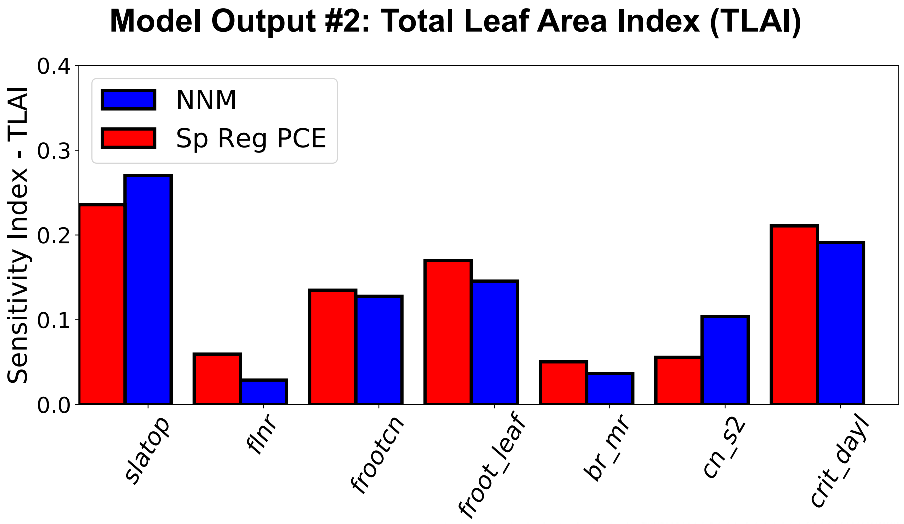


- ***Next steps*:**
  - Explore additional regularization techniques (l_1, adversarial networks), embedding physical constraints, …

Model Surrogates for sensitivity analysis

Land model simulation results for Kennedy Space Center Site (US-KS2)





- Goal: Determine key parameter sensitivities for dimensionality reduction
- Sensitivity Analysis selected a subset of 10 principal parameters (out of 65) that impact selected model outputs
  - *The Neural Network Model captured key sensitivities that were missed by previous methods*

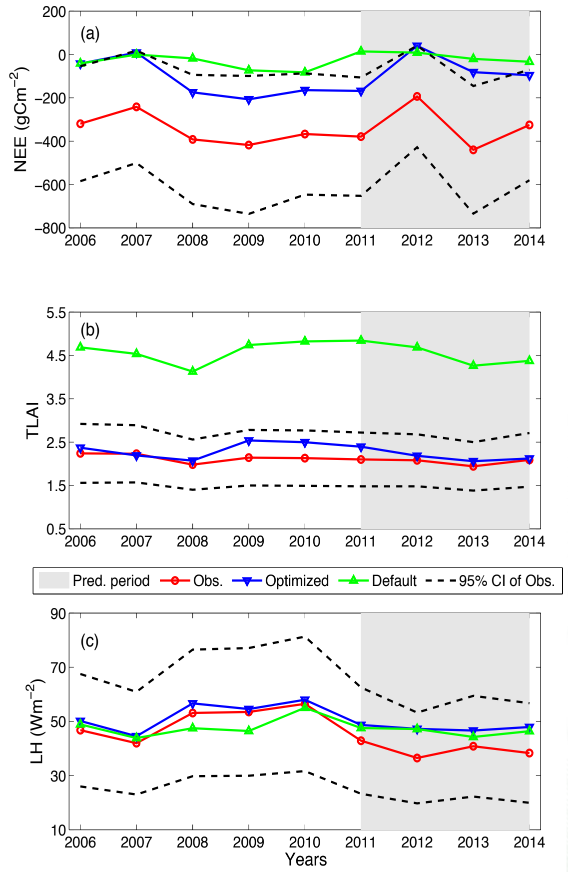
Single gridcell model calibration

OSCM goal: Calibrate E3SM land and atmosphere models using surrogate modeling approaches

Calibration of the E3SM land model using surrogate based global optimization

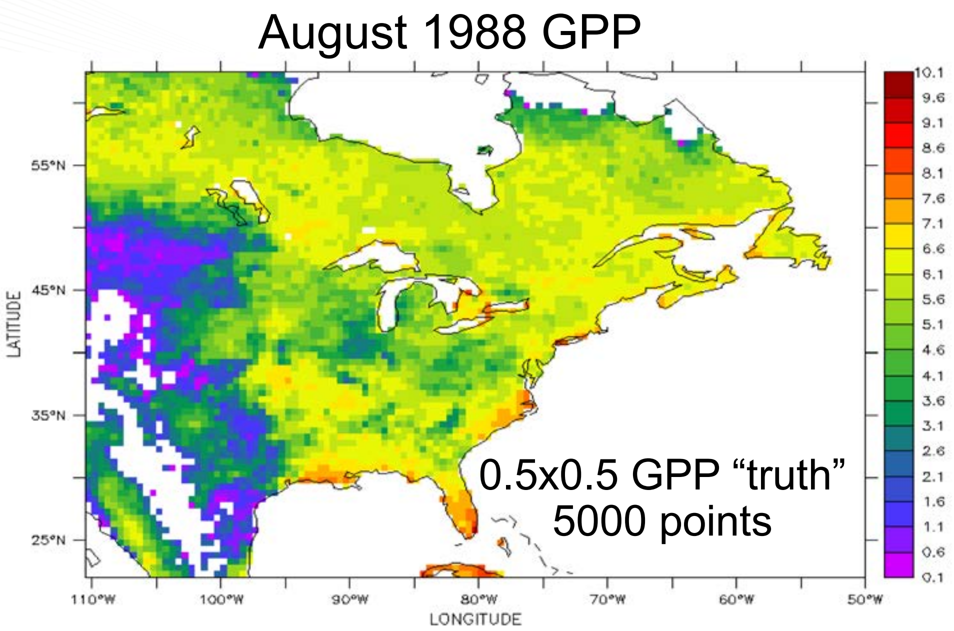
Objective	• Calibrate E3SM land model (ELM) to improve model projection of carbon fluxes.
New science	• Advanced sparse grid interpolation to construct a fast-to-evaluate surrogate system of the ELM. • Efficient global optimization algorithm to calibrate the ELM and find the optimal parameter values.
Significance	• An accurate surrogate model can be created for the ELM to reduce the model evaluation time. • Application of the optimized parameters improves ELM performance and predictive capability of carbon fluxes.

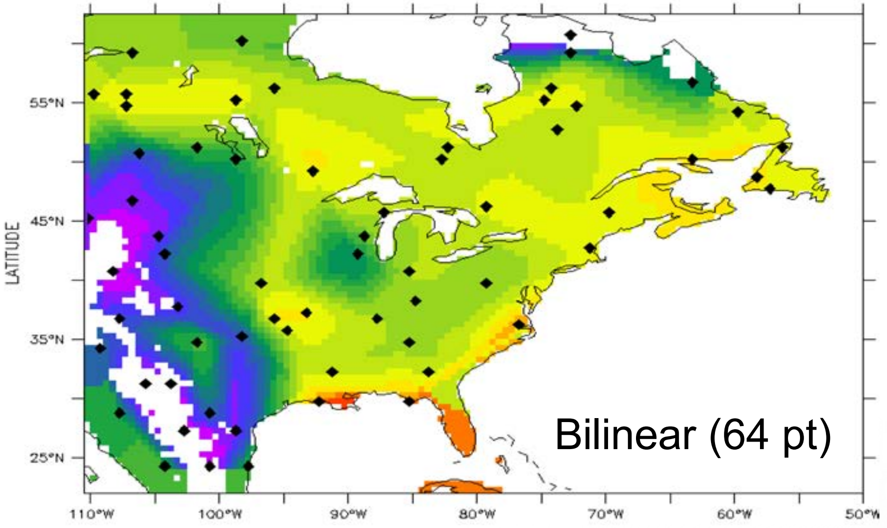
Lu, Ricciuto, Stoyanov, Gu (2018). “Calibration of the E3SM land model using surrogate based global optimization.” Journal of Advances in Modeling Earth Systems. Accepted.

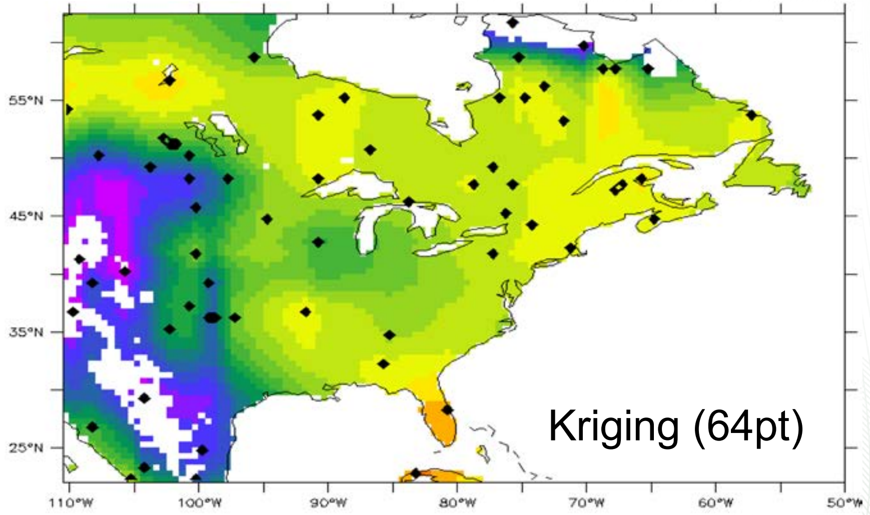


The optimized parameters improve model fit to observation (Obs.) compared to the default parameter values.

Reconstruction and network design







Goal: Reproduce model spatial variations using a network of single column simulations

- Simple ELM running at MIT and interface with OED system being developed

- OED problem using Bayesian linear-Gaussian setting, with "greedy" selection strategy for some simple synthetic problems.
- Provide performance guarantees for the greedy strategy that bound how much worse it could be than the combinatorial optimum

OAK RIDGE National Laboratory

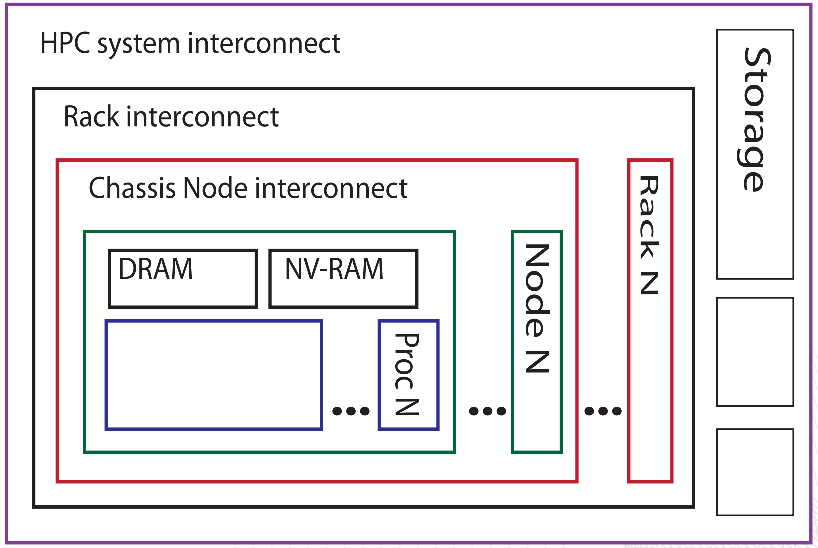
Atmosphere SCM ensemble infrastructure

Computational challenges

- Scaling from ~10s to ~10K+ SCM simulations
- I/O limitations (based on experience with ELM)
- How to optimize ensembles for pre-exascale architectures

Model development challenges (E3SM/CMDV)

- Eulerian dycore → SE dycore
- Development of relevant use cases
  - Locations of interest
  - Length of simulations

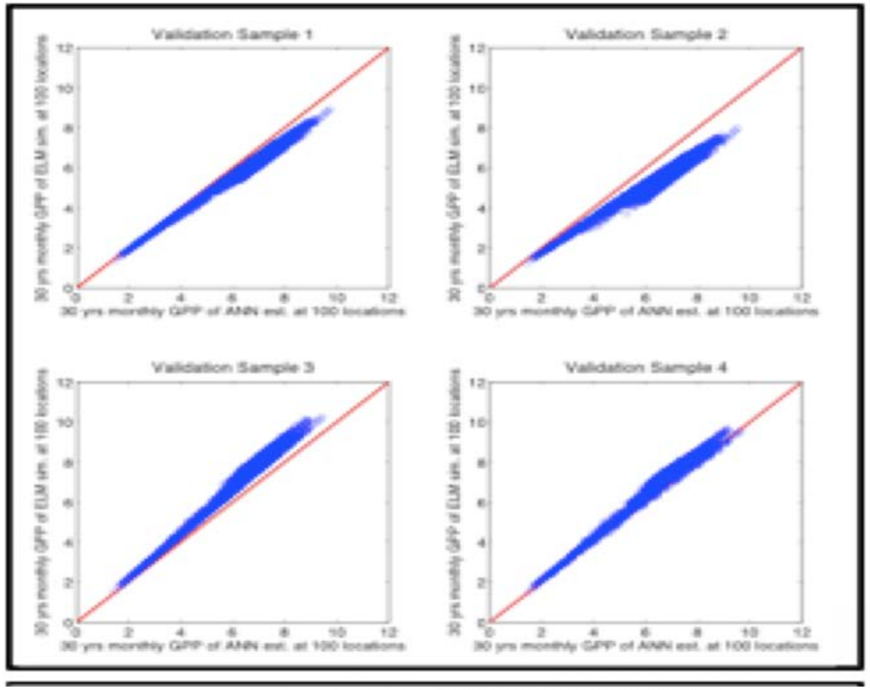


OAK RIDGE National Laboratory

Collaborations: Surrogate modeling and machine learning

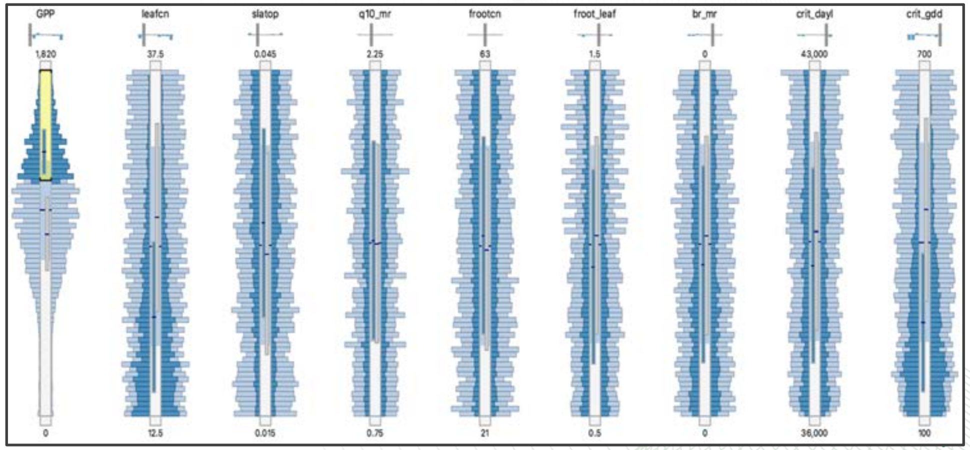
Machine learning

- Tuning of neural network hyperparameters
- Bayesian Neural network approaches
- Implementation on LCF facilities
- Neural networks with physical constraints (e.g. we can't have negative leaf area)



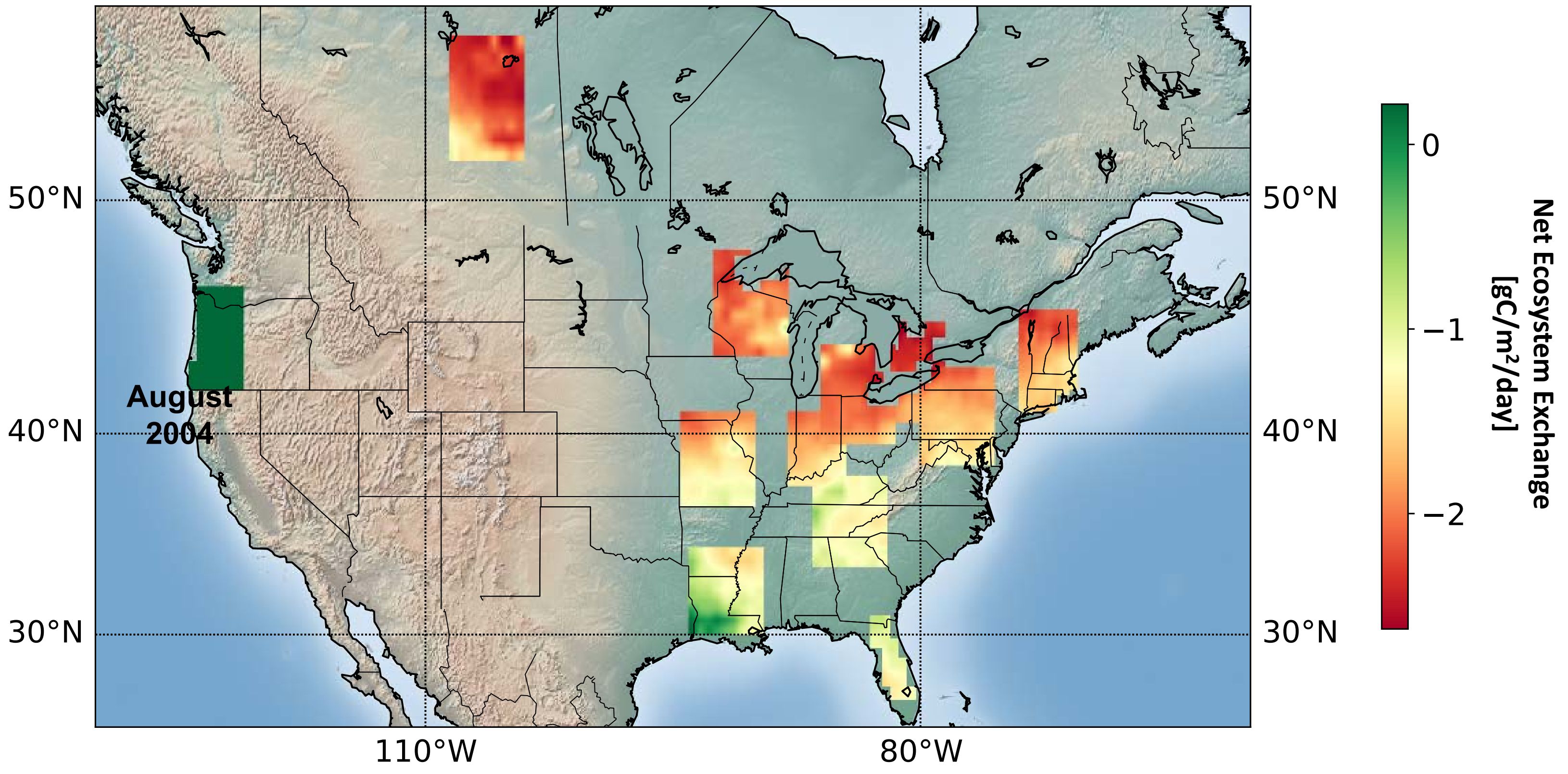
Ensemble analysis/visualization

- reducing the parameter space to explore
- Calibration "on the fly"
- Efficient calculation of ensemble statistics



EDEN visualization tool

sELM – In Progress: Exploring Model Structure and Calibration Work



- Partnership with FASTMath Institute to explore low-rank functional tensor train representations
  - discover model structure in the combined spatial-stochastic spaces
- Use sensitivity analysis and model error embedding to calibrate model components that are responsible the most for discrepancies between predictions and observations.
- Optimal experimental Design (OED) formulations

Next steps and potential collaborations

- Create surrogate models for land-atmosphere QOIs using efficient techniques.
- Further develop our OED system to link surrogate models with sensor placement optimization
- Use both land and land-atm coupled SCM in this framework
- Demonstrate E3SM successful use case for multi-fidelity uncertainty quantification.
- Potential collaboration with other BER SciDAC projects
  - I/O issues for large network ensembles
  - Coupling for ensemble simulations
  - UQ algorithms development