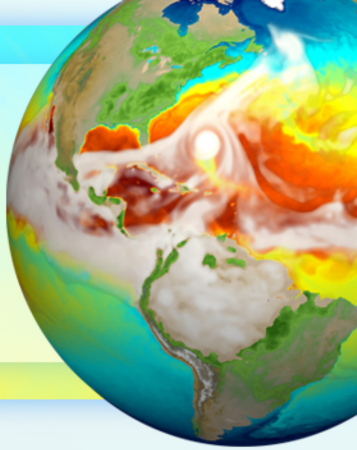
Uncertainty quantification methods and applications for the E3SM land model



Daniel M. Ricciuto (ORNL) and Khachik Sargsyan (SNL-CA)

Cosmin Safta, Vishagan Ratnaswamy (SNL-CA)
Dan Lu, Peter Thornton (ORNL), Jennifer Holm (LBNL)
Jayanth Jagalur Mohan, Youssef Marzouk (MIT)

CBGC Webinar
September 24th, 2019

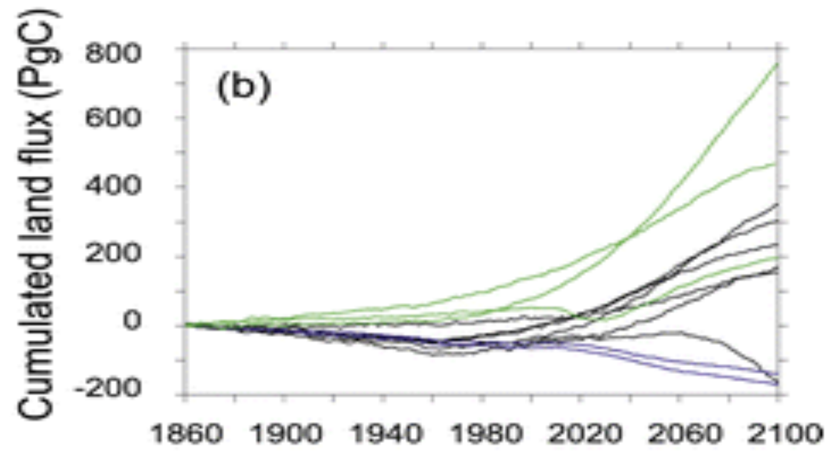
E3SM Energy Exascale Earth System Model
SciDAC Scientific Discovery through Advanced Computing
FASTMATH
U.S. DEPARTMENT OF ENERGY

Overview and motivation: CBGC models

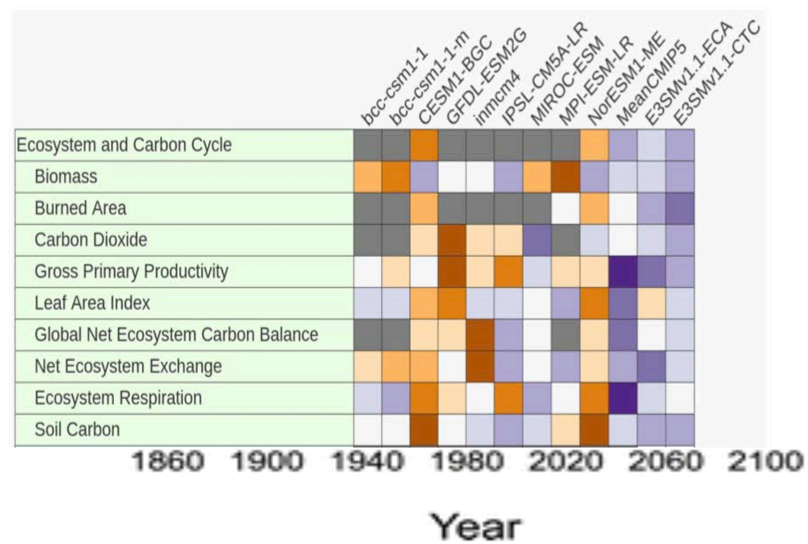
- Uncertainty from Multi-model ensembles
- Large spread in outputs
- Many quantities of interest
- Little formal uncertainty quantification (UQ)
  - Expensive model evaluation
  - High dimensionality

UQ challenges in E3SM and SciDAC:

- What processes drive uncertainty?
- What accounts for the key differences among models?
- Can model calibration using observations (e.g. satellite data) reduce uncertainty?

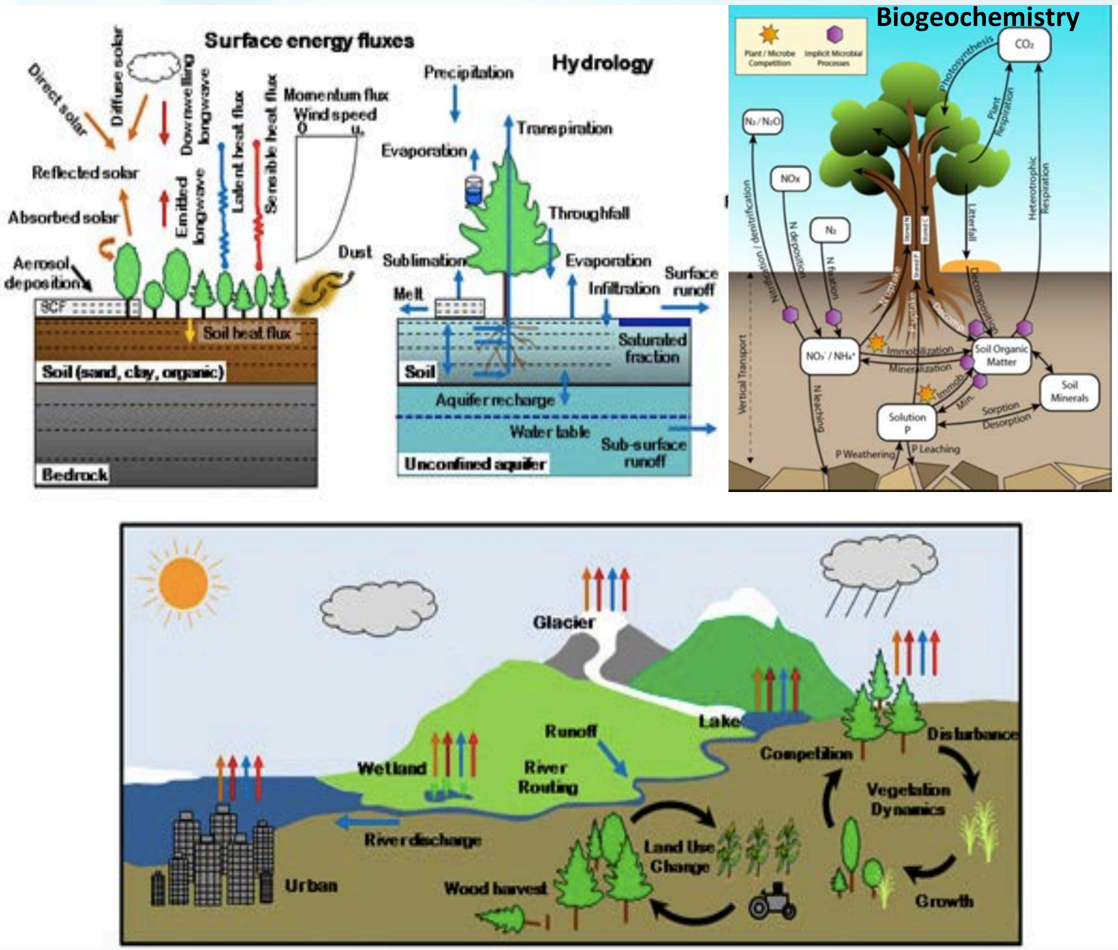


Friedlingstein et al (2014)



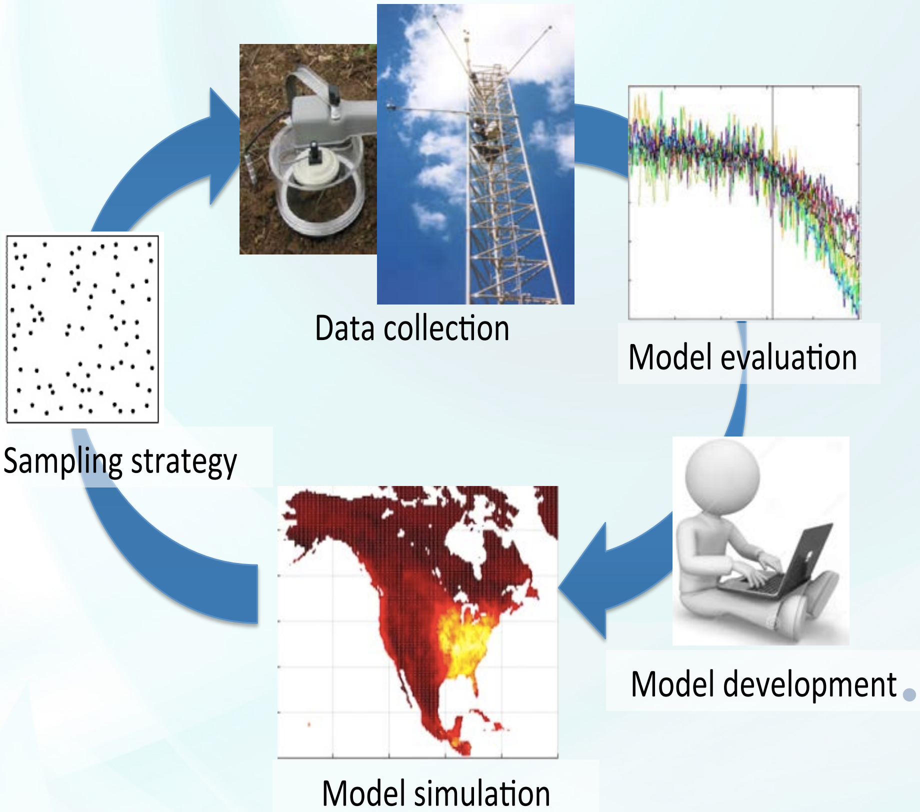
Burrows et al (in review)

Overview and motivation: Land BGC



- ELM is an increasingly complex model with many processes
- Large ensembles are needed for UQ
- Despite relatively small land contribution to computation, spinup limits ensemble size.
- Surrogate models can increase the efficiency of sensitivity analysis and calibration

SciDAC: Optimization of Sensor Networks for Improving Climate Model Predictions (OSCM)

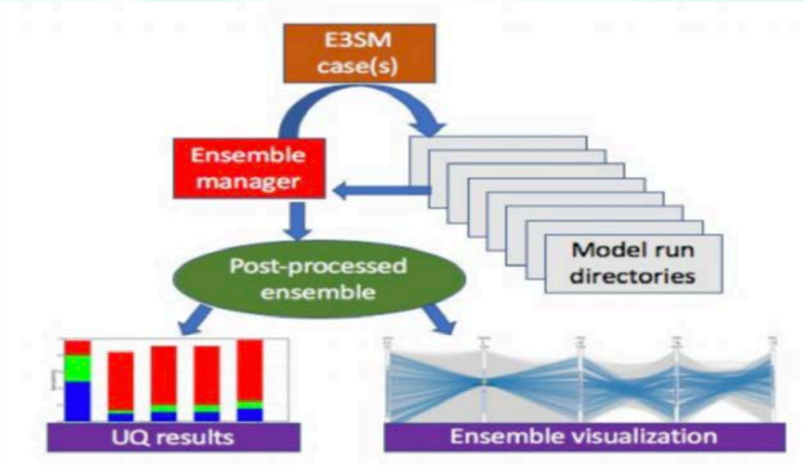


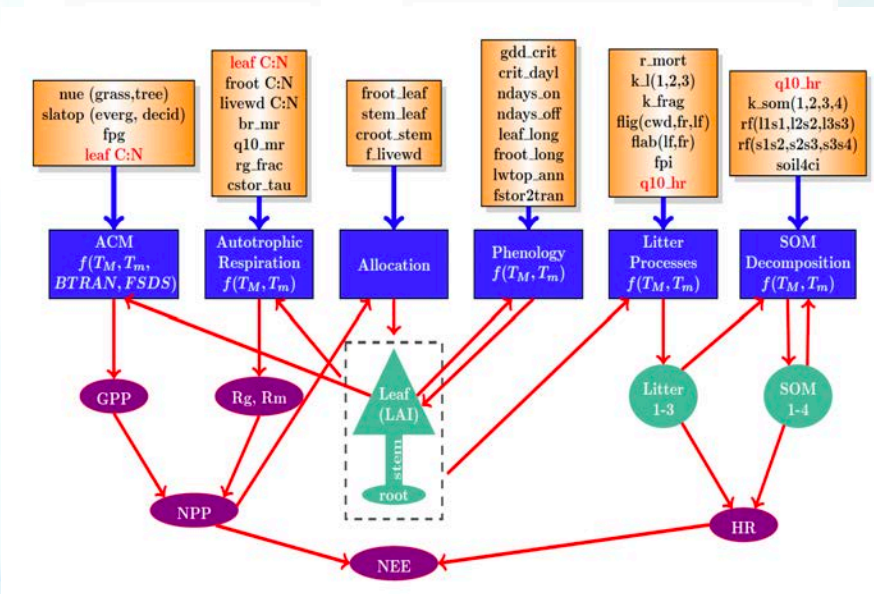
A “MODEX” loop

- *Formalizing MODEX*: A key DOE initiative for iterative feedback between models, experiments and observations
  - Develop UQ algorithms to characterize uncertainty and maximize uncertainty reduction from new observations
  - Apply to the E3SM land model for existing and proposed new observation networks

Key question: What is the *ideal placement of observation systems* to represent spatial and temporal variability in a signal of interest?

Model infrastructure to enable UQ





- E3SM single gridcell
  - Default v1 (FLUXNET sites)
  - FATES (Boreal Alaska site)
  - Using mpi4py as part of Offline Land Model Testbed (OLMT) to manage ensembles and post-process
- Simplified ELM
  - Carbon cycle only
  - 47 ELM parameters
  - Photosynthesis submodel – ACM or ELM NN fit
  - 100x faster than ELM for regional UQ testing

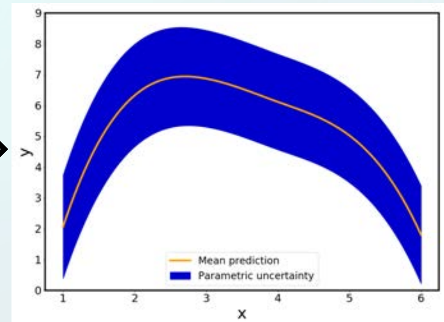
Uncertainty Propagation
… enabled by Surrogate Models

λ

Input parameters

Comp. model

$f(x;\lambda)$



Output prediction

- *Forward predictions:*
  - surrogate models,
  - sensitivity analysis,
  - parametric uncertainty

Work with the model as a black-box (**non-intrusive**):
create an ensemble of simulations with varying/perturbing λ

Never analyze the ensemble directly:
build a **surrogate** first
… otherwise called proxy, metamodel, emulator, response surface, supervised ML

$$f(x;\lambda) \approx f_s(x;\lambda)$$

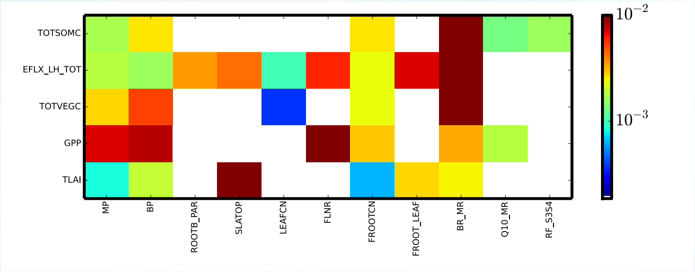
E3SM Energy Exascale Earth System Model

U.S. DEPARTMENT OF ENERGY

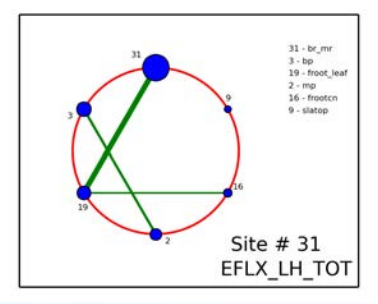
Global Sensitivity Analysis (GSA) enables parameter selection

… otherwise called Sobol indices, variance-based decomposition
Attribute fractions of output variance to input parameters

- i.e., how much output variance would reduce if a given parameter is fixed to its nominal value



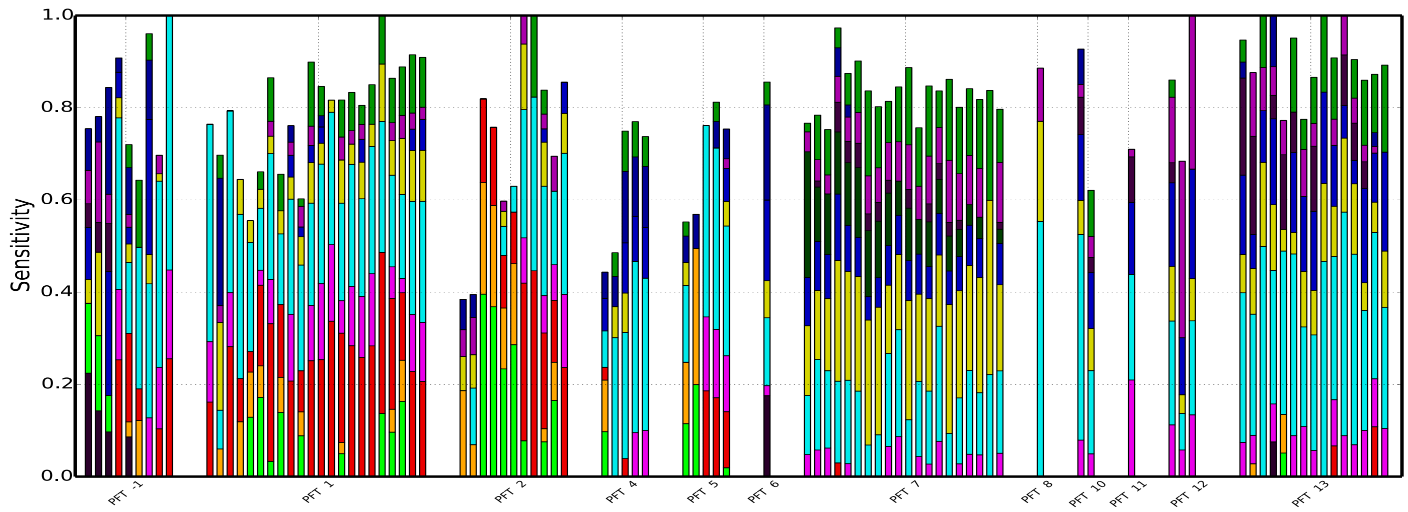
- also generalizes to joint sensitivities: joint parameter impact to a given QoI



E3SM Energy Exascale Earth System Model

U.S. DEPARTMENT OF ENERGY

The impact of parametric uncertainties on biogeochemistry in the E3SM v1 land model

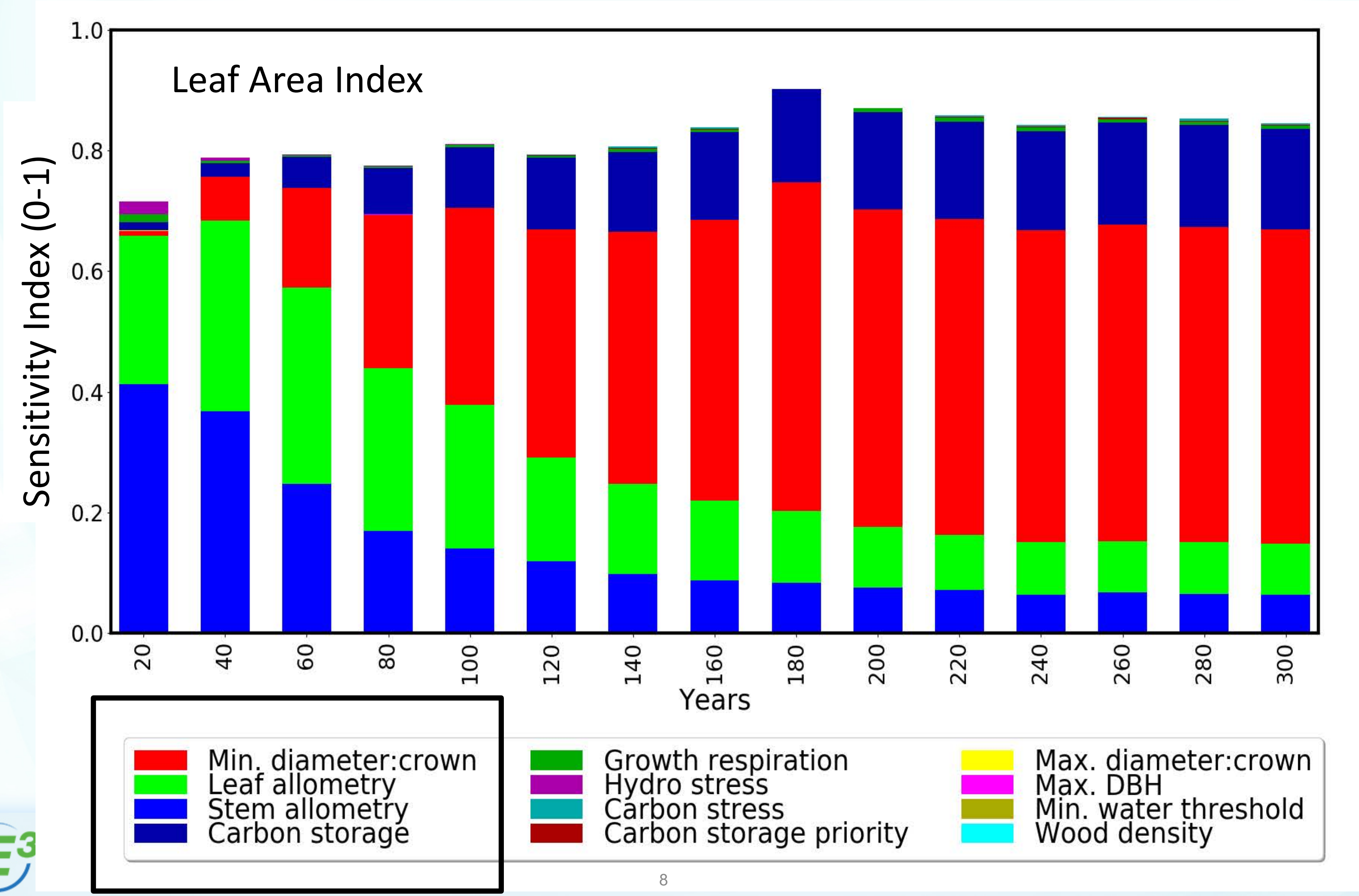


- Critical soil water potential
- Stomatal conductance slope
- Stomatal conductance intercept
- Specific leaf area (canopy top)
- Leaf carbon:nitrogen ratio
- Fraction of leaf N in RuBisCO
- Fine root carbon:nitrogen ratio
- Fine root:leaf allocation ratio
- Critical day length for senescence
- Fraction of C storage allocated
- Bulk dentrification rate
- Base rate for plant respiration
- T sensitivity for plant respiration

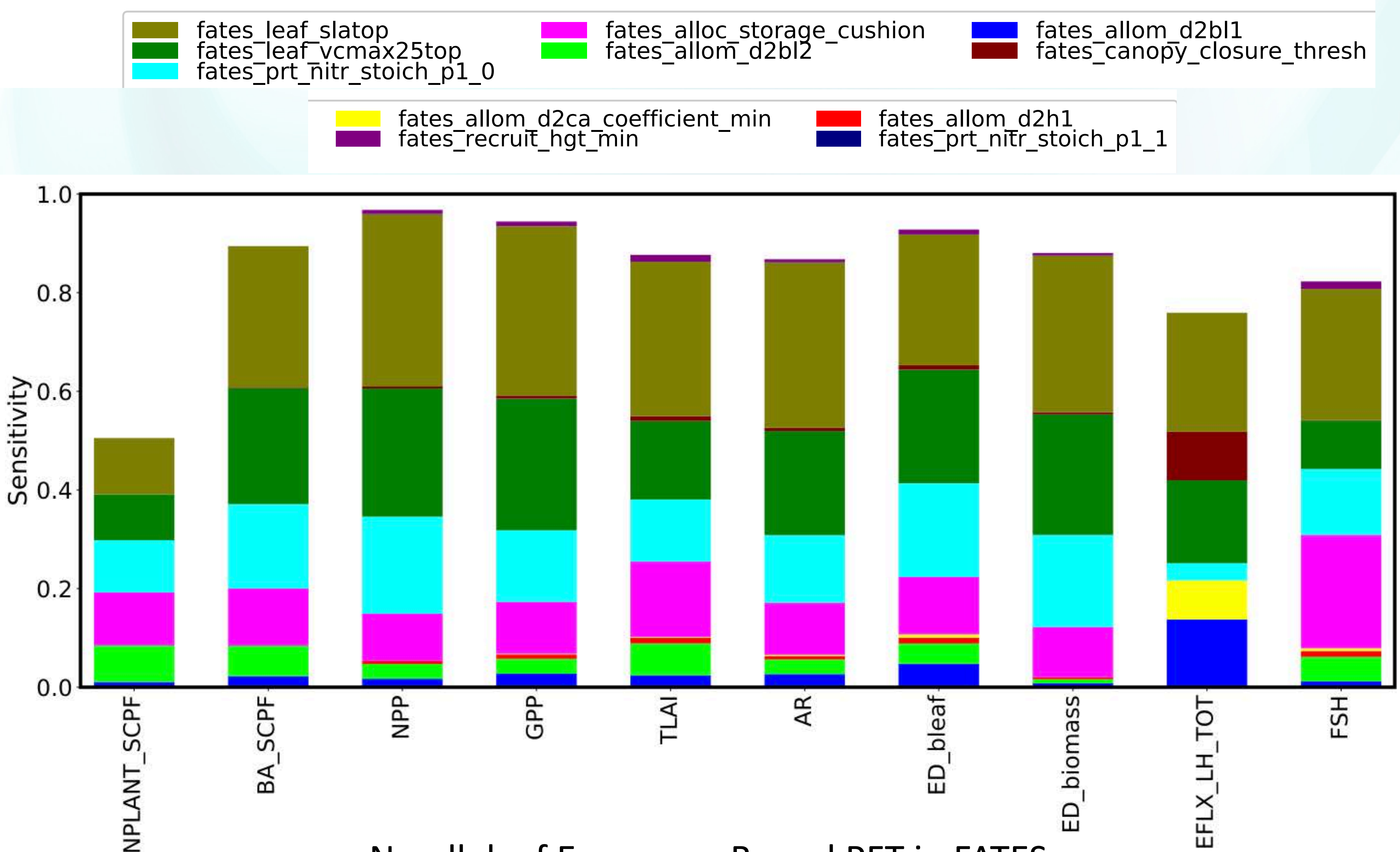
Out of 68 model parameters analyzed, fewer than 20 have a significant influence on land model quantities of interest. These parameter sensitivities depend on climate variables, and are largely consistent among sites within a biome.

Ricciuto, D., Sargsyan, K., & Thornton, P. (2018). The impact of parametric uncertainties on biogeochemistry in the E3SM land model. Journal of Advances in Modeling Earth Systems, 10. https://doi.org/10.1002/2017MS000962.

Boreal forest run over 300 years, influence on LAI



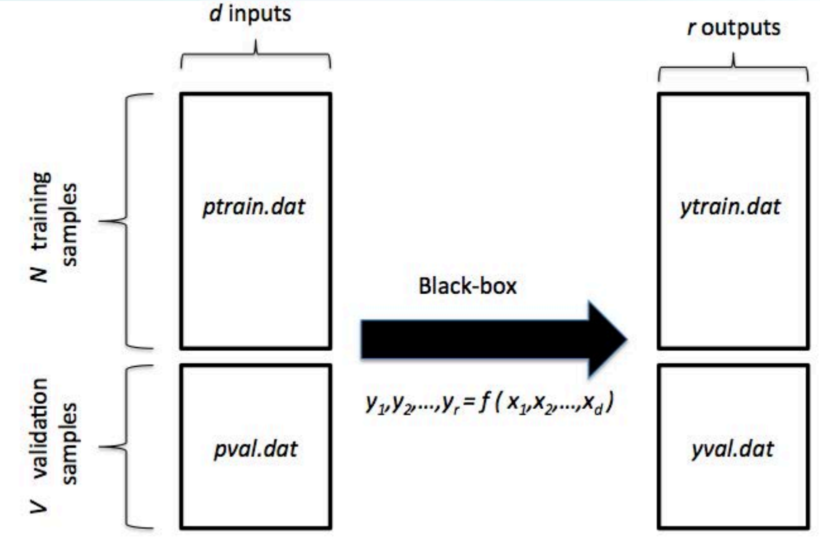
GSA across many QoIs



Needleleaf Evergreen Boreal PFT in FATES

Generic Forward UQ Workflow

- git clone git@github.com:E3SM-Project/Uncertainty-Quantification.git
- Python scripts utilizing UQTk Toolkit (www.sandia.gov/uqtoolkit)
  a lightweight C/C++ UQ toolkit from SNL-CA,
  part of tools within FASTMath SciDAC Institute
- Several demos available (uncertainty propagation, surrogate, sensitivity)
- Plotting scripts for quick automated analysis
- See confluence
- Bottom line:
  matrices of ensemble
  inputs and outputs
  and turnkey



Various surrogate types explored

UQ

- **Polynomial chaos (PC):**
  - Misnomer: nothing to do with chaos as in dynamical systems
  - Essentially a polynomial fit/regression to the black-box model
  - Extremely convenient for uncertainty propagation, moment estimation, global sensitivity analysis
    - e.g., PC surrogate allows extraction of sensitivity indices 'for free'
  - Can deal with highly non-linear models, but certain level of smoothness is assumed

- **Low-rank tensor representations:**
  - Nature is low-rank: only subset of inputs act together at the same time
  - More flexible than PC, but harder to construct

- **Neural networks:**
  - Can deal with non-smooth behaviors
  - Cons: much harder to train, even harder to interpret

ML

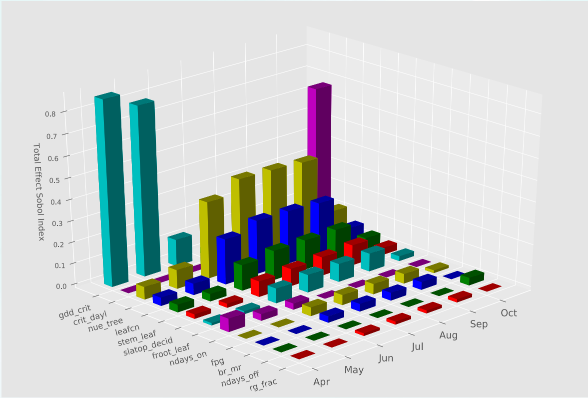
Sparse and Low-Rank Surrogates

- Exploit model structure to reveal sparse low-rank interactions between model components and associated parameters
  - Surrogate model accuracy 4-8%; improvement of by a factor of 2 over classical surrogate model approaches

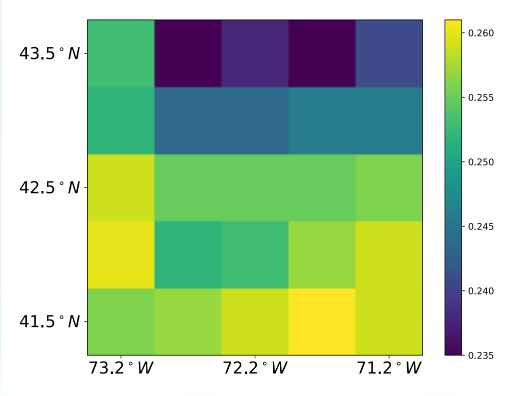
$$f(x_1, x_2, \ldots x_d) = \sum_{i_0=1}^{r_0} \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} f_1^{(i_0 i_1)}(x_1) f_2^{(i_1 i_2)}(x_2) \cdots f_d^{(i_{d-1} i_d)}(x_d)$$

- Explore parametric functional tensor train representations to augment the low-rank models over the Land Model inputs with spatio-temporal dependencies (Joint work with FASTMath)

Total Effect Sobol Indices at US-Ha1

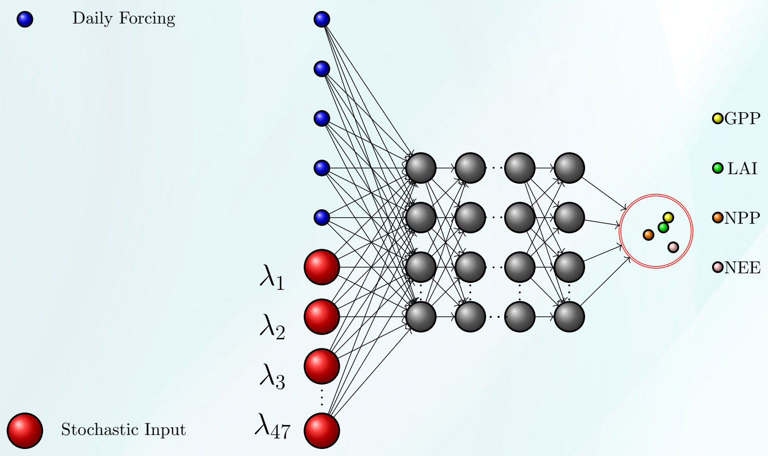


LEAFCN Sobol Index near US-Ha1

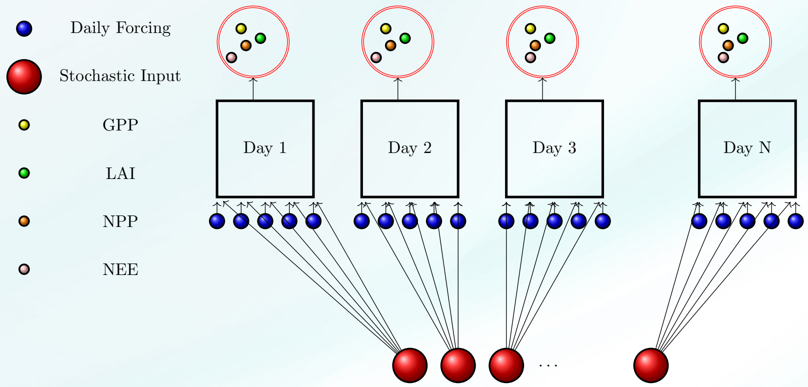


Neural Network surrogates allow more flexibility

Multilayer Perceptron (MLP)

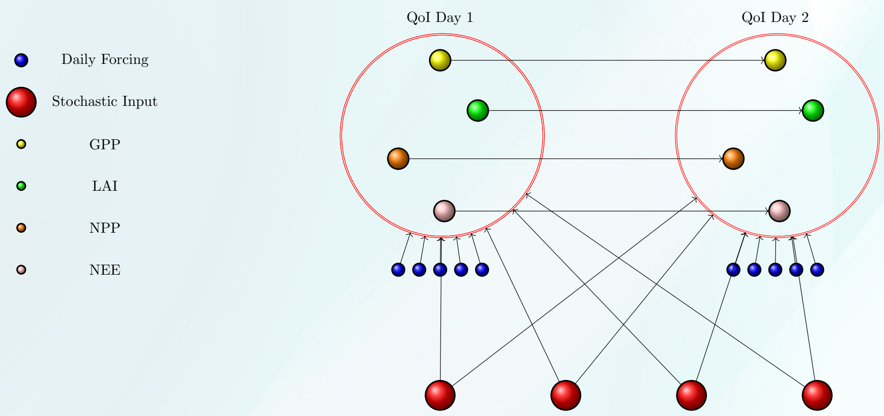


Recurrent Neural Network (RNN)

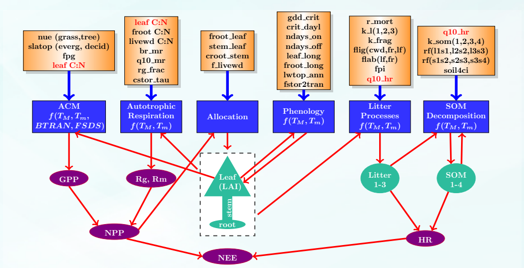


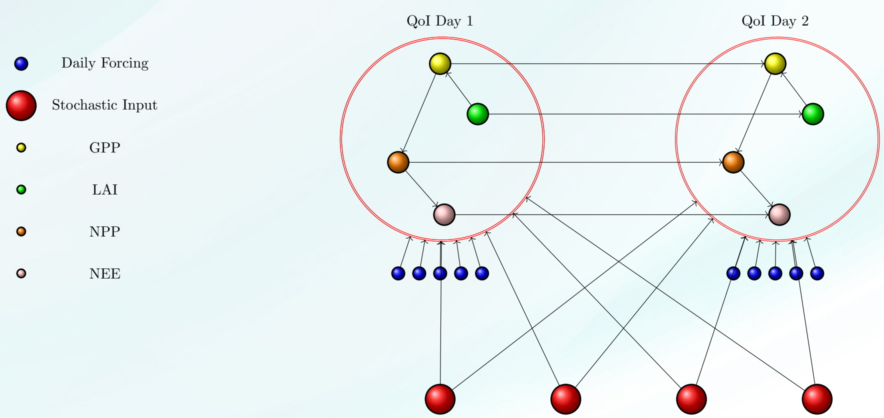
We have created specialized RNN architecture knowing the connections between processes

Vanilla long short-term memory (LSTM) network

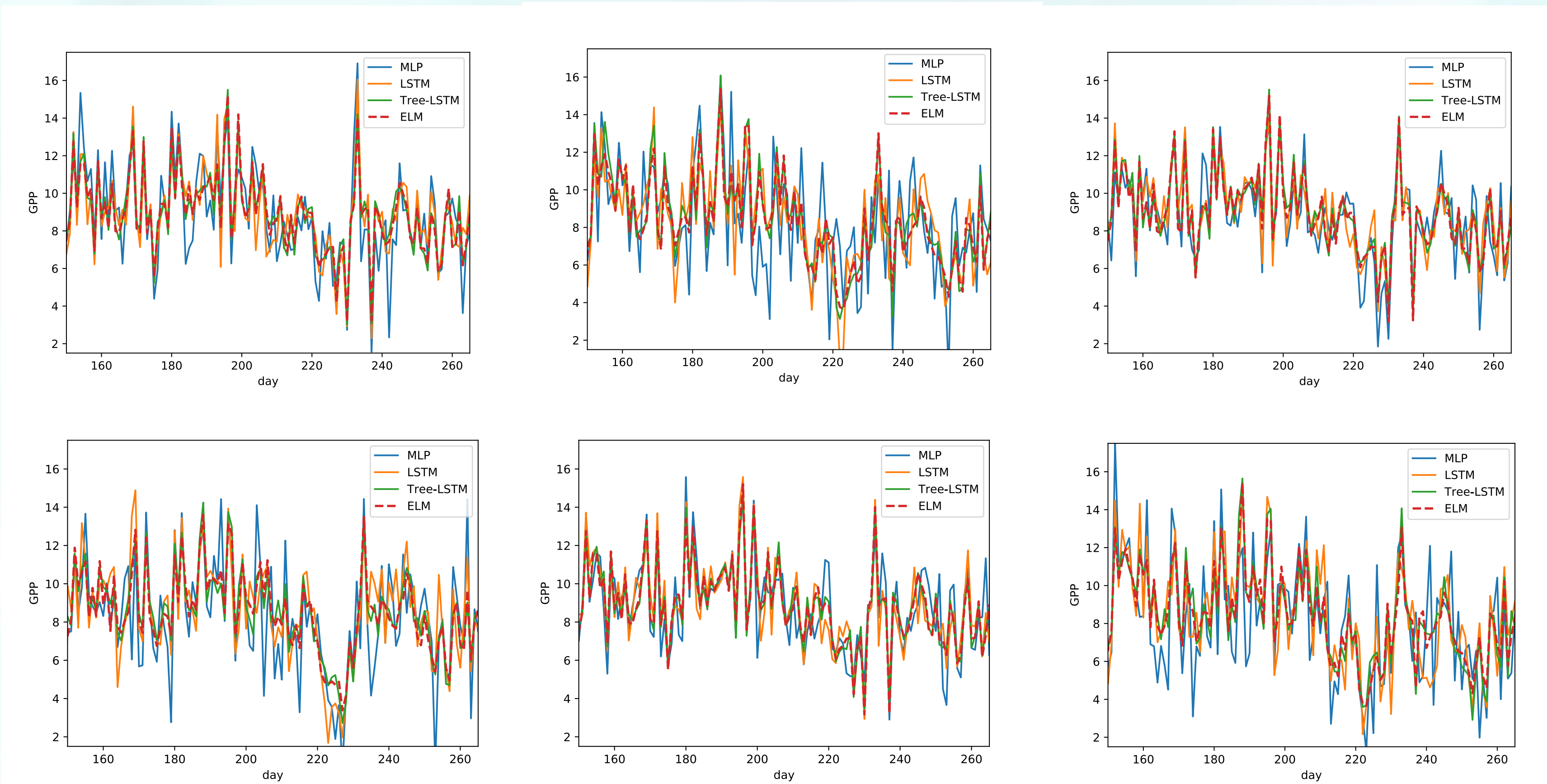


Physics-informed LSTM

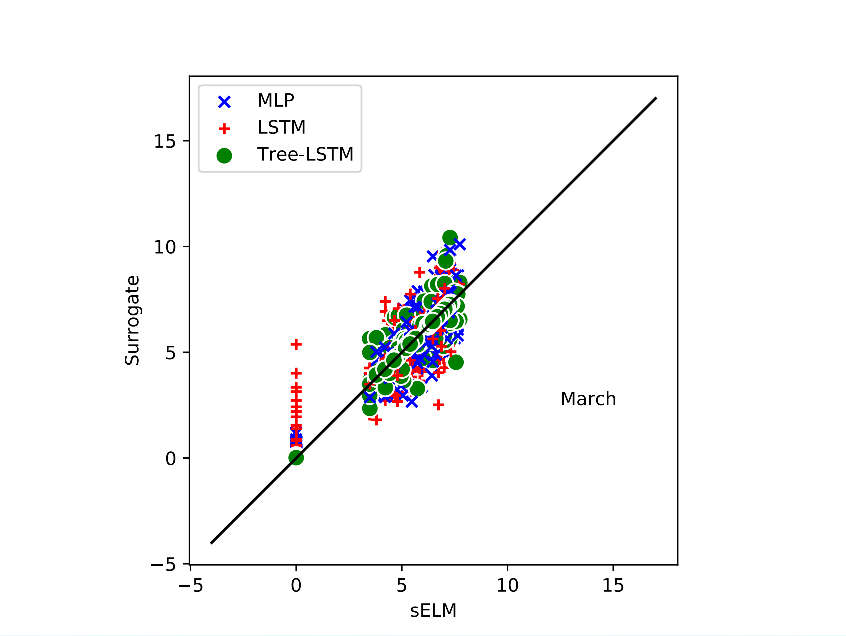


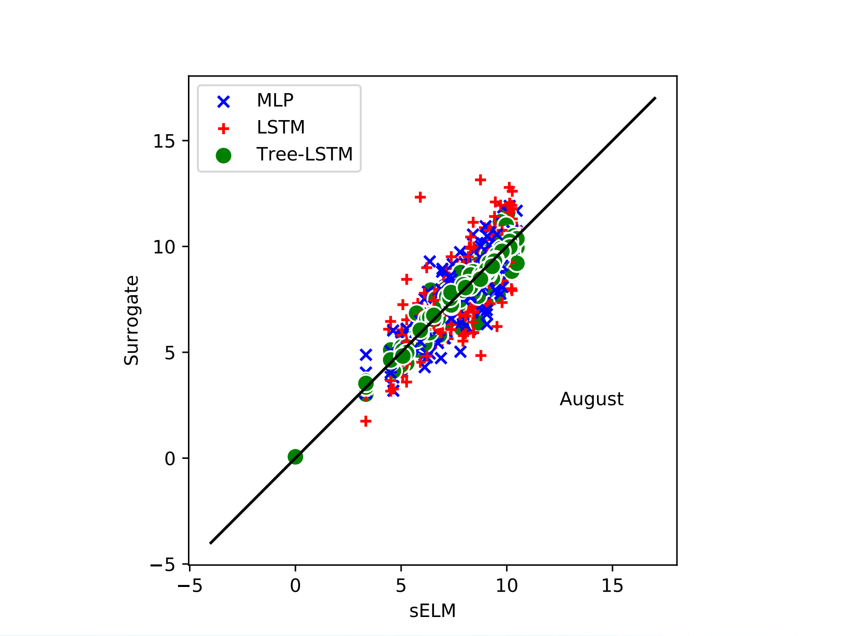


Physics-informed RNN architecture captures daily dynamics well with a fraction of the cost



Physics-informed RNN architecture captures daily dynamics well with a fraction of the cost





Physics-informed RNN architecture captures daily dynamics well with a fraction of the cost

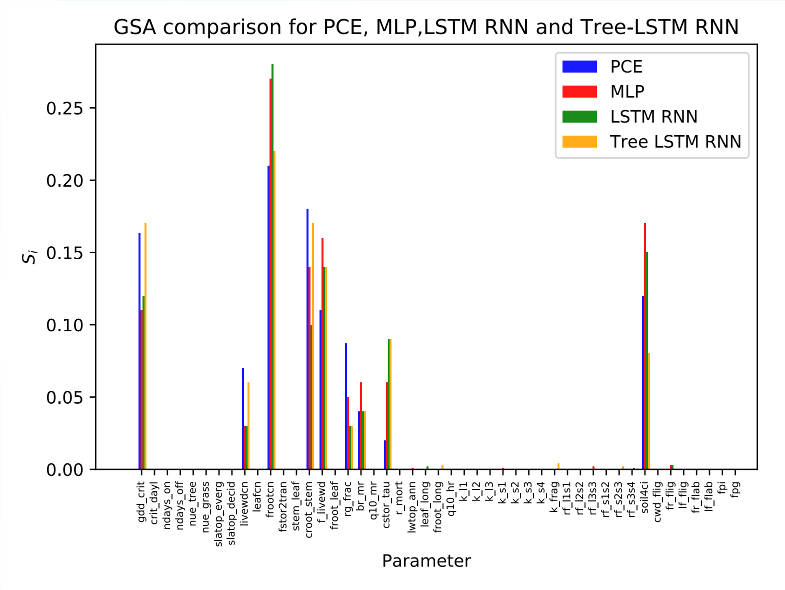
Price to pay?

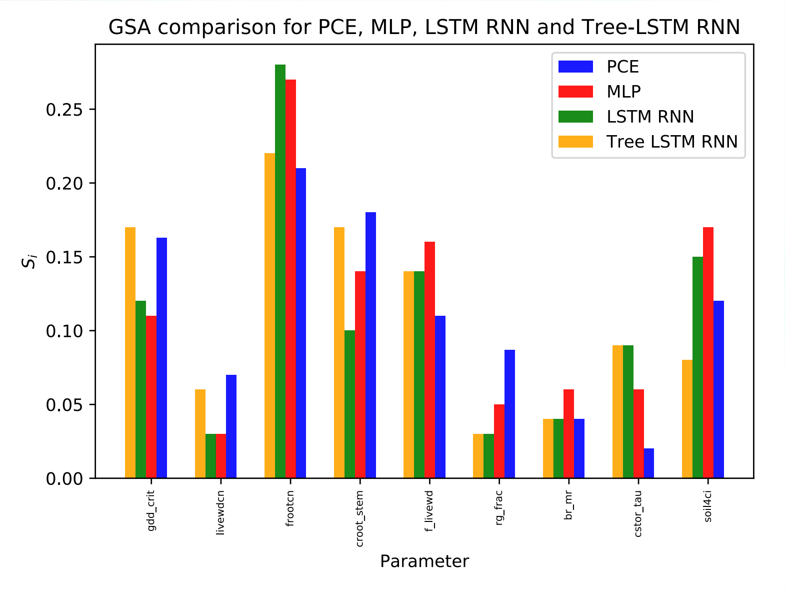
Compared to PC…

a) GSA is not ‘free’, and requires extensive sampling of the RNN surrogates.

* Not a big deal if the limiting factor is the ELM expense

b) Does not come with uncertainties

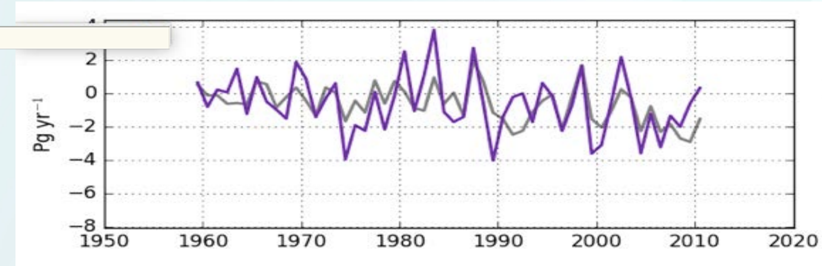




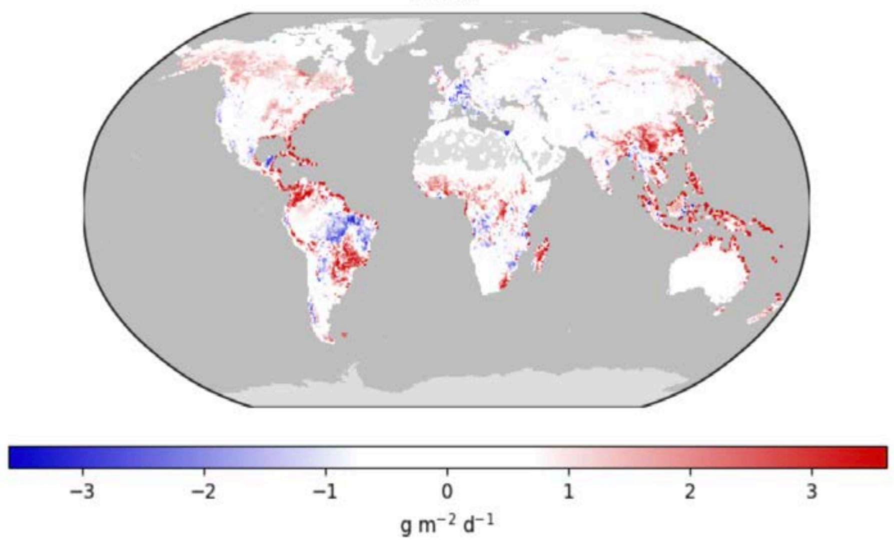
Offline land model benchmarking and validation

- Large biases remain in many regions – can we solve using calibration?
- ELM regional calibration is too expensive
- Site-level calibration results not globally relevant
- Regional sELM simulations for UQ methods development: 2000 member ensemble, 47 parameter

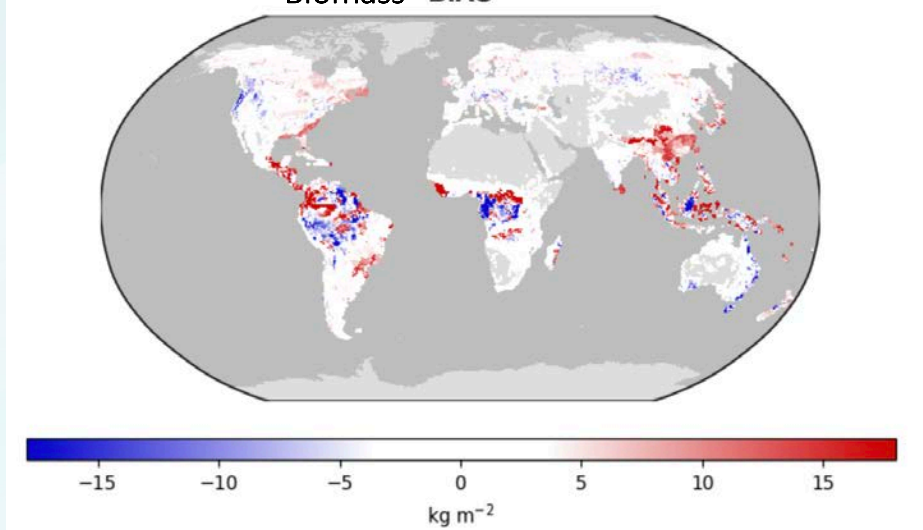
Global net land carbon flux



ELM v1 GPP **BIAS**



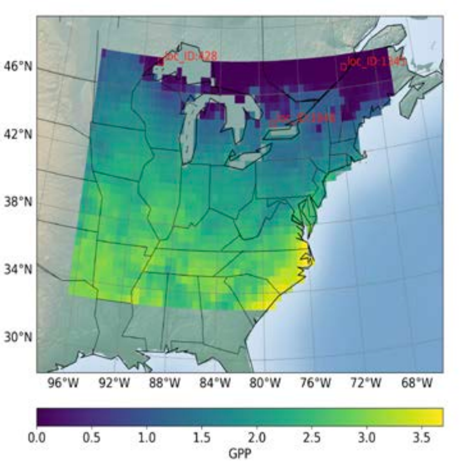
Biomass **BIAS**

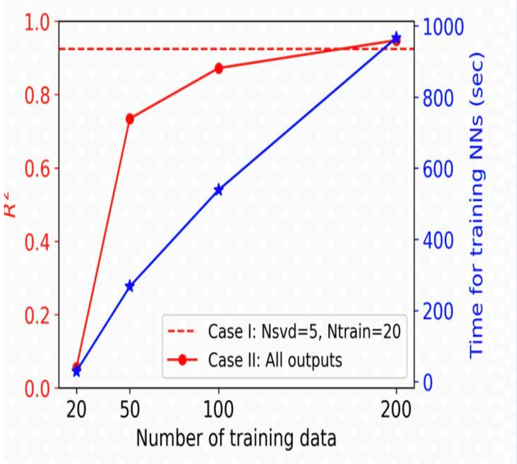


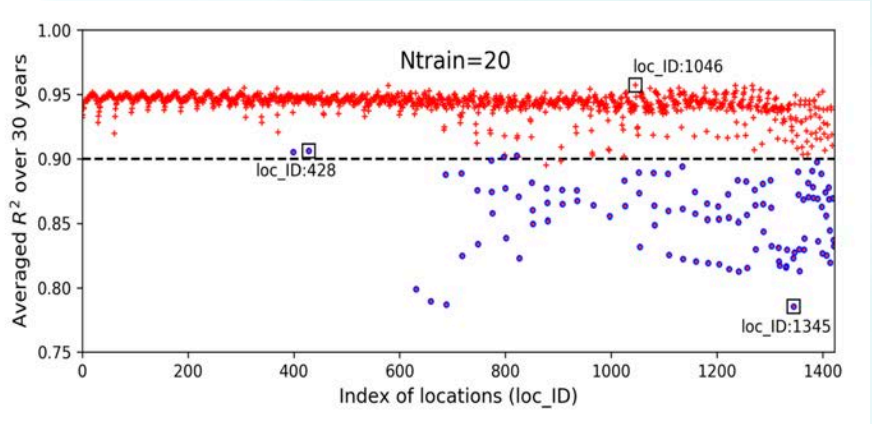
Energy Exascale Earth System Model

U.S. DEPARTMENT OF ENERGY

Surrogates with spatially and temporally varying outputs



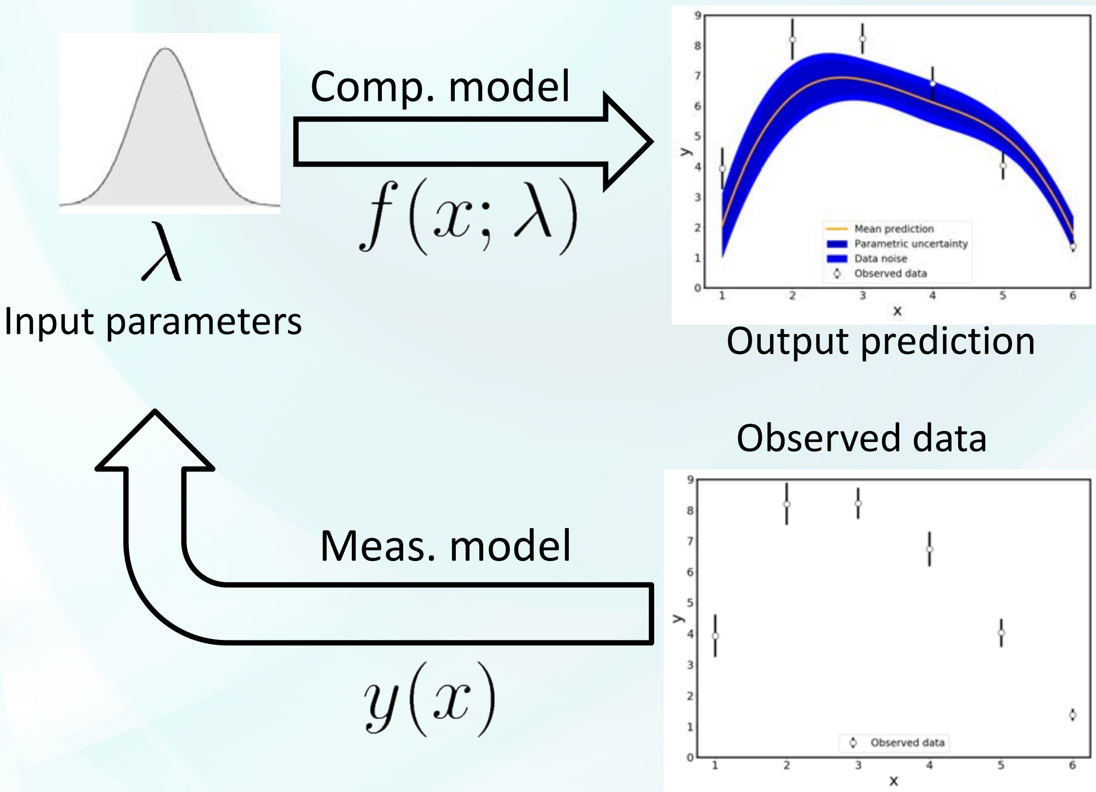




- In this example, we have 42660 GPP outputs (over space and time) – building surrogates for all of these is cumbersome!
- 8 model parameters varied
- Outputs highly correlated in space and time
- Singular value decomposition can be applied to reduce the dimensionality of our output
- NN with 5 singular values trained in 4 seconds – fewer samples and far less time than standard approach

- Not all NNs are created equally. Key to NN selection is selection of hyperparameters.
- Only 20 training samples (model simulations) are necessary for good surrogate accuracy at most locations.

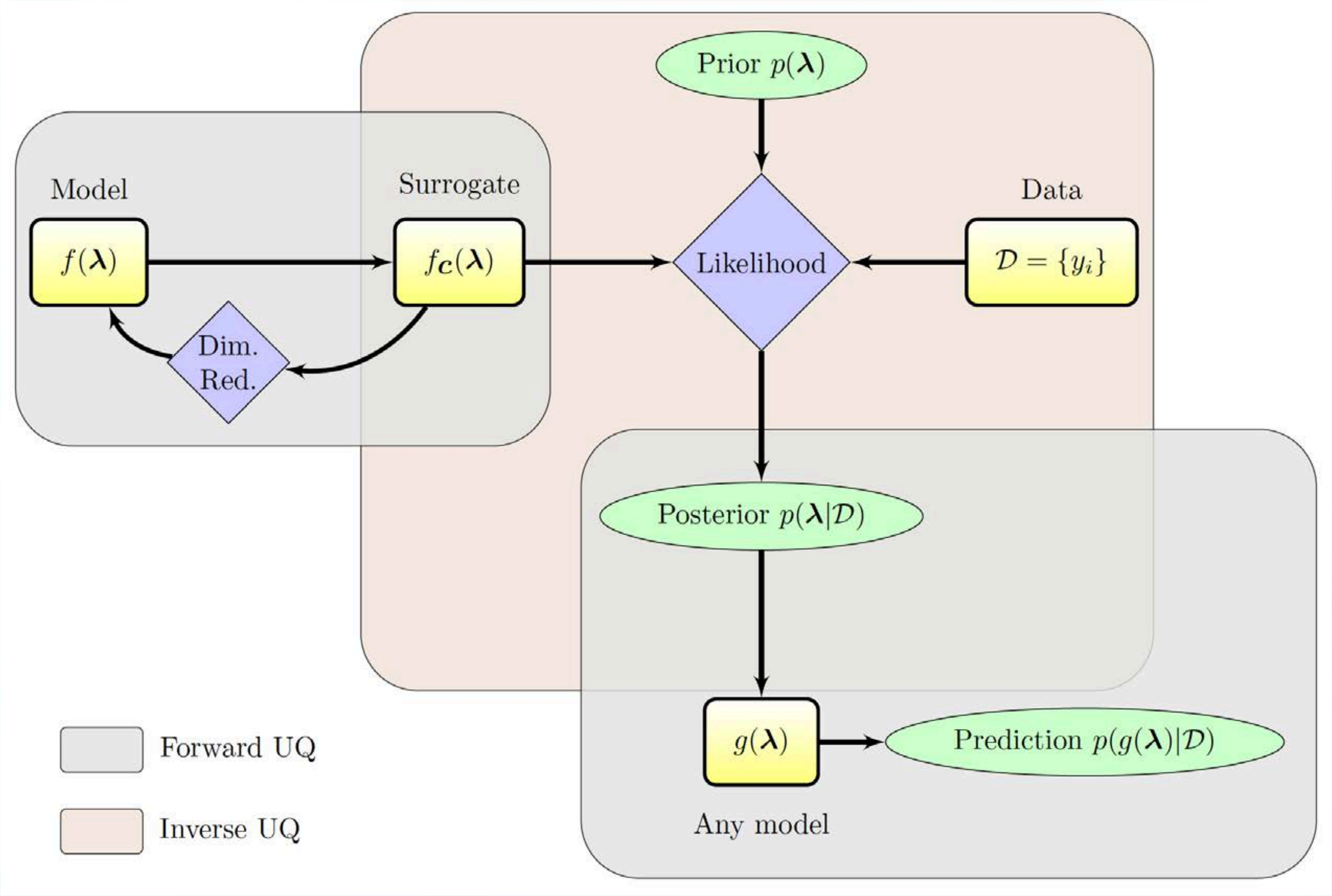
Inverse modeling: tuning model parameters with observational data



- *Forward predictions:*
  - surrogate models
  - sensitivity analysis,
  - parametric uncertainty

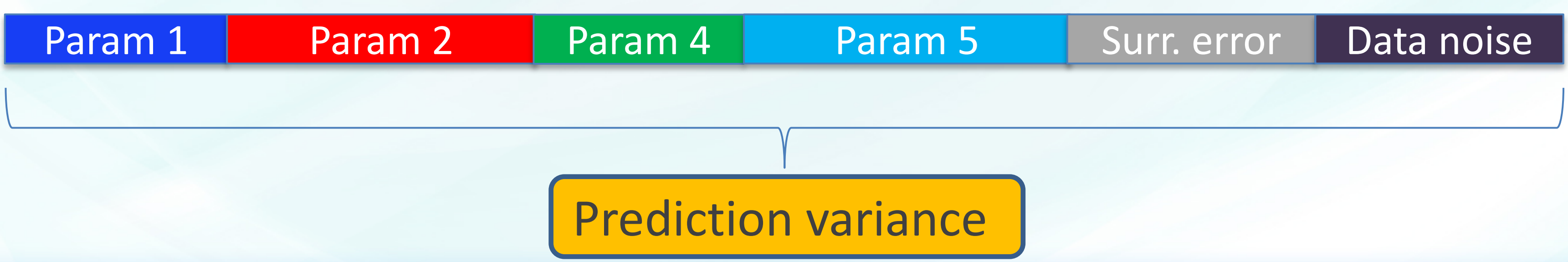
- *Inverse modeling:*
  - parameter tuning
  - calibration
  - data noise incorp.

Generic Forward+Inverse UQ Workflow



Bayesian approach is main tool for parameter calibration

- Bayesian inference allows incorporation of various sources of uncertainty
- Markov chain Monte Carlo (MCMC) for building posterior PDFs
  - Ugly high-dimensional parameter PDFs, but advanced MCMC methods are available
- Requires many online evaluations of the model
  - This is why surrogates are handy!
- Predictive uncertainty decomposition augmented with surrogate error and observational noise



Bayesian approach is main tool for parameter calibration

Elephant in the room:  **model *structural* error**

Uncertainty decomposition of model prediction
needs to account for model error –
often the dominant component of the uncertainty!

Model error

Param 1	Param 2	Param 4	Param 5	Surr. error	Data noise

Prediction variance

Model error is crucial, and often the dominant, component of predictive uncertainty

Ignoring model error leads to
- Biased parameter estimation
- Overconfident predictions

Data — Model — Data noise

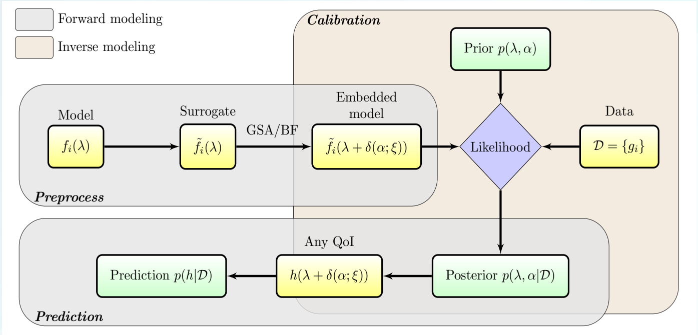
$$y(x_i) = f(x_i; \lambda) + \epsilon_i$$

Prediction variance = Param unc. | Surr. error | Data noise | Model error

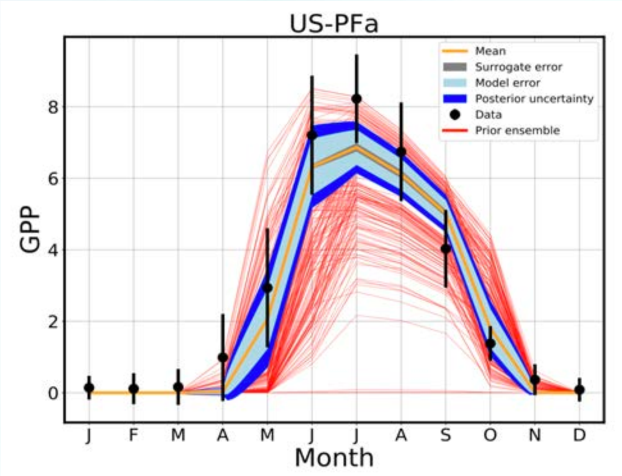
Representing and estimating model error is useful for
- Reliable computational predictions
- Model comparison, selection
- Scientific discovery and model improvement:
  - *“is it worth resolving details, or just parameterize empirically?”*
- Optimal resource allocation:
  - *“do I improve my model (e.g. high-res), or run more simulations?”*

Calibration with *Embedded* Model Structural Error

- Model structural error embedding approach *[Sargsyan et. al., 2015, 2018]*
  - Embedded, but not intrusive, i.e. black-box
  - Physics-driven model correction
  - Meaningful extrapolation to full set of QoI predictions
  - Disambiguation between model error and data noise
  - Core FASTMath capability

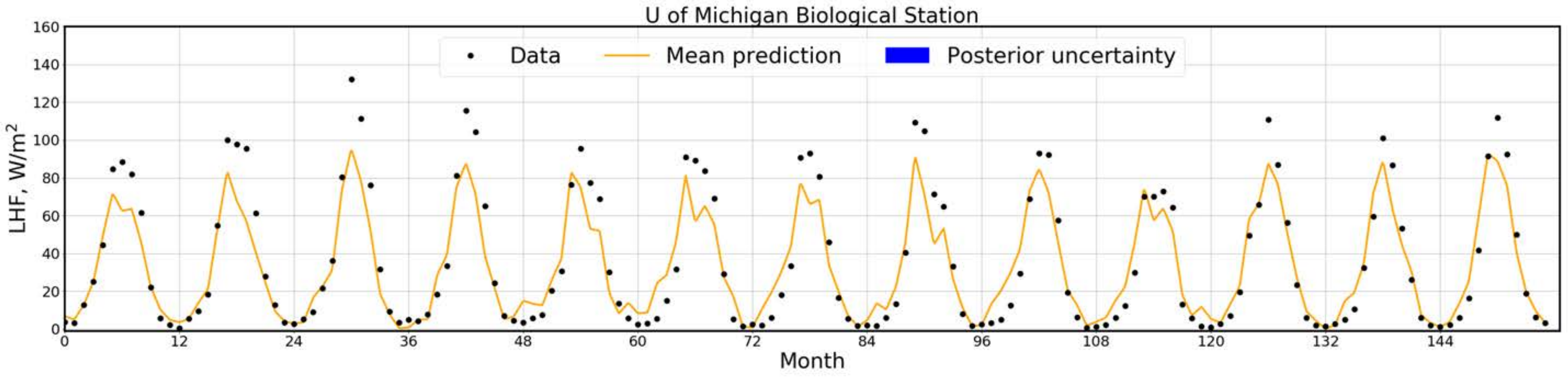


- Calibration of sELM with FLUXNET sites data
- Model error is the dominant uncertainty component
- Removes parameter biases and overfitting
- Points to submodels/parameters that are the culprit



Calibration of ELM given FLUXNET observations

Conventional calibration **without** model error

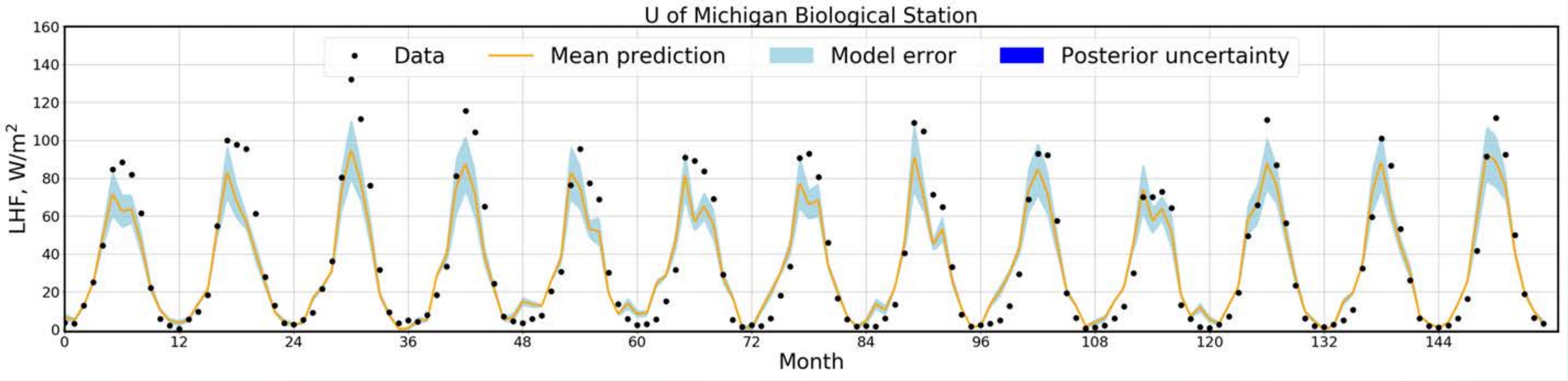


- Summer month peaks are not captured
- Posterior uncertainty negligible

LHF = Latent Heat Flux

Calibration of ELM given FLUXNET observations

Calibration **with** embedded model error

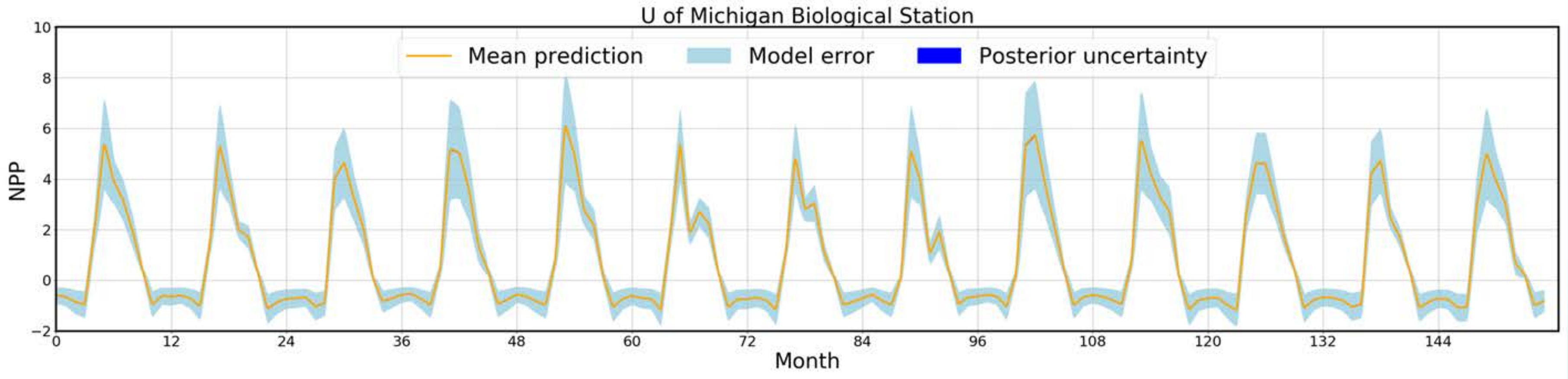


- Model error component dominates
- Captures model deficiency in summer months

LHF = Latent Heat Flux

Calibration of ELM given FLUXNET observations

Calibration **with** embedded model error

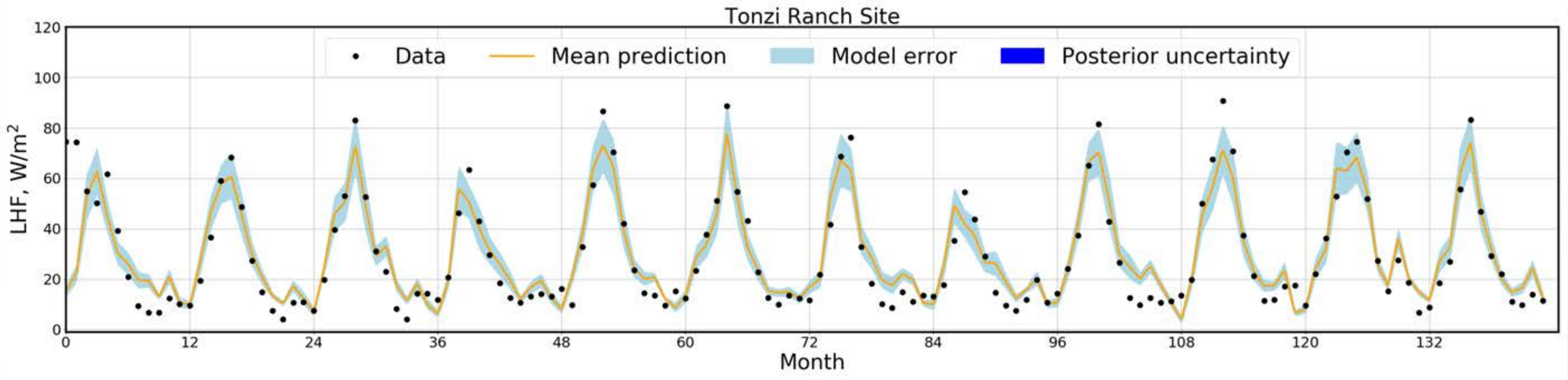


- Allows more accurate prediction of unobservable QoIs
- Can be piped to human component or atmosphere model as a boundary condition

NPP = Net Primary Productivity

Calibration of ELM given FLUXNET observations

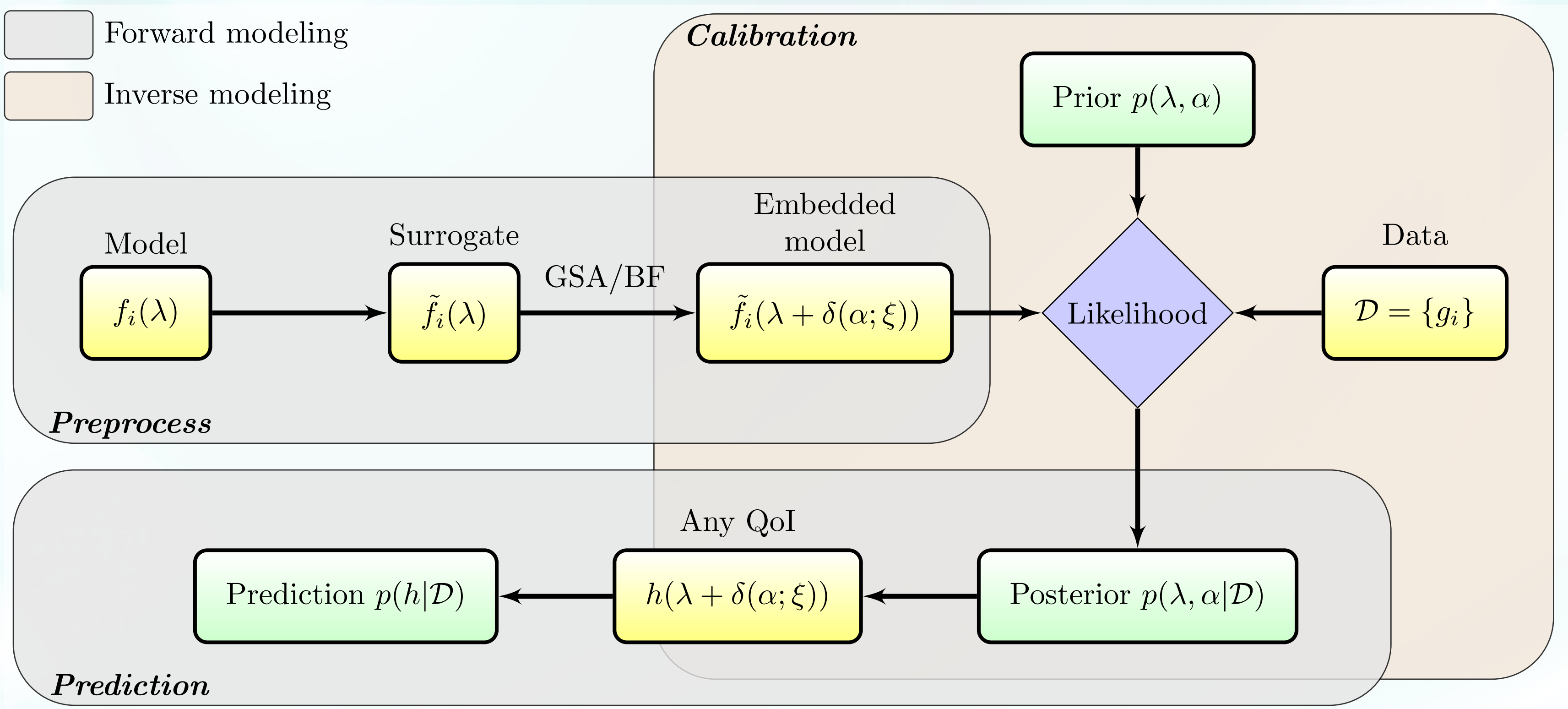
Calibration **with** embedded model error



- Allows prediction at other FLUXNET sites
- Assumption: model goes wrong in a similar way

LHF = Latent Heat Flux

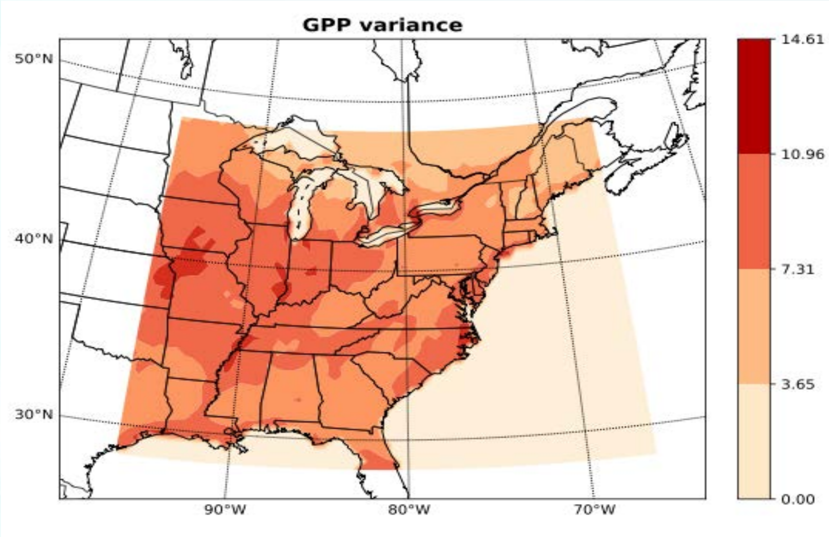
Forward+Inverse UQ Workflow with Embedded Model Structural Error



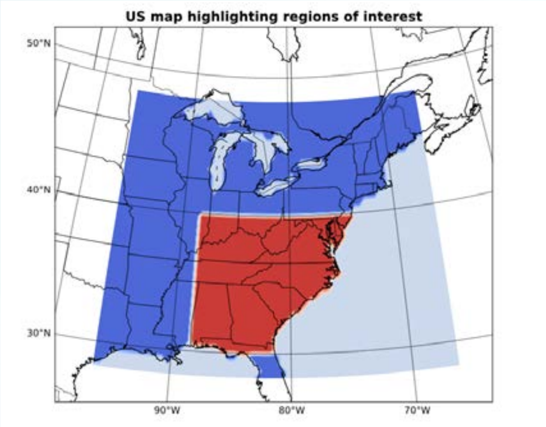
Prediction variance = Param unc. | Surr. error | Data noise | Model error

Using surrogate models to inform observation locations

- Given the uncertainty in our ensemble, where could new observations be placed to optimally reduce posteriorprediction uncertainty?



Variance of the Gross primary produce (GPP) output, determined using 2000 ensemble runs of the sELM model for 30 years.

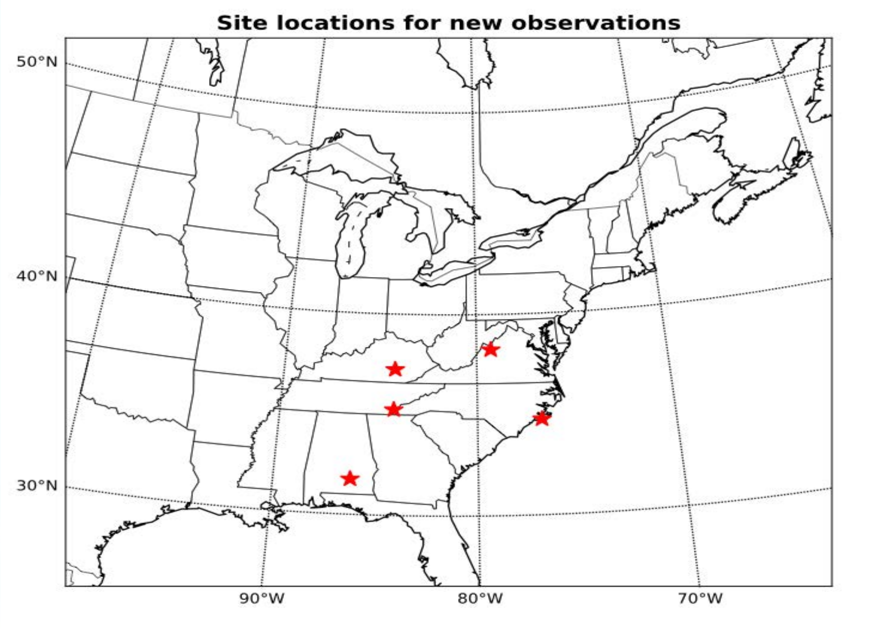


- The computational grid shown is 41x61 and this includes the lightly colored ocean and other water bodies.
- In red we have a hypothetical *region of interest* involving 430 grid points. A possible *QoI* (considered for illustration) is the aggregate GPP in this region
- The remaining land area in dark blue is where sELM output data is available. Grid points at any location are plausible solutions when performing experimental design owing to distant correlations.

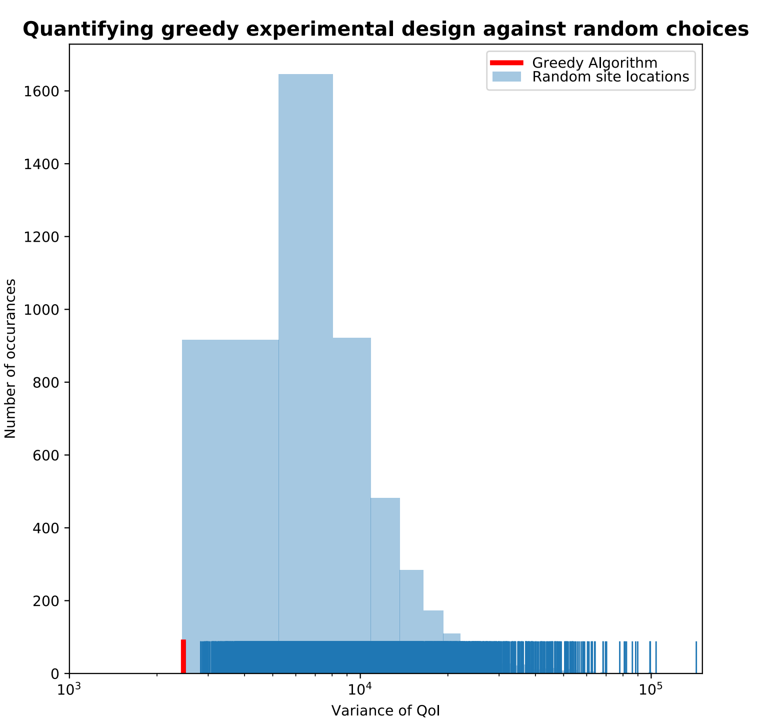
E3SM Energy Exascale Earth System Model

U.S. DEPARTMENT OF ENERGY

Developing experimental design procedures to aid selection of new site locations

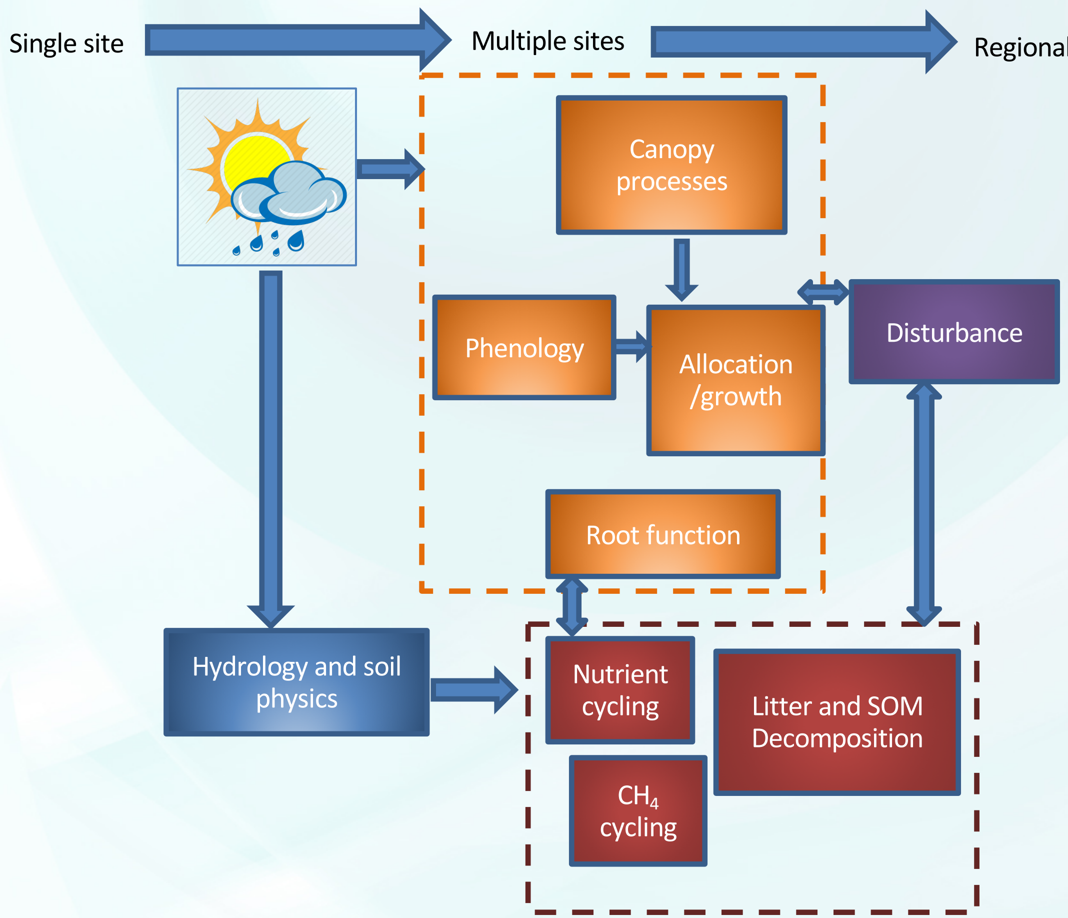


An experimental design study seeking five new sites for observations to reduce the uncertainty of aggregate GPP over the region of interest. Optimal points needn't necessarily be spread out, can can cluster this solution seems to suggest.



The site locations obtained using greedy heuristics are near optimal, and always perform better than random selection of sites. The rug plot indicates the reduced variance of the *QoI* after having factored in the new hypothetical field measurements.

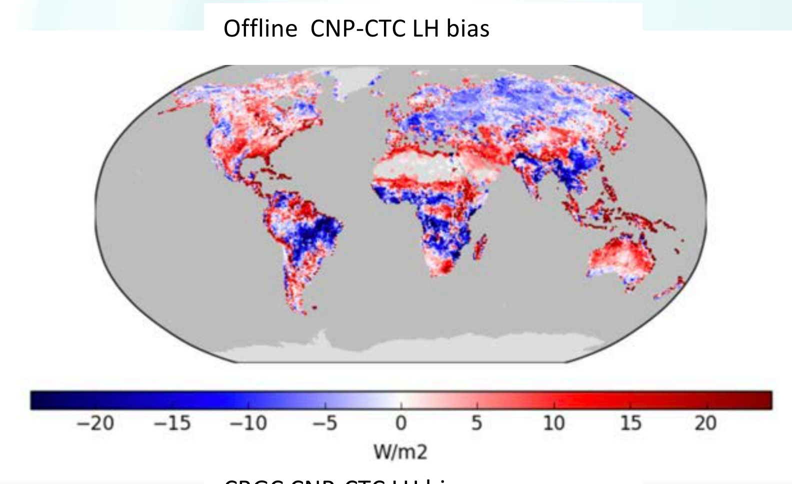
Next steps: ELM UQ testbed

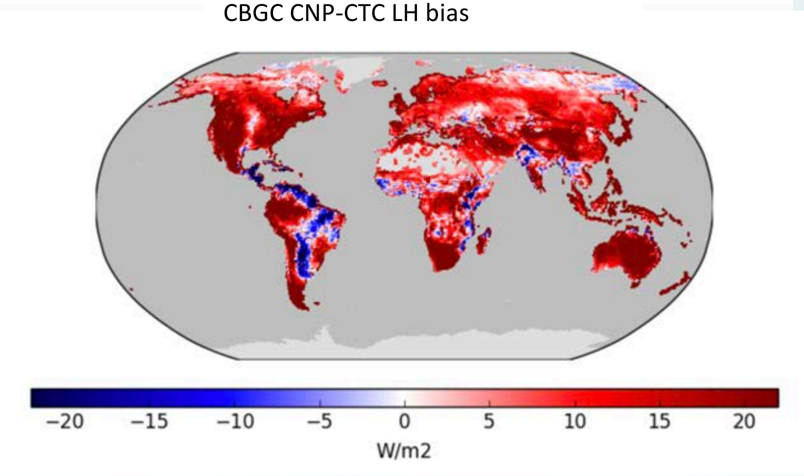


- Develop ELM "functional units" for process submodels
- Develop rapid evaluation capability using surrogates for:
  - Key individual model outputs
  - Each process submodel
- Hierarchical calibration
  - ELM complexity is high
  - Calibrate submodels using process-specific observations
  - Calibrate ELM using integrative observations (e.g. NEE).
- Enable ecological forecasting

Joint work with the ORNL TES SFA

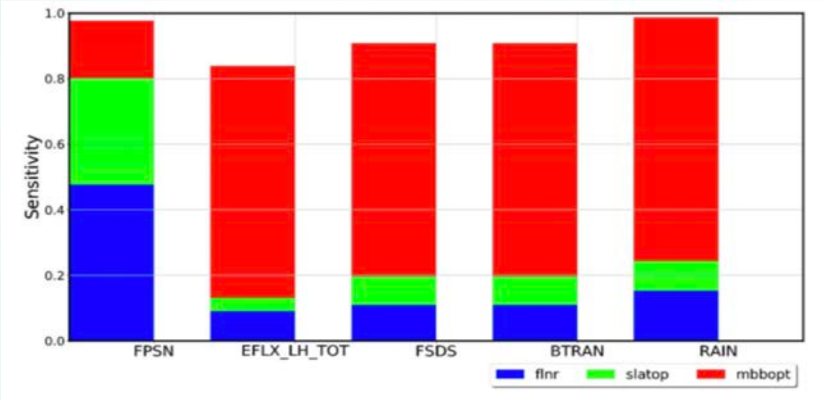
Next steps: Coupled UQ





- Well-tuned/calibrated offline component models may perform poorly in coupled system
- Biases related to other components or coupling between them

- Large computational demand for individual experiments (including spinup)
- Additional outputs/dimensions – even larger ensembles needed.
- Machine learning --> meaningful UQ?
- First step in OSCM : SCM land-atmosphere



Sensitivity analysis of SCM at GOAmazon site

Summary

- **Forward UQ (uncertainty propagation):**
  - Surrogate modeling is the key
    - For point/site simulations, we have well-developed workflows and willing to work with core, NGD and ecosystem projects.
    - Many options: from Polynomial Chaos to Low-Rank Tensors to Physics-Informed Recurrent Neural Networks
    - More advanced, space-and/or-time surrogate modeling tools available
    - With a combination of approaches, we can achieve high surrogate accuracy with small number of simulations.
  - Global sensitivity analysis or variance decomposition for parameter dimension reduction
- **Inverse UQ (parameter calibration):**
  - Bayesian model calibration with Markov chain Monte Carlo (MCMC) sampling
  - Expense is alleviated by using an accurate surrogate: makes ELM calibration feasible
  - Key advance: incorporating model structural error
- **"MODEX" loop enabler**: use attributable model prediction uncertainties to optimally locate new observation locations.