Surrogate-enabled Sensitivity Analysis and Parameter Inference of High-Dimensional Models

Khachik Sargsyan, Cosmin Safta

Sandia National Laboratories

Livermore, CA

ICIAM
Valencia, Spain
July 15-19, 2019

Thanks to: Habib Najm, Tiernan Casey, James Oreluk, Bert Debusschere (SNL), Daniel Ricciuto (ORNL), Jason Bender (LLNL), Youssef Marzouk, Chi Feng (MIT), Roger Ghanem (USC), Xun Huan (UMichigan)

Outline

- Forward Modeling
  - Surrogate construction via Polynomial chaos (PC)
  - Global sensitivity analysis (GSA) for parameters
  - Key challenge: high-dimensionality
    - **Bayesian compressive sensing**
- Inverse Modeling
  - Bayesian calibration of the surrogate
  - Key challenge: model structural error
    - **Embedded model error**
- Predictive uncertainty attribution
  - GSA with data noise, parameter uncertainty and model error
- Climate Land Model application

Polynomial chaos (PC) surrogate

- Model of interest $f(\cdot)$
  - Expensive to evaluate, e.g. climate land model
  - High-dimensional, i.e. $\boldsymbol{\lambda} \in \mathbb{R}^d$ with large d ($\sim 50 - 100$)
  - Usually not feasible to look under the hood

$$y = f(\boldsymbol{\lambda})$$

- Need to develop a parametric surrogate to replace the model in ensemble-intensive studies:
  - sensitivity analysis
  - calibration
  - optimal experimental design

$$f_{\boldsymbol{c}}(\boldsymbol{\lambda}) \approx f(\boldsymbol{\lambda})$$

- Polynomial chaos surrogates are convenient
  - moment estimation
  - uncertainty propagation
  - global sensitivity analysis (GSA)

$$f_{\boldsymbol{c}}(\boldsymbol{\lambda}) \approx \sum_{k=1}^{K} c_k \Psi_k(\boldsymbol{\lambda})$$

Global Sensitivity Analysis with PC

$$f(\boldsymbol{\lambda}) \simeq \sum_{k=0}^{K} c_k \Psi_k(\boldsymbol{\lambda})$$

- Main effect sensitivity indices

$$S_i = \frac{Var[\mathbb{E}(f(\boldsymbol{\lambda}|\lambda_i)]}{Var[f(\boldsymbol{\lambda})]} = \frac{\sum_{k\in\mathbb{I}_i} c_k^2 ||\Psi_k||^2}{\sum_{k>0} c_k^2 ||\Psi_k||^2}$$

  - $\mathbb{I}_i$ is the set of bases with only λ_i involved
  - S_i is the uncertainty contribution that is due to i-th parameter only

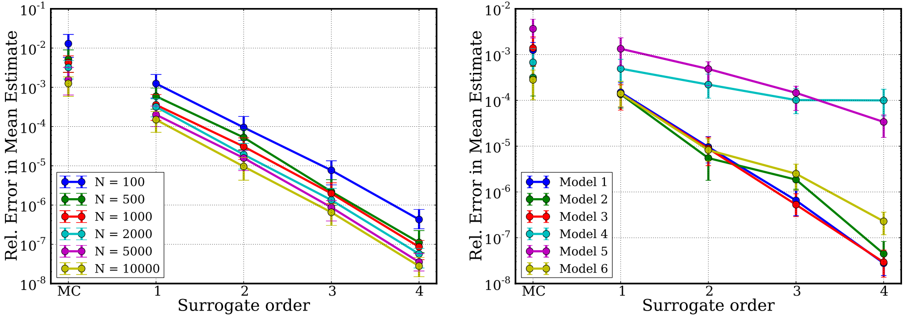
- Total effect sensitivity indices

$$T_i = 1 - \frac{Var[\mathbb{E}(f(\boldsymbol{\lambda}|\lambda_{-i})]}{Var[f(\boldsymbol{\lambda})]} = \frac{\sum_{k\in\mathbb{I}_i^T} c_k^2 ||\Psi_k||^2}{\sum_{k>0} c_k^2 ||\Psi_k||^2}$$

  - $\mathbb{I}_i^T$ is the set of bases with λ_i involved, including all its interactions.
  - T_i is the total uncertainty contribution due to i-th parameter

Moment estimation or GSA:

usually better to prebuild and work with the PC surrogate



... except when high-d, making surrogate construction challenging

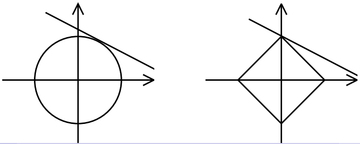
High-d PC surrogate with compressive sensing

$$f(\boldsymbol{\lambda}) \approx f_{\boldsymbol{c}}(\boldsymbol{\lambda}) = \sum_{k=1}^{K} c_k \Psi_k(\boldsymbol{\lambda})$$

- Given an ensemble of model evaluations $\boldsymbol{f} = f(\boldsymbol{\lambda}^{(i)})$, for $i = \overline{1, N}$
- PC coefficients are found by regression $\quad \arg\min_{\boldsymbol{c}} ||\boldsymbol{f} - \boldsymbol{P}\boldsymbol{c}||_2$
- Usually truncating PC bases up to a given total degree....
- ... leads to infeasibly large basis set: $K = (d+p)!/(d!p!) \gg N$
- Compressive sensing, LASSO, basis pursuit: regularized regression

$$\arg\min_{\boldsymbol{c}} \left\{ ||\boldsymbol{f} - \boldsymbol{P}\boldsymbol{c}||_2^2 + \alpha ||\boldsymbol{c}||_1 \right\}$$

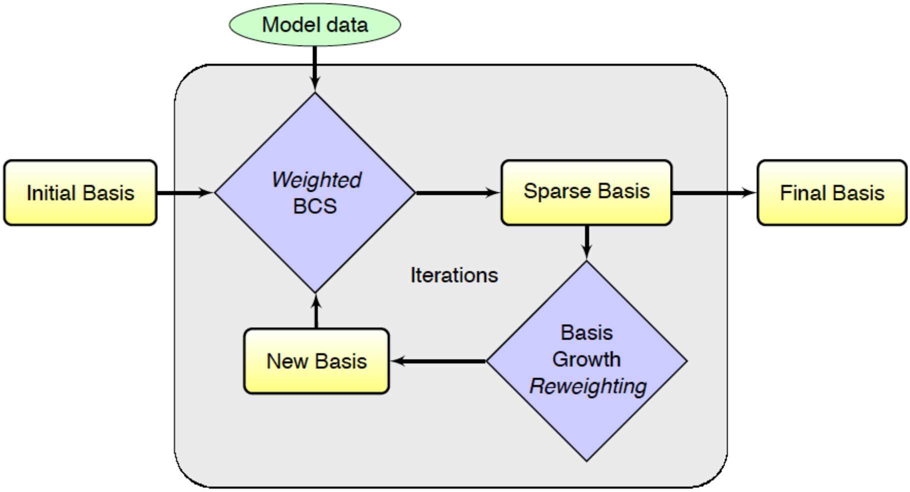
  - ... or $\arg\min_{\boldsymbol{c}} ||\boldsymbol{f} - \boldsymbol{P}\boldsymbol{c}||_2$ s.t. $||\boldsymbol{c}||_1 < \epsilon$
  - ... or $\arg\min_{\boldsymbol{c}} ||\boldsymbol{c}||_1$ s.t. $||\boldsymbol{f} - \boldsymbol{P}\boldsymbol{c}||_2 < \epsilon$



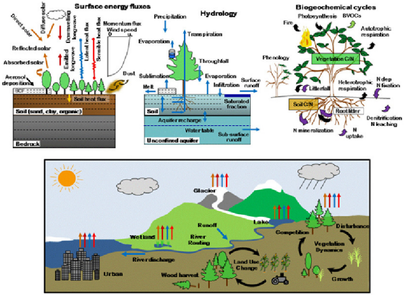
Compressive sensing: enhancements

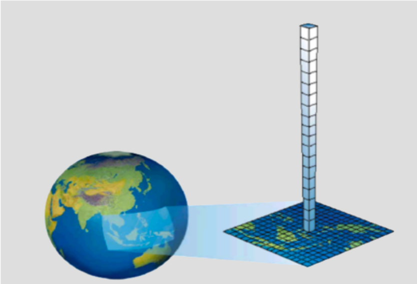
- Bayesian extension: $\arg\min_{\boldsymbol{c}}\{\overbrace{||\boldsymbol{Z}-\boldsymbol{P}\boldsymbol{c}||_2^2}^{\text{Likelihood}}+\overbrace{\alpha||\boldsymbol{c}||_1}^{\text{Prior}}\}$
  - Get coefficients with uncertainties
  - Fights overfitting better
  - Connections with relevance vector machine (RVM)
- Weighted regularization
  - Always better, if you know how to weigh
- Iterative growth of polynomial basis
  - Exploit the structure of polynomial bases for smarter search
  - An iterative procedure that allows increasing the order for the relevant basis terms while maintaining the dimensionality reduction [Sargsyan *et al.* 2014], [Jakeman *et al.* 2015].
  - Iterations inform the weighting procedure

BCS removes unnecessary basis terms



Application of Interest: E3SM Land Model

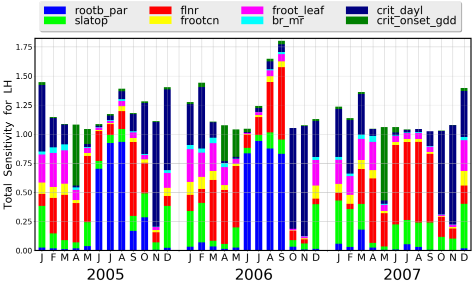




- US Department of Energy (DOE) sponsored Earth system model
- Land, atmosphere, ocean, ice, human system components
- High-resolution, employ DOE leadership-class computing facilities
- Large number of uncertain input parameters

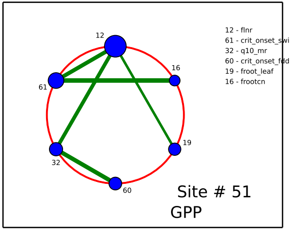
Sparse PC surrogate and uncertainty decomposition for the E3SM Land Model

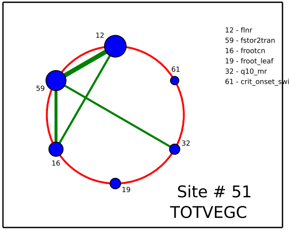
- Main effect sensitivities : rank input parameters
- Joint sensitivities : most influential input couplings
- About 200 polynomial basis terms in the 50-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data



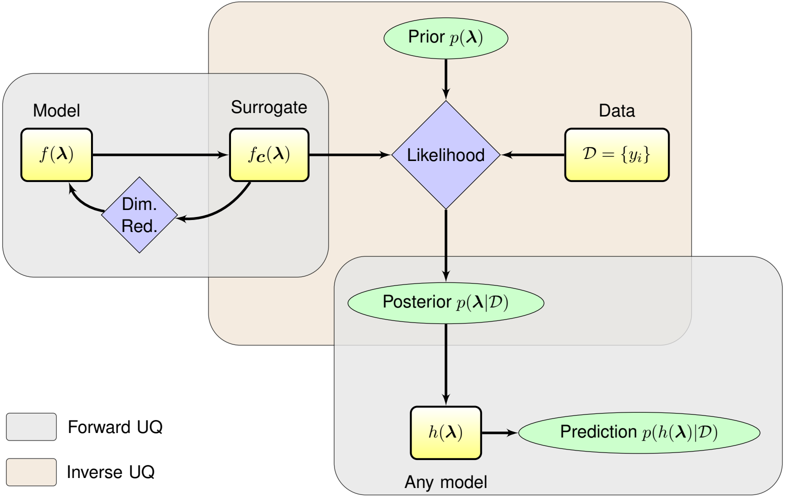
Sparse PC surrogate and uncertainty decomposition for the E3SM Land Model

- Main effect sensitivities : rank input parameters
- Joint sensitivities : most influential input couplings
- About 200 polynomial basis terms in the 50-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data





Surrogate-enabled calibration and prediction workflow



Elephant in the room: model error $g(x) \approx f(x;\lambda)$

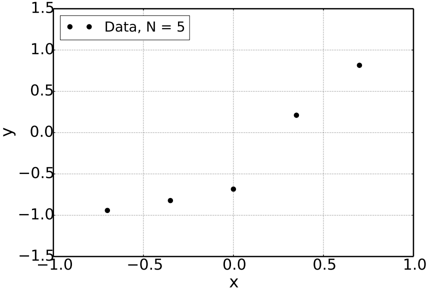
deviation from 'truth' or from a higher-fidelity model

- ... otherwise called (with slightly altered meanings): model discrepancy, model structural error, model inadequacy, model misspecification, model form error, model uncertainty

- Inverse modeling context
  - Given experimental or higher-fidelity model data, estimate the model error
- Represent and estimate the error associated with
  - Simplifying assumptions, parameterizations
  - Mathematical formulation, theoretical framework

- ...will be useful for
  - Model validation, comparison
  - Scientific discovery and model improvement
  - Reliable computational predictions

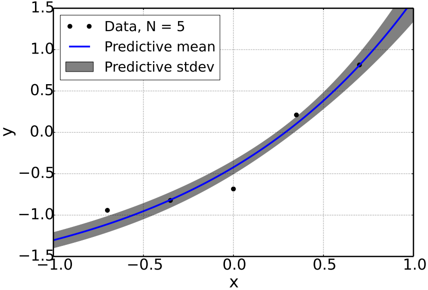
Ignoring model error leads to overconfident and biased predictions



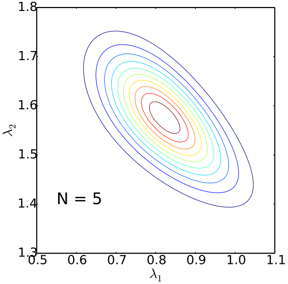
Model-data fit

- Given noisy data, calibrate an exponential model: $g(x) \approx f(x; \lambda)$

Ignoring model error leads to overconfident and biased predictions



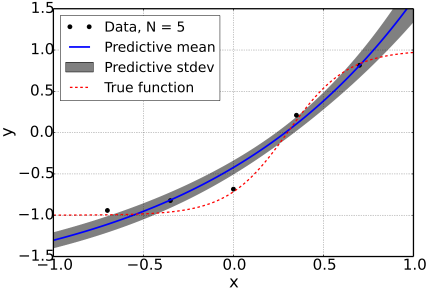
Model-data fit



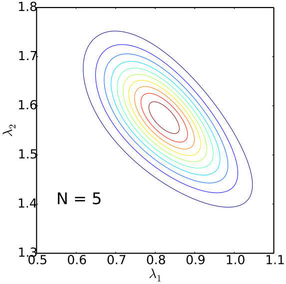
Posterior on parameters

- Given noisy data, calibrate an exponential model: $g(x) \approx f(x; \lambda)$
- Employ Bayesian inference to obtain posterior PDFs on λ

Ignoring model error leads to overconfident and biased predictions



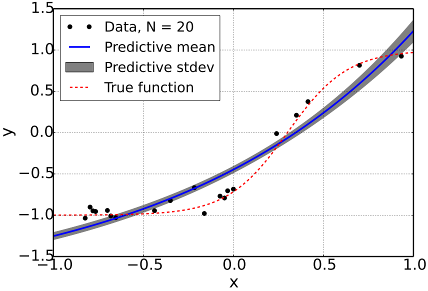
Model-data fit



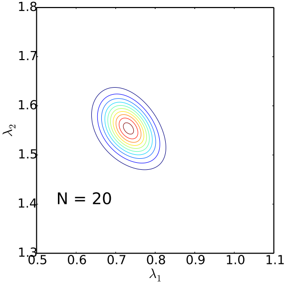
Posterior on parameters

- Given noisy data, calibrate an exponential model: $g(x) \approx f(x; \lambda)$
- Employ Bayesian inference to obtain posterior PDFs on λ
- True model – dashed-red – is *structurally* different from fit model $f(x, \lambda)$

Ignoring model error leads to overconfident and biased predictions



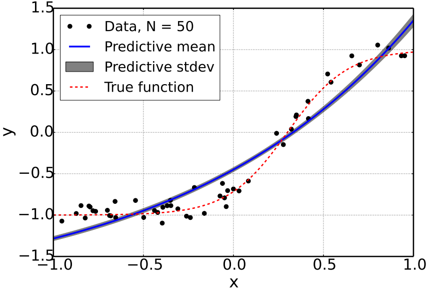
Model-data fit



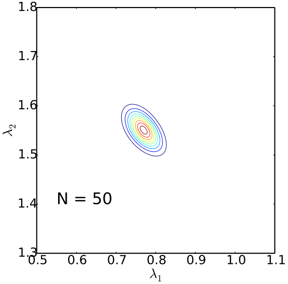
Posterior on parameters

- Given noisy data, calibrate an exponential model: $g(x) \approx f(x; \lambda)$
- Employ Bayesian inference to obtain posterior PDFs on λ
- True model – dashed-red – is *structurally* different from fit model $f(x, \lambda)$
- Higher data amount reduces posterior and predictive uncertainty
  - Increasingly sure about predictions based on the *wrong* model

Ignoring model error leads to overconfident and biased predictions



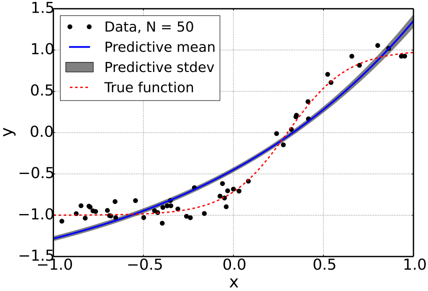
Model-data fit



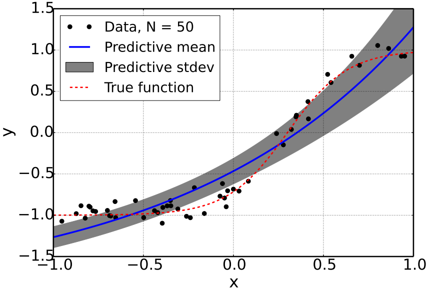
Posterior on parameters

- Given noisy data, calibrate an exponential model: $g(x) \approx f(x;\lambda)$
- Employ Bayesian inference to obtain posterior PDFs on λ
- True model – dashed-red – is *structurally* different from fit model $f(x,\lambda)$
- Higher data amount reduces posterior and predictive uncertainty
  - Increasingly sure about predictions based on the *wrong* model

Ignoring model error leads to overconfident and biased predictions



No model error treatment



Model error accounted for

- Given noisy data, calibrate an exponential model: $g(x) \approx f(x;\lambda)$
- Employ Bayesian inference to obtain posterior PDFs on λ
- True model – dashed-red – is *structurally* different from fit model $f(x,\lambda)$
- Accounting for model error allows extra uncertainty component to propagate through predictions

Calibrate $f(x;\lambda)$, given data $g(x)$

x are operating conditions, design parameters, various QoIs
λ are model parameters to be inferred/calibrated

- Default: Ignore model errors: $g(x) = f(x;\lambda) + \epsilon$
  - Biased or overconfident physical parameters
  - Wrong model predictions

- Conventional: Correct for model errors: $g(x) = f(x;\lambda) + \delta(x) + \epsilon$
  - Physical parameters are ok
  - Wrong model predictions (data-specific corrections)
  - Model and data errors mixed up

- What we do: Correct *inside* the model: $g(x) = f(x;\lambda + \delta(x)) + \epsilon$
  - Embedded model error
  - Preserves model structure and physical constraints
  - Disambiguates model and data errors
  - Allows meaningful extrapolation

Bayesian Framework for Model Error Estimation

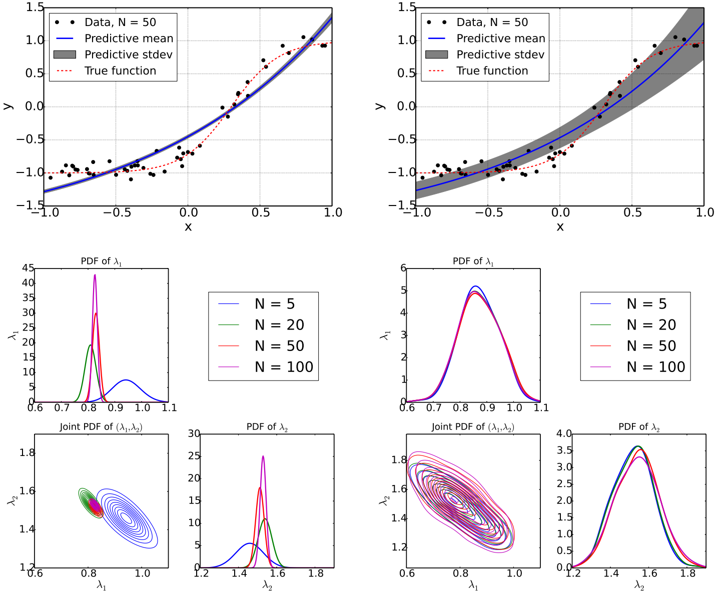
$$y_i = f(x_i; \lambda + \delta_\alpha) + \epsilon_i$$

- Given data y_i, perform *simultaneous* estimation of $\tilde{\alpha} = (\lambda, \alpha)$, i.e. model parameters λ and model-error parameters α.
- Bayes' theorem

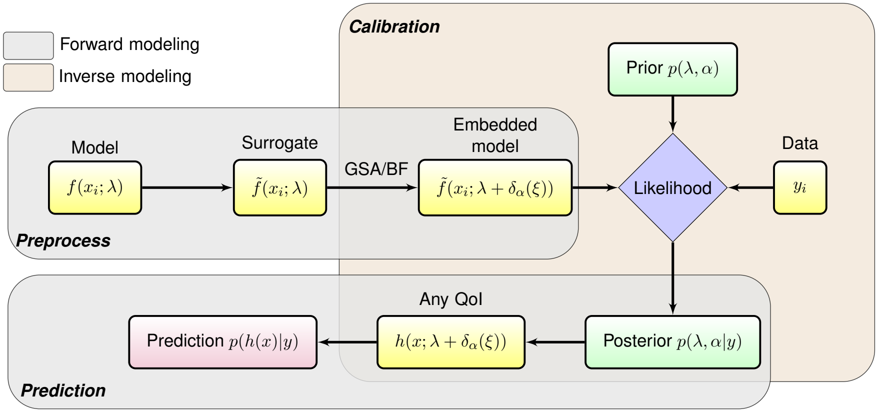
$$\overbrace{p(\tilde{\alpha}|y)}^{\text{Posterior}} = \frac{\overbrace{p(y|\tilde{\alpha})}^{\text{Likelihood}} \; \overbrace{p(\tilde{\alpha})}^{\text{Prior}}}{\underbrace{p(y)}_{\text{Evidence}}}$$

- In order to estimate the likelihood $L_y(\tilde{\alpha}) = p(y|\tilde{\alpha}) = p(y|\lambda, \alpha)$, one needs uncertainty propagation through $f(x_i; \underbrace{\lambda + \delta_\alpha}_{\text{stochastic}})$,
- ... hence, we employ Polynomial Chaos (PC) representation for δ_α.

.. back to toy example



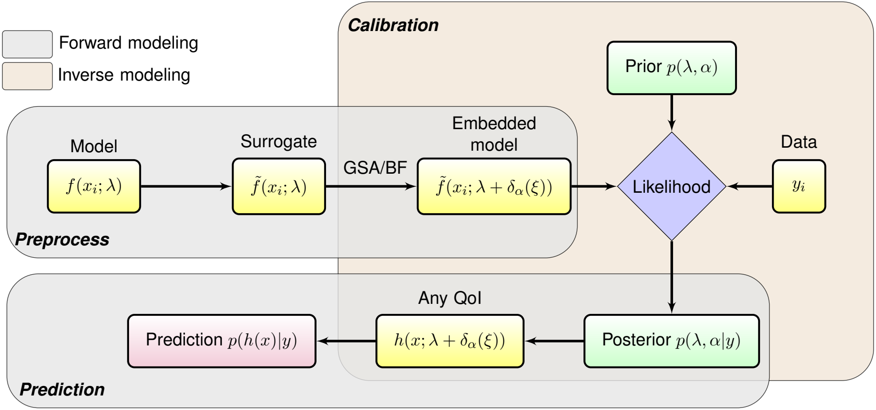
Model error embedding – workflow



Predictive uncertainty decomposition: Total Variance =

Parametric uncertainty + Data noise + Model error + Surrogate error

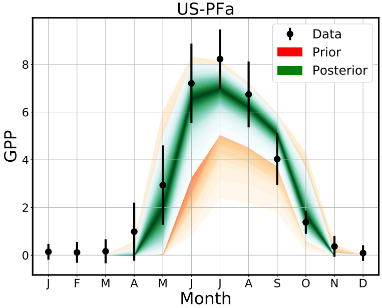
Model error embedding – workflow

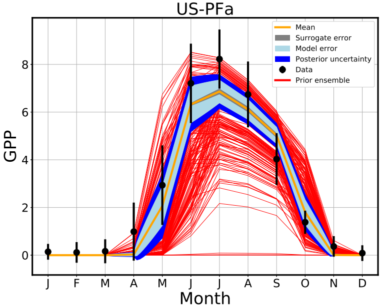


$$\sigma_i^2 = \underbrace{\mathbb{E}_{\tilde{\alpha}}\left[\sigma_i^2(\tilde{\alpha})\right]}_{\text{Model error}} + \underbrace{\mathbb{V}_{\tilde{\alpha}}\left[\mu_i(\tilde{\alpha})\right]}_{\text{Posterior uncertainty}} + \underbrace{(\sigma_i^{LOO})^2}_{\text{Surrogate error}} + \underbrace{s_i^2}_{\text{Data noise}}$$

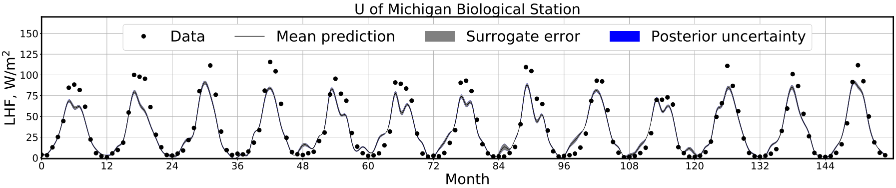
E3SM Land Model (ELM)

- Demo for Park Falls site (WI)
- Gross Primary Productivity (GPP) observations
- Predictive variance decomposition on the right: essentially a glorified GSA into components due to
  - Surrogate error
  - Posterior uncertainty
  - **Model error**



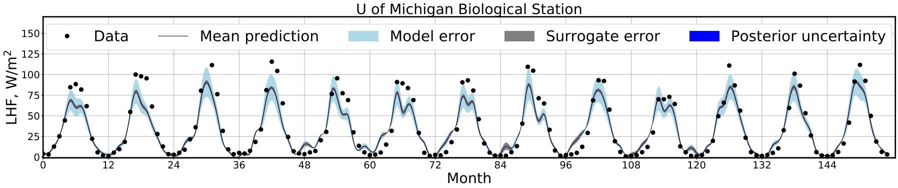


E3SM Land Model (ELM)



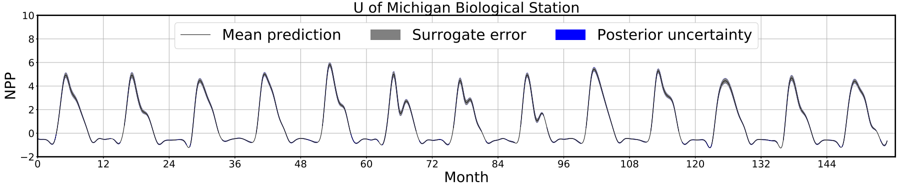
- Conventional calibration without model error

E3SM Land Model (ELM)



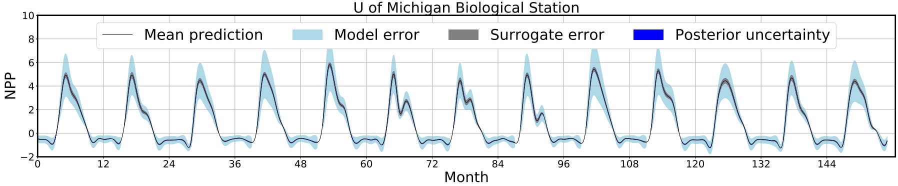
- Predictive variance decomposition with model-error component
- ... with predictive uncertainty that captures model error

E3SM Land Model (ELM)



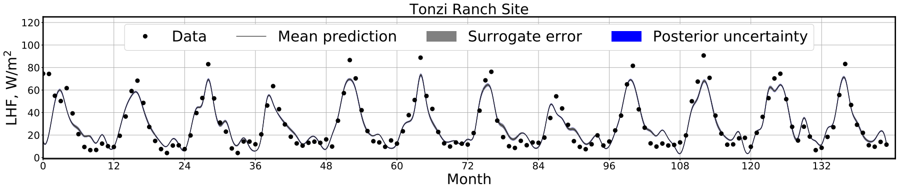
- Predictive variance decomposition with model-error component
- Allows meaningful prediction of other QoIs
  (e.g. no data/observable)

E3SM Land Model (ELM)



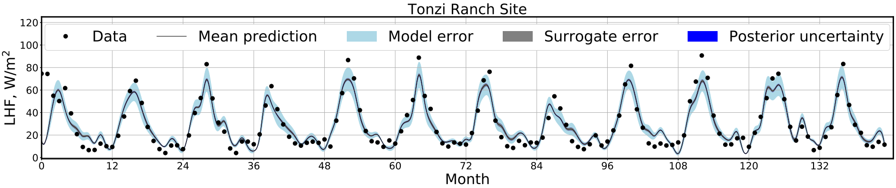
- Predictive variance decomposition with model-error component
- Allows meaningful prediction of other QoIs
  (e.g. no data/observable)
- ... with predictive uncertainty that captures model error

E3SM Land Model (ELM)



- Predictive variance decomposition with model-error component
- Allows prediction at other sites

E3SM Land Model (ELM)



- Predictive variance decomposition with model-error component
- Allows prediction at other sites
- ... with predictive uncertainty that captures model error

Summary

- Forward UQ:
  - Polynomial chaos surrogate construction for complex models
  - **High-D challenge: sparse PC via Bayesian compressive sensing**
- Inverse UQ:
  - Bayesian inference for parameter estimation
  - **Model error challenge: embedded model error representation**
- All developments done within UQTk, lightweight C++/Python library out of SNL-CA (*www.sandia.gov/uqtoolkit*) UQTk

---

We are hiring!

- **Postdoctoral Position** UQ-in-Climate at Sandia National Labs
- Go to Sandia careers' website and look for job ID 668176
- Experience with UQ, climate modeling, coding.
- Salary $90K+/year, in Livermore, CA

Literature

Thank you!

Bayesian compressive sensing

K. Sargsyan, C. Safta, H. Najm, B. Debusschere, D. Ricciuto, P. Thornton, “Dimensionality reduction for complex models via Bayesian compressive sensing”, *Int. J. Uncertainty Quantification*, 4(1), 63-93, (2014).

D. Ricciuto, K. Sargsyan, P. Thornton, “The Impact of Parametric Uncertainties on Biogeochemistry in the E3SM Land Model”, *J of Advances in Modeling Earth Systems*, 10(2), 297-319, (2018).

Embedded model error

K. Sargsyan, H. Najm, R. Ghanem, “On the Statistical Calibration of Physical Models”, *Int. J. Chem. Kinetics*, 47(4), 246-276, (2015).

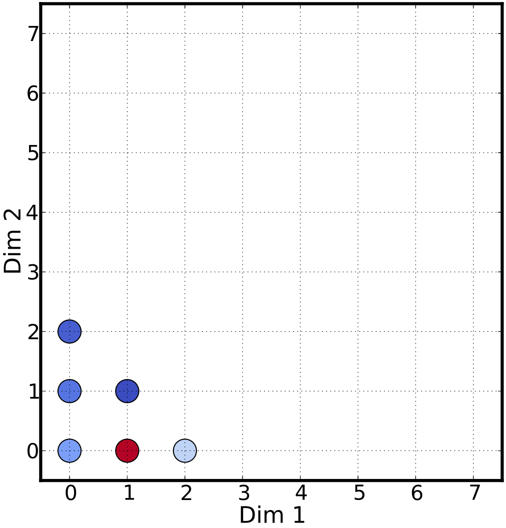
K. Sargsyan, X. Huan, H. Najm. “Embedded Model Error Representation for Bayesian Model Calibration”, *International Journal of Uncertainty Quantification*, in press (2019).

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research and Office of Biological and Environmental Research, Scientific Discovery through Advanced Computing (SciDAC) program.

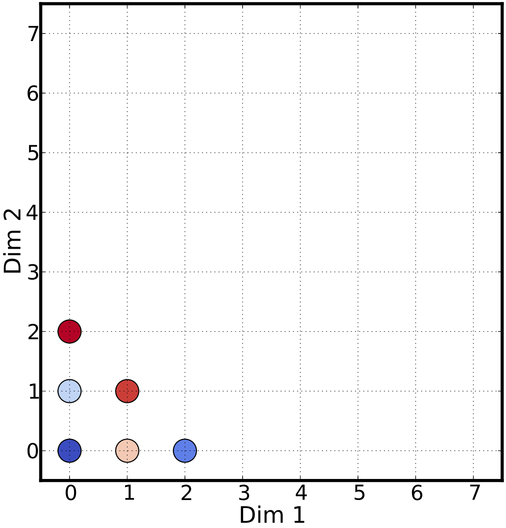
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA-0003525.

Additional Material

Basis set growth: simple anisotropic function



Basis set growth: ... added outlier term



Where to put model error?

$$y_i = f(x_i; \lambda) + \delta(x_i) + \epsilon_i$$

- Outside:
  - Explicit GP representation [Kennedy-O'Hagan, 2001]
  - See also [Higdon et. al, 2004], [Bayarri et. al, 2007]
  - Usage: too many to cite
  - Variants exist: multiplicative noise, non-linear maps etc.

$$y_i = \tilde{f}(x_i; \lambda, \delta(x_i)) + \epsilon_i$$

- Inside:
  - Increased use, especially in physical models: [Emory et. al, 2011] [Oliver and Moser, 2011], [Morrison et. al, 2016], [Sondak et. al, 2017], [Huan et. al, 2017], [Rizzi et. al, 2018]...
  - Engineering/statistical adjustment [Joseph and Melkote, 2009]
  - Additive corrections to submodels [Strong et. al, 2011]
  - Validation of extrapolative predictions [Oliver et. al, 2014]
  - Field inversion and machine learning [Duraisamy et. al, 2015-]
  - Hybrid correction [He and Xiu, 2016]
  - Random field correction [Brown and Atamturktur, 2016]
  - Hierarchical mixture model [Feng, 2017]
  - Parameter inflation [Pernot et. al, 2017]
  - Hierarchical stochastic model [Wu et. al, 2017]
  - Dynamic discrepancy [Bhat et. al., 2017]

More data leads to ‘leftover’ model error

Calibrating a quadratic $f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2$
w.r.t. ‘truth’ $g(x) = 6 + x^2 - 0.5(x+1)^{3.5}$ measured with noise $\sigma = 0.1$.

Summary of features:

- Well-defined model-to-model calibration
- Model-driven discrepancy correlations
- Respects physical constraints
- Disambiguates model and data errors
- Calibrated predictions of multiple QoIs