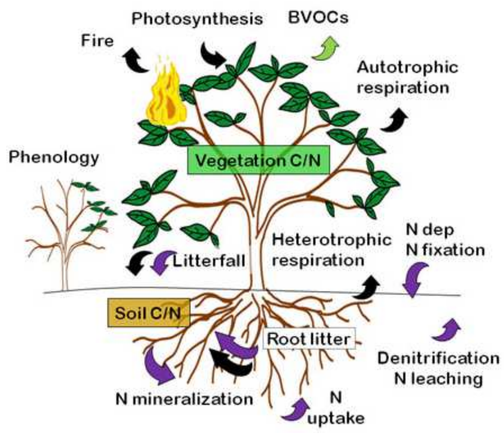
Surrogate Construction for High-Dimensional Climate Models

AGU Fall Meeting 2014
ksargsy@sandia.gov
SAND2014-20505PE

KHACHIK SARGSYAN, COSMIN SAFTA, BERT DEBUSSCHERE, HABIB NAJM — Sandia National Laboratories
DANIEL RICCIUTO, PETER THORNTON — Oak Ridge National Laboratory

Community Land Model (CLM)

- Nested computational grid hierarchy
- Represents spatial heterogeneity of the land surface
- Expensive simulations
- Involves $50-100$ input parameters
- Dependent parameters
- Non-smooth input-output dependence



http://www.cesm.ucar.edu/models/clm/

Surrogate $g_{\mathbf{c}}(\boldsymbol{\lambda}) \approx G(\boldsymbol{\lambda})$ is necessary for expensive models

The surrogate model can be queried instead of the CLM for a) Global sensitivity analysis, b) Optimization, c) Forward uncertainty propagation, d) Calibration.

Polynomial chaos (PC) as a surrogate model

- Interprets input parameters as random variables
- Propagates input uncertainties to outputs of interest

E.g., uniform inputs $\lambda_i \sim \text{Uniform}[a_i, b_i]$,

$$\lambda_i = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2}\eta_i.$$

Output is represented with respect to Legendre polynomials

$$G(\boldsymbol{\lambda}(\boldsymbol{\eta})) \approx g_{\mathbf{c}}(\boldsymbol{\eta}) \equiv \sum_{k=0}^{K} c_k \Psi_k(\boldsymbol{\eta}).$$

Bayesian inference of PC modes leads to a probabilistic surrogate

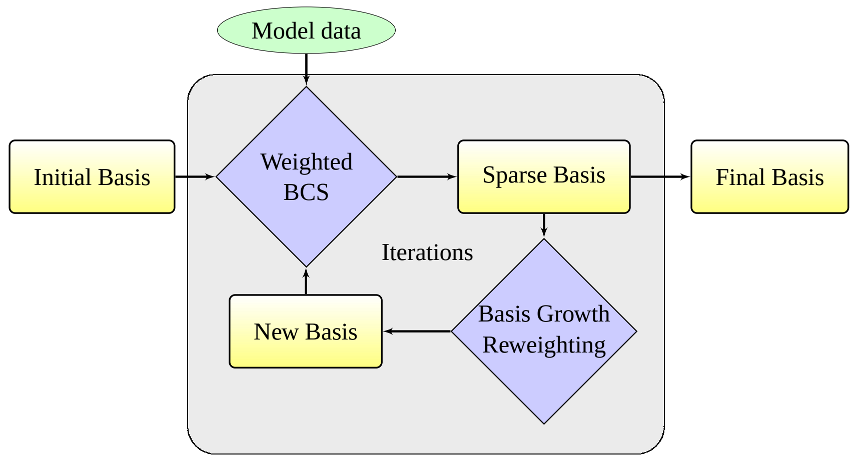
$$p(\mathbf{c}|\alpha, \mathcal{D}) \propto \quad L_{\mathcal{D}}(\mathbf{c}) \quad \times \quad p(\mathbf{c}|\alpha)$$

Data $\mathcal{D}$ is the set of all training runs $\mathcal{D} = (\boldsymbol{\lambda}_i, G(\boldsymbol{\lambda}_i))_{i=1}^N$. The size of $\mathbf{c}$, i.e. the number of polynomial basis terms grows fast; a p-th order, d-dimensional basis has a total of $(p+d)!/(p!d!)$ terms. Sparsity priors strive to detect the smallest set of basis.

Gaussian likelihood $$L_{\mathcal{D}}(\mathbf{c}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(G(\boldsymbol{\lambda}_i) - g_{\mathbf{c}}(\boldsymbol{\eta}_i))^2}{2s^2}\right)$$

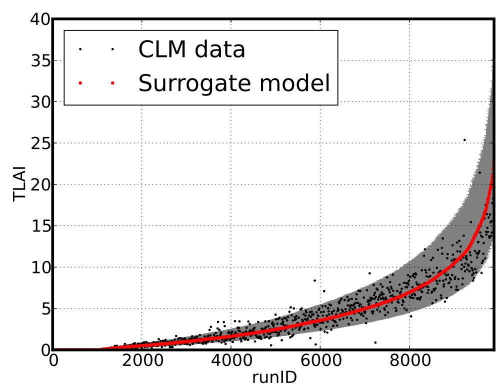
Laplace prior $$p(\mathbf{c}|\alpha) = \int \prod_{k=0}^{K-1} p(c_k|\sigma_k^2) p(\sigma_k^2|\alpha) d\sigma_k^2 = \prod_{k=0}^{K-1} \frac{\sqrt{\alpha}}{2} e^{-\sqrt{\alpha}|c_k|}$$

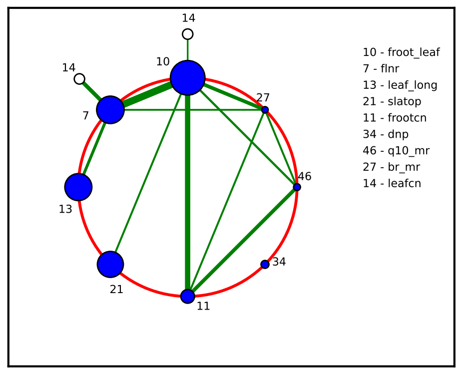
Iterative Bayesian compressive sensing (iBCS): dimensionality reduction using sparsity priors



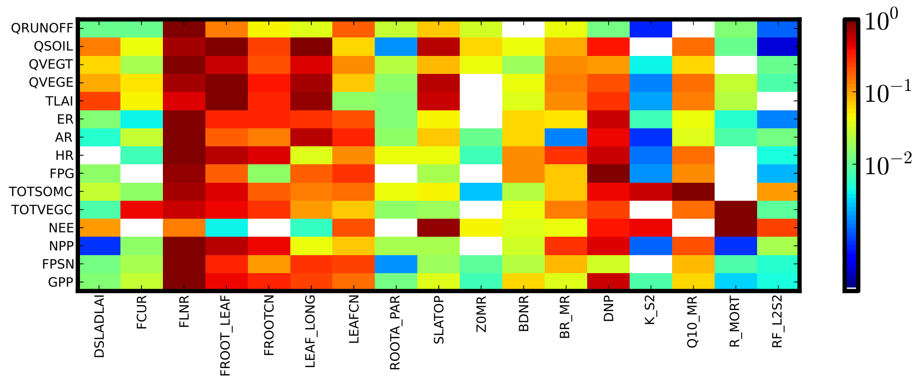
Uncertain surrogate model and global sensitivity indices

- Surrogate constructed with 10000 CLM simulations
- The iterative weighted BCS picks only ~ 500 PC bases
- Surrogate sensitivity indices computed via Monte Carlo
- Circle sizes correspond to main effect sensitivity indices
- Line widths correspond to joint sensitivity indices





Sensitivity ranking of the most important inputs



Highlights

- Surrogates are necessary for complex climate models
- Polynomial Chaos surrogate via Bayesian machinery
- High-dimensionality is addressed by the iterative weighted Bayesian compressive sensing algorithm
- Global sensitivity analysis applied to surrogate achieves dimensionality reduction
- Data clustering and classification employed for nonsmooth models to obtain a piecewise-PC surrogate model
- K. Sargsyan *et al.*, "Dimensionality Reduction for Complex Models via Bayesian Compressive Sensing", *Int. J. of Uncertainty Quantification*, 4(1), pp. 63-93, 2014.

Current and future work

- Optimal computational design to improve BCS efficiency
- Local surrogate construction and calibration of input parameters given observational data

This work was supported by the US Department of Energy, Office of Science, under the project "Climate Science for a Sustainable Energy Future", funded by the Biological and Environmental Research (BER) program.

Sandia National Laboratories

Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract No. DE-AC04-94AL85000.