SAND2014-18343PE

High-Dimensionality Challenges in Uncertainty Quantification

Habib N. Najm

hnnajm@sandia.gov

Sandia National Laboratories
Livermore, CA, USA

ICERM Workshop on Approximation, Integration, and Optimization,
Brown University, Providence, RI
Sep. 29 – Oct. 3, 2014

Acknowledgement

B.J. Debusschere, R.D. Berry, K. Sargsyan, C. Safta,
K. Chowdhary, M. Khalil — Sandia National Laboratories, CA

R.G. Ghanem — U. South. California, Los Angeles, CA
O.M. Knio — Duke Univ., Durham, NC
O.P. Le Maître — CNRS, Paris, France
Y.M. Marzouk — Mass. Inst. of Tech., Cambridge, MA

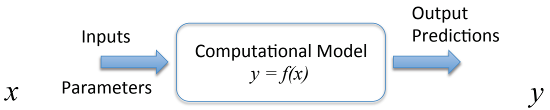
This work was supported by:

- DOE Office of Science (SC)
- DOE Office of Advanced Scientific Computing Research (ASCR), Scientific Discovery through Advanced Computing (SciDAC)
- DOE ASCR Applied Mathematics program.
- DOE Office of Basic Energy Sciences, Div. of Chem. Sci., Geosci., & Biosci.
- DOE Office of Biological & Environmental Research
- DOE/SC National Energy Research Scientific Computing Center (NERSC) under Contract No. DE-AC02-05CH11231

Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94-AL85000.

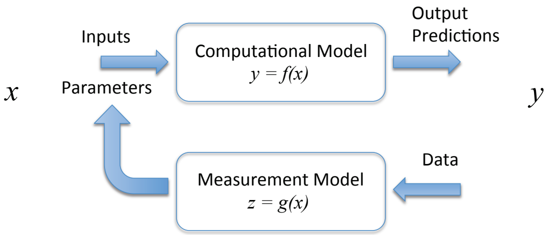
Outline

Uncertainty Quantification and Computational Science



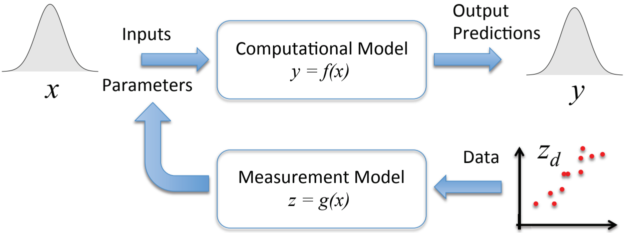
Forward problem

Uncertainty Quantification and Computational Science



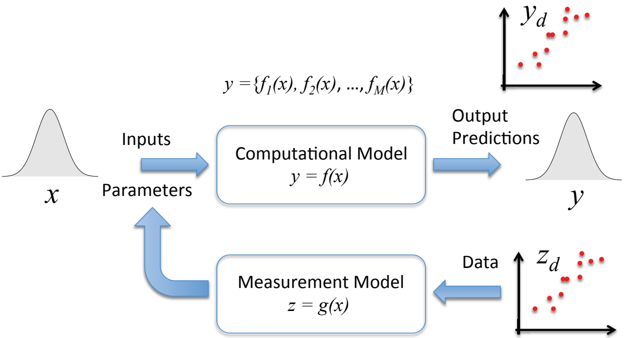
Inverse & Forward problems

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Model validation & comparison, Hypothesis testing

Relevance of hi-dimensionality in UQ

- Large number of degrees of freedom in uncertain model inputs/outputs
  - Random fields – material properties, initial/boundary conditions
  - Model parameters – reaction rate constants, physical constants, phenomenological models
  - State variables – model outputs
- Consequences
  - Joint densities
  - Functional representations of random variables/fields
  - Global sensitivity analysis
  - Forward and Inverse UQ

Probabilistic Forward UQ — $y = f(x)$

Represent uncertain quantities using probability theory

- Random sampling, MC, QMC
  - Generate random samples $\{x^i\}_{i=1}^N$ from the PDF of x, $p(x)$
  - Bin the corresponding $\{y^i\}$ to construct $p(y)$
  - Not feasible for computationally expensive $f(x)$
    - slow convergence of MC/QMC methods
      $\Rightarrow$ very large N required for reliable estimates
- Build a cheap surrogate for $f(x)$, then use MC
  - Collocation – interpolants
  - Regression – fitting
- Galerkin methods
  - Polynomial Chaos (PC)
  - Intrusive and non-intrusive PC methods

Probabilistic Forward UQ & Polynomial Chaos Representation of Random Variables

With $y = f(x)$, x a random variable, estimate the RV y

- Can describe a RV in terms of its
  - density, moments, characteristic function, or
  - as a function on a probability space
- Constraining the analysis to RVs with finite variance
  - $\Rightarrow$ Represent RV as a spectral expansion in terms of orthogonal functions of standard RVs
    - Polynomial Chaos Expansion
- Enables the use of available functional analysis methods for forward UQ

Polynomial Chaos Expansion (PCE)

- Model uncertain quantities as random variables (RVs)
- Given a *germ* $\boldsymbol{\xi}(\omega) = \{\xi_1, \cdots, \xi_n\}$ – a set of *i.i.d.* RVs
  - where $p(\boldsymbol{\xi})$ is uniquely determined by its moments

Any random variable $Z \in L^2(\Omega, \mathfrak{S}(\boldsymbol{\xi}), P)$ can be written as a PCE:

$$Z(\omega) = Z(\boldsymbol{\xi}) \simeq \sum_{k=0}^{P} z_k \Psi_k(\boldsymbol{\xi}(\omega))$$

- z_k are mode strengths
- $\Psi_k()$ are multivariate functions orthogonal w.r.t. $p(\boldsymbol{\xi})$

With dimension n and order p: $$P + 1 = \frac{(n+p)!}{n!p!}$$

Orthogonality

By construction, the functions $\Psi_k()$ are orthogonal with respect to the density of ξ

$$z_k = \frac{\langle Z\Psi_k \rangle}{\langle \Psi_k^2 \rangle} = \frac{1}{\langle \Psi_k^2 \rangle} \int Z(\lambda(\xi))\Psi_k(\xi)\, p_\xi(\xi)\, d\xi$$

Examples:

- Hermite polynomials with Gaussian basis
- Legendre polynomials with Uniform basis, ...
- Global versus Local PC methods
  - Adaptive domain decomposition of the support of ξ

Random Fields

- A random field (RF) $Z(x, \omega)$ is a function on a product space $D \times \Omega$
  - a much more complex object than a RV
  - an RV at an $x \in D$
  - an infinite dimensional object
- However, in many physical systems, uncertain field quantities, described by RFs, have an underlying *smoothness* due to correlations
- Can be represented with a small no. of stochastic degrees of freedom
- Optimal representation – second-order statistics
  - Karhunen-Loeve expansion

Random Fields – KLE

- Karhunen-Loeve Expansion (KLE) for a RF with a continuous covariance function

$$Z(x,\omega) = \mu(x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\zeta_i(\omega)\phi_i(x)$$

- $\mu(x)$ is the mean of $Z(x,\omega)$ at x
- λ_i and $\phi_i(x)$ are the eigenvalues and eigenfunctions of the covariance

$$C(x_1,x_2) = \langle [Z(x_1,\omega) - \mu(x_1)][Z(x_2,\omega) - \mu(x_2)] \rangle$$

- The ζ_i are uncorrelated zero-mean unit-variance RVs

$$\zeta_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D Z(x,\omega)\phi_i(x)\mathrm{d}x$$

Rosenblatt transform for Multi-D RVs

- Rosenblatt transformation maps any set of jointly distributed random variables $(\zeta_1, \ldots, \zeta_d)$ to uniform *i.i.d.* RVs $\eta_i \overset{iid}{\sim} U(0,1)$, $i = 1, \ldots, d$ (Rosenblatt, 1952).
- Relies on conditional CDFs $F_{k|k-1,\ldots,1}(\zeta_k|\zeta_{k-1}, \ldots, \zeta_1)$

$$\begin{aligned}
\eta_1 &= F_1(\zeta_1) \\
\eta_2 &= F_{2|1}(\zeta_2|\zeta_1) \\
&\vdots \\
\eta_d &= F_{d|d-1,\ldots,1}(\zeta_d|\zeta_{d-1}, \ldots, \zeta_1)
\end{aligned}$$

- A multi-D generalization of 1D CDF mapping.
- Conditional CDFs harder to evaluate in high dimensions

Rosenblatt transformation with samples – KDE

- Given samples $\{\zeta^i\}_{i=1}^N$ of the random vector $\zeta = (\zeta_1, \ldots, \zeta_d)$
- Kernel Density Estimation (KDE) is useful to compute the conditional CDFs

$$
\begin{aligned}
\eta_k &= F_{k|k-1,\ldots,1}(\zeta_k|\zeta_{k-1},\ldots,\zeta_1) \\
&= \int_{\infty}^{\zeta_k} p_{k|k-1,\ldots,1}(\zeta_k'|\zeta_{k-1},\ldots,\zeta_1)d\zeta_k' \\
&= \int_{\infty}^{\zeta_k} \frac{p_{k,\ldots,1}(\zeta_k',\zeta_{k-1},\ldots,\zeta_1)}{p_{k-1,\ldots,1}(\zeta_{k-1},\ldots,\zeta_1)}d\zeta_k'
\end{aligned}
$$

Rosenblatt transformation with samples – KDE

$$
\eta_k \approx \frac{1}{h\sqrt{2\pi}} \int_{-\infty}^{\zeta_k} \frac{\sum_{i=1}^{N} \exp\left(-\frac{(\zeta_1-\zeta_1^i)^2+\cdots+(\zeta_k^{'}-\zeta_k^i)^2}{2h^2}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{(\zeta_1-\zeta_1^i)^2+\cdots+(\zeta_{k-1}-\zeta_{k-1}^i)^2}{2h^2}\right)} d\zeta_k^{'}
$$

$$
= \int_{-\infty}^{\zeta_k} \frac{\sum_{i=1}^{N} \exp\left(-\frac{(\zeta_1-\zeta_1^i)^2+\cdots+(\zeta_{k-1}-\zeta_{k-1}^i)^2}{2h^2}\right) \times \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{(\zeta_k^{'}-\zeta_k^i)^2}{2h^2}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{(\zeta_1-\zeta_1^i)^2+\cdots+(\zeta_{k-1}-\zeta_{k-1}^i)^2}{2h^2}\right)} d\zeta_k^{'}
$$

$$
= \frac{\sum_{i=1}^{N} \exp\left(-\frac{(\zeta_1-\zeta_1^i)^2+\cdots+(\zeta_{k-1}-\zeta_{k-1}^i)^2}{2h^2}\right) \times \Phi\left(\frac{\zeta_k-\zeta_k^i}{h}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{(\zeta_1-\zeta_1^i)^2+\cdots+(\zeta_{k-1}-\zeta_{k-1}^i)^2}{2h^2}\right)}
$$

RF KLE PCE

- Define $\xi_i \overset{iid}{\sim} N(0,1)$, $i = 1, \ldots, d$, where $\eta_i = \Phi(\xi_i)$
- The Rosenblatt transform is then

$$\Phi(\xi_k) = F_{k|k-1,\ldots,1}(\zeta_k|\zeta_{k-1}, \ldots, \zeta_1), \quad k = 1, \ldots, d$$

- Using the inverse Rosenblatt transform, we have

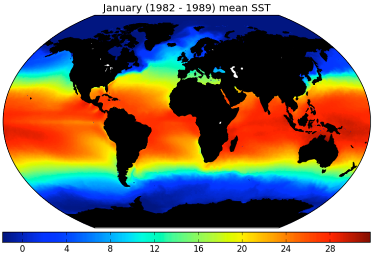
$$\zeta_k = F^{-1}_{k|k-1,\ldots,1}\left(\Phi(\xi_k)|\xi_{k-1}, \ldots, \xi_1\right) = f_k(\xi_1, \ldots, \xi_k)$$

which allows the construction of a WH PCE for each ζ_k and the representation of the RF $Z(x,\omega)$ as a WH PCE

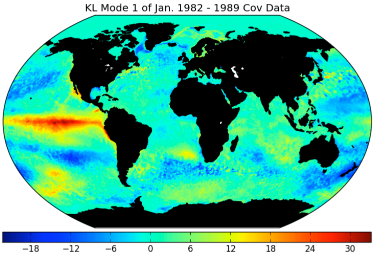
$$Z(x,\omega) = \sum_{j=0}^{\infty} Z_j(x)\Psi_j(\boldsymbol{\xi})$$

Random Fields – large scale NOAA data – SVD

- KLE for uncertain Sea Surface Temperature (SST)
  - 1/4-degree spatial resolution data
- 10^6-dimensional random field encompassing spatial and temporal uncertainty in SST data
- SVD using Trilinos / parallelized block Krylov Schur solver



Mean SST

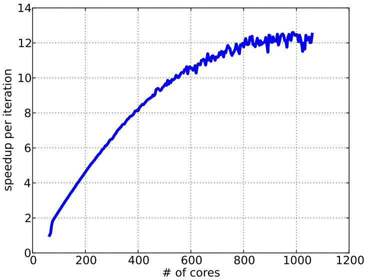


1^{st} KL mode

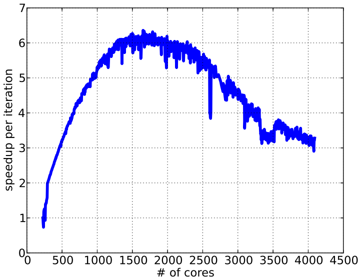
SVD on NOAA SST data – Scaling

- Parallel SVD: use Trilinos eigensolver library Anasazi
- Strong scaling study – fixed problem size
  - 3-10k samples of a 10^6-dimensional data set

3k samples

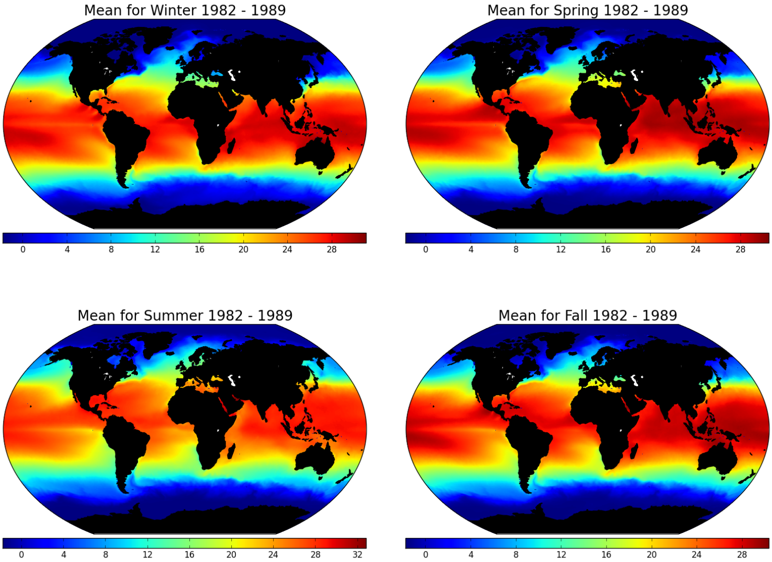


10k samples

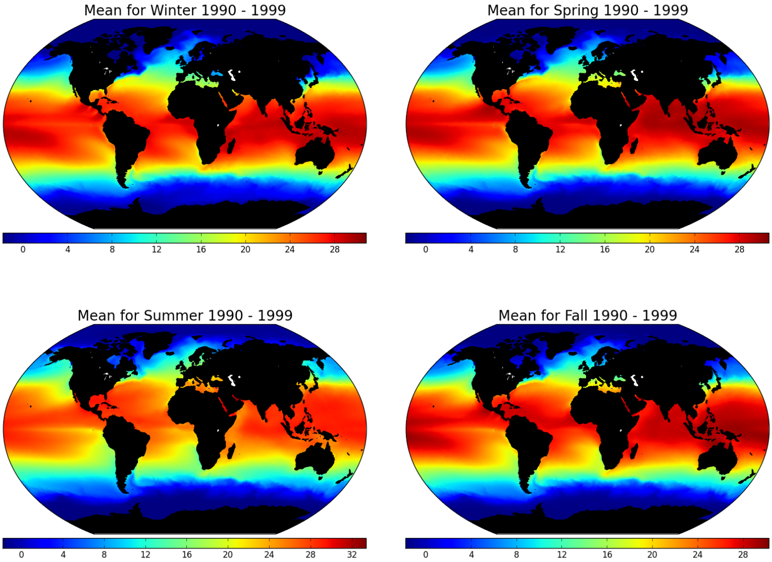


Strong-scaling speedup of the SVD solver
NERSC-Hopper

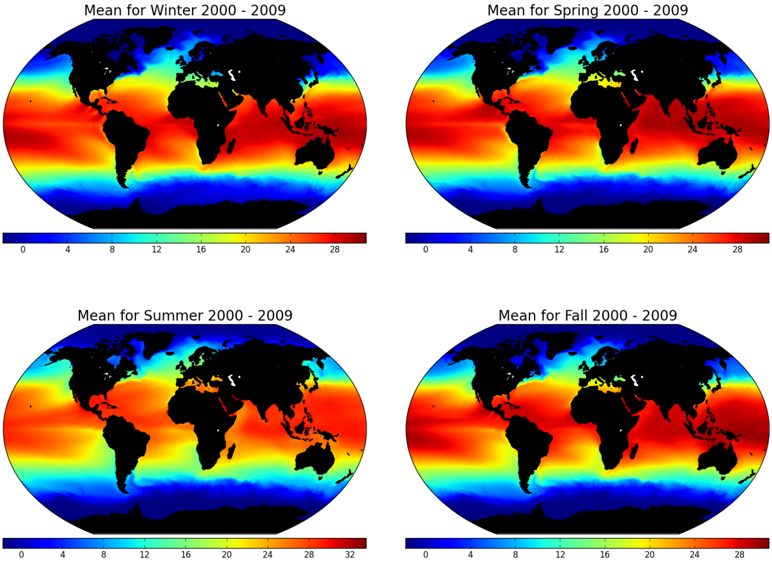
Mean SST for different seasons, 1982 - 1989



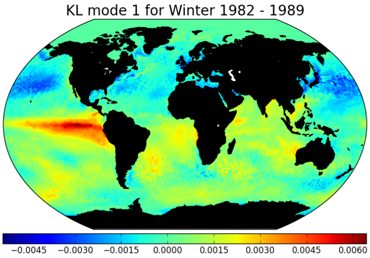
Mean SST for different seasons, 1990 - 1999

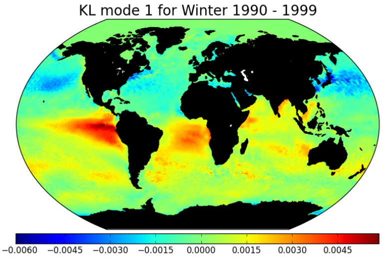


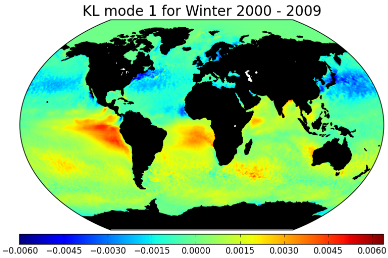
Mean SST for different seasons, 2000 - 2009



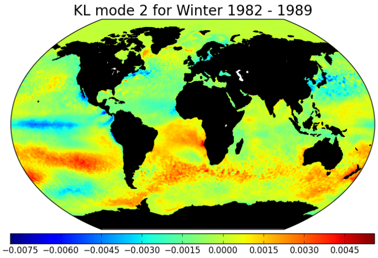
1st principal components for winter over decades

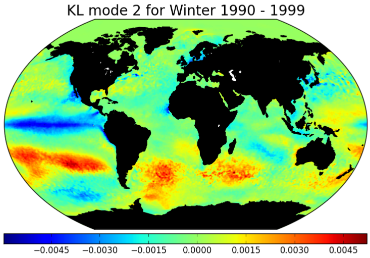


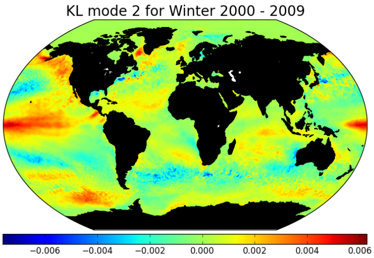




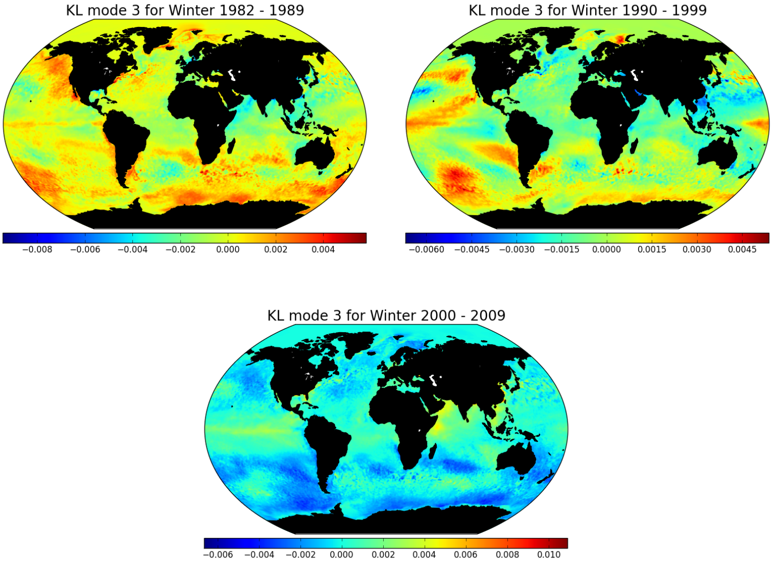
2nd principal components for winter over decades



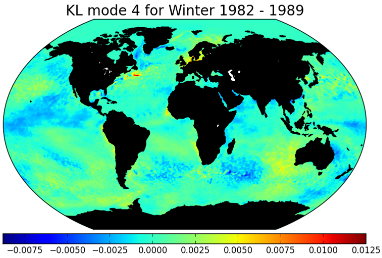


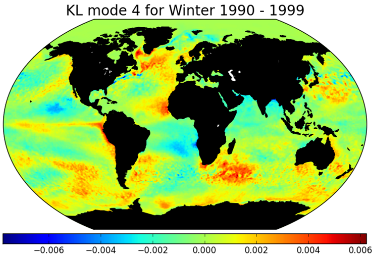


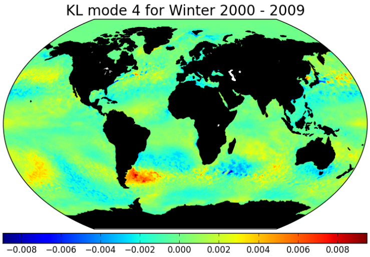
3rd principal components for winter over decades



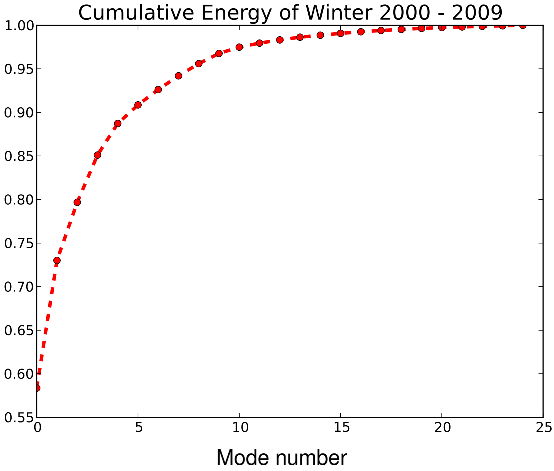
4th principal components for winter over decades





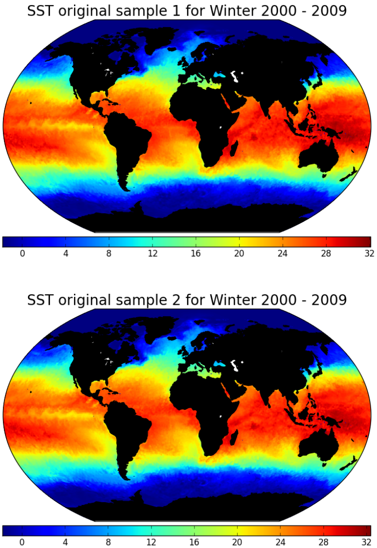


SST RF reconstruction, winter-2000s



Truncated KLE with ∼25 modes captures most of the energy

SST RF reconstruction, winter-2000s – 25 modes



Random Fields – sparse data

SNL

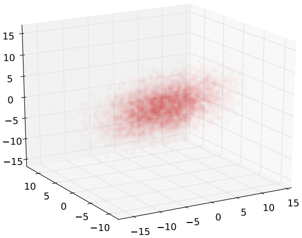
- Developed a Bayesian procedure for KLE construction given sparse data
  - Bayesian Principal Component Analysis (BPCA)/BKLE
- Address challenges arising due to
  - approximate knowledge of the covariance matrix
  - lack of positive definiteness of sample covariance matrix
- BPCA framework explores the space of orthonormal vectors, seeking those that best explain the data
  - Likelihood density $p(\Phi)$ is peaked at

$$\Phi^* = \operatorname*{argmin}_{\Phi \in V_k(\mathbb{R}^d)} \sum_{i=1}^{n} \|x^i - P_\Phi x^i\|^2$$

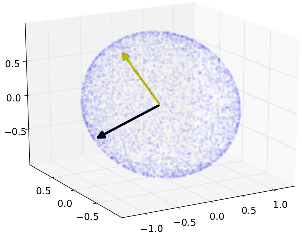
where $V_k(\mathbb{R}^d)$ is the space of k orthonormal d-dimensional vectors

- Resulting KLE incorporates uncertainty due to small number of samples

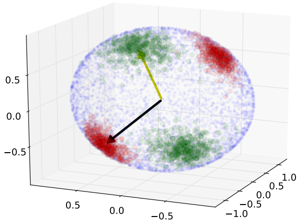
BPCA Example – Data from a 3D MVN



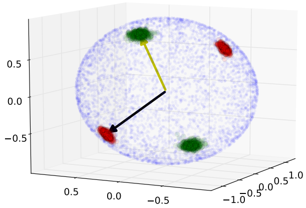
Samples of random variables from a
3D Multivariate Normal (MVN) distribution



First two principal components.
Black is the vector with maximum variance

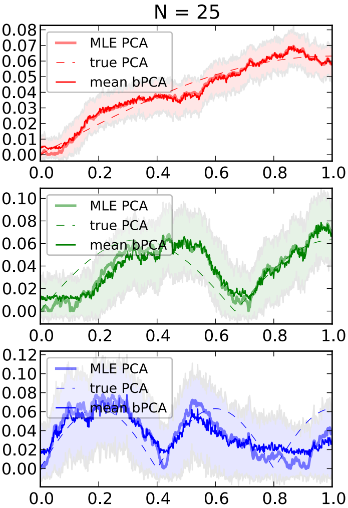


Samples from $p(\Phi)$ using 100 samples, x^i



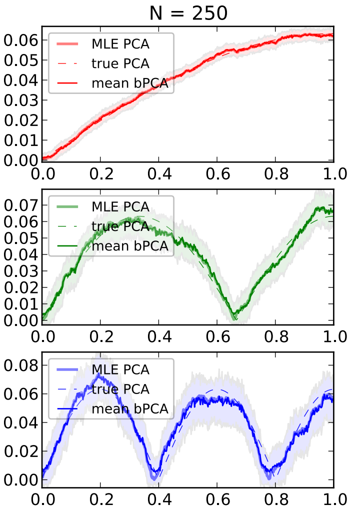
Samples from $p(\Phi)$ using 300 samples, x^i

BPCA Example – Brownian motion – 25 samples



- 500-dimensional Brownian motion stochastic process.
- Using only 25 samples, we compute samples from $p(\Phi)$ and plot the first three principal components.
- The dark solid lines represent the principal components and the shaded region represents error bars based on samples using the Bayesian PCA approach.

BPCA Example – Brownian motion – 250 samples



- Using 250 samples
- Modes are evaluated with improved accuracy
- Lower uncertainty

Essential Use of PC in UQ

Strategy:

- Represent model parameters/solution as random variables
- Construct PCEs for uncertain parameters
- Evaluate PCEs for model outputs

Advantages:

- Computational efficiency
- Utility
  - Moments: $\mathrm{E}(u) = u_0, \;\; \mathrm{var}(u) = \sum_{k=1}^{P} u_k^2 \langle \Psi_k^2 \rangle, \;\; \ldots$
  - Global Sensitivities – fractional variances, Sobol' indices
  - Surrogate for forward model

Requirement:

- RVs in L^2, *i.e.* with finite variance, on $(\Omega, \mathfrak{S}(\xi), P)$

Intrusive PC UQ: A direct *non-sampling* method

- Given model equations: $\boxed{\mathcal{M}(u(\boldsymbol{x},t);\lambda)=0}$
- Express uncertain parameters/variables using PCEs

$$u=\sum_{k=0}^{P}u_k\Psi_k;\quad \lambda=\sum_{k=0}^{P}\lambda_k\Psi_k$$

- Substitute in model equations; apply Galerkin projection
- New set of equations: $\boxed{\mathcal{G}(U(\boldsymbol{x},t),\Lambda)=0}$
  - with $U=[u_0,\dots,u_P]^T$, $\Lambda=[\lambda_0,\dots,\lambda_P]^T$
- Solving this <u>deterministic</u> system <u>once</u> provides the full specification of uncertain model ouputs

Non-intrusive PC UQ

- *Sampling*-based
- Relies on black-box utilization of the computational model
- Evaluate projection integrals *numerically*
- For any quantity of interest $\phi(\boldsymbol{x}, t; \lambda) = \sum_{k=0}^{P} \phi_k(\boldsymbol{x}, t)\Psi_k(\boldsymbol{\xi})$

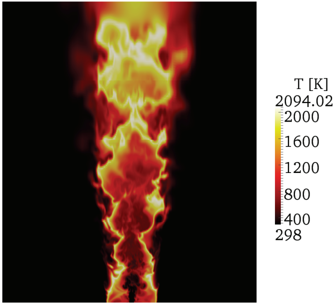
$$\phi_k(\boldsymbol{x}, t) = \frac{1}{\langle \Psi_k^2 \rangle} \int \phi(\boldsymbol{x}, t; \lambda(\boldsymbol{\xi}))\, \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \quad k = 0, \ldots, P$$

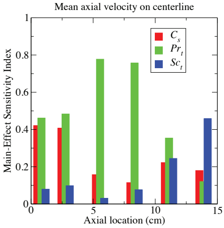
- Integrals can be evaluated using
  - A variety of (Quasi) Monte Carlo methods
    - Slow convergence; $\sim$ indep. of dimensionality
  - Quadrature/Sparse-Quadrature methods
    - Fast convergence; depends on dimensionality

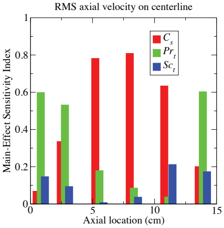
UQ in LES computations: turbulent bluff-body flame

with M. Khalil, G. Lacaze, & J. Oefelein, Sandia Nat. Labs

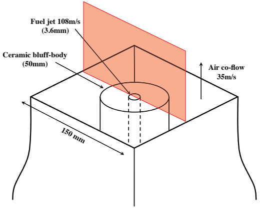
- CH_4-H_2 jet, air coflow, 3D flow
- Re=9500, LES subgrid modeling
- 12×10^6 mesh cells, 1024 cores
- 3 days run time, 2×10^5 time steps
- 3 uncertain parameters (C_s, Pr_t, Sc_t)
- 2^{nd}-order PC, 25 sparse-quad. pts







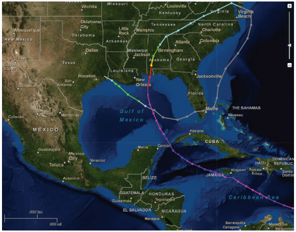
Main-Effect Sensitivity Indices

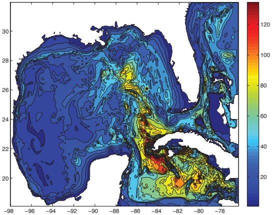


J. Oefelein & G. Lacaze, SNL

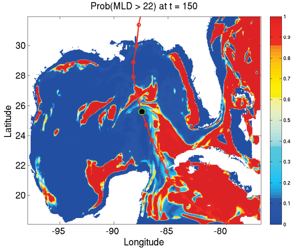
UQ in Ocean Modeling – Gulf of Mexico

A. Alexanderian, J. Winokur, I. Sraj, O.M. Knio, Duke Univ.
A. Srinivasan, M. Iskandarani, Univ. Miami;   W.C. Thacker, NOAA





- Hurricane Ivan, Sep. 2004
- HYCOM ocean model (hycom.org)
- Predicted Mixed Layer Depth (MLD)
- Four uncertain parameters, *i.i.d.* U
  - subgrid mixing & wind drag params
- 385 sparse quadrature samples



(Alexanderian *et al.*, Winokur *et. al.*, *Comput. Geosci.*, 2012, 2013)

PC and High-Dimensionality

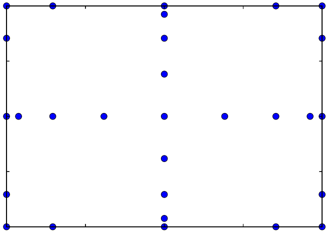
Dimensionality n of the PC basis: $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_n\}$

- $n \approx$ number of uncertain parameters
- $P + 1 = (n+p)!/n!p!$ grows fast with n

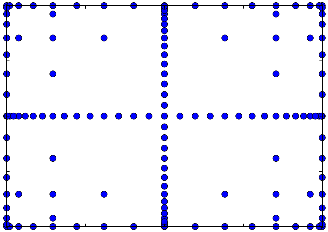
Impacts:

- Size of intrusive PC system
- Hi-D projection integrals $\Rightarrow$ large # non-intrusive samples
    - Sparse quadrature methods

Clenshaw-Curtis sparse grid, Level = 3



Clenshaw-Curtis sparse grid, Level = 5



PC Sparse Quadrature in hiD – Climate land model

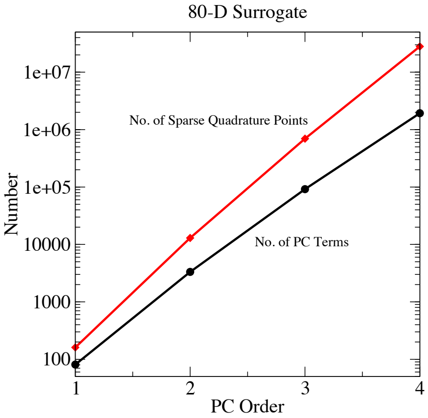
Full quadrature: $N = (N_{1D})^n$

Sparse Quadrature

- Wide range of methods
- Nested & hierarchical
- Clenshaw-Curtis: $N = \mathcal{O}(n^p)$
- Adaptive – greedy algorithms

Number of points can still be excessive in hi-D

- Large no. of terms
- Reduction/sparsity



Dimensionality Reduction

- Sensitivity analysis
  - Local vs. Global
  - Once-at-a-time; Random sampling; Response surface
- Known correlations in the input space
- Choice of basis
  - Karhunen-Loève Expansion
  - Proper Generalized Decomposition
  - Adaptive sparse representations
- Low dimensional manifolds in high dimensional space
  - Intrinsic dimension
  - PCA; Isometric map; Diffusion map
- HDMR methods (ANOVA)
  - cut-HDMR / Anchored-ANOVA
- Sparsification:
  - Compressive Sensing (CS), LASSO
  - Bayesian Compressive Sensing (BCS)

PC coefficients via sparse regression

PCE:

$$y = f(x) \simeq \sum_{k=0}^{K-1} c_k \Psi_k(x)$$

with $x \in \mathbb{R}^n$, Ψ_k max order p, and $K = (p+n)!/p!/n!$

- N samples $(x_1, y_1), \ldots, (x_N, y_N)$
- Estimate K terms $c_0, \ldots, c_{K-1}$, s.t.

$$\min ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}||_2^2$$

where $\boldsymbol{y} \in \mathbb{R}^N$, $\boldsymbol{c} \in \mathbb{R}^K$, $\boldsymbol{A}_{ik} = \Psi_k(x_i)$, $\boldsymbol{A} \in \mathbb{R}^{N\times K}$

With $N << K \Rightarrow$ under-determined

- Need some form of regularization

Regularization – Compressive Sensing (CS)

- ℓ_2-norm — Tikhonov regularization; Ridge regression:

$$\min \{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \|\boldsymbol{c}\|_2^2\}$$

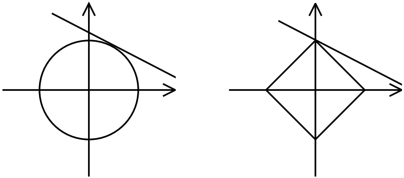
- ℓ_1-norm — Compressive Sensing; LASSO; basis pursuit

$$\min \{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \|\boldsymbol{c}\|_1\}$$

$$\min \{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2\} \quad \text{subject to } \|\boldsymbol{c}\|_1 \leq \epsilon$$

$$\min \{\|\boldsymbol{c}\|_1\} \quad \text{subject to } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 \leq \epsilon$$

$\Rightarrow$ discovery of sparse signals



Bayesian Regression

- Bayes formula

$$p(\boldsymbol{c}|D) \propto p(D|\boldsymbol{c})\pi(\boldsymbol{c})$$

- Bayesian regression: prior as a regularizer, *e.g.*
  - Log Likelihood $\Leftrightarrow \;\; \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2$
  - Log Prior $\Leftrightarrow \;\; \|\boldsymbol{c}\|_p^p$
- Laplace sparsity priors $\pi(c_k|\alpha) = \frac{1}{2\alpha} e^{-|c_k|/\alpha}$
- LASSO (Tibshirani 1996) ... formally:

$$\min \left\{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{c}\|_1 \right\}$$

  Solution $\sim$ the posterior mode of $\boldsymbol{c}$ in the Bayesian model

$$y \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{c}, I_N), \qquad c_k \sim \frac{1}{2\alpha} e^{-|c_k|/\alpha}$$

- Bayesian LASSO (Park & Casella 2008)
- Bayesian compressive sensing (Ji 2008)

Bayesian Compressive Sensing (BCS)

- Dimensionality reduction using hierarchical priors

$$p(c_k|\sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{c_k^2}{2\sigma_k^2}} \qquad\qquad p(\sigma_k^2|\alpha) = \frac{\alpha}{2} e^{-\frac{\alpha\sigma_k^2}{2}}$$

- Effectively, one obtains Laplace *sparsity* prior

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

- The parameter α can be further modeled hierarchically, or fixed.
- Evidence maximization dictates values for $\sigma_k^2, \alpha, \sigma^2$ and allows exact Bayesian solution

$$\boldsymbol{c} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with

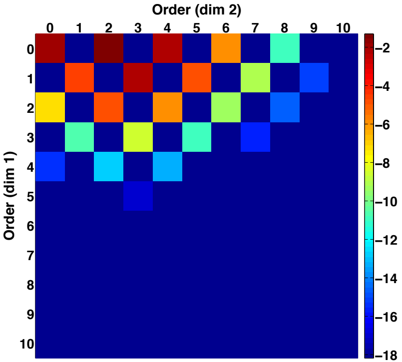
$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{P}^T\boldsymbol{u} \qquad\qquad \boldsymbol{\Sigma} = \sigma^2(\boldsymbol{P}^T\boldsymbol{P} + \mathrm{diag}(\sigma^2/\sigma_k^2))^{-1}$$

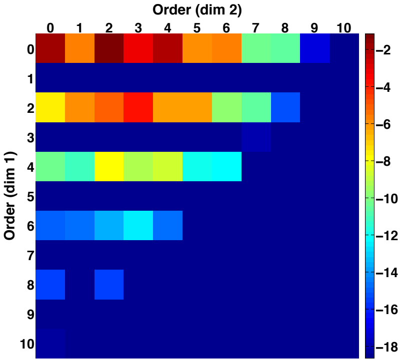
- **KEY**: Some $\sigma_k^2 \to 0$, hence the corresponding basis terms are dropped.

[Tipping, 2001, Ji *et al.*, 2008; Babacan *et al.*, 2010]

BCS removes unnecessary basis terms

$$f(x,y) = \cos(x+4y) \qquad f(x,y) = \cos(x^2+4y)$$

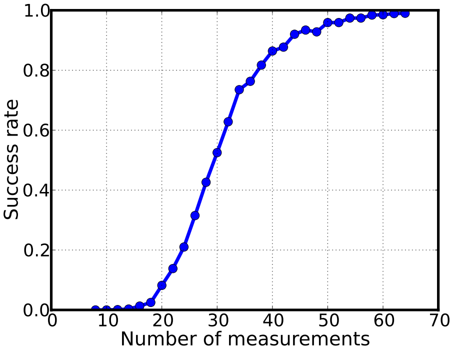




The square (i,j) represents the (log) spectral coefficient for the basis term $\psi_i(x)\psi_j(y)$.

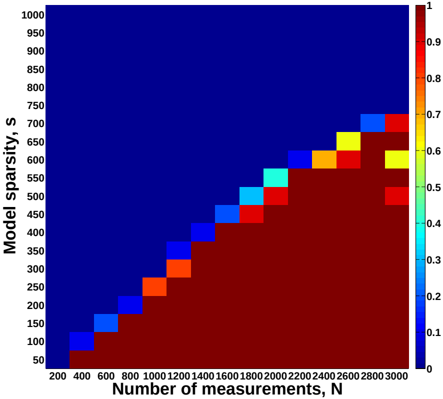
Success rate grows with more data and ‘sparser’ model

- Test function: $f(\boldsymbol{x}) = \sum_{k=0}^{K-1} c_k \Psi_k(\boldsymbol{x})$
- Only S coefficients c_k are non-zero, with $S < N < K$

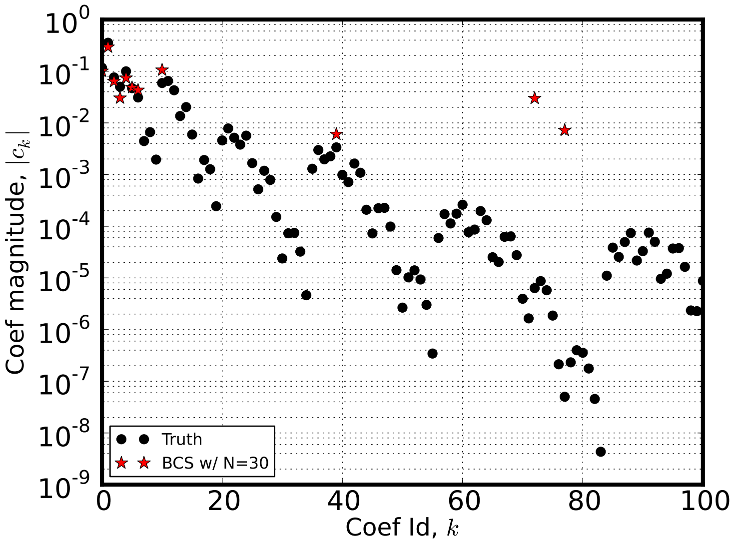


Success rate grows with more data and 'sparser' model

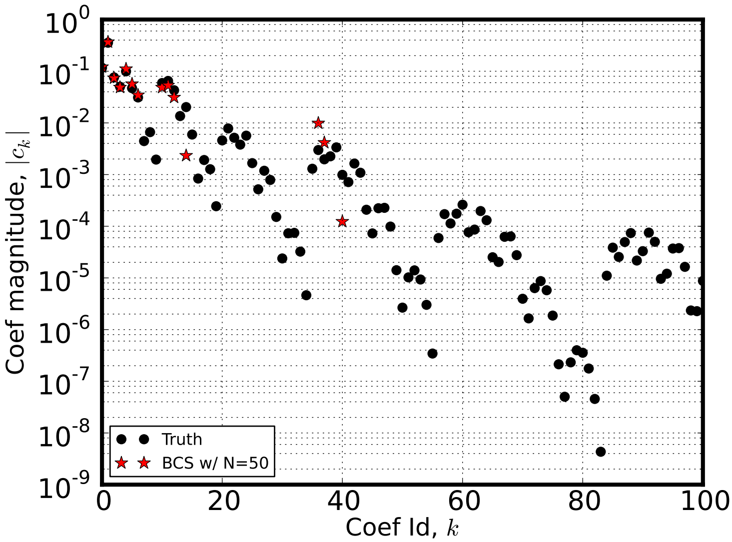
- Test function: $f(\boldsymbol{x}) = \sum_{k=0}^{K-1} c_k \Psi_k(\boldsymbol{x})$
- Only S coefficients c_k are non-zero, with $S < N < K$



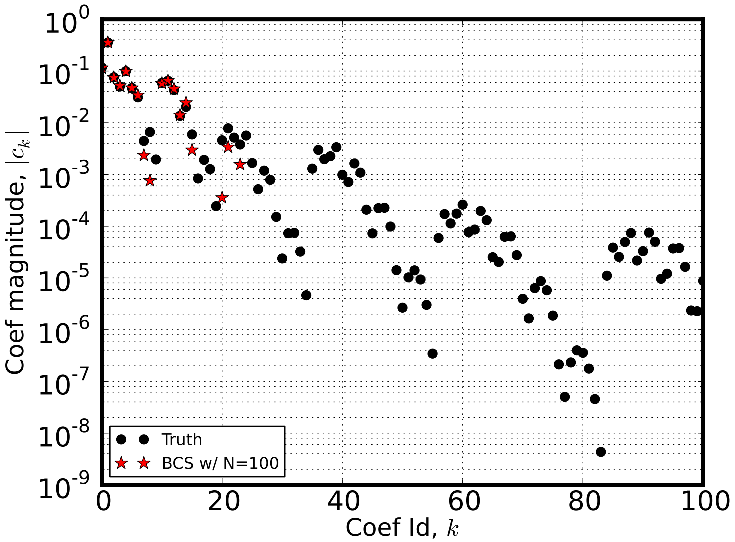
BCS recovers true coefficients with more data



BCS recovers true coefficients with more data



BCS recovers true coefficients with more data



Weighted Bayesian Compressive Sensing

- Dimensionality reduction using hierarchical priors

$$p(c_k|\sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{c_k^2}{2\sigma_k^2}} \qquad\qquad p(\sigma_k^2|\alpha_k) = \frac{\alpha_k}{2} e^{-\frac{\alpha_k \sigma_k^2}{2}}$$

- Effectively, one obtains Laplace *sparsity* prior

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

- The parameter α_k can be further modeled hierarchically, or fixed.
- Evidence maximization dictates values for $\sigma_k^2, \alpha_k, \sigma^2$ and allows exact Bayesian solution

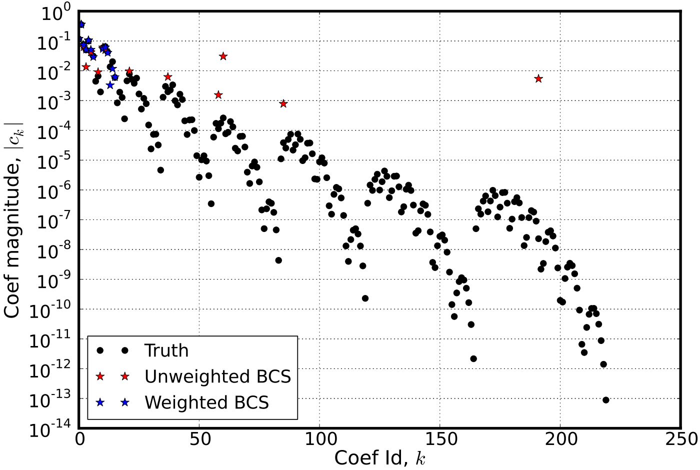
$$\boldsymbol{c} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

  with

$$\boldsymbol{\mu} = \sigma^{-2} \boldsymbol{\Sigma} \boldsymbol{P}^T \boldsymbol{u} \qquad\qquad \boldsymbol{\Sigma} = \sigma^2 (\boldsymbol{P}^T \boldsymbol{P} + \mathrm{diag}(\sigma^2/\sigma_k^2))^{-1}$$

- **KEY**: Some $\sigma_k^2 \to 0$, hence the corresponding basis terms are dropped.

WBCS recovers true coefficients better



$$f(\boldsymbol{x}) = x_0 \cos\left(e + \sum_{i=1}^{9} x_i/i\right)$$

Iteratively reweighting Compressive Sensing [Candes *et al.*, 2007]

Sparsest solution: $\min ||\boldsymbol{c}||_0$ such that $\boldsymbol{u} \approx \boldsymbol{P}\boldsymbol{c}$
Compressive sensing: $\min ||\boldsymbol{c}||_1$ such that $\boldsymbol{u} \approx \boldsymbol{P}\boldsymbol{c}$
Weighted compressive sensing: $\min ||\boldsymbol{W}\boldsymbol{c}||_1$ such that $\boldsymbol{u} \approx \boldsymbol{P}\boldsymbol{c}$

For sparse signals, $\boldsymbol{u} = \boldsymbol{P}\boldsymbol{c}$, with $||\boldsymbol{c}||_0 = S < K$, ideal weights are

$$\boldsymbol{W} = diag\left(\frac{1}{|c_k|}\right)$$ [i.e., $W_{kk} = +\infty$ if $c_k = 0$]

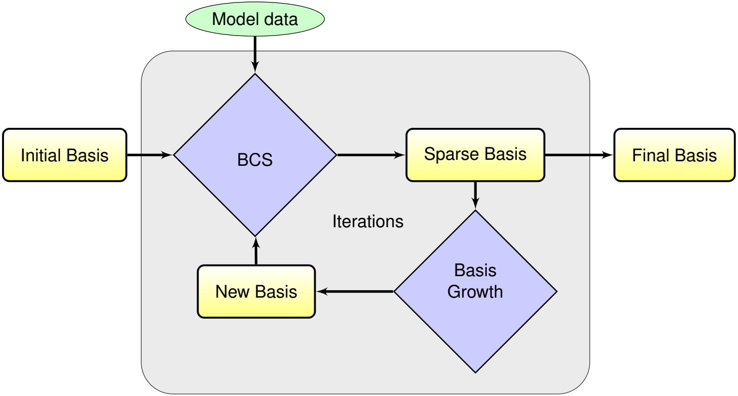
In practice, the true signal coefficients are not known, so ...

$\Rightarrow$ Iterative re-weighting

$$\boldsymbol{W}^{(i+1)} = diag\left(\frac{1}{|c_k^{(i)}| + \epsilon}\right)$$ [$\epsilon \ll 1$ for stability]

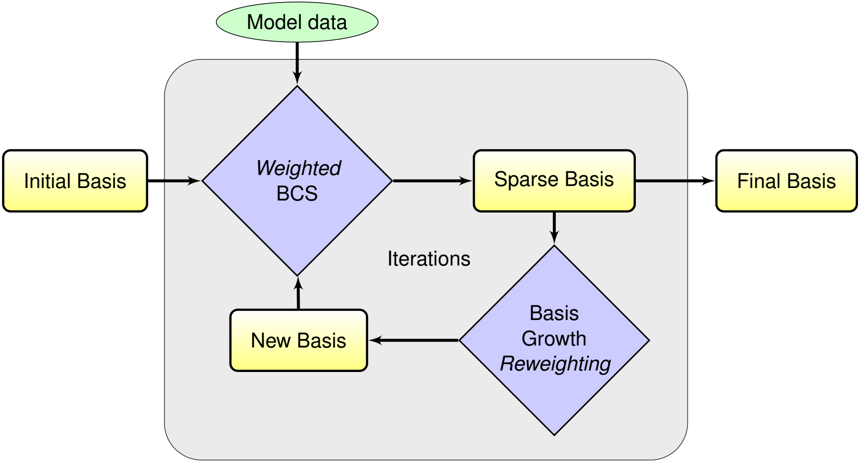
Iterative Bayesian Compressive Sensing (iBCS)

Iterative BCS: Implement an iterative procedure that allows increasing the order for the relevant basis terms while maintaining the dimensionality reduction. [Sargsyan *et al.* 2014]



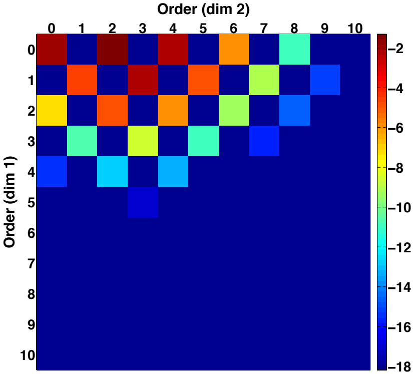
iBCS with reweighting

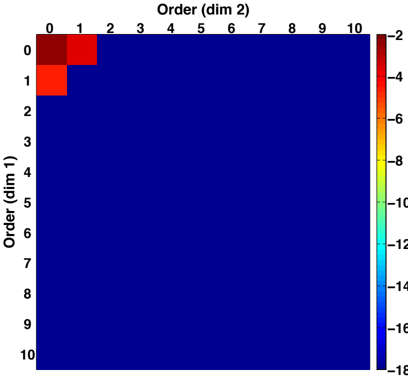
- Combine basis growth and reweighting!



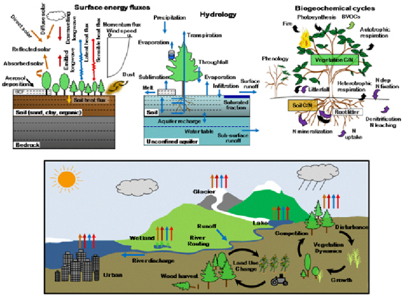
Basis set growth

$$f(x,y) = \cos(x+4y)$$





Application of Interest: Community Land Model

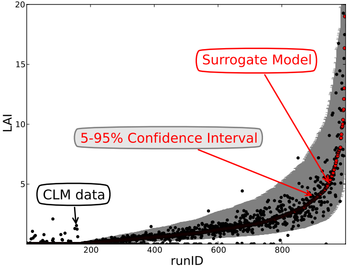


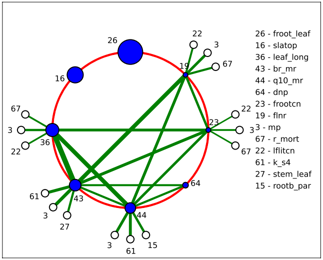
http://www.cesm.ucar.edu/models/clm/

- Nested computational grid hierarchy
- A single-site, 1000-yr simulation takes ~ 10 hrs on 1 CPU
- Involves ~ 70 input parameters; some dependent
- Non-smooth input-output relationship

Sparse PC surrogate for the Community Land Model

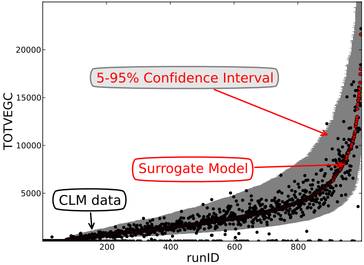
- Main effect sensitivities : rank input parameters
- Joint sensitivities : most influential input couplings
- About 200 polynomial basis terms in the 70-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data

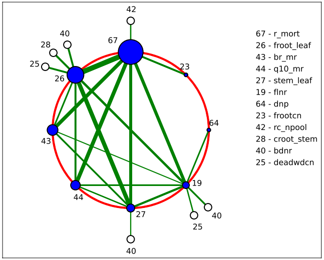




Sparse PC surrogate for the Community Land Model

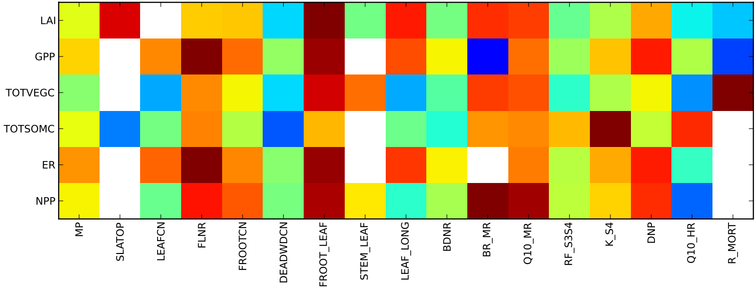
- Main effect sensitivities : rank input parameters
- Joint sensitivities : most influential input couplings
- About 200 polynomial basis terms in the 70-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data





Sparse PC surrogate for the Community Land Model

- Main effect sensitivities : rank input parameters
- Joint sensitivities : most influential input couplings
- About 200 polynomial basis terms in the 70-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data



Closure

- Probabilistic UQ framework
- Hi-dimensionality challenges
- Spectral PC representation of random variables
- Optimal KLE representation of random fields
- Dimensionality reduction – sparsity
- Bayesian compressive sensing