Quantification of Uncertainty in Extreme Scale Computations

www.quest-scidac.org

Habib N. Najm

Sandia National Laboratories,
Livermore, CA

2014 SciDAC-3 PI Meeting
July 30 – August 1, 2014
Washington, DC

Acknowledgement

QUEST Team:

SNL	M. Eldred, B. Debusschere, J. Jakeman, K. Chowdhary, C. Safta, K. Sargsyan
USC	R. Ghanem
Duke	O. Knio, O. Le Maître, J. Winokur
UT	O. Ghattas, R. Moser, C. Simmons, A. Alexanderian T. Bui-Thanh, N. Petra, G. Stadler
LANL	D. Higdon, J. Gattiker
MIT	Y. Marzouk, P. Conrad, T. Cui, A. Gorodetsky

This work was supported by:

- US Department of Energy (DOE), Office of Advanced Scientific Computing Research (ASCR), Scientific Discovery through Advanced Computing (SciDAC)

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

1. Introduction
2. Progress Highlights – Software
   - DAKOTA
   - QUESO
3. Progress Highlights – Algorithms
   - Sparsity
   - Random Fields
   - Adaptive Sparse Quadrature
   - Asympotoically Exact MCMC
4. Closure

Why UQ? Why in SciDAC?

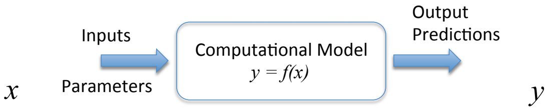
Why UQ?

- Assessment of confidence in computational predictions
- Validation and comparison of scientific/engineering models
- Design optimization
- Use of computational predictions for decision-support
- Assimilation of observational data and model construction

Why UQ in SciDAC?

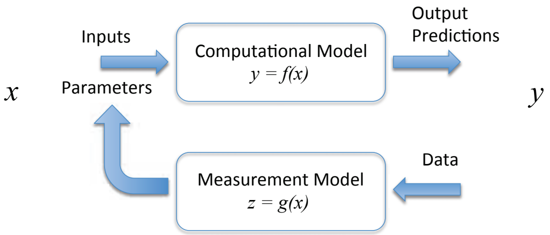
- Explore model response over range of parameter variation
- Enhanced understanding extracted from computations
- Particularly important given **cost** of SciDAC computations

Uncertainty Quantification and Computational Science



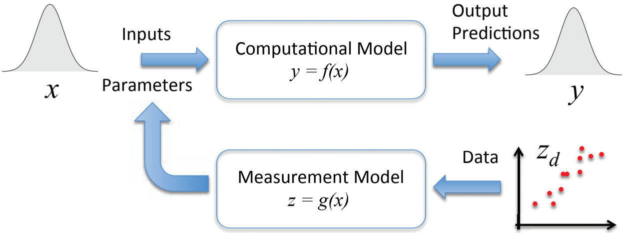
Forward problem

Uncertainty Quantification and Computational Science



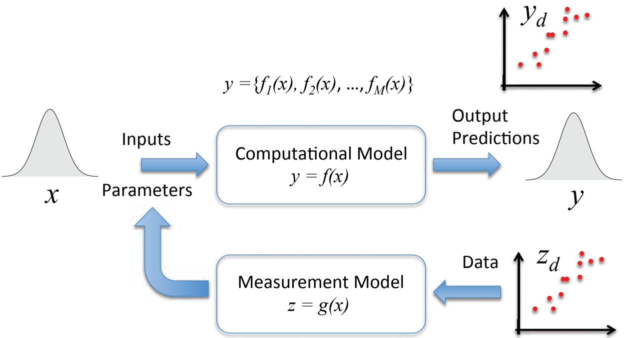
Inverse & Forward problems

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Model validation & comparison, Hypothesis testing

Key Elements of our UQ strategy

- Probabilistic framework
  - Uncertainty is represented using probability theory
- Parameter Estimation, Model Calibration
  - Experimental measurements
  - Regression, Bayesian Inference
    - Markov Chain Monte Carlo (MCMC) methods
- Forward propagation of uncertainty
  - Polynomial Chaos (PC) Stochastic Galerkin methods
    - Intrusive/non-intrusive
  - Stochastic Collocation methods
- Model comparison, selection, and validation
- Experimental design and uncertainty management

QUEST UQ Software tools

- **DAKOTA:** Optimization and calibration; non-intrusive UQ; global sensitivity analysis; ∼10K registered downloads.
- **QUESO:** Bayesian inference; multichain MCMC; model calibration and validation; decision under uncertainty.
- **GPMSA:** Bayesian inference; Gaussian process emulation; model calibration; model discrepancy analysis.
- **UQTk:** Intrusive and non-intrusive forward PC UQ; custom sparse PCE; random fields.
- **MUQ:** Adaptive forward PC UQ; advanced MCMC and variational methods for inference; efficient surrogates.

DAKOTA – Recent Developments – dakota.sandia.gov

High-dimensionality:

- Sparse representations
  - Memory conserving approaches to high-dimensional compressed sensing and variance-based decomposition
  - Multifidelity compressed sensing
- PCE regression with high dimensional basis adaptation.

Model Complexity:

- Orthogonal least interpolation
- Tail probability estimation – adaptive importance sampling
- Improved response QoI scalability

Software Integration:

- Bayesian calibration with QUESO/GPMSA/DREAM

Architecture:

- Dynamic multi-level job schedulers (MPI & hybrid)

QUESO – Recent Developments

UT Austin

- Migration to Github; expanded user base significantly
  - https://github.com/libqueso/queso
- Software quality and usability improvements
- Full user documentation and a large number of examples
- Developer documentation in development
- QUESO-Dakota interface
  - Ongoing effort to add Gaussian process (GP) based emulation capabilities to QUESO
  - Using GPMSA as a reference
  - Enabling Dakota to access such new capabilities in QUESO
- Inference of random fields
- Initial support for fault tolerant sampling
- Initial support for heterogenous architectures

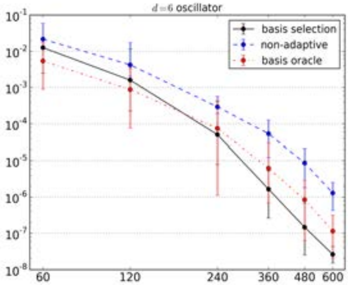
Sparsity and Compressive Sensing

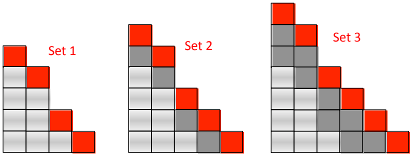
- Many physical models have a large # of uncertain inputs
- UQ in this high-dimensional setting is a major computational challenge
  - too many samples and/or large # PC modes
- Yet physical models typically exhibit sparsity
  - A small number of inputs are important
- Seek sparse PC representation on input space
  - Small number of dominant terms
- Compressed sensing (CS) is useful for discovering sparsity in high dimensional models
- Identify terms that contribute most to model output variation
- Ideal for when data is limited

Sparse Representations – developments

- CS algorithms have been developed for under-determined solutions of the coefficients of PC expansions (PCEs)
  - basis pursuit, basis pursuit denoising
  - orthogonal matching pursuit
  - least angle regression
  - least absolute shrinkage and selection operator (LASSO).
- Orthogonal least interpolation (OLI)
  - determines the lowest order PCE that can interpolate a given (unstructured) data set.
- New capabilities include:
  - support for gradient (adjoint) enhancement
  - fault tolerance
  - cross-validation of algorithm parameters
  - either structured (sub-sampled tensor product) or unstructured (Latin hypercube) grids

Adaptive Basis Selection





- Cardinality of total degree basis grows factorially with the number of uncertain inputs.
- Even for lower dimensional problems redundant basis terms can degrade accuracy
- To reduce redundancy and improve accuracy the PCE truncation can be chosen adaptively.

Random Fields – Relevance

- Many applications involve uncertain inputs/outputs that have spatial or time dependence
- Such an uncertain function, represented probabilistically, is a random field/process.
  - It is a random variable at each space/time location
  - Generally with some correlation structure in space/time
  - An infinite-dimensional object
- The Karhunen Loeve expansion (KLE) provides an optimal representation of random fields, employing a (small) number of eigenmodes of its covariance function

Random Fields – sparse data

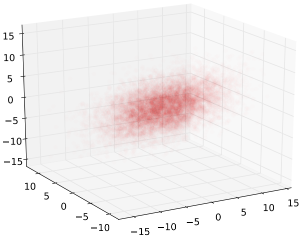
- Developed a Bayesian procedure for KLE construction given sparse data
  - Bayesian Principal Component Analysis (BPCA)
- Address challenges arising due to
  - approximate knowledge of the covariance matrix
  - lack of positive definiteness of sample covariance matrix
- BPCA framework explores the space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

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

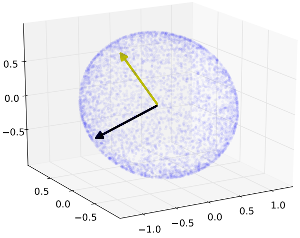
where $V_k(\mathbb{R}^d)$ is the space of k orthonormal d-dimensional vectors

- Resulting KLE incorporates uncertainty due to small number of samples

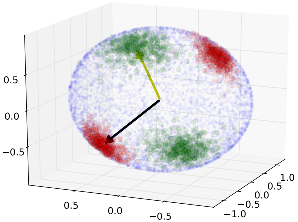
BPCA Example – Data from a 3D MVN



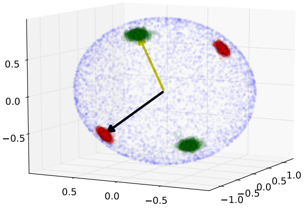
Samples of random variables from a
3D Multivariate Normal (MVN) distribution



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

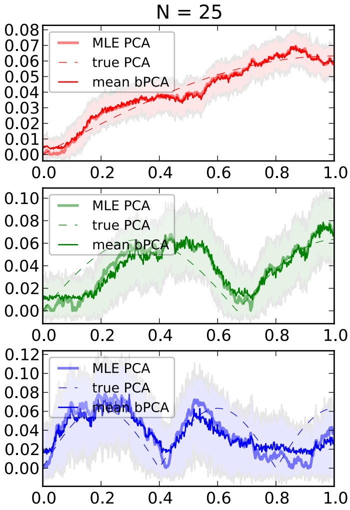


Samples from $p(\Phi)$ using 100 samples, x^i



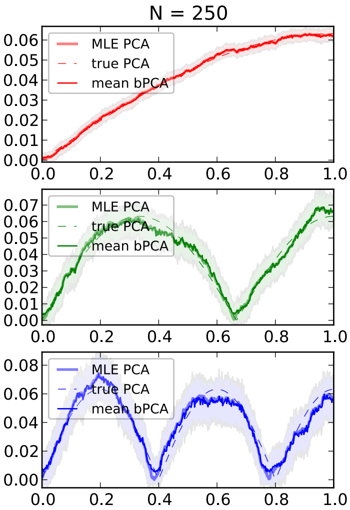
Samples from $p(\Phi)$ using 300 samples, x^i

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.

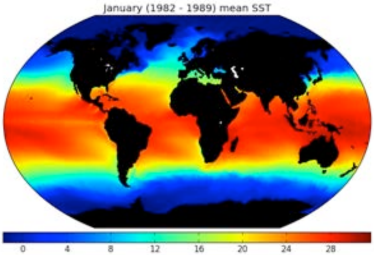
BPCA Example – Brownian motion – 250 samples



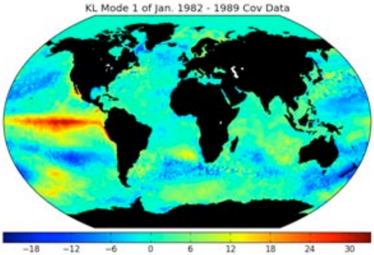
- Using 250 samples
- Modes are evaluated with improved accuracy
- Lower uncertainty

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
- Hopper / NERSC implementation



Mean SST

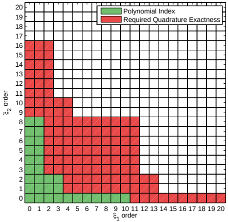


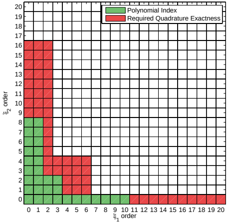
1^{st} KL mode

Adaptive Sparse Quadrature (ASQ) for UQ Duke/MIT

Non-Intrusive Pseudospectral projection using sparse tensorization of 1-D quadrature formulas:

- prevent internal aliasing
- improve accuracy
- reduce number of simulations



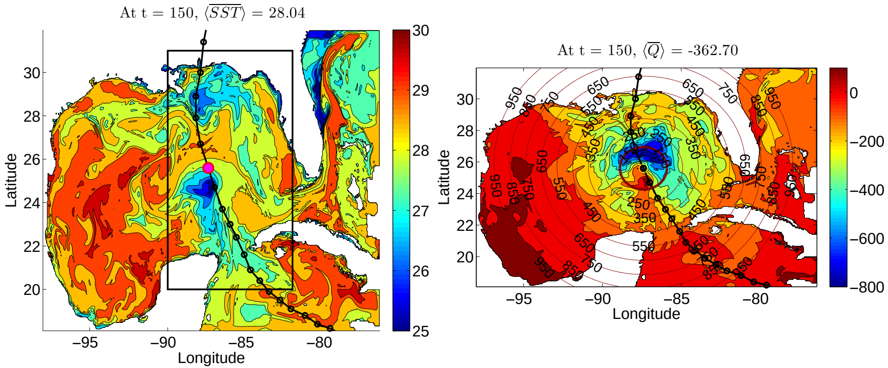


Adaptivity:

- progressive construction by introducing new tensorizations with cost control
- robust error indicator to guide the adaptation process
- nested hierarchical approximation (local dimension-wise error control)

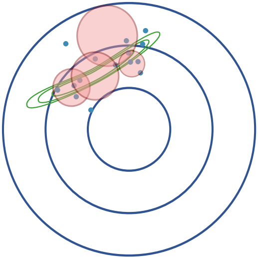
UQ with ASQ – Ocean Dynamics Simulation

Example of application: uncertainty in subgrid mixing and wind coupling parameterization (4-dimensions) in hurricane Ivan simulations ($\lesssim 400$ realizations)

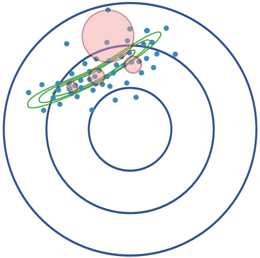


Asymptotically Exact MCMC – I

- Forward UQ yields useful surrogates for Bayesian inference
- Yet surrogates should be most accurate in regions of high posterior probability
- We have developed a new approach for incrementally constructing local approximations during MCMC



earlier times

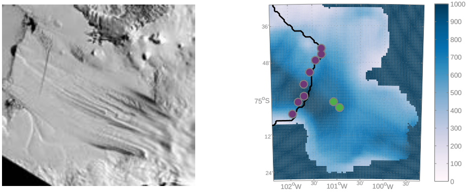


later times

Asymptotically Exact MCMC – II

- Algorithm applies *approximate* MCMC transition kernels, but is provably ergodic with respect to the *exact* posterior
- Probability of evaluating the full forward model during a given MCMC iteration approaches zero
- Speedups of several orders of magnitude over direct MCMC sampling

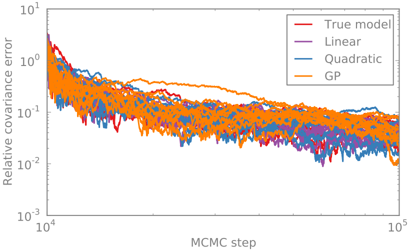
- Applied to large-scale inference problem with a *black-box forward model*: MITgcm for ice-ocean dynamics in the West Antarctic Ice Sheet (with P. Heimbach, MIT)
- Code available in the latest release of MUQ



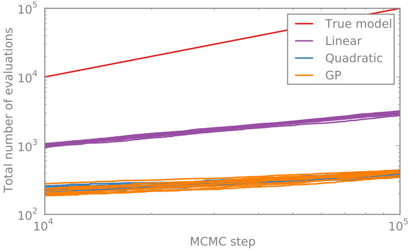
Satellite image and sample locations

Asymptotically Exact MCMC – III

- Elliptic PDE inverse problem: $\nabla \cdot (\kappa(x)\nabla u(x)) = -f$
- Infer permeability field $\kappa(x)$ from limited/noisy observations of pressure u



Accuracy of chains



Cost of chains

Only **300 model evaluations** needed for 10^5 MCMC samples!

Closure

- Highlights of recent progress
  - Software
  - Algorithms
- Refining and robustifying QUEST algorithms and software to address UQ challenges in large-scale problems
  - high dimensionality
  - large range of scales
  - complex models and high computational cost
- Addressing UQ needs of SciDAC application partnerships
  - Eight funded active partnerships

Read more at: **quest-scidac.org**