SAND2018-2934C

Uncertainty Quantification in Computational Models

Habib N. Najm

hnnajm@sandia.gov

Combustion Research Facility
Sandia National Laboratories
Livermore, CA

Oil & Gas HPC Conference
Rice University
Houston, TX
March 12-13, 2018

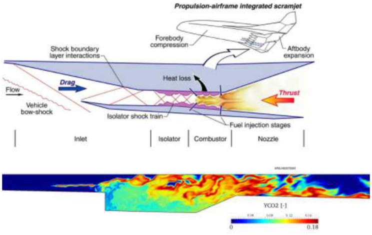
Outline

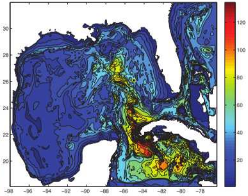
1. Introduction
   - Motivation
   - Computational perspective
2. Forward propagation of uncertainty
   - Polynomial Chaos
   - High dimensionality
   - Scramjet application
3. Inverse problem
   - Bayesian inference
   - High dimensionality
4. Closure

The Case for Uncertainty Quantification

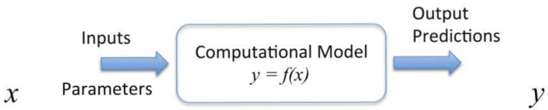
UQ is needed in:

- Assessment of confidence in computational predictions
- Validation and comparison of scientific/engineering models
- Robust design optimization under uncertainty
- Use of computational predictions for decision-support
- Assimilation of observational data and model construction
- Multiscale and multiphysics model coupling



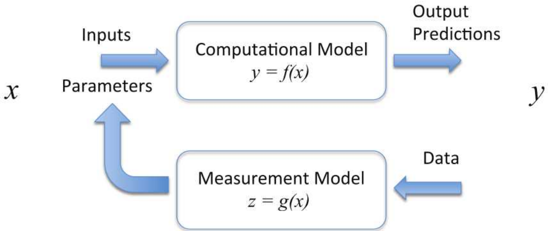


Uncertainty Quantification and Computational Science



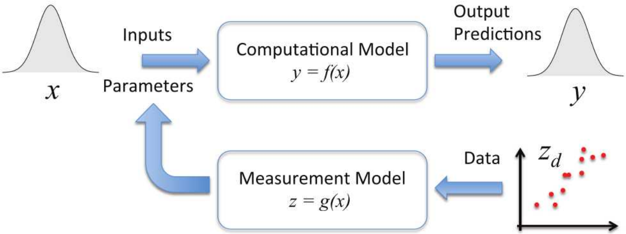
Forward problem

Uncertainty Quantification and Computational Science



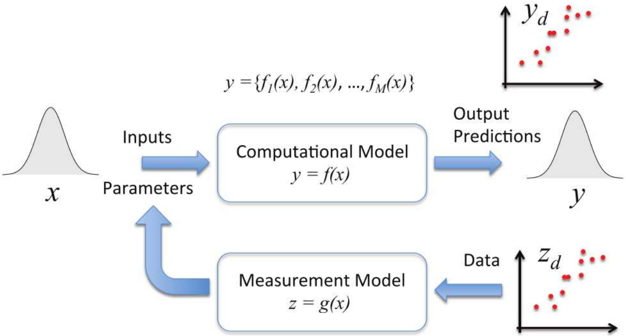
Inverse & Forward problems

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Uncertainty Quantification and Computational Science



Inverse & Forward UQ

Model validation & comparison, Hypothesis testing

UQ from a Computational Perspective

- UQ involves significant computational cost
  - Generally (many times) $\times$ the deterministic code
  - Run case cannot be the single largest capability problem
- Enables extraction of additional information on the physical system at hand
  - Effectively a parametric study over uncertain model inputs
  - Enhances scientific discovery from computations
- Computational elements of interest
  - Scalability and Performance
  - Fault tolerance
  - Code failures

Forward propagation of parametric uncertainty

Forward model: $y = f(x)$

- Local sensitivity analysis (SA) and error propagation

$$\Delta y = \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} \Delta x$$

  This is ok for:
  - small uncertainty
  - low degree of non-linearity in $f(x)$

- Non-probabilistic methods
  - Fuzzy logic
  - Evidence theory – Dempster-Shafer theory
  - Interval math

- Probabilistic methods – this is our focus

Probabilistic Forward UQ – $y = f(x)$

Represent uncertain quantities using probability theory

Random sampling, Monte Carlo (MC), Quasi-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)

Probabilistic Forward UQ & Polynomial Chaos Representation of Random Variables

With $y = f(x)$, x a random variable (RV), 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 $\boldsymbol{\xi} = \{\xi_1, \cdots, \xi_n\}$
  - Polynomial Chaos Expansion

$$u(\boldsymbol{x}, t, \omega) = f(\boldsymbol{x}, t, \boldsymbol{\xi}) \simeq \sum_{k=0}^{P} u_k(\boldsymbol{x}, t)\Psi_k(\boldsymbol{\xi}(\omega))$$

- Enables the use of functional analysis methods for forward UQ

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

Requirement:

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

Non-intrusive PC UQ

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

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

- 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

PC and High-Dimensionality

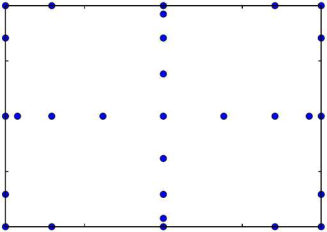
Dimensionality n of the PC basis: $\boldsymbol{\xi} = \{\xi_1, \dots, \xi_n\}$

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

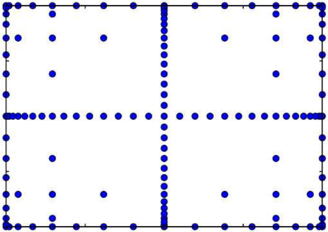
Impacts:

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

Clenshaw-Curtis sparse grid, Level = 3



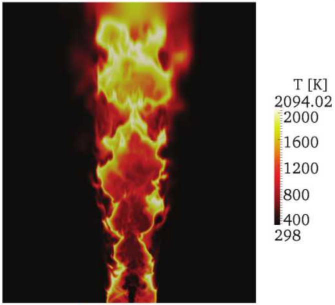
Clenshaw-Curtis sparse grid, Level = 5

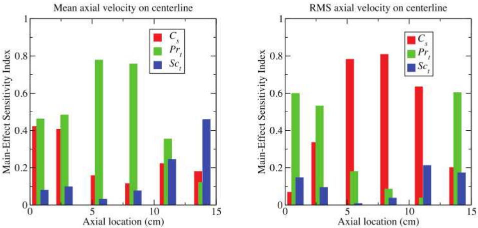


UQ in LES computations: turbulent bluff-body flame

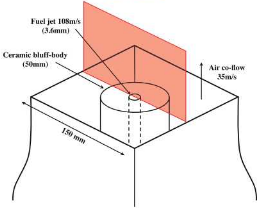
with M. Khalil, G. Lacaze, & J. Oefelein, Sandia Nat. Labs

- 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





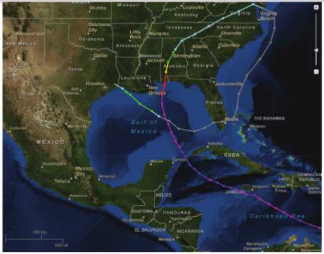
Main-Effect Sensitivity Indices

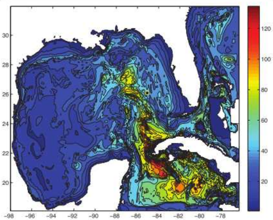


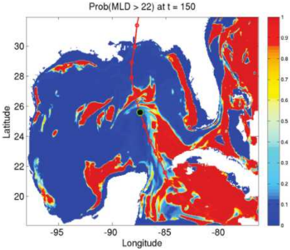
J. Oefelein & G. Lacaze, SNL

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)

High dimensionality is a dominant challenge in UQ

Forward UQ:

- High dimensionality is the result of
  - Large number of uncertain parameters/inputs
  - Large number of degrees of freedom in random field inputs
- PCE sparse-quadrature requires an unfeasible number of model evaluations for very high dimensional systems
- MC requires similarly large number of samples when the number of important dimensions is very high
- Typically, physical model output quantities of interest are *smooth*
  - Only a small number of inputs are important
- In this case, the way out is:
  - Use global sensitivity analysis (GSA) with MC to identify important parameters
  - Use PCE sparse-quadrature on the reduced dimensional space for accurate forward UQ

Hi-dimension with large-scale computational models

When the number of feasible samples for GSA is highly limited due to computational costs:

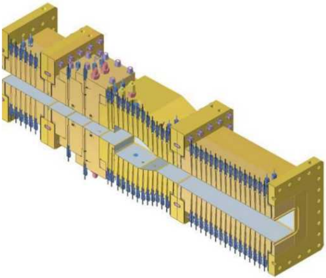
- Reliable MC-estimation of sensitivity indices requires regularization
- Presuming smoothness, use MC samples to fit a PCE, which is subsequently used to estimate the sensitivity indices
- Employ ℓ_1-norm constrained regression to discover a sparse PCE
  - compressive sensing
- Employ Multilevel Monte Carlo (MLMC), as well as Multilevel Multifidelity (MLMF) methods
  - Optimal combination of coarse/fine mesh and low/high fidelity models to minimize computational costs for a given accuracy

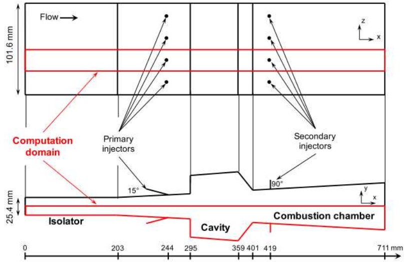
Similarly for forward PC UQ:

- Employ adaptive anisotropic sparse quadrature with MLMF methods on reduced dimensional input space

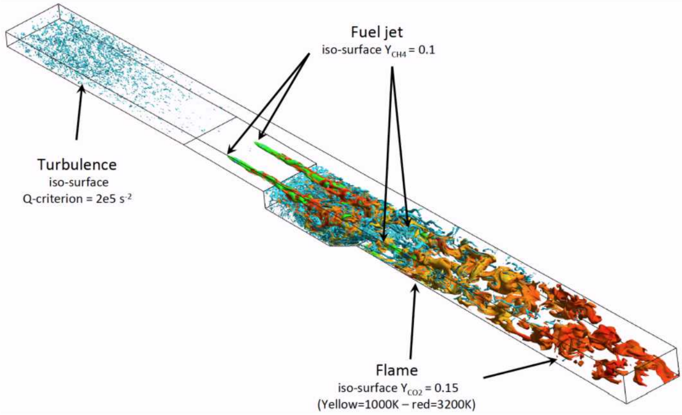
UQ in a Scramjet application

- NASA Langley Hypersonic International Flight Research and Experimentation (HIFiRE) direct connect rig (HDCR)
- LES computation of supersonic turbulent multiphase combustion
  - RAPTOR code by Joe Oefelein (Georgia Tech)
- GSA and forward UQ



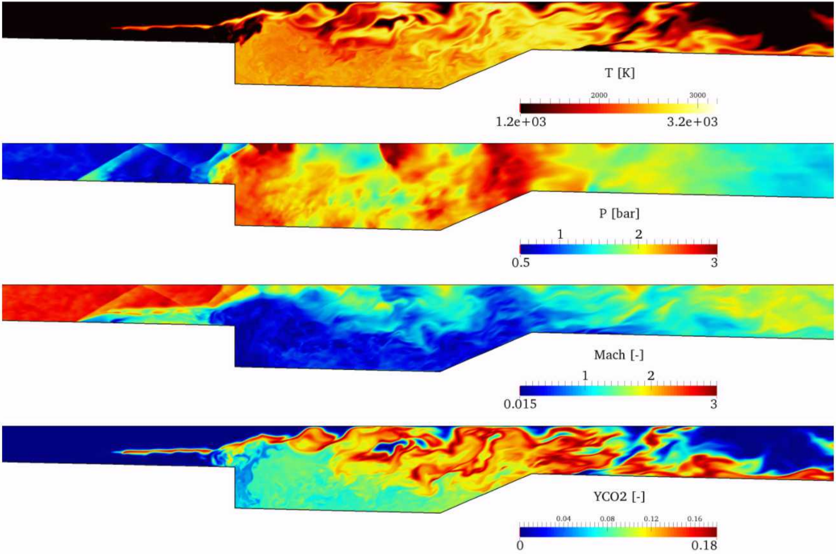


Full 3D reacting flow simulation of the HDCR

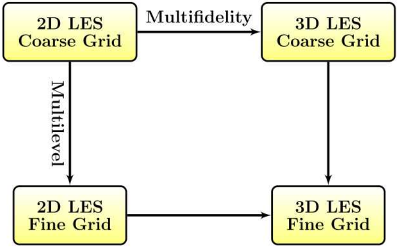


Resolution: $d_{inj}/\Delta_x = 16$
Number of cells: 66 M
Run time: 31 days on 2432 processors
CPU time for convergence: 1.8 M hours

Flowfield details in cavity region (3D d/16 resolution):



High-D – MLMF GSA

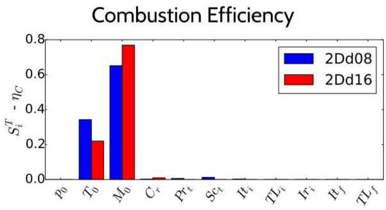


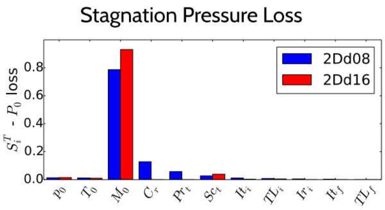
- Two model forms and two mesh discretization levels
  - 2D (LF) and 3D (HF) computations
  - $d/8$ and $d/16$ meshes
- 11 uncertain parameters
  - $p_0, T_0, M_0, C_r, Pr_t, Sc_t$
  - $It_i, TL_i, Ir_i, It_f, TL_f$
- Examine the impact on global quantities of interest
  - combustion efficiency η_c
  - stagnation pressure loss Δp_{stag}

Global Sensitivity Analysis - 2D – GSA + CS + ML-PCE

Setup

- QoIs: combustion efficiency (η_C) and stagnation pressure loss (ΔP_{stag}).
- 11 uncertain parameters: $p_0, T_0, M_0, C_r, Pr_t, Sc_t, \ldots$
- 256 simulations for each of $d/8$ and $d/16$





- Inlet Mach number (M_o) and stagnation temperature (T_o) are dominant
- Turbulence model parameters (C_R, Pr_t, Sc_t) play a minor role

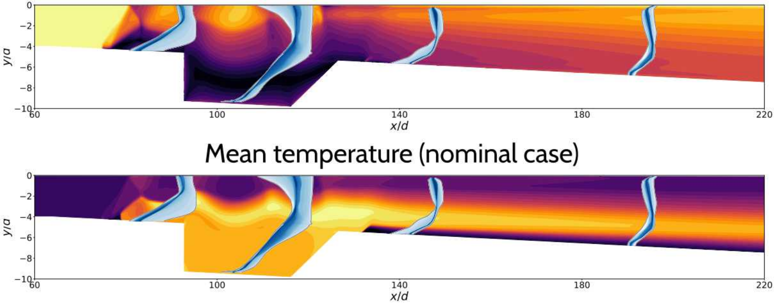
Forward UQ in the 2D Scramjet Configuration

- ML PCE sparse-quadrature relying on $d/8$ and $d/16$ computations

Axial velocity profiles with uncertainty shown with blue shades superposed on:

Mean Mach number (nominal case)

Mean temperature (nominal case)



Inverse UQ – Estimation of Uncertain Parameters

Forward UQ requires specification of uncertain inputs

Probabilistic setting

- Require joint PDF on input space
- Statistical inference – an inverse problem

Bayesian setting

- Given Data: PDF on uncertain inputs can be estimated using Bayes formula
  - Bayesian Inference
- Given Constraints: PDF on uncertain inputs can be estimated using the Maximum Entropy principle
  - MaxEnt Methods

Bayes formula for Parameter Inference

- Data Model (fit model + noise model): $y = f(\lambda) * g(\epsilon)$
- Bayes Formula:

$$p(\lambda, y) \quad = \quad p(\lambda|y)p(y) = p(y|\lambda)p(\lambda)$$

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

- Prior: knowledge of λ prior to data
- Likelihood: forward model and measurement noise
- Posterior: combines information from prior and data
- Evidence: normalizing constant for present context

Exploring the Posterior

- Given any sample λ, the un-normalized posterior probability can be easily computed

$$p(\lambda|y) \propto p(y|\lambda)p(\lambda)$$

- Explore posterior w/ Markov Chain Monte Carlo (MCMC)
  - Metropolis-Hastings algorithm:
    - Random walk with proposal PDF & rejection rules
  - Computationally intensive, $\mathcal{O}(10^5)$ samples
  - Each sample: evaluation of the forward model
    - Surrogate models
- Evaluate moments/marginals from the MCMC statistics

Bayesian inference – High Dimensionality Challenge

- Judgement on local/global posterior peaks is difficult
  - Multiple chains; Tempering
- Choosing a good starting point is very important
  - An initial optimization strategy is useful, albeit not trivial
- Choosing good MCMC proposals, and attaining good mixing
  - Likelihood-informed
    - Markov jump in those dimensions informed by data
    - Sample from prior in complement of dimensions
    - Adaptive proposal learning from MCMC samples
    - Log-Posterior Hessian $\Rightarrow$ local Gaussian approx.
    - Adaptive, Geometric, Langevin MCMC
  - Dimension independent
    - Proposal design: good MCMC performance in hiD
  - Literature: A. Stuart, M. Girolami, K. Law, T. Cui, Y. Marzouk

(Law 2014; Cui *et al.*, 2014,2015; Cotter *et al.*, 2013)

Closure

- Probabilistic UQ framework
  - Polynomial Chaos representation of random variables
- Forward UQ
  - Computational challenges in high dimension
  - GSA, CS, MLMF, etc ...
  - Basis adaptation, active subspace, manifold methods
- Inverse UQ
  - High dimensionality challenges
  - Bayesian estimation of model error
  - Model comparison and selection – Bayesian model evidence
- Open source tools from Sandia National Labs
  - Dakota – dakota.sandia.gov
  - UQTk – www.sandia.gov/UQToolkit

Acknowledgement

B.J. Debusschere, R.D. Berry, K. Sargsyan, C. Safta,
K. Chowdhary, M. Khalil, X. Huan, T. Casey,
J. Oefelein, G. Lacaze, Z. Vane, L. Hakim
– 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 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.
- DOD DARPA DSO

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.