Uncertainty Quantification in Computational Models

Habib N. Najm

Sandia National Laboratories, Livermore, CA, USA

Workshop on Understanding Climate Change from Data (UCC11)
University of Minnesota, Minneapolis, MN
August 15-16 2011

Acknowledgement

B.J. Debusschere, R.D. Berry, K. Sargsyan, C. Safta
— Sandia National Laboratories, CA
R.G. Ghanem — U. South. California, Los Angeles, CA
O.M. Knio — Johns Hopkins Univ., Baltimore, MD
O.P. Le Maître — CNRS, France
Y.M. Marzouk — Mass. Inst. of Tech., Cambridge, MA

This work was supported by:

- US DOE Office of Basic Energy Sciences (BES) Division of Chemical Sciences, Geosciences, and Biosciences.
- US DOE Office of Advanced Scientific Computing Research (ASCR), Applied Mathematics program & 2009 American Recovery and Reinvesment Act.
- US DOE Office of Biological and Environmental Research (BER)

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

The Case for Uncertainty Quantification

UQ enables:

- enhanced scientific understanding from computations
  - exploration of model predictions over range of uncertainty
- Assessment of confidence in computational predictions
- Validation and comparison of scientific/engineering models employing (noisy) data
- Design optimization
- Use of computational predictions for decision-support
- Assimilation of observational data and model construction

Sources of Uncertainty in computational models

- Lack of knowledge, and data noise
- model structure
  - participating physical processes
  - governing equations
  - constitutive relations
- model parameters
  - transport properties
  - thermodynamic properties
  - constitutive relations
  - rate coefficients
- initial and boundary conditions, geometry
- numerical errors, bugs
- faults, data loss, silent errors

Overview of UQ Methods

Estimation of model/parametric uncertainty

- Expert opinion, data collection
- Regression analysis, fitting, parameter estimation
- Bayesian inference of uncertain models/parameters

Forward propagation of uncertainty in models

- Local sensitivity analysis (SA) and error propagation
- Fuzzy logic; Evidence theory — interval math
- Probabilistic framework — Global SA / stochastic UQ
    - Random sampling, statistical methods
    - Polynomial Chaos (PC) methods
        - Collocation methods — sampling — non-intrusive
        - Galerkin methods — direct — intrusive

Polynomial Chaos Methods for UQ

- Model uncertain quantities as random variables (RVs)
- Any RV with finite variance can be represented as a Polynomial Chaos expansion (PCE)

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

– $u_k(\mathbf{x},t)$ are mode strengths
– $\boldsymbol{\xi}(\omega) = \{\xi_1, \cdots, \xi_n\}$ is a vector of standard RVs
– $\Psi_k()$ are functions orthogonal w.r.t. the density of $\boldsymbol{\xi}$

- 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 the basis/*germ* ξ

$$u_k(\boldsymbol{x}, t) \quad = \quad \frac{\langle u\Psi_k \rangle}{\langle \Psi_k^2 \rangle} = \frac{1}{\langle \Psi_k^2 \rangle} \int u(\boldsymbol{x}, t; \lambda(\boldsymbol{\xi}))\, \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

Examples:

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

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
- Sensitivity information

Requirement:

- Random variables in L^2, i.e. with finite variance

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

- Given model equations: $\mathcal{M}(u(\mathbf{x},t);\lambda) = 0$
- 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: $\mathcal{G}(U(\mathbf{x},t),\Lambda) = 0$
  - with $U = [u_0, \dots, u_P]^T$, $\Lambda = [\lambda_0, \dots, \lambda_P]^T$
- Solving this system *once* provides the full specification of uncertain model ouputs

Intrusive PC UQ ODE example

$$\boxed{\frac{du}{dt} = f(u;\lambda)}$$

$$\lambda = \sum_{i=0}^{P} \lambda_i \Psi_i \qquad u(t) = \sum_{i=0}^{P} u_i(t)\Psi_i$$

$$\boxed{\frac{du_i}{dt} = \frac{\langle f(u;\lambda)\Psi_i\rangle}{\langle \Psi_i^2\rangle} \qquad i = 0,\ldots,P}$$

Say $f(u;\lambda) = \lambda u$, then

$$\frac{du_i}{dt} = \sum_{p=0}^{P}\sum_{q=0}^{P} \lambda_p u_q C_{pqi}, \quad i = 0,\cdots,P$$

where the tensor $C_{pqi} = \langle \Psi_p\Psi_q\Psi_i\rangle / \langle \Psi_i^2\rangle$ is readily evaluated

Non-intrusive Spectral Projection (NISP) PC UQ

Sampling-based; black-box use of the computational model.
For any model output of interest $\phi(\cdot;\lambda(\boldsymbol{\xi})) = \sum_k \phi_k(\cdot)\Psi_k(\boldsymbol{\xi})$:

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

- Integrals can be evaluated *numerically* using
  - A variety of (Quasi) Monte Carlo methods
  - Quadrature/Sparse-Quadrature methods
- PC surface $\sum_k \phi_k(\cdot)\Psi_k(\boldsymbol{\xi})$ can be fitted using regression or Bayesian Inference employing computational samples
  - Discovering/exploiting sparsity via L^1-norm minimization
    - (Bayesian) compressed sensing
      – Use in CESM/CLM4 study in progress
    - Lasso

Challenges in PC UQ – High-Dimensionality

- Dimensionality n of the PC basis: $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_n\}$
  - number of degrees of freedom
  - $P + 1 = (n + p)!/n!p!$ grows fast with n
- Impacts:
  - Size of intrusive system
  - # non-intrusive (sparse) quadrature samples
- Generally $n \approx$ number of uncertain parameters
- Reduction of n:
  - Sensitivity analysis
  - Dependencies/correlations among parameters
  - Identification of dominant modes in random fields

    Karhunen-Loéve, PCA, ...
  - ANOVA/HDMR methods
  - L^1 norm minimization

Challenges in PC UQ – Non-Linearity

- Bifurcative response at critical parameter values
  - Rayleigh-Bénard convection
  - Transition to turbulence
  - Chemical ignition
- Discontinuous $u(\lambda(\boldsymbol{\xi}))$
  - Failure of global PCEs in terms of smooth $\Psi_k()$
  - $\Leftrightarrow$ failure of Fourier series in representing a step function
- Local PC methods
  - Subdivide support of $\lambda(\boldsymbol{\xi})$ into regions of smooth $u \circ \lambda(\boldsymbol{\xi})$
  - Employ PC with compact support basis on each region
  - A spectral-element vs. spectral construction
  - Domain-mapping for arbitrary discontinuity shapes
    - Application in climate – AMOC ON/OFF switching

Challenges in PC UQ – Time Dynamics

- Systems with limit-cycle or chaotic dynamics
- Large amplification of phase errors over long time horizon
- PC order needs to be increased in time to retain accuracy
- Time shifting/scaling remedies

- Futile to attempt representation of detailed turbulent velocity field $\boldsymbol{v}(\boldsymbol{x}, t; \lambda(\boldsymbol{\xi}))$ as a PCE
  - Fast loss of correlation due to energy cascade
  - Problem studied in 60's and 70's
- Focus on flow statistics, *e.g.* Mean/RMS quantities
  - Well behaved
  - Argues for non-intrusive methods with DNS/LES of turbulent flow

Estimation of Input/Parametric Uncertainties

- Need the joint PDF on the input space
  - Published data is frequently inadequate
- Bayesian inference can provide the joint PDF
  - Requires raw data ... frequently not available
- At best: (legacy) nominal parameter values and error bars

- Fitting hypothesized PDFs to each parameter nominals/bounds independently is not a good answer
  - Correlations and joint PDF structure can be crucial to uncertainty in predictions
- Bayesian methods that make other available information *explicit* in the PDF structure ... “Data-Free" Inference

Closure

- UQ is increasingly important in computational modeling
- Probabilistic UQ framework
  - PC representation of random variables
  - Utility in forward UQ
    - Intrusive PC methods
    - Non-intrusive methods
  - Challenges
    - Nonlinearity
    - High-dimensionality
    - Time dynamics
    - Probabilistic characterization of uncertain inputs
- UQ is particularly relevant in climate modeling
  - consequential predictions

Outlook

Ongoing research on various fronts

- Dimensionality reduction
  - Sensitivity, PCA, ANOVA/HDMR, low-D manifolds, CS, ...
- Discontinuities in high-D spaces
  - Efficient tiling of high-D spaces
- Adaptive anisotropic sparse quadrature
- Adaptive sparse tensor representations
- Long-time oscillatory dynamics in field variables
- Intrusive solvers ... stability, convergence, preconditioning
- Methods for characterization of uncertain inputs
  - Absence of data, dependencies among observations
- Model comparison, selection, validation

A Selection of PC Literature

Ghanem, R., Spanos, P., “Stochastic Finite Elements: A Spectral Approach", Springer Verlag, (1991).

Debusschere, B., Najm, H., Pébay, P., Knio, O., Ghanem, R., Le Maître, O., “Numerical Challenges in the Use of Polynomial Chaos Representations for Stochastic Proecesses", *SIAM J. Sci. Comp.*, **26**(2):698-719, (2004).

Le Maître, O., Knio, O., “Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics", Springer-Verlag, (2010).

Ganapathysubramanian, B., Zabaras, N., “Sparse Grid Collocation Schemes for Stochastic Natural Convection Problems", *J. Comp. Phys.*, **225**(1):652-685, (2007).

Le Maître, O., Najm, H., Ghanem, R., Knio, O., “Multi Resolution Analysis of Wiener-type Uncertainty Propagation Schemes", *J. Comp. Phys.*, **197**:502-531, (2004).

Marzouk, Y., Najm, H., “Dimensionality Reduction and Polynomial Chaos Acceleration of Bayesian Inference in Inverse Problems", *J. Comp. Phys.*, **228**(6):1862-1902, (2009).

Najm, H., “Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics", *Ann. Rev. Fluid Mech.*, **41**(1):35-52, (2009).

Nobile, F., Tempone, R., Webster, C., “An Anisotropic Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data", *SIAM J. Num. Anal.*, **46**(5):2411-2442, (2008).

Soize, C. Ghanem, R., “Physical Systems with Random Uncertainties: Chaos Representations with Arbitrary Probability Measure", *SIAM J. Sci. Comp.*, **26**(2):395-410, (2004).

Xiu, D., Karniadakis, G., “The Wiener-Askey Polynomial Chaos for Stocahstic Differential Equations", *SIAM J. Sci. Comp.*, **24**(2):619-644, (2002).

Xiu, D., “Numerical Methods for Stochastic Computations: A Spectral Method Approach", Princeton U. Press (2010).