SAND2011-6400C

Uncertainty Quantification

Habib N. Najm

Sandia National Laboratories, Livermore, CA, USA

SAMSI Methodology Workshop
Raleigh, NC
Sep 7-10, 2011

Acknowledgement

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

This work was supported by:

- The DOE Office of Basic Energy Sciences (BES) Division of Chemical Sciences, Geosciences, and Biosciences.
- The US Department of Energy (DOE), Office of Advanced Scientific Computing Research (ASCR), Applied Mathematics and SciDAC programs.
- 2009 American Recovery and Reinvesment Act.

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

- model structure
  - participating physical processes
  - governing equations
  - constitutive relations
- model parameters
  - transport and thermodynamic properties
  - constitutive relations, equations of state
  - source term rate parameters
- 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

Different Types of Uncertainty?

- Epistemic versus Aleatoric uncertainty
- Both *can* be handled equally well with probability theory
  - Bayesian viewpoint encompasses both
  - Probabilistic math structure is self-consistent for both
- When interval methods are used in practical problems:
  - Challenges with blow up of interval ranges – Singer, SISC 2006
  - Resort to random sampling – Kreinovich, RC 2007
- Any quantity can be estimated probabilistically
  - Expert opinion
  - Maximum Entropy
  - Bayes rule
- Epistemic parameters at interval limits can be easily accomodated using conditional probability densities

Parameter Estimation

- Model calibration — Inverse problem – Bayes rule
  - Data:
    - Experimental observations
    - Computational predictions – high-fidelity “truth" model
  - Missing data — Bayesian Imputation
    - Simulates missing data using posterior predictive distribution given observed values
    - Observed data posterior
  - No data – but given summary statistics
    - Simulate data satisfying summary statistics/constraints
    - Pooled posterior
- Expert elicitation
- Computational predictions — Forward UQ

Bayes formula for Parameter Inference

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

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

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

- 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

Surrogate Models for Bayesian Inference

- Need an inexpensive response surface for
  - Observables of interest y
  - as functions of parameters of interest λ
- Gaussian Process (GP) surrogate
  - GP goes through all data points with probability 1.0
  - Uncertainty between the points
- Fit a convenient polynomial to $y = f(\lambda)$
  - over the range of uncertainty in λ
  - Employ a number of samples (λ_i, y_i)
  - Fit with interpolants, regression, ... global/local
  - With uncertain λ :
    - Construct polynomial chaos response surface

Prior Modeling

- Informative prior
- (Mostly) Uninformative prior
  - Improper prior
  - Objective prior
  - Maxent prior
  - Reference prior
  - Jeffreys prior
- The choice of prior can be crucial when there is little information in the data relative to the number of degrees of freedom in the inference problem
- When there is sufficient information in the data, the data can overrule the prior

Likelihood Modeling

- This is frequently the *core* modeling challenge
  - Error model: a statistical model for the discrepancy between the forward model and the data
  - composition of the error model with the forward model
- Hierarchical Bayes modeling, and dependence trees

$$p(\phi, \theta) = p(\phi|\theta)p(\theta)$$

- Choice of observable – constraint on Quantity of Interest?
- Stochastic versus Deterministic forward models
  - Intrinsic noise term, *e.g.* Langevin eqn.
  - Specified uncertain parameter in fit model
- Error model:
  - Composed of discrepancy between
    - data and the truth – (data error)
    - model prediction and the truth – (model error)
  - Mean bias and correlated/uncorrelated noise structure

Experimental Data

- Empirical data error model structure can be informed based on knowledge of the experimental apparatus
- Both bias and noise models are typically available from instrument calibration
- Noise PDF structure
  - A counting instrument would exhibit Poisson noise
  - A measurement combining many noise sources would exhibit Gaussian noise
  - Noise correlation structure
    - Point measurement
    - Field measurement
- Error model composed of model error + data error

Computational Data

- Computational predictions from a high-fidelity model; presumed “true" — No data error
- Model error due to discrepancy between (simple) forward model prediction and truth
- Not statistical (for deterministic forward models)
  - Yet modeled statistically
- Where/How to incorporate this error term with the model
  - On model output observables
    - Not consistent with *physical* models
    - Not useful for one out of multiple observables
  - In a sub-model — Constitutive law, closure relation
  - In an existing model parameter

Parameter Estimation in the Absence of Data

- Frequently:
  - we know summary statistics about parameters from previous fitting
  - the raw data used to arrive at these statistics is not available
- How can we construct a joint PDF on the parameters?
- In the absence of data, the structure of the fit model, combined with the summary statistics, implicitly inform the joint PDF on the parameters
- Goal: Make available information *explicit* in the joint PDF
- Data Free Inference (DFI):
  - Discover a consensus joint PDF on the parameters consistent with given information in the absence of data

DFI Algorithm Structure (Berry *et al.*, JCP)

Basic idea:

- Explore the space of hypothetical data sets
  - MCMC chain on the data
  - Each state defines a data set
- For each data set:
  - MCMC chain on the parameters
  - Evaluate statistics on resulting posterior
  - Accept data set if posterior is consistent with given information
- Evaluate pooled posterior from all acceptable posteriors
  Logarithmic pooling:

$$p(\lambda|y) = \left[\prod_{i=1}^{K} p(\lambda|y_i)\right]^{1/K}$$

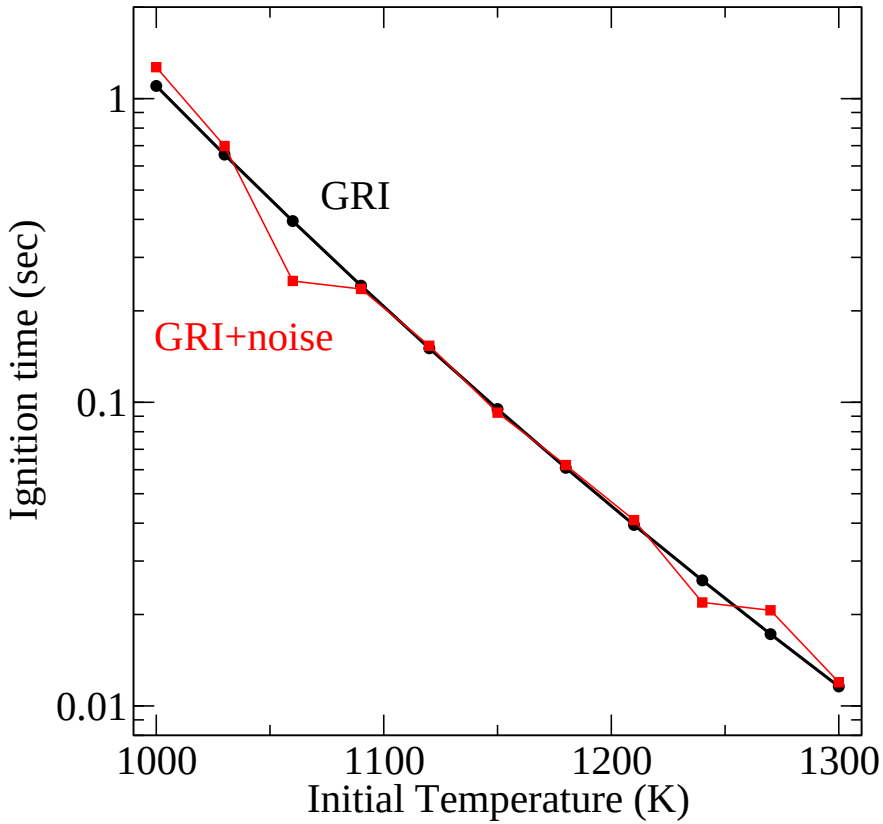
Example: Parameter Estimation in Chemical Systems

- Forward UQ requires the joint PDF on the input space
  - Published data is frequently inadequate
- Bayesian inference can provide the joint PDF
  - Requires raw data ... which is not available
- At best: 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

Generate ignition "data" using a detailed model+noise

- Ignition using a detailed chemical model for methane-air chemistry
- Ignition time versus Initial Temperature
- Multiplicative noise error model
- 11 data points:

$$
\begin{aligned}
d_i &= t_{\mathrm{ig},i}^{\mathrm{GRI}}(1+\sigma\epsilon_i) \\
\epsilon &\sim N(0,1)
\end{aligned}
$$



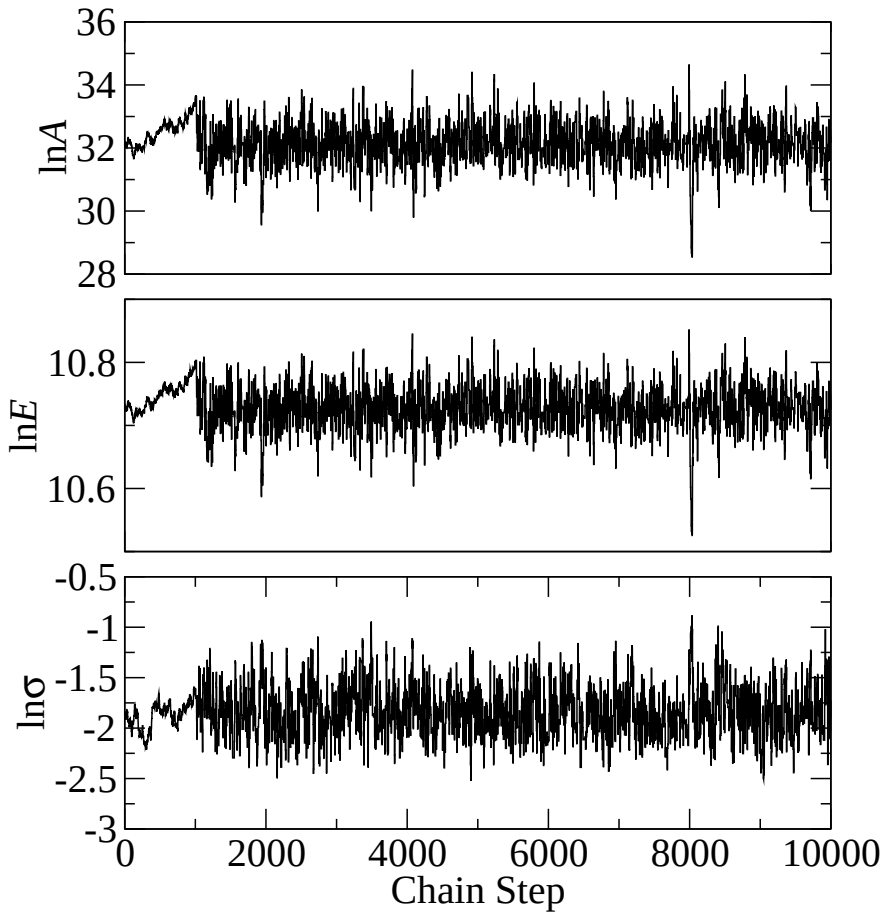
Fitting with a simple chemical model

- Fit a global single-step irreversible chemical model

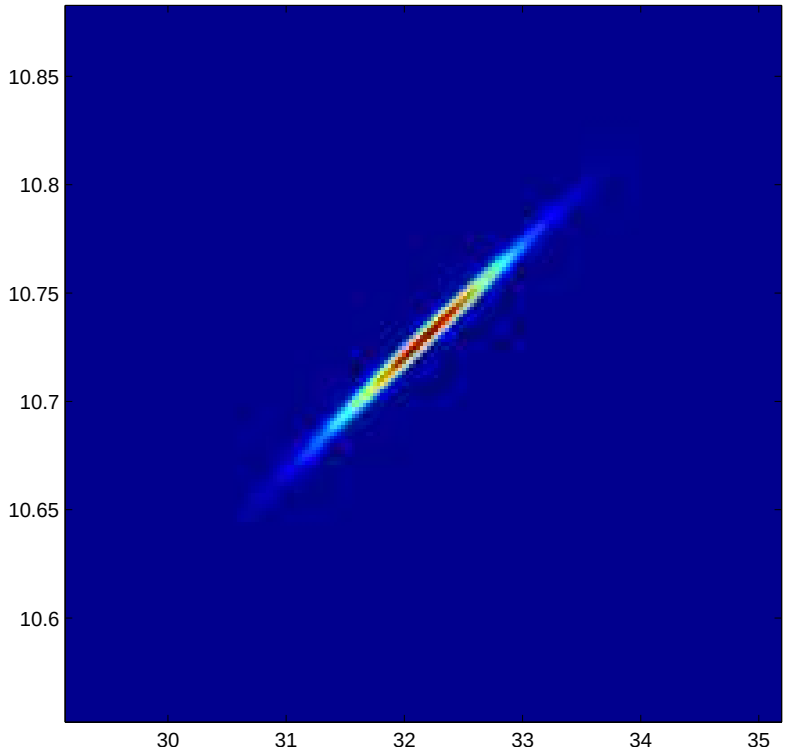
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

$$\begin{aligned} \Re &= [CH_4][O_2]k_f \\ k_f &= A\exp(-E/R^oT) \end{aligned}$$

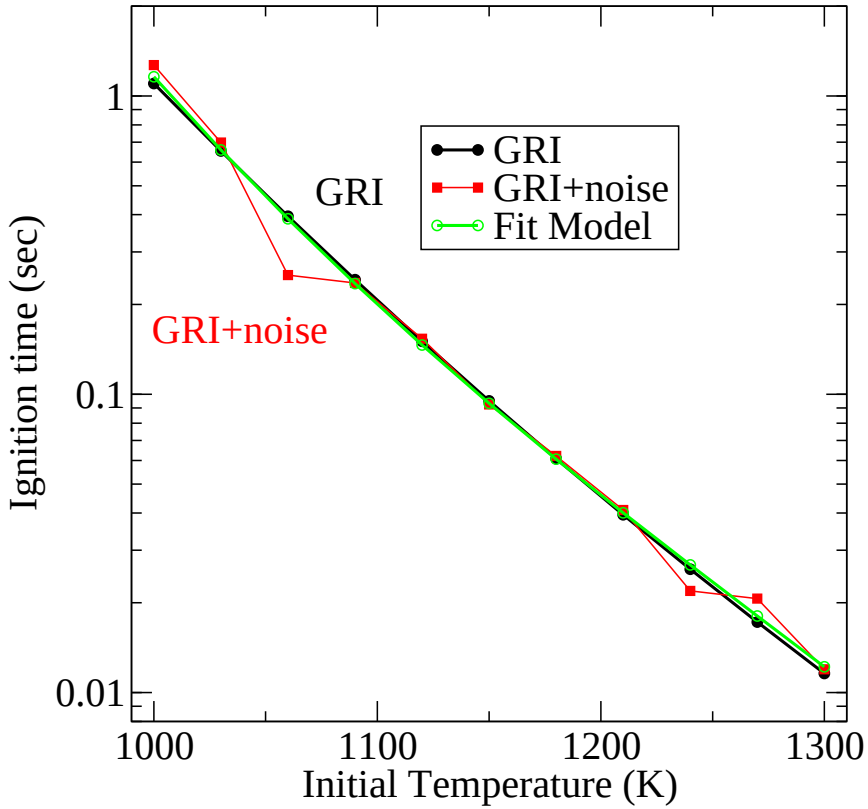
- Infer 3-D parameter vector $(\ln A, \ln E, \ln \sigma)$
- Good mixing with adaptive MCMC when start at MLE



Bayesian Inference Posterior and Nominal Prediction



Marginal joint posterior on $(\ln A, \ln E)$ exhibits strong correlation



Nominal fit model is consistent with the true model

Data Free Inference Challenge

Discarding initial data, reconstruct marginal $(\ln A, \ln E)$ posterior using the following information

- Form of fit model
- Range of initial temperature
- Nominal fit parameter values of $\ln A$ and $\ln E$
- Marginal 5% and 95% quantiles on $\ln A$ and $\ln E$

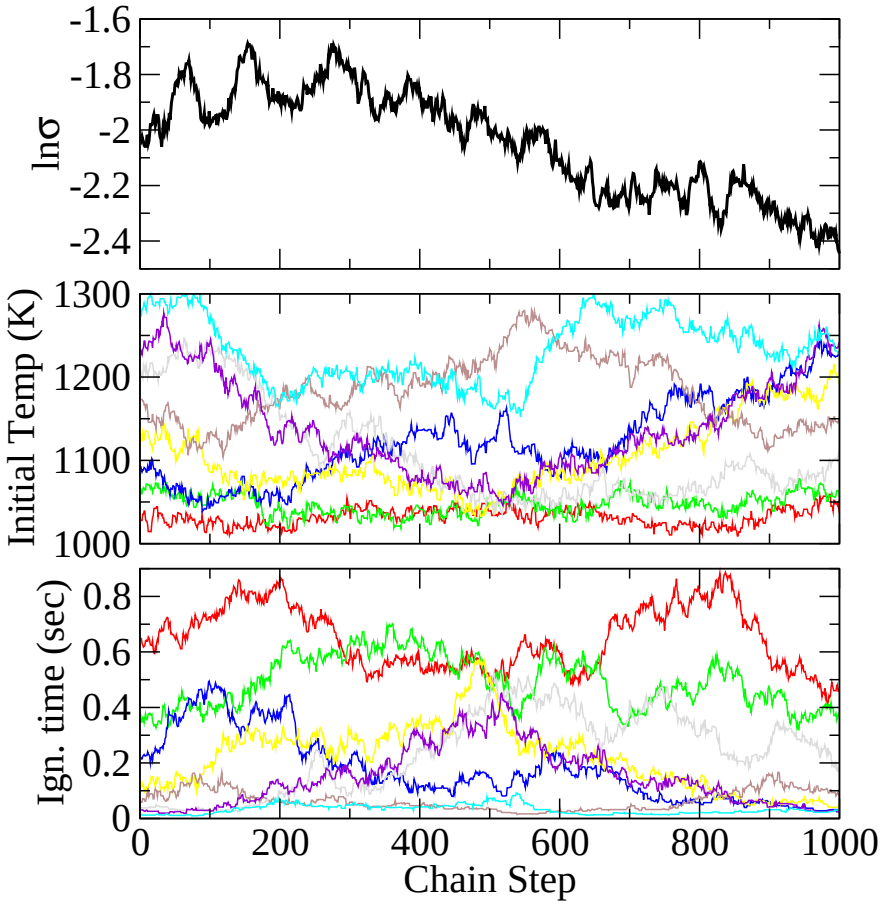
Further, for now, presume

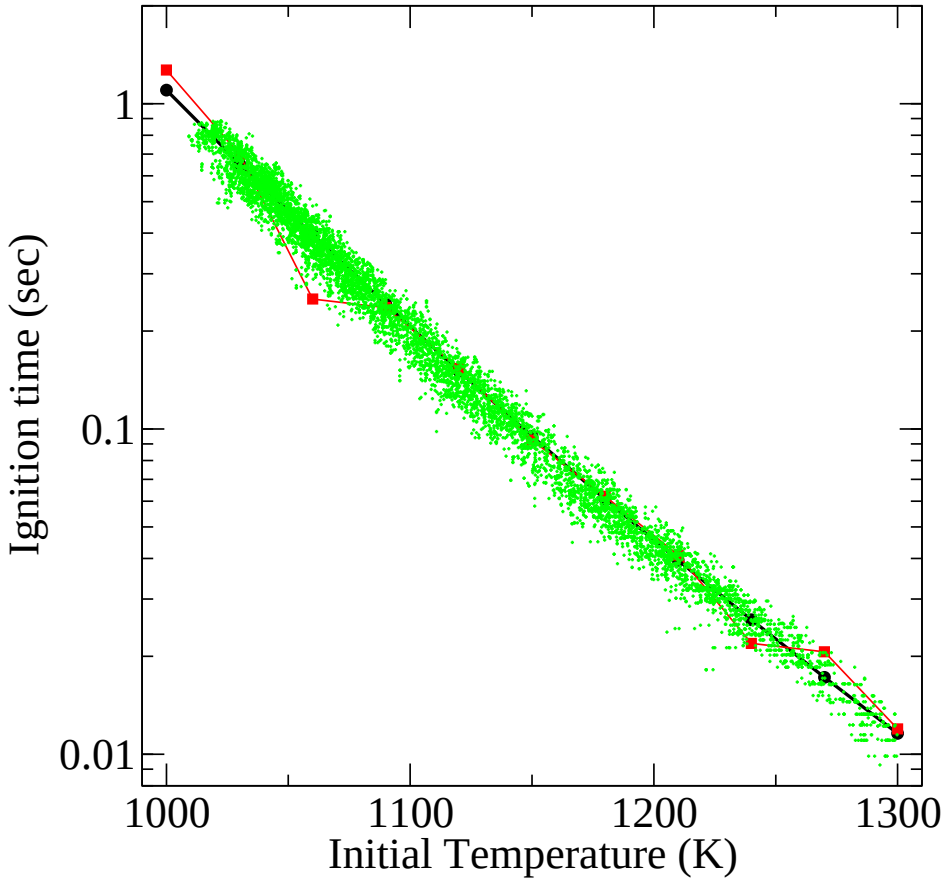
- Multiplicative Gaussian errors
- $N = 8$ data points

DFI Uses two nested MCMC chains

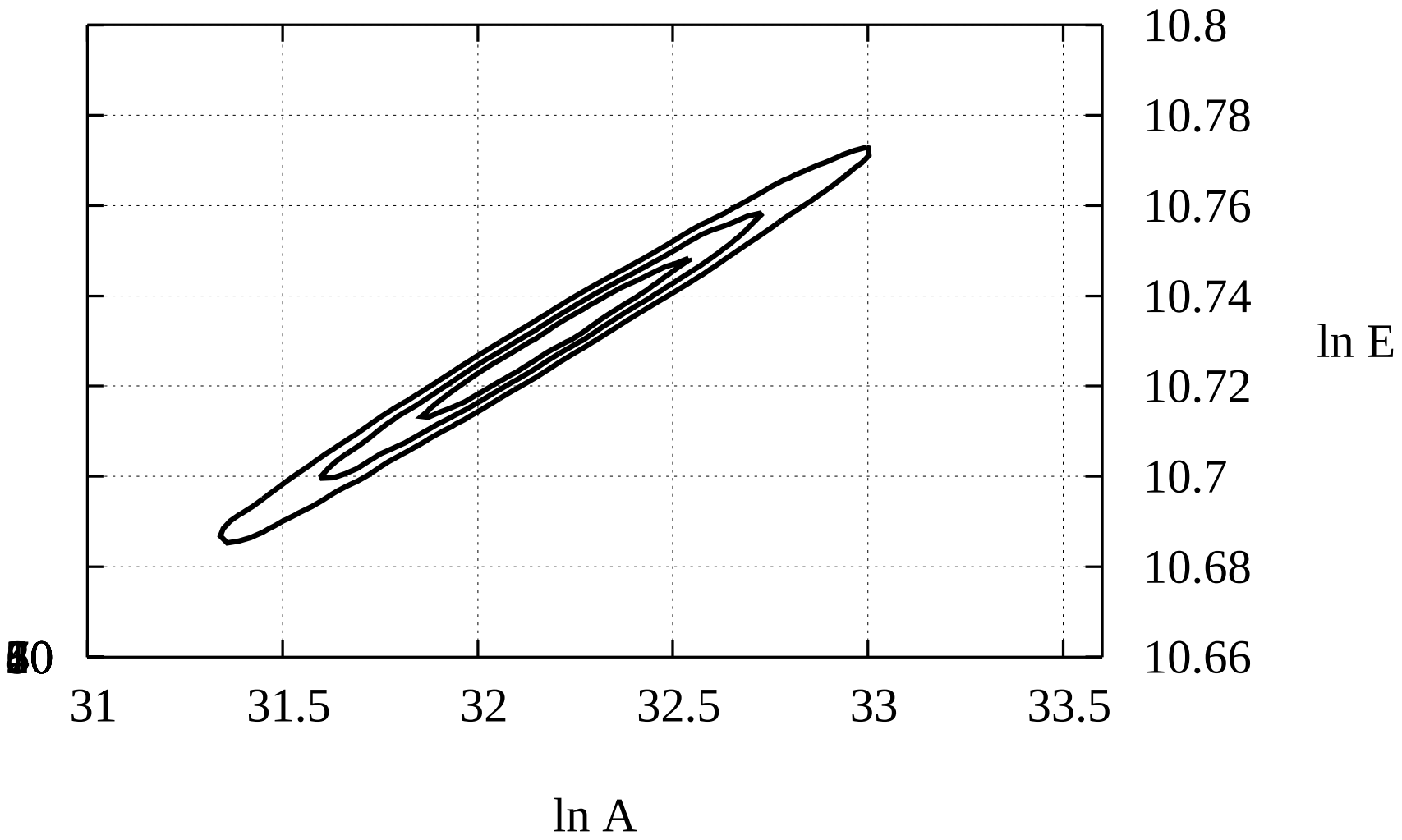
- An outer chain on the data, $(2N+1)$–dimensional
  - Generally high-dimensional
  - N data points (x_i, y_i) + σ
  - Likelihood function captures constraints on parameter nominals+bounds
- An inner chain on the model parameters
  - Conventional MCMC for parameter estimation
  - Likelihood based on fit-model
  - parameter vector $(\ln A, \ln E, \ln \sigma)$
- Computationally challenging
  - Single-site update on outer chain
  - Adaptive MCMC on inner chain
  - Run multiple outer chains in parallel, and aggregate resulting acceptable data sets

Short sample from outer/data chain

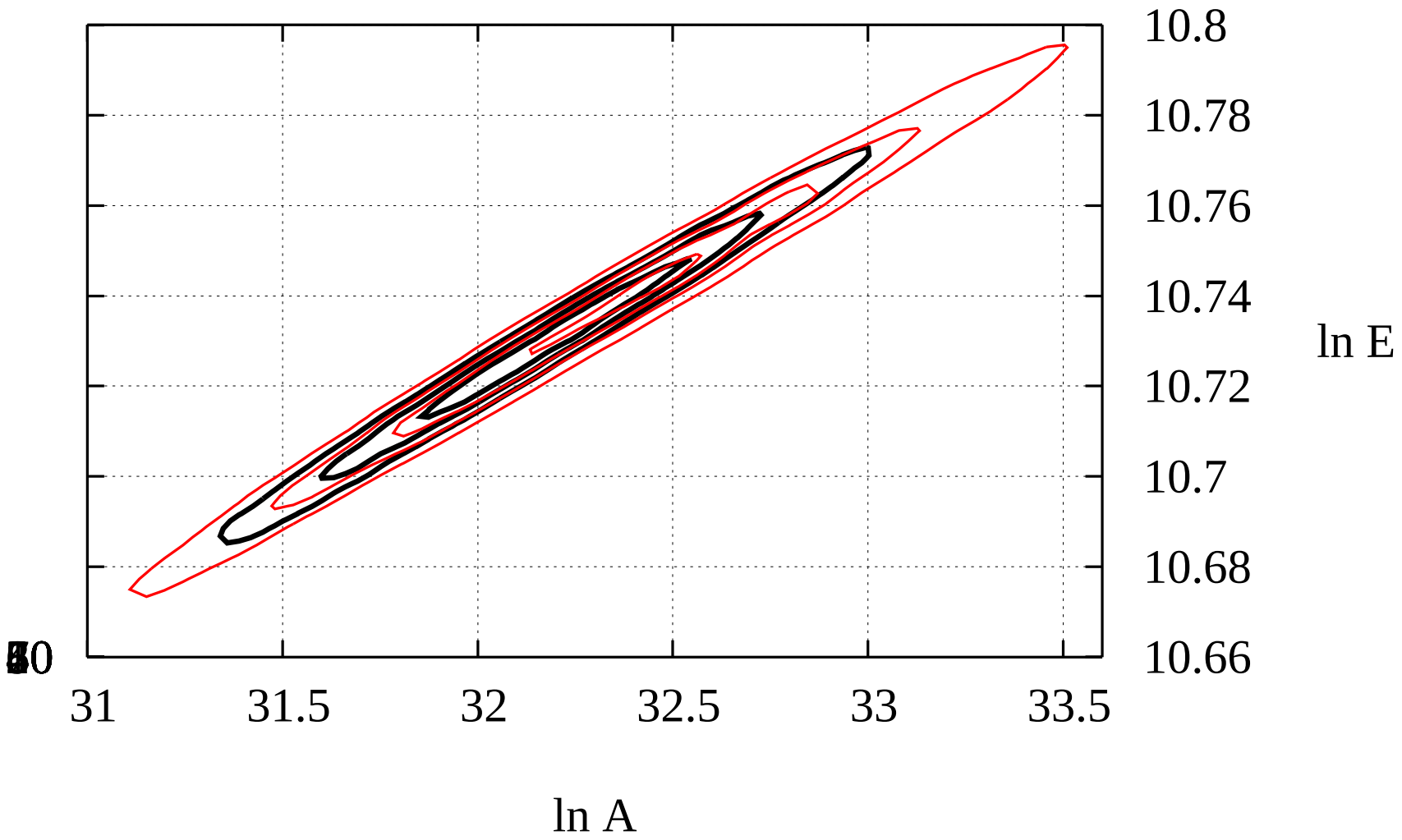




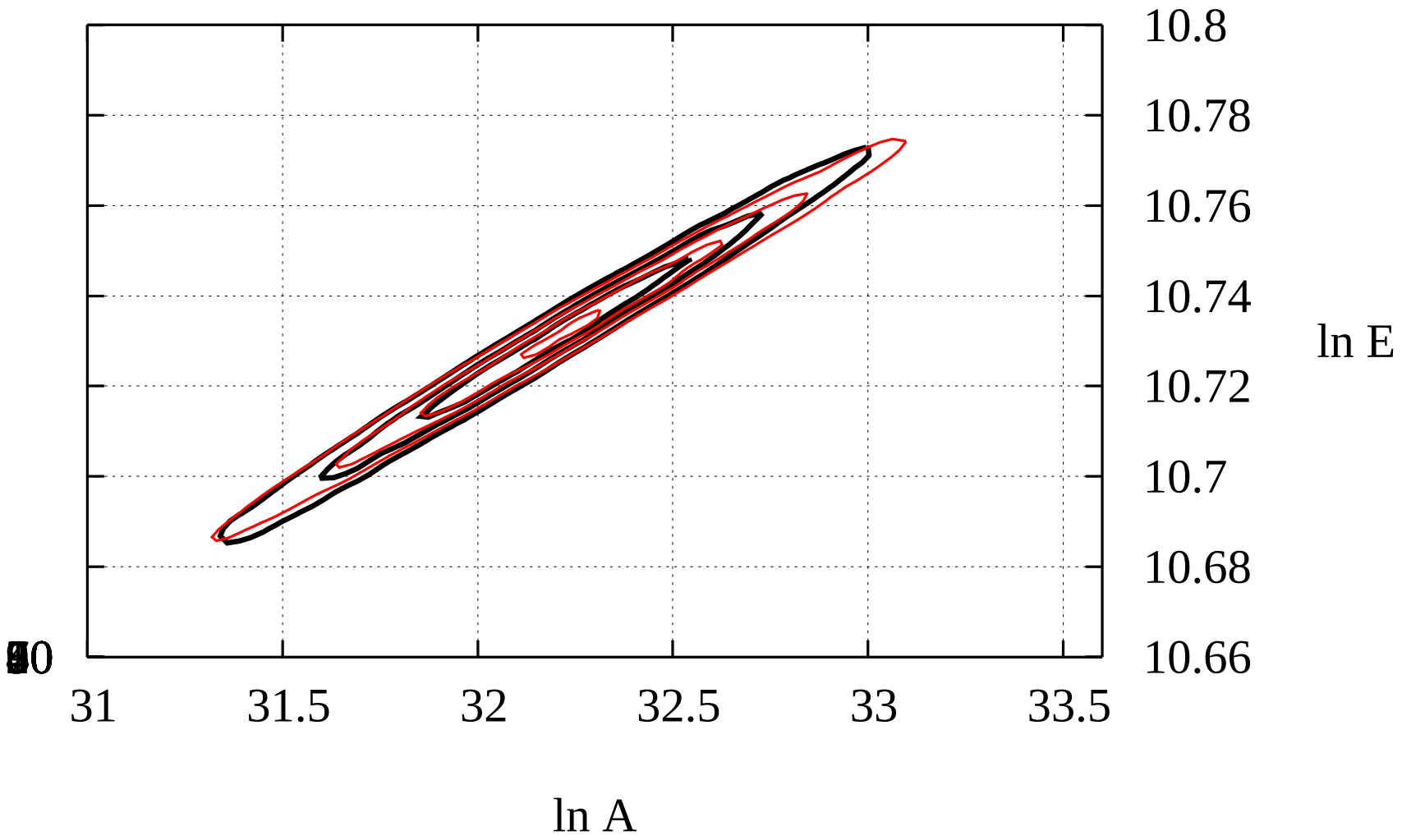
Reference Posterior – based on actual data



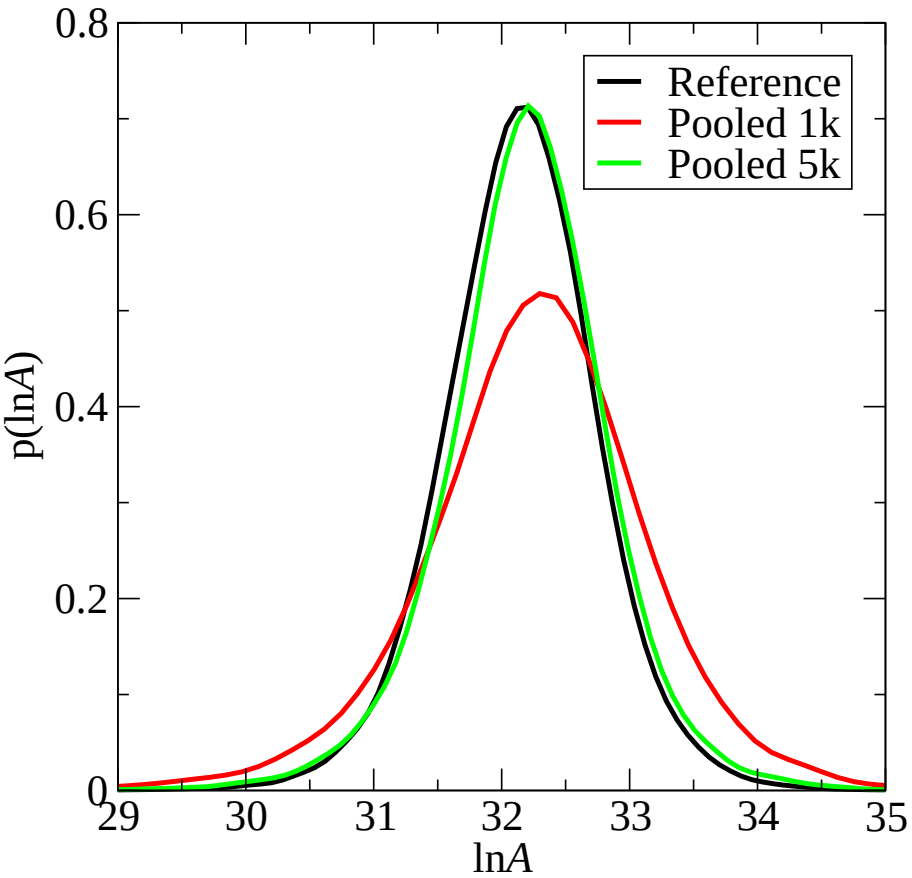
Ref + DFI posterior based on a 1000-long data chain

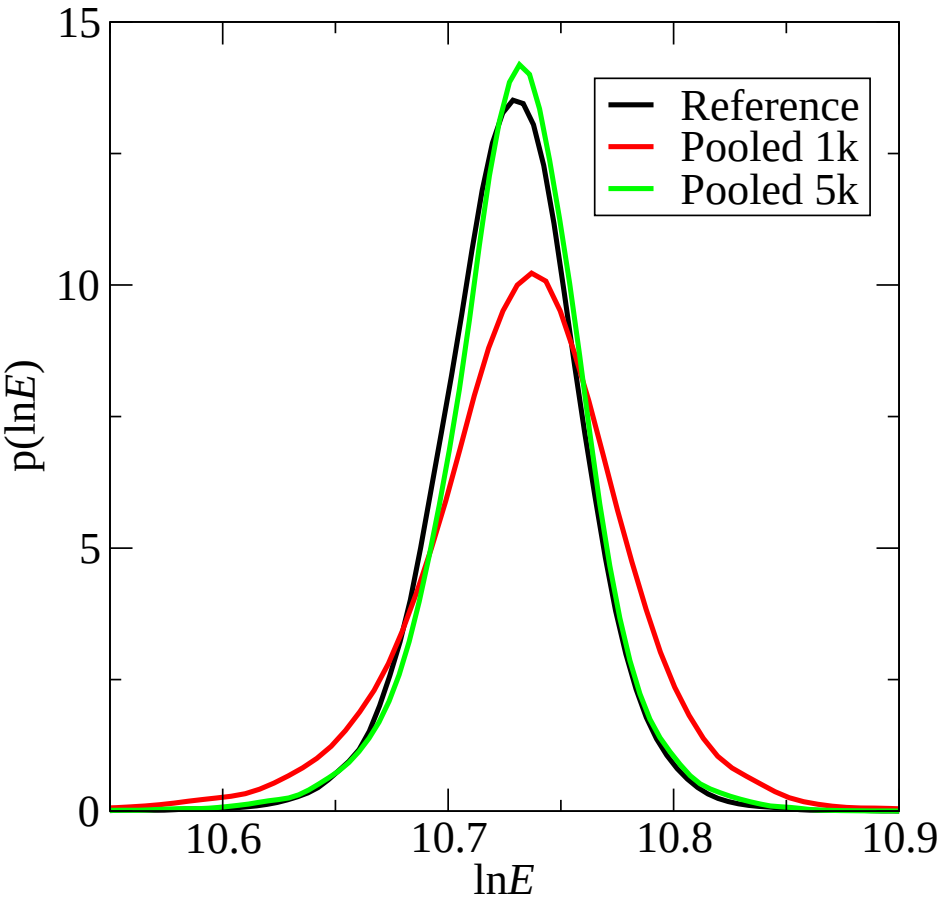


Ref + DFI posterior based on a 5000-long data chain



Marginal Pooled DFI Posteriors on $\ln A$ and $\ln E$





Probabilistic Forward UQ

- 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 most fundamentally as a function on a probability space
- Constraining the analysis to functions in L^2, *i.e.* to RVs with finite variance, enables the representation of a RV as a spectral expansion in terms of orthogonal functions of standard RVs.
  - Polynomial Chaos
- Enables the use of available functional analysis methods for forward UQ

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)

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

  - $u_k(\boldsymbol{x}, t)$ 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* $\boldsymbol{\xi}$

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

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

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
    - Lasso

Challenges in Forward 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 Forward 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

Challenges in Forward 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

Model UQ

- No model of a physical system is strictly true
- The probability of a model being strictly true is zero
- Given limited information, some models may be relied upon for describing the system

Let $\mathcal{M} = \{M_1, M_2, \ldots\}$ be the set of all models

- $p(M_k|I)$ is the probability that M_k is the model behind the available information
  - Model Plausibility
- Parameter estimation from data is conditioned on the model

$$p(\theta|D, M_k) = \frac{p(D|\theta, M_k)\pi(\theta|M_k)}{p(D|M_k)}$$

Bayesian Model Comparison

Evidence (marginal likelihood) for M_k:

$$p(D|M_k) = \int p(D|\theta, M_k)\pi(\theta|M_k)\mathrm{d}\theta$$

Bayes Factor B_{ij}:

$$B_{ij} = \frac{p(D|M_i)}{p(D|M_j)}$$

Plausibility of M_k:

$$p(M_k|D, \mathcal{M}) = \frac{p(D|M_k)\ \pi(M_k|\mathcal{M})}{\sum_s p(D|M_s)\pi(M_s|\mathcal{M})} \qquad k = 1, \ldots$$

Posterior odds:

$$\frac{p(M_i|D, \mathcal{M})}{p(M_j|D, \mathcal{M})} = B_{ij}\ \frac{\pi(M_i|\mathcal{M})}{\pi(M_j|\mathcal{M})}$$

Validation

- Validity is a statement of model utility for predicting a given observable under given conditions
- Inspection of model utility requires accounting for uncertainty
- Statistical tool-chest for model validation
  - Cross-validation
  - Bayes Factor
  - Model Plausibility
  - Posterior Odds
  - Posterior predictive:

$$p(\tilde{D}|D, M_k) = \int p(\tilde{D}|\theta, M_k) p(\theta|D, M_k) d\theta$$

Closure

- UQ is increasingly important in computational modeling
- Probabilistic UQ framework
- Bayesian parameter estimation, model calibration
  - Data sources
  - Error modeling
  - Missing/absent data
- Forward PC UQ
  - Representation
  - Propagation
- Model uncertainty
  - Model comparison