SAND2011-3749C

Uncertainty Quantification in Reacting Flow

Habib N. Najm

Combustion Research Facility
Sandia National Laboratories, Livermore, CA

Stanford UQ Seminar
May 26, 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:

- 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 program.
- 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

1. Basics
2. Validation
3. Estimation of Uncertain Inputs
4. Challenges in Forward PC UQ
5. Closure

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 versus Frequentist
  - 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
  - Expert opinion
  - Maximum Entropy
  - Bayes formula

Validation

- No model is “true"
- 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
  - Calibration based on a data subset and analysis of fit to its complement
  - Model comparison – Bayes Factors, Model Plausibility
  - Posterior predictive

Bayes formula for Parameter Inference

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

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

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

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

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 $U = [u_0,\ldots,u_P]^T$, $\Lambda = [\lambda_0,\ldots,\lambda_P]^T$
- Solving this system *once* provides the full specification of uncertain model ouputs

Non-intrusive Spectral Projection (NISP) PC UQ

- *Sampling*-based
- Relies on black-box utilization of the computational model
- Evaluate projection integrals *numerically*
- For any model output of interest $\phi(\boldsymbol{x}, t; \lambda)$:

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

- Integrals can be evaluated using
  - A variety of (Quasi) Monte Carlo methods
  - Quadrature/Sparse-Quadrature methods

PC Surrogate

$$\lambda(\boldsymbol{\xi}) = \sum_{k=0}^{P} \lambda_k \Psi_k(\boldsymbol{\xi})$$

$$u = f(\lambda(\boldsymbol{\xi})) \qquad u = \sum_{k=0}^{P} u_k \Psi_k(\boldsymbol{\xi})$$

- Surrogate can be constructed with any presumed $p_\lambda(\lambda)$
  - Convenient linear option: $\lambda = \lambda_0 + \lambda_1 \xi$
- PDF(ξ) controls local accuracy of the surrogate over λ
  - A uniform ξ implies uniform weighting of the error residual over λ
- Any forward-UQ method of choice can be used to construct the surrogate

PC Surrogate Construction for Noisy Functions

- Quadrature formulae presume a degree of smoothness
  - No convergence for a noisy function

$$u_k = \frac{1}{\langle \Psi_k^2 \rangle} \int u(\lambda(\xi))\, \Psi_k(\xi) p_{\xi}(\xi) d\xi, \quad k = 0, \ldots, P$$

- Sparse-Quadrature formulae are *ill-conditioned* and highly-sensitive to noise
  - No convergence with order
  - Error grows with increased dimensionality
- Options in the presence of noise:
  - RMS fitting for PC coefficients
  - Bayesian inference of PC coefficients

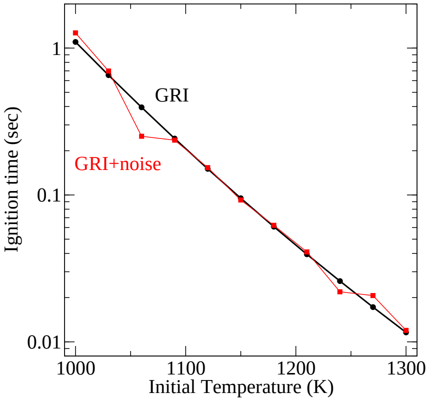
Parameter Estimation in Chemical Systems

- 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:

$$d_i = t_{\text{ig},i}^{\text{GRI}}(1 + \sigma\epsilon_i)$$
$$\epsilon \sim N(0, 1)$$



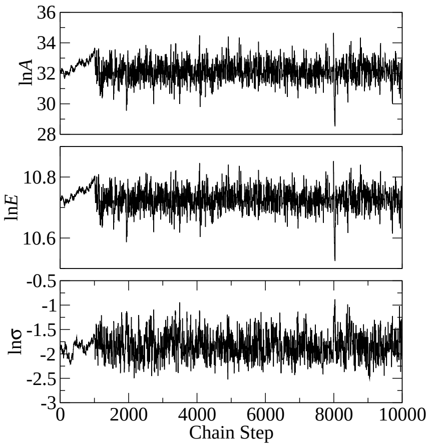
Fitting with a simple chemical model

- Fit a global single-step irreversible chemical model

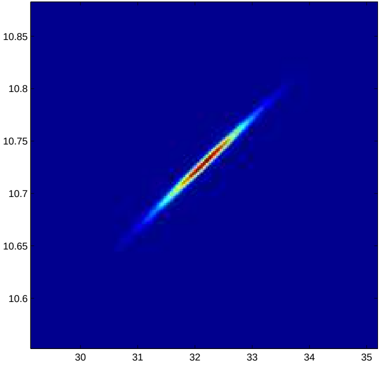
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

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

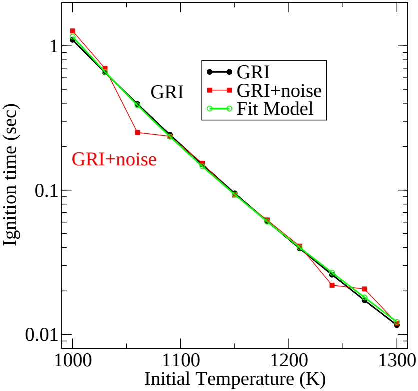
- Infer 3-D parameter vector $(\ln A, \ln E, \ln \sigma)$
- Good mixing with adaptive MCMC when start at MLE



Bayesian Inference Posterior and Nominal Prediction

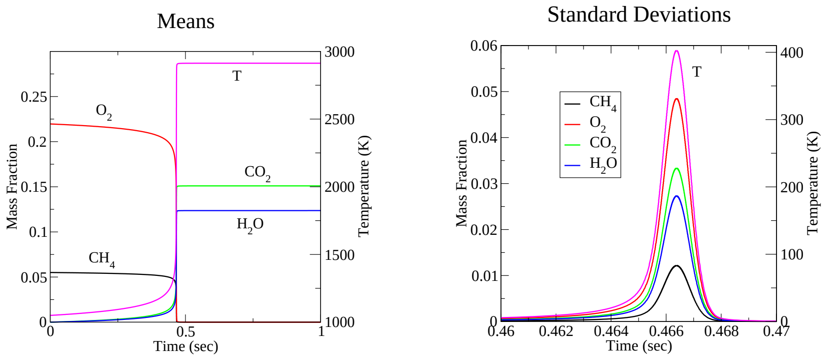


Marginal joint posterior on $(\ln A, \ln E)$ exhibits strong correlation



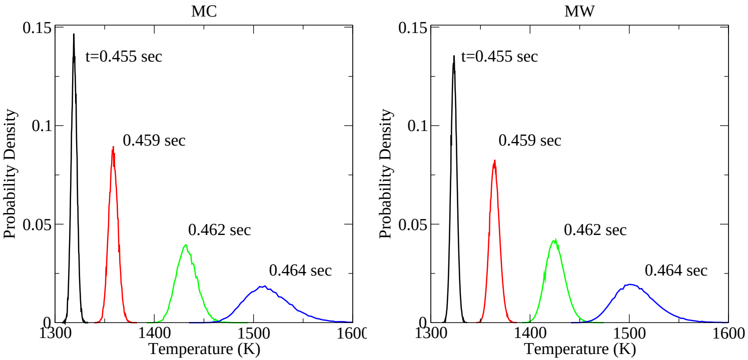
Nominal fit model is consistent with the true model

Correlation Slope χ and Chemical Ignition



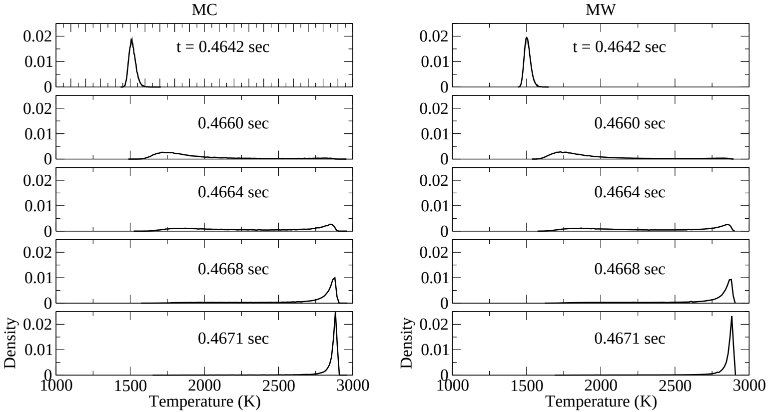
- 4^{th} Order Multiwavelet PC, Multiblock, Adaptive
- $\sigma_{T,\max} \sim$ 400 K during ignition transient, $\chi \sim 0.03$

Time evolution of Temperature PDFs in preheat stage



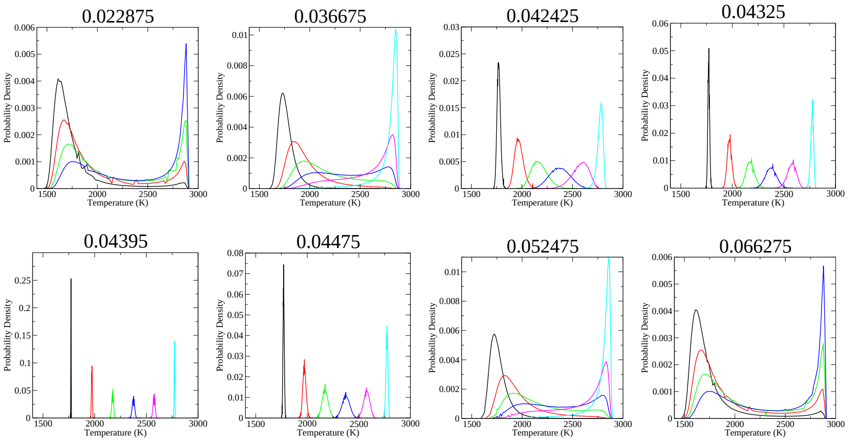
- Similar results from MC (20K samples) and MW PC
- Increased uncertainty, and long high–T PDF tails, in time

Evolution of Temp. PDF – Fast Ignition Transient



- Transition from unimodal to bimodal PDFs
- Leakage of probability mass from pre-heat PDF high–T tail

Time evolution of Temperature PDFs for different χ



- Bimodal solution PDFs for high uncertainty growth
- Unimodal for low uncertainty growth, with $\chi \approx 0.044$

Central Challenge for UQ in Chemical Kinetic Models

- Need joint PDF on model parameters for forward UQ
- Joint PDF structure is crucial
- Joint PDF not available for chemical kinetic parameters
- At best, have
  - Nominal parameter values
  - Bounds, *e.g.* marginal 5%, 95% quantiles
- PDF ***can*** be constructed by repeating experiments or access to original raw data
  - Neither is feasible
- Is there a way to construct an approximate PDF ***without*** access to raw data?
  - Yes!

Data Free Inference (DFI)

(Berry *et al.*, JCP, in review)

- Intuition: In the absence of data, the structure of the fit model, combined with the nominals and bounds, implicitly inform the correlation between the parameters
- Goal: Make this information *explicit* in the joint PDF
- DFI: discover a consensus joint PDF on the parameters consistent with given information:
  - Nominal parameter values
  - Bounds
  - The fit model
  - The data range
  - ... potentially other/different constraints

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 Algorithm Structure

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

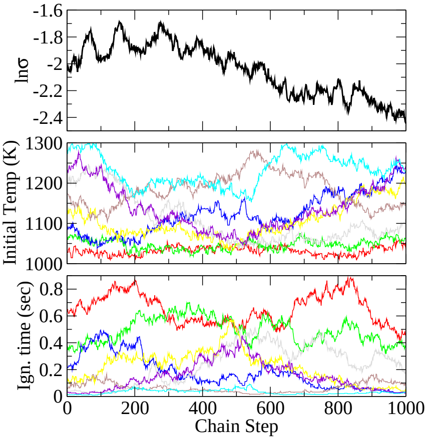
  Logarithmic pooling:

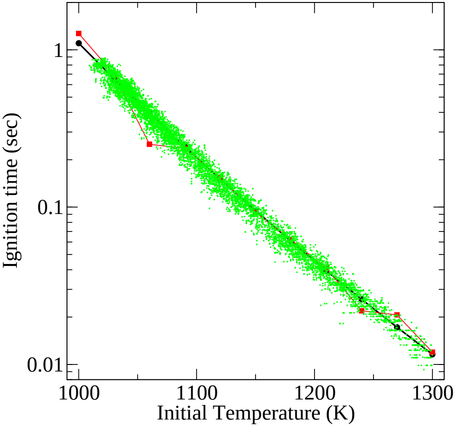
$$p(\lambda|y) = \left[\prod_{i=1}^{K} p(\lambda|y_i)\right]^{1/K}$$

DFI Uses two nested MCMC chains

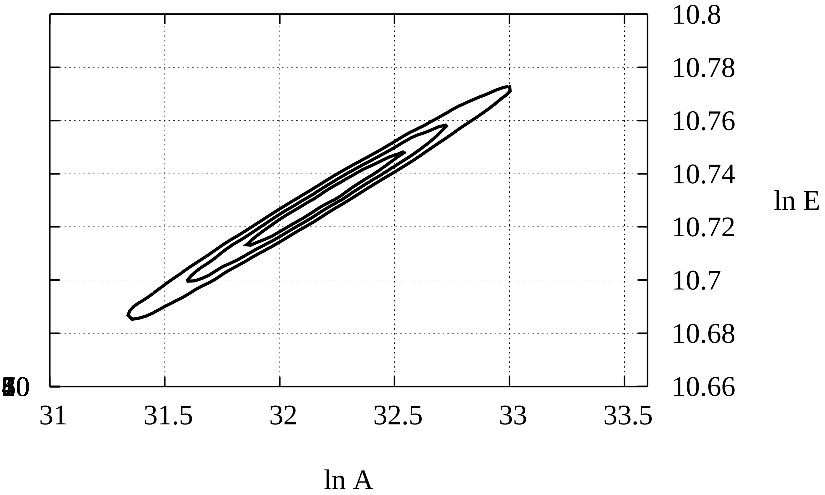
- An outer chain on the data, $(2N+1)$–dimensional
  - Generally high-dimensional
  - N data points (x_i, y_i) + σ
  - Likelihood function captures constraints on parameter nominals+bounds
- An inner chain on the model parameters
  - Conventional MCMC for parameter estimation
  - Likelihood based on 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

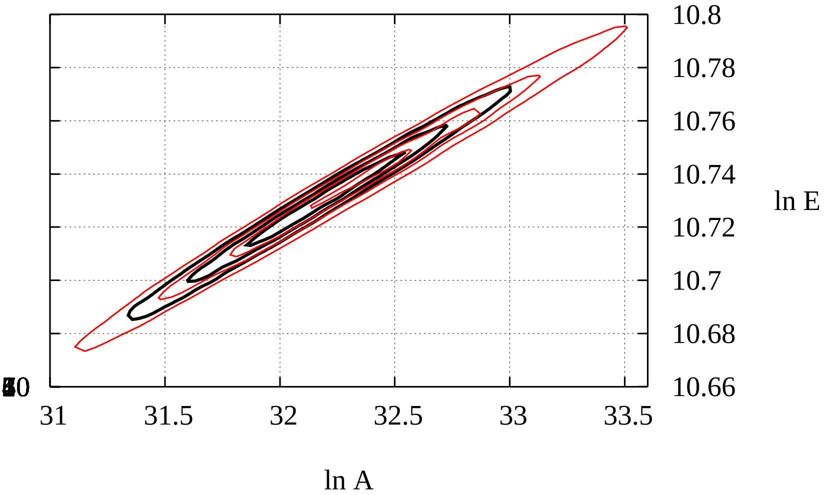




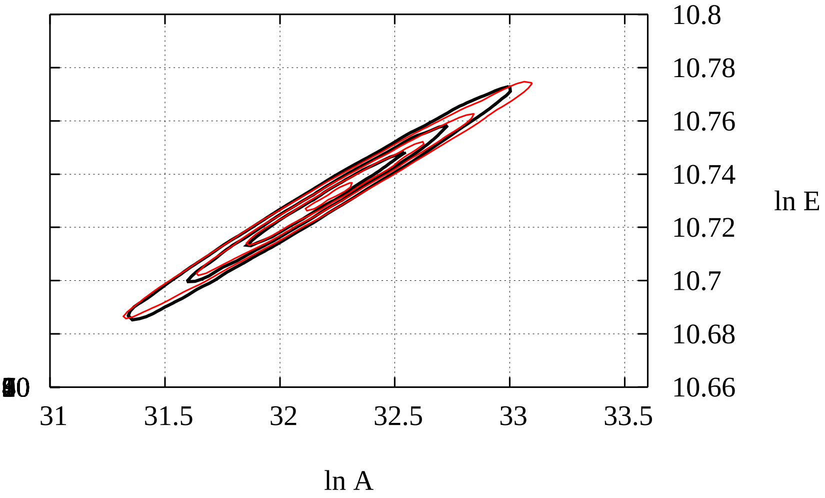
Reference Posterior – based on actual data



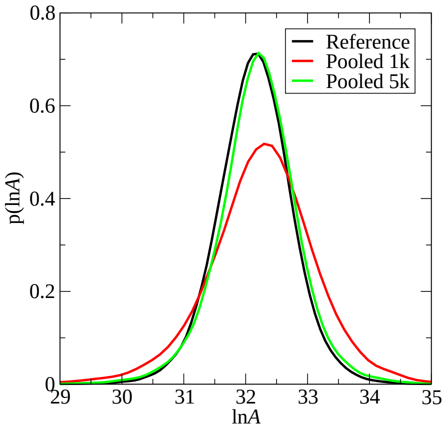
Ref + DFI posterior based on a 1000-long data chain

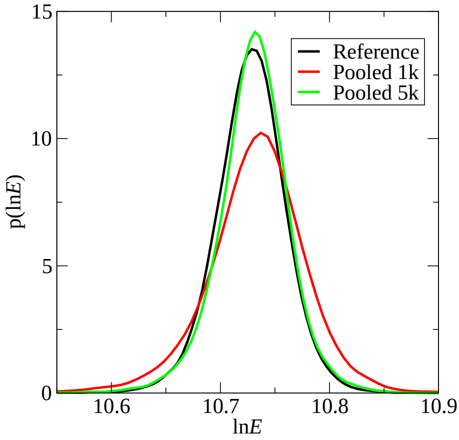


Ref + DFI posterior based on a 5000-long data chain



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





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, ...

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

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
- Remedies
  - Time shifting/scaling
  - Choose smooth observables
- 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
  - $\Rightarrow$ Use non-intrusive methods with DNS/LES of turbulence

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
  - Utility in inverse problems – surrogates
    - Bayesian inference
    - Model validation
- Need for probabilistic characterization of uncertain inputs
  - Correlations important for uncertainty in predictions
  - DFI $\Rightarrow$ joint PDF consistent with available information