SAND2011-5426C

Uncertainty Quantification in Reacting Flow

Habib N. Najm[1], Robert D. Berry[1], Bert J. Debusschere[1], and Youssef M. Marzouk[2]

[1]Sandia National Laboratories, Livermore, CA, USA
[2]Massachusetts Institute of Technology, Cambridge, MA, USA

USA/South America Symposium on Stochastic Modeling and Uncertainty Quantification
August 1-5 2011, Rio de Janeiro

Acknowledgement

K. Sargsyan, C. Safta
— Sandia National Laboratories, CA, USA
R.G. Ghanem — U. South. California, Los Angeles, CA, USA
O.M. Knio — Johns Hopkins Univ., Baltimore, MD, USA
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 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

Need for UQ in Reacting Flow Modeling

- Combustion: dominant means of utilization of fossil fuels
  - Power generation
  - Transportation
  - Industrial processing & residential use
- Chemical models involve much empiricism
  - Models: choice of species and reactions
  - Parameters:
    - Chemical rate constants
    - Thermodynamic parameters
- Flow models rely on empiricism and approximations
  - mass/energy transport and fluid constitutive laws
  - turbulence/subgrid models
- Focus on uncertainty in chemical model parameters

Challenges with UQ in reacting flow

- Estimation of uncertainty in parameters (and models)
  - Published data is inadequate
  - Raw data is not available
  - Data on correlations among parameters is not available
    {forward, backward rate constants, thermo. props}
  - Ongoing community efforts to address this:
    $\Rightarrow$ PRIME, Active tables
- Non-linearity
  - Amplif. of uncertainty. Bifurcation
  - oscillatory dynamics
- Stiffness
  - Large range of time scales. Low dimensional manifolds
- High dimensionality

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
- Any quantity can be estimated
  - Expert opinion
  - Maximum Entropy
  - Bayes formula

Bayes formula for Parameter Inference

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

$$p(x,y) = p(x|y)p(y) = p(y|x)p(x)$$

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

- Prior: knowledge of x 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 x, the un-normalized posterior probability can be easily computed

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

- 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 x
- Gaussian Process (GP) surrogate
  - GP goes through all data points with probability 1.0
  - Uncertainty between the points
- Fit a convenient polynomial to $y = f(x)$
  - over the range of uncertainty in x

  - Employ a number of samples (x_i, y_i)
  - Fit with interpolants, regression, ... global/local
  - With uncertain x :
    - Construct Polynomial Chaos response surface

Marzouk *et al.* 2007; Marzouk & Najm, 2009

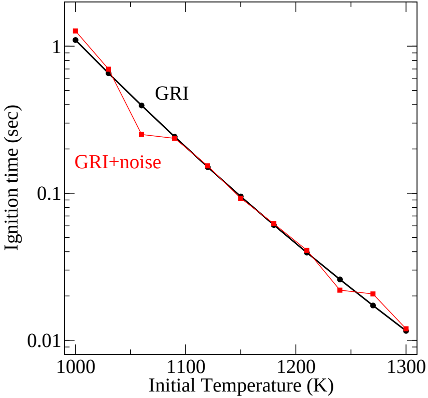
Parameter Estimation in Chemical Systems

- 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:

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



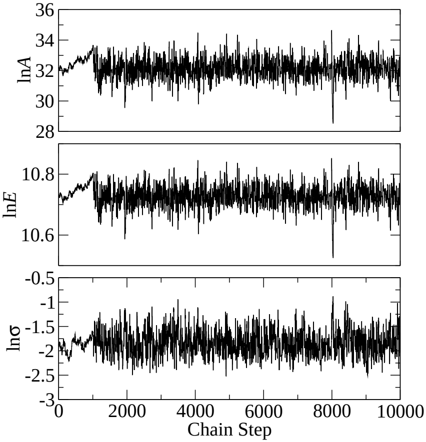
Fitting with a simple chemical model

- Fit a global single-step irreversible chemical model

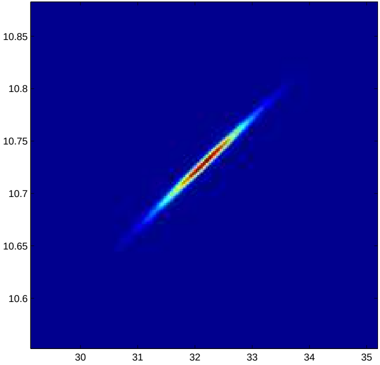
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

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

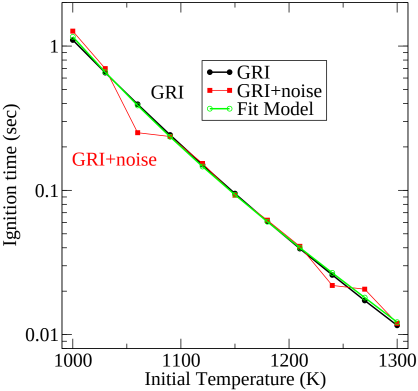
- Infer 3-D parameter vector $(\ln A, \ln E, \ln \sigma)$
- Good mixing with adaptive MCMC when start at MLE



Bayesian Inference Posterior and Nominal Prediction

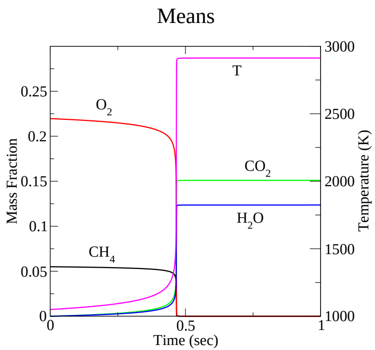


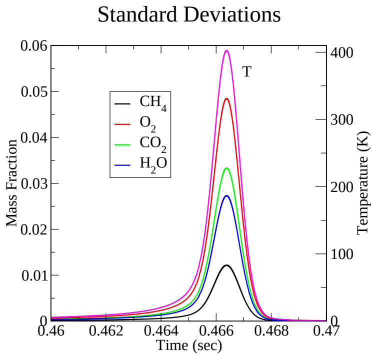
Marginal joint posterior on $(\ln A, \ln E)$ exhibits strong correlation



Nominal fit model is consistent with the true model

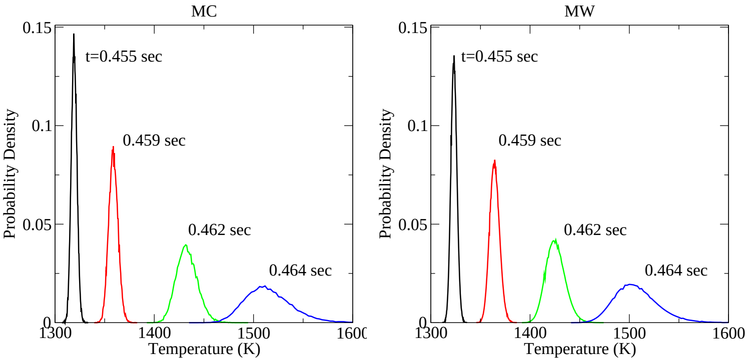
Correlation Slope χ and Chemical Ignition





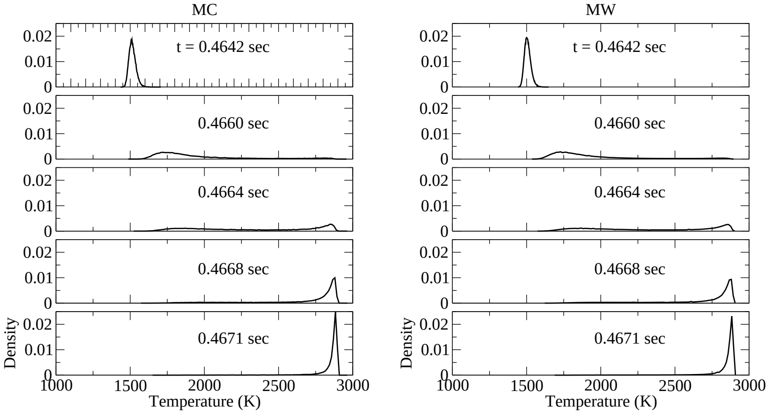
- 4^{th} Order Multiwavelet PC, Multiblock, Adaptive
- $\sigma_{T,\max} \sim$ 400 K during ignition transient, $\chi \sim 0.03$

Time evolution of Temperature PDFs in preheat stage



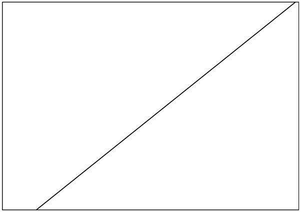
- Similar results from MC (20K samples) and MW PC
- Increased uncertainty, and long high–T PDF tails, in time

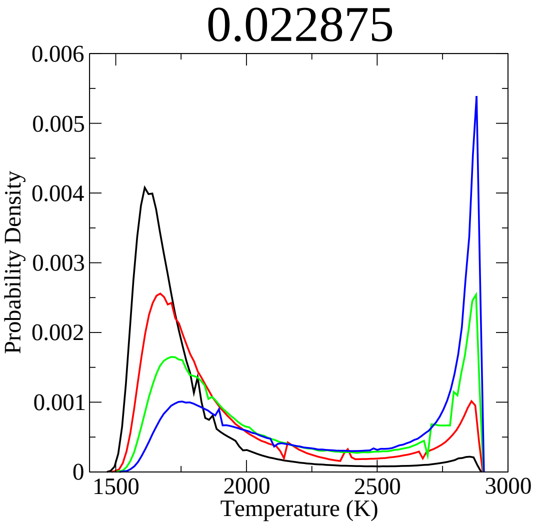
Evolution of Temp. PDF – Fast Ignition Transient



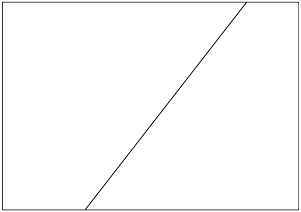
- Transition from unimodal to bimodal PDFs
- Leakage of probability mass from pre-heat PDF high–T tail

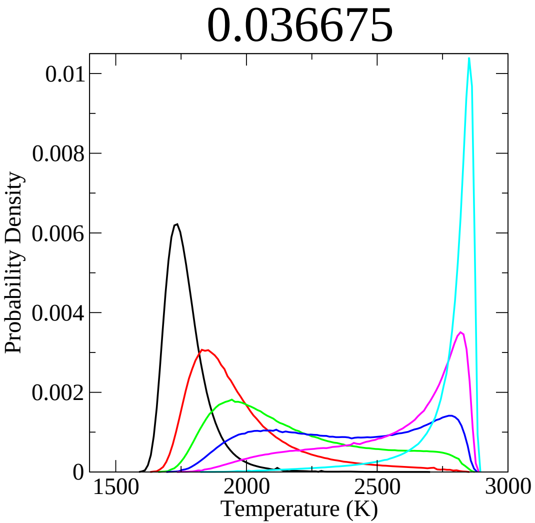
Time evolution of Temperature PDFs f(slope)



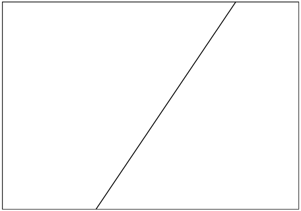


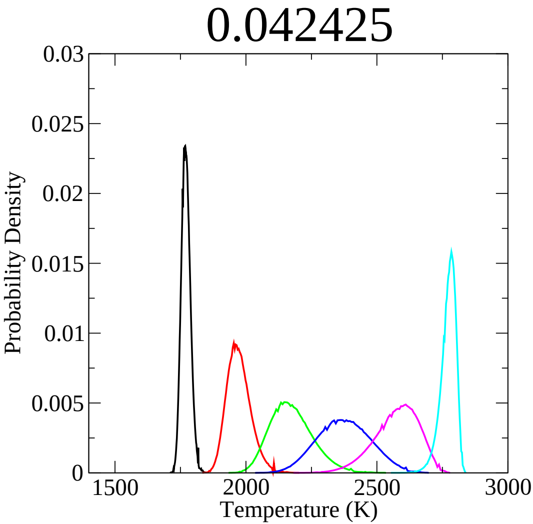
Time evolution of Temperature PDFs f(slope)



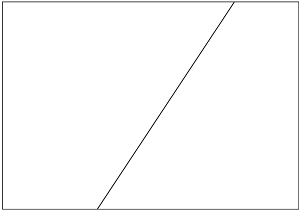


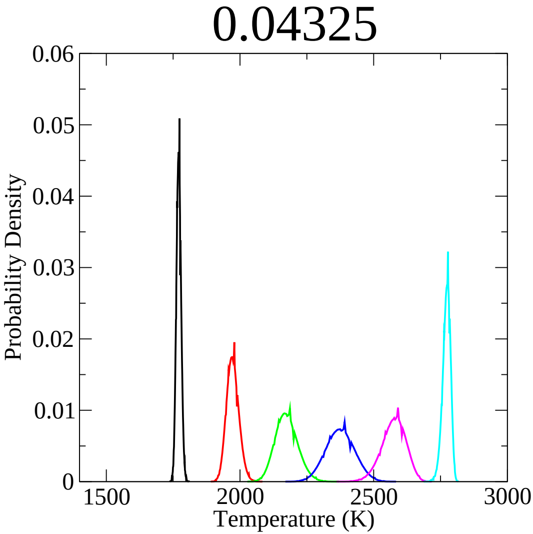
Time evolution of Temperature PDFs f(slope)



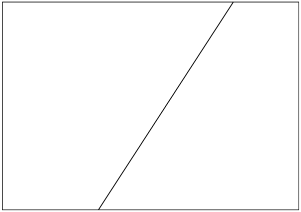


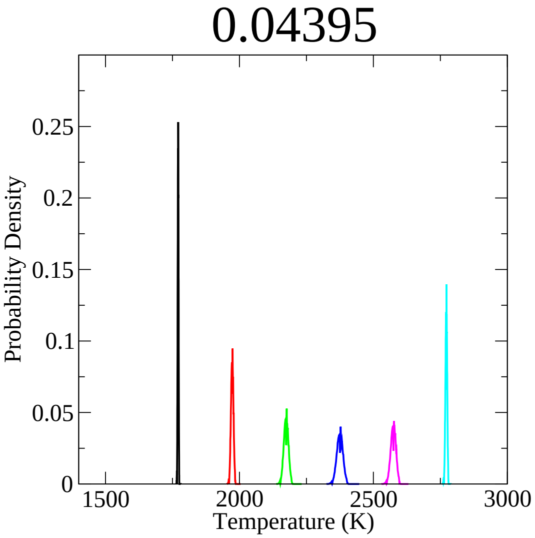
Time evolution of Temperature PDFs f(slope)



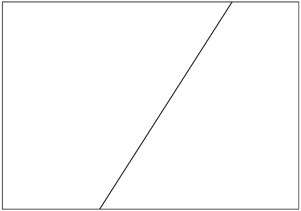


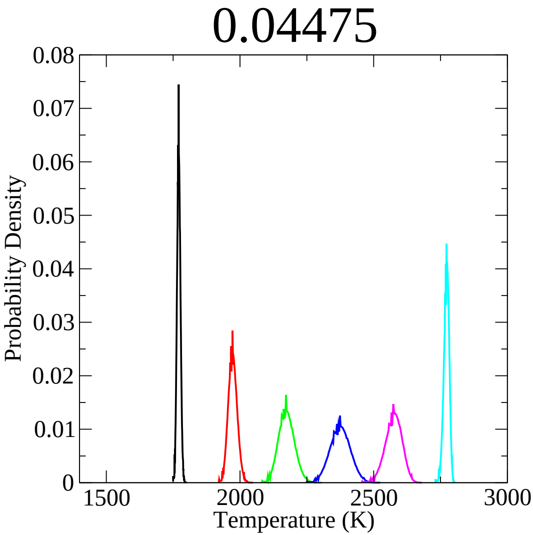
Time evolution of Temperature PDFs f(slope)



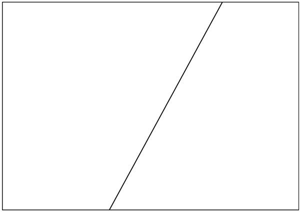


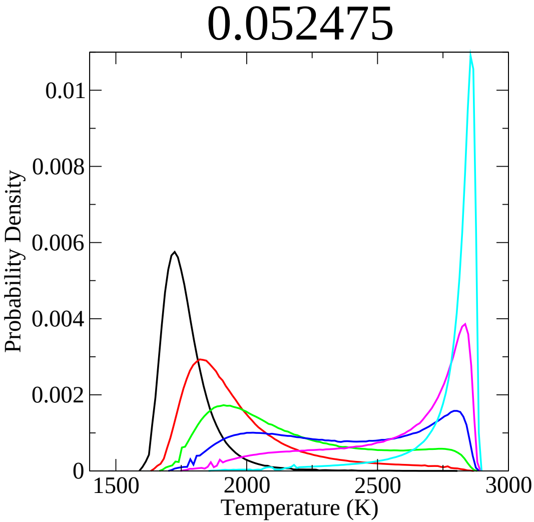
Time evolution of Temperature PDFs f(slope)



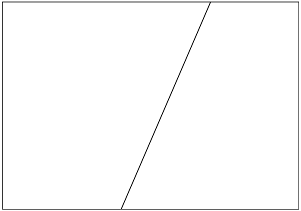


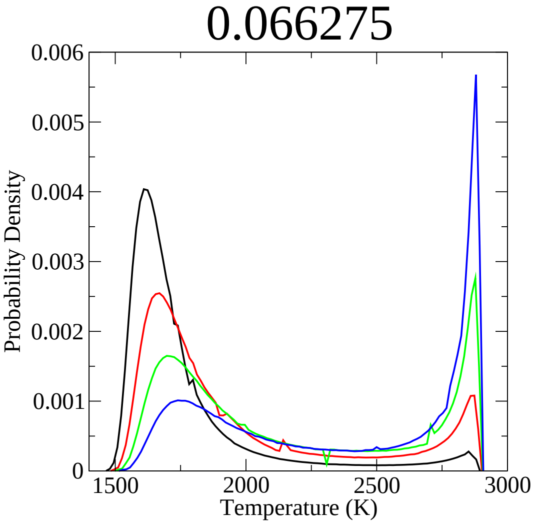
Time evolution of Temperature PDFs f(slope)





Time evolution of Temperature PDFs f(slope)





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

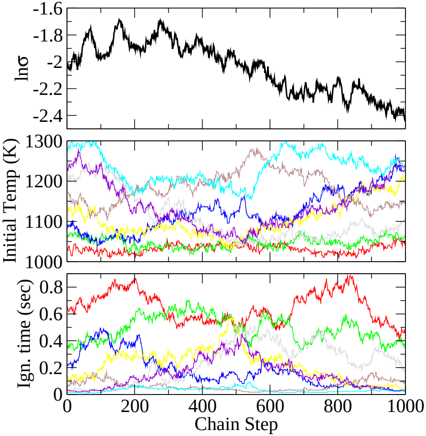
  Logarithmic pooling:

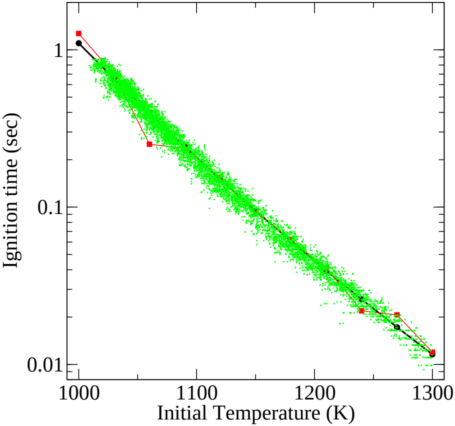
$$p(\lambda|y) = \left[\prod_{i=1}^{K} p(\lambda|y_i)\right]^{1/K}$$

DFI Uses two nested MCMC chains

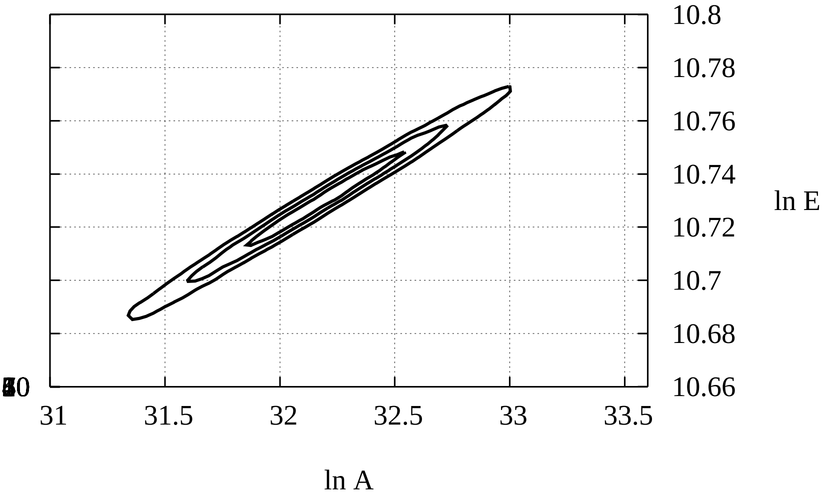
- An outer chain on the data, $(2N+1)$–dimensional
  - Generally high-dimensional
  - N data points (x_i, y_i) + σ
  - Likelihood function captures constraints on parameter nominals+bounds
- An inner chain on the model parameters
  - Conventional MCMC for parameter estimation
  - Likelihood based on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

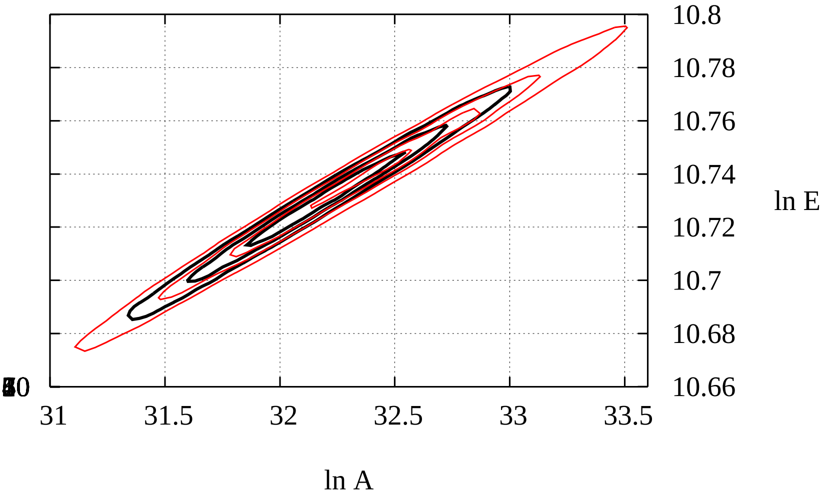




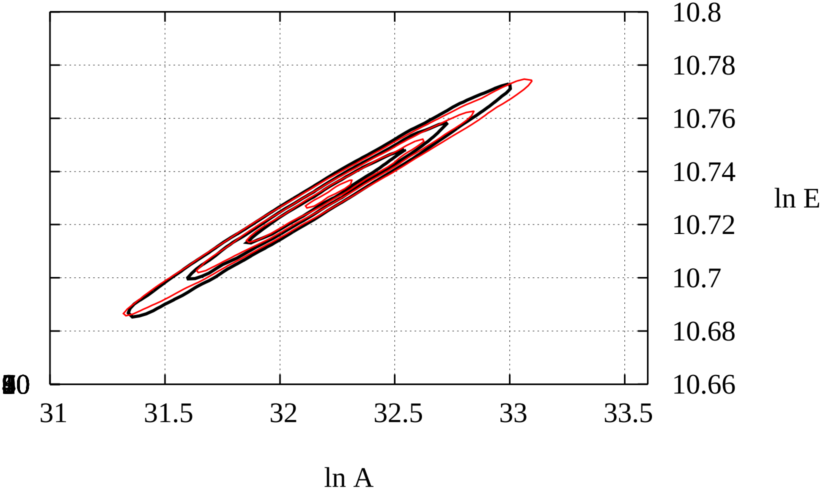
Reference Posterior – based on actual data



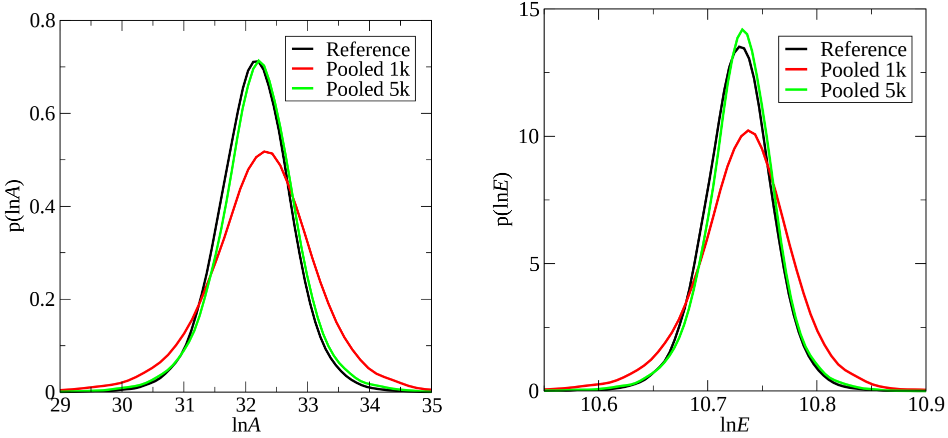
Ref + DFI posterior based on a 1000-long data chain



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



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



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

Outlook

Ongoing research on various fronts

- Dimensionality reduction
  - Sensitivity, PCA, ANOVA/HDMR, low-D manifolds, ...
- 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