SAND2015-7606PE

# Hybrid approach to surrogate modeling

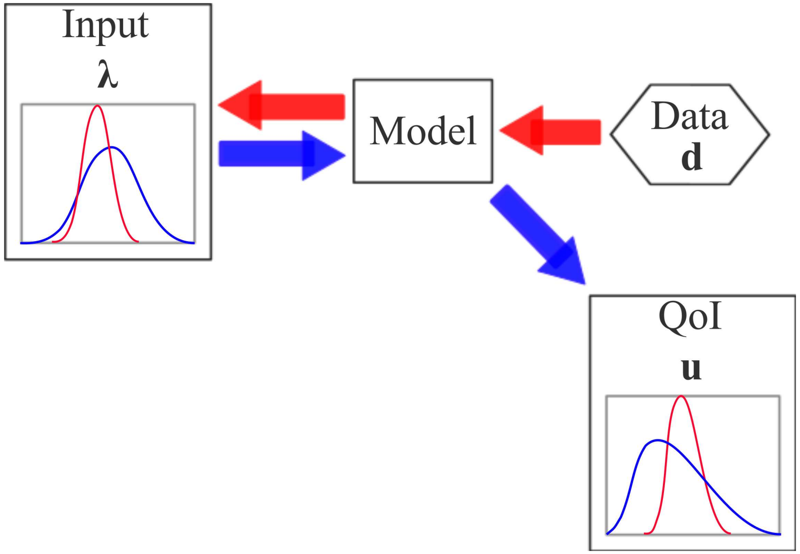
Moe Khalil

July 24, 2015

# Overview

# Why Uncertainty Quantification?






- UQ in computational models:
  - Validation of physical models
  - Statistical inverse problems
  - Calibration and design optimization
  - Use of computational predictions for decision-support
  - Assessment of confidence in computational predictions
- UQ methods:
  - Local sensitivity analysis and error propagation
  - Fuzzy logic; Evidence theory
  - Probabilistic framework; Measure theory
    - Sampling (Monte Carlo)
    - Polynomial Chaos

# Computational Challenges



- Solution of statistical inverse problems and optimization problems can be computationally taxing
  - Large dimensionality of the input space
  - Large-scale forward model (arising from PDEs)
  - Complex (non-linear) behavior is input-output map
- Reducing computational cost:
  - Reducing dimension of input space
    - Parameter-space reduction
    - Truncated Karhunen-Loève expansions for stochastic processes
  - Reduction in number of forward simulations
    - Smarter sampling schemes for inverse problems
    - More efficient optimization algorithms
  - Reducing cost of forward simulations
    - Surrogate models
    - Reduced-order models
    - Multigrid/multiscale approaches

# Surrogates for Forward Models



- Classically, surrogate models can be categorized into [Eldred et al., 2004]
  - Data-fit models
    - Based on interpolation or regression
    - Rely on repeated forward simulations (Black-box)
    - Examples: Gaussian processes, radial basis functions
  - Reduced-order models
    - Derived using a projection framework
    - Rely on full forward problem simulations, or ”snapshots”
    - Examples: Proper orthogonal decomposition, reduced basis methods
  - Hierarchical surrogate models
    - Based on constructing a series of physics-based models of decreasing accuracy
    - Rely on using high and low accuracy models in tandem
    - Examples: mesh coarsening methods, alternative basis expansions

# *Polynomial chaos expansion surrogates*



- Represent model inputs/outputs as random variables
- Construct PCEs for uncertain parameters
- Obtain PCEs for model outputs using *Non-intrusive Spectral Projection*:
  - *Sampling*-based
  - Relies on black-box utilization of the computational model
  - Evaluate projection integrals *numerically* using
    - A variety of Monte Carlo methods
      - Slow convergence; $\sim$ indep. of dimensionality
    - Quadrature/Sparse-Quadrature methods
      - Fast convergence; depends on dimensionality

# Polynomial Chaos Expansion (PCE)



- Model uncertain quantities as random variables (RVs)
- Any RV in $L^2(\Omega, \mathfrak{S}(\boldsymbol{\xi}), P)$ can be written as a PCE:

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

- $\boldsymbol{\xi} = \{\xi_1, \cdots, \xi_n\}$ a set of *i.i.d.* RVs
- $u_k(\boldsymbol{x}, t)$ are coefficients (deterministic):

$$u_k(\boldsymbol{x}, t) = \frac{\langle u(\boldsymbol{x}, t, \boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) \rangle}{\langle \Psi_k^2(\boldsymbol{\xi}) \rangle}$$

- $\Psi_k()$ are functions orthogonal w.r.t. $p(\boldsymbol{\xi})$
  - e.q. Legendre polynomials with Uniform germ
- Advantage: numerous functional analysis methods
  - Computational efficiency
  - Sensitivity information

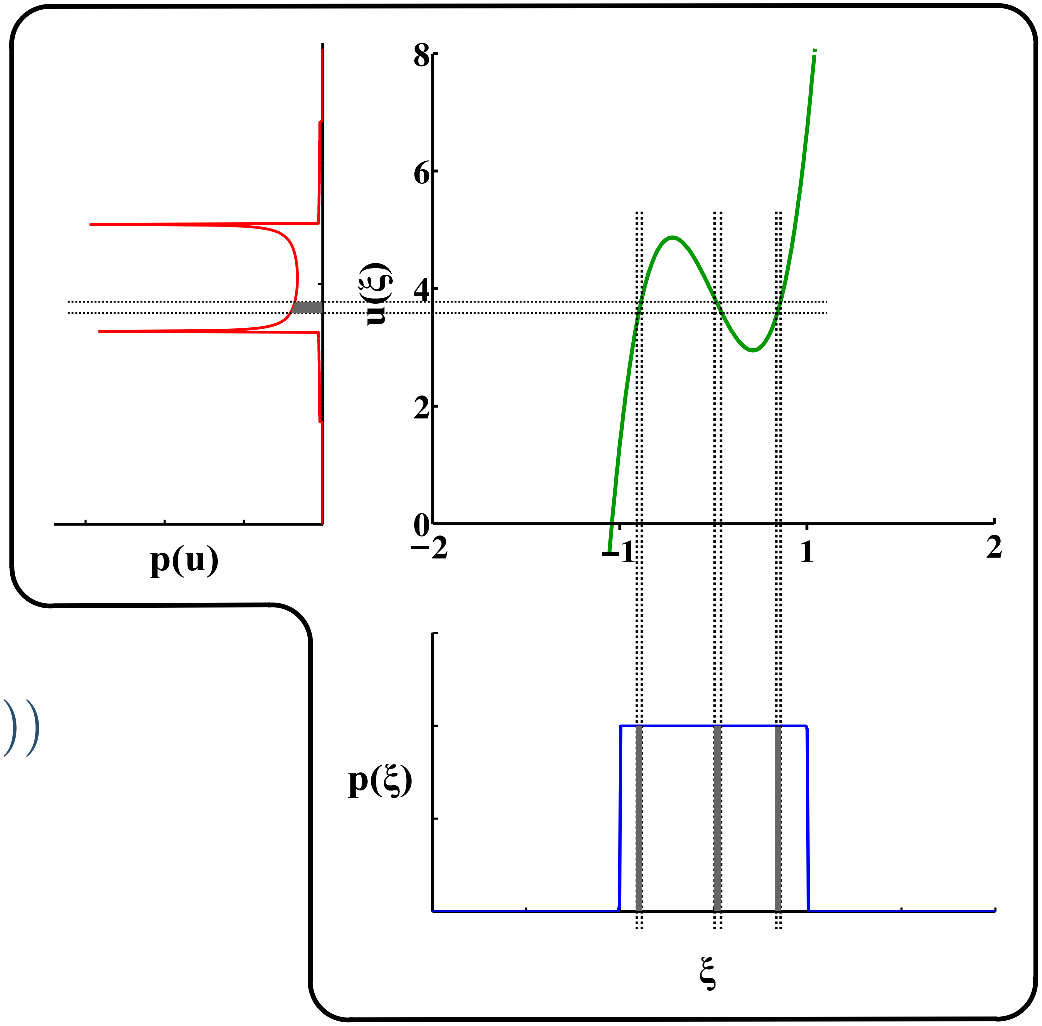
# PCE example: a map from standard RV to QoI



$$\mathrm{u} = \sum_{k=0}^{P} u_k \psi_k(\xi)$$

- $\mathrm{u}$: random variable
- $u_k$: PC coefficients
- $\psi_k$: 1D Legendre polynomial of order $k$
- $\xi$: Uniform RV

$$\mathrm{u} = 4 + 2.25\left(\frac{1}{2}\left(5\xi^3 - 3\xi\right)\right)$$



# Application: Turbulent bluff-body non-premixed flame

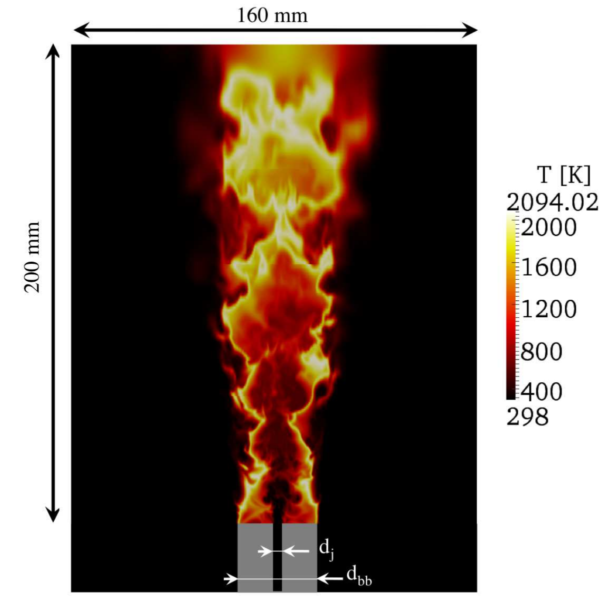


Collaborators:

- Guilhem Lacaze, Joe Oefelein, and Habib Najm (Sandia)

Flame specifications:

- Sydney Bluff-body burner
- Fuel: 50/50 mixture of methane and hydrogen injected at 108 m/s with coflow at 35 m/s at the combustor section




Large-Eddy simulation theoretical framework [Oefelein 2006]:

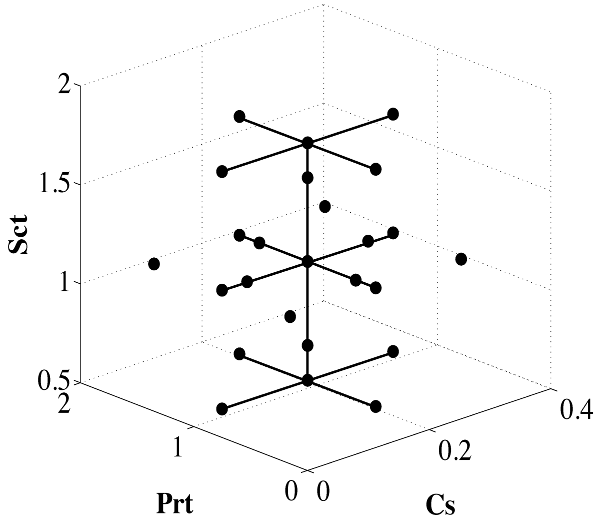
- Fully-coupled, compressible conservation equations
- Transport of mass, momentum, energy and mixture fraction
- Combustion closure: Flamelet (tabulation of heat-release) with presumed-pdf
- Dynamic SGS modeling turned off

# Application: Numerics



Current UQ investigation focuses on

- $0.065 < C_s < 0.346$ [Piomelli 1989]
- $0.5 < Pr_t < 1.7$ [Erlebacher 1992]
- $0.5 < Sc_t < 1.7$ [Reynolds 1976]



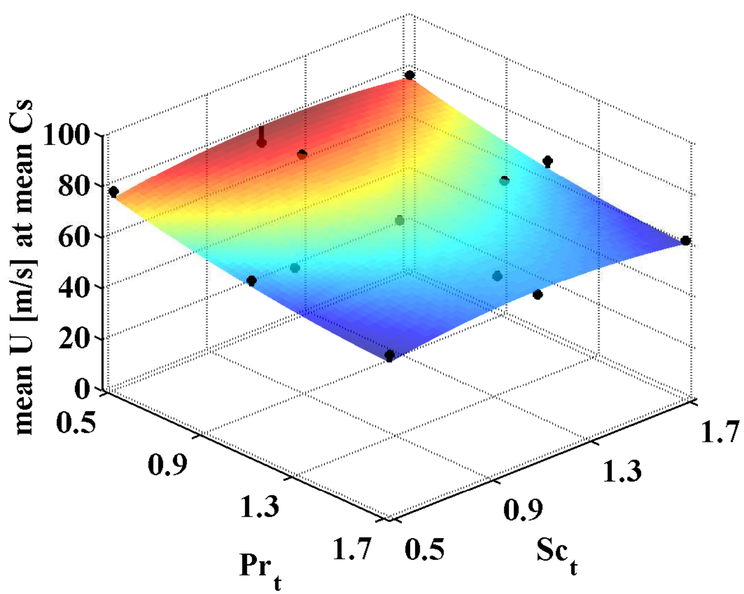

Computational aspects:

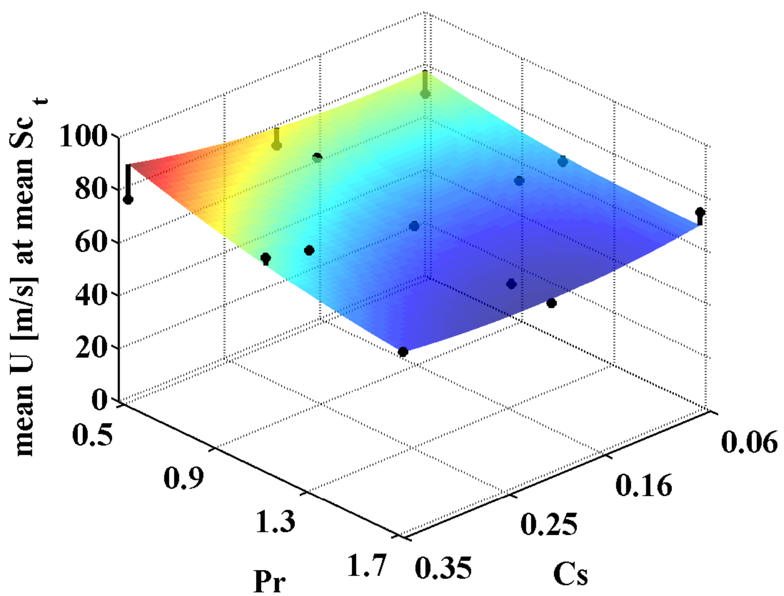
- 25 LESs: explore the parametric space and build surrogate
- 18 LESs: test surrogate accuracy
- Mesh: 12 million hexahedral cells
- Simulation is run for $400$ ms to attain steady state and the subsequent $30$ ms used to gather statistics
- One forward simulation $\sim 62$ hours on 1024 processors
- Total CPU consumption $\sim 2.7 \times 10^6$ cpu hours
- NERSC's newest supercomputer Edison (Cray XC30)

# Response Surfaces



- From the 25 simulations at quadrature points, the following selected second order PC surrogate models are obtained:



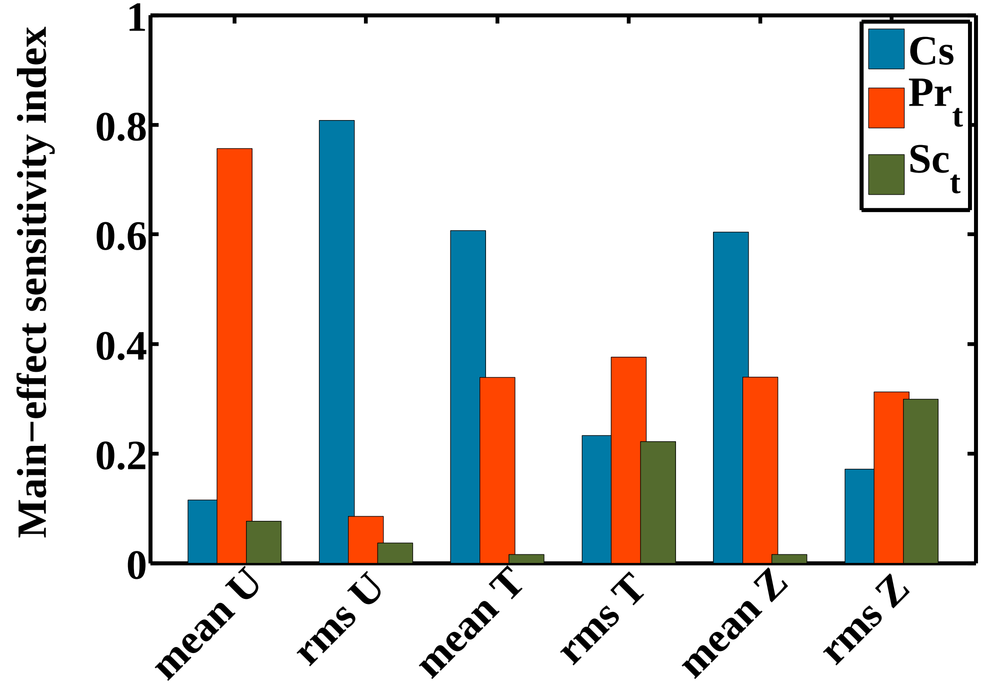

NRMSE error estimated from 18 additional LESs to be $4.4\%$

# Global Sensitivity Analysis



- The PC surrogate models allow us to perform global sensitivity analysis analytically in order to quantify the uncertainty in the QoIs due to each uncertain parameter:

| | | | | | | |
|---|---|---|---|---|---|---|
| mean | 70.9 | 38.6 | 479 | 90.3 | 0.809 | 0.098 |
| sd | 5.27 | 3.70 | 54.8 | 12.2 | 0.062 | 0.011 |
| cov (%) | 7.4 | 9.6 | 11 | 14 | 7.7 | 11 |



Sandia National Laboratories

# A hybrid approach to surrogate modeling



- Constructing an accurate PCE surrogate over the entire parameter space may be prohibitive due to strong nonlinearities in the quantities of interest
- Two approaches to increase the accuracy of PCE surrogates
  - Domain decomposition
    - Split the parameter space into non-overlapping subdomains
    - For each subdomain, construct a low-order local PCE surrogate
  - Hybrid construction using Padé Approximants
    - Expand output using PCE along $N-1$ dimensions
    - Use Padé Approximant representation along $N$th dimension exhibiting strong nonlinearity
    - Padé Approximant is formulated as a ratio of PCEs

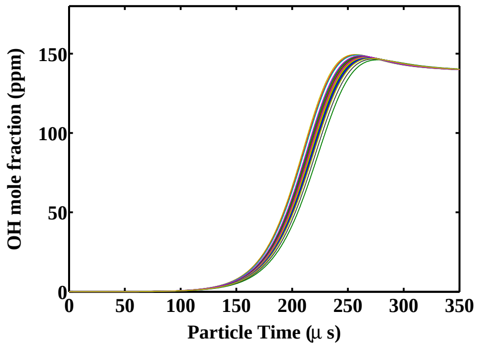
$$\begin{aligned} u(\boldsymbol{x}, t, \boldsymbol{\xi}) = u(\boldsymbol{\theta}) &\simeq \sum_{n=0}^{N} u_n(\theta_i)\Psi_n(\boldsymbol{\theta}_{(\sim i)}) \\ &= \sum_{n=0}^{N} \frac{\sum_{m=0}^{M} p_{m,k}\Phi_m(\theta_i)}{\sum_{l=0}^{L} q_{l,k}\Phi_l(\theta_i)} \Psi_n(\boldsymbol{\theta}_{(\sim i)}) \end{aligned}$$

# Application: Chemical Kinetics

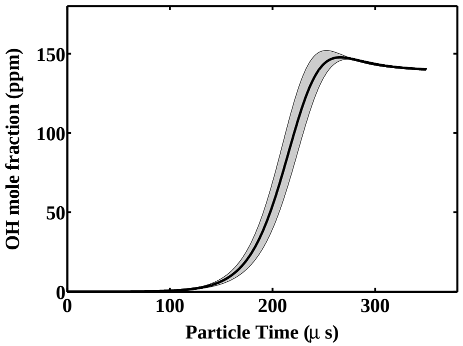


- Bayesian inference of chemical kinetic rate parameters from measured species concentration profiles
- Unknown rate coefficients for two reactions

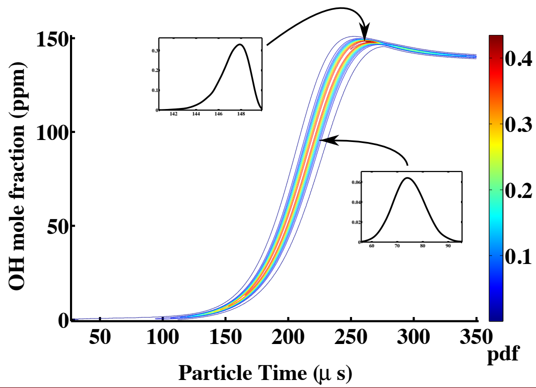
$$H + O_2 \rightarrow OH + O$$
$$OH + H_2 \rightarrow H_2O + H$$




realizations




confidance interval

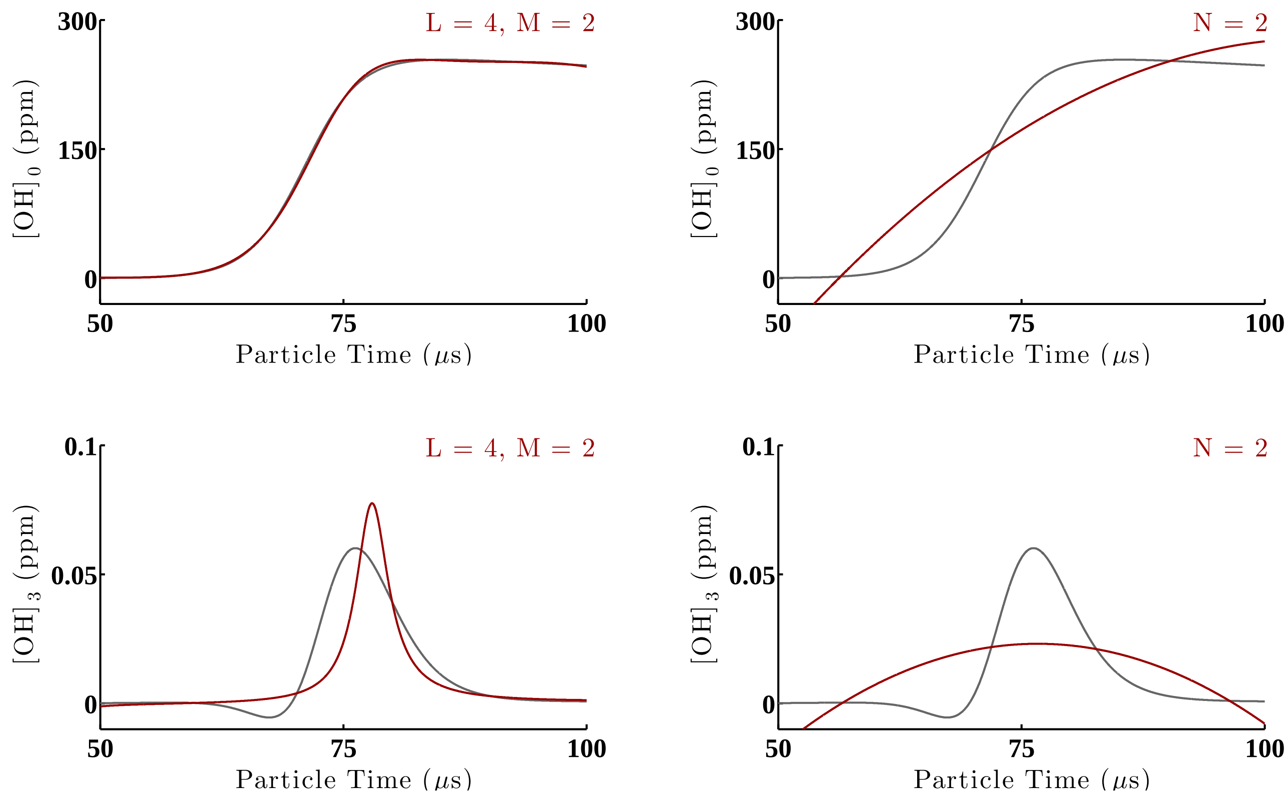


# Convergence Study



$$[\mathrm{OH}](k_1,k_2,t) \simeq \sum_{n=0}^{N} [\mathrm{OH}]_n(t)\Psi_n(k_1,k_2)$$

$$= \sum_{n=0}^{N} \frac{\sum_{m=0}^{M} p_{m,k}\Phi_m(t)}{\sum_{l=0}^{L} q_{l,k}\Phi_l(t)} \Psi_n(k_1,k_2)$$



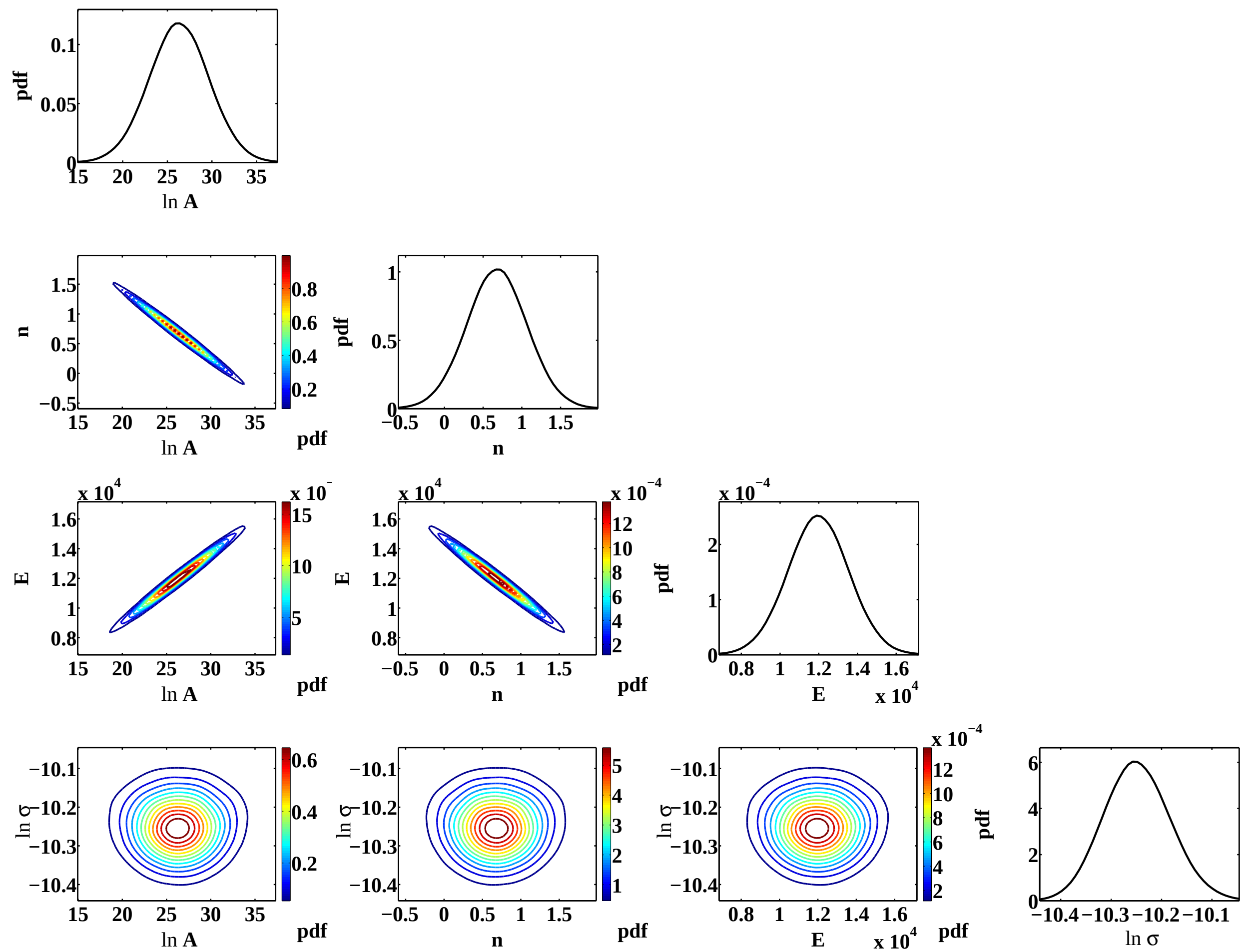

Padé

PCE

# Bayesian Inference



- The PC surrogate is used to expedite MCMC sampling of parameters from their associated posterior pdf



# Conclusion



- Polynomial chaos expansion surrogates for large-scale systems via efficient sparse-quadrature
  - Forward UQ in LES of a turbulent non-premixed hydrocarbon flame
- Hybrid surrogate model formulation using Padé approximants and polynomial chaos expansion for strong nonlinear behavior
  - Application: Inference of the Arrhenius parameters for the rate coefficient of the H2/O2 mechanism branching reaction $H+O2 \rightarrow OH+O$
- Methodology can be extended to combine PCE with other techniques
  - Radial basis functions
  - Neural networks