Uncertainty Quantification in Climate Modeling

Khachik Sargsyan, Cosmin Safta,
Bert Debusschere, Habib Najm

Sandia National Laboratories
Livermore, CA

UQ for Multiscale Systems
Baltimore, MD
July 20, 2010

Uncertainties in Climate Modeling

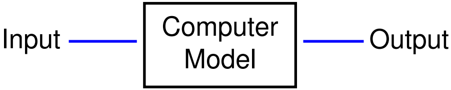
- Uncertainty sources
  - Parameter uncertainty
    - Model parameters
    - Initial/boundary conditions
    - Model geometry/structure
  - Model/structural uncertainty
    - Unknown physics
    - Reduced order models
  - Scenario uncertainty
    - Policy restrictions
    - Technology improvement
  - Intrinsic variability
    - Stochastic physics
  - Numerical errors

Uncertainties in Climate Modeling

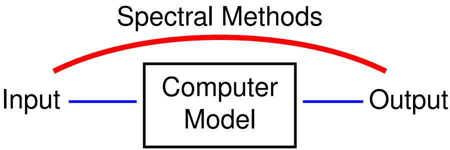
- Uncertainty sources
  - Parameter uncertainty
    - Model parameters
    - Initial/boundary conditions
    - Model geometry/structure
  - Model/structural uncertainty
    - Unknown physics
    - Reduced order models
  - Scenario uncertainty
    - Policy restrictions
    - Technology improvement
  - Intrinsic variability
    - Stochastic physics
  - Numerical errors

- Need UQ for...
  - Model validation
  - Confidence assessment
  - Optimal design
  - Data assimilation

UQ components and methods

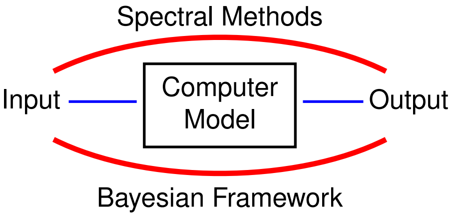


UQ components and methods



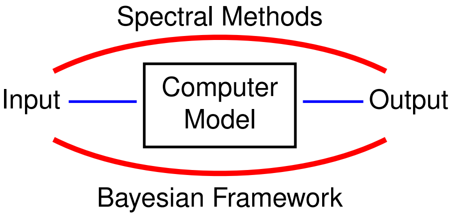
- Sensitivity analysis
  - Small parameter perturbations
- Predictability assessment
  - Larger parameter uncertainties

UQ components and methods



- Sensitivity analysis
  - Small parameter perturbations
- Predictability assessment
  - Larger parameter uncertainties
- Parameter estimation/calibration
  - Inverse problems

UQ components and methods



- Forward UQ methods
  - Direct (intrusive)
    - Derive new forward model
    - Intrusive Spectral Projection (ISP)
  - Sampling (non-intrusive)
    - Monte-Carlo, Quasi Monte-Carlo
    - Non-intrusive Spectral Projection (NISP)

Non-Intrusive Spectral Projection (NISP)

- Polynomial Chaos expansions for input γ and output Z

$$\gamma \approx \sum_k \gamma_k \Psi_k(\xi)$$

$$Z = f(\gamma) \approx \sum_k f_k \Psi_k(\xi)$$

- Orthogonal projection via quadrature to obtain PC modes

$$f_k = \int f(\gamma)\Psi_k(\xi)\mathrm{pdf}(\xi)d\xi \approx \sum_* f(\gamma(\xi^*))w^*$$

Non-Intrusive Spectral Projection (NISP)

- Polynomial Chaos expansions for input γ and output Z

$$\gamma \approx \sum_k \gamma_k \Psi_k(\xi)$$

$$Z = f(\gamma) \approx \sum_k f_k \Psi_k(\xi)$$

- Orthogonal projection via quadrature to obtain PC modes

$$f_k = \int f(\gamma)\Psi_k(\xi)\mathrm{pdf}(\xi)d\xi \approx \sum_* f(\gamma(\xi^*))w^*$$

Challenges tackled in more detail today

- non-linearities/bifurcations

Non-Intrusive Spectral Projection (NISP)

- Polynomial Chaos expansions for input γ and output Z

$$\gamma \approx \sum_k \gamma_k \Psi_k(\xi)$$

$$Z = f(\gamma) \approx \sum_k f_k \Psi_k(\xi)$$

- Orthogonal projection via quadrature to obtain PC modes

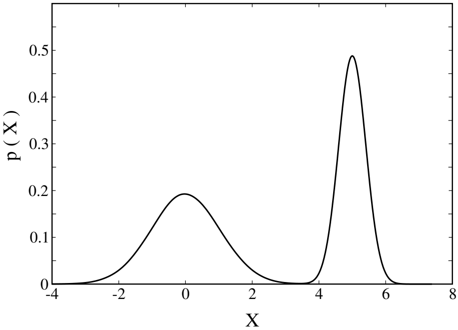
$$f_k = \int f(\gamma)\Psi_k(\xi)\text{pdf}(\xi)d\xi \approx \sum_* f(\gamma(\xi^*))w^*$$

Challenges tackled in more detail today

- non-linearities/bifurcations
- low-probability/high-impact events

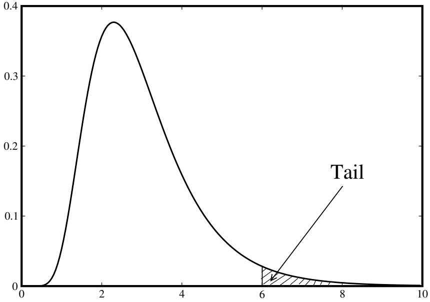
UQ methods are challenged by..

- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Intrinsic stochasticity
- Curse of dimensionality



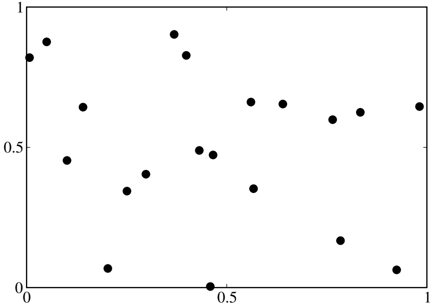
UQ methods are challenged by..

- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Intrinsic stochasticity
- Curse of dimensionality



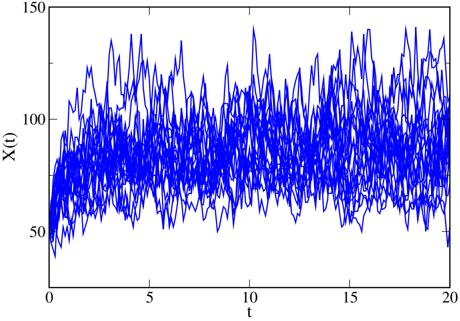
UQ methods are challenged by..

- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Intrinsic stochasticity
- Curse of dimensionality



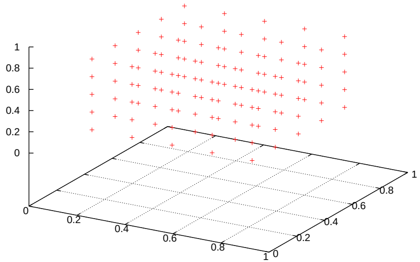
UQ methods are challenged by..

- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Intrinsic stochasticity
- Curse of dimensionality



UQ methods are challenged by..

- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Intrinsic stochasticity
- Curse of dimensionality



UQ Challenges

- non-linearities/bifurcations
- low-probability/high-impact events
- limited data
- intrinsic stochasticity
- high dimensionality

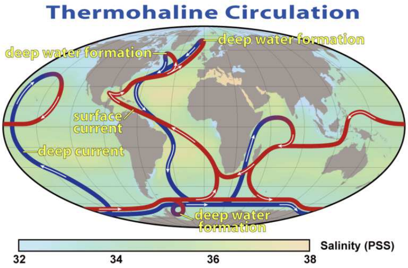
Discontinuities/Nonlinearities/Bifurcations

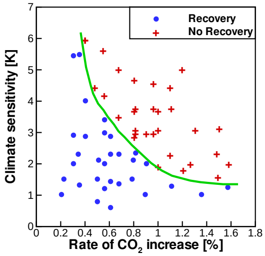
- Stochastic domain decomposition
  - Wiener-Haar Expansions, Multiwavelets [Le Maître *et al*, 2004,2007]
  - Multielement PC [Wan & Karniadakis, 2009]
- Data domain decomposition [Sargsyan *et al*, 2009,2010]
  - Data clustering
  - Mixture PC expansions

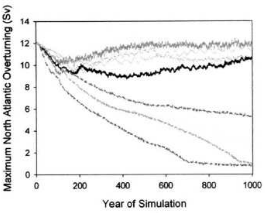
- Adaptive setting
- Does not scale with dimensionality
- For expensive models, can not split much
- *Need a 'smart' domain decomposition*

Webster et al - J. Environ. Syst. 31: 39-59, 2007

- Computational model - EMIC
- Input parameters
  - Rate of CO_2 increase (r)
  - Climate sensitivity (λ)
- Output observable
  - Overturning streamfunction (Z)

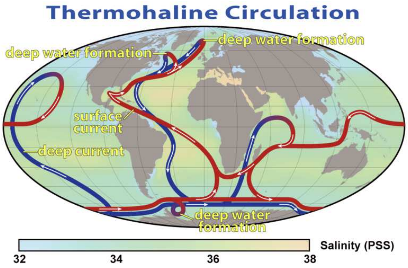


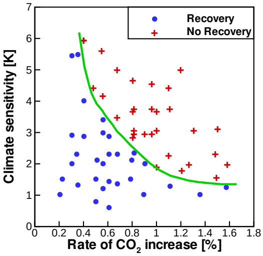


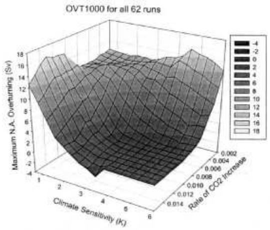


Global representations fail to capture discontinuities

- Computational model - EMIC
- Input parameters
  - Rate of CO_2 increase (r)
  - Climate sensitivity (λ)
- Output observable
  - Overturning streamfunction (Z)







UQ & Discontinuities - Proposed Methodology

Our approach locates the discontinuity first so the domain can be subdivided into regions with smooth model response where spectral uncertainty quantification methods can be used.

Two-step approach:

- **Bayesian inference of the location of the discontinuity**
- **Polynomial chaos representation via parameter domain mapping at each side of the discontinuity**

Bayesian Inference of the Location of Discontinuity

- Parameterize the discontinuity: $r \approx p_{\boldsymbol{c}}(\lambda) = \sum_{k=0}^{K} c_k P_k(\lambda)$

- Approximation model:

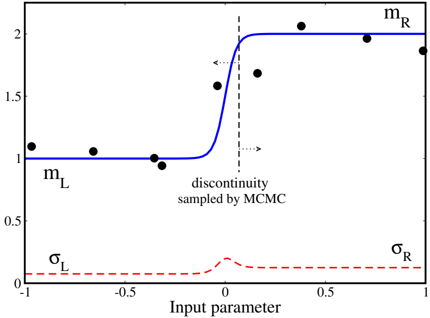
$$\mathcal{M}_{\boldsymbol{c}} \equiv g(\lambda, r) = m_L + (m_R - m_L)\frac{1 + \tanh\left(\alpha(r - p_{\boldsymbol{c}}(\lambda))\right)}{2}$$

- Noise model postulated: $\sigma(\lambda, r)$
- Likelihood function:

$$\log P(\mathcal{D}|\mathcal{M}_{\boldsymbol{c}}) = \sum_{i=1}^{N} \log\left(P(z_i|\mathcal{M}_{\boldsymbol{c}})\right) = -\sum_{i=1}^{N} \frac{(z_i - g(\lambda, r))^2}{2\sigma(\lambda, r)^2}.$$

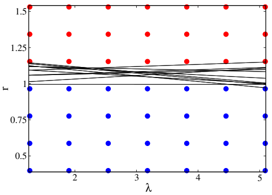
Bayesian Inference of the Location of Discontinuity

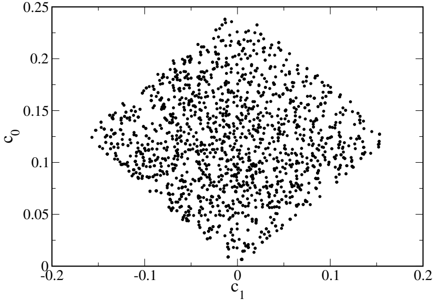
- Parameterize the discontinuity: $r \approx p_{\mathbf{c}}(\lambda) = \sum_{k=0}^{K} c_k P_k(\lambda)$
- Bayes' formula: $P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$

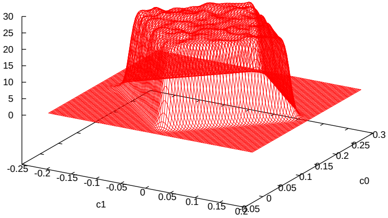


Highlights

- Any distribution of input points
- Generalizes to multiple dimensions
- Probabilistic representation







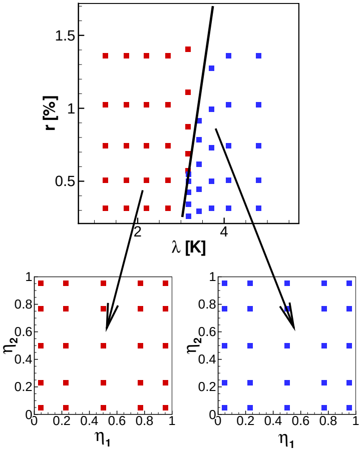
Discontinuity curve samples and their pdf

Parameter Domain Mapping

- Assume linear discontinuity
- Use Rosenblatt Transformation (RT) to map the pair of uncertain parameters (λ,r) to i.i.d. uniform random variables η_1 and η_2:

$$\lambda = F_{\lambda}^{-1}(\eta_1),$$
$$r = F_{r|\lambda}^{-1}(\eta_2|\eta_1)$$

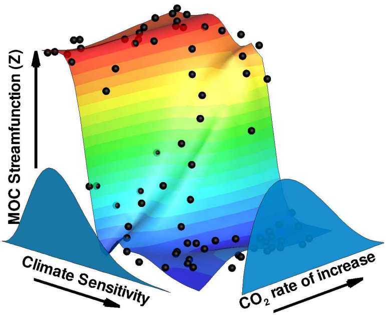
- Apply the RT mapping to both sides of the discontinuity



ROSENBLATT TRANSFORMATION: $(\lambda, r) \rightarrow (\eta_1, \eta_2)$

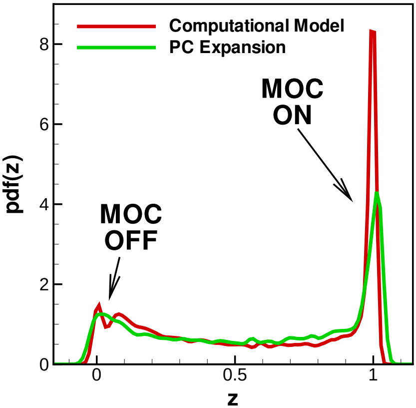
Discontinuous data represented well with the averaged PC

PCE IN (η_1, η_2) DOMAIN



Discontinuous data represented well with the averaged PC.

OUTPUT PDF



Resulting output PDF given input parameter joint PDF.

Summary

- A methodology for uncertainty quantification in climate models with limited data and discontinuities was proposed:
  - Bayesian approach to detect and parameterize the discontinuity as well as the uncertainty associated with it.
  - Rosenblatt transformation maps each of the irregular domains to rectangular ones where the application of the local spectral methods of uncertainty propagation is feasible.
- “Knowledge Discovery from Climate Data: Prediction, Extremes, and Impacts” Workshop Proceedings - 9th IEEE International Conference on Data Mining, 2009.
- Full paper in preparation.

UQ Challenges

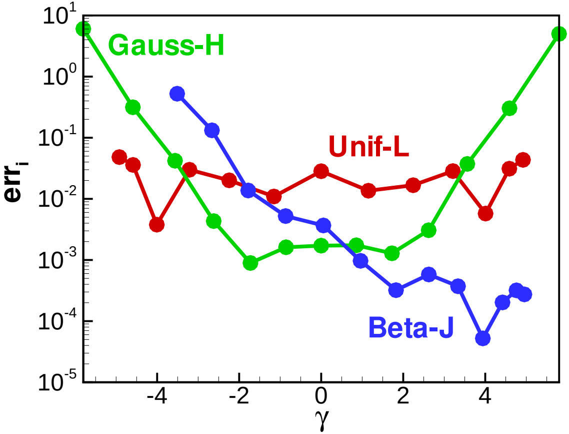
- non-linearities/bifurcations
- low-probability/high-impact events
- limited data
- intrinsic stochasticity
- high dimensionality

Dealing with 'fat' tails

- Several climate observables (e.g. climate sensitivity) exhibit heavy tails
  - require a significant number of simulations to obtain a good sampling of these regions
- Construct spectral expansions based on...
  - Non-classical bases that cluster points in the tail region
  - Bases tailored to the expected behavior of the output
- Use spectral expansions for...
  - Propagating distributions from input parameters to output observables
  - Surrogate models to accelerate the inference process in inverse problems

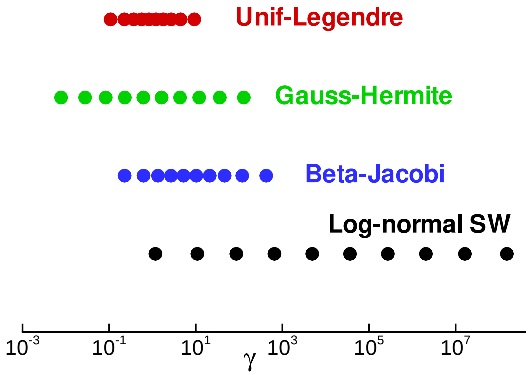
Pointwise error is large at low-probability regions

$$Z = f(\gamma) \approx \sum_k f_k \Psi_k(\xi) \Longrightarrow f_k = \int f(\gamma)\Psi_k(\xi)\mathrm{pdf}(\xi)d\xi \approx \sum_* f(\gamma(\xi^*))w^*$$



Non-classical quadrature points span the tails better

$$Z = f(\gamma) \approx \sum_k f_k \Psi_k(\xi) \Longrightarrow f_k = \int f(\gamma)\Psi_k(\xi)\mathrm{pdf}(\xi)d\xi \approx \sum_* f(\gamma(\xi^*))w^*$$



Build a custom PC based on input distribution

- Classical PCEs for input γ and output Z
  - ξ is normal, $\Psi_k(\cdot)$ are Hermite - standard!

$$\gamma \approx \sum_k \gamma_k \Psi_k(\xi)$$

$$Z = f(\gamma) \approx \sum_k f_k \Psi_k(\xi)$$

- *Customized* PCE for output Z with respect to input distribution:
  - γ is *any*, $\Phi_k(\cdot)$ are found by orthogonalization.

$$\gamma = \gamma \qquad \text{(as 'optimal' as it gets)}$$

$$Z = f(\gamma) \approx \sum_k f_k \Phi_k(\gamma) \qquad \text{(hopefully, near optimal)}$$

Build a custom PC based on input distribution

- Classical PCEs for input γ and output Z
  - ξ is normal, $\Psi_k(\cdot)$ are Hermite - standard!

$$\gamma \approx \sum_k \gamma_k \Psi_k(\xi)$$

$$Z = f(\gamma) \approx \sum_k f_k \Psi_k(\xi)$$

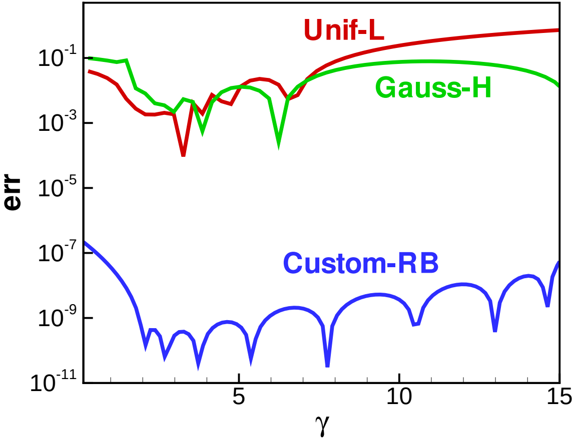
- *Customized* PCE for output Z with respect to input distribution:
  - γ is *any*, $\Phi_k(\cdot)$ are found by orthogonalization.

$$\gamma = \gamma \qquad \text{(as `optimal' as it gets)}$$

$$Z = f(\gamma) \approx \sum_k f_k \Phi_k(\gamma) \qquad \text{(hopefully, near optimal)}$$

Custom PC Expansions show much better convergence than standard PCE

Input γ belongs to Roe-Baker climate sensitivity distribution.
Synthetic forward model: $f(\gamma) = \cos(\gamma)$



Switching gear...

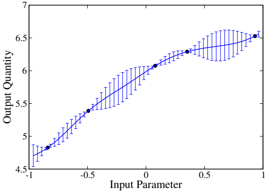
UQ Challenges

- non-linearities/bifurcations
- low-probability/high-impact events
- limited data
- intrinsic stochasticity
- high dimensionality

Limited Data

Both observational experiments and computer model simulations are expensive.

- Need to infer functional representation based on limited number of model runs/experiments.
  - Interpolation (kriging)
  - Gaussian Process emulation to assess the lack-of-knowledge [O'Hagan]
  - Extended to stochastic model setting



- *Bayesian* experimental design
  - What are the best locations to take observations?
  - At which parameter sets to run climate models to gain maximal information?

UQ Challenges

- non-linearities/bifurcations
- low-probability/high-impact events
- limited data
- intrinsic stochasticity
- high dimensionality

Intrinsic stochasticity

- Stochastic computer models
- Probabilistic representation of the lack-of-knowledge
- *Climate buzzword:* Stochastic Physics [Palmer & Williams, 2009]

- See previous talk
  (Bert Debusschere on Stochastic Reaction Networks)
- Extend Gaussian Process emulation of deterministic codes to the stochastic case

be Bayesian!!

-

Intrinsic stochasticity

- Stochastic computer models
- Probabilistic representation of the lack-of-knowledge
- *Climate buzzword:* Stochastic Physics [Palmer & Williams, 2009]

- See previous talk
  (Bert Debusschere on Stochastic Reaction Networks)
- Extend Gaussian Process emulation of deterministic codes to the stochastic case

be Bayesian!!

-

UQ Challenges

- non-linearities/bifurcations
- low-probability/high-impact events
- limited data
- intrinsic stochasticity
- high dimensionality

Curse of Dimensionality

- (Dimension-adaptive) Sparse quadrature integration
- High Dimensional Model Representation (HDMR)
  - would not handle discontinuities
  - tried cut-HDMR in a chemical kinetics context: fails!
- Proper Generalized Decomposition [Nuoy, 2010]

- Turn it into the *blessing of dimensionality* [Donoho, 2000]
- Compressive Sensing in spectral methods' [Doostan *et al*, 2009]

short answer: no free lunch

Curse of Dimensionality

- (Dimension-adaptive) Sparse quadrature integration
- High Dimensional Model Representation (HDMR)
  - would not handle discontinuities
  - tried cut-HDMR in a chemical kinetics context: fails!
- Proper Generalized Decomposition [Nuoy, 2010]

- Turn it into the *blessing of dimensionality* [Donoho, 2000]
- Compressive Sensing in spectral methods' [Doostan *et al*, 2009]

short answer: no free lunch

-

Summary

- Nonlinearities, Bifurcations, Bimodalities
  - Probabilistic detection of discontinuities followed by domain mapping and polynomial chaos expansions to construct model "surrogates".
- Tail regions
  - Employ spectral basis that cluster quadrature points in the tail to construct surrogate models.
  - Construct custom spectral basis based on "expected" shape of the climate model output to improve convergence of the spectral expansion.

Current and future work

- Bring in real climate model data
- Still prohibitively many model runs required: possibly give up orthogonal projection in favor of Bayesian inference
- Experimental design: inform climate modelers on the optimal parameter sets to run simulations
- Gaussian process emulation
  - Couple with PC, either
    a) PC as the mean trend, or
    b) uncertain integration via Bayes-Hermite quadrature.
- Patching PC expansions: capture both the mean and the tail regions

- Climate Science for a Sustainable Energy Future (CSSEF)
  - Multi-lab, multi-year project
  - UQ needed for calibration, validation and prediction

Acknowledgements

- Sandia's internal
  Laboratory Directed Research & Development,

- partially funded by combined support from the DOE
  National Nuclear Security Administration and Basic Energy Research,

- DOE Office of Science, Advanced Scientific Computing Research,
  Applied Mathematics.

Thank You!

Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

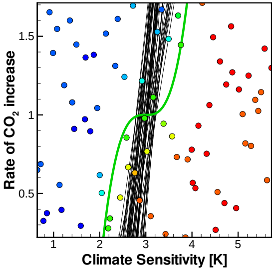
Inference of Discontinuity - 3^{rd} order polynomial

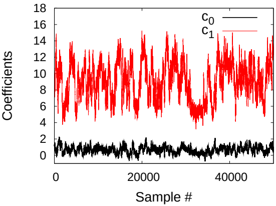
- Synthetic discontinuous data

$$z_i = (1 + \sigma\xi)\,\mathrm{erf}\left(\beta(r_i - \tilde{r}(\lambda_i))\right).$$

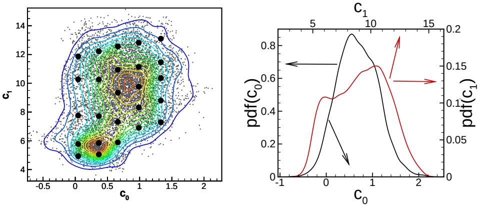
- Use straight lines to infer the discontinuity

$$\tilde{r}(\lambda) = c_0 + c_1\lambda.$$





MCMC samples



Joint and Marginal Posterior Distributions

PC expansion, averaged over discontinuity curves

- PC expansion for each discontinuity curve sample:

$$Z_{\boldsymbol{c}}^{L,R}(\lambda, r) = \tilde{Z}_{\boldsymbol{c}}(\eta_1, \eta_2) = \sum_{p=0}^{P} z_p \Psi_p^{(2)}(\eta_1, \eta_2)$$

- Model expansion depends on the parameter location:

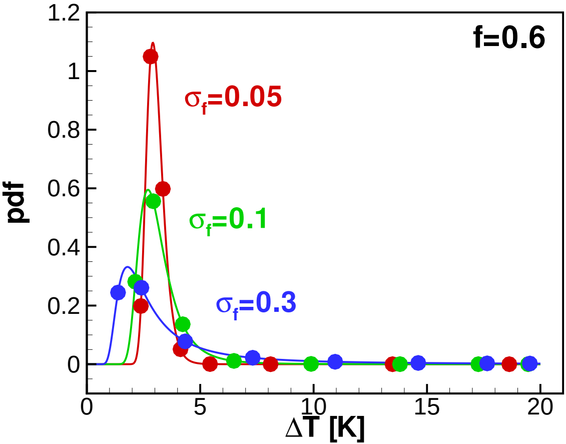
$$Z_{\boldsymbol{c}}(\lambda, r) = \begin{cases} Z_{\boldsymbol{c}}^{L}(\lambda, r) & \text{if } (\lambda, r) \in D_L \\ Z_{\boldsymbol{c}}^{R}(\lambda, r) & \text{if } (\lambda, r) \in D_R \end{cases}.$$

- Average over all PC expansions via RT:

$$\hat{Z}(\lambda, r) = \int_C p(\boldsymbol{c}) Z_{\boldsymbol{c}}(\lambda, r) d\boldsymbol{c} = \int_{[0,1]^{K+1}} Z_{R^{-1}(\vec{\eta})}(\lambda, r) d\vec{\eta}$$

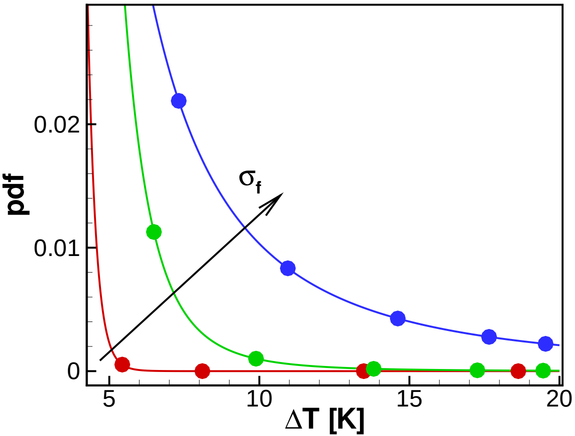
Custom Basis' Quad Points Extend to the Tail

(pdf shape from Roe & Baker, Science 2007)



Custom Basis' Quad Points Extend to the Tail

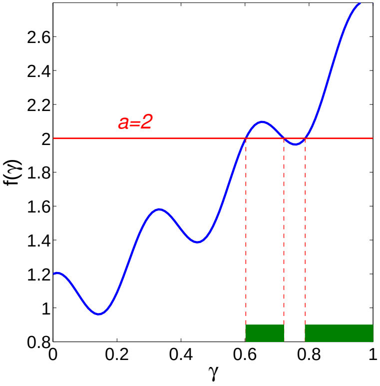
(pdf shape from Roe & Baker, Science 2007)



Accuracy of tail probabilities

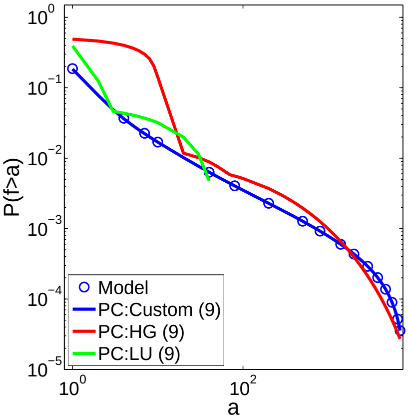
$$P(f > a) = \int_{\gamma:f(\gamma)>a} \mathrm{pdf}(\gamma)d\gamma$$

- Compare $P(f_{exact} > a)$ against $P(f_{PC} > a)$.
- Accuracy depends both on
  - $|f_{exact} - f_{PC}|$.
  - how accurate are the the regions where $f > a$.



Test probability errors

- Forward model:
  $f(\gamma) = \exp(0.5\gamma - 1)$
  - γ is a truncated log-normal
- PC expansions:
  - Legendre polynomials
  - Hermite polynomials
  - Custom basis (truncated log-normal)



Test probability errors