# *Advanced Tools for Uncertainty Quantification in Climate Modeling*

Khachik Sargsyan, Cosmin Safta,
Bert Debusschere, Habib Najm

Sandia National Laboratories
Livermore, CA

TriLab UQ Meeting
Breckenridge, CO
June 28, 2010

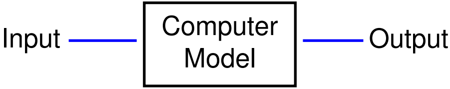
# Uncertainty Quantification in Climate Modeling

- Uncertainty sources
  - Parameter uncertainty
    - Model parameters
    - Initial/boundary conditions
    - Model geometry/structure
  - Model uncertainty
    - Unknown physics
    - Reduced order models
  - Scenario uncertainty
    - Policy restrictions
    - Technology improvement
  - Intrinsic variability
- Need UQ for...
  - Model validation
  - Confidence assessment
  - Optimal design
  - Data assimilation

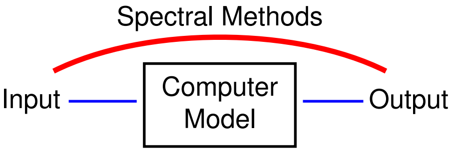
# Uncertainty Quantification in Climate Modeling

- Uncertainty sources
  - Parameter uncertainty
    - Model parameters
    - Initial/boundary conditions
    - Model geometry/structure
  - Model uncertainty
    - Unknown physics
    - Reduced order models
  - Scenario uncertainty
    - Policy restrictions
    - Technology improvement
  - Intrinsic variability
- Need UQ for...
  - Model validation
  - Confidence assessment
  - Optimal design
  - Data assimilation

# UQ components and methods

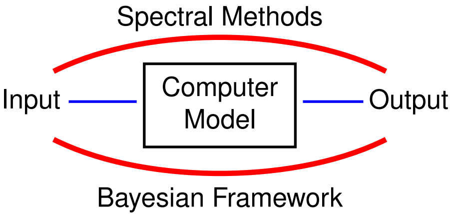


# UQ components and methods



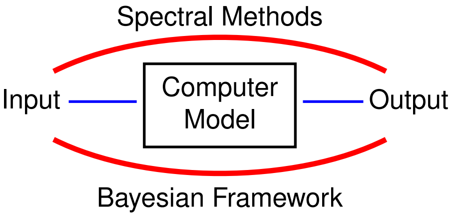

- Sensitivity analysis
  - Small parameter perturbations
- Predictability assessment
  - Larger parameter uncertainties

# UQ components and methods



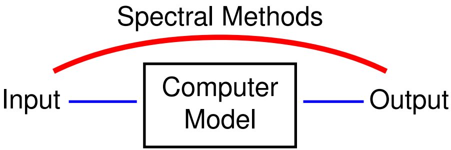

- Sensitivity analysis
  - Small parameter perturbations
- Predictability assessment
  - Larger parameter uncertainties
- Parameter estimation
  - Inverse problems

# UQ components and methods




- Sensitivity analysis
  - Small parameter perturbations
- Predictability assessment
  - Larger parameter uncertainties
- Parameter estimation
  - 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 $\gamma$ 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 $\gamma$ 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 this talk

- non-linearities/bifurcations

# Non-Intrusive Spectral Projection (NISP)

- Polynomial Chaos expansions for input $\gamma$ 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

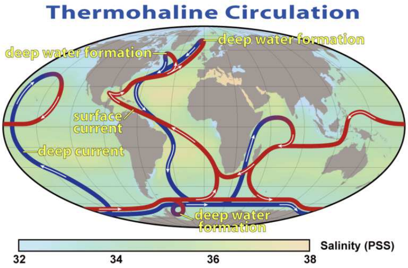
$$f_k = \int f(\gamma)\Psi_k(\xi)\mathrm{pdf}(\xi)d\xi \approx \sum_* f(\gamma(\xi^*))w^*$$

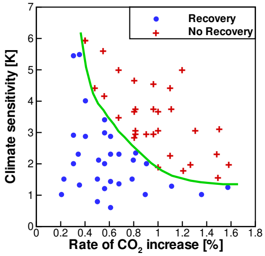
## Challenges tackled in this talk

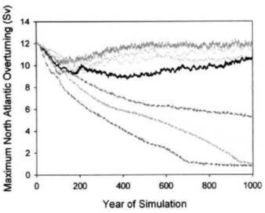
- non-linearities/bifurcations
- low-probability/high-impact events

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

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

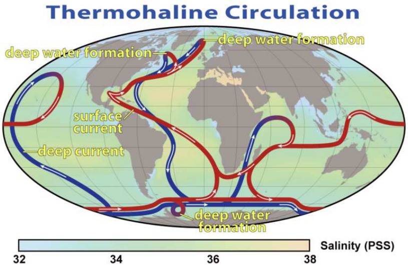


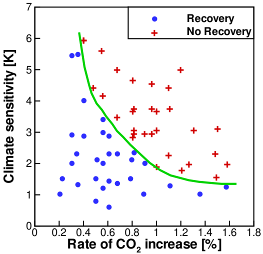



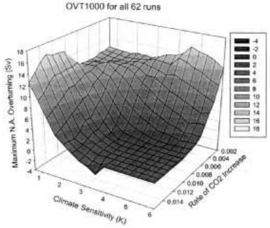



# Global representations fail to capture discontinuities

- Computational model - EMIC
- Input parameters
  - Rate of $CO_2$ increase ($r$)
  - Climate sensitivity ($\lambda$)
- 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*

- Need to represent model output in a problem-independent fashion that takes into account the bifurcations
  - **Bayesian inference of the location of the discontinuity**
- Need to perform uncertainty quantification with only a limited set of sample points, due to the computational cost of the forward model
  - **Polynomial chaos representation via parameter domain mapping**

# 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

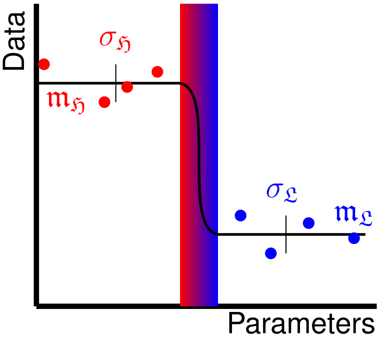
- Parameterize the discontinuity: $r \approx p_{\mathbf{c}}(\lambda) = \sum_{k=0}^{K} c_k P_k(\lambda)$

"Likelihood" "Prior"

**Bayes' formula:**

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$$

"Posterior" "Evidence"

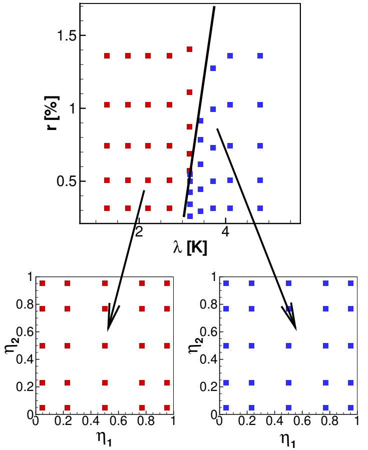


# Parameter Domain Mapping

- Assume linear discontinuity
- Use Rosenblatt Transformation (RT) to map the pair of uncertain parameters ($\lambda$,$r$) to i.i.d. uniform random variables $\eta_1$ and $\eta_2$:

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

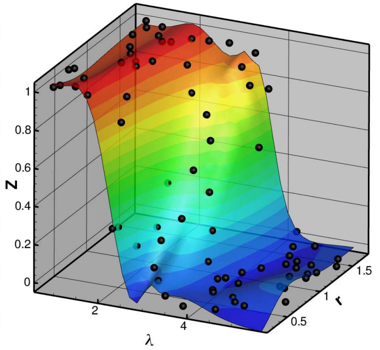
- Apply the RT mapping to both sides of the discontinuity



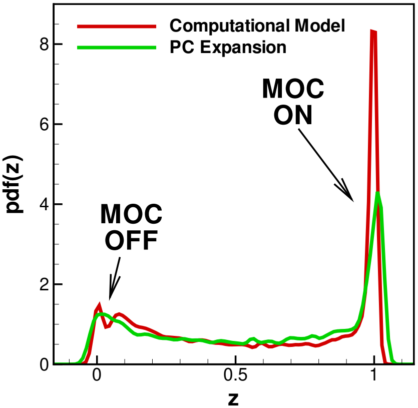

ROSENBLATT TRANSFORMATION: $(\lambda, r) \rightarrow (\eta_1, \eta_2)$

# Discontinuous data represented well with the averaged PC

PCE IN $(\eta_1, \eta_2)$ DOMAIN




OUTPUT PDF




Discontinuous data represented well with the averaged PC.

Resulting output PDF given input parameter joint PDF.

Challenges tackled in this talk

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

# Switching gear...

Challenges tackled in this talk

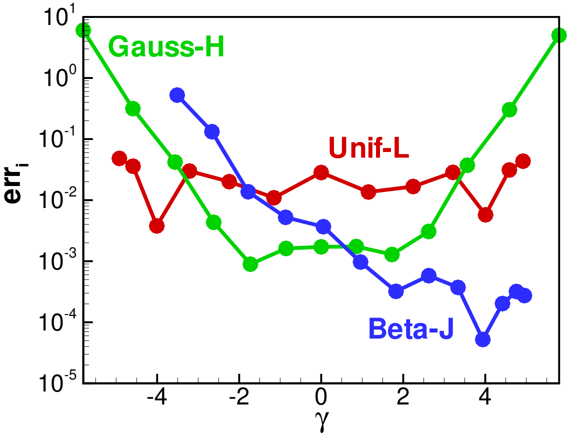
- non-linearities/bifurcations
- low-probability/high-impact events

# 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

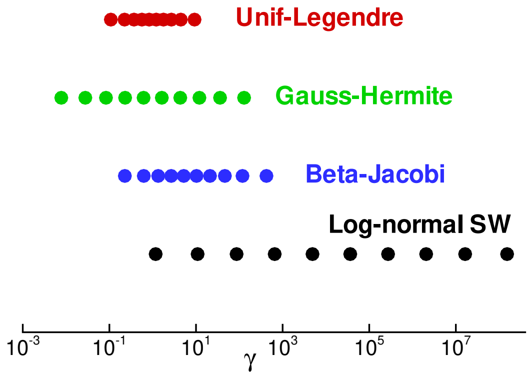
# Pointwise error is large at low-probability regions

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



# 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 $\gamma$ and output $Z$
  - $\xi$ 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:
  - $\gamma$ 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 $\gamma$ and output $Z$
  - $\xi$ is normal, $\Psi_k(\cdot)$ are Hermite - standard!

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

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

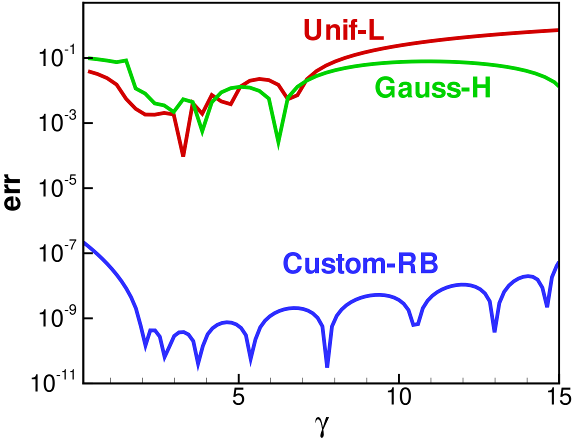
- *Customized* PCE for output $Z$ with respect to input distribution:
  - $\gamma$ is *any*, $\Phi_k(\cdot)$ are found by orthogonalization.

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

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

# Custom PC Expansions show much better convergence than standard PCE

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



# 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.

# Acknowledgements


- Sandia LDRD
- DOE BER
- DOE ASCR

*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.*

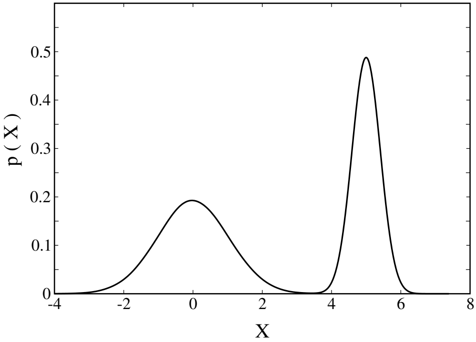
# Acknowledgements


- Sandia LDRD
- DOE BER
- DOE ASCR


Thank You!

# UQ methods are challenged by..

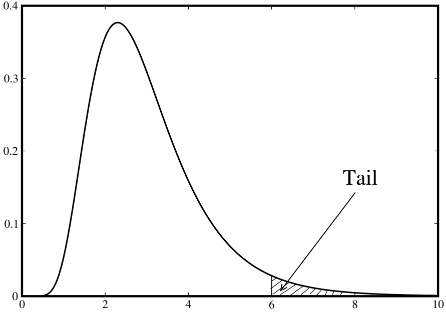
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity




*Smart* domain decomposition

# UQ methods are challenged by..

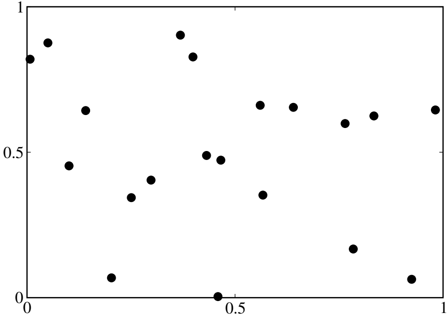
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity




*Custom* spectral bases

# UQ methods are challenged by..

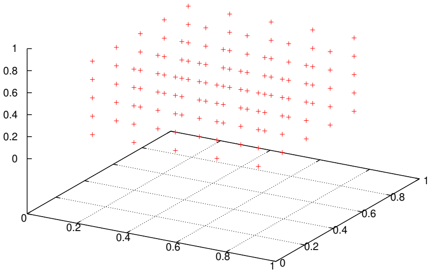
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity




Bayesian approach, O'Hagan's work

# UQ methods are challenged by..

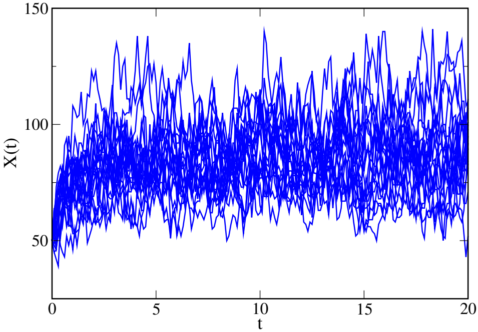
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity




*Sparse* quadrature/cubature

# UQ methods are challenged by..

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




Work with moments, Bayesian?

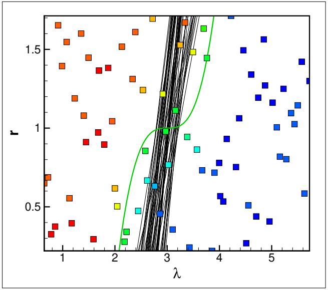
# Inference of Discontinuity - $3^{rd}$ order polynomial

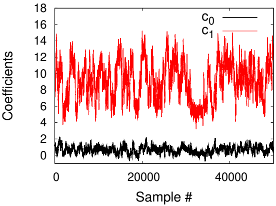
- Synthetic discontinuous data

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

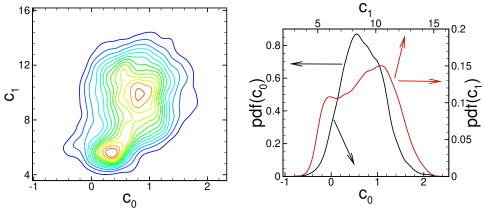
- Use straight lines to infer the discontinuity

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




**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:

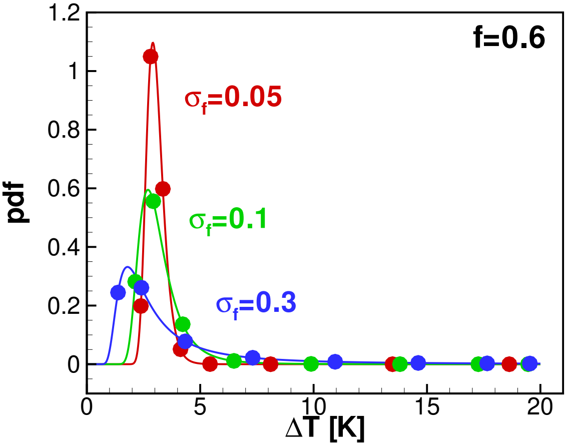
$$\hat{Z}(\lambda, r) = \int_C p(\boldsymbol{c}) Z_{\boldsymbol{c}}(\lambda, r) d\boldsymbol{c} = \int_{[0,1]^{K+1}} Z_{R^{-1}(\vec{\eta})}(\lambda, r) d\vec{\eta}$$

# As a conclusion..

- 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

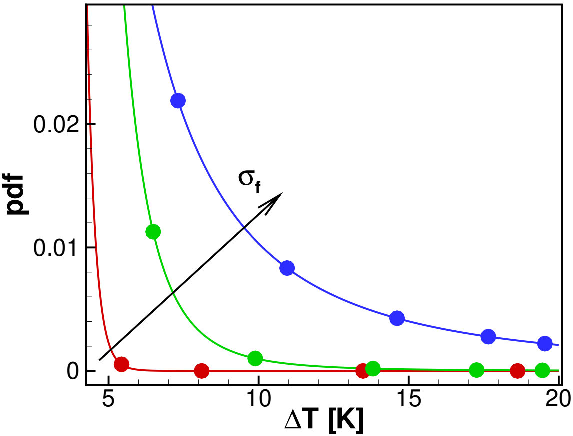
# Custom Basis' Quad Points Extend to the Tail

(pdf shape from Roe & Baker, Science 2007)



# Custom Basis' Quad Points Extend to the Tail

(pdf shape from Roe & Baker, Science 2007)



# Summary and Future Climate-related Research

- Used our expertise to deal with...
  - Non-linearities/bifurcations/bimodalities in climate modeling
  - Low-probability/high impact events
- Would like to leverage our expertise in spectral UQ /inverse problems to...
  - improve predictability of climate models
  - reduce uncertainties in source attribution ("surrogate" models, Bayesian experimental design)