SAND2011-4870C

Uncertainty Quantification given Discontinuities, Long-tailed Distributions, and Computationally Intensive Models

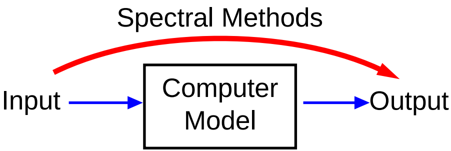
Habib Najm, Cosmin Safta, Khachik Sargsyan
Bert Debusschere

Sandia National Laboratories
Livermore, CA

7th International Congress on Industrial
and Applied Mathematics
Vancouver, BC

Thanks: DOE NNSA BER, SNL LDRD

UQ components and methods



Employ sampling-based methods:

- Non-intrusive Spectral Projection (NISP)
- Bayesian Inference

Objective

Tackle two of the challenges encountered in forward UQ:

1. Output observables exhibit discontinuities for smooth changes in the input parameters

2. Model predictions exhibit fat-tailed distributions.

Objective

Tackle two of the challenges encountered in forward UQ:

1. Output observables exhibit discontinuities for smooth changes in the input parameters
2. Model predictions exhibit fat-tailed distributions.

Expensive Computational Model

Objective

Tackle two of the challenges encountered in forward UQ:

1. Output observables exhibit discontinuities for smooth changes in the input parameters

Polynomial Chaos expansion represents random variables as a polynomials of standard random variables

- Truncated PCE: finite dimension n and order p

$$X(\boldsymbol{\lambda}(\boldsymbol{\eta})) \simeq \sum_{k=0}^{P} c_k \Psi_k(\boldsymbol{\eta})$$

with the number of terms $P + 1 = \frac{(n+p)!}{n!p!}$.

[Wiener, 1938; Ghanem & Spanos, 1991; Xiu & Karniadakis, 2002; Le Maître & Knio, 2010]

Polynomial Chaos expansion represents random variables as a polynomials of standard random variables

- Truncated PCE: finite dimension n and order p

$$\text{Output } X \longrightarrow X(\boldsymbol{\lambda}(\boldsymbol{\eta})) \simeq \sum_{k=0}^{P} c_k \Psi_k(\boldsymbol{\eta}) \longleftarrow \text{Input } \boldsymbol{\eta}$$

with the number of terms $P + 1 = \frac{(n+p)!}{n!p!}$.

- $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)$ standard i.i.d. r.v.
Ψ_k standard orthogonal polynomials
c_k spectral modes.

[Wiener, 1938; Ghanem & Spanos, 1991; Xiu & Karniadakis, 2002; Le Maître & Knio, 2010]

Polynomial Chaos expansion represents random variables as a polynomials of standard random variables

- Truncated PCE: finite dimension n and order p

$$\text{Output } X \longrightarrow X(\boldsymbol{\lambda}(\boldsymbol{\eta})) \simeq \sum_{k=0}^{P} c_k \Psi_k(\boldsymbol{\eta}) \longleftarrow \text{Input } \boldsymbol{\eta}$$

  with the number of terms $P + 1 = \frac{(n+p)!}{n!p!}$.

- $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)$ standard i.i.d. r.v.
  Ψ_k standard orthogonal polynomials
  c_k spectral modes.

- Most common standard Polynomial-Variable pairs:
  Gauss-Hermite, Legendre-Uniform

[Wiener, 1938; Ghanem & Spanos, 1991; Xiu & Karniadakis, 2002; Le Maître & Knio, 2010]

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: $\lambda_2 \approx p_{\boldsymbol{c}}(\lambda) = \sum_{k=0}^{K} c_k P_k(\lambda_1)$

- Approximation model:

$$\mathcal{M}_{\boldsymbol{c}} \equiv g(\lambda_1, \lambda_2) = m_L + (m_R - m_L)\frac{1 + \tanh\left(\alpha(\lambda_2 - p_{\boldsymbol{c}}(\lambda_1))\right)}{2}$$

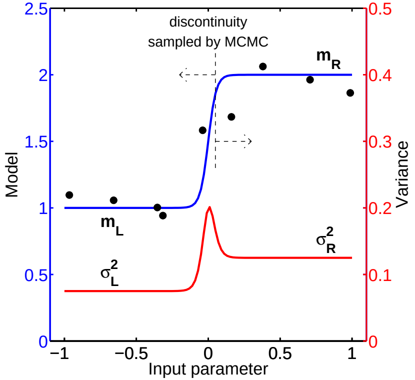
- Noise model postulated: $\sigma(\lambda_1, \lambda_2)$

- Likelihood function:

$$\log P(\mathcal{D}|\mathcal{M}_{\boldsymbol{c}}) = \sum_{i=1}^{N} \log P(z_i|\mathcal{M}_{\boldsymbol{c}}) = -\frac{1}{2}\sum_{i=1}^{N}\left(\frac{(z_i - g_i)^2}{\sigma_i^2} + \log\left(2\pi\sigma_i^2\right)\right)$$

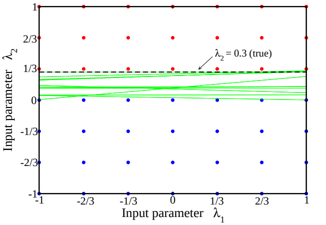
Bayesian Inference of the Location of Discontinuity

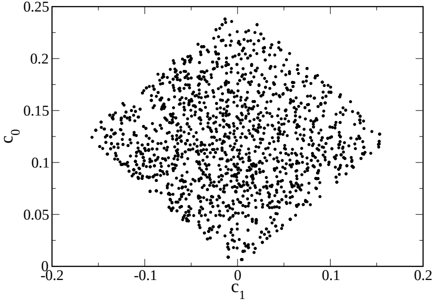
- Parameterize the discontinuity: $\lambda_2 \approx p_{\boldsymbol{c}}(\lambda) = \sum_{k=0}^{K} c_k P_k(\lambda_1)$
- Bayes' formula: $P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$

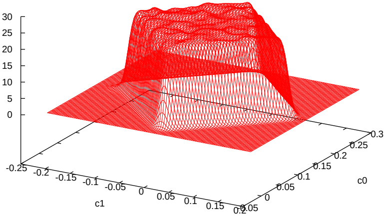


Discontinuity Detection - Highlights

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







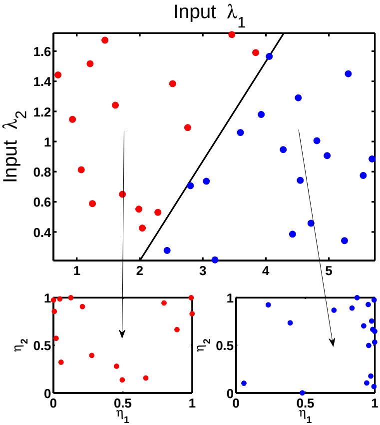
Discontinuity curve samples and their pdf

Parameter Domain Mapping via Rosenblatt Transformation

- Assume linear discontinuity
- Use Rosenblatt Transformation (RT) to map the pair of uncertain parameters (λ_1,λ_2) to i.i.d. uniform random variables η_1 and η_2:

$$\begin{aligned} \lambda_1 &= F_{\lambda}^{-1}(\eta_1), \\ \lambda_2 &= F_{\lambda_2|\lambda_1}^{-1}(\eta_2|\eta_1) \end{aligned}$$

- Apply the RT mapping to both sides of the discontinuity

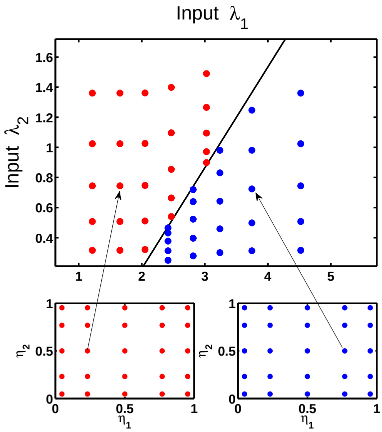


ROSENBLATT TRANSFORMATION: $(\lambda_1, \lambda_2) \rightarrow (\eta_1, \eta_2)$

Parameter Domain Mapping via Rosenblatt Transformation

- Assume linear discontinuity
- Use Rosenblatt Transformation (RT) to map the pair of uncertain parameters (λ_1,λ_2) to i.i.d. uniform random variables η_1 and η_2:

$$\begin{aligned} \lambda_1 &= F_{\lambda}^{-1}(\eta_1), \\ \lambda_2 &= F_{\lambda_2|\lambda_1}^{-1}(\eta_2|\eta_1) \end{aligned}$$



ROSENBLATT TRANSFORMATION: $(\lambda_1, \lambda_2) \rightarrow (\eta_1, \eta_2)$

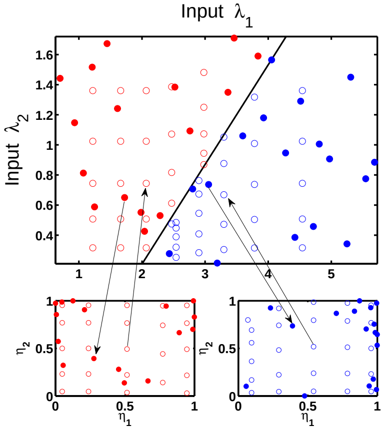
- Apply the RT mapping to both sides of the discontinuity

Parameter Domain Mapping via Rosenblatt Transformation

- Assume linear discontinuity
- Use Rosenblatt Transformation (RT) to map the pair of uncertain parameters (λ_1,λ_2) to i.i.d. uniform random variables η_1 and η_2:

$$\begin{aligned} \lambda_1 &= F_{\lambda}^{-1}(\eta_1), \\ \lambda_2 &= F_{\lambda_2|\lambda_1}^{-1}(\eta_2|\eta_1) \end{aligned}$$

- Apply the RT mapping to both sides of the discontinuity



ROSENBLATT TRANSFORMATION: $(\lambda_1, \lambda_2) \rightarrow (\eta_1, \eta_2)$

Compute the Polynomial Chaos Expansions

- PC expansion for the output observable $Z = f(\lambda)$

$$Z \simeq \sum_{k=0}^{K} c_k \Psi_k(\xi)$$

with

$$\langle \Psi_i(\xi)\Psi_j(\xi)\rangle \equiv \int \Psi_i(\xi)\Psi_j(\xi) p_\xi(\xi) d\xi = \delta_{ij}\langle \Psi_i(\xi)^2\rangle$$

- Spectral Projection

$$c_k = \frac{\langle f(\lambda(\xi))\Psi_k(\xi)\rangle}{\langle \Psi_k^2(\xi)\rangle}$$

- Bayesian Inference:

$$P(c_k|f(\lambda_i)) \approx P(f(\lambda_i)|c_k)P(z_k)$$

PC Expansion, Averaged Over Discontinuity Curves

- PC expansion for each discontinuity curve sample:

$$Z_{\boldsymbol{c}}^{L,R}(\vec{\lambda}) = \tilde{Z}_{\boldsymbol{c}}(\vec{\eta}) = \sum_{p=0}^{P} c_p \Psi_p^{(2)}(\vec{\eta})$$

- Model expansion depends on the parameter location:

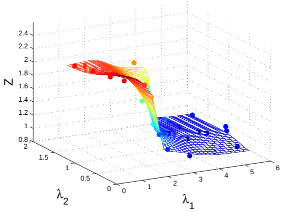
$$Z_{\boldsymbol{c}}(\vec{\lambda}) = \begin{cases} Z_{\boldsymbol{c}}^{L}(\vec{\lambda}) & \text{if } (\vec{\lambda}) \in D_L \\ Z_{\boldsymbol{c}}^{R}(\vec{\lambda}) & \text{if } (\vec{\lambda}) \in D_R \end{cases}.$$

- Average over all PC expansions via RT:

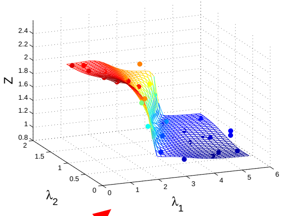
$$\hat{Z}(\vec{\lambda}) = \int_C p(\boldsymbol{c}) Z_{\boldsymbol{c}}(\vec{\lambda}) d\boldsymbol{c} = \int_{[0,1]^{K+1}} Z_{R^{-1}(\vec{\eta})}(\vec{\lambda}) d\vec{\eta}$$

Discontinuous Data Represented Well with the Averaged PC

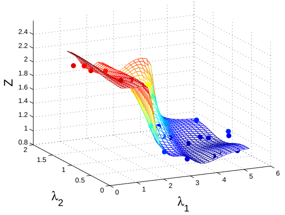
PCE Coefficients via Bayesian Inference

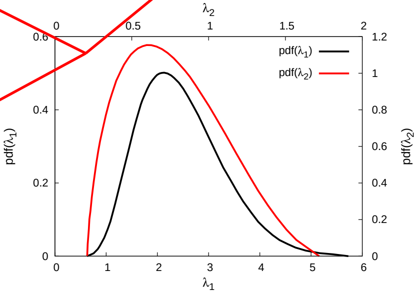


PCE Coefficients via Hybrid Approach



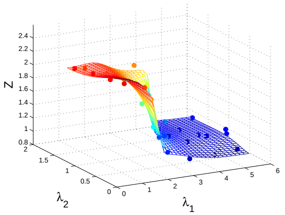
PCE Coefficients via Quadrature Projection



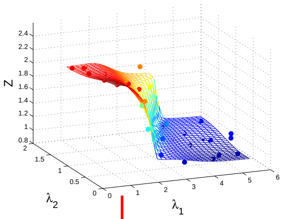


Discontinuous Data Represented Well with the Averaged PC

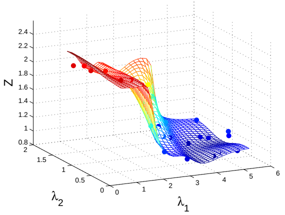
PCE Coefficients via Bayesian Inference

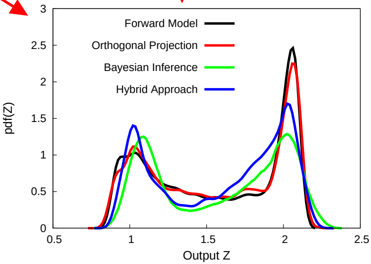


PCE Coefficients via Hybrid Approach



PCE Coefficients via Quadrature Projection



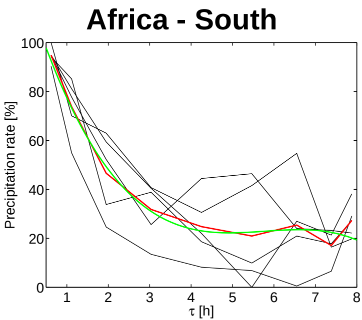


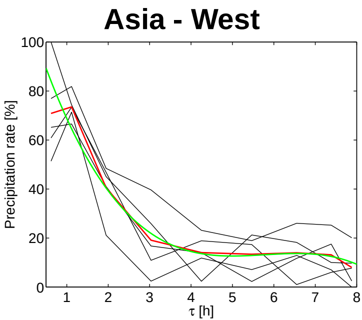
Objective

Tackle two of the challenges encountered in forward UQ:

2. Model predictions exhibit fat-tailed distributions.

Precipitation Data from Climate Simulations



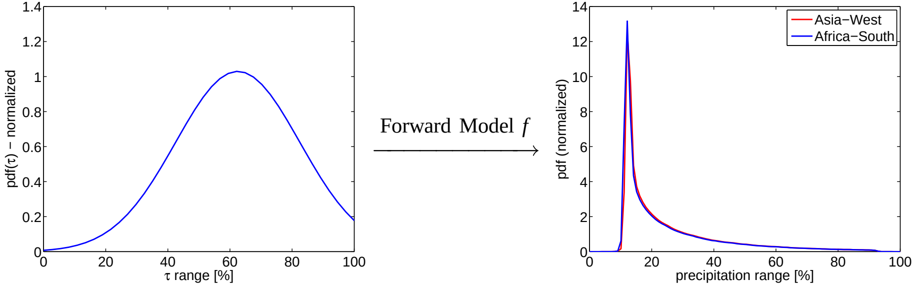


- Precipitation rate scaled between the minimum and maximum encountered in the simulation
- Input parameter τ - consumption rate of CAPE (convective available potential energy)

Black lines - 2 year averages; *Red* lines - 10 year averages; *Green* lines - 3^{rd}-order PC expansions.

(*data courtesy of Mike Levy & Mark Taylor, Sandia National Labs*)

Forward UQ: Input Parameter PDF $\longrightarrow$ Output Observable PDF



- Compute the probability that average precipitation exceeds a certain amount:

$$P(precip. > p_r) = \int_{\tau : f(\tau) > p_r} \text{pdf}(\tau) d\tau$$

Polynomial Chaos Expansions and Galerkin Projection

- PC expansion for the output observable $Z = f(\lambda)$

$$Z \simeq \sum_{k=0}^{K} c_k \Psi_k(\xi)$$

with

$$\langle \Psi_i(\xi)\Psi_j(\xi)\rangle \equiv \int \Psi_i(\xi)\Psi_j(\xi)p_\xi(\xi)d\xi = \delta_{ij}\langle \Psi_i(\xi)^2\rangle$$

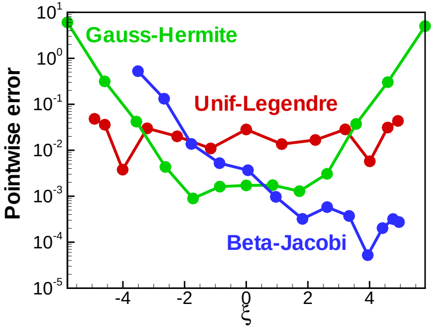
- Galerkin (orthogonal) projection

$$c_k = \frac{\langle f(\lambda(\xi))\Psi_k(\xi)\rangle}{\langle \Psi_k^2(\xi)\rangle}$$

is weighted-L_2 optimal, i.e. it minimizes

$$\int \left| f(\lambda(\xi)) - \sum_{k=0}^{K} c_k \Psi_k(\xi) \right|^2 p_\xi(\xi)d\xi$$

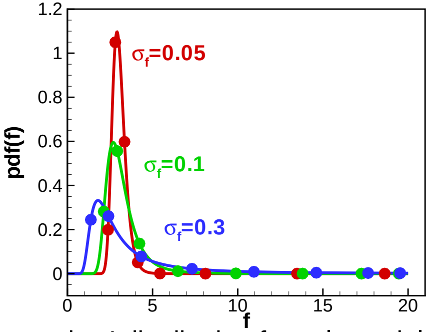
Conventional Basis Functions



- Legendre basis : error is 'independent' of position
- Hermite basis : error is worse in "tails", away from the origin
- Jacobi basis: error is small in desired region, i.e. it is controllable!

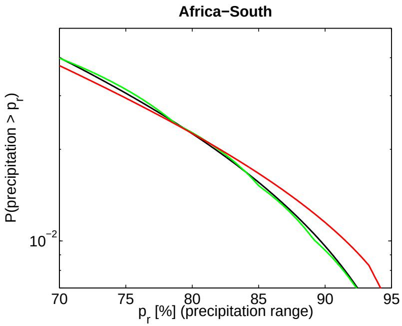
Custom Basis Functions

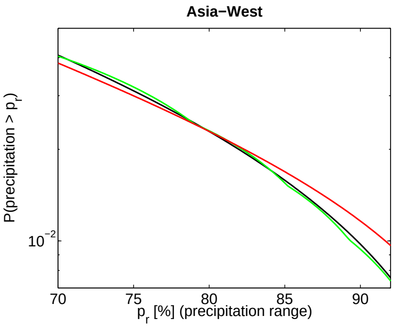
- Design custom polynomials that are orthogonal with respect to fat tailed distributions to get a better accuracy in the tail region.
  - Use Stieltjes or Chebyshev's algorithms. Algorithms are ill-conditioned $\rightarrow$ use arbitrary precision arithmetic software.



(Quadrature points' distribution for polynomials orthogonal w.r.t. truncated log-normal pdf.)

"Tail" Probabilities Based on PC Basis Surrogates





- *Black* lines - "Exact" values; *Red* lines - Hermite PC basis (9^{th} order); *Green* lines - Custom PC basis (9^{th} order).
- The set of quadrature points corresponding to the custom PDF have a better coverage of the distribution's tail compared to the set corresponding to a Gaussian PDF.

Summary and Future Work

- *Nonlinearities, Bifurcations, Bimodalities*
  - Probabilistic detection of discontinuities followed by domain mapping and polynomial chaos expansions to construct model "surrogates"
  - Extend this approach to incorporate optimal experimental design, i.e. find parameter values at which the model should be simulated to give maximum information

Summary and Future Work

- *Nonlinearities, Bifurcations, Bimodalities*
  - Probabilistic detection of discontinuities followed by domain mapping and polynomial chaos expansions to construct model “surrogates”
  - Extend this approach to incorporate optimal experimental design, i.e. find parameter values at which the model should be simulated to give maximum information
- *Tail regions*
  - Construct custom spectral basis based on “expected” shape of the computer model output to improve convergence of the spectral expansion.
  - Extend this methodology to multi-dimensional parameter dependencies.
  - Develop surrogate models as mixed PC expansions: accurate both near the mean as well as in the tail regions.