Uncertainty Quantification Methodologies for Climate Model Data with Discontinuities

Khachik Sargsyan, Cosmin Safta,
Bert Debusschere, Habib Najm

Sandia National Laboratories
Livermore, CA

SIAM Annual Meeting
Pittsburgh, PA
July 14, 2010

Uncertainties in Climate Modeling

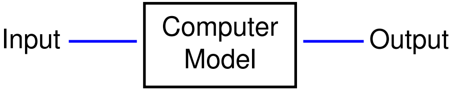
- Uncertainty sources
  - Parameter uncertainty
    - Model parameters
    - Initial/boundary conditions
    - Model geometry/structure
  - Model/structural uncertainty
    - Unknown physics
    - Reduced order models
  - Scenario uncertainty
    - Policy restrictions
    - Technology improvement
  -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

Uncertainties in Climate Modeling

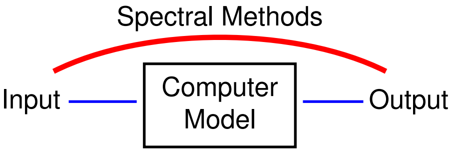
- Uncertainty sources
  - Parameter uncertainty
    - Model parameters
    - Initial/boundary conditions
    - Model geometry/structure
  - Model/structural uncertainty
    - Unknown physics
    - Reduced order models
  - Scenario uncertainty
    - Policy restrictions
    - Technology improvement
  -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

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

UQ components and methods

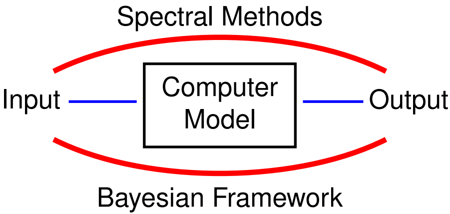


UQ components and methods



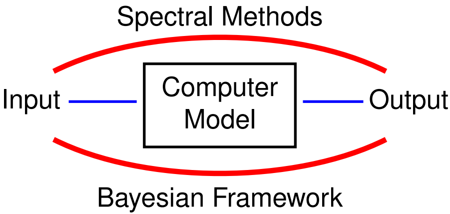
- Sensitivity analysis
  - Small parameter perturbations
- Predictability assessment
  - Larger parameter uncertainties

UQ components and methods



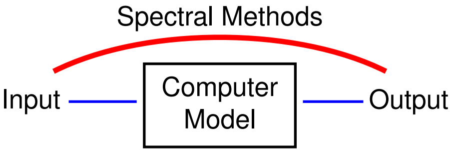
- Sensitivity analysis
  - Small parameter perturbations
- Predictability assessment
  - Larger parameter uncertainties
- Parameter estimation
  - Inverse problems

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)

Polynomial Chaos expansion represents any random variable as a polynomial of a standard random variable

- 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 $P+1 = \frac{(n+p)!}{n!p!}$.

- $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)$ standard i.i.d. r.v.
  Ψ_k standard orthogonal polynomials
  c_k spectral modes.
- Most common standard Polynomial-Variable pairs:
  (continuous) Gauss-Hermite, Legendre-Uniform,
  (discrete) Poisson-Charlier.

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

Polynomial Chaos expansion represents any random variable as a polynomial of a standard random variable

- 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 $P + 1 = \frac{(n+p)!}{n!p!}$.

- $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)$ standard i.i.d. r.v.
  Ψ_k standard orthogonal polynomials
  c_k spectral modes.
- Most common standard Polynomial-Variable pairs: (continuous) Gauss-Hermite, Legendre-Uniform, (discrete) Poisson-Charlier.

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

Polynomial Chaos expansion represents any random variable as a polynomial of a standard random variable

- Truncated PCE: finite dimension n and order p

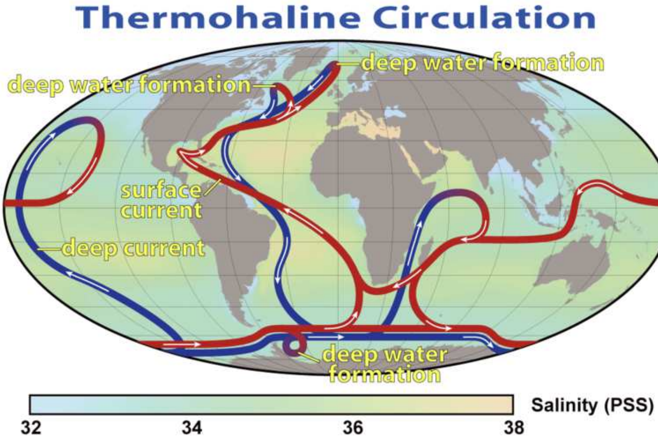
$$X(\boldsymbol{\lambda}(\boldsymbol{\eta})) \simeq \sum_{k=0}^{P} c_k \Psi_k(\boldsymbol{\eta})$$

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

- $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)$ standard i.i.d. r.v.
Ψ_k standard orthogonal polynomials
c_k spectral modes.
- Most common standard Polynomial-Variable pairs:
(continuous) Gauss-Hermite, Legendre-Uniform,
(discrete) Poisson-Charlier.

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

Meridional Overturning Circulation



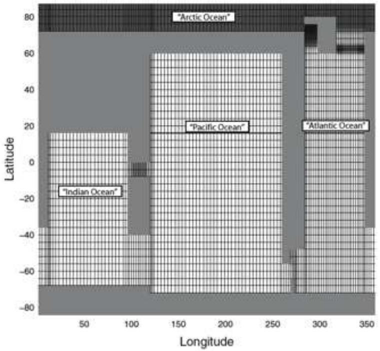
MOC transports heat from warm to cold regions

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

- Computational model
  - 3D Ocean general circulation model
  - Zonally-averaged atmospheric model
  - Thermodynamic sea-ice model
  - Simplified models for river runoff

- Input parameters
  - Rate of CO_2 increase (r)
  - Climate sensitivity (λ)

- Output observable
  - Overturning streamfunction (Z)

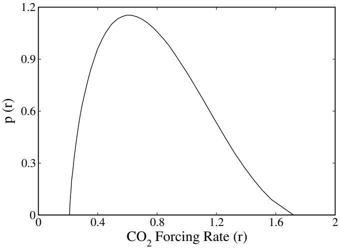


Webster et al - J. Environ. Syst. 31: 39-59, 2007

- Computational model
  - 3D Ocean general circulation model
  - Zonally-averaged atmospheric model
  - Thermodynamic sea-ice model
  - Simplified models for river runoff

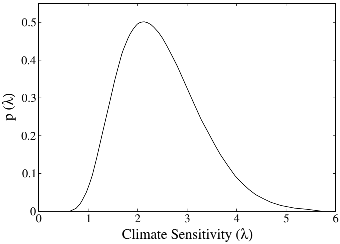
- Input parameters
  - Rate of CO_2 increase (r)
  - Climate sensitivity (λ)

- Output observable
  - Overturning streamfunction (Z)



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

- Computational model
  - 3D Ocean general circulation model
  - Zonally-averaged atmospheric model
  - Thermodynamic sea-ice model
  - Simplified models for river runoff
- Input parameters
  - Rate of CO_2 increase (r)
  - Climate sensitivity (λ)
- Output observable
  - Overturning streamfunction (Z)

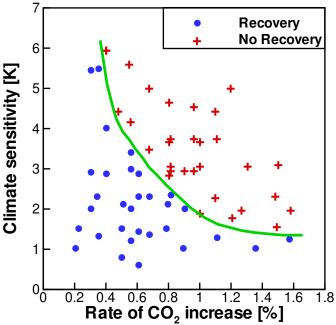


Webster et al - J. Environ. Syst. 31: 39-59, 2007

- Computational model
  - 3D Ocean general circulation model
  - Zonally-averaged atmospheric model
  - Thermodynamic sea-ice model
  - Simplified models for river runoff

- Input parameters
  - Rate of CO_2 increase (r)
  - Climate sensitivity (λ)

- Output observable
  - Overturning streamfunction (Z)

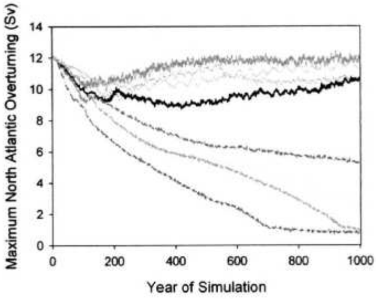


Webster et al - J. Environ. Syst. 31: 39-59, 2007

- Computational model
  - 3D Ocean general circulation model
  - Zonally-averaged atmospheric model
  - Thermodynamic sea-ice model
  - Simplified models for river runoff

- Input parameters
  - Rate of CO_2 increase (r)
  - Climate sensitivity (λ)

- Output observable
  - Overturning streamfunction (Z)

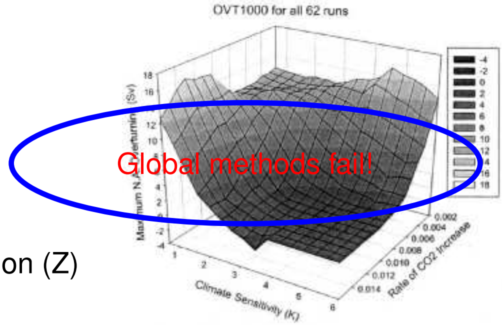


Webster et al - J. Environ. Syst. 31: 39-59, 2007

- Computational model
  - 3D Ocean general circulation model
  - Zonally-averaged atmospheric model
  - Thermodynamic sea-ice model
  - Simplified models for river runoff

- 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

- 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:

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

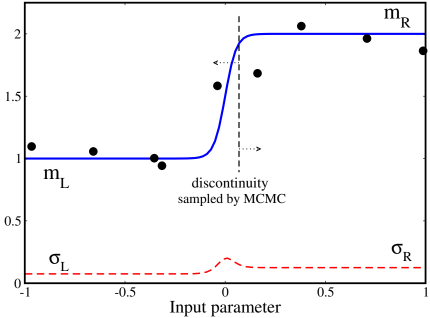
- Noise model postulated: $\sigma(\lambda, r)$
- Likelihood function:

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

Bayesian Inference of the Location of Discontinuity

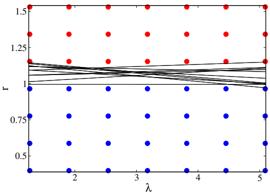
- Parameterize the discontinuity: $r \approx p_{\boldsymbol{c}}(\lambda) = \sum_{k=0}^{K} c_k P_k(\lambda)$

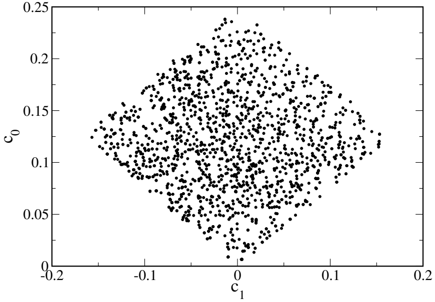
- Bayes' formula: $P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}$

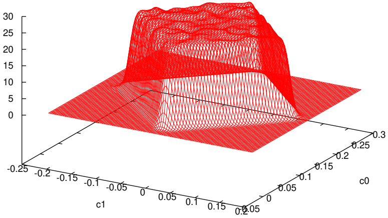


Highlights

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







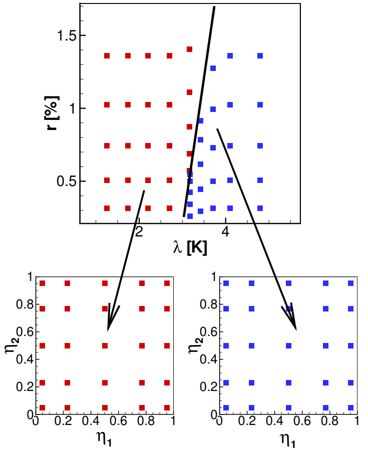
Discontinuity curve samples and their pdf

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:

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

- Apply the RT mapping to both sides of the discontinuity



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

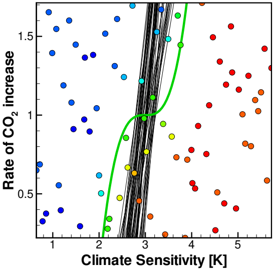
Inference of Discontinuity - 3^{rd} order polynomial

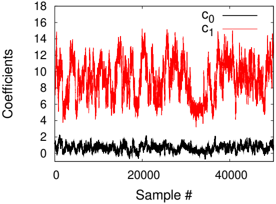
- Synthetic discontinuous data

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

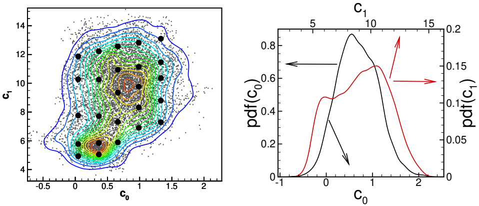
- Use straight lines to infer the discontinuity

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





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:

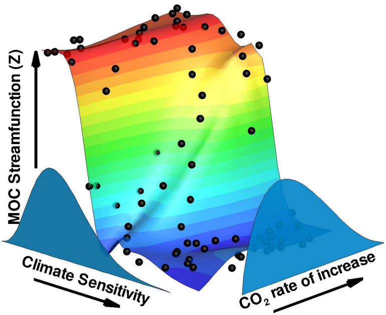
$$Z_{\boldsymbol{c}}(\lambda, r) = \begin{cases} Z_{\boldsymbol{c}}^{L}(\lambda, r) & \text{if } (\lambda, r) \in D_L \\ Z_{\boldsymbol{c}}^{R}(\lambda, r) & \text{if } (\lambda, r) \in D_R \end{cases}.$$

- Average over all PC expansions via RT:

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

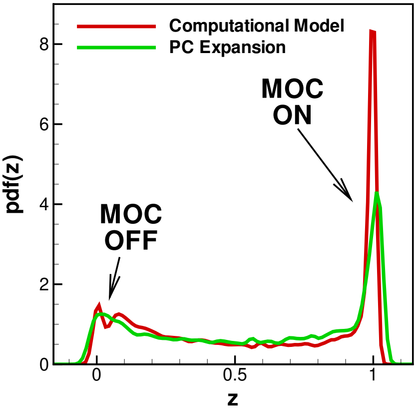
Discontinuous data represented well with the averaged PC

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

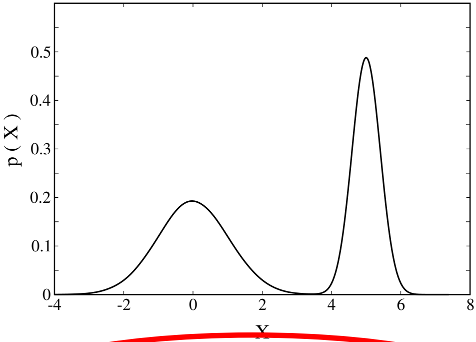
- A methodology for uncertainty quantification in climate models with limited data and discontinuities was proposed:
  - Bayesian approach to detect and parameterize the discontinuity as well as the uncertainty associated with it.
  - Rosenblatt transformation maps each of the irregular domains to rectangular 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.

Work in progress

- Bring in real climate model data.
- Still prohibitively many model runs required: possibly give up orthogonal projection in favor of Bayesian inference.
- Gaussian process emulation to implement uncertainties due to the lack of knowledge at non-sampled points.
- Experimental design: inform climate modelers on the optimal parameter sets to run simulations.

UQ methods are challenged by..

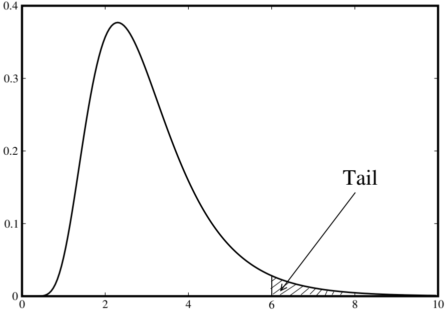
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity



Smart domain decomposition

UQ methods are challenged by..

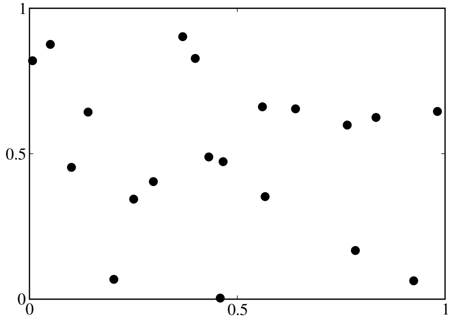
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity



Custom spectral bases

UQ methods are challenged by..

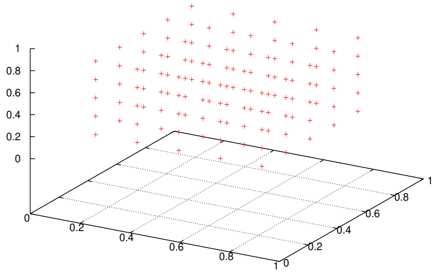
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity



Bayesian approach, O'Hagan's work

UQ methods are challenged by..

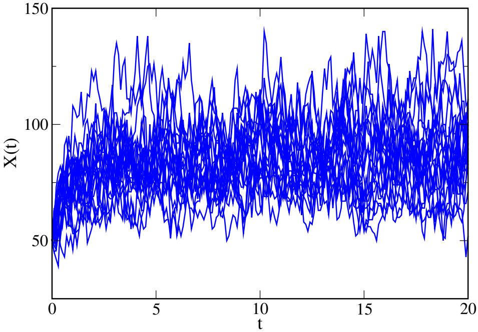
- Nonlinearities, Bifurcations, Bimodalities
- Tail regions
- Limited data
- Curse of dimensionality
- Intrinsic stochasticity



Sparse quadrature/cubature

UQ methods are challenged by..

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



Work with moments, Bayesian?

Acknowledgements

- Sandia's internal
  Laboratory Directed Research & Development

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

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's internal
  Laboratory Directed Research & Development

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

Thank You!

Literature

- N. Wiener, “Homogeneous Chaos”, *American Journal of Mathematics*, 60(4), pp. 897-936, 1938.
- R. Ghanem and P. Spanos, “Stochastic Finite Elements: a Spectral Approach”, Springer, 1991.
- D. Xiu and G.E. Karniadakis, “The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations”, *SIAM J. Sci. Comput.*, 24(2), pp. 619-644, 2002.
- O. Le Maître and O. Knio, “Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics”, Springer, 2010.

- M. Rosenblatt, “Remarks on a multivariate transformation”, *Annals of Mathematical Statistics*, 23(3), pp. 470-472, 1952.
- M. Webster, J. Scott, A. Sokolov, S. Dutkiewicz, and P. Stone, “Estimating probability distributions from complex models with bifurcations: The case of ocean circulation collapse”, *Journal of Environmental Systems*, 31(1), pp. 39-59, 2007.

- K. Sargsyan, C. Safta, B. Debusschere, and H. Najm, “Uncertainty Quantification in the Presence of Limited Climate Model Data with Discontinuities”, *proceedings of IEEE International Conference on Data Mining*, pp. 241-247, 2009.