Uncertainty Quantification given Discontinuous Climate Model Response and a Limited Number of Model Runs

Poster ID: GC23F-0973

KHACHIK SARGSYAN, COSMIN SAFTA, BERT DEBUSSCHERE, HABIB NAJM
Sandia National Laboratories, Livermore, CA

Uncertainty Quantification challenges in complex models

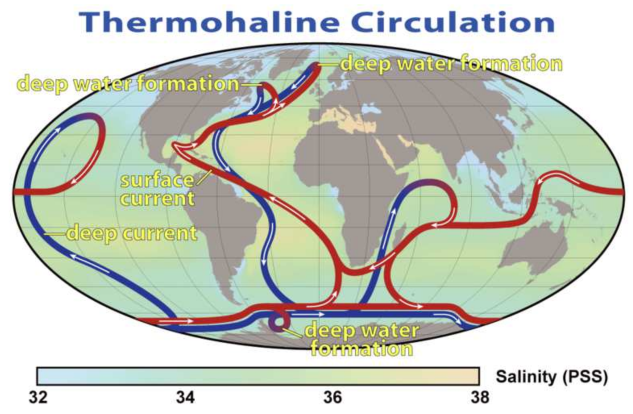
- Limited number of model simulations
- Discontinuities/nonlinearities in model response

Spectral methods for Uncertainty Quantification with global, smooth bases are challenged by discontinuities in model response. Domain decomposition reduces the impact of nonlinearities and discontinuities. However, while gaining more smoothness in each subdomain, the current domain refinement methods require prohibitively many simulations. Therefore, discontinuity detection *up front* provides huge improvement to the current methodologies.

Two-step approach

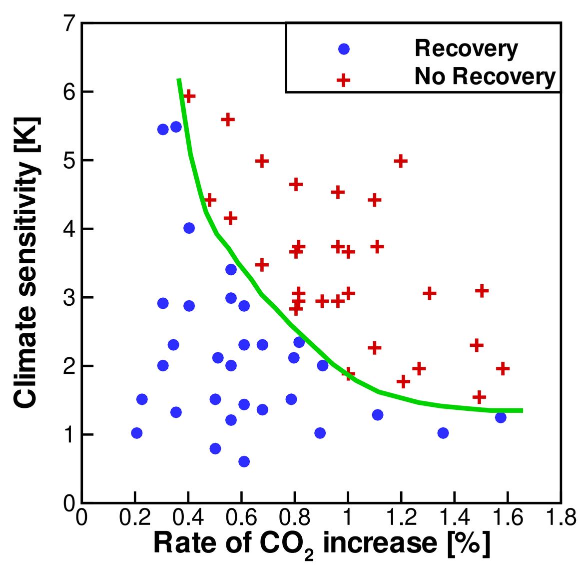
1. Detect the discontinuity location
2. Obtain spectral expansion on each side

Motivational problem: Global Conveyor Belt



The *Meridional Overturning Circulation* (MOC) is one of the most discussed environmental phenomena that can potentially collapse as a result of increased greenhouse gas concentrations. In the MOC, warmer surface currents flow from the tropics northward, towards the North Atlantic and then cool down due to the heat exchange with the cooler atmosphere and the ice. As the water becomes dense enough, it sinks to larger depths and becomes a part of the global deep water formation, the "global conveyor belt", returning to the Southern Ocean.

MOC Streamfunction as a function of CO_2 forcing rate (r) and climate sensitivity (λ)



Consider the uncertainty in the MOC as a function of two uncertain parameters λ and r, each characterized by a probability distribution function (PDF).

- Certain (λ, r) parameter pairs lead to MOC shut-off
- MOC streamfunction $Z(\lambda, r)$ has a sharp gradient across a 'discontinuity' curve $r = \tilde{r}(\lambda)$
- Global methods fail to properly obtain the response surface for $Z(\lambda, r)$

For details, see Webster *et al.* [5]. The data is used with permission from the publisher, Baywood Publishing Co., Inc.

Output uncertainty quantification via PC expansions

To propagate input uncertainties to output distributions, Polynomial Chaos (PC) spectral expansions are used; see Ghanem and Spanos [1].

PC expansion of the input parameters via the inverses of their cumulative distribution functions (CDF)

$$\lambda = F_\lambda^{-1}(\eta_1) = \sum_{k=0}^{K} \lambda_k \Psi_k(\eta_1),$$

$$r = F_r^{-1}(\eta_2) = \sum_{k=0}^{K} r_k \Psi_k(\eta_2)$$

with Legendre polynomials $\Psi_k(\cdot)$ of independent, Uniform[0,1] random variables η_1, η_2.

Bivariate PC expansion for the output

$$Z(\lambda, r) = \sum_{p=0}^{P} z_p \Psi_p(\eta_1, \eta_2)$$

can be found by a Galerkin (orthogonal) projection

$$z_p = \frac{\langle Z(\lambda(\eta_1), r(\eta_2)) \Psi_p(\eta_1, \eta_2)\rangle}{\langle \Psi_p^2(\eta_1, \eta_2)\rangle}$$

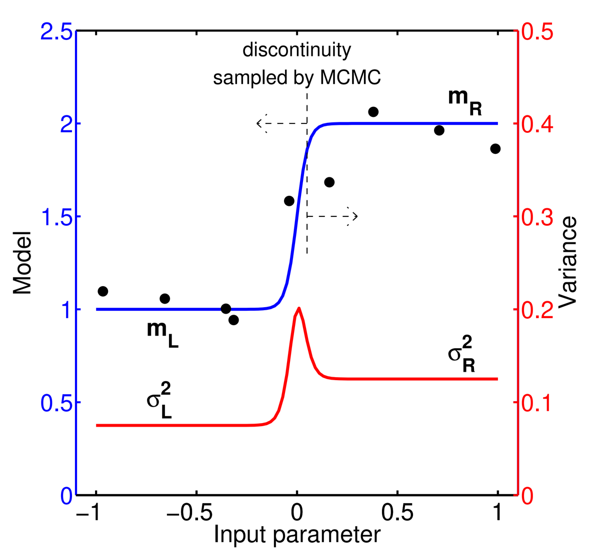
Global methods fail, since the output $Z(\lambda, r)$ has a steep gradient or discontinuity across a curve $r = \tilde{r}(\lambda)$

Bayesian inference of the discontinuity curve

- Parameterize the discontinuity curve, e.g.

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

- Approximation model $\mathcal{M}_{\boldsymbol{\gamma}}$ with parameters $\boldsymbol{\gamma} = (\boldsymbol{c}, m_L, m_R, \alpha)$:



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

- Statistical noise model assumes larger discrepancy near the discontinuity:

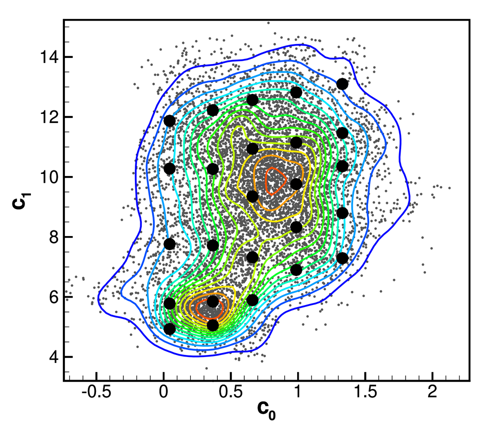
$$\sigma^2(\lambda, r) = \sigma_L^2 \left(\frac{1 - \tanh(\alpha(r - p_{\boldsymbol{c}}(\lambda)))}{2}\right)^2 + \sigma_R^2 \left(\frac{1 + \tanh(\alpha(r - p_{\boldsymbol{c}}(\lambda)))}{2}\right)^2 + \frac{\beta}{\cosh^4(\alpha(r - p_{\boldsymbol{c}}(\lambda)))}$$

- Likelihood function:

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

- Bayes' formula:

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



Bayesian inference leads to posterior distribution for *all* parameters $\boldsymbol{\gamma}$. Marginalize over *hyperparameters* $(m_L, m_R, \alpha, \sigma_L, \sigma_R, \beta)$ to obtain posterior distribution on the parameters $\boldsymbol{c}$ of the discontinuity curve.

Parameter domain mapping via Rosenblatt transformation

- Rosenblatt transformation [2] (RT) to map the pair of uncertain parameters (r, λ) to i.i.d. Uniform[0,1] random variables η_1 and η_2:

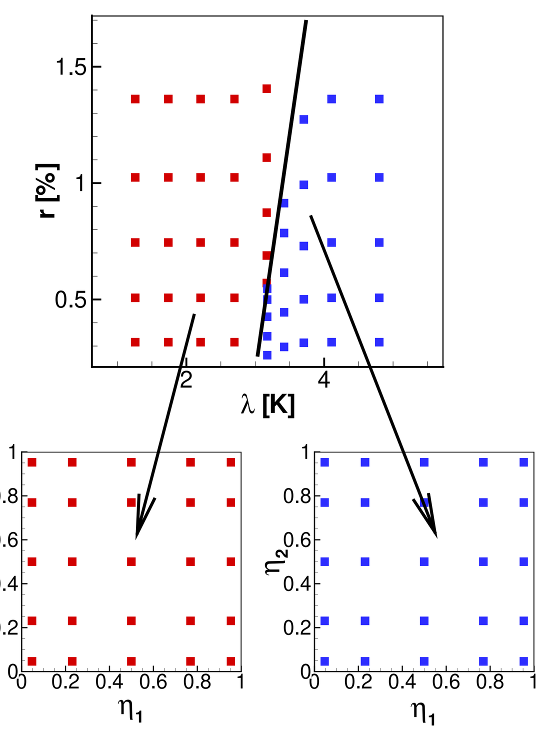
$$\lambda = F_\lambda^{-1}(\eta_1), \quad r = F_{r|\lambda}^{-1}(\eta_2|\eta_1)$$

- Apply the RT mapping to both sides of the discontinuity to obtain PC expansion of the model output (employing Galerkin projection or Bayesian inference, see [4]):

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

- Model expansion depends on the parameter location with respect to the discontinuity:

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



Averaging PC expansions with respect to sample curves

The expectation of the representation $Z_{\boldsymbol{c}}(\lambda, r)$ with respect to the K-variate probability distribution function $p(\boldsymbol{c})$ of the coefficient vector $\boldsymbol{c}$:

$$\hat{Z}(\lambda, r) = \int_C p(\boldsymbol{c}) Z_{\boldsymbol{c}}(\lambda, r) d\boldsymbol{c}.$$

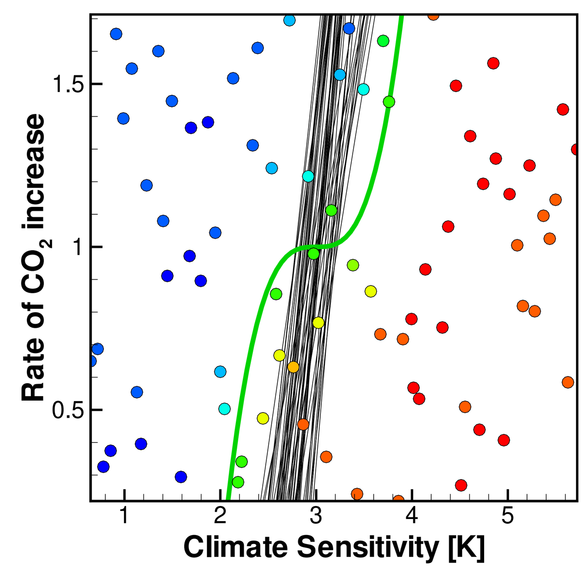
We will use the Rosenblatt transformation $\boldsymbol{\nu} = R(\boldsymbol{c})$ to map the unknown domain of integration C to a rectangular domain $[0, 1]^K$.

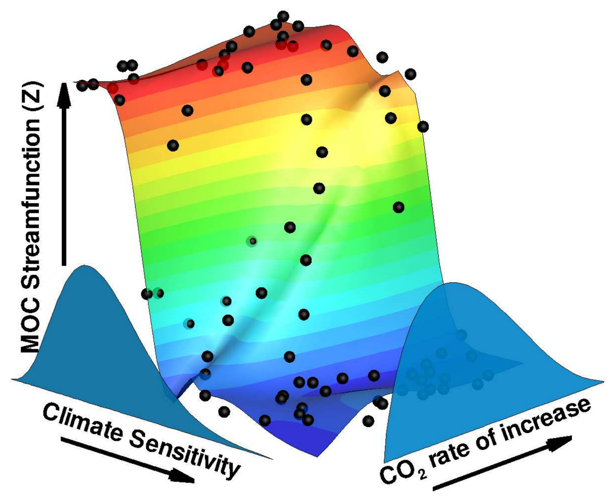
The vector of random variables $\boldsymbol{\nu}$ defined with the help of the conditional distributions of $\boldsymbol{c}$ has independent Uniform[0,1]-distributed components. Then $\hat{Z}(\lambda, r)$ can be rewritten as

$$\hat{Z}(\lambda, r) = \int_{[0,1]^K} Z_{R^{-1}(\boldsymbol{\nu})}(\lambda, r) d\boldsymbol{\nu}.$$

The latter integral can be taken by quadrature rules

$$\hat{Z}(\lambda, r) \approx \sum_{\boldsymbol{\nu}^*} Z_{R^{-1}(\boldsymbol{\nu}^*)}(\lambda, r) w^*.$$

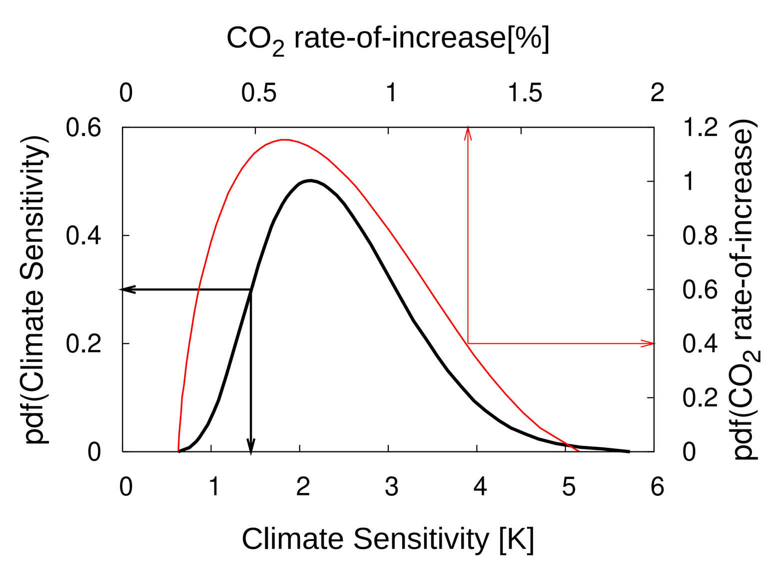


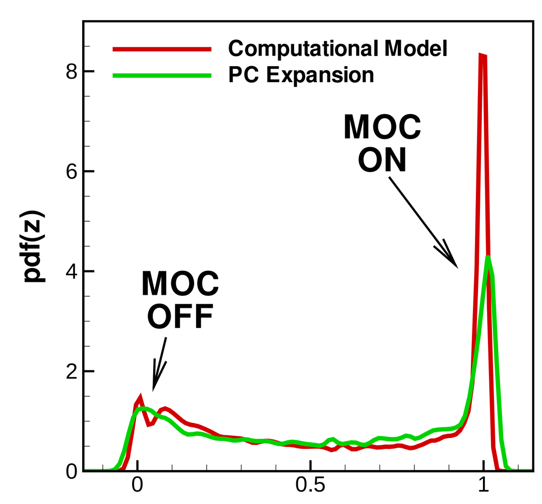


Compare densities from the averaged PC expansion with analytical model

- Sample the parameter space of climate sensitivity and CO_2 forcing rate according to input PDF
- Evaluate the MOC streamfunction Z using both the forward model and the averaged PC
- Obtain the PDF of Z by a smooth kernel density estimation
- The MOC exhibits a bimodal behavior and the averaged PC expansion is in good agreement with the analytical model; details can be found in Sargsyan *et al.* [3, 4].

Input PDF ⟶ Output PDF





Ongoing and Future Work

- Extend this approach to incorporate optimal experimental design, i.e. find parameter values at which the model should be simulated to give maximum information.
- Demonstrate the methodology with data from climate research groups.

References

[1] R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, New York, 1991.

[2] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3):470 – 472, 1952.

[3] K. Sargsyan, C. Safta, B. Debusschere, and H. Najm. Uncertainty quantification in the presence of limited climate model data with discontinuities. In *ICDMW '09: Proceedings of the 2009 IEEE International Conference on Data Mining Workshops*, pages 241–247, 2009.

[4] K. Sargsyan, C. Safta, B. Debusschere, and H. Najm. Uncertainty quantification given discontinuous model response and a limited number of model runs. *SIAM Journal on Scientific Computing*, 2010, submitted.

[5] M.D. 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):1–21, 2004.

Sandia National Laboratories

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.