SAND2009-4937C

Uncertainty Quantification in the Presence of Limited Climate Model Data with Discontinuities

Submitted for Blind Review

Abstract

Uncertainty quantification in climate models is challenged by the sparsity of the available climate data due to the high computational cost of these model runs. Another feature that prevents classical uncertainty analyses from being easily applicable is the bifurcative behavior in the climate data with respect to certain parameters. A typical example is the Meridional Overturning Circulation in the Atlantic Ocean. The maximum overturning stream function exhibits discontinuity across a curve in the space of two uncertain parameters, namely climate sensitivity and CO_2 *forcing. We develop a methodology that performs uncertainty quantification in this context in the presence of limited data.*

1 Introduction

One remarkable feature of climate models is that global climate phenomena potentially exhibit discontinuous behavior with respect to various parameters. Specifically, consider the *Meridional Overturning Circulation* (MOC), as one of the most discussed environmental phenomena that can potentially collapse [17] 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” [1], returning to the Southern Ocean. The MOC plays an extremely important role in the northward heat transport in the Atlantic Ocean [3] and its weakening or collapse would cause major climatic change and, consequently, economic disruptions. Both local and global effects on the climate and the environment are significant. Experts argue [25] that cooling of the North Atlantic region could lead to the weakening of the hydrological cycle due to reduced evaporation. It is also likely that the frequency of extreme cold events would increase as the MOC weakens. Globally, the cooling of the Northern Hemisphere and warming of the Southern Hemisphere will influence large-scale atmospheric circulation (e.g tropical weather patterns and intensity, hurricane activity, etc.).

It is self-evident therefore that the assessment of fidelity of such high-consequence predictions as MOC weakening or collapse, is of crucial importance. Because of limited observations and the difficulties associated with high-resolution modeling, the current state of predictive power is quite restricted, necessitating the need to account for uncertainty in climate models and associated model parameters. Uncertainty quantification (UQ) approaches have been used in this context, however, hitherto used interpolation-based approaches require a prohibitively large number of samples, are extremely sensitive to the choice of sampling points, and cannot accurately handle discontinuities. We develop a methodology that allows applying the UQ techniques in the presence of limited data, with respect to two model parameters.

Many studies suggest that the atmospheric CO_2 increase will damage the MOC significantly, and a potential collapse of the circulation is possible. The rate of CO_2 increase, denoted by r, is one of the most important parameters in MOC modeling, together with the climate sensitivity (λ) parameter, defined as the equilibrium change in the temperature that corresponds to a hypothetical doubling of the atmospheric CO_2 concentration. Evidence from paleoclimatic reconstructions [11] and from recent model simulations reveals that MOC behavior exhibits a bifurcation with respect to the pair of parameters (r, λ). Namely, the MOC collapses as the parameters cross a curve $r = \tilde{r}(\lambda)$ in the parametric (r, λ) space. In recent papers [18, 22], a detailed analysis of the MOC dependence on the two parameters r and λ has been carried out. The authors use a forward model that consists of a three-dimensional global circulation model (GCM) coupled with both a zonally averaged atmospheric and a thermodynamic sea-ice model. This joint model is formally categorized as an Earth Model of Intermediate Complexity (EMIC). A single run of a time span of a couple of centuries can take several weeks to simulate on a supercomputer. Webster *et al.* [22] explore the uncertainty in the MOC as a function of two uncertain parameters r and λ, each characterized by a probability distribution function. However, the computational cost of the forward model restricts the uncertainty quantification analysis, since only a small number of sampling runs are plausible. By obtaining simulation data on maximum overturning from a small number of hypothetical pairs of parameters, the response surface is then constructed using the Deterministic Equivalent Modeling Method (DEMM) [21, 23]. It relies on the expansion of the response in terms of global orthogonal polynomials, and therefore is bound to fail in case of bifurcations, since no linear combination of polynomials can properly represent a discontinuous response function. Although eventually the authors achieved a reasonable representation of the response surface by using ad-hoc information about the structure of the response, they recognized the necessity of determining the response surface characteristics in a more general manner, as well as the importance of identifying optimal sampling points *prior* to the simulations.

Advanced uncertainty quantification algorithms have been developed in the literature, using basis functions with *local* support to adequately handle systems with bifurcation behavior [7–9]. These methods were used to characterize problems with bifurcative convective [10] and chemical [13] structure. A predictability analysis was performed in [16], for stochastic reaction networks that are modeled as continuous-time, discrete-state Markov processes. Bayesian inference was used to obtain spectral expansions for state variables that are consistent with the full range of the uncertain parameters. An adaptive domain refinement algorithm that leads to a multi-domain spectral expansion with basis functions of local support represents the bifurcation structure efficiently. However, limited data prohibits the use of the local basis refinement methods.

We developed a methodology to detect and properly parameterize discontinuity locations in the presence of limited data. Coupled with random domain mapping strategy, our technique enables propagation of the uncertainty from climate model parameters to model outputs. The approach relies on an efficient representation of uncertain quantities using spectral expansions of random variables/fields with bases of compact support. Section 2 introduces the basic tools and concepts used throughout the work. Then, Section 3 illustrates preliminary results based on a toy problem. Finally, some conclusions are drawn in Section 4, together with the discussion of the current and future challenges.

2 Spectral Uncertainty Quantification in Climate Model Predictions

Given the clearly recognized need for uncertainty analysis of the MOC response by the climate modeling community, we outline here two major issues that we will tackle: a) the need to represent model output in a (problem-independent) fashion that takes into account the bifurcations/discontinuities and b) the need to perform uncertainty quantification with only a limited set of sample points, due to the computational cost of the forward model. Hence, here and thereafter, by referring to the sparsity of the data we relate to the computational challenges associated with obtaining the data from complicated climate models. Below we focus on the technical approaches that will help us overcome the above difficulties in the MOC uncertainty analysis.

2.1 Bayesian Inference of the Loction of the Discontinuity in Climate Model Predictions

We propose a methodology that finds a probabilistic description of the discontinuity given finitely many data points. Assuming the discontinuity curve is a polynomial, the algorithm is based on Bayesian inference of its coefficients.

For clarity, let us focus on a two-dimensional case, since generalization to multidimensional data is straightforward. Assume the function $Z(r,\lambda) : I_r \times I_\lambda \to \mathbb{R}$ changes sharply as the arguments cross a smooth curve $r = \tilde{r}(\lambda)$. Given N values $\{z_i\}_{i=1}^N$ of this discontinuous function at points $\{(r_i,\lambda_i)\}_{i=1}^N \subset I_r \times I_\lambda$, we postulate $\tilde{r}(\lambda)$ to be a K-th order polynomial $p_{\boldsymbol{c}}(\lambda)$ and infer the coefficients $\boldsymbol{c} = (c_0,\dots,c_K)$ of an orthogonal expansion of it with respect to Legendre polynomials, i.e.

$$r \approx p_{\boldsymbol{c}}(\lambda) = \sum_{k=0}^{K} c_k \Psi_k^{(1)}(\lambda), \qquad (1)$$

where we use the superscript to denote the number of arguments of the polynomials, i.e. $\Psi^{(i)}(\cdot)$ is an i-variate Legendre polynomial.

The assumption that r and λ are related through a single-valued function can be lifted by parameterizing the curve along its length s, say, $r = r(s)$, $\lambda = \lambda(s)$. However, for the time being we will keep the simple assumption $r = \tilde{r}(\lambda)$, for the clarity of presentation.

Bayes' formula for the posterior probability for Model $\mathcal{M}$ given the Data $\mathcal{D} = \{z_1,\dots,z_N\}$ is

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}, \qquad (2)$$

where the prior probability $P(\mathcal{M})$ and the posterior probability $P(\mathcal{M}|\mathcal{D})$ represent degrees of belief about Model $\mathcal{M}$, before and after having the particular Data $\mathcal{D}$, respectively [19]. The likelihood function, viewed as a function of the Model, $L(\mathcal{M}) = P(\mathcal{D}|\mathcal{M})$ represents the probability of obtaining the data set $\mathcal{D}$ *if* it was drawn from the Model $\mathcal{M}$. The *evidence* $P(\mathcal{D})$ is the same probability marginalized over $\mathcal{M}$ and plays the role of a normalizing constant. We presume a simplified approximation model for the function $Z(r,\lambda)$ as a piecewise-constant function with two values m_L and m_R on each side of the discontinuity. Also, for the sake of constructing a well-defined likelihood function, we assume the discrepancy between the piecewise-constant model and the data to be Gaussian with standard deviations σ_L and σ_R on opposite sides of the discontinuity curve. This model is then parameterized by the coefficients of the discontinuity polynomial, expanded in terms of Legendre basis functions, $p_{\boldsymbol{c}}(\lambda) = \sum_{k=0}^{K} c_k \Psi_k^{(1)}(\lambda)$ up to order K, as well as by the *hyperparameter* set $\boldsymbol{h} = (m_L, \sigma_L, m_R, \sigma_R)$:

$$\mathcal{M}_{\boldsymbol{c},\boldsymbol{h}} \equiv g(r,\lambda,\boldsymbol{h}) = \begin{cases} m_L + \sigma_L \xi_L & \text{if } r \leq p_{\boldsymbol{c}}(\lambda) \\ m_R + \sigma_R \xi_R & \text{if } r > p_{\boldsymbol{c}}(\lambda) \end{cases}. \tag{3}$$

Here ξ_L and ξ_R are independent, standard normal random variables. In principle, one cares only about the values of the model parameter vector $\boldsymbol{c}$, and any joint distribution over $(\boldsymbol{c}, \boldsymbol{h})$ should be marginalized with respect to hyperparameters $\boldsymbol{h}$.

In the absence of additional information, we choose a uniform prior on the polynomial coefficient vector $\boldsymbol{c}$, i.e., $P(\mathcal{M}) = P(\boldsymbol{c}) = 1/\Delta^{K+1} = \text{const}$ on a hypercube $[-\Delta/2, +\Delta/2]^{K+1}$ for sufficiently large Δ. Also, constant, but positive-ranged priors are chosen for the hyperparameters.

The likelihood function $L(\mathcal{M}_{\boldsymbol{c},\boldsymbol{h}})$ is the probability of having the particular Data set $\mathcal{D} = \{z_1, \ldots, z_N\}$ *if* it was drawn from Model $\mathcal{M}$ with coefficients $\boldsymbol{c}$ and hyperparameters $\boldsymbol{h}$. In terms of the log-Likelihood,

$$\begin{aligned} \log P(\mathcal{D}|\mathcal{M}_{\boldsymbol{c},\boldsymbol{h}}) &= \textstyle\sum_{i=1}^{N} \log\left(P(z_i|\mathcal{M}_{\boldsymbol{c},\boldsymbol{h}})\right) = \\ &= -\textstyle\sum_{r_i \leq p_{\boldsymbol{c}}(\lambda_i)} \left(\log(\sigma_L\sqrt{2\pi}) + \frac{(z_i - m_L)^2}{2\sigma_L^2}\right) - \\ &\quad - \textstyle\sum_{r_i > p_{\boldsymbol{c}}(\lambda_i)} \left(\log(\sigma_R\sqrt{2\pi}) + \frac{(z_i - m_R)^2}{2\sigma_R^2}\right). \end{aligned} \tag{4}$$

Given the likelihood $P(\mathcal{D}|\mathcal{M})$ and the prior $P(\mathcal{M})$, we then draw samples from the posterior distribution $P(\mathcal{M}|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M})P(\mathcal{M})$ via Markov Chain Monte Carlo (MCMC) sampling. MCMC is a class of techniques that allows sampling from a posterior distribution by constructing a Markov Chain that has the posterior as its stationary distribution [2, 4]. In particular, we are using the adaptive Metropolis algorithm [5], which is an improvement over the original Metropolis algorithm [12], since it uses the covariance of the previously visited chain states to find better proposal directions, thus exploring the posterior distribution in a more efficient manner by accounting for the correlations between the parameters, see [5] for details.

The outcome of the MCMC chain is a distribution over the space of possible discontinuity curves. Specifically, we have a set of sample parameter vectors $\boldsymbol{c}$, each generating a specific discontinuity curve. To emphasize the stochasticity of these curves and their corresponding parameter sets, we will use the argument ω as an element of the sample space of the MCMC chain:

$$p_{\boldsymbol{c}}(\lambda;\omega) = \sum_{k=0}^{K} c_k(\omega)\Psi_k^{(1)}(\lambda). \tag{5}$$

2.2 Polynomial Chaos Representation via Parameter Domain Mapping

We will use Polynomial Chaos (PC) expansions to represent the uncertain parameters r and λ as random variables with probability distribution functions that are found from historical records and from expert opinions [22]. Specifically, we will use Legendre-Uniform PC expansions (PCEs), that are compatible with *local* PC implementations for handling discontinuities [7–9]. The uncertain parameters r and λ are assumed to be independent, and their PCEs

$$\begin{aligned} r &= \textstyle\sum_{k=0}^{P} r_k \Psi_k^{(1)}(\eta_1) = F_r^{-1}(\eta_1),\\ \lambda &= \textstyle\sum_{k=0}^{P} \lambda_k \Psi_k^{(1)}(\eta_2) = F_\lambda^{-1}(\eta_2) \end{aligned} \tag{6}$$

will be used to represent the uncertainty in the parameters in terms of independent Uniform$[0,1]$ random variables η_1 and η_2 and Legendre polynomials $\Psi_k^{(1)}(\cdot)$. A one-to-one correspondence between η_1 and r, as well as between η_2 and λ can be realized through the respective inverse cumulative distribution functions F_r^{-1} and F_λ^{-1}. This is a consequence of a simple result from probability theory stating that $F(X)$ is a Uniform$[0,1]$ random variable *if* $F(\cdot)$ is the cumulative distribution function of the random variable X. Next, we express the forward model output - in this case, the maximum of the overturning stream function - as a PC expansion of order p_{ord} in terms of bivariate Legendre polynomials $\Psi_p^{(2)}(\cdot,\cdot)$ and the random variable pair $(\eta_1,\eta_2) = (F_r(r), F_\lambda(\lambda))$ that is related to the two uncertain parameters through the cumulative distribution functions. Specifically, the maximum of the overturning stream function $Z(r,\lambda)$ can be written as

$$Z(F_r^{-1}(\eta_1), F_\lambda^{-1}(\eta_2)) \equiv \tilde{Z}(\eta_1,\eta_2) = \sum_{p=0}^{P} z_p \Psi_p^{(2)}(\eta_1,\eta_2), \tag{7}$$

up to bivariate polynomial order p_{ord}, with the number of terms $P+1 = (p_{\text{ord}}+2)(p_{\text{ord}}+1)/2$. The PC coefficients z_p can be obtained using orthogonal projection formulas

$$z_p = \frac{\langle \tilde{Z}(\eta_1,\eta_2)\Psi_p^{(2)}(\eta_1,\eta_2)\rangle}{\langle \Psi_p^{(2)}(\eta_1,\eta_2)\Psi_p^{(2)}(\eta_1,\eta_2)\rangle}. \tag{8}$$

The expectation $\langle\cdot\rangle$ is taken with respect to the probability distribution of the variables (η_1,η_2), which conveniently is a constant on $[0,1]$ in the Legendre-Uniform case.

The spectral representation (7) assumes a certain degree of smoothness in the output Z with respect to its arguments (r,λ). However, in the present context, the maximum overturning stream function undergoes a bifurcation with respect to the climate sensitivity λ and the rate of CO_2 increase r, essentially making the use of the global expansion (7) infeasible.

In principle, multi-domain PC expansions [7–9, 15] allow local representations that effectively overcome this issue. However, these local representations would be strongly challenged by the computational cost of obtaining climate model data, since they are generally obtained by adaptive refinement of domains with new data samples required for each level of refinement. We will take advantage of the fact that there is a single discontinuity curve that can be parameterized using Bayesian inference as described in Section 2.1.

The presence of discontinuities suggests that separate PC expansions of the maximum overturning stream function Z on each side of the discontinuity could lead to more accurate respresentations compared to global PC expansions. However, compact-support PC representations require rectangular domains and will be challenged by the irregularity of the resulting domains on either side of the discontinuity curve.

For each sample discontinuity curve $r = p_{\boldsymbol{c}}(\lambda;\omega)$ we find maps from irregular parameter domains (r,λ) on either side of the discontinuity curve to rectangular domains (r',λ'). These mappings are constructed via Laplace equations similar to the method of generating curvilinear body-fitted coordinates in computational fluid dynamics [6, 20, 24].

Let us focus on one of the irregular domains, the approach is exactly the same on the other side of the diconutinuity. Denote the stochastic map that depends on the ω-sampled discontinuity curve by $\mathcal{G}_\omega : (r,\lambda) \mapsto (r',\lambda')$. To be precise, $\mathcal{G}_\omega \equiv \mathcal{G}_{\omega,L}$ if (r,λ) lies on one side of the discontinuity, and $\mathcal{G}_\omega \equiv \mathcal{G}_{\omega,R}$ otherwise. In fact, to save an interim step, we can assume $(r',\lambda') = (\eta_1',\eta_2') \in [0,1]\times[0,1]$. Together with the map $\mathcal{F} : (\eta_1,\eta_2) \mapsto (r,\lambda)$ in Eq. (6), we effectively have a stochastic map $\mathcal{G}_\omega \circ \mathcal{F} \equiv \mathcal{G}'_\omega : (\eta_1,\eta_2) \mapsto (\eta_1',\eta_2')$ from $[0,1]\times[0,1]$ to itself. Again, it is worth emphasizing, that we have two such maps from $[0,1]\times[0,1]$ to itself, $\mathcal{G}'_{\omega,L}$ and $\mathcal{G}'_{\omega,R}$, for the domains on one or the other side of the discontinuity.

Now the integral in the numerator of the orthogonal projection formulae (8) will undergo a change of variables:

$$\begin{aligned}\langle \tilde{Z}(\eta_1,\eta_2)\Psi_p^{(2)}(\eta_1,\eta_2)\rangle &= \int \tilde{Z}(\eta_1,\eta_2)\Psi_p^{(2)}(\eta_1,\eta_2)d\eta_1\eta_2 \\ &= \int \tilde{Z}_\omega(\eta_1',\eta_2')\Psi_{p,\omega}^{(2)}(\eta_1',\eta_2')J_\omega(\eta_1',\eta_2')d\eta_1'\eta_2',\end{aligned} \tag{9}$$

where $J_\omega(\eta_1',\eta_2') = \frac{\partial(\eta_1,\eta_2)}{\partial(\eta_1',\eta_2')}$ is the Jacobian of the inverse map $(\mathcal{G}'_\omega)^{-1}$. Postponing the discussion of the possible ways and limitations of taking the above integral to Section 4, we note that, with z_p evaluated by (9), the PC representation (7) is written as (on either side of the discontinuity curve)

$$\tilde{Z}(\eta_1,\eta_2;\omega) = \sum_{p=0}^{P} z_p(\boldsymbol{c}(\omega))\Psi_p^{(2)}(\eta_1,\eta_2) \tag{10}$$

for each discontinuity curve coefficient vector $\boldsymbol{c}(\omega)$, where ω represents a sample element of the MCMC chain. In order to properly parameterize the space of sample elements ω, we will use the inverse Rosenblatt transformation [14, 15], which maps any (possibly dependent) random variable vector $\boldsymbol{c}$ to a vector $\boldsymbol{\zeta}$ of i.i.d. Uniform$[0,1]$ random variables of the same size. That will allow a well-defined integration

$$z_{pm} = \frac{\langle z_p(\boldsymbol{c}(\omega))\Psi_m^{(K+1)}(\boldsymbol{\zeta})\rangle}{\langle \Psi_m^{(K+1)}(\boldsymbol{\zeta})\Psi_m^{(K+1)}(\boldsymbol{\zeta})\rangle} \tag{11}$$

for the coefficients of the PC expansion of the z_p's

$$z_p = \sum_{m=0}^{M} z_{pm}\Psi_m^{(K+1)}(\boldsymbol{\zeta}), \tag{12}$$

where the $\Psi_m^{(K+1)}(\cdot)$'s are $(K+1)$-variate Legendre polynomials up to order m_{ord}, hence the number of terms is $M+1 = (K+1+m_{\text{ord}})!/(K+1)!m_{\text{ord}}!$.

To this end, the model output, i.e. the maximum overturning stream function Z is expressed as a PC expansion in terms of two stochastic variables (η_1,η_2) corresponding to parametric uncertainties and $K+1$ additional stochastic variables $\boldsymbol{\zeta} = (\zeta_0,\zeta_1,\ldots,\zeta_K)$ that account for the uncertainty in the discontinuity curve

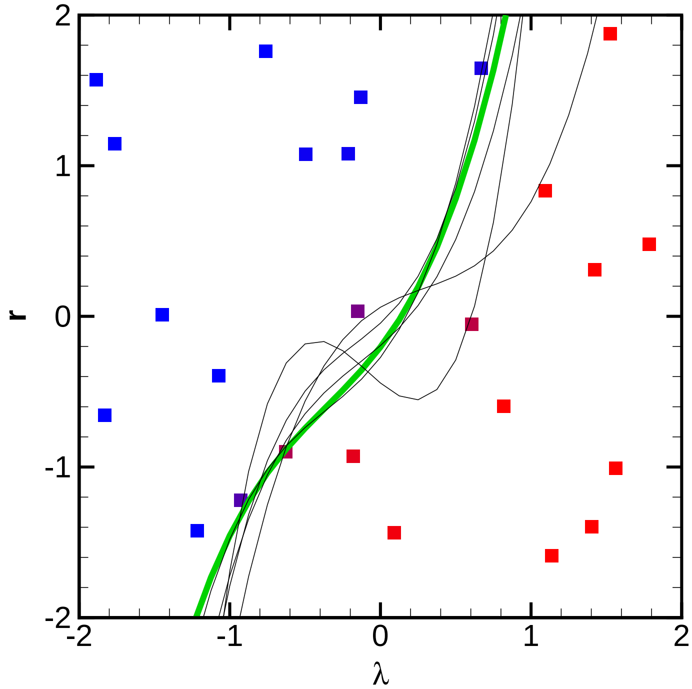

Figure 1. Sample discontinuity curves obtained through Bayesian inference. The thick green line corresponds to the maximum a posteriori estimate. Data samples and the color-coded output z_i values are artificially generated according to (15). Discontinuity strength parameter is set to $\alpha = 1$, while the data discontinuity curve is $\tilde{r}(\lambda) = 2\lambda^3$.

$$\begin{aligned} Z(\eta_1, \eta_2, \zeta) &= \sum_{p=0}^{P} z_p(\zeta)\Psi_p^{(2)}(\eta_1, \eta_2) \\ &= \sum_{p=0}^{P}\sum_{m=0}^{M} z_{pm}\Psi_m^{(K+1)}(\zeta)\Psi_p^{(2)}(\eta_1, \eta_2) \\ &= \sum_{l=0}^{L} z_l \Psi_l^{(K+3)}(\zeta, \eta_1, \eta_2), \end{aligned} \tag{13}$$

where the last equality is obtained by reordering the terms and $L + 1 = (M + 1)(P + 1)$. Note that we returned to the notation Z for the function to emphasize that it has the original climate model data values, i.e. the values for the maximum overturning stream function.

3 Preliminary Results

Let us demonstrate the methodology on test cases with artificially generated data, as if they were generated from climate models. For clarity of the presentation, we postulate simple, first-order Legendre-Uniform PC expansions for both parameters

$$\begin{aligned} r &= -2 + 4\eta_1, \\ \lambda &= -2 + 4\eta_2, \end{aligned} \tag{14}$$

i.e. the uncertain parameters r and λ are assumed to be uniformly distributed between -2 and 2. A simple way to generate a discontinuous data set $z_1, \ldots, z_N$ is to use the bivariate function with discontinuity strength parameter α and discontinuity curve $r = \tilde{r}(\lambda)$:

$$z_i = 1 + \tanh\left(\alpha(r_i - \tilde{r}(\lambda_i))\right). \qquad (15)$$

Figure 1 illustrates the inference of the discontinuity curve with color-coded data values. By assuming a third order polynomial, the discontinuity curve is parameterized by a four-dimensional vector, and the MCMC chain provides samples of this four-dimensional random vector. Clearly, there are infinitely many curves that are equally likely to be a discontinuity curve. Indeed, as long as two curves do not have any data points between them, their corresponding posterior values will be the same by construction of the likelihood function (4). Our methodology provides samples of all these curves and parameterizes them according to a PC expansion (1).

It is worth noting that the four dimensional random vector $\boldsymbol{c}$ has dependent components. As an example, Figure 2 shows the contours of joint posteriors of the first parameter c_0 with c_1, c_2 and c_3.

Figure 3 shows sample domain mappings for two discontinuity curves extracted from Figure 1. In the computational domain (*CD*) (right frame) the quadrature points corresponding to the roots of 4-th order Gauss-Legendre polynomials are shown with diamond symbols. For each instance of the discontinuity curve, there is a set of points in the physical domain (*PD*) corresponding to the quadrature points in the *CD*. In order to evaluate the integrals in Eq. (9), the forward climate model needs to be sampled in the *PD* at locations corresponding to the quadrature points in the *CD*. Moreover, this should be carried out for every sample curve with coefficient vector $\boldsymbol{c}(\omega)$ that correspond to *another* set quadrature point locations, in the $\boldsymbol{\zeta}$-space, in order to properly compute the integral in (11).

4 Conclusions and Work-in-Progress

In this paper, we are developing a methodology for uncertainty quantification in climate models with limited data and discontinuities. The sparsity of available data does not allow using the well-established domain-decomposition methods in order to tackle the discontinuity. Therefore, we propose a Bayesian approach to detect and parameterize the discontinuity as well as the uncertainty associated with it. Next, we use a random domain mapping strategy to map each of the two uncertain and irregular domains to rectangular ones where the application of the local spectral methods of uncertainty propagation is feasible.

The solution of the actual UQ problem depends on the implementation of the computation of the integral in (9). One can sample the data at the quadrature points of the computational domain(η_1', η_2') to perform the integration. However, this approach puts a restriction on the positions of the sampling points in the physical domain (r, λ) or (η_1, η_2). Moreover, for every sample of the discontinuity curve, new samples in the physical domain would be needed. For this reason,we are currently also considering interpolative methods of integration.

Another issue that we are tackling is the restrictions posed by the Rosenblatt transformation. Namely, a very large number of samples are required in order to properly capture the conditional distributions in the Rosenblatt transformation [14, 15]. Also, the effectiveness of the implementation of the Rosenblatt transformation relies on a certain

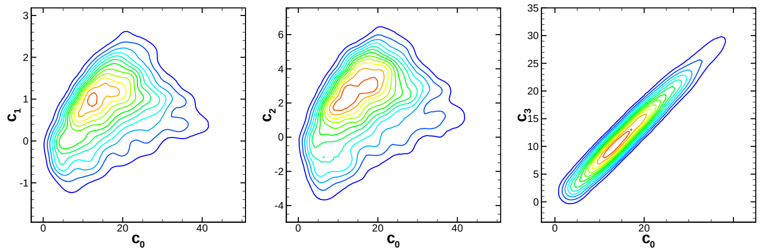

Figure 2. Contours of the joint posterior distributions of the first parameter c_0 coupled with c_1, c_2 and c_3. Clearly, there is strong dependence between various components of the random vector c.

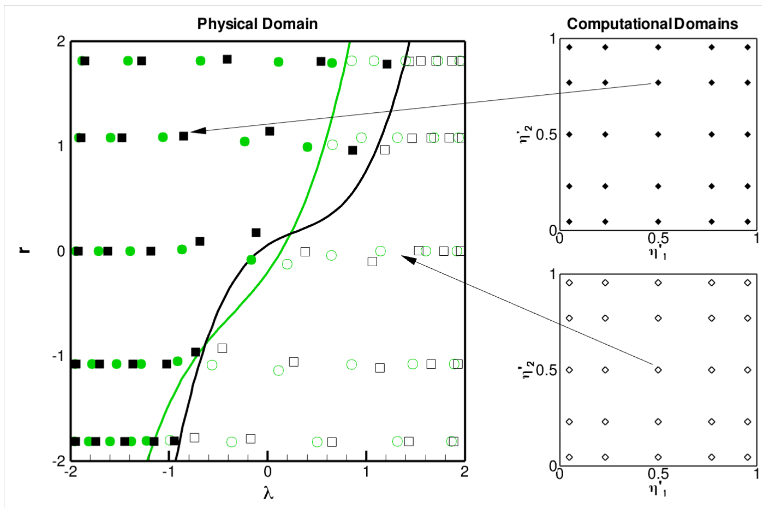

Figure 3. Sample domain mappings between physical and computational domain. The discontinuity curves are shown with thick lines - green for the most likely location, black for another instance. The quadrature points in the computational domain are shown with diamonds and their corresponding locations in the physical domain are shown with filled and open symbols for sub-domains left and right of the discontinuity curves, respectively.

importance ordering of the component random variables. For that reason, we consider enhancing our methodology to parameterize the discontinuity curve in terms of Karhunen-Loève expansions that automatically order the expansion coefficients.

Acknowledgment

This work was supported by Sandia National Laboratories Senior's Council LDRD (Laboratory Directed Research and Development) Program. Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy under contract No. DE-AC04-94-AL85000.

References

[1] W. Broecker. The great ocean conveyor. *Oceanography*, 4:79–89, 1991.

[2] D. Gamerman. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, London, 1997.

[3] A. Ganachaud and C. Wunsch. The oceanic meridional overturning circulation, mixing, bottom water formation and heat transport. *Nature*, 408:453–457, 2000.

[4] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

[5] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

[6] K. Hoffmann and S. Chiang. *Computational Fluid Dynamics*, volume 1, chapter 9, pages 358–426. EES, 2000.

[7] O. Le Maître, R. Ghanem, O. Knio, and H. Najm. Uncertainty propagation using Wiener-Haar expansions. *J. Comput. Phys.*, 197(1):28–57, 2004.

[8] O. Le Maître, H. Najm, R. Ghanem, and O. Knio. Multiresolution analysis of Wiener-type uncertainty propagation schemes. *J. Comput. Phys.*, 197:502–531, 2004.

[9] O. Le Maître, H. Najm, P. Pébay, R. Ghanem, and O. Knio. Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM J. Sci. Comput.*, 29(2):864–889, 2007.

[10] O. Le Maître, M. Reagan, B. Debusschere, H. Najm, R. Ghanem, and O. Knio. Natural convection in a closed cavity under stochastic, non-Boussinesq conditions. *SIAM J. Sci. Comp.*, 26(2):375–394, 2004.

[11] J. F. McManus, R. Francois, J. M. Gherardi, L. D. Keigwin, and S. Brown-Leger. Collapse and rapid resumption of atlantic meridional circulation linked to deglacial climate changes. *Nature*, 428:834–837, 2004.

[12] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[13] H. Najm, B. Debusschere, Y. Marzouk, S. Widmer, and O. L. Maître. Uncertainty Quantification in Chemical Systems. *Int. J. Num. Meth. Eng.*, 2008. in press.

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

[15] K. Sargsyan, B. Debusschere, H. Najm, and O. L. Maître. Spectral representation and reduced order modeling of the dynamics of stochastic reaction networks via adaptive data partitioning. *SIAM Journal on Scientific Computing*, submitted.

[16] K. Sargsyan, B. Debusschere, H. Najm, and Y. Marzouk. Bayesian inference of spectral expansions for predictability assessment in stochastic reaction networks. *Journal of Computational and Theoretical Nanoscience*, in press.

[17] A. Schmittner and T. F. Stocker. The stability of the thermohaline circulation in global warming experiments. *Journal of Climate*, 12(4):1117–1133, 1999.

[18] J. R. Scott, A. P. Sokolov, P. H. Stone, and M. D. Webster. Relative roles of climate sensitivity and forcing in defining the ocean circulation response to climate change. *Climate Dynamics*, 30(5):441–454, 2008.

[19] D. Sivia. *Data Analysis: A Bayesian Tutorial*. Oxford Science, 1996.

[20] D. Tartakovsky and D. Xiu. Stochastic analysis of transport in tubes with rough walls. *J. Comput. Phys.*, 217:248–259, 2006.

[21] M. Tatang, W. Pan, R. Prinn, and G. McRae. An efficient method for parametric uncertainty analysis of numerical geophysical models. *J. Geophys. Res.*, 102:21925–21932, 1997.

[22] 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):39–59, 2007.

[23] M. Webster and A. Sokolov. A Methodology for Quantifying Uncertainty in Climate Projections. *Climatic Change*, 46:417–446, 2000.

[24] D. Xiu and D. Tartakovsky. Numerical Methods for Differential Equations in Random Domains. *SIAM J. Sci. Comput.*, 28(3):1167–1185, 2006.

[25] K. Zickfeld, A. Levermann, M. G. Morgan, T. Kuhlbrodt, S. Rahmstorf, and D. W. Keith. Expert judgements on the response of the atlantic meridional overturning circulation to climate change. *Climatic Change*, 82:235–265, 2007.