Ocean Modelling 43–44 (2012) 52–63

Contents lists available at SciVerse ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Propagating boundary uncertainties using polynomial expansions

W.C. Thacker [a,b], A. Srinivasan [c,d,*], M. Iskandarani [d], O.M. Knio [e,1], M. Le Hénaff [a]

[a] Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL 33149, United States
[b] Atlantic Oceanographic and Meteorological Laboratory, Miami, FL 33149, United States
[c] Center for Computational Science, University of Miami, Miami, FL 33149, United States
[d] MPO Division, Rosenstiel School for Marine and Atmospheric Sciences, University of Miami, Miami, FL 33149, United States
[e] Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218, United States

ARTICLE INFO

Article history:
Received 13 June 2011
Received in revised form 25 November 2011
Accepted 27 November 2011
Available online 10 December 2011

Keywords:
Polynomial chaos
Ocean modeling
Error propagation
Uncertainty quantification
Spectral stochastic expansions
Gulf of Mexico
Loop Current
HYCOM

ABSTRACT

The method of polynomial chaos expansions is illustrated by showing how uncertainties in boundary conditions specifying the flow from the Caribbean Sea into the Gulf of Mexico manifest as uncertainties in a model's simulation of the Gulf's surface elevation field. The method, which has been used for a variety of engineering applications, is explained within an oceanographic context and its advantages and disadvantages are discussed. The method's utility requires that the spatially and temporally varying uncertainties of the inflow be characterized by a small number of independent random variables, which here correspond to amplitudes of spatiotemporal modes inferred from an available boundary climatology.

1. Introduction

The object of this paper is to point out how uncertainties of oceanographic simulations might be explored using the method of polynomial chaos expansions. This method was first introduced by Wiener (1938), who addressed the question of efficiently estimating uncertainties of a dynamical simulation stemming from uncertainties in its defining parameters. He realized that, in principle, a probability density describing the uncertainty of the parameters might be propagated dynamically to provide distributional information about any aspects of the simulation, but there was the issue of how to do it in practice. By using polynomial expansions to express the simulation's dependence on the uncertain parameters, he reduced the problem of propagating uncertainties to the task of determining expansion coefficients. The phrase "polynomial chaos", which has become popular in the engineering literature, stems from Wiener's referring to uncertainty as "chaos" and from his use of a polynomial expansion.[2,3] When the outputs of a simulation are well-approximated by polynomials of the inputs, polynomial expansions are appropriate, but when they are not, the expansions may converge slowly or may not converge at all.[4] The "chaos" part of the method relates to the choice of the polynomial basis: as the probability density function describing the uncertainty of the inputs appears in all expectation integrals, it is best to choose polynomials that are orthogonal when weighted by that density.

The method certainly should be of interest, as oceanographic simulations have many uncertain inputs.[5] For example, they depend on initial values of temperature, salinity, and other state variables at each point within the model's domain, on temporally varying values characterizing forcing fluxes everywhere on the

* Corresponding author at: Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL 33149, United States.
E-mail addresses: w.thacker@miami.edu, carlisle.thacker@noaa.gov (W.C. Thacker), asrinivasan@rsmas.miami.edu (A. Srinivasan), miskandarani@rsmas.miami.edu (M. Iskandarani), knio@jhu.edu (O.M. Knio), mlehenaff@rsmas.miami.edu (M. Le Hénaff).
[1] Present address: Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, United States.

[2] Chaos within this context should not be confused with its more modern usage to indicate sensitivity to small perturbations (Lorenz, 1963).
[3] For an introduction to the engineering literature see the reviews by Xiu (2009) and Najm (2009).
[4] While the Cameron–Martin theorem (Cameron and Martin, 1948) guarantees convergence for any finite variance process, in practice convergence is tested by checking the impact of retaining more terms in the expansion.
[5] Other approaches to oceanographic uncertainty can be found in the books of Bennett (2002), Evensen (2009), and Wunsch (2006). For discussions of uncertainty in fields other than oceanography, see the article in the special issue of Journal of Computational Physics (Karniadakis and Glimm, 2006) in which Lermusiaux (2006) presents his view of oceanographic uncertainties to a wider audience.

1463-5003/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2011.11.011

air-sea boundary, on values used for a variety of transport coefficients, and when there are open lateral boundaries on the details of their specification. Quantitative information about the impacts of their mis-specification could be quite valuable. Not only would it reveal the limitations of the utility of a simulation, it would also suggest which inputs must be better known to achieve a more useful simulation. It is important to recognize that the method of polynomial chaos expansions, like all methods for dealing with uncertainty, suffers from what Bellman (1957) called the "curse of dimensionality", namely the inescapable fact that computational complexity increases geometrically with increasing numbers of uncertain parameters. Thus, in practice, the method is used to examine the consequences of a limited number of uncertain inputs.

As Kalman filtering (e.g. Evensen, 2009) is better known to oceanographers, especially within the context of data assimilation where its role is to characterize the dynamically evolving uncertainties of the model state, comparing it with the method of polynomial chaos expansions can be instructive. The Kalman filter owes much of its utility to its characterization of the uncertainties using only an evolving mean state and an evolving matrix of covariances characterizing the state's uncertainty. The curse of dimensionality manifests in the size of the error-covariance matrix, which is unmanageably large, so much effort has been devoted to its approximation. For example, the ensemble Kalman filter approximates it using covariances inferred from a manageable number of simulations chosen to sample important aspects of the state's uncertainty. The method of polynomial chaos expansions as illustrated here also uses an ensemble of simulations to characterize the input uncertainties. However, the purpose of the ensemble is to provide quadrature information needed for evaluating the expansion coefficients, so the ensemble members are chosen to optimize the accuracy of the coefficients. The resulting expansions provide not just means and covariances but provide complete distributional information about the model's outputs.

It is also useful to note that Monte Carlo methods (e.g. Gilks et al., 1996), which also seek general distributional information about outputs, generally require a much larger ensemble of simulations to achieve the same accuracy that might be obtained from polynomial chaos expansions with a small quadrature ensemble. Polynomial interpolation between simulations in effect provides additional implicit sampling. While large Monte Carlo ensembles are unachievable for computationally intensive simulations, smaller quadrature ensembles might be affordable using today's computational resources.

If alternative choices for the uncertain parameters are regarded as perturbations of the favorite choice, then this method might be regarded as a perturbation method. However, as there is no requirement that the perturbations be small, the method of polynomial chaos expansions can accommodate information about large but unlikely perturbations. Within the context of automatic differentiation, propagation of infinitesimal perturbations is accomplished using the forward method and tangent-linear codes for accomplishing this can be generated automatically, but unfortunately they have to be run once for each perturbed input (e.g. Griewank and Corliss, 1991). On the other hand, sensitivities of a single output to infinitesimal perturbations of all uncertain inputs can be computed with automatically generated codes that implement the reverse or adjoint method.[6]

To illustrate the method of polynomial chaos expansions, we examine how uncertainties in the inflow through the Yucatan Straits manifest in the Gulf of Mexico's surface-elevation field and in the behavior of the Loop Current. Because of the Gulf's semi-enclosed geography with the Loop Current being the principal dynamical feature, we thought that the consequences of misspecifying the inflow should be interesting. Our challenge was to find a way to reduce the uncertainties of the spatially and temporally varying inflow to a few parameters, as we could find no published example of a similar problem. As the circulation in the Gulf is simulated using a high-resolution numerical model, the major computational expense is the ensemble of simulations needed to evaluate the coefficients of the polynomial expansions; the cost of evaluating the coefficients and using them to examine the output uncertainties is trivial in comparison.

Section 2 describes the methodology. After describing the numerical model used to simulate the Gulf's circulation, Section 3 explains our approach to reducing the inflow uncertainties to two random parameters. Section 4 discusses how the expansions are truncated and the ensemble of simulations needed for evaluating the coefficients of the polynomials. Then Section 5 presents the mean and standard deviation of the surface elevation field resulting from assumed distribution of possible boundary conditions and discusses surface-elevation covariances. By showing probability densities characterizing the non-Gaussian nature of the model's response, Section 6 illustrates how the polynomial expansions can be used to emulate the numerical model. And section 7 examines the convergence of the polynomial expansion. Finally, Section 8 concludes with comments about what the method might offer for oceanographic applications.

2. The methodology

The objective of the method is to assess how uncertainties of inputs of a dynamical system manifest in its outputs. To see how it works, consider the simple case of only a single uncertain input x, as generalization to two or more is relatively straightforward.[7] To express its uncertainty quantitatively, x can be expressed in terms of a central value x_0, which when not accounting for uncertainty would be used as input, and a spread x_1 characterizing the likely range of values around x_0:

$$x = x_0 + x_1\xi, \tag{1}$$

where ξ is a standardized random variable with probability density function $p(\xi)$.[8] For most problems we might have some idea what values to use for x_0 and x_1, but there may be little empirical basis for our choice of $p(\xi)$. When there are no fixed bounds on the range of x, the probability density might be taken as Gaussian. That was in fact the choice made by Wiener (1938), and that will also be ours, but other, possibly empirical, densities might be used.

Again for simplicity it is useful to focus on a single output $y = y(\xi)$, which might be thought of as the surface elevation at a particular space–time point.[9] The method centers on the assumption that output y can be efficiently described by series of polynomi-

[6] Adjoint codes are typically used to compute the gradient of a cost function for use in algorithms seeking to optimize the choice of a model's uncertain input parameters.

[7] When there is more than one uncertain parameter of interest, x in Eq. (1) becomes a vector, as do x_0 and ξ, while x_1 becomes a matrix.

[8] When constructing software that might be used for a variety of applications, it is useful to standardize ξ so that it has zero for its central value and a spread of unity.

[9] Another approach to polynomial chaos expansion (e.g. Knio and Le Maître, 2006; Le Maître and Knio, 2010) does require that the uncertainty of *all* evolving state variables be computed. In that case the polynomial chaos expansions for all state variables, each similar to Eq. (2), are inserted into the dynamical equations and the condition that the residuals be small in a statistical sense produces a system of equations for the expansion coefficients similar to but more complicated than the original dynamical system. As this would require software at least as demanding to construct as that already existing for the numerical model, this option was not considered for this study. Finette (2006) has proposed this approach for studying uncertainties of underwater acoustics, Ge et al. (2008) for nonlinear shallow-water equations, and Shen et al. (2010) for the Lorenz (1984) model. Somewhat similarly, Sapsis and Lermusiaux (2009) have suggested using a temporal evolving set of basis functions rather than a fixed polynomial basis.

als of the input x, or equivalently by polynomials of the standardized input ξ:

$$y(\xi) = \sum_{k=0}^{K} y_k P_k(\xi) + \epsilon_K(\xi), \tag{2}$$

where $P_0(\xi) = 1$, $P_1(\xi) = \xi$, and P_k for $k > 1$ are orthogonal polynomials of degree k and y_k are coefficients that must be determined; K indicates where the series is truncated and ϵ_K represents the truncation error. Rearranging could cast the expansion as a power series, but the polynomial grouping is preferred in order to exploit the orthogonality of the polynomials when evaluating expectation integrals:

$$\int P_j(\xi) P_l(\xi) p(\xi) d\xi = N_k \delta_{j,k}, \tag{3}$$

where N_k is a normalization constant. The probability density $p(\xi)$ governs the choice of polynomials to be used for the expansion. For example, a Gaussian density requires Hermite polynomials, the first few of which are listed in Table 1 together with their normalization constants. Similarly, a uniform density on a finite interval would require Legendre polynomials. And polynomials for an empirical density might be constructed using a Gram–Schmidt procedure (Witteveen and Bijl, 2006).

Before addressing the issues of how many terms are needed and how their coefficients can be determined, consider how the polynomial expansion (2) can be used to examine the uncertainty of model outputs. First, it allows statistics of the uncertain output y to be computed in a straightforward manner. For example, the mean of y for all possible values of x is simply the first term of the expansion y_0:

$$\langle y \rangle = \int y(\xi) p(\xi) d\xi = \sum_{k=0}^{K} y_k \int P_k(\xi) p(\xi) d\xi + \int \epsilon_K(\xi) p(\xi) d\xi = y_0, \tag{4}$$

where the truncation error term vanishes because it could have been represented by extending K to ∞. Thus, the mean is independent of the number of terms retained in the expansion. Note that the mean output is generally not the same as the output corresponding to the mean input: $y_0 \neq y(\xi_0)$. The variance of y involves all coefficients except y_0:

$$\begin{aligned} \langle (y - y_0)^2 \rangle &= \sum_{k,l=1}^{\infty} y_k y_l \int P_k(\xi) P_l(\xi) p(\xi) d\xi \\ &= \sum_{k=1}^{K} N_k y_k^2 + \text{truncation error}. \end{aligned} \tag{5}$$

As the estimate for variance reflects the number of retained terms, it can be used to monitor convergence.[10] And, for a second output variable z, perhaps the surface elevation at another space–time point, which would have expansion coefficients z_k, the covariance is provided by a similar series:

$$\langle (y - y_0)(z - z_0) \rangle = \sum_{k=1}^{K} N_k y_k z_k + \text{truncation error}. \tag{6}$$

Higher statistical moments can be computed in a similar fashion.

Second, the expansion (2) can also be used to generate an output for any desired input *without the need for solving the dynamic system*: just neglect the truncation error ϵ_K and evaluate $y(\xi)$ for the value of ξ corresponding to the desired input x.[11] Thus, knowing the expansion coefficients y_k allows you to synthesize an ensemble of outputs and thus to build a histogram characterizing the likelihood of any given value. The major cost of this convenience is in the evaluation of the expansion coefficients. The accuracy with which the outputs can be evaluated depends on the degree to which the polynomial expansion has converged. There appears to be no *a priori* way to know how many terms are required, so an *a posteriori* examination of the impact of the last retained term will be needed. Clearly, for high-resolution ocean modeling, computational resources limit the number of terms that can practically be considered.

Table 1
The first few Hermite polynomials P_n and their normalization factors N_n, which are appropriate when ξ has a standard normal probability density $p(\xi) = (1/\sqrt{2\pi}) \exp(-\xi^2/2)$.

n	$H_n(\xi)$	N_n
0	1	1
1	ξ	1
2	$\xi^2 - 1$	2
3	$\xi^3 - 3\xi$	6
4	$\xi^4 - 6\xi^2 + 3$	24
5	$\xi^5 - 10\xi^3 + 15\xi$	120
6	$\xi^6 - 15\xi^4 + 45\xi^2 - 15$	720

Now turn to the issue of evaluating the coefficients. Rather than following the original approach of Wiener (1938), which requires first deriving and then solving a coupled set of equations for the temporally evolving expansion coefficients, we take the simpler approach suggested by Le Maître et al. (2002) of determining them from a specifically designed ensemble of simulations that sample the possible inputs.[12] After multiplying (2) by $P_k(\xi)p(\xi)$ and integrating, orthogonality of the polynomials provides expressions for the coefficients:

$$y_k = \frac{1}{N_k} \int y(\xi) P_k(\xi) p(\xi) d\xi. \tag{7}$$

If $y(\xi)$ were known at sufficiently many values ξ_q, then the coefficients could be evaluated by quadrature:

$$\int y(\xi) P_k(\xi) p(\xi) d\xi = \sum_q y(\xi_q) P_k(\xi_q) w_q + \text{quadrature error}, \tag{8}$$

where w_q is the weight associated with quadrature point ξ_q and the summation is over all quadrature points (e.g. Abramowitz and Stegun, 1970). Computing the expansion coefficients z_k for a second output variable z requires little additional expense, as exactly the same quadrature points ξ_q can be used and the values $z(\xi_q)$ can be obtained simultaneously with $y(\xi_q)$. The major computational expense of examining *all* outputs of an ocean circulation model is the storage of their values for each quadrature point.

As obtaining the value of $y(\xi_q)$ for each quadrature point would require solving the dynamical system for the corresponding input $x_q = x_0 + x_1\xi_q$, when the coefficients are determined by quadrature, the polynomial chaos method might be regarded as a special type of Monte Carlo method. However, a distinction can be drawn based on the number of ensemble members needed for accurately portraying the statistics of y. When the expansion (2) for $y(\xi)$ converges rapidly, the number of quadrature points needed for accurate evaluation of the coefficients is also small, requiring considerably fewer model integrations than would be needed to achieve the same accuracy with Monte Carlo methods.

Quadrature presents the issues of how many points are needed and where they should be located so that they efficiently evaluate the integrals for the expansion coefficients. The presence of the

[10] Because each term in the expression (5) is positive, truncating the expansion necessarily *underestimates* the variance.

[11] In this regard polynomial chaos expansions resemble the Bayesian emulator of Conti and O'Hagan (2010).

[12] Tatang et al. (1997) have suggested a similar approach for studying uncertainties of radiative forcing in atmospheric models and Webster and Sokolov (2000) for quantifying uncertainties in climate projections.

probability density $p(\xi)$ in the integrand can be exploited by Gaussian quadrature to guarantee that Q quadrature points approximate the integral exactly when the rest of the integrand is a polynomial of degree $2Q-1$ or less. It is the product of a polynomial of degree K or less with the output, so the coefficients $y_0, y_1, \ldots, y_K$ can be computed exactly using $K+1$ Gaussian quadrature points as long as the output is a polynomial of degree K or less. If the neglected terms are small, y is almost a polynomial of degree K, so $K+1$ quadrature points should give good approximations for the integrals. In practice, as the rate of convergence of the polynomial expansion is *a priori* unknown, computational costs limit the number of model runs and thus the number of the quadrature points that can be used, limiting the accuracy of the expansion coefficients. Confirming the convergence of the quadrature integrals, like confirming the convergence of the polynomial expansion, requires an *a posteriori* analysis.

When the probability density $p(\xi)$ is Gaussian, the best locations for the points are those appropriate for Hermite–Gauss quadrature (Abramowitz and Stegun, 1970). These locations depend on the number of points used. Table 2 shows the locations and weights when the number of points range from 2 to 7. So the decision of where they should be located reduces to that of how many to use to get accurate evaluations of the integrals.[13]

Because practicality dictates that the consequences of only a few uncertain inputs can be analyzed, it is important to be clear about exactly which are to be propagated. For this study the dynamical system is the circulation of the Gulf of Mexico as described by a numerical model based on the partial-differential equations of fluid dynamics, and the focus is on the uncertainties of the inflow through the Yucatan Straits, which will be characterized in Section 3 below using two uncertain parameters. To a lesser extent practicality also limits the number of outputs that can be analyzed.[14] Here, they are limited to the evolving surface elevation field sampled at 15-day intervals.

3. Modeling the Gulf of Mexico and its inflow

The flow in the Gulf of Mexico (Fig. 1) is simulated using the Hybrid Coordinate Ocean Model, which is commonly known as HYCOM.[15] The configuration used here is the same as that being used operationally by the US Navy for ocean prediction.[16] The computational domain is open along portions of its southern and eastern boundaries, where values are generally provided by a lower-resolution (1/12° vs. 1/25°) simulation of HYCOM configured for the Atlantic Ocean (similar to Chassignet et al., 2007). As boundary conditions were available for the period from September 9, 2004 through December 31, 2007, that period was chosen for demonstrating the method of polynomial chaos expansions. To illustrate the method of polynomial chaos expansions, we ask how uncertainties associated with flow from the Caribbean manifest within the Gulf.

Table 2
Quadrature points ξ_q and weights w_q for Hermite–Gauss integration with n points.

n	ξ_q	w_q
2	±1.000000	0.5000000
3	0	0.6666667
	±1.732044	0.1666667
4	±0.741964	0.4541241
	±2.334414	0.0458758
5	0	0.5333335
	±1.355626	0.2220757
	±2.869694	0.0112574
6	±0.616706	0.4088287
	±1.889176	0.0886155
	±3.324257	0.0025558
7	0	0.4571431
	±1.154406	0.2401230
	±2.366760	0.0307574
	±3.750439	0.0005484

We are immediately faced with the issue of how uncertainties in these boundary conditions might be quantified. Ideally, they would be taken from a large ensemble of Atlantic simulations carefully prepared by varying all inputs over their likely values, but unfortunately no such ensemble exists. What was available was a "climatology" of the open boundary conditions (Kourafalou et al., 2009), and its spatial and temporal variability provides a proxy for the statistics of the uncertainties of the boundary conditions. Without having empirical evidence of the nature of the boundary uncertainty, it seems best to guarantee that they have a similar spread and co-variability as the boundary fields themselves, so that alternative boundary conditions produce reasonable flows.

The next issue is reducing the boundary climatology to just a few parameters that might be propagated using the method of polynomial chaos expansions. These few parameters should characterize the uncertainties of each of the model's state variables at every point on the open southern boundary. In addition there is the issue of characterizing how these uncertainties change with time. Even if the deviation of the boundary state at a given time might differ from the favorite boundary conditions for that time, its deviation at another time is not likely to be the same, but it should be related. So there is the need to account for temporal variations in the uncertainty without unduly increasing the number of random parameters.

Our solution was to analyze the boundary climatology into a sum of products of spatial patterns and time series using singular-value decomposition. The spatial patterns are multivariate empirical orthogonal functions (EOFs), *i.e.* eigenvectors of the boundary-data correlation matrix, the times series are the corresponding principal components, *i.e.* linear combinations of the boundary variables, and the singular values indicate the amounts of variability associated with each spatiotemporal mode. If the first few singular values are sufficiently large, then most of the boundary climatology's multivariate spatial and temporal co-variability can be described by just a few modes, and these modes of co-variability can be used to model the uncertainties of the boundary conditions.

So a class of reasonable boundary conditions can be generated by adding to the favorite boundary conditions some amounts of each of these modes, with a probability density governing how much of each mode is reasonable. The density's central values would be zero, so that the favorite boundary conditions would be the most likely boundary conditions, and the multidimensional spread could be estimated from the singular values, which are related to the fraction of variance of each mode occurring in the climatology. Thus, if X_0 is a matrix containing data for the favorite boundary conditions, with each column corresponding to

[13] These quadrature points and weights are appropriate when the integrands involve the normal density $(1/\sqrt{2\pi})\exp(-\xi^2/2)$. Tabulated values are often for integrands having instead a factor $\exp(-\xi^2)$ with the points shifted by a factor $\sqrt{2}$ and the weights larger by a factor $\sqrt{\pi}$.

[14] For our example, the outputs comprise the hydrodynamic and thermodynamic fields at all points within the model's domain as they evolve in time as well as Lagrangian quantities such as centers of eddies or the maximum northward extent of the Loop Current.

[15] HYCOM's distinguishing feature is a generalized vertical coordinate system that optimizes the distribution of vertical computational layers by making them isopycnic in stratified regions, terrain-following in shallow coastal regions, and isobaric in the unstratified mixed layer (Bleck, 2002). It serves a large community, who use it for a variety of applications. More information about HYCOM can be found at http://www.hycom.org.

[16] Details of the surface forcing, mixing parameterizations, etc. can be found in the paper of Prasad and Hogan (2007).

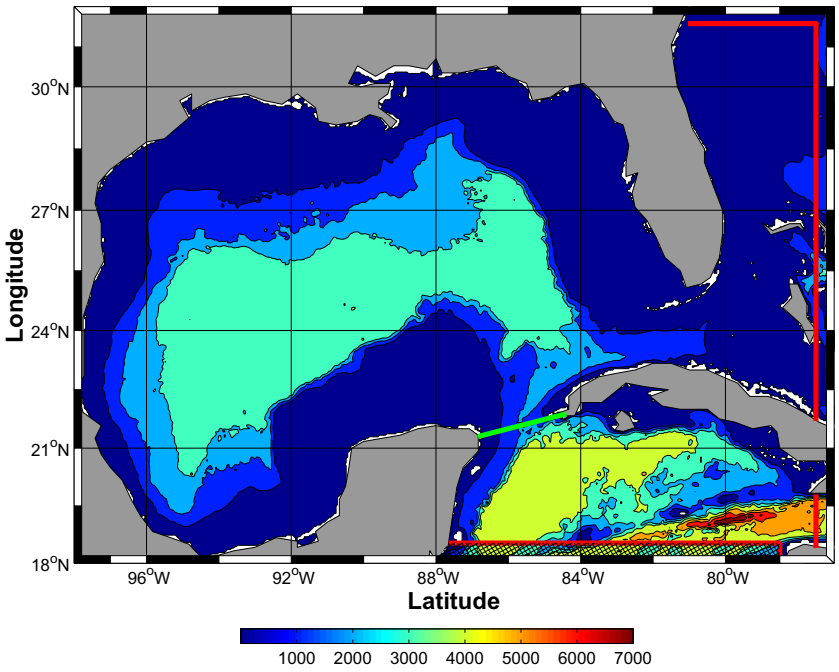

Fig. 1. Model domain. Color-filled contours indicate bathymetry in meters. Computational information from larger domain is provided along the red line along the open boundary. Uncertainty of the inflow is modeled in the cross-hatched region at the southern boundary.

a particular time and each row to a particular variable at a point on the open boundary, then other possible boundary conditions could be represented by a similar matrix X:

$$X = X_0 + \alpha \sum_{\kappa}^{N} \xi_\kappa \lambda_\kappa c_\kappa r_\kappa^T, \tag{9}$$

where the column vectors c_κ are the EOFs, the row vectors r_κ^T are the principal components, λ_κ are the singular values, N is the number of modes used to characterize the boundary variability, ξ_κ are unit-variance random amplitudes reflecting the uncertainty of the boundary values.[17,18,19] The coefficient α controls the spread of likely boundary values relative to the boundary's climatological variability; we use $\alpha = 1$ for the examples discussed here, but if we had considered our favorite boundary conditions to be more reliable, a smaller value would have been be more appropriate.

Because the principal components are uncorrelated, their random amplitudes should also be uncorrelated, and their multivariate probability density function p_N should be the product of univariate densities p:

$$p_N(\xi_1, \xi_2, \ldots, \xi_N) = \prod_{\kappa}^{N} p(\xi_\kappa). \tag{10}$$

And because the principal components reflect the variance of the climatology about its mean, the univariate densities can be taken to be standard normal densities.

The singular values λ_k are shown in Fig. 2. Their squares are proportional to the fraction of variance of the climatology represented by the linear combinations of the boundary variables corresponding to each mode. As the first these eight modes account for 90% of the variance, it would be good to use these eight modes to model the uncertainty of the boundary conditions. However, because computational costs increases geometrically with each additional uncertain parameter, we have chosen to use only the first two,

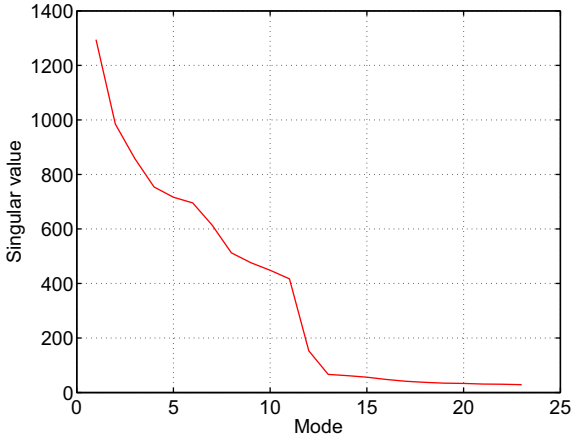

Fig. 2. Singular values λ_κ of the boundary climatology. As the climatology provides bi-weekly values and the means have been removed, there are only 25 non-zero singular values.

[17] When computing the singular-value decomposition, the boundary data were first standardized by removing each variable's mean and dividing by its standard deviation, so that all could be represented on a common scale. The components of the EOFs c_κ were then multiplied by the corresponding standard deviations to restore their units.

[18] The climatological boundary data were available bi-weekly for 26 weeks, and as means had been removed, there were 25 non-zero singular values.

[19] It is interesting to compare Eq. (9) with its univariate counterpart (1). X, which represents all variables at all points on the open southern boundary at all times, is the multivariate generalization of the single uncertain input x, and X_0 is the generalization of the central value x_0. While the many elements of X are all uncertain, their uncertainties are not independent, as they are tied to the N parameters ξ_κ, the multivariate generalizations of the standardized parameter ξ in (1). The spread x_1 generalizes to the matrix products $\alpha\lambda_\kappa c_\kappa\ r_\kappa$ assigning a spread to each boundary variable for each of the parameters ξ_κ. If the matrices X and X_0 are all unfolded into column vectors $\mathbf{x}$ and $\mathbf{x}_0$, and if the parameters are regarded as elements of a (shorter) column vector ξ, then when unfolded the spread matrices $\alpha\lambda_\kappa c_\kappa r_\kappa$ can be organized as columns of a grand spread matrix X_1, and Eq. (9) can be rewritten in a form analogous to (1): $\mathbf{x} = \mathbf{x}_0 + X_1\xi$.

which together account for 42% of the variance, as that is sufficient to illustrate the method. Thus, Eq. (9) becomes:

$$X = X_0 + \alpha\xi_1\lambda_1 c_1 r_1^T + \alpha\xi_2\lambda_2 c_2 r_2^T. \tag{11}$$

The lower-left panel of Fig. 3 shows the first two principal components r_1 and r_2, which characterize the temporal behavior of the two modes of southern-boundary variability. The upper two panels show the corresponding spatial patterns of meridional velocity. These patterns, along with those for zonal velocity, temperature, salinity and pressure which are not shown, are contained in the column vectors c_1 and c_2. For comparison, the mean meridional velocity is shown in the lower-right panel.

4. Quadrature ensemble

With two random inputs the polynomial chaos expansion involves polynomials of two variables. Because the parameters ξ_1 and ξ_2 are, by construction, uncorrelated over the one-year interval for which there were data to define them, they can be regarded as statistically independent. Consequently, the bi-variate probability density describing their distribution is the product of two univariate densities $p(\xi_1)$ and $p(\xi_1)$, in this case standard Gaussians, and the expansion polynomials can be chosen as products of Hermite polynomials $P_{k_1}(\xi_1)$ and $P_{k_2}(\xi_2)$. Thus, the polynomial expansion for a single output becomes:

$$y(\xi_1,\xi_2) = \sum_{k_1,k_2}^{k_1+k_2\leqslant K} y_{k_1,k_2} P_{k_1}(\xi_1) P_{k_2}(\xi_2) + \epsilon_K(\xi_1,\xi_2), \tag{12}$$

where triangular truncation retains polynomials of total degree no greater than $K = 6$.[20] Exploiting orthogonality of the polynomials as in section 2 expresses the expansion coefficients as double integrals:

$$y_{k_1,k_2} = \frac{1}{N_{k_1}N_{k_2}} \iint y(\xi_1,\xi_2) P_{k_1}(\xi_1) p(\xi_1) d\xi_1 P_{k_2}(\xi_2) p(\xi_2) d\xi_2, \tag{13}$$

which can be evaluated by Gauss–Hermite quadrature one integral at a time:

$$\begin{aligned}\iint & y(\xi_1,\xi_2) P_{k_1}(\xi_1) p(\xi_1) d\xi_1 P_{k_2}(\xi_2) p(\xi_2) d\xi_2 \\ &\approx \sum_{q_1}\sum_{q_2} y(\xi_{q_1},\xi_{q_2}) P_{k_1}(\xi_{q_1}) w_{q_1} P_{k_2}(\xi_{q_2}) w_{q_2}.\end{aligned} \tag{14}$$

The quadrature points q_1 and q_2 and weights w_1 and w_2 (Table 2) are for integrating over ξ_1 and ξ_2, respectively.

Now comes the question of how many quadrature points to use. First, should the same number be used for both integrations? If the polynomial expansion converges faster for one uncertain input than for the other, the output is more nearly a low-order polynomial of that variable and fewer quadrature points may suffice. However, as the rates of convergence are not known *a priori*, it seems best to use the same number for both. With the two uncertain variables treated symmetrically, the locations of the quadrature points needed for one integration are the same as for the other, just along a different axis; because the quadrature approximates a double integral, the points are not confined to the axes but are spread over the plane (Fig. 4). So if Q points are needed for each one-dimensional integral, then Q^2 pairs (ξ_{q_1},ξ_{q_1}) are need for the double integral. As truncation ignores polynomials of degree $k > 6$, it is reasonable to use $Q = 7$ quadrature points in each direction or 49 in all, each requiring a separate run of the HYCOM model. And as the model has a horizontal resolution of roughly 4 km, the 64 runs needed for $Q = 8$ would tax our available computational resources, so we chose to proceed with 7 quadrature points in each direction.

Fig. 4 shows the locations of the 49 quadrature points relative to contours of the bi-variate normal density function. There is a 90% probability that an open southern boundary conditions corresponds to points (ξ_1,ξ_2) within the smallest circle. The next larger circle encloses an additional 9% of the possible boundary conditions, and each larger circle adds a smaller fraction, leaving only 0.0001% outside the largest circle. Note that many of the quadrature points correspond to boundary conditions that are highly unlikely. Thus, the ensemble of HYCOM runs providing values at the quadrature points includes what might be regarded as quite extreme events. Monte Carlo methods would require an ensemble of 1,000,000 randomly drawn boundary conditions to have a reasonable chance of sampling beyond the largest circle where the much smaller quadrature ensemble has four points. Note however that each of these remote cases has a quadrature weight of only 3.0074×10^{-7}.[21]

Each of the 49 quadrature points provides a different specification of the open southern boundary. And with initial conditions, surface forcing, mixing parameters and all other inputs being the same, the boundary conditions determine an ensemble of 49 HYCOM simulations. Fig. 5 illustrates how the sea-surface-elevation field differs across the members of this ensemble. Because even the quadrature points within the inner circle of Fig. 4 correspond to significant departures from the favorite boundary conditions, the red contours do not cluster around the thicker black contour. With the other members of the quadrature ensemble corresponding to rather unlikely boundary conditions, it appears that the boundary ensemble does not provide much direct information about the consequences of more likely situations; instead, that sort of information might be obtained from the polynomial chaos expansions once the coefficients have been evaluated using the quadrature ensemble. Nevertheless, one thing is immediately clear: the variability after 15 days is much less than that after a longer time. The reason is that all members of the ensemble start from the same initial conditions, which did not reflect previous uncertainties in the flow through the open southern boundary. There is a transient period during which the uncertainties develop, and only after this period can the consequences of the boundary uncertainty be fully appreciated. However, because natural dynamical variability is also present, the ensemble of simulations evolve in time even after the uncertainties are fully developed.

5. Means, standard deviations, and covariances of surface elevation

For this study the surface elevation for each grid cell was saved at 15-day intervals, and we focused on just six of these days.[22] For each grid cell there were 49 values of surface elevation on each of those days, one value from each member of the quadrature ensemble, which were used to evaluate the coefficients of the polynomial expansions for each cell's surface elevation.[23] Once the expansion

[20] Because the principal components are uncorrelated, so are their sums and differences. So the expansions equally well might have been in polynomials of $\xi_\pm = (\xi_1 \pm \xi_2)/\sqrt{2}$, which if truncated at degree K would contain terms involving ξ_1^{2K} and ξ_2^{2K}. Triangular truncation $K_1 + K_2 \leqslant K$ better respects the isotropy of the random variables and regards the products $K_1 + K_2 > K$ as being of higher order.

[21] For practical purposes these weights might be taken to be zero and the simulations corresponding to the four most unlikely members of the ensemble could be avoided. The two-dimensional array of quadrature points does not appear to be optimal for sampling likely situations and other approaches to two- and higher-dimensional quadrature might be more cost-effective.

[22] Subsurface variables might also have been saved in order to avoid redoing the ensemble of simulations in case uncertainties of some aspects of subsurface circulation might be a focus of future interest.

[23] Note that coefficients for all outputs need not be computed at once. The appropriate strategy is first to compute the coefficients needed to get the statistics you think might be the most useful; if other statistics are subsequently desired, then more coefficients can be computed at that time.

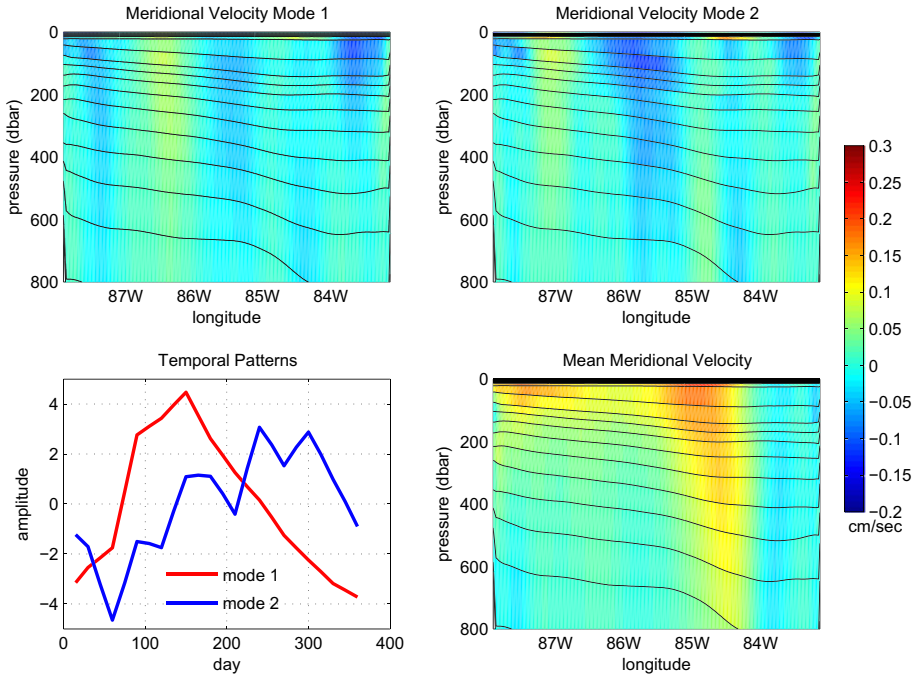

Fig. 3. Principal components (lower-left panel), 1st EOF (upper-left) and 2nd EOF (upper-right) for the meridional velocity, and mean meridional velocity at the southern open boundary. The inserts show the bathymetry at the open southern boundary.

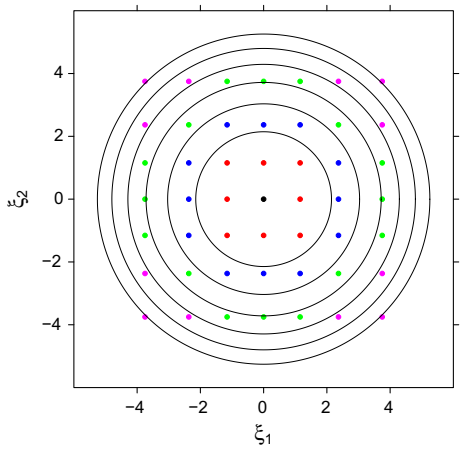

Fig. 4. Circles enclose regions of 90%, 99%,...,99.9999% probability. Dots mark locations of the Gauss–Hermite quadrature points, with red dots corresponding to relatively likely, blue less likely, green unlikely, and magenta highly unlikely boundary conditions. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

coefficients for surface elevation have been computed, computing mean and variance is straightforward as the generalization of Eqs. (4) and (5) to the case of two uncertain inputs is obvious: the mean is given by the constant term in the expansion, and the variance is approximated by the sum of the squares of the coefficients of the other terms.

Fig. 6 shows the mean surface elevation for each grid cell at six different times. These means are averages over the uncertainties in the boundary conditions as characterized by the two boundary modes. The plots look much like those of the central member of the ensemble at the corresponding times, but somewhat less sharp. Notice that 15 days after the boundary uncertainty was initiated, when each member of the ensemble retains much of their common heritage of the same initial conditions, gradients of the mean surface elevation are stronger than they are in the other panels, where sufficient time has passed for the cumulative effect of past uncertainties to be felt.

While each output's mean can be estimated from a single expansion coefficient, estimates of its variance involve all the remaining coefficients and are thus subject to truncation error.[24] Fig. 7 shows estimates of the standard deviations of each grid cell's surface elevation at the same times as the means were shown in Fig. 6. These standard deviations reflect the uncertainties in the positions of the Loop Current and its rings in the HYCOM simulation stemming from the uncertainties in the flow from the Caribbean into the Gulf. The fact that the uncertainties are smaller for day 15 illustrates again the fact that boundary uncertainties predating the quadrature simulations were not reflected in the common initial conditions. The fact that the magnitudes of the standard deviations of surface elevation are comparable in size to the means of Fig. 6 reflects the choice $\alpha = 1$ in Eq. (9), which characterizes uncertainties in the inflow comparable to its climatic range. A smaller value of α more appropriate for a situation where the inflow is relatively well known could have led to smaller values of standard deviation.

Estimates of covariances of two outputs are computed from sums of products of their expansion coefficients. Fig. 8 shows plots of covariance of surface elevation for a target cell at the coordinates (86°E, 24.1°N), with that of all grid cells. That target cell, which is marked by a white star, was chosen because it appears to be at the center of variability of the Loop Current circulation. Just as in the plots for the means and standard deviations, the spin-up of uncertainty is evident. To de-emphasize the magnitude of the uncertainty, plots of correlation coefficients (not shown) could also be drawn; they would be expected to show a maximum at the target point and a similar general structure with high correlations over the region for which the Loop Current penetrates.

[24] With triangular truncation at 6th degree, there are $(1 + 2 + \ldots + 7)/2 = 28$ terms retained in the polynomial expansion and thus 27 terms contributing to the variance.

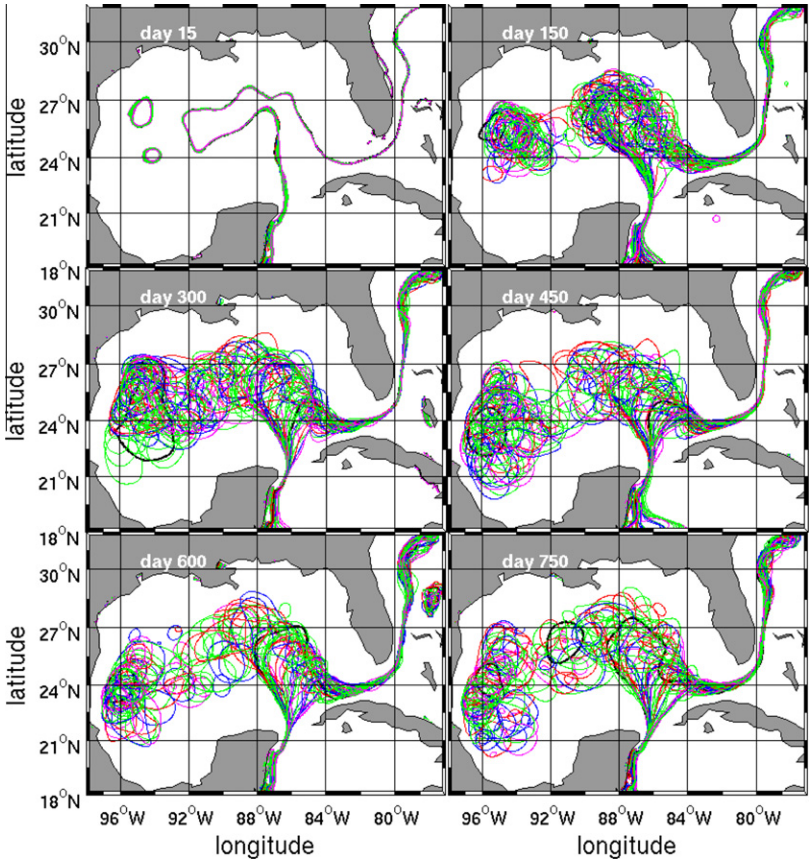

Fig. 5. Locations of the Loop Current and its eddies from the 49 HYCOM quadrature runs as indicated by 17 cm sea-surface-height contours. The panels, from upper left to lower right, show the contours at 15, 150, 300, 450, 600, and 750 days after the boundary uncertainties were initiated. The colors of the contours correspond to the colors of the dots in Fig. 4 with the thick black contour indicating the central member of the ensemble.

The covariances of Fig. 8 represent a subset of all covariances that might be estimated. If all variables could be saved during the quadrature runs, then it would be possible in principle to estimate the evolving covariance matrix for the complete model state. As a covariance matrix quantifying the spread of uncertainty of the model state is central to statistical methods for assimilating data, it is useful to consider whether those estimated from polynomial chaos expansions might be useful for that task. While the covariances of Fig. 8 characterize only the uncertainties attributable to uncertainties associated with the specification of the open southern boundary, their counterparts needed for assimilating data should account for all sources of uncertainty. On the other hand, because the covariances associated with all sources of uncertainty are not known, much of the art of assimilation is in their approximation. Perhaps uncertainties from other sources, *e.g.* wind stress or mixing parameters, when propagated give similar patterns of co-variability, indicating that dynamical evolution causes uncertainties to forget their origins and to manifest in patterns reflecting the likely state of the system. If so, then polynomial chaos expansions might offer a way to explore useful approximations to the error-covariance matrix.

6. Emulation and kernel density estimates for surface elevation

A great advantage of the method of polynomial expansion is that it allows for a detailed view of the distributions of the outputs of the dynamical system. Once an output's expansion coefficients have been evaluated, then the expansion in Eq. (12) can be used

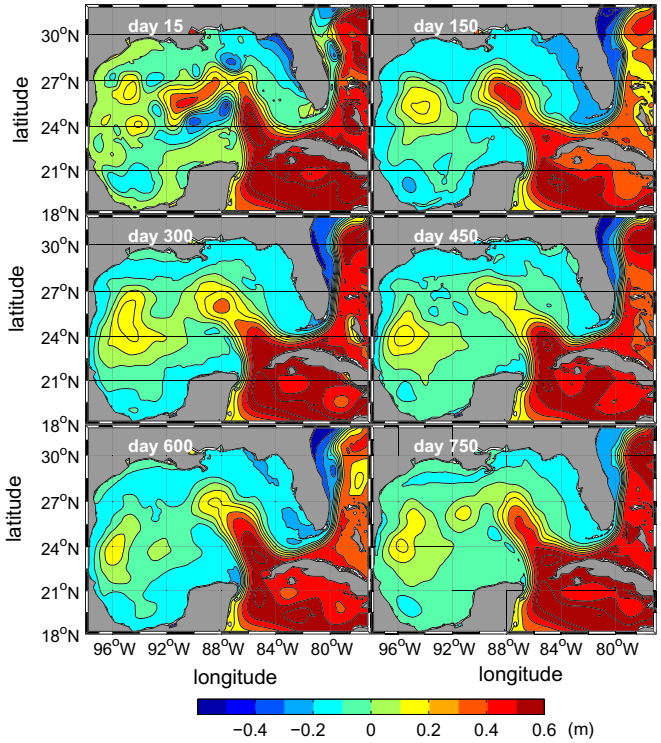

Fig. 6. Mean (m) of the sea-surface-height field from the polynomial chaos expansion at 15, 150, 300, 450, 600, and 750 days after the boundary uncertainties were initiated.

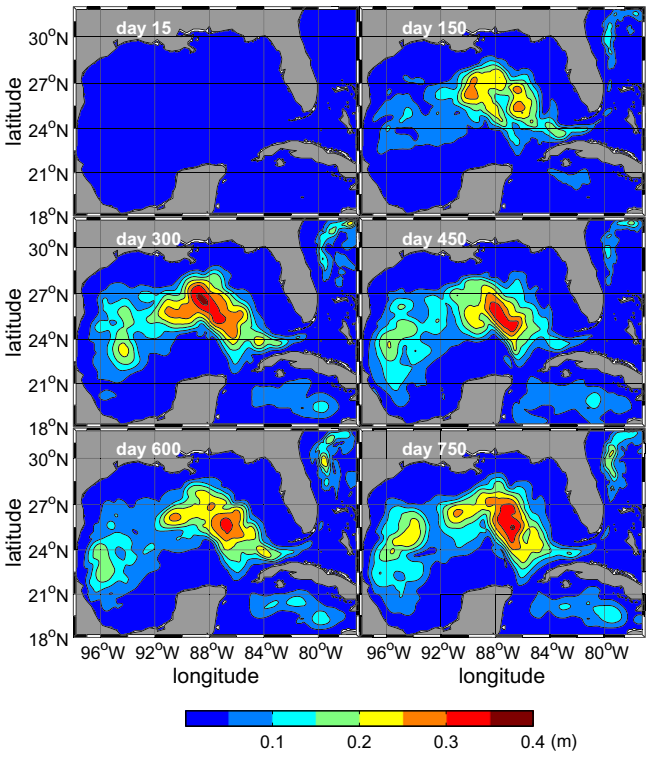

Fig. 7. Standard deviation (m) of the sea-surface-height field from the polynomial chaos expansion at 15, 150, 300, 450, 600, and 750 days.

to emulate possible values of the output y without the need for additional expensive simulations.

Fig. 9 illustrates how emulated values can provide information about the consequences of different boundary conditions for any

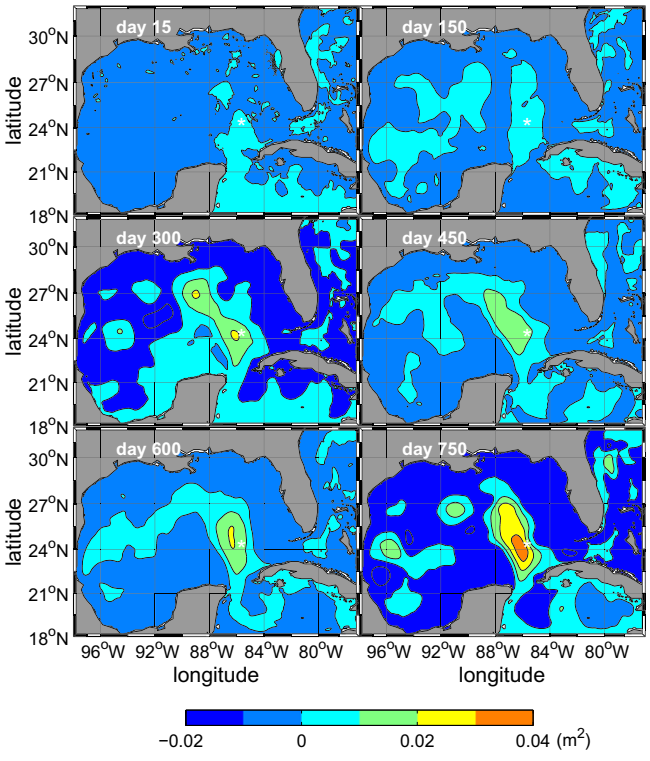

Fig. 8. Estimates of covariance (m^2) of the surface elevation for each grid cell with the surface elevation for the cell at the point (86°E, 24.1°N) marked by the white star. The panels correspond to the same days as those in Figs. 6 and 7.

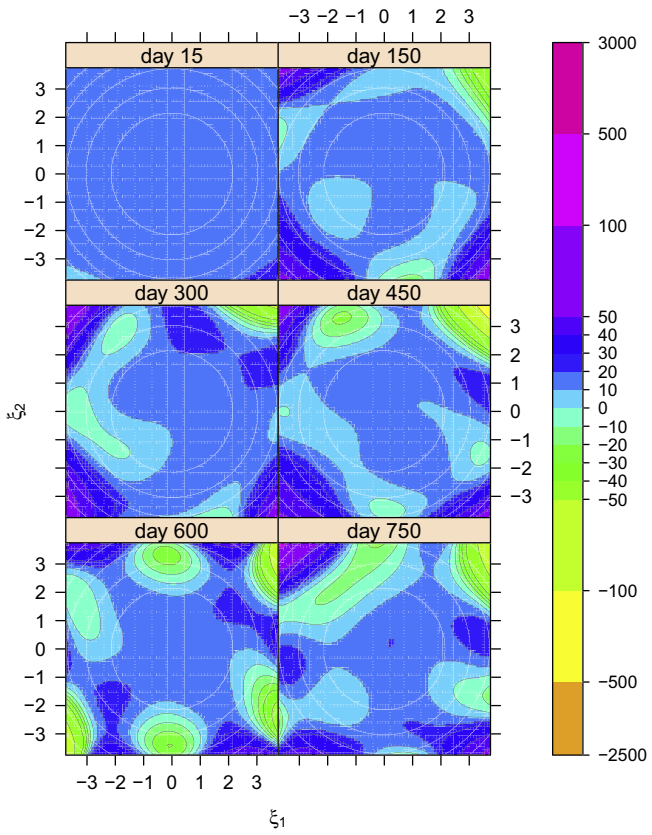

Fig. 9. Surface elevation (cm) at (86°E, 24.1°N) as a function of random variables ξ_1 and ξ_2. Note the compressed color scale used to distinguish more likely from less likely responses: Contours are at 10 cm intervals from −50 to 50 cm, and more extreme values are represented with a compressed scale. The circles, which are the same as those of Fig. 4, indicate that the extreme values are highly unlikely.

output. It shows the response of surface elevation at (86°E, 24.1°N) as a function of ξ_1 and ξ_2. Recall from Fig. 4 that the likely boundary conditions are near the center of the plots, so the responses within the innermost circle are generally more useful. Note that the response does not necessarily vary monotonically with ξ_1 and ξ_2; instead it can exhibit maxima and minima consistent with its assumed polynomial nature. Note also that, for extremely unlikely boundary flows as indicated by the circular contours, the response can be extremely large, while the range of responses for more likely boundary flows is quite reasonable.

Emulated values of surface elevation from a large ensemble of randomly generated (ξ_1, ξ_2) pairs can be used to construct a histogram that can be smoothed to provide an approximate probability density. An example for surface elevation at the same location (86°E, 24.1°N) is shown in Fig. 10. The histograms correspond to 50,000 randomly drawn (ξ_1, ξ_2) pairs, each determining a different boundary condition that is propagated through the Gulf by emulation, as an ensemble of numerical simulations of this size is clearly unfordable. The range of emulated data actually range from −2.20 to 3.25 m, but as there are very few extreme values, the plots have been restricted to a more reasonable range. For comparison, there are ticks marking the 49 simulated values from the quadrature ensemble. Kernel density functions are superimposed in red as estimates of the probability densities. The densities are clearly evolving. The small spread at day 15 is another reflection of the memory of the common initial conditions shared by all members of the quadrature ensemble, and the evolving shape of the density is a consequence of the different responses to surface forcing of the different ensemble members. For reference, superimposed in black, are Gaussian densities with means and standard deviations given

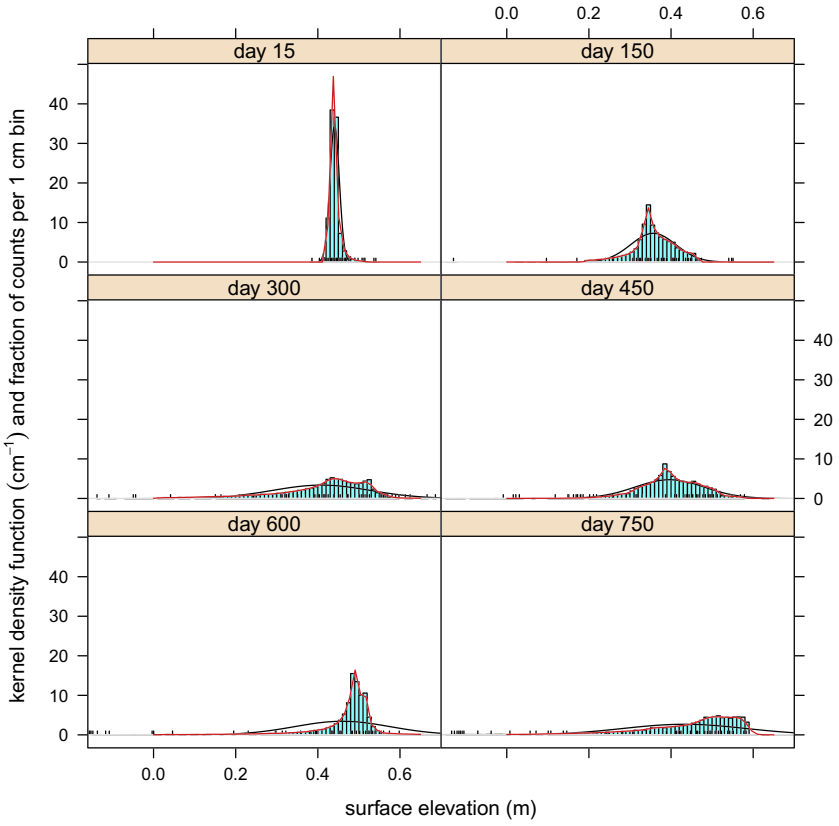

Fig. 10. Kernel density estimates for surface elevation (m) at the point (86°E, 24.1°N) for days 15, 150, 300, 450, 600, and 750. They were derived from histograms generated using polynomial chaos expansion for the mixed layer depths corresponding to 50,000 random boundary conditions. Ticks along the bottom indicate values for the 49 HYCOM simulations. Red curves are kernel density estimates, and black curves are Gaussian densities with means and standard deviations from the polynomial expansions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

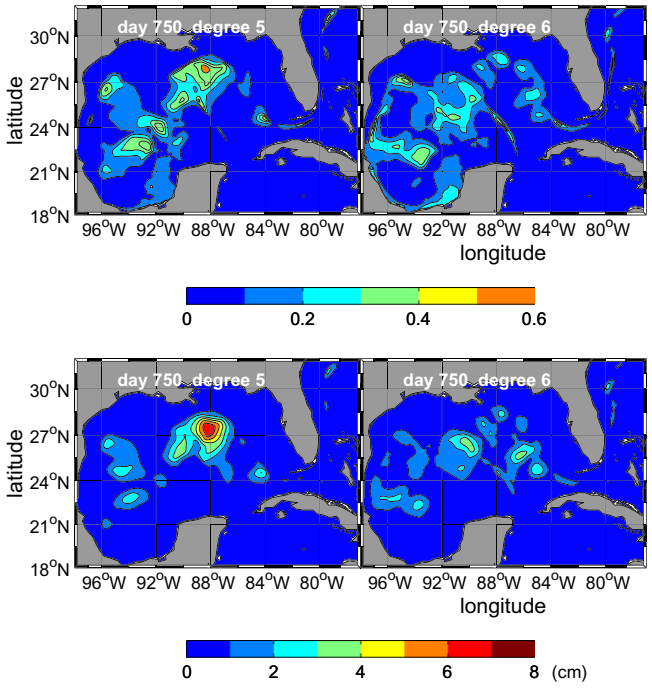

Fig. 11. Upper panels show fraction of variance of survace elevation at day 750 due to the retained polynomial terms of highest degree: (left) contribution of the 6 5th-degree terms relative to the total contributed by the 21 terms of degree less than 6; (right) contribution of the 7 6th-degree terms relative to the total contributed by the 28 terms of degree less than 7. Lower panels show incrementalal contribution to standard deviation (cm) of surface elevation at day 750.

by the polynomial expansions, which agree quite closely with their counterparts estimated from the data comprising the histograms. The kernel densities are clearly not Gaussian; note their narrower peaks, their skewness, and at day 300 an indication of bi-modality.

7. Convergence

One method for assessing convergence of polynomial chaos expansions is by examining the convergence of variance. This can be done by examining the fraction of variance associated with the polynomials of highest degree that are retained:

$$f_K = \frac{\sigma_K^2 - \sigma_{K-1}^2}{\sigma_K^2}, \tag{15}$$

where σ_K^2 is the variance at the Kth level of truncation. Note that the number of terms contributing to the numerator increases with increasing K; with triangular truncation used here there are $K+1$ terms with total degree K. Nevertheless, if the series is converging, then f_K should decrease with increasing K.

The upper panels of Fig. 11 shows contour plots of f_5 and f_6 for the terminal fraction of the variance of surface elevation in each grid cell for day 750. The left panel, which shows the fraction of variance associated with the 5th-degree terms when 6th-degree and higher terms are dropped. Similarly, the right panel shows that associated with the 6th-degree terms, which are the highest that we computed. Overall, the terminal fraction of variance is seen to decrease, but there are a few small regions where, although the fraction is small at both truncations, it can be seen to increase. This is an indication that, if they could be afforded, more terms should be used.

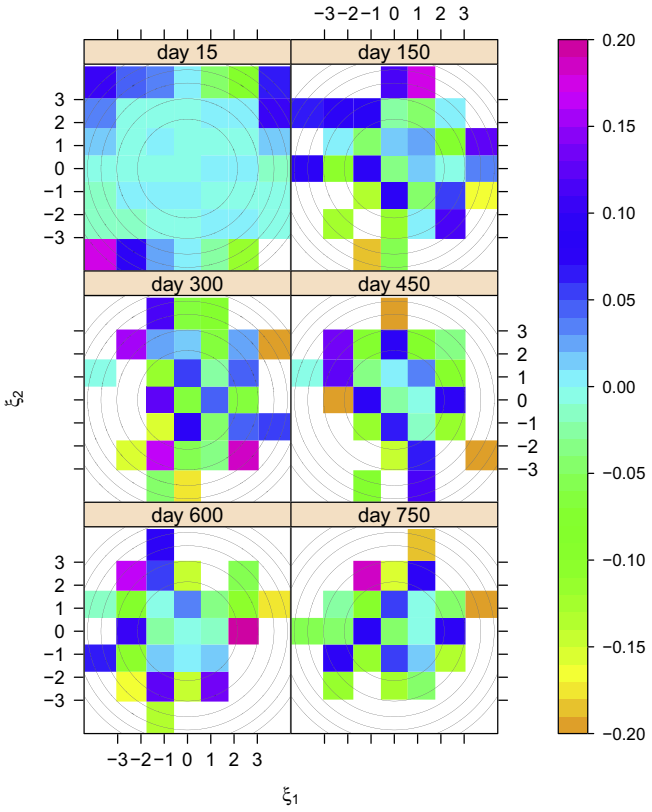

Fig. 12. Errors (m) of the polynomial chaos expansion for sea-surface-height at (86°E, 24.1°N). The color of each rectangle indicates the difference between the HYCOM simulation and its approximation by the polynomial chaos expansion at each quadrature point, with white indicating errors larger than 20 cm. The circles are contours of the probability density function as in Fig. 4.

Another method for assessing convergence is the incremental contribution of the terms of highest degree to the standard deviation. As each term in the expansion for variance is positive, the incremental contribution to the standard deviation is necessarily positive, and it should be seen to decrease to indicate convergence. The lower panels of Fig. 11 indicate that the contributions to the standard deviation of surface elevation are generally smaller when going from 5th to 6th degree, but this is not the case everywhere. These small increases again suggest that more terms might be needed in the polynomial expansions for surface elevations in some parts of the Gulf. However, as retaining polynomials of higher degree requires a larger quadrature ensemble and increases the computational costs, the benefits of greater convergence must be balanced against the imprecision with which the boundary uncertainties can be quantified. For practical purposes convergence might be regarded as being adequate almost everywhere.

A third way to judge convergence is by how well the polynomial expansions reproduce values computed during the 49 quadrature runs. For example, Fig. 12 shows the errors in surface elevation at (86°E, 24.1°N) when the polynomial chaos expansions are truncated to exclude terms with polynomials of total degree greater than six. The low-probability boundary conditions, which were associated with unreasonable extremes of surface elevation in Fig. 9 are also associated with the largest errors. In order to focus on the more likely boundary conditions, the color scale was limited to errors of ±20 cm. Errors at all quadrature points within the inner circle indicating the region of 90% most likely boundary conditions are within the limits of the color scale, as are most of those within the 99% annulus. The larger disagreements at the quadrature points corresponding to unlikely boundary conditions is another indication that more terms are needed for convergence, while the reasonable agreement at the central points suggests that convergence is sufficient for practical use.

8. Discussion and conclusions

The objective of this paper has been to introduce to the oceanographic community the method of polynomial chaos expansions for propagating uncertainties through a dynamical system. In particular, we wanted to demonstrate that it could be used with state-of-the-art numerical models such as HYCOM. The examples illustrated how uncertainties in the specification of the flow from the Caribbean Sea into the Gulf of Mexico manifest as uncertainties in surface elevation as characterized by the mean and standard deviation at each grid cell, by their covariances, and by estimates of their probability density functions.

The novel aspect of the paper is its treatment of the uncertainties of the open boundary conditions. Because practicality limits the number of parameters characterizing the uncertainties that are to be propagated, it was essential to reduce the spatially and temporally varying multivariate description of the inflow to a small number of random variables. The solution presented here has been to assume first that the patterns of uncertainty are similar to the patterns of variability of a climatology of the boundary flow and second that they can be approximated by the first two terms of a modal decomposition. As the two modes are uncorrelated, their random amplitudes could be assumed to be statistically independent and the joint probability density function for the amplitudes could be the product of individual densities. Furthermore, as the modal decomposition partitions climatological variance about the mean, it seemed most appropriate to take the individual densities to be Gaussians.

This treatment of the boundary uncertainties is simply one of convenience necessitated by the lack of a quantitative understanding of the actual uncertainties of the specification of the evolving state of the inflow. Conceptually, a similar but better treatment would be based not on the available climatology but on a large ensemble of runs of the outer model, which provides the boundary conditions for the Gulf model, with each member of that ensemble reflecting uncertainties in the outer model's specification. A modal decomposition of the boundary conditions provided by that ensemble would reflect the uncertainties of the actual flow. Unfortunately, such computations are prohibitively expensive. The available climatology provided a proxy for that ideal ensemble of simulations.

The uncertainties in the boundary conditions were propagated as though the initial conditions were perfectly known, even though uncertainties in earlier flow through the open boundary would have contributed to the uncertainties of the initial conditions. Consequently, there was an initial period during which the effect of the boundary uncertainties accumulated within the Gulf. If the hypothetical large ensemble of outer-model simulations of the previous paragraph had been initiated much earlier to allow for an uncertainty spin-up, then there would be uncertainties in the interior of the Gulf that would be compatible with those at the open boundary, and a modal decomposition might reduce this larger set of data to a few parameters that could characterize both the open boundary and the initial conditions. In the absence of that hypothetical ensemble, it seemed appropriate to spin-up the interior uncertainties from a completely certain initial state.

The polynomial expansions are most valuable when they converge rapidly. Here, the issue of convergence of the polynomial expansions has been illustrated by showing how retaining more terms impacts the variance of the surface elevation. The results indicated that convergence was adequate everywhere within the Gulf of Mexico, given the paucity of information about the nature

of the uncertainties of the inflow and the computational cost of achieving higher precision.

The method of polynomial chaos expansions can be regarded as occupying the middle ground between Monte Carlo methods and Kalman filtering (Evensen, 2009). Like Monte Carlo methods, polynomial chaos expansion attempt to obtain complete information about the statistical distributions of the outputs of a dynamical system, whereas Kalman filtering presumes those distributions can be characterized adequately by an error-covariance matrix. But while Monte Carlo methods require a very large ensemble of solutions to the dynamical system, polynomial expansions need a much smaller ensemble to provide values for evaluating expansion coefficients.

The essential difficulty faced by all methods is the curse of dimensionality: the computational burden grows geometrically with increasing number of uncertain parameters. While Kalman filtering treats each state variable as being uncertain, for oceanographic applications where models have huge numbers of uncertain inputs, the focus is on finding computationally efficient ways to capture important aspects of the uncertainties using as few parameters as possible. The ensemble Kalman filter, for example, uses an ensemble of simulations comparable in size to the quadrature ensemble used in this study, but the appropriate composition of the ensemble is still a matter of study.

One question that was not explored here is that of the convergence of the quadrature approximations of the integrals. This becomes increasingly important as the number of uncertain inputs increases. Here we illustrated how two-dimensional quadrature requires the square of the number of simulations needed for the individual one-dimensional quadratures, which is simply another manifestation of the curse of dimensionality. In the future we plan to explore the use of sparse quadrature methods for handling several uncertain parameters and to take a closer look at convergence in quadrature.

Acknowledgments

The research herein was supported by the Office of Naval Research, under Grant No N000141010498. Dr. Iskandarani was also supported by NSF-OCE0622662. All calculations were performed at the Center for Computational Science at the University of Miami. The authors would like to thank the reviewers for their helpful suggestions.

References

Abramowitz, M., Stegun, I., 1970. Handbook of Mathematical Functions. Dover.
Bellman, R., 1957. Dynamic Programming. Princeton University Press, Princeton NJ.
Bennett, A.F., 2002. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, Cambridge.
Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modell. 37, 55–88.
Cameron, R.H., Martin, W.T., 1948. Transformations of Wiener integrals under translations. Ann. Math. 45 (2), 386–396.
Chassignet, E.P., Hurlburt, H.E., Smedstad, O.M., Halliwell, G.R., Hogan, P.J., Wallcraft, A.J., Baraille, R., Bleck, R., 2007. The HYCOM (Hybrid Coordinate Ocean Model) data assimilative system. J. Marine Syst. 65, 60–83.
Conti, S., O'Hagan, A., 2010. Bayesian emulation of complex multi-output and dynamic computer models. J. Stat. Plan. Infer. 140, 640–651.
Evensen, G., 2009. Data Assimilation: The Ensemble Kalman Filter. Springer-Verlag, Berlin.
Finette, S., 2006. A stochastic representation of environmental uncertainty and its coupling to acoustic wave propagation in ocean waveguides. J. Acoust. Soc. Am. 12, 2567–2579.
Ge, L., Cheung, K.F., Kobayashi, M.H., 2008. Stochastic solution for uncertainty propagation in nonlinear shallow-water equations. J. Hydraulic. Eng. 134, 1732–1743.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, London.
Griewank, A., Corliss, G.F. (Eds.), 1991. Automatic Differentiation of Algorithms: Theory, Implementation, and Application. SIAM Proceeding Series. Society for Industrial and Applied Mathematics, Philadelphia.
Karniadakis, G.E., Glimm, J., 2006. Uncertainty quantification in simulation science. J. Comput. Phys. 217.
Knio, O.M., Le Maître, O.P., 2006. Uncertainty propagation in CFD using polynomial chaos decomposition. Fluid Dyn. Res. 38, 616–640.
Kourafalou, V.H., Pang, G., Kang, H., Hogan, P.J., Smedsted, O., Weisberg, R.H., 2009. Evaluation of global ocean data assimilation experiment products on south Florida nested simulations with hybrid coordinate ocean model. Ocean Dyn. 59, 47–66.
Le Maître, O.P., Knio, O.M., 2010. Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics. Scientific Computation Series. Springer, Berlin.
Le Maître, O.P., Reagan, M.T., Najm, H.N., Ghanem, R.G., Knio, O.M., 2002. Uncertainty propagation in CFD using polynomial chaos decomposition. J. Comput. Phys. 181, 9–44. doi:10.1006/jcph.2002.7104.
Lermusiaux, P.F.J., 2006. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. J. Comput. Phys. 217, 176–199. doi:10.1016/j.jcp.2006.02.010.
Lorenz, E.N., 1963. Deterministic non-periodic flows. J. Atmos. Sci. 20, 130–141.
Lorenz, E.N., 1984. Irregularity: a fundamental property of the atmosphere. Tellus 36A, 98–110.
Najm, H.N., 2009. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. Ann. Rev. Fluid Mech. 41, 35–52.
Prasad, T.G., Hogan, P.J., 2007. Upper-ocean response to Hurricane Ivan in a 1/25° nested Gulf of Mexico HYCOM. J. Geophys. Res. 112. doi:10.1029/2006JC003695.
Sapsis, T.P., Lermusiaux, P.F.J., 2009. Dynamically orthogonal field equations for continuous stochastic dynamical systems. Physica D 238, 2347–2360. doi:10.1016/j.physd.2009.09.017.
Shen, C.Y., Evans, T.E., Finette, S., 2010. Polynomial chaos quantification of the growth of uncertainty investigated with a lorenz model. J. Atmos. Oceanic Technol. 27, 1059–1071. doi:10.1175/2009JTECHO727.1.
Tatang, M.A., Pan, W., Prinn, R.G., McRae, G.J., 1997. An efficient method for parametric uncertainty analysis of numerical geophysical models. J. Geophys. Res. 102 (21), 925–21,932. doi:10.1029/97JD01654.
Webster, M.D., Sokolov, A.P., 2000. A methodology for quantifying uncertainty in climate projections. Climate Change 46, 417–446. doi:10.1023/A:1005685317358.
Wiener, N., 1938. The homogenous chaos. Am. J. Math. 60, 897–936.
Witteveen, J., Bijl, H., 2006. Modeling arbitrary uncertainties using Gram–Schmidt polynomial chaos. In: 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, AIAA-2006-896.
Wunsch, C., 2006. Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications. Cambridge University Press, Cambridge.
Xiu, D., 2009. Fast numerical methods for stochastic computations: a review. Commun. Comput. Phys. 5, 242–272.