Propagating boundary uncertainties using polynomial expansions

W. C. Thacker — UM/CIMAS & NOAA/AOML
A. Srinivasan — University of Miami, CCS
M. Iskandarani — University of Miami, MPO
O. M. Knio — Duke University
M. Le Henaff — University of Miami, CIMAS

1. Introduction

This study explores the uncertainty of the Circulation within the Gulf of Mexico resulting from the uncertainty of the inflow through the Yucatan Straits. It requires first to characterize the uncertainties of the inflow and then to propagate them dynamically so that they manifest in the circulation at later times. Here, the nature of the uncertainty of the inflow is assumed to be similar to its simulated climatological variability, and polynomial expansions for each simulated variable are used to propagate this inflow uncertainty.

2. Characterizing the uncertainty of the Yucatan inflow

As our high-resolution HYCOM simulations of the Gulf's circulation requires open boundary conditions from a lower-resolution simulation of a larger region, the flow specified at the southern open boundary provide a convenient proxy for the Yucatan inflow. Quantifying its uncertainty requires deciding the nature and likelihood of possible deviations from these boundary conditions. As there is no observational basis for quantifying this uncertainty, we proceed under the assumption that deviations from the boundary conditions are proportional to deviations from a long-term mean of simulated boundary flow. Because the computational requirements for propagating the boundary uncertainties increase geometrically with increasing numbers of parameters used to characterize them, it is important to use as few parameters as possible to describe their multivariate and spatiotemporal nature. This is achieved by decomposing the deviations into spatiotemporal patterns and using the first two, which account for 42% of the boundary variability, as proxies for the uncertainty of the specified boundary conditions. Thus, the possible boundary values are described as:

$$X(\xi_1,\xi_2) = X_0 + \alpha\left(\xi_1\lambda_1 c_1 r_1^T + \xi_2\lambda_2 c_2 r_2^T\right) , \qquad (1)$$

where X is a matrix of boundary values, each row corresponding to a simulated time and each column to one of HYCOM's boundary values at that time, where the column vectors c_1 and c_2 are the EOFs, the row vectors r_1^T and r_2^T are the principal components, λ_1 and λ_2 are the singular values. Uncertainty is parametrized through ξ_1 and ξ_2, which are unit-variance, uncorrelated, normally distributed random amplitudes reflecting the uncertainty of the boundary values. 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 X_0 to be more reliable, a smaller value would have been be more appropriate.

3. Propagating uncertainty with polynomial expansions

Any simulated output y is linked to the boundary uncertainties through the assumption that it can be well approximated by a polynomial of the random amplitudes ξ_1 and ξ_2:

$$y(\xi_1,\xi_2) = \sum_{k_1,k_2}^{k_1+k_2\le K} y_{k_1,k_2} P_{k_1}(\xi_1) P_{k_2}(\xi_2) + \epsilon_K(\xi_1,\xi_2) , \qquad (2)$$

where triangular truncation retains polynomials of total degree no greater than $K = 6$ and where ϵ_K represents truncation error. The orthogonality of the Hermite polynomials P_k provides an expression for the expansion coefficients:

$$y_{k_1,k_2} = \frac{1}{N_{k_1}N_{k_2}} \int\int y(\xi_1,\xi_2) P_{k_1}(\xi_1) p(\xi_1) d\xi_1 P_{k_2}(\xi_2) p(\xi_2)\; d\xi_2 , \qquad (3)$$

where the constants N_k account for the normalization convention and where $p(\xi)$ is the standard Gaussian probability density. Because the time series r_1 and r_2 are uncorrelated, the random variables ξ_1 and ξ_2 are presumed to be independent and to have joint density $p(\xi_1)p(\xi_2)$.

4. Quadrature ensemble

The expansion coefficients for any quantity simulated by HYCOM can be evaluated by Gauss-Hermite quadrature one integral at a time:

$$\int\int y(\xi_1,\xi_2) P_{k_1}(\xi_1) p(\xi_1) d\xi_1 P_{k_2}(\xi_2) p(\xi_2) d\xi_2 \approx \sum_{q_1}\sum_{q_2} y(\xi_{q_1},\xi_{q_2}) P_{k_1}(\xi_{q_1}) w_{q_1} P_{k_2}(\xi_{q_2}) w_{q_2} , \qquad (4)$$

where (ξ_{q_1},ξ_{q_2}) are quadrature points where values of y must be known and where w_{q_1} and w_{q_2} are corresponding weights. The principal cost of evaluating the coefficients is that of the HYCOM simulations needed to get values y at the quadrature points.

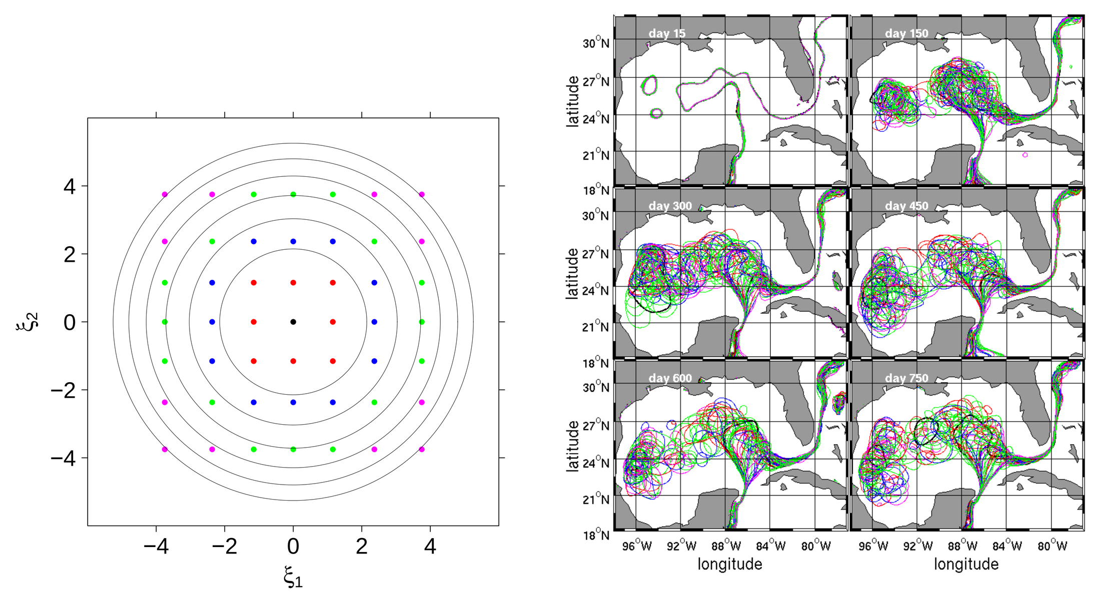

Figure 1: *Left: Circles enclose regions of 90%, 99%, . . ., 99.9999% probability. Dots mark locations of 49 Gauss-Hermite quadrature points, with red dots corresponding to relatively likely, blue less likely, green unlikely, and magenta highly unlikely boundary conditions. Right: Locations of the Loop Current and its eddies from 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 figure 1 with the thick black contour indicating the central member of the ensemble.*

5. Uncertainty of the surface elevation field

After the ensemble of quadrature simulations has been run, it is simple to evaluate the uncertainty of any variables that have been saved. For example, an estimate of the mean of y is given by the polynomial expansion's constant term $y_{0,0}$. An estimate of its variance is given by the sum of the squares of the coefficients of the non-constant terms. Similarly its covariance with another variable is estimated by sums of products of their coefficients. Note that these estimates are *not* the same as simple averages of the quadrature ensemble; the choice of the quadrature run allows the polynomial expansions to offer considerably more accurate statistics than the ensemble average.

Such estimates are shown below for the surface-elevation field at six different times. Because all members of the quadrature ensemble start from the same initial conditions, there has been no accounting for previous boundary uncertainties. Consequently there is an initial period during which boundary uncertainties propagate into the Gulf, and out the Florida Straits while building uncertainty distributions for all model variables. This build-up of uncertainty can be seen as increasing standard deviations and covariances for days 15 and 150.

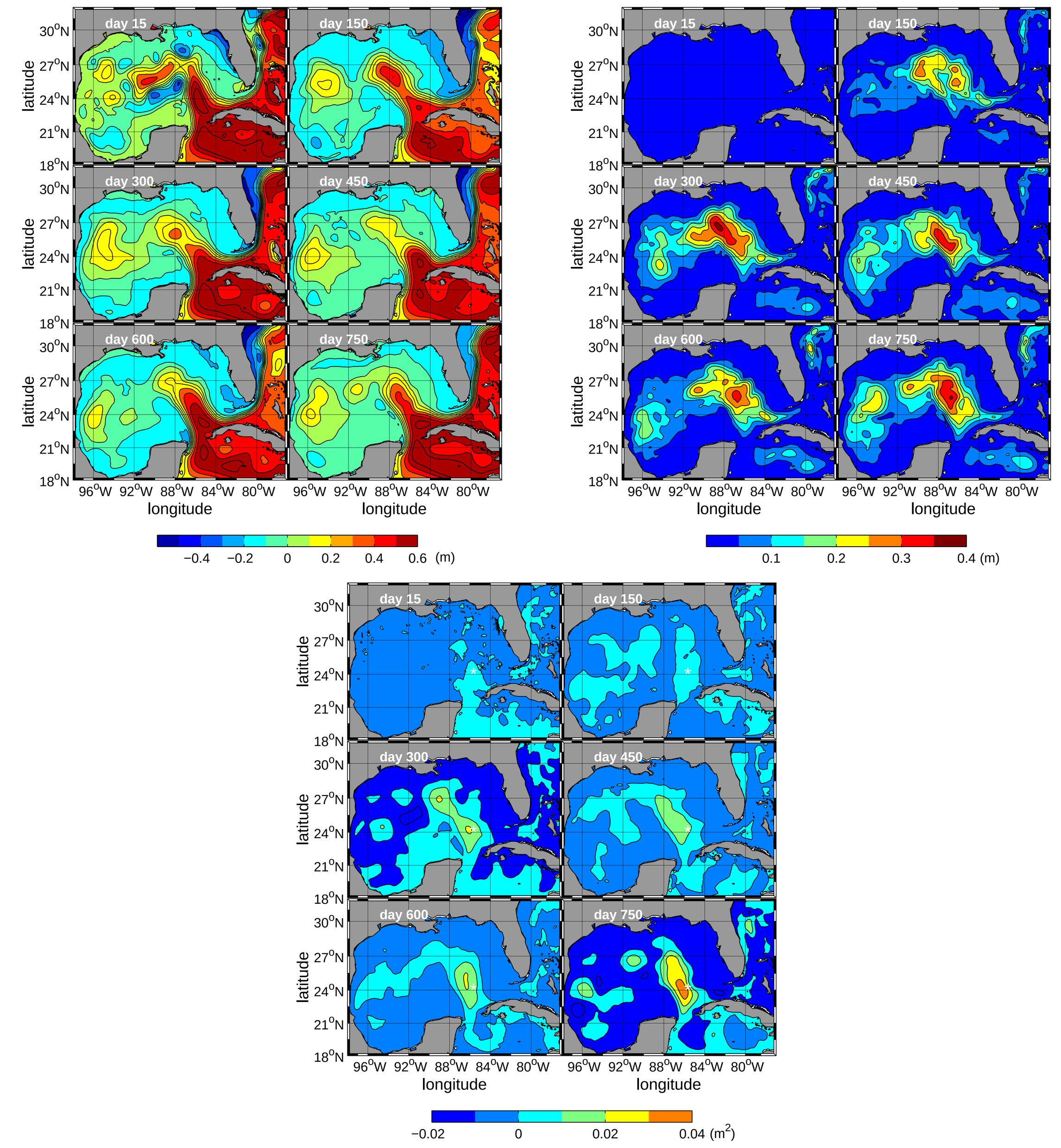

Figure 2: *Left: Mean (m) of the sea-surface-height field from the polynomial chaos expansion at 15, 150, 300, 450, 600, and 750 days after the boundary uncertainties were initiated. Right: Standard deviation (m) of the sea-surface-height field from the polynomial chaos expansion at same days. Bottom: 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.*

6. Convergence of the polynomial expansion

As variance and standard deviation increase as more terms are retained in the polynomial expansion, convergence can be examined by comparing different levels of truncation. Figure 3 shows that going from 5th to 6th degree polynomials contributes significantly less to the standard deviation of surface elevation than does going from 4th to 5th.

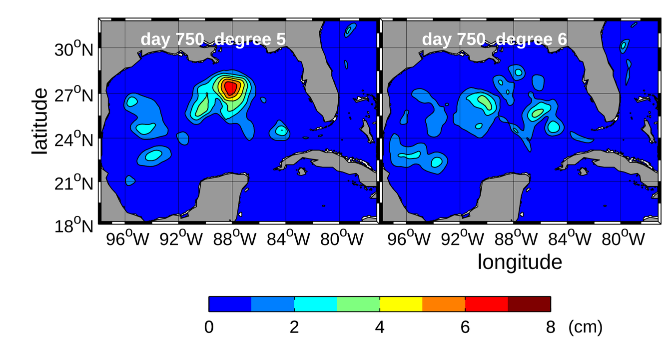

Figure 3: *Incremental contribution to standard deviation (cm) of surface elevation at day 750. Left: contribution of the 6 5th-degree terms relative to the total contributed by the 21 terms of degree less than 6. Right: contribution of the 7 6th-degree terms relative to the total contributed by the 28 terms of degree less than 7.*

7. Uncertainty at one field point

The polynomial expansions provide inexpensive alternatives to the simulations for boundary inflows that were not in the quadrature ensemble. For any value of the amplitudes ξ_1 and ξ_2, there is a polynomial approximation for y. The left panel of figure 4 shows the response surface for the surface elevation at a high-variance point through which the Loop Current often passes for both likely and unlikely boundary flows. To give an idea of the accuracy of this response surface the right panel shows difference between the simulated surface elevation at this point and for the 49 ensemble members and corresponding values from the polynomial expansions; errors for likely inflows are reasonable in magnitude in spite of choosing $\alpha = 1$ while unlikely inflows are poorly approximated. The polynomial expansions also allow for an inexpensive synthetic ensemble, which provides information about the non-Gaussian nature of the probability densities for simulated variables. An example is shown in the lower panel.

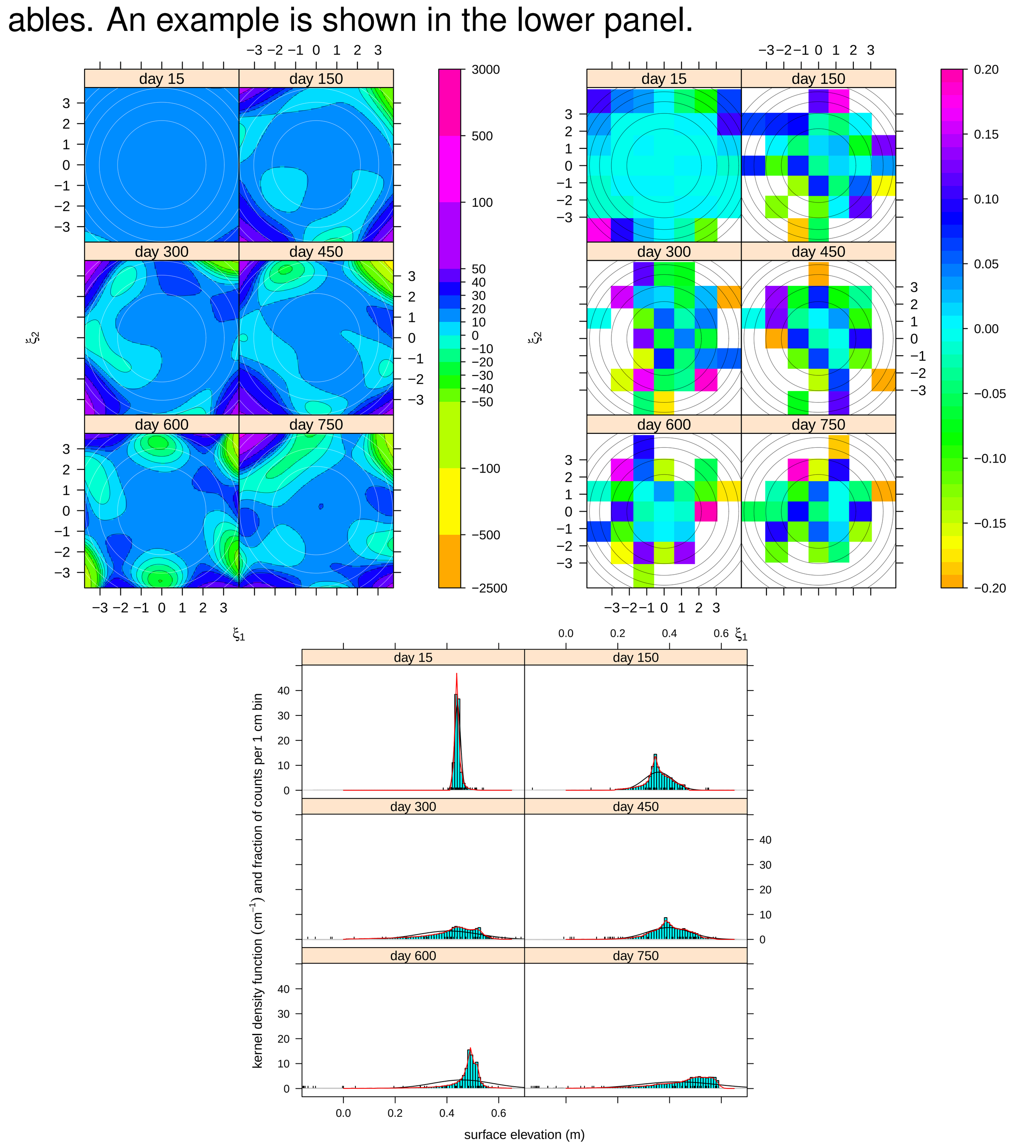

Figure 4: *Left: Surface elevation (cm) at (86°E,24.1°N) as a function of random variables ξ_1 and ξ_2. Note the compressed color scale used to distinguish more likely from less likely responses: Contours are at 10 cm intervals from -50 to 50 cm, and more extreme values are represented with a compressed scale. The circles indicate that the extreme values are highly unlikely. Right: 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. Bottom: Kernel density estimates for surface elevation (m) at the point (86°E, 24.1°N) derived from histograms generated using polynomial chaos expansion 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.*

8. Conclusion

Polynomial expansion provide a relatively inexpensive way to explore the consequences of uncertainties in a model's inputs. The approach taken here illustrates the need for quantifying the uncertainties of the inputs. In particular, because the expense increases geometrically with the number of uncertain inputs to be examined, it is important to focus on only the few that are the most important. In this regard, the method has similarities to Kalman filtering as applied to oceanographic and meteorological models. And like Kalman filtering, polynomial expansions offer the possibility of updating prior estimates of input uncertainties by exploiting observational data, although that aspects has not been explored here.

Carlisle.Thacker@noaa.gov