Application and validation of polynomial chaos methods to quantify uncertainties in simulating the Gulf of Mexico circulation using HYCOM.

Mohamed Iskandarani  Matthieu Le Hénaff  Carlisle Thacker
University of Miami

Guotu Li  Omar Knio
Duke University

Ashwanth Srinivasan
Tendral LLC

June 2, 2015
Outline

- Use Polynomial Chaos (PC) expansions to quantify uncertainties in oceanic forecasts
  - Approximate model with an accurate surrogate
  - Compute series coefficients via an ensemble
  - Validate the accuracy of the surrogate
  - Mine the series for statistical information in lieu of model
- Initial Condition uncertainties in the Gulf of Mexico
  - How to perturb a field?
  - How to localize perturbations around a dynamical process?
  - Series Validation
  - Statistical Outputs
- Uncertainty Experiments to date
  - Initial conditions uncertainties (small ensemble)
  - Initial conditions and wind forcing uncertainties
HYCOM surrogate idea

- $M(x, t, \xi)$ is a model output
- $\xi$ is a stochastic variable that represents the dependence of $M$ on the uncertain input data
- $\xi$ is characterized by its probability density function $\rho(\xi)$
- The mean of $M$ is: $\overline{M}(x, t) = \int M(x, t, \xi) \rho(\xi) \, d\xi$
- Its variance is
  $$\sigma^2(M) = \overline{(M - \overline{M})^2} = \int (M - \overline{M})^2 \rho(\xi) \, d\xi$$
- A surrogate embodies the dependence of $M$ on the uncertain data $\xi$ via a spectral series in $\xi$:
  $$M(x, t, \xi) = \sum_{n=0}^{P} \hat{M}_n(x, t) \psi_n(\xi)$$
PC surrogate series $M(x, t, \xi) = \sum_{n=0}^{P} \hat{M}_n(x, t) \psi_n(\xi)$

- $\hat{M}_n(x, t)$: series coefficients
- $\psi_k(\xi)$: orthonormal basis functions w.r.t. $\rho(\xi)$

\[
\overline{\psi_m \psi_n} = \int \psi_m(\xi) \psi_n(\xi) \rho(\xi) d\xi = \delta_{m,n}
\]

- $\psi_m(\xi)$ consists of Legendre polynomials when $\rho(\xi)$ is a uniform distribution
- Mean: $\overline{M} = \sum_{n=0}^{P} \hat{M}_n(x, t) \overline{\psi_n} = \hat{M}_0(x, t)$
- Variance: $(M - \overline{M})^2 = \sum_{n=1}^{P} \hat{M}_n^2(x, t)$
- Where to truncate the series, $P$? Monitor Variance
- How to determine the coefficients $\hat{M}_n$?
Calculating the PC series coefficients

- Minimize the norm of the approximation error by using either Galerkin projection, interpolation, least square, or compressed sensing. All can be implemented via ensemble.
- The least square approaches are useful when model response includes model noise.
- Galerkin projection exploits orthogonality

\[
\hat{M} = \overline{M\psi_m} = \int M(x, t, \xi)\psi_m(\xi)\rho(\xi)d\xi \\
\approx \sum_{q=1}^{Q} M(x, t, \xi_q)\psi_m(\xi_q)\omega_q
\]

- \(\xi_q, \omega_q\) are appropriate quadrature roots, weights
- \(M(x, t, \xi_q)\) requires an ensemble at the quadrature roots
Gulf of Mexico Circulation

Figure: Sea Surface Height in cm from AVISO

Target Loop Current Eddy separation during May-June 2010. Capture the role of the frontal anticyclones in eddy detachment.
Uncertainty in Initial Boundary Conditions

- Computational challenge: Number of sample grows exponentially with the number of stochastic variables.
- Rely on EOFs to characterize uncertainty and reduce the number of stochastic variables. For 2 EOFs mode we have:

\[
M(x, 0, \xi_1, \xi_2) = \overline{M}(x, 0) + \left[ \sqrt{\lambda_1} M_1 \xi_1 + \sqrt{\lambda_2} M_2 \xi_2 \right] \quad (1)
\]

- \((\lambda_k, M_k)\): are eigenvalues/eigenvectors of covariance matrix obtained from free-run simulation
- \(\overline{M}\): unperturbed initial condition
- \(M(x, 0, \xi)\): Stochastic initial condition input
- \(\xi_1, \xi_2\) are the amplitudes of the perturbations
Figure: First and Second SSH modes from a 14-day series. The 2 modes account for 50% of variance during these 14 days.

- Characterize local uncertainty: get perturbation from short, 14-day, simulation.
- Uncertainty dominated by Loop Current (LC) dynamics
- Mode 1 seems associated with a frontal eddy
**Figure:** Vertical slice along 26.4N showing Temperature perturbations. The first mode shows a strong 2.5°C cooling in the vicinity of the frontal cyclones. The "warm" perturbation around 90W is at the southern edge of a small anticyclone NW of the LC.
PC representation

- \((\xi_1, \xi_2)\) independent and uniformly distributed random variables
- PC basis: Legendre polynomials of max degree 6, \(P = 28\)
- Ensemble of 49 realizations for Gauss-Legendre quadrature

**Figure**: Quadrature/Sample points in \(\xi_1, \xi_2\) space. Center black circle corresponds to unperturbed run, while blue circles correspond to largest negative and positive perturbations.
Col 1: SSH of realization (1,1) with weakest frontal eddy

Col 2: SSH of unperturbed realization (4,4) has medium strength frontal eddy

Col 3: SSH of realization (7,7) has strongest frontal eddy and earliest LC separation

Col 4: Loop current edge in ensemble
SSH stddev (cm) grows in time with maximum in LC region
PC-error: \[ \| \epsilon \|_2^2 = \sum_q [\eta(\vec{x}, t, \xi_q) - \eta_{PC}(\vec{x}, t, \xi_q)]^2 \omega_q \]

SSH PC-errors (cm) grow in time with maxima in LC region.

On day 60 PC-error is about 38% of stddev.
Figure: Variance Analysis: The majority of the variance in the deep part of the Eastern Gulf of Mexico can be attributed to the 1st EOF, while the second mode plays a secondary role, particularly during the time span when the series is reliable (< 40 days, x-axis tick marks interval is 5-days).
Figure: **Predictability Limit**: Spatial distribution of the ratio of the forecast standard deviation to climatology standard deviation for SSH (from AVISO). The magenta lines show areas where the ratio > 1.
Initial Conditions & Wind Forcing Uncertainties

Focus on the following two Quantities of Interest (QoIs):

- Sea Surface Height (SSH) averaged over a square area near the loop current (LC) region: \([-86.04^\circ, -85.20^\circ]\) in longitude and \([25.19^\circ, 26.23^\circ]\) in latitude

- Mixed Layer Depth (MLD) averaged over the DeepWater Horizon (DWH) region: \([-88.44^\circ, -88.28^\circ]\) in longitude and \([28.68^\circ, 28.79^\circ]\) in latitude

Our HYCOM simulations start from 05-01-2010 to 05-30-2010. Fig. 7 on the left shows HYCOM results on the last day of simulation using unperturbed initial and wind forcing fields.

Figure: Field snapshots on day 30: (Top) SSH; (Bottom) MLD
Variance Analysis

1st Order Sensitivity:

\[ S_i = \frac{\sum_{\alpha \in S_i} c_\alpha^2 < \psi_\alpha, \psi_\alpha >}{\sum_{P} c_i^2 < \psi_i, \psi_i >} \]

Total Order Sensitivity:

\[ T_i = \frac{\sum_{\alpha \in S_T} c_\alpha^2 < \psi_\alpha, \psi_\alpha >}{\sum_{i=1}^P c_i^2 < \psi_i, \psi_i >} \]

**Figure:** Sensitivities: (Top) 1st Order; (Middle) total Order; (Bottom) Initial and Wind forcing sensitivities.
SSH: Joint Sensitivity

(a) Interaction

(b) Initial Condition

(c) Wind Forcing
MLD: Joint Sensitivity

(d) Interaction

(e) Initial Condition

(f) Wind Forcing
Conclusions

- Polynomial Chaos are a promising approach in analyzing oceanic uncertainties.
- It can monitor representation fidelity and adequacy of sampling.
- The 14-day time series EOF analysis helped in designing perturbations focused on a Loop Current separation event.
- The PC-Series performs reasonably well in the 20–40 days range.
- Predictability in SSH lost after about 20 days.
- It is probably enough to perturb only the 2 leading EOF modes.
- Including more uncertainty sources is more important for the current experiment than including more modes.
- SSH uncertainty due mainly to initial conditions except in shelf areas.
- MLD uncertainty mainly caused by winds.
Publications


