Quantifying Initial Condition and Wind Forcing Uncertainties in the Gulf of Mexico

Guotu Li [1], Mohamed Iskandarani [2], Matthieu Le Henaff [2], Olivier Le Maître [3] and Omar M. Knio [1]

[1]Duke University, Durham, NC, USA [2]University of Miami, Miami, FL, USA [3]LIMSI CNRS, Orsay, France

Abstract

This study aims at understanding the combined effects of uncertainties originating from initial condition and wind forcing fields in Ocean General Circulation Models (OGCM) using Polynomial Chaos (PC) expansions. Empirical Orthogonal Functions (EOF) are used to formulate both initial condition and wind forcing perturbations in forms of superposition of modal components with uniformly distributed random amplitudes. The forward deterministic hybrid coordinate ocean model (HYCOM) is used to propagate input uncertainties in simulation of the Gulf of Mexico (GoM). A Latin Hypercube (LH) ensemble of model realizations is used to construct PC surrogates; a basis pursuit denoising (BPDN) is used for this purpose, and the resulting expansions are validated through various statistical measures. Surrogates are generated for both localized and global Quantities of Interest (QoIs). Attention is focused on the sea surface height (SSH) and mixed layer depth (MLD). Global sensitivity analyses are used to quantify the dynamical implications of the interactions among different input uncertainties. Finally, a sub-sampling reconstruction test shows the possibility of utilizing a small number of OGCM realizations to build PC surrogates for field quantities with considerable fidelity in prediction statistics.

Setup

Dimension Reduction: Direct propagation of uncertainties in field quantities (e.g. initial and wind forcing fields) can lead to millions of stochastic dimensions depending on the model resolution. To make the problem tractable, EOF decompositions are used to represent uncertainties in initial condition and wind forcing fields.

$$\begin{cases} u(\boldsymbol{x}, t=0, \boldsymbol{\xi}_a) &= \bar{u}(\boldsymbol{x}, \boldsymbol{\xi}_a = \boldsymbol{0}) + 0.8 \times \left[\sqrt{\lambda_1}\mathcal{U}_1, \sqrt{\lambda_2}\mathcal{U}_2, \sqrt{\lambda_3}\mathcal{U}_3, \sqrt{\lambda_4}\mathcal{U}_4\right]\boldsymbol{\xi}_a^T \\ f(\boldsymbol{x}, t, \boldsymbol{\xi}_b) &= \bar{f}(\boldsymbol{x}, \boldsymbol{\xi}_b = \boldsymbol{0}) + 0.8 \times \left[\sqrt{\eta_1}\mathcal{F}_1, \sqrt{\eta_2}\mathcal{F}_2, \sqrt{\eta_3}\mathcal{F}_3, \sqrt{\eta_4}\mathcal{F}_4\right]\boldsymbol{\xi}_b^T \end{cases} \quad (1)$$

- $\bar{u}$, $\bar{f}$: unperturbed initial and wind forcing fields, respectively.
- $\boldsymbol{x}$: spatial coordinates.
- $\boldsymbol{\xi}_a = [\xi_1, \xi_2, \xi_3, \xi_4]$, $\boldsymbol{\xi}_b = [\xi_5, \xi_6, \xi_7, \xi_8]$: stochastic random amplitude vectors corresponding to initial condition and wind forcing modes, respectively; ξ_i's are independent and uniformly distributed over $[-1, 1]$.
- $(\lambda_i, \mathcal{U}_i)$, $(\eta_i, \mathcal{F}_i)$: eigenvalue/eigenvector pairs of covariance matrices in initial condition and wind forcing, respectively.
- Note that wind forcing EOFs are time-dependent as well.

Realization Ensemble: An ensemble of HYCOM realizations over a Latin Hypercube Sample (LHS) set ($\mathcal{P}_{LHS}$, with 798 samples) is generated. All simulations are from May-01 to May-30 in 2010.

Quantities of Interest (QoIs)

We focus on sea surface height (SSH) and mixed layer depth (MLD) in this study and construct PC surrogate models for both localized and global SSH/MLD defined below.

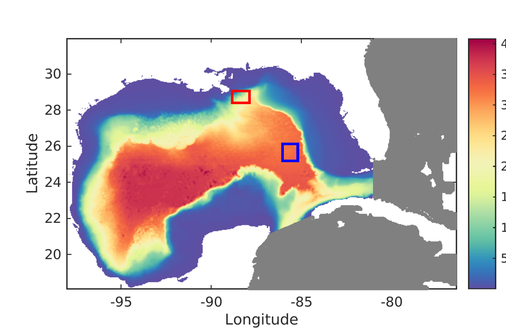

Figure 1: Bathymetry of the Gulf of Mexico, showing the depth in meters. The boxes indicate locations where localized QoIs are estimated.

QoI Definitions:

- Global QoIs: SSH and MLD fields inside the Gulf of Mexico (excluding the grey area in the figure).
- Localized QoIs:
  1. mean SSH averaged over the blue box bounded by $[-86.04°, -85.20°]$ in longitude and $[25.19°, 26.23°]$ in latitude near the loop current (LC) region;
  2. mean MLD averaged over the red box bounded by $[-88.84°, -87.88°]$ in longitude and $[28.40°, 29.07°]$ in latitude centered at the deep water horizon (DWH);

PC Surrogates for Local QoIs

PC Approximation: The real-valued random variable $\mathcal{Y}(\boldsymbol{\xi}) \in L_2(\Theta, P)$ is approximated by a truncated series of orthogonal polynomial basis functions,

$$\mathcal{Y}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} c_k \Psi_k(\boldsymbol{\xi}) \approx \sum_{k=0}^{N_p} c_k \Psi_k(\boldsymbol{\xi}) \triangleq \mathcal{S}(\boldsymbol{\xi}) \quad (2)$$

BPDN Reconstruction: To recover the coefficients c_k's, the following BPDN problem is solved,

$$minimize \quad ||\boldsymbol{c}||_{L_1} \quad subject\ to \quad ||\boldsymbol{Y} - [\Psi]\boldsymbol{c}||_{L_2} \le \sigma \quad (3)$$

where $\boldsymbol{Y}$ is a QoI realization vector and $\boldsymbol{c}$ is the PC expansion coefficient vector. $[\Psi]$ is the polynomial matrix in which each element $[\Psi]_{i,j} = \Psi_j(\boldsymbol{\xi}_i)$. σ is the error tolerance/noise level estimated by a cross-validation process.

PC Statistics:

$$E(\mathcal{S}) = <\mathcal{S}, \Psi_0> = c_0 \quad (4)$$

$$var(\mathcal{S}) = \sum_{k=1}^{N_p} c_k^2 < \Psi_k, \Psi_k > \ \text{ and } std(\mathcal{S}) = \sqrt{var(\mathcal{S})} \quad (5)$$

Error Estimates for Localized QoIs

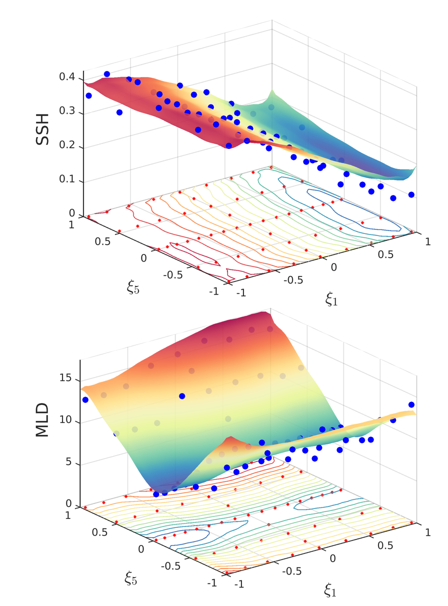

Figure 2: PC response surfaces (day-30) over the selected 2D plane ξ_1-ξ_5, comparing with HYCOM realizations (blue dots)

Estimate errors over the LHS ensemble ($\boldsymbol{\xi}_i \in \mathcal{P}_{LHS}$):

- Normalized Local Error: $\epsilon(\boldsymbol{\xi}_j) = \frac{|\mathcal{S}(\boldsymbol{\xi}_j) - Y(\boldsymbol{\xi}_j)|}{max(\boldsymbol{Y}) - min(\boldsymbol{Y})}$
- Global Error: $Err_{95} = \epsilon|_{CDF=0.95}$

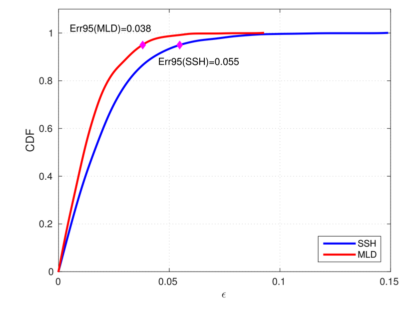

Figure 3: CDFs of local ϵ (day-30)

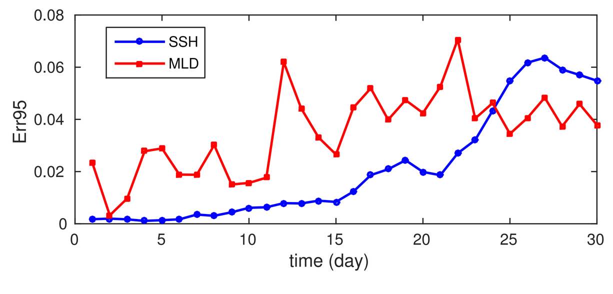

Figure 4: Time evolution of global PC errors

Remarks:

- The relatively small global errors ($Err95$) guarantee faithful PC predictions.
- PC response surfaces in Fig. 2 indicate that localized SSH is more sensitive to ξ_1 rather than ξ_5, while the MLD response shows the opposite.

Prediction Statistics for Localized QoIs

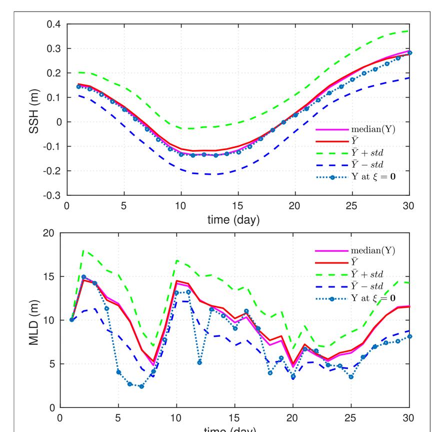

Figure 5: PC predictions of mean and STD bounds, HYCOM realization at $\boldsymbol{\xi} = \boldsymbol{0}$, median QoIs from HYCOM ensemble (Top) SSH; (Bottom) MLD;

Mean and STD:

- SSH curves exhibit smooth evolution in time.
- MLD shows highly oscillatory time dependence, indicating possible fast dynamics in MLD evolution.

Sensitivity Indices:

$$SI = \frac{\sum_{\alpha \in \mathbb{I}} c_\alpha < \Psi_\alpha, \Psi_\alpha >}{var(\mathcal{S})}$$

- The initial condition uncertainty dominates local SSH response; and local MLD is mostly influenced by wind forcing perturbations.
- Interaction between initial condition and wind forcing remains negligibly small (except for MLD on day-22).

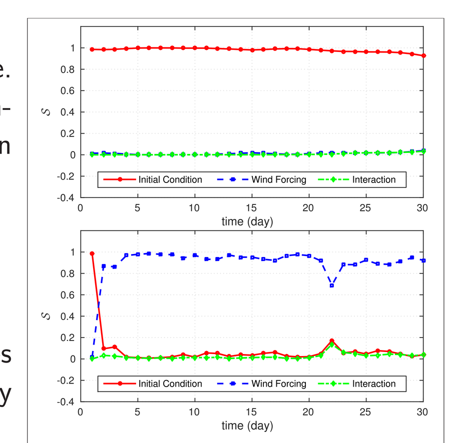

Figure 6: 1st order sensitivities to initial condition, wind forcing and their interactions (Top) SSH; (Bottom) MLD;

PC-Based EOF Reconstruction of Field Quantities

Let $\mathcal{Y}(\boldsymbol{x}_i, \boldsymbol{\xi}_j)$ denotes the HYCOM realization of a field variable at given spatial and stochastic coordinates. Decompose $\mathcal{Y}$ into stochastic mean $\overline{\mathcal{Y}}$ and fluctuation $\tilde{\mathcal{Y}}$:

$$\mathcal{Y}(\boldsymbol{x}_i, \boldsymbol{\xi}_j) = \overline{\mathcal{Y}}(\boldsymbol{x}_i) + \tilde{\mathcal{Y}}(\boldsymbol{x}_i, \boldsymbol{\xi}_j) \quad \text{(or in matrix form:} [\mathcal{Y}]_{i,j} = [\overline{\mathcal{Y}}]_{i,} + [\tilde{\mathcal{Y}}]_{i,j}) \quad (6)$$

The EOF decomposition starts with the following eigenvalue problem:

$$[C]\boldsymbol{\Phi}_k = \lambda_k [W] \boldsymbol{\Phi}_k \quad \text{(where } [W] = \frac{1}{N_{LHS}}[I], [C] = \left([\tilde{\mathcal{Y}}][W]\right)^T \left([\tilde{\mathcal{Y}}][W]\right)) \quad (7)$$

Stochastic modes $\boldsymbol{\Phi}_k$'s are scaled such that $\boldsymbol{\Phi}_k^T [W] \boldsymbol{\Phi}_{k'} = \delta_{kk'}$ and spatial modes are then obtained by:

$$\boldsymbol{U}_k = \frac{[\tilde{\mathcal{Y}}][W]\boldsymbol{\Phi}_k}{\sqrt{\lambda_k}} \quad \text{(Note: } \boldsymbol{U}_k^T \boldsymbol{U}_{k'} = \delta_{kk'}) \quad (8)$$

The truncated EOF decomposition of a target fluctuation field is:

$$\tilde{\mathcal{Y}}(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{k=1}^{r} \sqrt{\lambda_k} \boldsymbol{U}_k(\boldsymbol{x}) \boldsymbol{\Phi}_k(\boldsymbol{\xi}) \quad \text{(where } r = min\{l \mid \sum_{k=1}^{l} \lambda_k \ge p \sum_{k=1}^{N_{LHS}} \lambda_k\}) \quad (9)$$

The above EOF decomposition aims at minimizing the mean-square-error evaluated over the realization ensemble:

$$\epsilon_{L_2}^2 = \frac{1}{N_{LHS}} \sum_{i=1}^{N_{LHS}} || \sum_{k=1}^{r} \sqrt{\lambda_k} U_k(\boldsymbol{x}) \Phi_k(\boldsymbol{\xi}_i) - \tilde{\mathcal{Y}}(\boldsymbol{x}, \boldsymbol{\xi}_i) ||_{\Omega}^2 / \sum_{j=1}^{N_{LHS}} \lambda_j \quad (10)$$

SSH & MLD Field Properties

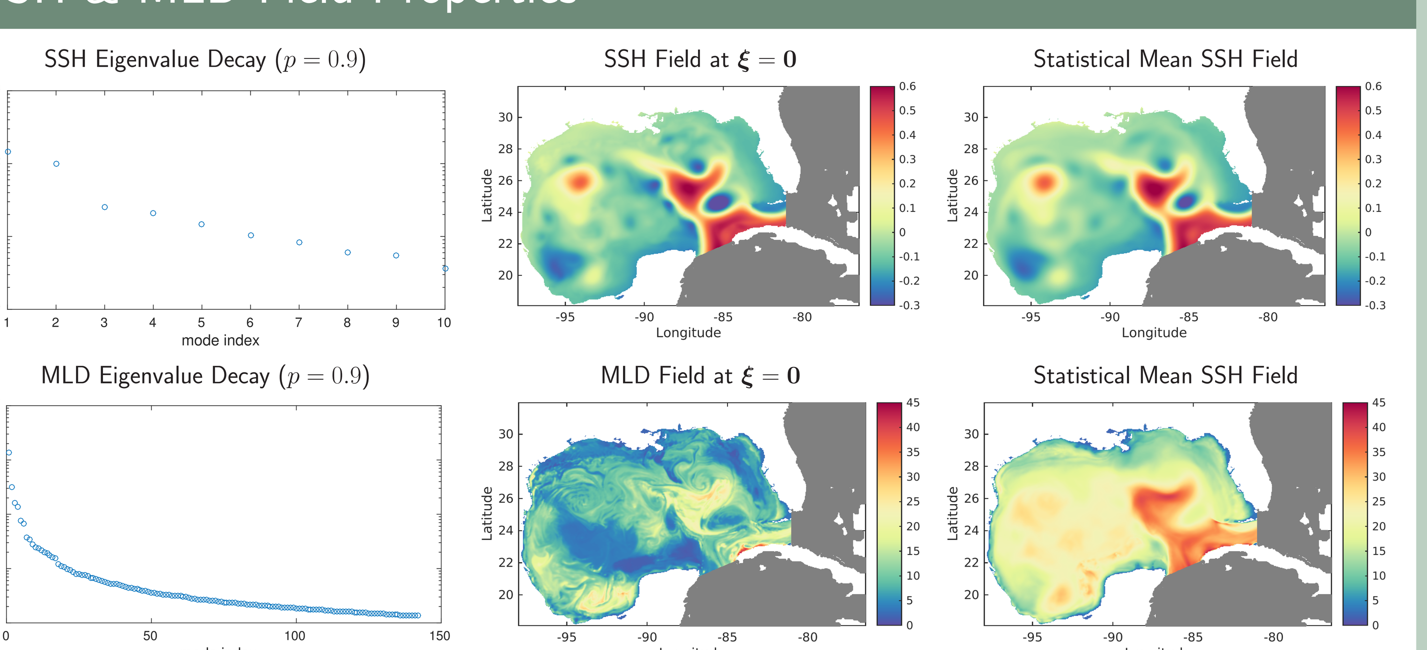

Figure 7: Illustration of different properties in SSH and MLD fields

- For SSH, first 10 EOF modes capture 90% of the field variability. Capturing 90% of MLD variability requires 142 EOF modes. Note in our SSH field reconstruction, 10 EOF modes were included; whereas for MLD fields, we only kept the first 27 modes (corresponds to $p = 0.8$).
- The SSH field at $\boldsymbol{\xi} = \boldsymbol{0}$ shows smooth spatial distribution of SSH, and is very close to the statistical mean field.
- The MLD field at $\boldsymbol{\xi} = \boldsymbol{0}$ exhibits fine spatial structures throughout the GoM. These structures are smoothed out in the mean MLD field.

Field Predictions

- SSH: STD fields obtained from the HYCOM realization ensemble and PC surrogate show almost identical distribution. The largest uncertainty occurs around the LC region.
- MLD: The PC surrogate slightly "underestimates" the empirical prediction from the HYCOM realization ensemble but captures dominant spatial structures of the STD field. Fine spatial structures observed in individual realizations are smoothed out in PC reconstructions.

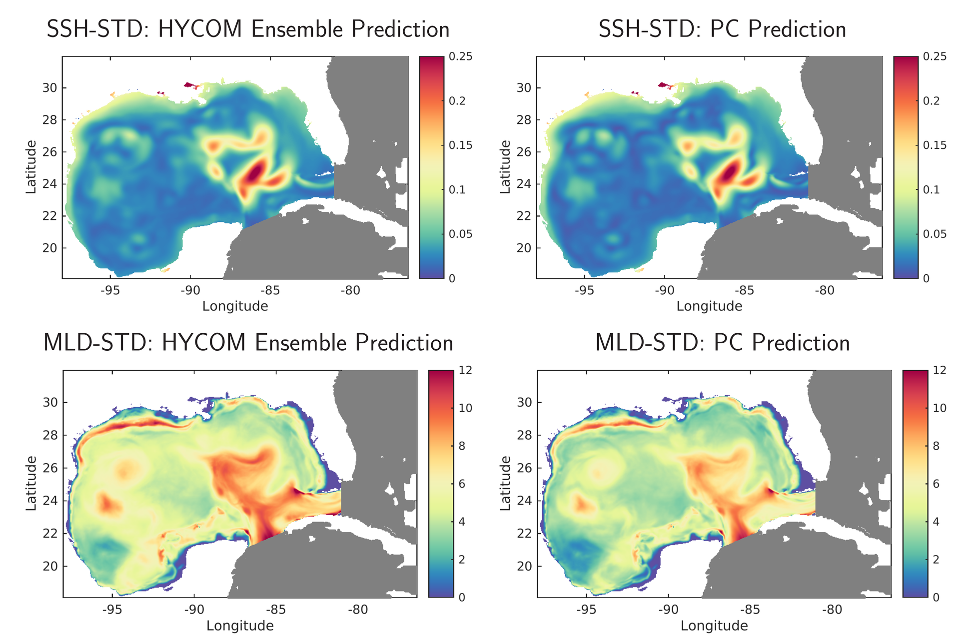

Figure 8: Comparison of standard deviation (STD) fields

Sensitivity Fields

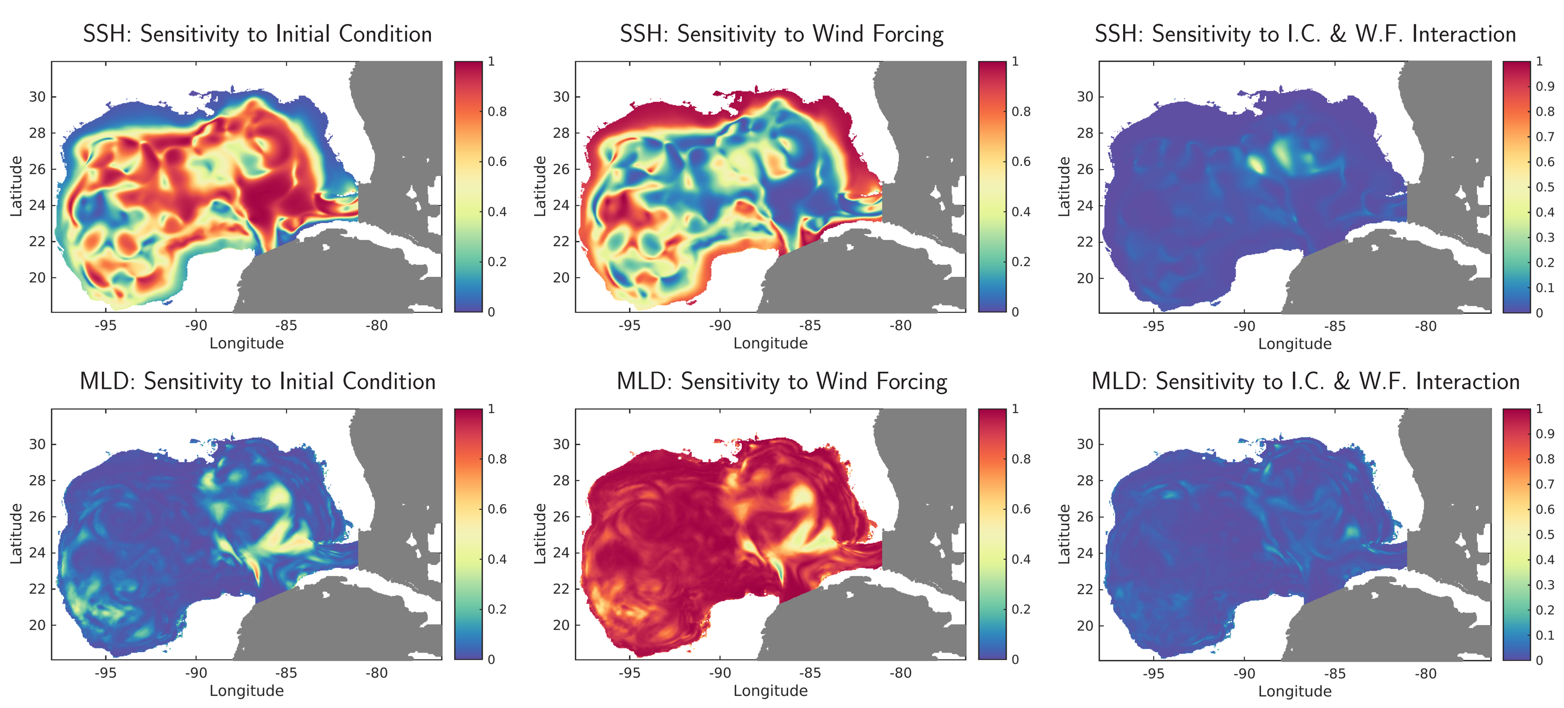

Figure 9: Sensitivity fields to initial condition, wind forcing and their interaction, respectively

- SSH: In the near shore region, SSH is more sensitive to the wind forcing, while in the inner region of the GoM, it is more sensitive to initial condition perturbations.
- MLD: The MLD field is only sensitive to wind forcing perturbations almost everywhere in the GoM, influence of initial condition appears only around the LC.
- Interactions between initial condition and wind forcing perturbations seem to be negligible in both SSH and MLD field response.

Small Sample Field Reconstructions

Motivation: Explores the possibility of utilizing a small number of realizations to reconstruct field statistics.

Sub-sampling: Initial ensemble is denoted as $\mathcal{O}$. We wish to extract a subset $\mathcal{O}_s$ with considerable uniformity. To this end, we start with an independent LHS set, $\mathcal{O}_I$, of the same size as $\mathcal{O}_s$. For each point in $\mathcal{O}_I$, we find its closest neighbor in $\mathcal{O}$ and move to $\mathcal{O}_s$. (e.g Fig. 10)

Results: STD and sensitivity fields given by PC surrogates reconstructed with 50 realizations show excellent agreement with those obtained from the original large ensemble .

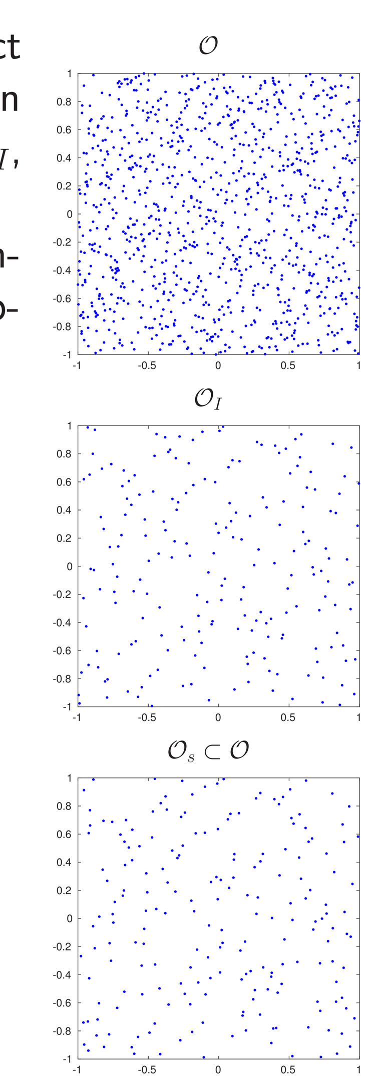

Figure 10: Illustration of the sub-sampling process

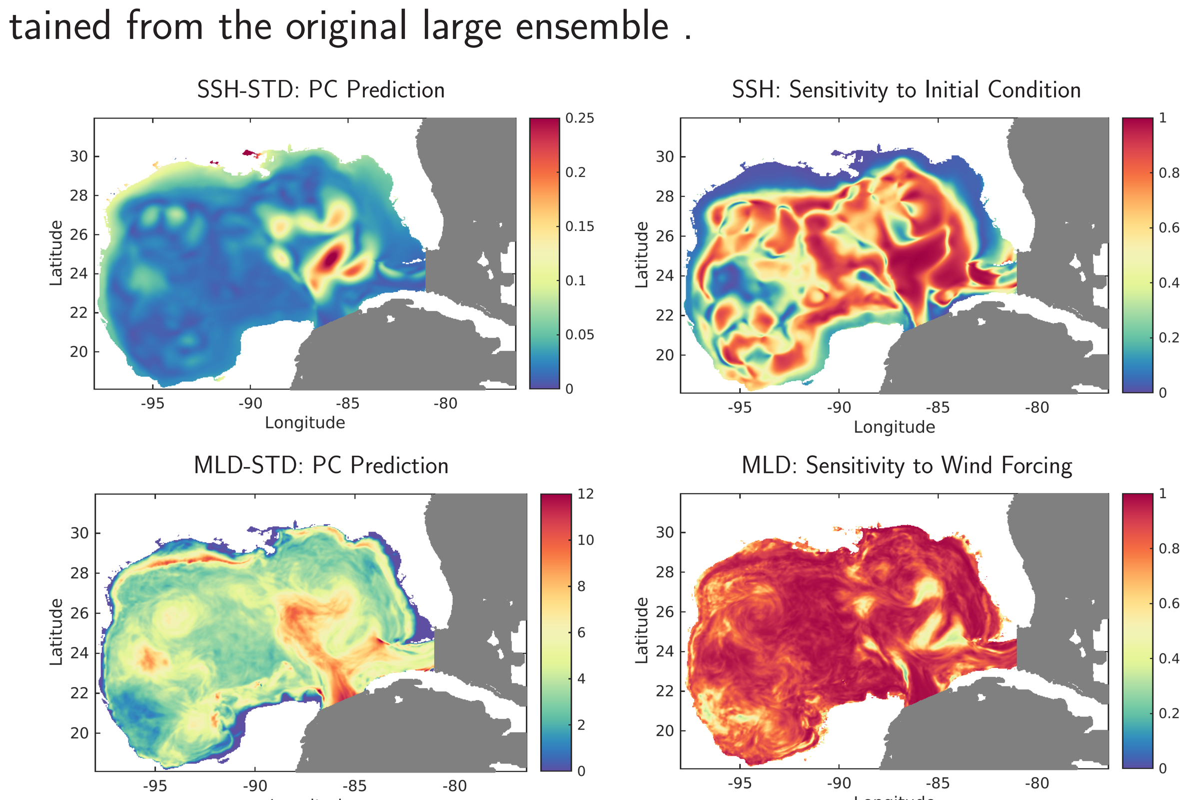

Figure 11: Predictions given by PC-based EOF surrogates reconstructed from 50 HYCOM realizations

Conclusions

- PC surrogates for localized QoIs provide faithful approximations of individual realizations and robust estimates of solution statistics.
- PC-based reconstructions of SSH/MLD fields provide useful insight into field sensitivities.
- Sub-sampling analysis indicates that ensembles of the order of 50 realizations may be sufficient for the purpose of variance and sensitivity predictions.
- Work is underway to extend the present analysis to coupled ocean-wave-atmosphere model.

This work was supported in part by a grant from BP/The Gulf of Mexico Research Initiative, and by the US Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE- SC0008789.