Uncertainty Quantification Using Polynomial Chaos Methods

April 15, 2015
Outline

Uncertainty Quantification (UQ)
What is Polynomial Chaos
Sturm-Liouville Problems
Numerical Quadrature
Example of Advection-Diffusion Equation
Forward Propagation and Analysis
Bayesian Inference
UQ-Initial Conditions
What is UQ

Uncertainty Quantification Tasks:

- **Identification**: What are the uncertainty sources?,
- **Characterization**: aleatoric (intrinsically random) or epistemic (fixed but have unknown values)
  Characterization may be scale dependent
- **Forward Propagation**: Propagate input uncertainty through numerical model to calculate output uncertainty
- **Inverse Propagation**: Use observations/experiments to correct input uncertainties
- **Sensitivity Analysis**: Which uncertainties contribute the most to output uncertainties
- **Reduction**: Improve forecast by assimilating observations

UQ assesses confidence in model predictions and allows resource allocation for fidelity improvements
Uncertainty Sources in Models

- Model equations
- Initial Conditions: Observation sparse in space-time
- Boundary Conditions
  - Momentum, heat and fresh water fluxes
  - Lateral Boundary Conditions in Regional Models
  - Bottom boundary conditions
- Parameterization of small scale processes
  - mixed layer and bottom boundary layer parameters
  - bulk formula for air-sea fluxes

Reliable predictions require a quantitative assessment of dominant sources of error and uncertainty
The Uncertainty Space

- The uncertainty sources generate an **uncertainty space** (sometimes referred to as stochastic space)
- The dimension of the space depends on the number of uncertain inputs, typically very large
- Forward Propagation requires sampling the uncertain space, this can be done using Monte Carlo sampling.
- **Monte Carlo Sampling:**
  - Draw random samples from the uncertain space and construct histograms of Quantities of Interests
  - Typically large number of samples is required for reliable estimates of statistics, this is challenging for computationally expensive models.
  - Nagging question: have enough samples been drawn?
  - Is there another way?
The Model Proxy Approach

- Construct a model **proxy** using a small ensemble
- Use the proxy to compute output statistics
- Proxy must be **efficient** to draw $O(10^4)$–$O(10^7)$ samples
- Proxy must be **reliable** to approximate model output
Sample Problem

- Steady-State Advection-Diffusion

\[
\begin{align*}
    uC_x &= \alpha C_{xx}, \quad 0 \leq x \leq 1 \\
    C(0) &= 0, \quad C(1) = C_m
\end{align*}
\] (1)

- For constant \( u \) and \( \alpha \) the solution is

\[
    C(x) = C_m \frac{e^{\frac{u}{\alpha}x} - 1}{e^{\frac{u}{\alpha}} - 1}
\] (2)

- Actually \( C \) is a function of \( x \) and the problem data:

\[
    C = C(x, u, \alpha, C_m)
\]

- The solution is uncertain if its data is uncertain
**Dependence of $C(x)$ on its data**

**Figure:** Top: $C(x)$ profiles for different values of uncertain parameters. Bottom: Response Surfaces for $C(x = \frac{1}{2}, u, \alpha, C_m = 1)$ as a function of $u$ alone (left), and as a function of $(u, \alpha)$ (middle); and response surface for $C(x = \frac{8}{10}, u_i, \alpha, C_m = 1)$.
Probabilistic Output

- Since the data is uncertain, one would like to estimate mean, variance, PDF of output given the PDF of the input.
- Assume PDFs of inputs are known \( p_u(u), p_{\alpha}(\alpha) \) and \( p_{C_m}(C_m) \) and that the variable are independent, then the joint probability is the product of the individual PDFs.

\[
p_{u,\alpha,C_m} = p_u(u)p_{\alpha}(\alpha)p_{C_m}(C_m)
\]

- The expected value of \( C \) is

\[
\langle C(x) \rangle = \int C(x, u, \alpha, C_m) p_{u,\alpha,C_m}(u, \alpha, C_m) \, du \, d\alpha \, dC_m
\]

- The variance is

\[
\sigma^2 = \int (C(x, u, \alpha, C_m) - \langle C \rangle)^2 p_{u,\alpha,C_m}(u, \alpha, C_m) \, du \, d\alpha \, dC_m
\]
Probabilistic Output

- For the simple example with analytically tractable solution

\[
\langle C(x) \rangle = \int C_m \frac{e^{\frac{u}{\alpha} x} - 1}{e^{\frac{u}{\alpha}} - 1} \rho_{u,\alpha,C_m}(u, \alpha, C_m) \, du \, d\alpha \, dC_m
\]

- Only the integral involving \( C_m \) is tractable analytically since \( C \) depends linearly on \( C_m \) but nonlinearly on \( u \) and \( \alpha \)

- The computations of the mean and variance require numerical computations even when the solution can be written in closed form.

- In most situations the solution is only available after solving PDEs numerically.
Some Notation

- It is useful to regard $u$, $\alpha$ and $C_m$ as random variables.
- It is also useful to work with standardized random variables.
- Map the interval $u_{\text{min}} \leq u \leq u_{\text{max}}$ into $-1 \leq \xi \leq 1$

$$\xi = 2 \frac{u - u_{\text{min}}}{u_{\text{max}} - u_{\text{min}}} - 1$$

- Inverse transformation: $u = (u_{\text{max}} - u_{\text{min}}) \frac{\xi + 1}{2} + u_{\text{min}}$
- The pdf becomes $p_u(u) = p_u(u(\xi)) = \rho(\xi)$
- The other parameters can be treated similarly to obtain the vector of standardized random variables $\xi = (\xi_1, \xi_2, \xi_3)^T$
What is Polynomial Chaos (PC)

PC combines probabilistic and approximation frameworks to express dependency of model outputs on uncertain model inputs

- Series representation:

\[ M(x, t, \xi) \approx M_P \doteq \sum_{k=0}^{P} \tilde{M}_k(x, t) \psi_k(\xi) \] (3)

- \( \xi \): uncertain input characterized by its PDF \( \rho(\xi) \)
- \( M(x, t, \xi) \): model output aka observable
- \( \tilde{M}_k(x, t) \): series coefficients
- \( \psi_k(\xi) \): basis (shape) functions in \( \xi \)-space

- Basic Questions
  - How to choose \( \psi_k \)?
  - How to determine the coefficients \( \tilde{M}_k \)?
  - Where to truncate the series?
Benefit of functional representation

What can you do with a series?

- **Sum** series to interpolate in $\xi$-space
  - series is computationally (much) cheaper than a complex model
  - can sum it millions of time to build histogram or effect Monte Carlo sampling
- **Integrate** in $\xi$-space for statistical moments
  - Mean: $E[M] = \int M \rho(\xi) d\xi = \sum_k \hat{M}_k \int \rho(\xi) \psi_k(\xi) d\xi$
  - Variance: $\text{var}[M] = \int \left( \sum_k \hat{M}_k \psi_k(\xi) - E[M] \right)^2 \rho(\xi) d\xi$
- **Differentiate** in $\xi$-space (no adjoint code!)
  \[
  \frac{\partial M}{\partial \xi} = \sum_k \hat{M}_k \frac{\partial \psi_k}{\partial \xi}
  \]

Series must be reliable to reap benefits
Choice of Polynomial Chaos Basis

- Identify $\rho(\xi)$ with weight function of a singular Sturm-Liouville eigenproblem (see section 3).
- An inner product in $\xi$-space can be defined:
  $$\langle \psi_j, \psi_k \rangle = \int \psi_k(\xi) \psi_j(\xi) \rho(\xi) d\xi$$
- A norm can also be defined: $\| \psi_i \| = \langle \psi_i, \psi_i \rangle$
- The eigenfunctions $\psi_k(\xi)$ form a complete orthogonal basis for the space of square integrable functions $\langle \psi_j, \psi_k \rangle = \delta_{i,j}$
- The series converge spectrally fast for smooth functions.
- The basis choice can then rely on the well-known theory of orthogonal polynomials.
### Choice of Polynomial Chaos Basis

<table>
<thead>
<tr>
<th>$\xi$-distribution</th>
<th>Domain</th>
<th>weight $\rho(\xi)$</th>
<th>basis $\psi_k(\xi)$</th>
<th>parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss</td>
<td>$(-\infty, \infty)$</td>
<td>$\frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}}$</td>
<td>Hermite</td>
<td>none</td>
</tr>
<tr>
<td>Gamma</td>
<td>$(0, \infty)$</td>
<td>$\frac{\xi^\alpha e^{-\xi}}{\Gamma(\alpha+1)}$</td>
<td>Laguerre</td>
<td>$\alpha &gt; 1$</td>
</tr>
<tr>
<td>Beta</td>
<td>$[-1, 1]$</td>
<td>$\frac{(1+\xi)^\alpha (1-\xi)^\beta}{2^{\alpha+\beta+1} B(\alpha+1, \beta+1)}$</td>
<td>Jacobi</td>
<td>$\alpha, \beta &gt; 1$</td>
</tr>
<tr>
<td>Uniform</td>
<td>$[-1, 1]$</td>
<td>$\frac{1}{2}$</td>
<td>Legendre</td>
<td>none</td>
</tr>
</tbody>
</table>

- For many common distributions the basis functions are polynomials.
- A simple scaling can turn the standard weight functions into standard distributions.
- The table lists some common distributions and their corresponding orthogonal polynomials.
- Most commonly used input distribution
- Support on \((-\infty, \infty)\)
• Useful to represent uncertainties in semi-infinite intervals.
• Support on \((0, \infty)\)
- Useful for uncertainties that varies between set quantities.
- Can be tailored to weigh some values more than others
- Support on $[-1, 1]$
• Useful for uncertainties with sharp bounds
• or not much is known about input distribution
• Support on $[-1, 1]$
Statistic with Polynomial Chaos when $w(\xi) = \rho(\xi)$

- Series: $M(x, t, \xi) = \sum_{k=0}^{P} \tilde{M}_k(x, t) \psi_k(\xi)$
- Expectation operator is equivalent to projection on $\psi_0$

$$E[\psi_k] = \int \psi_k(\xi) \rho(\xi) d\xi = \langle \psi_k, \psi_0 \rangle = \delta_{k,0}$$

- mean:

$$E[M] = \sum_{k=0}^{P} \tilde{M}_k(x, t) E[\psi_k(\xi)] = \tilde{M}_0(x, t)$$

- Variance:

$$E\left[ (M - E[M])^2 \right] = \sum_{k=1}^{P} \tilde{M}_k^2(x, t)$$

- Covariance:

$$E[ (u - E[u]) (v - E[v]) ] = \sum_{k=1}^{P} u_k(x) v_k(x, t)$$
How do we determine PC coefficients?

- **Series**: \( M(x, t, \xi) = \sum_{k=0}^{P} \hat{M}_k(x, t)\psi_k(\xi) \)

- **Galerkin Projection** on \( \psi_k \) basis (minimizes \( L_2 \)-error norm)

\[
\hat{M}_k(x, t) = \langle M, \psi_k \rangle = \int M(x, t, \xi)\psi_k(\xi)\rho(\xi)d\xi
\]

- **Non Intrusive Spectral Projection**: Approximate integral numerically via **quadrature**

\[
\hat{M}_k(x, t) \approx \sum_{q=1}^{Q} M(x, t, \xi_q)\psi_k(\xi_q)\omega_q
\]

- \( \xi_q/\omega_q \) quadrature points/weights
- Quadrature requires an ensemble run at \( \xi_q \).
- No code modification is necessary
Choice of Quadrature

- Gauss quadrature most accurate/point ($\psi_{p+1}(\xi_q) = 0$) but Naive tensorization cost grows exponentially: $p^N$.
- Rely on Nested Sparse Smolyak Quadrature Tempers the curse of dimensionality
- Adaptive Quadrature
Multidimensional basis

Multi-dimensional basis functions $\psi_k(\xi_1, \xi_2, \ldots, \xi_n)$ are tensor products of 1D basis functions:

$$
\psi_k(\xi_1, \xi_2, \ldots, \xi_n) = \psi_{\alpha_1}^k(\xi_1)\psi_{\alpha_2}^k(\xi_2) \ldots \psi_{\alpha_n}^k(\xi_n)
$$

- 1D Legendre basis: $L_0(\xi) = 1$, $L_1(\xi) = \xi$, $L_2(\xi) = \frac{3\xi^2-1}{2}$
- 2D Example
  - $\psi_0 = L_0(\xi_1)L_0(\xi_2)$
  - $\psi_2 = L_0(\xi_1)L_1(\xi_2)$
  - $\psi_4 = L_1(\xi_1)L_1(\xi_2)$
  - $\psi_5 = L_0(\xi_1)L_2(\xi_2)$
  - $\psi_6 = L_2(\xi_1)\psi_0(\xi_2)$
  - $\psi_7 = L_2(\xi_1)L_1(\xi_2)$
  - $\psi_8 = L_1(\xi_1)L_2(\xi_2)$
  - $\psi_9 = L_0(\xi_1)L_3(\xi_2)$
- Triangular truncation is common, max order=3
- number of coefficient is $P + 1 = \frac{(N+p)!}{N!p!}$
  - $N$ is the number of stochastic variables
  - $p$ is the max polynomial degree in 1D
Polynomial Chaos Expansions Summary

- Paradigm shift from statistical to combined probabilistic/approximation view
- Can quantify approximation error and “convergence” to solution
- No a-priori restriction/assumption on output statistics
- Approach robust to model non-linearity and model differentiability
- Can be done non-intrusively via ensembles.
- Multiple independent stochastic variables can be handled by multi-dimensional tensorization of 1D basis functions and quadratures.
- Sampling Challenges for high $N$ or $p$
Basis functions are eigenfunctions of a singular Sturm-Liouville problem

The Sturm-Liouville operator $\mathbf{L}$ is

$$
\mathbf{L} = \frac{1}{w} \left[ -\frac{d}{d\xi} \left( p \frac{d}{d\xi} \right) + q \right]
$$

(4)

- Sturm-Liouville eigenvalue problems are of the form: $\mathbf{L}\psi = \lambda\psi$ + appropriate BCs
- $\psi = \psi(\xi)$ and $\lambda$ are eigenvectors/eigenvalues pairs
- $p, q$ and $w$ are given real-valued functions
- $p > 0$ is continuously differentiable in $(-1,1)$ and continuous at $x = \pm 1$. The problem is called singular when $p(\pm 1) = 0$.
- $q \geq 0$ is continuous, and bounded in $(-1,1)$
- $w \geq 0$ is continuous and integrable over $(-1,1)$. It is known as the weight function.
Remarks on the weight function

- Since $w(\xi) \geq 0$ it can be used to define an inner product between 2 functions $f$ and $g$

$$ (f, g) = \int_{-1}^{1} f(\xi) g(\xi) w(\xi) d\xi $$

(5)

- It is also useful in defining a norm of a function:

$$ \|f\|^2 = (f, f) $$

(6)

- $L^2$ is the vector space of square integrable functions

$$ \|f\|^2 = \int_{-1}^{1} w(\xi) f^2 \, d\xi < \infty $$

(7)
Properties of Sturm-Liouville eigenproblems

- Eigenvectors of \( L \) form a complete basis for \( \mathcal{L}^2 \):
  - Any function \( M(\xi) \in \mathcal{L}^2 \) can be written as
    \[
    M_P = \sum_{k=0}^{P} \hat{M}_k \psi_k(\xi)
    \]
  - The series converges in \( |||\)-norm
    \[
    \lim_{P \to \infty} ||M - M_P|| \longrightarrow 0
    \]
- Operator \( L \) is self adjoint: \( (Lu, v) = (u, Lv) \)
- Operator \( L \) is positive definite: \( (Lu, u) > 0 \)
- Eigenfunctions \( \psi_j(\xi) \) are orthogonal: \( (\psi_i, \psi_j) = \delta_{ij} ||\psi_i||^2 \)
- \( \infty \) of real positive eigenvalues
- Spectral series converges spectrally fast for \( \infty \)-differentiable functions
\( \psi_n(\xi) \) are Legendre Polynomials

- \( p = (1 - \xi^2), \ q = 0 \) and \( w = 1 \).
- SL PDE: \( \frac{d}{d\psi} 
\left[(1 - \psi^2) \frac{d\psi}{d\xi}\right] + \lambda \psi = 0 \)
- Eigenvalues \( \lambda_n = n(n + 1), \ n = 0, 1, \ldots \)
- Eigenvectors are Legendre Polynomials \( \psi_n = L_n(x) \)
- Inner product \( (u, v) = \int_{-1}^{1} u(\xi) \, v(\xi) d\xi \)
- Eigenvector norms \( \|L_n\|_2 = (L_n, L_n) = \left(n + \frac{1}{2}\right)^{-1} \)
- Expansion \( M(\xi) = \sum_{n=0}^{P} \hat{M}_n L_n(x) \), and

\[
\hat{M}_n = \frac{(M, L_n)}{(L_n, L_n)} = \left(n + \frac{1}{2}\right) \int_{-1}^{1} u(\xi) L_n(\xi) d\xi
\]
Some properties of Legendre Polynomials

- $L_0(x) = 1$, $L_1(x) = x$, $L_2(x) = \frac{3x^2-1}{2}$
- Recurrence $L_{n+1}(x) = \frac{2n+1}{n+1}x L_n(x) - \frac{n}{n+1} L_{n-1}(x)$
- $|L_n(x)| \leq 1$ for $|x| \leq 1$ and $L_n(\pm 1) = (\pm 1)^n$.

Figure: The first 6 Legendre polynomials.
Approximation of integrals by quadrature

- Many definite integrals cannot be computed in closed form, and must be approximated numerically.
- Basic building block

\[ \int_{-1}^{1} u(\xi) w(\xi) d\xi \approx \sum_{i=0}^{N} u(\xi_i) \omega_i \]  

- \( u(\xi) \) is any function
- \( w(\xi) > 0 \) is a weight function
- \( \xi_i \) are the quadrature roots
- \( \omega_i \) are the quadrature weights

- Integration can be done with function evaluations at \( \xi_i \)
- Pick \( \xi_i \) and \( \omega_i \) to minimize errors and maximize efficiency
Gauss quadrature

- Choose $\xi_i$ and $\omega_i$ to maximize order of exact integration.
- Let $p_m(x)$ be the set of orthogonal polynomials on the interval $a \leq x \leq b$ w.r.t. weight function $w(x)$ and of degree $m$

$$\int_{a}^{b} p_m(\xi)p_n(\xi)w(\xi)d\xi = \delta_{m,n}\|p_m\|^2 \quad (9)$$

Let the quadrature points be the roots of $p_{N+1}(\xi_i) = 0$ and the weights:

$$\sum_{i=0}^{N} \xi_i^k \omega_i = \int_{a}^{b} \xi^k w(\xi)d\xi, \quad 0 \leq k \leq N \quad (10)$$

then...
Gauss quadrature

• The weights are all positive $\omega_i > 0$.
• Gauss quadrature is exact for all polynomials, $q$, of degree less or equal to $2N + 1$

$$\sum_{i=0}^{N} q(\xi_i)\omega_k = \int_{a}^{b} q(\xi)w(\xi)d\xi$$  \hspace{1cm} (11)

• It is not possible to find a $\xi_i, \omega_i$ combination where the integration is exact for polynomials of degree $2N + 2$. 
Example: Legendre Gauss Quadrature

- Roots must be computed numerically \( L_{N+1}(\xi_i) = 0 \)
- Quadrature weights are

\[
\omega_i = \frac{2}{(1 - \xi_i)^2 L'_{N+1}(\xi_i)}
\]  

(12)

<table>
<thead>
<tr>
<th>( Q = N + 1 )</th>
<th>( \xi_i^q )</th>
<th>( \omega_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>( \pm \sqrt{\frac{1}{3}} )</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>( \frac{8}{9} )</td>
</tr>
<tr>
<td></td>
<td>( \pm \sqrt{\frac{3}{5}} )</td>
<td>5/9</td>
</tr>
<tr>
<td>4</td>
<td>( \pm \sqrt{\frac{3}{7} - \frac{2}{7} \sqrt{\frac{6}{5}}} )</td>
<td>( \frac{18 + \sqrt{30}}{36} )</td>
</tr>
<tr>
<td></td>
<td>( \pm \sqrt{\frac{3}{7} + \frac{2}{7} \sqrt{\frac{6}{5}}} )</td>
<td>( \frac{18 - \sqrt{30}}{36} )</td>
</tr>
</tbody>
</table>
Figure: Gauss–Legendre quadrature points and weights (shown in red) for various number of quadrature order. The quadrature roots clustering near the edges prevent spurious polynomial wiggles known as Runge oscillations. The weights associated with the edges decrease with increasing order.
Fix $C_m = 1$ and let $u$ and $\alpha$ be uncertain with $0.2 \leq u \leq 0.8$ and $0.05 \leq \alpha \leq 0.45$

Their input uncertainties is assumed uniform:

$$p_u(u) = \frac{1}{0.8-0.2} \cdot p_\alpha(\alpha) = \frac{1}{0.45-0.05}.$$

The weight functions are $\rho(\xi_1) = \frac{1}{2}$ and $\rho(\xi_2) = \frac{1}{2}$

The corresponding orthogonal polynomials are those of Legendre $\psi_k(\xi_1) = L_k(\xi_1) \psi_k(\xi_2) = L_k(\xi_2)$

Use polynomials of degree 6 so that $\psi_k = L_{k_1}(\xi_1)L_{k_2}(\xi_2)$ with $k_1 + k_2 \leq 6$

The basis has 28 members and 28 unknowns.

Number of 1D quadrature points is $Q = 6 + 1 = 7$ and its tensorization requires samples at $(\xi_{1,j}, \xi_{2,j})$ for a total of $7^2$ realizations
The series coefficients at $x = \frac{1}{2}$

**Figure:** Left: magnitude of the spectral coefficients as a function of polynomial degree $(m, n)$ in the $(\xi_1, \xi_2)$ directions. Right: Same but presented as 1D curves. The corresponding polynomial basis is $\psi(\xi_1, \xi_2) = L_m(\xi_1)L_n(\xi_2)$. 
Response Curves

Validation requires testing the PC approximation against an independent set of realizations.

Figure: Response curves $C(x = \frac{1}{2}, u = 0.5, \alpha, 1)$ (left) and $C(x = \frac{1}{2}, u, \alpha = 0.25, 1)$ (right) produced by the validation run (black), by the PC proxy (dashed red curve overlaps the validation curve). The red crosses show the quadrature points used to build the PC proxy.
**Response Surfaces**

**Figure:** Response Surface comparisons of \( C(x = 1/2, u, \alpha, 1) \) of the validation run (left), of the PC proxy (middle) and the difference between the two (right). The largest PC errors appear for low values of \( \alpha \).
Figure: Response Surface comparisons of $C(x = 1/2, u, \alpha, 1)$ of the validation run (left), of the PC proxy (middle) and the difference between the two (right). The largest PC errors appear for low values of $\alpha$. 
Figure: Location of $10^5$ random samples (left) in parameter space, and histogram (right) of $C$ at $x = \frac{1}{2}$. The red line is a KDE estimate of the output PDF. The empirical mean is $\overline{C} = 0.25$ and the standard deviation is 0.12. The output PDF is not Gaussian and exhibits two peaks.
Figure: Location of $10^5$ normally-distributed random samples (left) in parameter space, and histogram (right) of $C$ at $x = \frac{1}{2}$. The surrogate was resampled after assuming a Gaussian distribution for input parameters, now new model runs were needed. The new output PDF looks more Gaussian.
### Forward Problem: Parametric Sensitivity

#### Hurricane Ivan track

![Hurricane Ivan Track](image)

#### Comparison of mean & observed SST

- Legendre basis with $p = 5$
- 210 unknown coefficients
- Nested sparse Smolyak Ensemble size 385 ($6^4 = 1,296$ Gauss quadrature)

#### Table: HYCOM uncertain inputs

<table>
<thead>
<tr>
<th>Description</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>critical Richardson #</td>
<td>$p_1 \in [0.25, 0.7]$</td>
</tr>
<tr>
<td>background viscosity</td>
<td>$p_2 \in [10^{-4}, 10^{-3}]$</td>
</tr>
<tr>
<td>background diffusivity</td>
<td>$p_3 \in [10^{-5}, 10^{-4}]$</td>
</tr>
<tr>
<td>drag coefficient factor</td>
<td>$p_4 \in [0.2, 1.0]$</td>
</tr>
</tbody>
</table>

- Comparing mean & observed SST. Vertical lines show when Ivan enters GoM and when it is nearest buoy.
Variance Analysis

\[ T_i = \frac{\text{Variance due to parameter } p_i}{\text{Total variance}} \]

Figure: Evolution of the global sensitivity indices \( T_1, \ldots, T_4 \) for SST and MLD (bottom). The first vertical line indicates the time the hurricane enters the GOM whereas the second indicates a time at which the hurricane is close to the buoy.
Figure: $T_3$ (left) and $T_4$ (right) sensitivity contours for SST. Drag dominates uncertainty during high winds, otherwise it is background diffusivity.
\[ \bar{\tau} = \rho a C_D V \tilde{V} \]

\[ C_D = C_{D0} + C_{D1} (T_s - T_a) \]

\[ C_{D0} = a_0 + a_1 \tilde{V} + a_2 \tilde{V}^2 \]

\[ C_{D1} = b_0 + b_1 \tilde{V} + b_2 \tilde{V}^2 \]

\[ \tilde{V} = \max \, [ V_{\min}, \min (V_{\max}, V) ] \]

\( C_D \) is drag coefficient
\( V \) is wind speed at 10 m.
\( C_D \) saturates for \( V > V_{\max} \)

- Blue circles: aircraft observations
- red: wind tunnel
- green: drop sondes
- magenta: HYCOM fit to COARE 2.5,
- Problem: \( V_{\max} \) and \( C_D^{\max} \) are not well-known and does \( C_D \) decrease for \( V > V_{\max} \) as drop sondes suggest?
Inverse Modeling Problem

• Perturb $C_D$ by introducing 3 control variables $(\alpha, V_{\text{max}}, m)$

$$C_D' = \alpha C_D \text{ for } V < V_{\text{max}}$$

$$C_D' = \alpha [C_D + m(V - V_{\text{max}})] \text{ for } V > V_{\text{max}}$$

• multiplicative factor $0.4 \leq \alpha \leq 1.1$
• vary $V_{\text{max}}$ between 20 and 35 m/s
• $m$ is a linear slope modeling decrease for $V > V_{\text{max}}$ with $-3.8 \times 10^{-5} \leq m \leq 0$
• Use ITOP data to learn about likely distribution of $\alpha$, $V_{\text{max}}$ and $m$. 

Bayes Theorem: \( p(\theta \mid T) \propto p(T \mid \theta) \, p(\theta) \)

- Likelihood: \( \epsilon = T - M \) is normally distributed

\[
p(T \mid \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(T_i - M_i)^2}{2\sigma^2} \right)
\]  
(15)
Bayes Theorem: $p(\theta \mid T) \propto p(T \mid \theta) p(\theta)$

- Likelihood: $\epsilon = T - M$ is normally distributed

\[
p(T \mid \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{(T_i - M_i)^2}{2\sigma^2} \right) \tag{15}
\]

- $\sigma^2$ unknown, treated as hyper-parameter. Assume a Jeffreys prior

\[
p(\sigma^2) = \begin{cases} \frac{1}{\sigma^2} & \text{for } \sigma^2 > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{16}
\]
Bayes Theorem: $p(\theta \mid T) \propto p(T \mid \theta) \ p(\theta)$

- Likelihood: $\epsilon = T - M$ is normally distributed

$$p(T \mid \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(T_i - M_i)^2}{2\sigma^2}\right)$$  \hspace{1cm} (15)

- $\sigma^2$ unknown, treated as hyper-parameter. Assume a Jeffreys prior

$$p(\sigma^2) = \begin{cases} \frac{1}{\sigma^2} & \text{for } \sigma^2 > 0, \\ 0 & \text{otherwise.} \end{cases}$$  \hspace{1cm} (16)

- Uninformed priors for $\alpha$, $V_{\text{max}}$ and $m$:

$$p(\{\alpha, V_{\text{max}}, m\}) = \begin{cases} \frac{1}{b_i-a_i} & \text{for } a_i \leq \{\alpha, V_{\text{max}}, m\} \leq b_i, \\ 0 & \text{otherwise,} \end{cases}$$  \hspace{1cm} (17)

where $[a_i, b_i]$ denote the parameter ranges.
Final Form of Bayes theorem

\[ p(\{\alpha, V_{\text{max}}, m\}, \sigma^2 | T) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(T_i - M_i)^2}{2\sigma^2}\right) \]

\[ p(\sigma^2) p(\alpha) p(V_{\text{max}}) p(m) \]
Final Form of Bayes theorem

\[ p(\{\alpha, V_{\text{max}}, m\}, \sigma^2 | T) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(T_i - M_i)^2}{2\sigma^2}\right) \]

\[ p(\sigma^2) p(\alpha) p(V_{\text{max}}) p(m) \]

- Build full posterior with Markov Chain Monte Carlo (MCMC)
  MCMC requires \( O(10^5) \) estimates of \( M_i \): prohibitive
Final Form of Bayes theorem

\[
p(\{\alpha, V_{\text{max}}, m\}, \sigma^2 | T) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(T_i - M_i)^2}{2\sigma^2}\right) \]

\[
p(\sigma^2) p(\alpha) p(V_{\text{max}}) p(m)
\]

- Build full posterior with Markov Chain Monte Carlo (MCMC)
  MCMC requires \(O(10^5)\) estimates of \(M_i\): prohibitive
- Solve for center and spread of posterior
  minimization problem requiring access to cost function
  gradient and Hessian: Needs an adjoint model
Final Form of Bayes theorem

\[ p(\{\alpha, V_{\text{max}}, m\}, \sigma^2 | T) \propto \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( \frac{-(T_i - M_i)^2}{2\sigma^2} \right) \]

\[ p(\sigma^2) p(\alpha) p(V_{\text{max}}) p(m) \]

- Build full posterior with Markov Chain Monte Carlo (MCMC)
  MCMC requires \( O(10^5) \) estimates of \( M_i \): prohibitive

- Solve for center and spread of posterior
  minimization problem requiring access to cost function
  gradient and Hessian: Needs an adjoint model

- Rely on Polynomial Chaos expansions to replace HYCOM
  by a polynomial series that could be either summed for
  MCMC or differentiated for the gradients.
Figure: Fanapi’s JTWC track (black curve) and paths of C-130 flights. The yellow circles on the track represent the typhoon center at 00:00 UTC. The circles on the flight paths mark the 119 AXBT drops. The $42 \times 42$ km$^2$ analysis box is also shown.
Figure: Comparison of HYCOM vertical temperature profiles with AXBT observations on Sep 14 (left), 15 (center) and 17 (right). Temperature averages over the first 50 m are shown in the legend.
Evolution of the area-averaged SST realizations (blue) and of the corresponding PC estimates (red). The normalized rms error (right panel) remains below 0.1% for the duration of the simulation.
Figure: Normalized error between realizations and the corresponding PC surrogates at different depths; Top row: 00:00 UTC Sep 15; bottom row: 00:00 UTC Sep 18.
50m-deep mixed layer
2°C cooling after Fanapi arrives
Uncertainties confined to top 50 m.
Figure: SST response surface as function of $\alpha$ and $V_{max}$, with fixed $m = 0$. Plots are generated on different days, as indicated. SST’s dependence on $V_{max}$ decreases after 09/17.
Figure: Top row: chain samples for $V_{max}$, $m$ and $\alpha$. Bottom row: chain samples for $\sigma^2$ generated for different days, as indicated.
Figure: Posterior distributions for the drag parameters (top) and the variance between simulations and observations (bottom). The numbers show the Kullback-Liebler divergence quantifying the distance between 2 prior and posterior pdfs, i.e. the information gain.
Remarks on posteriors

- $V_{\text{max}}$ exhibits a well-defined peak at 34 m/s.
- Posterior of $m$ resembles prior. Data added little to our knowledge of $m$.
- $\alpha$ shows a definite peak at 1.03 with a Gaussian like-distribution.
- $\sqrt{\sigma^2}$ is a measure of the temperature error expected. This error grows with time from about 0.75$^\circ$C to 1$^\circ$C.
Joint posterior PDFs

Figure: Left: joint posterior distribution of $\alpha$ (left) and $V_{\text{max}}$; right: joint posterior of $\alpha$ and $\sigma^2$, generated for Sep 17-Sep 18. Single peak located at $V_{\text{max}} = 34$ m/s and $\alpha = 1.03$. The posterior shows a tight estimate for $\alpha$ with little spread around it.
Figure: Optimal wind drag coefficient $C_D$ using MAP estimate of the three drag parameters. The symbols refer to AXBT data used in the Bayesian inference.
Variational Form

- maximize the posterior density, or equivalently, minimize the negative of its logarithm

\[ J(\alpha, V_{\text{max}}, m, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = \sum_{d=1}^{5} \left[ J_d + \left( \frac{n_d}{2} + 1 \right) \ln(\sigma_d^2) \right], \]  

where \( J_d \) is the misfit cost for day \( d \), the \( \ln(\sigma_d^2) \) terms come from the normalization factors of the Gaussian likelihood functions and from the Jeffreys priors.

- The expression for \( J_d \) is:

\[ J_d(\alpha, V_{\text{max}}, m, \sigma_d^2) = \frac{1}{2\sigma_d^2} \sum_{i \in \mathcal{I}_d} [M_i - T_i]^2, \]  

where \( \mathcal{I}_d \) is the set of \( n_d \) indices of the observations from day \( d \).
Adjoint-Free Gradients

Minimization requires cost function gradients

\[
\begin{bmatrix}
\frac{\partial J}{\partial \alpha}, \frac{\partial J}{\partial V_{\text{max}}}, \frac{\partial J}{\partial m}
\end{bmatrix}
= \sum_{d=1}^{5} \frac{1}{\sigma_d^2} \left( \sum_{i \in I_d} (M_i - T_i) \begin{bmatrix}
\frac{\partial M_i}{\partial \alpha}, \frac{\partial M_i}{\partial V_{\text{max}}}, \frac{\partial M_i}{\partial m}
\end{bmatrix} \right)
\]

Compute them from PC expansion

\[
\begin{bmatrix}
\frac{\partial M}{\partial \alpha}, \frac{\partial M}{\partial V_{\text{max}}}, \frac{\partial M}{\partial m}
\end{bmatrix}
= \sum_{k=0}^{P} \hat{M}_k(x, t) \begin{bmatrix}
\frac{\partial \psi_k}{\partial \alpha}, \frac{\partial \psi_k}{\partial V_{\text{max}}}, \frac{\partial \psi_k}{\partial m}
\end{bmatrix}.
\]

- \( \frac{\partial \psi_k}{\partial \alpha} \) easy to compute
- No adjoint model needed
- For Hessian just differentiate above again.
Figure: Posterior probability distributions for (top) drag parameters and (bottom) variances $\sigma_d^2$ at selected days using variational method and MCMC. The vertical lines correspond to the MAP values from MCMC and optimal parameters found using the variational method.
Rely on EOFs to characterize uncertainty and reduce the number of stochastic variables. For 2 EOFs mode we have:

\[ u(\tilde{x}, 0, \xi_1, \xi_2) = \bar{u}(\tilde{x}, 0) + \alpha \left[ \sqrt{\lambda_1} U_1 \xi_1 + \sqrt{\lambda_2} U_2 \xi_2 \right] \]  

(20)

- \((\lambda_k, U_k)\): are eigenvalues/eigenvectors of covariance matrix obtained from free-run simulation
- \(\bar{u}\): unperturbed initial condition
- \(u(\tilde{x}, 0, \xi)\): Stochastic initial condition input
- \(\alpha\): multiplicative factor to control size of “kick”
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
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 Hermite 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.
T-section along 25N, stddev grows in time with maxima coinciding with Frontal Eddy during days 20–40.
PC-error: $\|\epsilon\|_2^2 = \sum_q [T(\tilde{x}, t, \xi_q) - T_{PC}(\tilde{x}, t, \xi_q)]^2 \omega_q$

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

On day 60 PC-error is about 50% of stddev
Distribution of SSH PC coefficients
Figure: Temperature (left) and Salt (right) profiles for extreme realizations at DWH
Examples of 2D tensorizations

Full-Tensor (49)

Classic Smolyak (17)

Arbitrary Multi-Index (33)
Varying polynomial order

Figure: Relative L2 error between the area-averaged SST and the Latin Hypercube Samples.

<table>
<thead>
<tr>
<th>Simple Truncation</th>
<th>P</th>
<th># of realizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = (5, 5, 5, 5)$</td>
<td>126</td>
<td>385</td>
</tr>
<tr>
<td>$p = (5, 5, 7, 7)$</td>
<td>168</td>
<td>513</td>
</tr>
<tr>
<td>$p = (2, 2, 5, 5)$</td>
<td>36</td>
<td>73</td>
</tr>
<tr>
<td>$p = (2, 2, 7, 7)$</td>
<td>59</td>
<td>169</td>
</tr>
</tbody>
</table>
Smolyak Projections

- Apply Smolyak’s algorithm directly to construct the PCE instead of purely generating the quadrature. Thus, the final projection becomes a weighted sum of aliasing-free sub-projections. This is an extension of the Smolyak tensor construction from quadrature operators to projection operators.

- Smolyak projection allows a refinement approach based on successive inclusion of any admissible multi-index, $\mathcal{F}$, of quadrature rules while maintaining the representation free of internal aliasing.

- A larger number of polynomials can be integrated than is possible with a classical dimensional truncation / quadrature using the same ensemble. The 513 HYCOM realizations yields 402 coefficient with Smolyak projection compared to 168 using Smolyak quadrature.
Adaptive Projections

- Rewrite projection as tensor products of projection differences:
  \[(\Delta_{k_1} \otimes ... \otimes \Delta_{k_d}) \, U,\]

- The \(L_2\) norm of this difference can be readily used to define an error indicator for multi-index \(k\),
  \[\epsilon(k) = \| (\Delta_{k_1} \otimes ... \otimes \Delta_{k_d}) \, U \|\]

  The indicator represents the variance surplus due to the \(k\) sub-projection.

- The surplus is computed for each \(k \in \mathcal{F}\) and the sub-projection with the highest \(\epsilon(k)\) is selected for subsequent refinement, which generally consists of inclusion of admissible forward neighbors.
Figure: Relative L2 difference between the PCE of the averaged SST and the LHS sample. Plotted are curves generated with (i) the adaptive Smolyak projection adapted at $t = 60$ hr, (ii) the Smolyak projection with the full database, and (iii) Smolyak classical quadrature using the full database. For the adapted solution, the refinement is stopped after iteration 5, leading to 69 realizations and a PCE with 59 polynomials. The full 513 database curves have 402 polynomials for the pseudo-spectral construction and 168 polynomials for the Smolyak quadrature.
Publications


