

# Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach

Alen Alexanderian · Justin Winokur · Ihab Sraj ·  
Ashwanth Srinivasan · Mohamed Iskandarani ·  
William C. Thacker · Omar M. Knio

Received: 1 October 2011 / Accepted: 2 February 2012 / Published online: 4 March 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** Polynomial chaos (PC) expansions are used to propagate parametric uncertainties in ocean global circulation model. The computations focus on short-time, high-resolution simulations of the Gulf of Mexico, using the hybrid coordinate ocean model, with wind stresses corresponding to hurricane Ivan. A sparse quadrature approach is used to determine the PC coefficients which provides a detailed representation of the stochastic model response. The quality of the PC representation is first examined through a systematic refinement of the number of resolution levels. The PC representation of the stochastic model response is then utilized to compute distributions of quantities of interest (QoIs) and to analyze the local and global sensitivity of these QoIs to

uncertain parameters. Conclusions are finally drawn regarding limitations of local perturbations and variance-based assessment and concerning potential application of the present methodology to inverse problems and to uncertainty management.

**Keywords** Ocean circulation model · Parametric uncertainty · Sensitivity analysis · Polynomial chaos · Sparse quadrature

**Mathematics Subject Classifications (2010)** 86A05 · 65D32 · 60-08 · 60H35

---

A. Alexanderian · J. Winokur · I. Sraj · O. M. Knio  
Department of Mechanical Engineering, Johns Hopkins  
University, Baltimore, MD 21218, USA

A. Srinivasan · M. Iskandarani  
Rosenstiel School of Marine and Atmospheric Science,  
University of Miami, Miami, FL 33149, USA

A. Srinivasan  
Center for Computational Science, University of Miami,  
Miami, FL 33149, USA

W. C. Thacker  
Cooperative Institute for Marine and Atmospheric Studies,  
University of Miami, Miami, FL 33149, USA

W. C. Thacker  
Atlantic Oceanographic and Meteorological Laboratory,  
Miami, FL 33149, USA

O. M. Knio (✉)  
Department of Mechanical Engineering and Materials  
Science, Duke University, Durham, NC 27708, USA  
e-mail: omar.knio@duke.edu

## 1 Introduction

Deterministic ocean models such as hybrid coordinate ocean model (HYCOM) [3, 5, 50], regional oceanic modeling system [13, 40], and MIT general circulation model [27] have been instrumental in ocean weather forecast. Substantial efforts have been made in the past few decades to improve those models for the purpose of achieving high resolution, better physical representation, and relatively fast and robust computations. In this paper, we extend the traditional deterministic ocean modeling, namely HYCOM, by considering uncertainty in the model parameters. Our goal is to provide a detailed statistical characterization of the model variables in space and time, namely one that enables us to conduct a global sensitivity analysis of the impact of uncertain parameters. The quantification of the uncertainties in the boundary conditions is another important aspect of this problem which is addressed in the recent work [45].

We have been motivated to adopt a polynomial chaos (PC) formalism, which appears to be ideally suited for the present purpose. PC methods [7, 12, 22–25, 38, 48, 49] have become increasingly popular in the past two decades and have been extensively used to propagate and quantify uncertainties in various physical problems, including fluid [23, 25, 30] and solid [11, 12] mechanics, as well as chemical systems [24, 31, 38]. These methods provide an approximation of the model variables in terms of a spectral expansion in an orthogonal polynomial basis of an underlying probability space. Once available, the PC expansion can be used to efficiently approximate the statistical properties of the model variables such as their distribution, moments, sensitivities, etc.

There are two major approaches for computing the coefficients in a PC expansion: (1) the *intrusive* method and (2) the *non-intrusive* methods. The intrusive method requires a reformulation of the original random dynamical system (and hence the deterministic solvers), through Galerkin projection onto the PC basis [12, 22]. Subsequently, one has to solve a larger system for the time/space evolution of the PC coefficients. Non-intrusive methods, on the other hand, provide a means to determine the requisite spectral representation through direct application of existing deterministic solvers.

There are different types of non-intrusive methods, including the non-intrusive spectral projection (NISP), the collocation method (CM), and the regression-like approaches. In the NISP method, PC coefficients are computed through an  $L^2$ -projection of variables onto a PC basis [22]. On the other hand, in the CM method, one computes the PC coefficients by using the PC basis as a set of interpolants [1, 28, 35, 47]. Finally, the regression-like approaches find the PC coefficients by minimizing the distance between the PC expansion and a set of observations [2].

In this paper, we will follow a NISP approach to compute the coefficients in a PC expansion. Non-intrusive PC methods in general, and NISP in particular, suffer from the so-called curse of dimensionality. In the NISP method, this phenomenon is seen through the rapid increase of the number of deterministic realizations needed for computations of the PC coefficients as the order of expansion and the number of stochastic dimensions increase. To tackle this issue, a sparse quadrature techniques will be explored, and the quality of the representation will be monitored as the sparse grid is successively refined.

We restrict our attention on the impact of parametric uncertainties in HYCOM, specifically those appearing in parameterizations of subgrid mixing and wind drag.

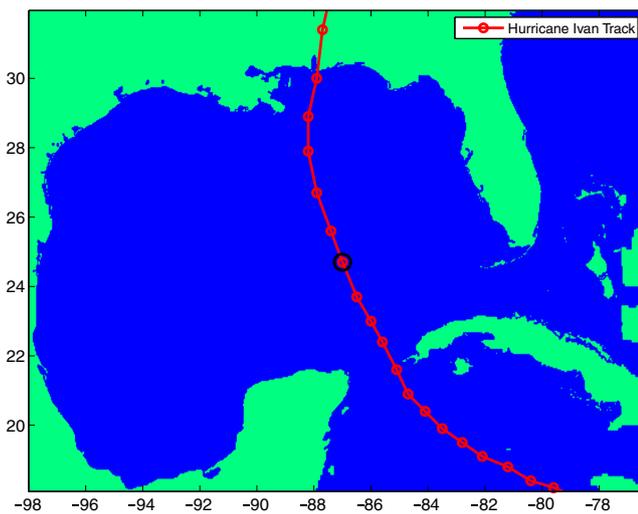
Uncertainties are propagated through short-time high-resolution simulations in the Gulf of Mexico, with wind stresses corresponding to hurricane Ivan. We exploit spectral representations to conduct a systematic assessment of the effect of uncertainties on quantities of interest, including global and local sensitivities with respect to the uncertain input parameters. Both local field variables as well as integral measures are considered in the analysis. In addition to quantifying global and local sensitivities, we develop a simplified measure transform technique that enables us to efficiently assess the impact of restricting the range of selected input parameters and thus demonstrate the potential of PC representations in managing uncertainties in ocean models.

The structure of this paper is as follows: Section 2 provides a brief discussion of HYCOM initialization, forcing, and wind drag parameterization. In Section 3, we introduce essential notation and review relevant PC concepts, formulate the stochastic problem, outline the sparse grid NISP approach, and briefly examine the effect of sparse grid refinement. In Section 4, we utilize the PC representation to gain insight into the the stochastic response of field variables to uncertain model inputs. In Section 5, we provide global and local analysis of sensitivities in field variables and also illustrate the implementation of measure transform computations. In Section 6, we investigate statistical properties of integral QoIs, namely the regionally averaged sea surface temperature and the average heat flux in a circular area around the eye of the hurricane. Finally, in Section 7, we provide concluding remarks and discuss the possible extension of the present methodology.

## 2 HYCOM in the Gulf of Mexico

Our study focuses on quantifying ocean model uncertainties associated with mixed layer and air–sea momentum exchange parameterizations, when the ocean is forced by hurricane-strength winds. The wind stress and the mixed layer parameterization modulate the ocean sea surface temperature (SST) response to the hurricane forcing, a key parameter in determining the exchanges of energy and momentum between the ocean and atmosphere and in impacting hurricane intensity forecast.

The setting chosen for our experiments is the circulation in the Gulf of Mexico during the passage of hurricane Ivan from 9–16 Sep 2004. Ivan's track is shown in Fig. 1. Several “classical” ocean model sensitivity studies have already been carried out (e.g.,



**Fig. 1** Hurricane Ivan entered the Caribbean on Sep 9 as a category 5 with winds of 160 mph; its intensity decreased to a category 4 upon crossing the Yucatan Strait on Sep 14; it made landfall as a category 3, with sustained winds of near 120 mph, on Sep 16 just west of Gulf Shores, Alabama

[15, 37, 50]); here we focus on using PC expansions to systematically characterize the entire response surface of the ocean model to parametric uncertainties in its subgrid parameterizations. Below, we briefly describe the “baseline” setup and specify the random inputs used in the analysis.

### 2.1 HYCOM

The model used is HYCOM, a free-surface general circulation ocean model that solves the hydrostatic Navier–Stokes equations. These include equations for horizontal momentum and mass conservation in addition to scalar transport equations for temperature and salinity. HYCOM relies on a hybrid vertical coordinate system to discretize the partial differential equations [3]. This system is designed to be quasi-optimal throughout the various flow regimes encountered in the ocean: isopycnic in the open stratified ocean to minimize spurious diapycnal mixing, terrain-following in coastal regions to faithfully represent flow-topography interactions, and  $z$ -coordinate near the surface to resolve mixed layer processes. A major advantage of this system is the ability to use advanced subgrid scale parameterizations for the ocean mixed layer while retaining the advantages of an isopycnal model in the ocean interior. For additional detail concerning the model, see, e.g., [3–5, 14].

Our Gulf of Mexico nested HYCOM configuration is very similar to the one adopted in [15, 37, 50]. The computational domain encompasses the Gulf of Mexico

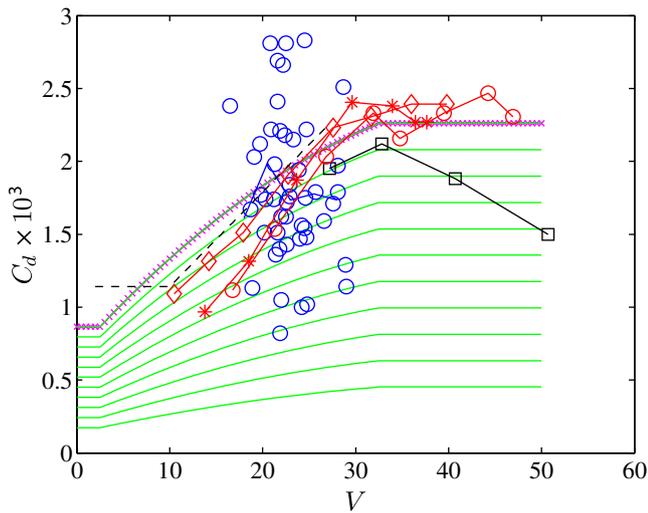
and portions of the Caribbean Sea; its grid resolution is  $\frac{1}{25}^\circ$  ( $\approx 4$  km) in the horizontal and 20 layers are used to discretize the vertical. The initial and boundary conditions are taken from a data assimilative  $1/12^\circ$  global HYCOM simulation so that major oceanic features, such as the loop current and its associated warm and cold core eddies, are positioned at their correct locations during Ivan’s transit. Atmospheric forcing fields are taken from the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) [16] at 27-km/three-hourly resolution; we note that the spatial coverage is too coarse to resolve all features of interest but should be adequate for our purposes. All the simulations performed herein covered the time window 9–16 Sep 2004, and model outputs were recorded at 3-h time intervals.

### 2.2 Mixed layer uncertainties

The subgrid scale parameterization adopted in the current experiments for the ocean mixed layer is K-profile parameterization (KPP) [20], a widely used vertical mixing scheme. KPP predicts an ocean boundary layer depth within which turbulent mixing is parameterized using a nonlocal bulk Richardson number, defined relative to the surface, and the similarity theory of turbulence. This boundary layer depth is determined by the depth at which the bulk Richardson number reaches a critical Value,  $Ri_c$ . Below the boundary layer, the vertical mixing is parameterized through the local gradient Richardson number and a background mixing. Here we follow [26] in perturbing three KPP’s parameters: the critical Richardson number, and the background viscosity and diffusivity; a combination of 0.3,  $10^{-4}$  m<sup>2</sup>/s, and  $10^{-5}$  m<sup>2</sup>/s for these parameters proved adequate in reproducing climatological observations of the zonal amplitude of equatorial currents. The range of background diffusion explored herein,  $10^{-5}$ – $10^{-4}$ , spans the entire range of (low) observed background diffusion to the (high) amount deemed necessary to maintain the observed oceanic stratification [29].

### 2.3 Wind drag parametric uncertainties

Momentum exchange between the ocean and atmosphere is effected through the wind stresses at their interface, and the latter play a key role in ocean mixed layer dynamics and in determining the SST. The wind stresses are commonly computed from a drag law of the form  $\tau = \rho C_d \|V\| V$  where  $C_d$  is the drag coefficient,  $\rho$  the air density, and  $V$  is the difference between air velocity at 10 m and ocean current velocity (the latter is usually neglected since air speed is much larger than



**Fig. 2** Observations and parameterizations of wind drag coefficient  $C_d$ . The *green lines* show our stochastic parameterization of  $C_d$  which is seen to fall within the envelope of various observations/fits at high-wind speeds. The *blue lines and circles* are the aircraft observation of [9], the *three-red lines* are fitted curves to the laboratory experiments of [8], and the *black lines* are the dropsondes observation of [36]. The *dashed line* refers to [21] parameterization that does not take into account  $C_d$ 's saturation at high  $V$ . The unperturbed HYCOM parameterization of  $C_d$ , due to [19] for  $\delta T = 0$ , is shown in *magenta*

current speed). The drag coefficient, in turn, is inferred from bulk parameterization formulas that depend on a number of atmospheric variables. Here we have used a fit due to Kara [19] to estimate the drag coefficient given the air speed at 10 m height,  $V$ , and the air–sea temperature difference,  $\delta T$ :

$$V_a = \max(2.5, \min(32.5, V)) \tag{1}$$

$$C_d^0 = a_0 + a_1 V_a + a_2 V_a^2 \tag{2}$$

$$C_d^1 = b_0 + b_1 V_a + b_2 V_a^2 \tag{3}$$

$$C_d = C_d^0 + C_d^1 \delta T \tag{4}$$

where  $a_i, b_i$  are coefficients determined from a least square fit to COARE v2.5.  $C_d^1$  is usually small and decreases to zero with air speed  $V$ ; hence air–sea temperature differences are not important in hurricane conditions. We represent the uncertainties in  $C_d$  with a multiplicative factor in the interval  $[0.2, 1]$ , a range that fits within the uncertainties in observational and experiment data as shown in Fig. 2.

### 3 The polynomial chaos framework

We begin this section by fixing the notation used throughout the paper and collecting the background

results needed in what follows. We denote by  $(\Omega, \mathcal{F}, \mu)$  a probability space, where  $\Omega$  is the sample space,  $\mathcal{F}$  is an appropriate  $\sigma$ -algebra on  $\Omega$ , and  $\mu$  is a probability measure. For a random variable  $\xi$  on  $\Omega$ , we write  $\xi \sim \mathcal{U}(a, b)$  to mean that  $\xi$  is uniformly distributed on the interval  $[a, b]$ . We use the term iid for a collection of random variables to mean that they are independent and identically distributed. The distribution function [18, 46] of a random variable  $\xi$  on  $\Omega$  is given by  $F_\xi(x) = \mu(\xi \leq x)$  for  $x \in \mathbb{R}$ .

#### 3.1 Polynomial chaos

In the present work, we consider models with finitely many uncertain parameters. We parameterize these uncertain parameters by a finite collection of real-valued iid random variables  $\xi_1, \dots, \xi_d$  on  $\Omega$ . We refer to  $d$  the number of random parameters, as the *dimension of the stochastic problem*. Let  $F_\xi$  denote the joint distribution function of the random vector  $\xi = (\xi_1, \dots, \xi_d)^T$ . Since the  $\xi_j$  are iid,  $F_\xi(\mathbf{x}) = \prod_{j=1}^d F(x_j)$  for  $\mathbf{x} \in \mathbb{R}^d$ , where  $F$  is the common distribution function for  $\xi_1, \dots, \xi_d$ .

For computational purposes, it is convenient to work in the image probability space  $(\Omega^*, \mathcal{B}(\Omega^*), F_\xi)$ , where  $\Omega^* \subseteq \mathbb{R}^d$  is the image of  $\Omega$  under  $\xi$ , and  $\mathcal{B}(\Omega^*)$  is the Borel  $\sigma$ -algebra on  $\Omega^*$ . We denote the expectation of a random variable  $X : \Omega^* \rightarrow \mathbb{R}$  by

$$\langle X \rangle = \int_{\Omega^*} X(\mathbf{s}) dF_\xi(\mathbf{s}).$$

The space of square integrable random variables on  $\Omega^*$ ,  $L^2(\Omega^*)$ , is endowed with the inner product:

$$(X, Y) = \int_{\Omega^*} X(\mathbf{s})Y(\mathbf{s}) dF_\xi(\mathbf{s}) = \langle XY \rangle,$$

and the norm of  $X$  is given by  $\|X\|_{L^2(\Omega^*)} = (X, X)^{1/2} = \langle X^2 \rangle^{1/2}$ .

In the present work, we will be dealing with uncertain inputs ranging on finite intervals. Therefore, we will parameterize these inputs with canonical random variables  $\xi_i \stackrel{iid}{\sim} \mathcal{U}(-1, 1)$ . Consequently, we will rely on the basis formed by the  $d$ -variate Legendre polynomials  $\{\Psi_k\}_0^\infty$ . Each  $\Psi_k$  is obtained through a product of 1D Legendre polynomials according to:

$$\Psi_k(\xi) = \prod_{i=1}^d \psi_{\alpha_i^k}(\xi_i), \quad \xi \in \Omega^*, \tag{5}$$

where  $\alpha^k = (\alpha_1^k, \alpha_2^k, \dots, \alpha_d^k)$  is a multi-index, with  $\alpha_i^k$  being the order of the 1D Legendre polynomial,  $\psi$ ,

in  $\xi_i$ . With this basis, any  $X \in L^2(\Omega^*)$  admits an expansion of the form:

$$X = \sum_{k=0}^{\infty} c_k \Psi_k, \tag{6}$$

known as a generalized polynomial chaos [48] expansion of  $X$ . The multi-index construction appearing in Eq. 5 will be also exploited in the computation of the Sobol (global) sensitivity indices, as discussed in Section 5.1 below.

In computations, we approximate  $X(\xi)$  with a truncated series,

$$X(\xi) \doteq \sum_{k=0}^P c_k \Psi_k(\xi), \tag{7}$$

where  $P$  is finite and depends on the truncation strategy adopted. In the present work, we consider truncations based on the total degree of the polynomials in the series. In this case,  $P$  depends of the stochastic dimension  $d$  and expansion “order,”  $p$ , according to

$$P = \frac{(d + p)!}{d!p!} - 1, \tag{8}$$

where  $p$  is the largest polynomial degree in the expansion.

With  $X$  expanded as in Eq. 7, using the orthogonality of the basis  $\{\Psi_k\}_0^P$  and the convention that  $\Psi_0 \equiv 1$ , we have immediate access to the moments

$$\begin{aligned} \langle X \rangle &= c_0, \\ \langle X^2 \rangle &= \sum_{k=0}^P c_k^2 \langle \Psi_k^2 \rangle. \end{aligned}$$

Moreover, we have

$$\text{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = \sum_{k=1}^P c_k^2 \langle \Psi_k^2 \rangle,$$

and the covariance of the two random variables,  $X = \sum_{k=0}^P x_k \Psi_k$  and  $Y = \sum_{k=0}^P y_k \Psi_k$ , is given by,

$$\text{cov}(X, Y) = \sum_{k=1}^P x_k y_k \langle \Psi_k^2 \rangle.$$

### 3.2 Non-intrusive spectral projection

Let  $X$  belong to  $L^2(\Omega^*)$ . As mentioned in the introduction, non-intrusive methods aim at computing the PC coefficients in the finite expansion (Eq. 7) via a set of deterministic evaluations of  $X(\xi)$  for specific

realizations of  $\xi$ . Observe that since  $\{\Psi_k\}_0^P$  form an orthogonal system, we have:

$$(X, \Psi_k) = \left( \sum_{l=0}^P c_l \Psi_l, \Psi_k \right) = \sum_{l=0}^P c_l (\Psi_l, \Psi_k) = c_k (\Psi_k, \Psi_k), \tag{9}$$

so that the coefficient  $c_k$  is given by

$$c_k = \frac{\langle X \Psi_k \rangle}{\langle \Psi_k^2 \rangle}. \tag{10}$$

The moments  $\langle \Psi_k^2 \rangle$  of the multivariate Legendre polynomials in Eq. 10 can be computed analytically [22], and hence, the determination of coefficients  $c_k$  amounts to the evaluation of the moments  $\langle X \Psi_k \rangle$ . We note that

$$\langle X \Psi_k \rangle = \int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_{\xi}(\mathbf{s}), \quad k = 0, \dots, P,$$

leading to the evaluation of a set of  $P + 1$  integrals over  $\Omega^* \subseteq \mathbb{R}^d$ . In the NISP approach, these integrals are discretized as finite sums of the form

$$\int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_{\xi}(\mathbf{s}) \doteq \sum_{j=1}^{N_q} w_j X(\xi_j) \Psi_k(\xi_j), \tag{11}$$

where  $\xi_j \in \Omega^*$  and  $w_j$  are the nodes and weights of an appropriate quadrature formula. Note that the same set of nodes is used to compute all the coefficients  $c_k$ , so the complexity of NISP scales with  $N_q$ , the number of nodes where one has to compute  $X$ . Therefore, the challenge is to design quadrature formulas yielding the lowest integration error for the minimal number of nodes. In general, this is a difficult problem, and one often proceeds by tensorization of 1D quadrature rules. Considering a 1D quadrature rule with  $n$  nodes, its full tensorization gives a  $d$ -variate formula having  $N_q = n^d$  nodes, showing that this approach is limited to low  $d$ . This exponential scaling with  $d$  is often referred to as *the curse of dimensionality*. An effective way of mitigating the curse of dimensionality is through the use of sparse tensorizations of sequences of 1D formulas using Smolyak’s formula [41], leading to sparse quadrature techniques [10, 32, 33]. This is the approach adopted in this paper. Irrespective of the integration formula considered, the set of integration nodes comprise what we call the NISP sample; we denote it by

$$\mathcal{S} = \{\xi_j\}_{j=1}^{N_q} \subset \Omega^*.$$

Thus, to evaluate Eq. 11, we need to compute  $X(\xi_q)$  for all  $\xi_q \in \mathcal{S}$ . Let  $\mathbf{\Pi} \in \mathbb{R}^{(P+1) \times N_q}$  be the matrix given by

$$\Pi_{k,j} = \frac{w_j \Psi_k(\xi_j)}{\langle \Psi_k^2 \rangle}, \quad k = 0, \dots, P, \quad j = 1, \dots, N_q.$$

**Table 1** The random input parameters for HYCOM

Parameter	Description	Distribution
$p_1$	Critical Richardson number	$\mathcal{U}(0.25, 0.7)$
$p_2$	Background viscosity (m <sup>2</sup> /s)	$\mathcal{U}(10^{-4}, 10^{-3})$
$p_3$	Background diffusivity (m <sup>2</sup> /s)	$\mathcal{U}(10^{-5}, 10^{-4})$
$p_4$	Stochastic wind drag coefficient	$\mathcal{U}(0.2, 1.0)$

We call  $\mathbf{\Pi}$  the NISP projection matrix. If we denote by  $\zeta$  the vector with coordinates  $\zeta_j = X(\xi_j)$ , then the vector  $\mathbf{c} = (c_0, \dots, c_k)$  of the spectral coefficients in Eq. 10 is given by  $\mathbf{c} = \mathbf{\Pi}\zeta$ , or in components,

$$c_k = \sum_{j=1}^{N_q} \Pi_{kj} \zeta_j = \sum_{j=1}^{N_q} \Pi_{kj} X(\xi_j), \quad k = 0, \dots, P.$$

### 3.3 HYCOM uncertainties recast as stochastic variables

Let us denote by  $\mathbf{p} = (p_1, p_2, p_3, p_4)^T$  the vector of random model inputs. Following the discussion in Section 2, we model these using uniform random variables, as specified in Table 1.

Specifically, the inputs  $p_i$  are parameterized by  $\xi_i \sim \mathcal{U}(-1, 1)$ ,  $i = 1, \dots, 4$ , through

$$p_i(\xi) = \mu_i + \sigma_i \xi_i, \quad i = 1, \dots, 4,$$

where  $\xi = (\xi_1, \xi_2, \xi_3, \xi_4)^T$ ,

$$\mu_i = \frac{1}{2}(a_i + b_i), \quad \text{and} \quad \sigma_i = \frac{1}{2}(b_i - a_i),$$

so that  $p_i \sim \mathcal{U}(a_i, b_i)$  as in Table 1.

Let  $\mathcal{G}$  denote the physical domain, which in our case is the Gulf of Mexico (see Fig. 1). At a given time  $t$  and a point  $\mathbf{x} \in \mathcal{G}$  and a given vector of random inputs  $\mathbf{p}(\xi)$ , we denote the model output by the random vector  $\mathbf{X}(t, \mathbf{x}, \xi)$ , which is given by

$$\mathbf{X}(t, \mathbf{x}, \xi) = H(t, \mathbf{x}; \mathbf{p}(\xi)),$$

where  $H(t, \mathbf{x}; \mathbf{p}(\xi))$  signifies the result of a deterministic HYCOM solve at time  $t$  and point  $\mathbf{x}$  with input

**Table 2** Number of quadrature nodes versus Smolyak quadrature level

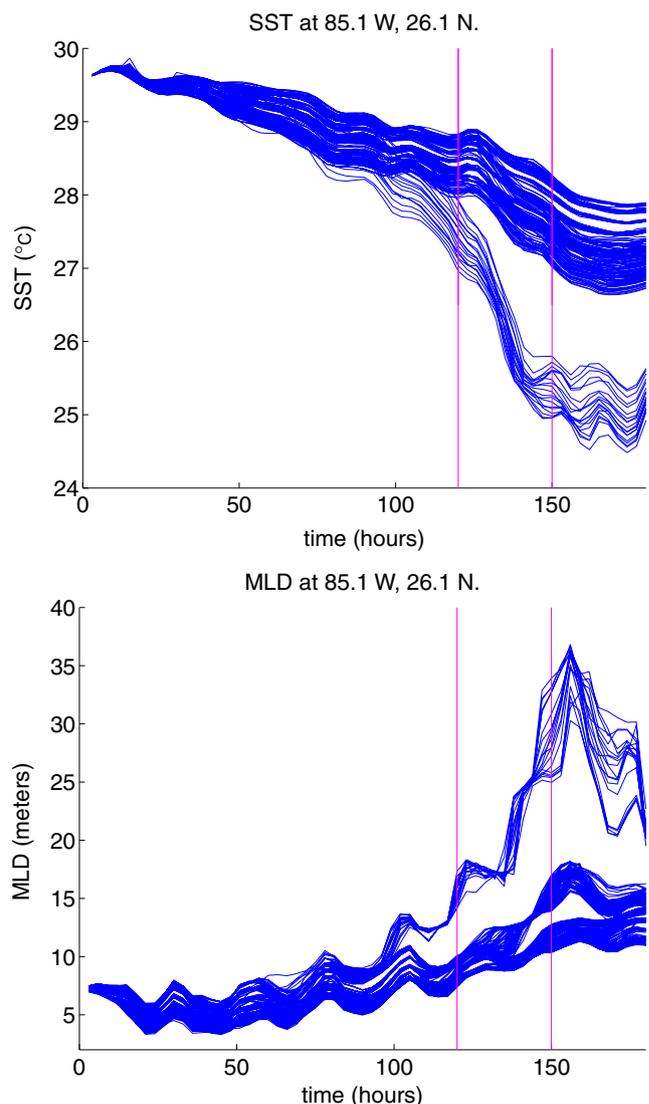
Quadrature level	Number of nodes
1	9
2	33
3	81
4	193
5	385

The Gauss–Kronrod–Patterson quadrature is used as the underlying rule

parameters  $\mathbf{p}(\xi)$ . Presently, we are mostly interested in SST, mixed layer depth (MLD), and to lesser extent sea surface height (SSH);  $\mathbf{X}$  may thus correspond to any of these QoIs.

### 3.4 Smolyak quadrature on random parameter space

As noted in Section 3.2, the NISP method proceeds by computing the PC coefficients via numerical integration as in Eq. 11. To mitigate the computational cost of numerical integration, we rely on a Smolyak quadrature [10, 32–34, 41], which is based on nested



**Fig. 3** Evolution of SST (*top*) and MLD (*bottom*) at the buoy location. The *curves* depict the 385 realizations corresponding to a level 5 Smolyak quadrature. The *first vertical line* shows the time when the hurricane enters the Gulf of Mexico, whereas the *second vertical line* corresponds to a time when the center of the hurricane is close to the buoy

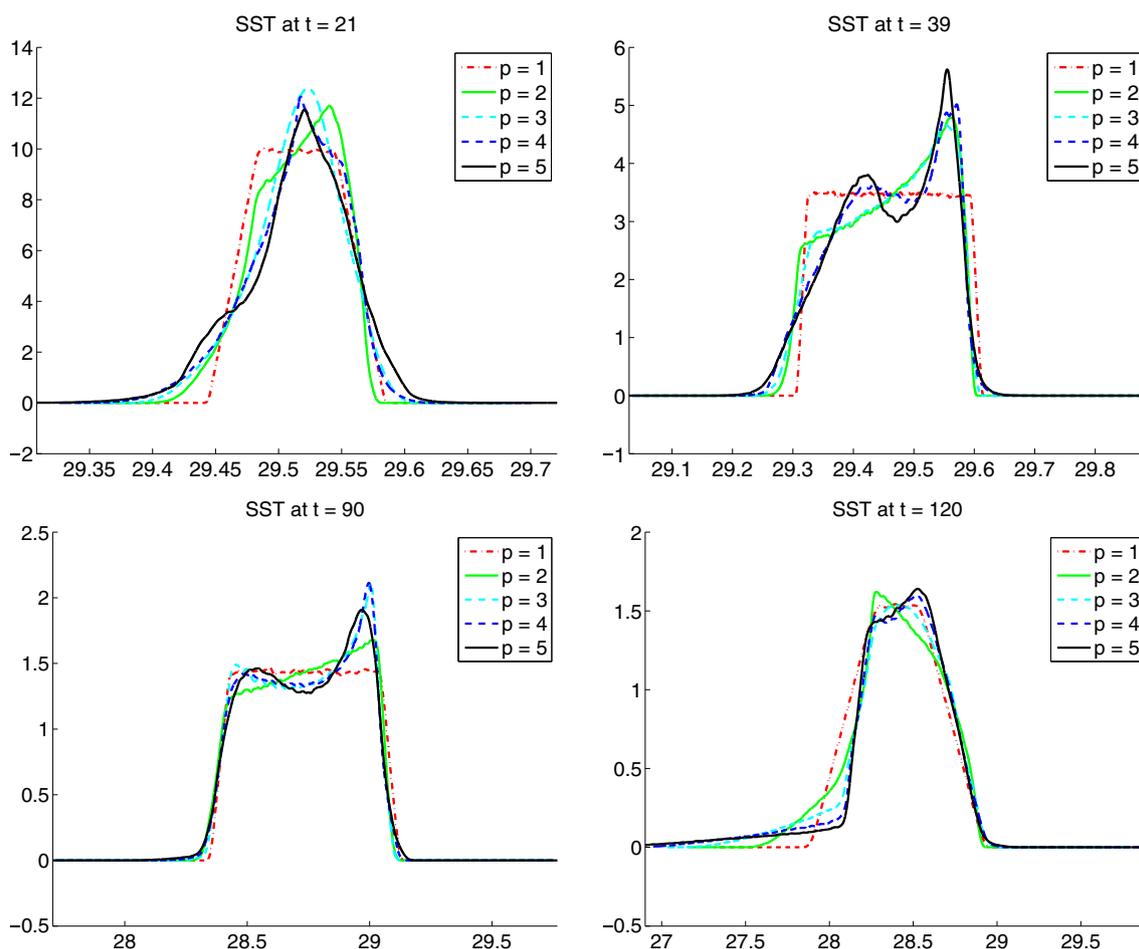
sparse grids. Increasing the resolution *level* leads to finer grids resulting in higher degrees of precision for the quadrature; for details, see [10, 32–34, 41]. In particular, a detailed account is provided in [10] on how to determine the Smolyak sparse grid (the set of integration nodes) and the corresponding weights. In our computations, we relied on SMOLPACK [32–34], with Gauss–Kronrod–Patterson as the basic 1D quadrature rule. For the present 4D case, we report in Table 2 the number of quadratures nodes associated with levels 1, 2, . . . , 5.

With the current truncation strategy used and the current sparse quadrature, one observes that the expansion coefficients of any function  $U \in \text{span}\{\Psi_k\}_0^p$  can be recovered exactly if the order,  $p$ , of the expansion is less or equal to the level of the quadrature. Thus, to integrate exactly a polynomial of order 5, 385 quadra-

ture nodes would be required (Table 2). In contrast, using a fully tensorized Gauss–Legendre quadrature would require six nodes in each of the four stochastic dimensions and hence a total of  $6^4 = 1,296$  nodes for the same degree of precision. This illustrates the savings afforded by the sparse quadrature.

### 3.5 PC representation of the model variables

In the computations, we used NISP with a level 5 Smolyak quadrature to compute the spectral expansion of the model output in the PC basis. In Fig. 3, we plot the time evolution of SST and MLD at 85.1 W/26.1 N. This corresponds to the location of buoy 42003 in the Gulf of Mexico, where observational data are available. To ascertain that a reasonable representation is



**Fig. 4** Distribution of SST at the buoy location at selected times, as indicated. Different scales are used in each panel to facilitate comparison of the different curves.  $p$  denotes the highest poly-

nomial degree in the truncated expansion. In each case, the PC expansion was sampled  $10^6$  times to generate the distribution curve

achieved with a level 5 quadrature, we have examined the convergence of the PC representation as the number of levels was increased. This was facilitated by the nested nature of the sparse grids, which naturally minimized the number of realizations required. A brief highlight of this analysis is provided below.

Figure 4 shows instantaneous distributions of SST at the buoy location. These are obtained by sampling the PC expansion at selected times. The distributions are generated using first-, second-, third-, fourth-, and fifth-order PC expansions computed using NISP with a level five sparse grid. Note that for the  $p = 1$  case, we have a linear combination of iid uniform random variables plus the mean. As seen in the plots, the distributions seem to level off as the PC order is increased to  $p = 5$  suggesting that a fifth-order expansion is sufficient.

To get further confidence in the spectral representations, approximate relative  $L^2$  errors were computed.

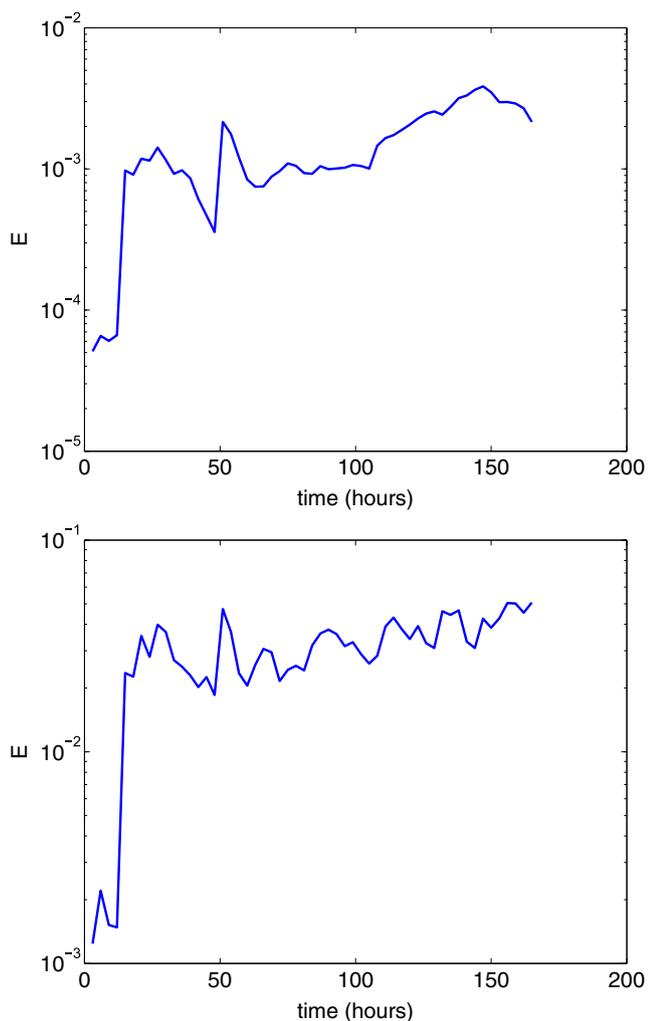


Fig. 5 Evolution of  $E$  for SST (top) and MLD (bottom)

With  $X$  in  $L^2(\Omega^*)$ , the relative  $L^2$  error between  $X$  and its (truncated) PC representation is given by,

$$\frac{\|X - \sum_{k=0}^P X_k \Psi_k\|_{L^2(\Omega^*)}}{\|X\|_{L^2(\Omega^*)}} = \frac{\left(\int_{\Omega^*} |X - \sum_{k=0}^P X_k \Psi_k|^2 dF_{\xi}\right)^{1/2}}{\left(\int_{\Omega^*} |X|^2 dF_{\xi}\right)^{1/2}},$$

which can be approximated through Monte Carlo (MC) integration. Such sampling, however, is prohibitively demanding. Instead, we approximate this relative error by:

$$E = \frac{\left(\sum_{\xi \in \mathcal{S}} |X(\xi) - \sum_{k=0}^P X_k \Psi_k(\xi)|^2\right)^{1/2}}{\left(\sum_{\xi \in \mathcal{S}} |X(\xi)|^2\right)^{1/2}},$$

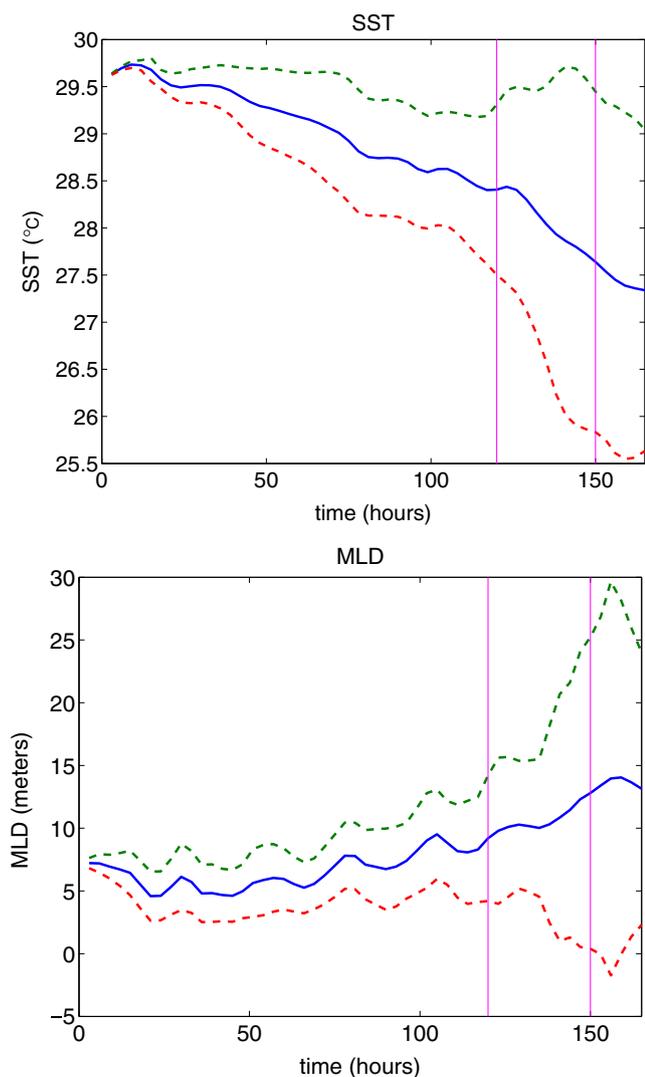
where  $\mathcal{S} \subseteq \Omega^*$  is the NISP sample set. Note that  $E$  is the relative  $\ell^2$  error on the NISP sample set.

Figure 5 shows the evolution of  $E$ , computed for SST and MLD at the buoy location. The NISP sample corresponding to a level 5 quadrature  $|\mathcal{S}| = 385$  was used for this purpose. The results indicate that  $E$  is the largest in the case of MLD, which still falls below about 5% throughout the computation. Thus, the PC expansion coincides reasonably well the underlying realization data.

### 4 Basic illustrations

The realizations illustrated in the previous section indicate that the stochastic model response to uncertain data can generally be quite complex, even on relatively short time scales. This can be appreciated from the evolution of SST and MLD at the buoy location, which reveal the occurrence of distinct bands. In this section, we briefly illustrate how the PC representations can be used to quantify the uncertainty in the response of QoIs. We first focus on the local behavior of SST and MLD at the buoy location and provide basic illustration of the computation of low-order moments of the solution. We then provide examples, illustrating the use of the PC representations in estimating the probability of a QoI being in a given interval.

Figure 6 shows the evolution of the mean SST and MLD at the buoy location. To quantify the impact

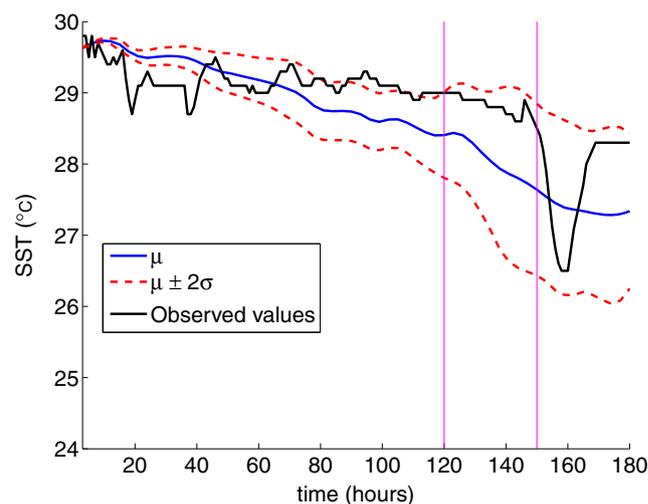


**Fig. 6** Mean and variances of the model quantities at the buoy location for SST (*top*) and MLD (*bottom*). The *first* vertical line indicates the time the hurricane enters the Gulf whereas the *second* line indicates a time at which the hurricane is close to the buoy

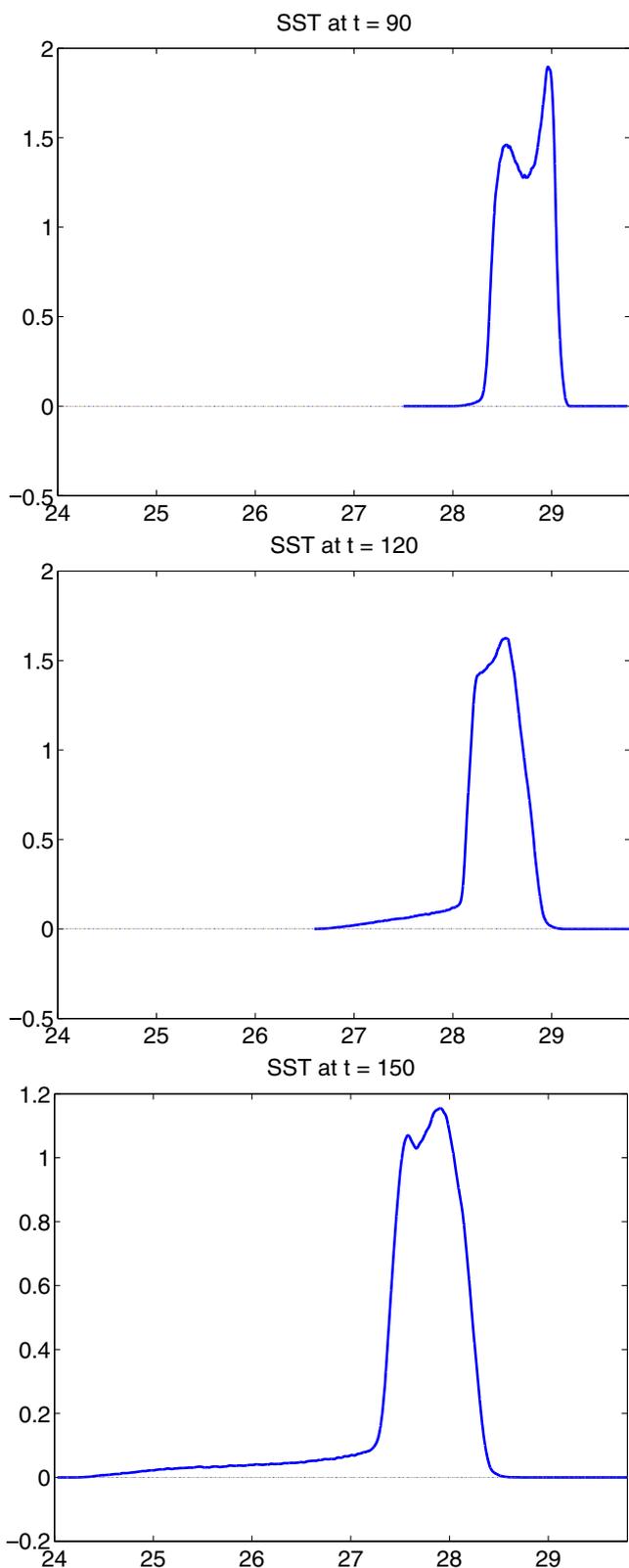
of uncertainty in the model inputs, curves indicating two standard deviation bounds are also depicted. The results indicate that the standard deviation generally increases as time evolves, though its evolution is evidently non-monotonic. Its behavior, however, is consistent with the spread of the individual realizations reported in Fig. 3.

Of course, direct comparison of the local predictions obtained by ocean global circulation model (OGCMs) to local field observations is generally not appropriate. This is the case because, even in the absence of any model and numerical errors, OGCM predictions are inherently filtered, whereas field observations would

reflect spatial and temporal fluctuations that cannot be captured by the model. Notwithstanding these limitations, in Fig. 7, we compare the field observations of SST at the location of buoy 42003, to the corresponding stochastic model predictions. Plotted in Fig. 7 are the evolution of the mean SST, as well as curves depicting two standard deviation bounds. Superficial comparison of the results would appear to indicate that the field observation generally fall within the predicted uncertainty band, though local excursions can be observed, especially at small times. However, one should caution against making such an assessment, namely because the stochastic model response is evidently complex and the distribution functions are highly non-Gaussian (this is illustrated in Fig. 8, which shows the distribution of SST at selected time instants). Clearly, in these situations, it is generally not appropriate to characterize the uncertainty in QoIs in terms of means and variances only. This is clearly the case in the present setting, as none of the individual model realizations depicted in Fig. 3 appears to provide a very good match with the observational data. Specifically, the model simulations do not reproduce the quick cooling and heating recorded in the observation around hour 160. This could be due to the local nature of the buoy measurements whereas the ocean model cannot resolve features below 3 km (this motivates us to look for global QoIs for model-data inter-comparisons). Furthermore, the COAMPS winds used as forcing are coarse spatially (27 km) whereas Ivan’s radius of maximum winds is only about 45 km. It is thus likely that COAMPS smears the real winds by decreasing their amplitude and broadening their spatial



**Fig. 7** Comparing stochastic mean  $\mu$  with two standard deviation bounds with observed SST data. The *first* vertical line indicates the time Ivan enters the Gulf, and the *second* one indicates the time the hurricane is closest to the buoy

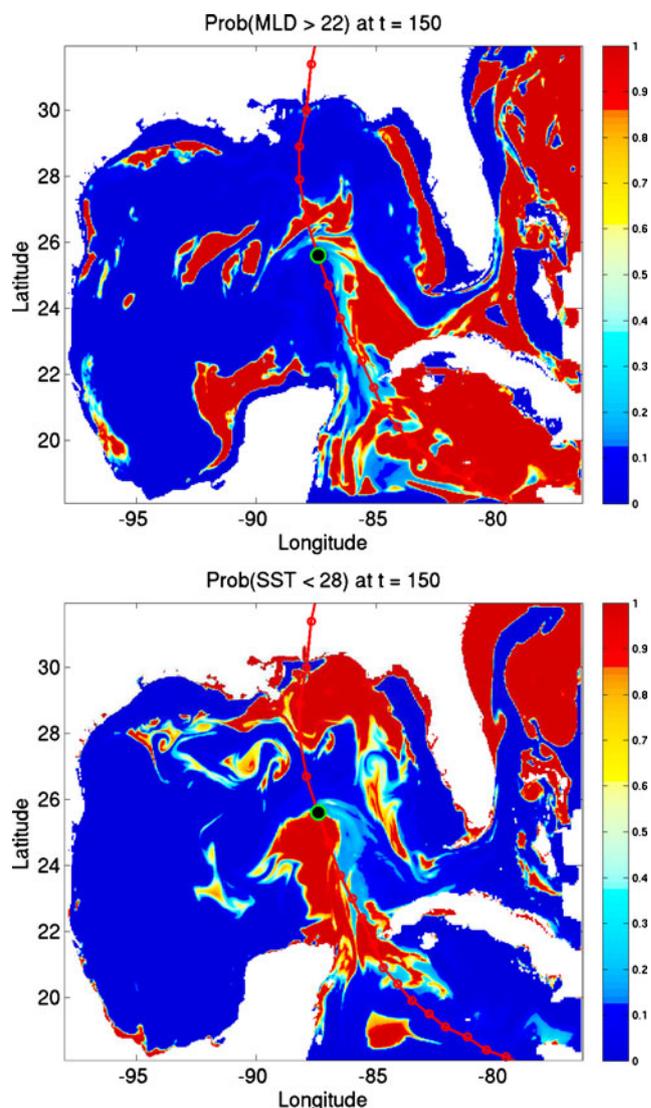


**Fig. 8** Distribution of SST at the buoy location at different times, as indicated. Note that as the hurricane traverses the Gulf a cooling trend can be observed, reflected by the shift of the distribution function to the left. Also, note the emergence of long tails for the distributions stretching toward cooler temperatures

extent. This explanation is consistent with the early cooling trend seen in the HYCOM realizations and with the absence of quick heating following the storm's passage.

One of the advantages of PC representations is that they can be sampled efficiently. This was exploited earlier in the generation of probability density functions, obtained through extensive sampling of the polynomial basis. We conclude this section by providing an additional example, namely concerning evaluation of expressions of the form:

$$\text{Prob}\{X > \beta\}.$$



**Fig. 9** Probability contours at  $t = 150$ . *Top*  $\text{Prob}\{\text{MLD} > 22\}$ ; *bottom*:  $\text{Prob}\{\text{SST} < 28^\circ\text{C}\}$ . The contours are generated at  $t = 150$  h. In both cases,  $10^4$  samples are drawn from the corresponding PC representations. The *bottom plot* highlights the cool fronts approaching the coast as the hurricane nears landfall

Such estimates are typically obtained through MC sampling. Specifically, one chooses a sample  $S \subset \Omega^*$  and uses

$$\text{Prob}\{X > \beta\} \approx \frac{\#\{\xi \in S \mid X(\xi) > \beta\}}{\#(S)}, \tag{12}$$

where  $\#(S)$  denotes the number of elements of the set  $S$ . The approximation improves as the sample size  $\#(S) \rightarrow \infty$ . Of course, the sampling becomes prohibitively costly if the realization appearing in Eq. 12 was to rely on independent model evaluations, especially when individual model evaluations are quite involved as is the case in ocean computations. Alternatively, by sampling the PC representation of  $X$ , such estimates can be obtained at a minute fraction of direct MC sampling. A sample of such computations is shown in Fig. 9, which depicts instantaneous contours of the probability that MLD exceeds 22 m and the probability that SST is cooler than 28 C. In both cases,  $10^4$  samples are drawn from the PC representation of these QoIs.

### 5 Stochastic sensitivity analysis

PC representations enable the analysis of model output sensitivities with modest computational cost. This includes both global sensitivity analysis, particularly using Sobol indices [6, 17, 22, 42–44], and local sensitivity analysis by differentiating the PC expansion [39]. Global sensitivity analysis aims to quantify the contribution of different random input parameters to the model variability, whereas local sensitivity analysis aims at quantifying the response of the model to local changes around a particular realization. In this section, we briefly outline the mathematical setup of the global and local sensitivity analyses we use in the remainder of the present study.

#### 5.1 Global sensitivity analysis via Sobol indices

The variance-based, global sensitivity analysis conducted in this section is inspired by the ANOVA (or Sobol) decomposition of square integrable functions of several variables [17, 42, 43]. As described in [6, 22, 44], PC representations of random variables enable efficient approximation of the corresponding variance-based, global sensitivity indices. Below, we briefly describe the computation of the so-called *total* sensitivity indices of random variables from their PC representations.

Suppose  $U$  is a square integrable random variable with PC representation given by

$$U(\xi) = \sum_{k=0}^P U_k \Psi_k(\xi), \quad \xi \in [-1, 1]^d,$$

where  $d$  denotes the stochastic dimension. We denote by  $\alpha^k$  the multi-index associated with  $k$  – th term in the PC expansion [22]. For each index set  $u \subseteq \{1, \dots, d\}$ , we define

$$K_u = \left\{ k \in \{1, \dots, P\} : \begin{cases} \alpha_i^k > 0, & \text{if } i \in u, \\ \alpha_i^k = 0, & \text{if } i \notin u \end{cases} \right\}.$$

Note that  $K_u$  picks multi-indices that have non-zero entries at the indices specified by  $u = (i_1, \dots, i_s)$  and zero entries everywhere else. Then, the Sobol sensitivity indices [6, 22, 44] for  $u$  are given by,

$$S_u = \frac{\sum_{k \in K_u} U_k^2 \langle \Psi_k^2 \rangle}{\sum_{k=1}^P U_k^2 \langle \Psi_k^2 \rangle}, \quad u \subseteq \{1, \dots, d\}.$$

To get the total sensitivity corresponding to the  $i$  – th input  $\xi_i$ , we compute the total index [6, 22, 44]:

$$T_i = \sum_{u \ni i} S_u = \frac{\sum_{u \ni i} \sum_{k \in K_u} U_k^2 \langle \Psi_k^2 \rangle}{\sum_{k=1}^P U_k^2 \langle \Psi_k^2 \rangle}. \tag{13}$$

The above formula can be simplified as follows: Let  $\mathcal{J}_i$  be the index set given by

$$\mathcal{J}_i = \{k \in \{1, \dots, P\} : \alpha_i^k > 0\},$$

which picks all modes with an input from  $\xi_i$ . Note that  $\mathcal{J}_i = \bigcup_{u \ni i} K_u$  for  $i = 1, \dots, n$ . Therefore,

$$T_i = \frac{\sum_{k \in \mathcal{J}_i} U_k^2 \langle \Psi_k^2 \rangle}{\sum_{k=1}^P U_k^2 \langle \Psi_k^2 \rangle}. \tag{14}$$

Using Eq. 14, the computation of  $T_i$  is straightforward. Note that the total sensitivity index  $T_i$  measures the contribution of the  $i$  – th random input to total model variability by computing the fraction of the total variance due to all the terms in the PC expansion which involve  $\xi_i$ .

It is also worth noting that for random variables expanded in a PC basis, all the index sets above are dictated by the basis alone. Thus, we need to find

the index sets  $K_u$  for computation of  $S_u$  and  $\mathcal{I}_i$  for computation of  $T_i$  only once. Hence, the computation of  $S_u$  or  $T_i$  for a random variable is immediate once its PC coefficients are determined.

### 5.1.1 Global sensitivity analysis

Using the total sensitivity indices introduced in Section 5.1, we can assess the contribution of each of the random inputs, parameterized by  $\xi_1, \dots, \xi_4$ , to the model variability. For clarity, we recall the significance of the sensitivity indices  $T_1, \dots, T_4$  in Table 3.

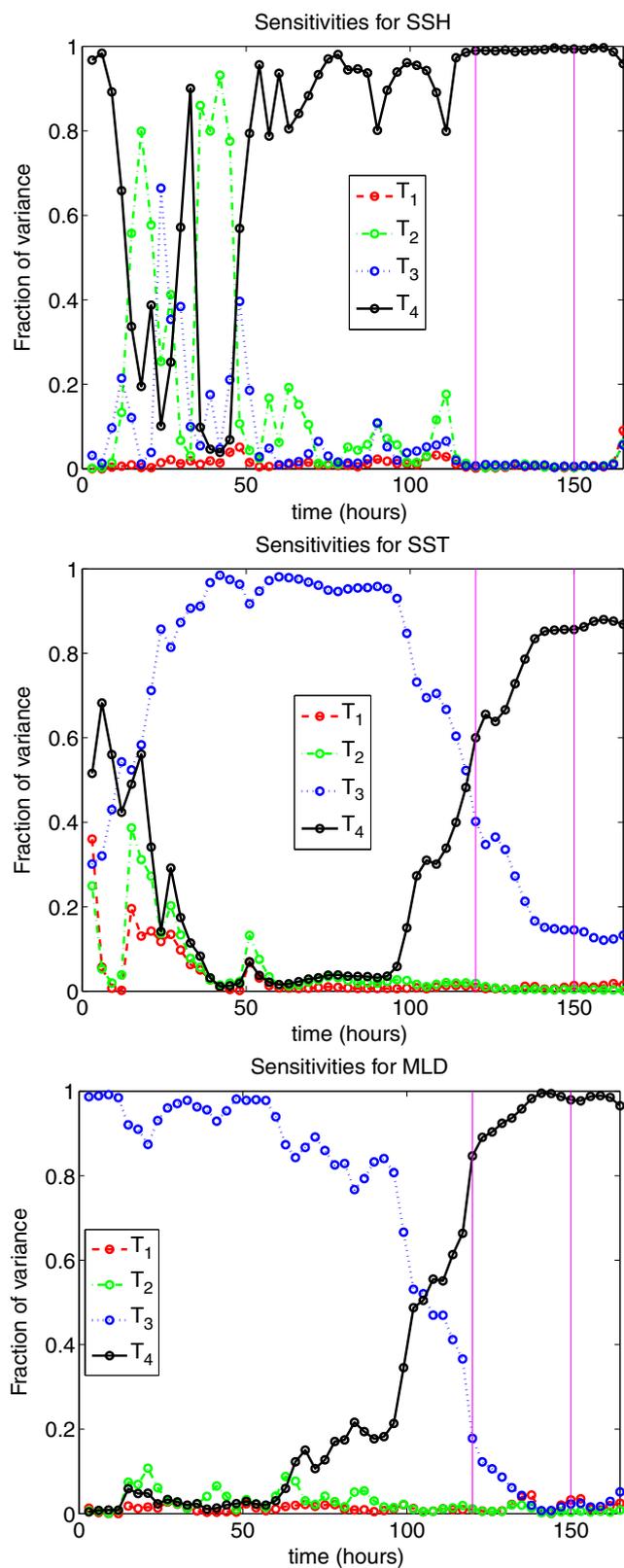
Figure 10 depicts the global sensitivity indices  $T_i$ ,  $i = 1, \dots, 4$  for SSH, SST, and MLD. The results were computed using the PC expansions of model variables at the location of the buoy 42003. A basic observation from these results is that as time evolves,  $T_4$  becomes clearly dominant. This appears to coincide with the emergence of distinct bands in the model realizations, seen in Fig. 3. For the SST and MLD, the impact of background diffusivity is dominant in the initial stages, but uncertainty in wind drag coefficient clearly prevails as the hurricane enters the Gulf of Mexico (GOM) and approaches the buoy location. In the case of sensitivities for SSH, we note that  $T_4$  begin to dominate the other sensitivity indices much earlier in time. Specifically, for times larger than 50 h,  $\xi_1$ ,  $\xi_2$ , and  $\xi_3$  contribute very little to the total variance in SSH. Below, we explore how measure transforms can be used to quantify the range of wind drag uncertainty in which these trends remain valid. These transforms are also useful to re-assess the output uncertainty when the input uncertainty range is restricted without incurring additional model realizations. For example, one can explore a range for the drag coefficient that is wider than physically possible and then restrict its range a posteriori for more realistic assessments of output uncertainties and variance analyses.

### 5.1.2 Restriction of the wind drag coefficient

We saw in the previous subsection that the stochastic wind drag coefficient,  $p_4$ , parameterized by  $\xi_4$ , ultimately dominates the variance of the selected QoIs. Here we address the question of how much would one

**Table 3** Physical meaning of ocean model sensitivity indices

Index	Significance
$T_1$	Sensitivity due to critical Richardson number
$T_2$	Sensitivity due to background viscosity
$T_3$	Sensitivity due to background diffusivity
$T_4$	Sensitivity due to wind drag coefficient



**Fig. 10** Evolution of the global sensitivity indices  $T_1, \dots, T_4$  for SSH (top), SST (middle), 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

need to restrict the range of  $p_4$  so that its sensitivity is no longer dominant. We focus our attention to reducing the range of  $p_4$  around its mean value, namely according to

$$p_4^\alpha(\xi) = \frac{1}{2}(a_4 + b_4) + \frac{1}{2\alpha}(b_4 - a_4)\xi_4,$$

where  $\alpha > 1$  is a restriction factor. Table 4 shows how increasingly narrower intervals  $p_4$  are generated by increasing  $\alpha$ .

To determine sensitivities corresponding to restricted range, we exploit the existing PC representation as a *surrogate* for the actual random fields and thus avoid the need to generate new realizations. A simple means of achieving this is by relating a generic output associated with the scaled parameter range,  $Y_t$ , to the corresponding value associated with the original interval,  $X_t$ , using the transform:

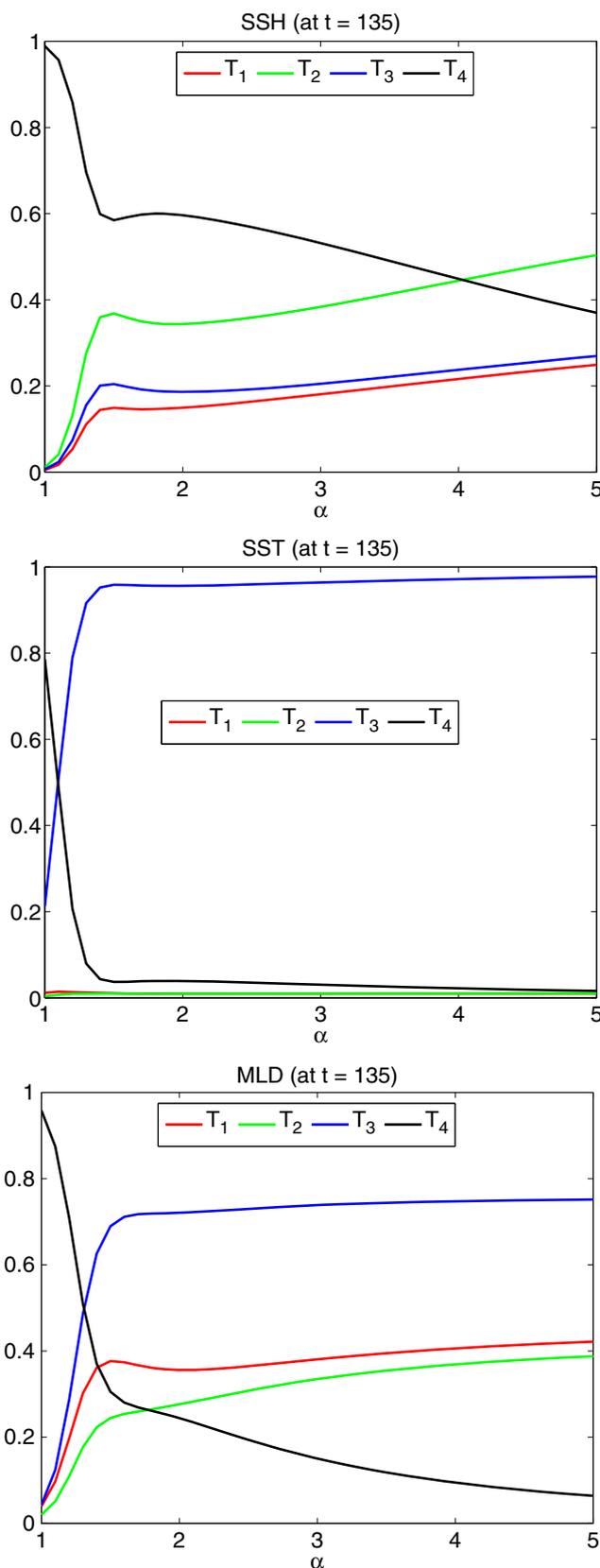
$$Y_t(\xi_1, \xi_2, \xi_3, \xi_4) = X_t(\xi_1, \xi_2, \xi_3, \xi_4/\alpha) \\ = \sum_{k=0}^P X_t^k \Psi_k(\xi_1, \xi_2, \xi_3, \xi_4/\alpha).$$

$Y_t$  can then be projected into the PC basis to obtain a new set of PC coefficients.

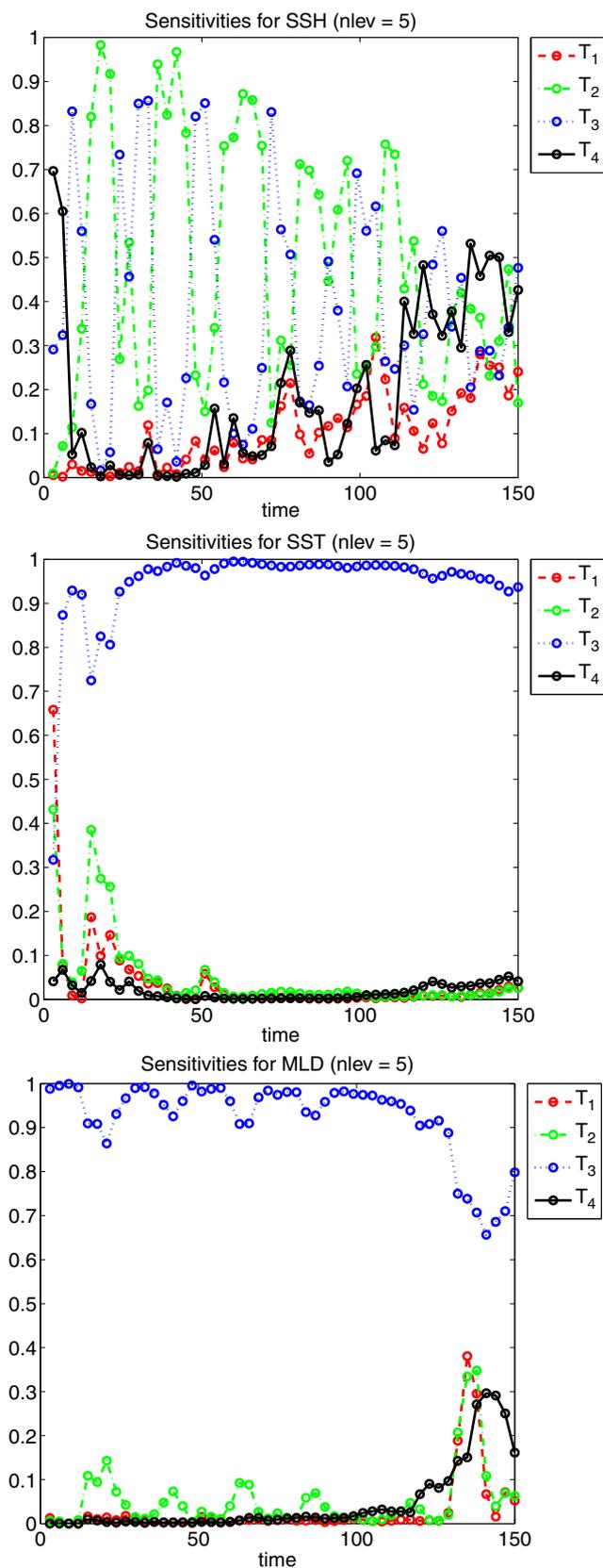
Figure 11 shows the dependence of  $T_1, \dots, T_4$  for increasing values of  $\alpha$ . Plotted are curves depicting the global sensitivities of SSH, SST, and MLD at  $t = 135$  h, a time at which  $\xi_4$  has become dominant. The results readily yield quantitative predictions, for each of these QoIs, of how much  $\xi_4$  must be restricted for its global sensitivity index to cease dominating the others. In particular, it is interesting to note that a restriction factor  $\alpha \simeq 4$  is needed to achieve this effect for SSH, whereas significantly smaller values would be needed for SST and MLD. Also note that, as  $\alpha$  increases, the uncertainty in SSH becomes dominated by  $\xi_2$ , whereas for SST and MLD, the dominant parameter is  $\xi_3$ . This may only hold, however, at this particular time instant. To illustrate the time-dependent behavior of the sensitivity indices, we plot in Fig. 12 the evolution of  $T_1, \dots, T_4$  for a single restriction factor,  $\alpha = 3$ . In this

**Table 4** Successively smaller intervals for the wind drag coefficient

$\alpha$	Wind drag coefficient range
1.00	[0.20, 1.00]
1.50	[0.33, 0.87]
2.00	[0.40, 0.80]
2.50	[0.44, 0.76]
3.00	[0.47, 0.73]
3.50	[0.49, 0.71]
4.00	[0.50, 0.70]



**Fig. 11** The decline of model sensitivity to  $\xi_4$  as the interval of wind drag coefficients shrinks (i.e.,  $\xi_4$  approaches zero). The indices  $T_i$  are as in Table 3



**Fig. 12** Model sensitivities over time. Interval for the stochastic wind drag coefficient (parameterize by  $\xi_4$ ) is shrunk with a factor of  $1/3$  (i.e.,  $\alpha = 3$ ). The indices  $T_i$  are as in Table 3

case, one observes that all random inputs eventually become significant contributors to the total variance of SSH, whereas the variability of SST and MLD becomes dominated by the background diffusivity.

The present exercise further illustrates the utility of the PC representations of the model outputs. In general, input uncertainties may not be known exactly, so it may be prudent to consider wide parameter ranges and then restrict these ranges as data and observations become available. Based on the corresponding PC representations, one can efficiently determine whether slight changes in parameter ranges, e.g., removing certain extreme parameter values, would have a material impact on the balance of model sensitivities. Alternatively, as illustrated above, one can assess under what ranges of uncertainty a specific parameter ceases to play a dominant role in the model variability. The present experiences also highlight the advantages of the global sensitivity analysis over the local sensitivity approach. Specifically, whereas large sensitivities may prevail in a given parameter range or in the neighborhood of a given selected parameter vector, these sensitivities may also drop rapidly as the parameter range is refined or as one moves away from the selected parameter vector. The ability to detect and quantify these situations without the need to generate additional model runs is one of the advantages of applying change of measure techniques to PC representations of model outputs.

### 5.1.3 Sensitivity regions

In Section 5.1.1, we provided a global sensitivity analysis at the buoy location. Here, we illustrate the spatial distribution of these global indices for the whole computational domain to provide additional insight into their impact on the predictions.

We have previously noted the interplay between  $\xi_3$  and  $\xi_4$  on the global sensitivity of SST. For brevity, we focus our attention on the spatial distribution of the corresponding indices.

Figures 13 and 14, respectively, depict spatial contours of  $T_3$  and  $T_4$ ; in both cases, plots are generated for  $t = 90$  h and  $t = 147$  h. One observes from Fig. 13 that at  $t = 90$  h,  $\xi_3$  is dominant in most of the gulf region but that its impact greatly diminishes with the passage of the hurricane, particularly around the path of the latter. As can be seen from Fig. 14,  $\xi_4$  becomes dominant in precisely the same regions. Thus, concerning the sensitivity of SST and for the present range of parameters, the interplay between  $\xi_3$  and  $\xi_4$  that was observed locally at the buoy location also occurs in most of the gulf. It is also interesting to note that the sensitivity

contours exhibit structures that are reminiscent of the evolution of ocean eddies. Similar observations have in fact been made in internal vortex dominated flows [23].

### 5.2 Local sensitivity

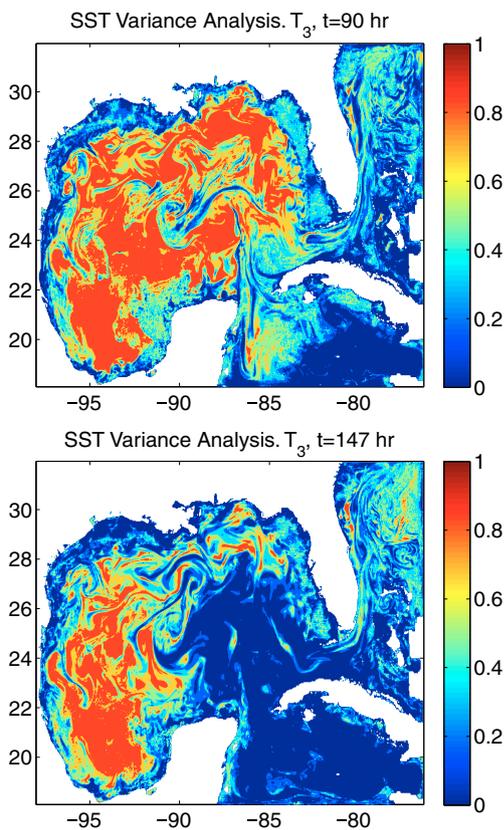
As noted earlier, one can readily estimate local sensitivities by differentiation of the PC expansion with respect to random variables. Letting  $U(\xi) = \sum_{k=0}^P U_k \Psi_k(\xi)$ , one can compute the partial derivatives of  $U$  with respect to  $\xi_i, i = 1, \dots, d$ , which results in:

$$\frac{\partial U}{\partial \xi_j} \Big|_{\xi} = \sum_{k=0}^P U_k \frac{\partial \Psi_k}{\partial \xi_j} \Big|_{\xi}.$$

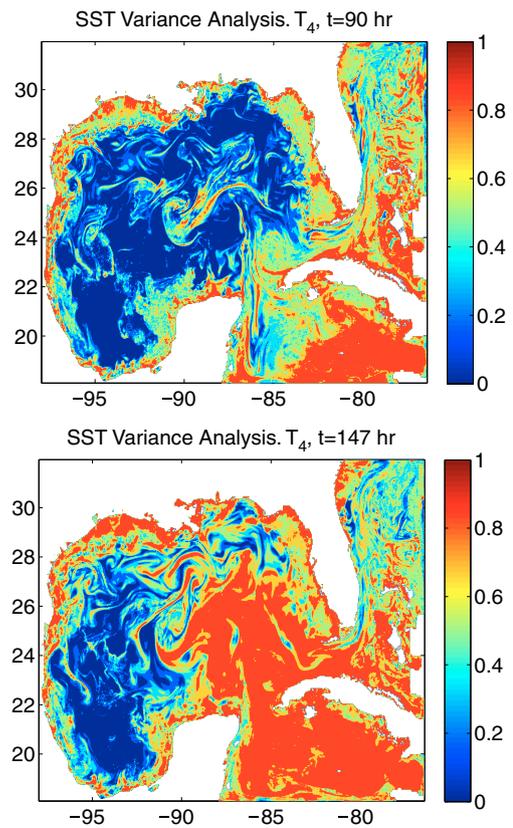
For each of the basis functions  $\Psi_k$ , which are  $d$ -variate polynomials defined in Eq. 5, we have

$$\frac{\partial \Psi_k}{\partial \xi_j} = \frac{\partial}{\partial \xi_j} \prod_{i=1}^d \psi_{\alpha_i^k}(\xi_i) = \frac{d\psi_{\alpha_j^k}}{d\xi_j} \prod_{i \neq j} \psi_{\alpha_i^k}.$$

Thus, with the multi-index construction, the computation is reduced to the differentiation of 1D polynomials.



**Fig. 13**  $T_3$  sensitivity contours for SST. Plots are generated for  $t = 90$  h (top) and  $t = 147$  h (bottom)

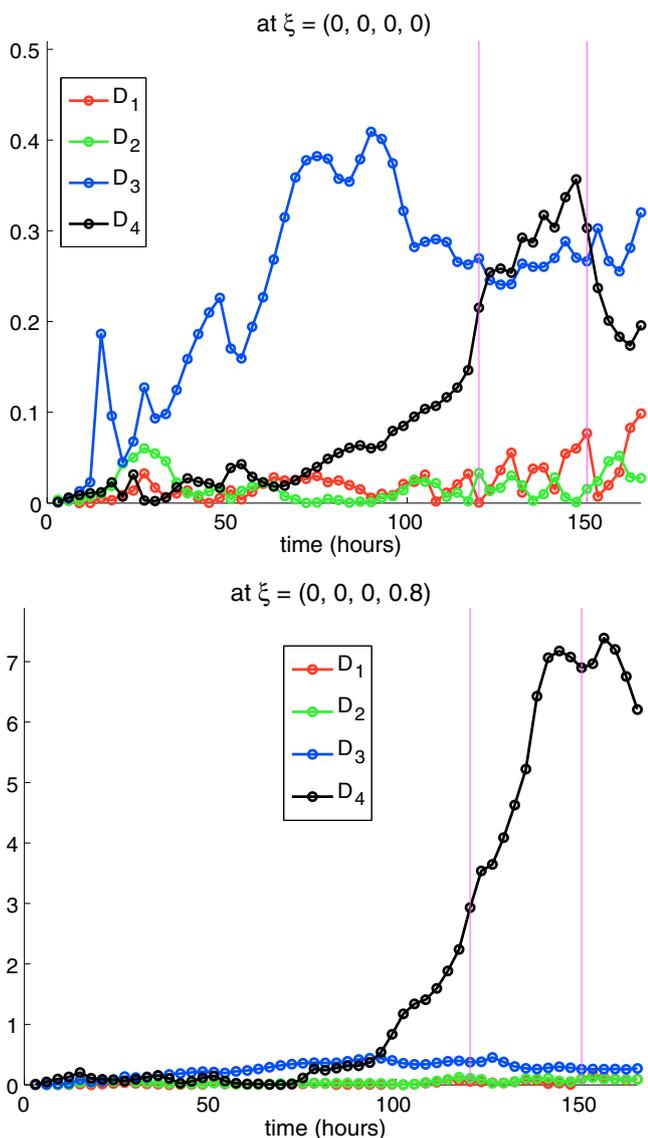


**Fig. 14**  $T_4$  sensitivity contours for SST. Plots are generated for  $t = 90$  h (top) and  $t = 147$  h (bottom)

As an illustration, we examine the local sensitivities of SST at the buoy location, specifically by computing the local indices,

$$D_j(\xi) = \left| \frac{\partial \text{SST}(t, \xi)}{\partial \xi_j} \right|, \quad j = 1, \dots, 4.$$

Figure 15 shows the evolution of  $D_i$  for  $\xi = (0, 0, 0, 0)$  and  $\xi = (0, 0, 0, 0.8)$ . The first point,  $\xi = \mathbf{0}$ , corresponds to the expected value of the random parameter vector, whereas the second point corresponds to a large value of the  $\xi_4$ . Thus, the results contrast two realizations, one corresponding to the mean value of the parameters and a second corresponding to a high drag coefficient. In the first case, the local sensitivity of the solution is initially dominated by the background diffusivity ( $\xi_3$ ) though both  $D_3$  and  $D_4$  (wind drag coefficient) become comparable as the hurricane approaches the buoy location. In contrast,  $\xi = (0, 0, 0, 0.8)$ , the wind drag coefficient rapidly becomes dominant, with its sensitivity factor achieving amplitudes several folds large than the remaining ones. The present results thus highlight the large variability that occurs with the selected parameter ranges and the utility of conducting a global



**Fig. 15** Evolution of the local sensitivity indices,  $D_i$ ,  $i = 1 \dots 4$ , at  $\xi = (0, 0, 0, 0)$  (top) and  $\xi = (0, 0, 0, 0.8)$  (bottom). The first vertical line indicates the time at which the hurricane enters the Gulf, whereas the second one indicates a time at which the hurricane is close to the buoy

assessment of the sensitivity of the solution to uncertain parameters.

### 6 Integral quantities

We now turn our attention to analyzing the global sensitivity of the predicted circulation to the uncertain model inputs by examining spatially averaged (integral) quantities. Our motivation is two-fold; on one hand, the behavior of integral quantities provides a meaningful and well-founded means of comparing predictions to

observations, as well as comparing the predictions of different models. On the other hand, the averaging that is inherent in the definition of integral quantities is often expected to lead to observables that exhibit a smoother dependence on uncertain inputs than local, time-varying signals. Accordingly, one can anticipate that integral QoIs would exhibit simpler distributions and thus admit simpler representations than instantaneous field variables. In this section, we briefly address this question in the context of the present ocean circulation database.

Following [15], we focus specifically on the average sea surface temperature in a region enclosing the hurricane track, as well as the average heat flux in a circular region around the center of the hurricane. Such QoIs are frequently used both to characterize the predicted circulation, as well as analyze the predictive skill of various models.

As in [15], a rectangular *control box* is selected for the purpose of defining a “regionally-averaged” SST. As depicted in Fig. 16, the hurricane track cuts through the control box. To obtain the mean sea surface temperature within this region, we start from corresponding PC representation,

$$SST(\mathbf{x}, t, \xi) \doteq \sum_{k=0}^P SST_k(\mathbf{x}, t) \Psi_k(\xi),$$

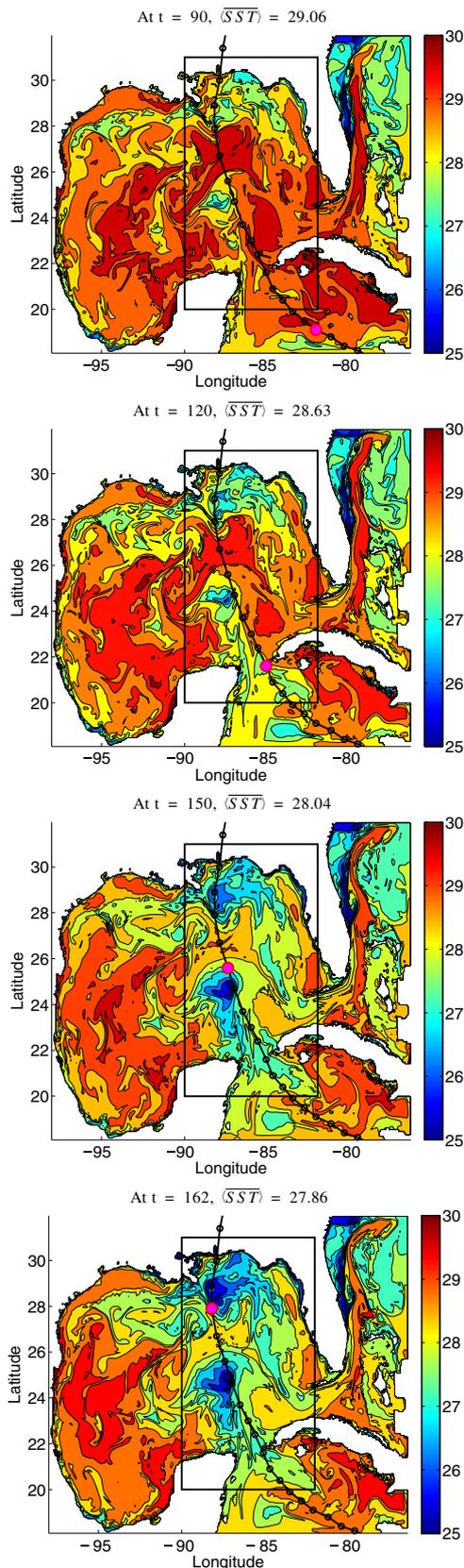
and estimate the box-averaged temperature,  $\overline{SST}$ , through:

$$\overline{SST}(t, \xi) = \frac{1}{N} \sum_{i=1}^N SST(\mathbf{x}_i, t, \xi),$$

where  $\mathbf{x}_i$ ,  $i = 1, \dots, N$  are the grid points lying within the control box. We derive a spectral representation for  $\overline{SST}$  by inserting the PC representation of SST into the above definition:

$$\begin{aligned} \overline{SST}(t, \xi) &= \frac{1}{N} \sum_{i=1}^N SST(\mathbf{x}_i, t, \xi) \\ &\doteq \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^P SST_k(\mathbf{x}_i, t) \Psi_k(\xi) \\ &= \sum_{k=0}^P \left( \frac{1}{N} \sum_{i=1}^N SST_k(\mathbf{x}_i, t) \right) \Psi_k(\xi). \end{aligned}$$

Thus, the spectral coefficients of  $\overline{SST}(t)$  are the averages of the corresponding coefficients of  $SST(\mathbf{x}, t)$



**Fig. 16** Contours of expected SST at selected time instants. The track of the hurricane and its instantaneous location (red dot) are also indicated. Also depicted is the boundary of the control box

over the control box. The mean of  $\overline{SST}$  is computed, respectively, through:

$$\langle \overline{SST} \rangle (t) = \overline{SST}_0(t) = \frac{1}{N} \sum_{i=1}^N SST_0(x_i, t),$$

or alternatively,

$$\langle \overline{SST} \rangle (t) = \overline{SST}_0(t) = \frac{1}{N} \sum_{i=1}^N \langle SST(x_i, t) \rangle.$$

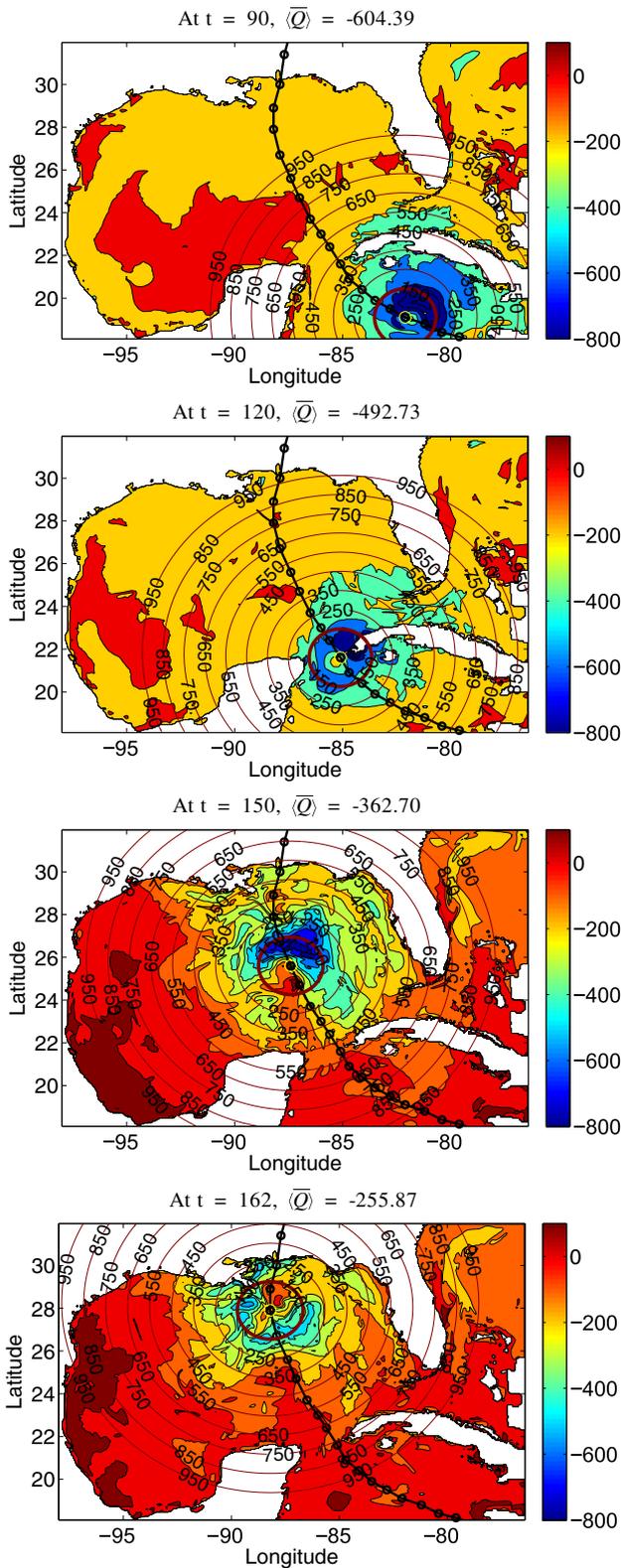
Similarly, the variance of the box-averaged temperature is readily estimated using:

$$\text{Var}(\overline{SST}) = \sum_{k=1}^P \overline{SST}_k^2(t) \langle \Psi_k^2 \rangle.$$

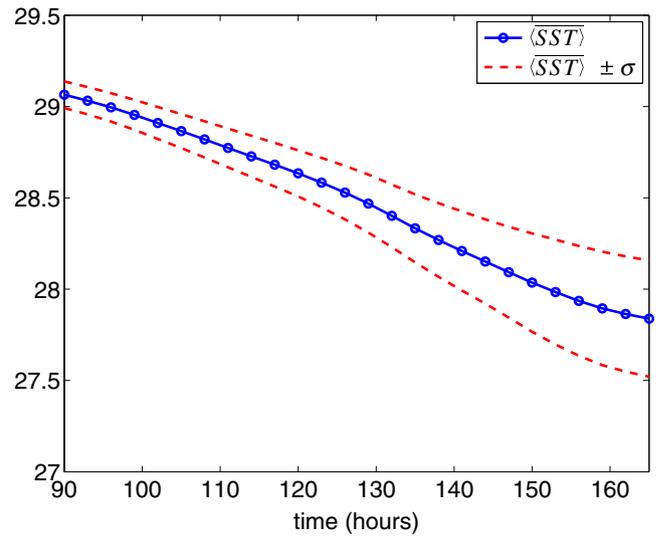
In addition to the mean temperature in the control box, we examine the heat flux in a circular region around the center of the hurricane, a quantity which is frequently used as a measure of potential strengthening or weakening of the hurricane [15]. Since our primary interest is the global sensitivity of the stochastic solution to the uncertain inputs, a simplified approach is adopted by integrating the heat flux over a circle of fixed radius,  $R_t = 150$  km, around the center of the hurricane. The resulting stochastic average heat flux,  $\overline{Q}$ , is function of time and of the germ  $\xi$ . An analogous methodology to that used for  $\overline{SST}$  is used in order to estimate its PC representation and consequently its statistical properties.

As previously discussed, we are primarily interested in characterizing the mean, variance, and parametric sensitivities for the averaged SST and heat flux. Before proceeding to this analysis, we illustrate in Figs. 16 and 17 the evolution of the mean temperature and of the mean heat flux. The figures show instantaneous contours of the mean temperature field and of the mean heat flux, respectively, and depict regions over which the spatial averages are obtained.

Figures 18 and 19 show the evolution of the  $\langle \overline{SST} \rangle$  and  $\langle \overline{Q} \rangle$ . Also plotted in both cases are curves that lie 1 standard deviation away from the mean prediction. Consistent with [15], the passage of the hurricane leads to a monotonic decrease in the average temperature in the control box, and this behavior is consistent with evolution of the mean temperature field shown in Fig. 16. This is accompanied by a monotonic increase in the standard deviation. The mean and standard deviation of  $\overline{Q}$  generally increase over the time interval depicted, though their behavior is clearly not monotonic.

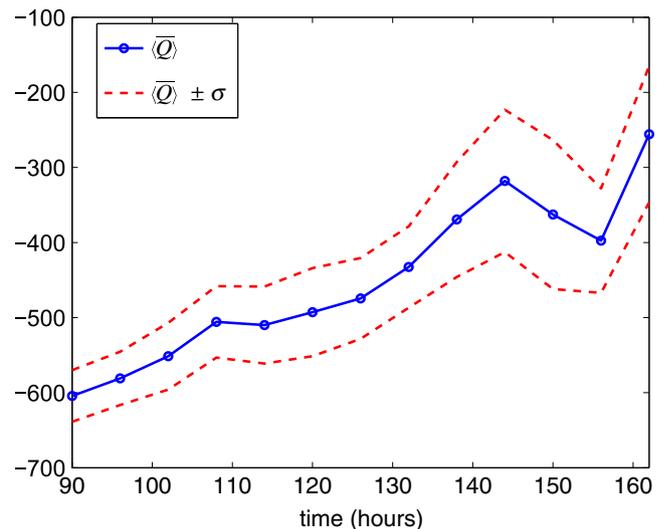


**Fig. 17** Contours of the mean heat flux at selected times as indicated. The concentric circular lines denote the distance (in kilometers) from the center of the hurricane. The thicker line denotes the boundary of the circle of 150-km radius, centered over the eye of the hurricane, over which the heat flux is averaged. The average value is shown in the label

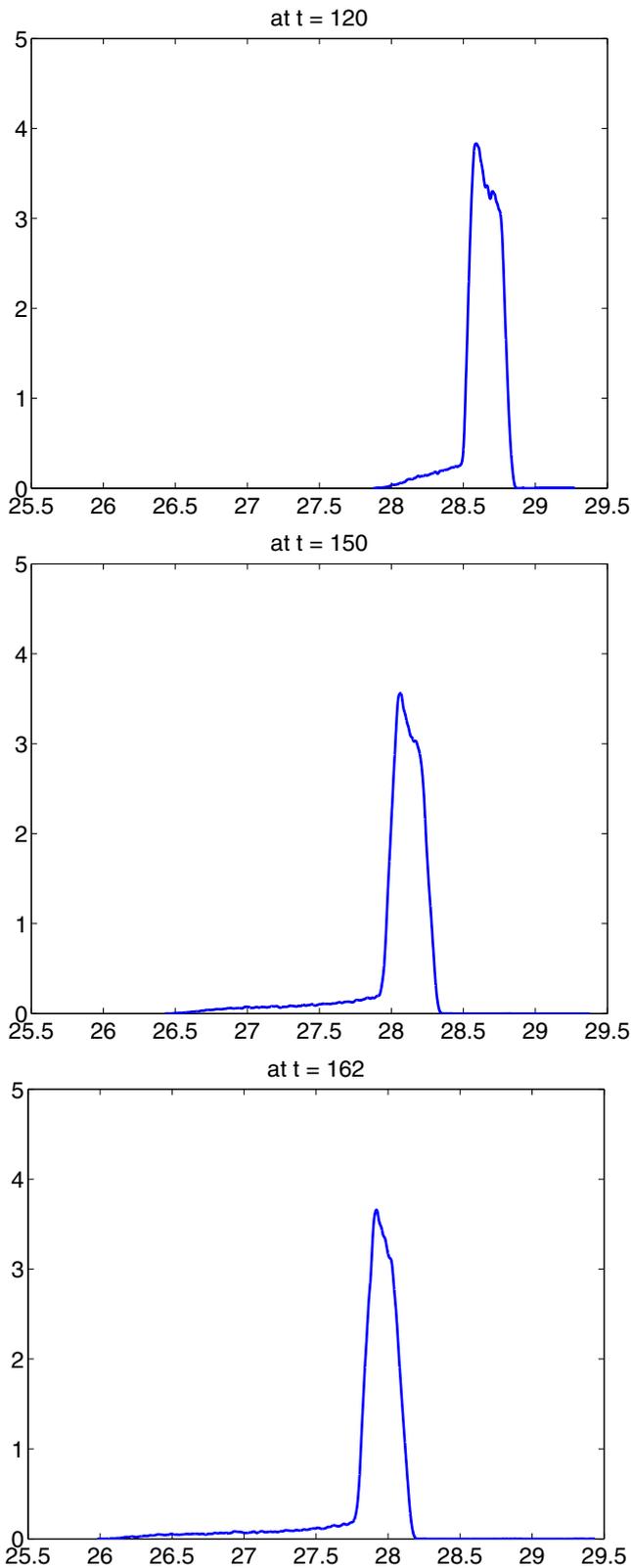


**Fig. 18** Average expected SST in the control box,  $\langle SST \rangle$ , over time with 1 standard deviation bounds. Here we can see the average cooling in the control box over time

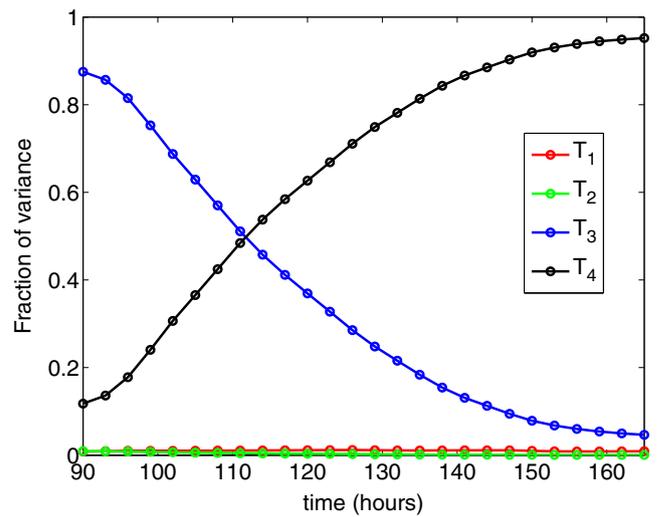
To further examine the behavior of the solution, we plot in Fig. 20 instantaneous PDFs of the box-averaged temperature,  $\overline{SST}$ . The PDFs are seen to gradually shift toward the left, consistent with the monotonic decrease of the mean shown in Fig. 18. Note that long tails extending toward lower temperature develop in these distributions and that these tails become more pronounced as time evolves. We have also examined the PDFs of the average heat flux  $\overline{Q}$ , which were found to exhibit similar trends to the PDFs of  $\overline{SST}$ . However, in this case, the PDFs shift to the right as time evolves,



**Fig. 19** Average expected heat flux in  $R_r$ ,  $\langle \overline{Q} \rangle$ , over time, with 1 standard deviation bounds



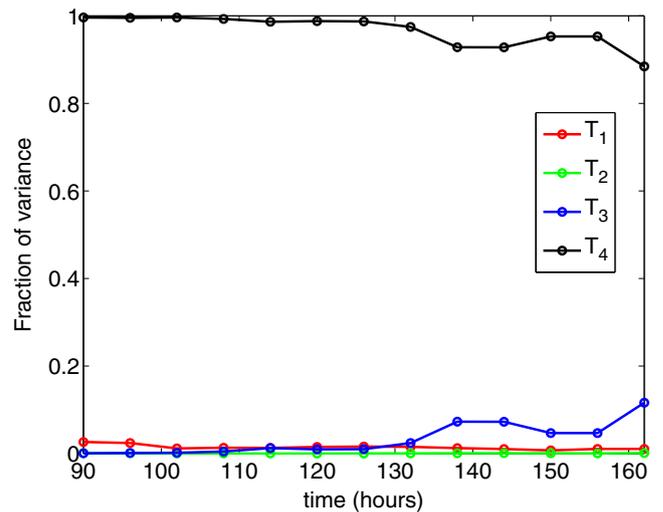
**Fig. 20** Distributions of  $\overline{SST}$ , at selected times



**Fig. 21** Global sensitivity indices for  $\overline{SST}$  over time. The sensitivity indices  $T_i$  are as in Table 3

and long tails extend toward higher heat flux values. For brevity, these results are omitted.

The observed skewed PDFs with long tails in integral quantities are somewhat unexpected because the smoothing associated with spatial averaging is typically expected to lead to simpler distribution functions. Evidently, the present experiences indicate that this may not always hold true. Specifically, in the present case, the PDFs of the box-averaged temperature (Fig. 20) exhibit essentially the same features as those determined locally (Fig. 8). This highlights a rather complex response of the solution to the uncertain parameters and



**Fig. 22** Global sensitivity indices for  $\overline{Q}$  over time. The sensitivity indices  $T_i$  are as in Table 3. Note that here the stochastic wind drag coefficient dominates the variability in  $\overline{Q}$  because  $\overline{Q}$  is always computed around the center of the hurricane

suggests that in realistic scenarios, simplified stochastic representations may not be readily applicable.

Figures 21 and 22 show the evolution of the global sensitivity indices for  $\overline{SST}$  and  $\overline{Q}$ , respectively. Consistent with previous observations, the results indicate that the wind drag coefficient dominates the variability in the average heat flux,  $\overline{Q}$ . Meanwhile, consistent with previous observations, the variance in  $\overline{SST}$  is dominated by the background diffusivity in the early stages of the computation. However, the effect of the wind drag coefficient becomes clearly dominant when the hurricane enters the control region.

## 7 Conclusions

A sparse spectral projection approach was implemented to propagate and quantify parametric uncertainties in an OGCM. The simulations focused on the impact of the subgrid mixing parameters and wind drag coefficient on the circulation in the GOM. High-resolution HYCOM simulations, forced by hurricane Ivan winds, were used for this purpose.

A non-intrusive spectral projection scheme, based on a Smolyak sparse quadrature grid, was used to derive the PC representation of the stochastic response of selected QoIs. A brief numerical study was initially conducted to analyze the impact of the sparse grid refinement. For the presently considered conditions and uncertain parameter ranges, a level 5 sparse quadrature was found to be sufficient for adequate representation of the model response. This amounted to the construction of a sparse database consisting of 385 independent HYCOM realizations.

Computations were then used to demonstrate the application of the resulting PC representation to estimate various statistical quantities. These included the analysis of the behavior of means and variances of both field and integral quantities, as well as the use of spectral representation to efficiently estimate probabilities of model variables being in a given interval. Also illustrated was the use of the PC representation as a surrogate model, namely through the implementation of simplified change of measure technique which analyzed the effect of restricting the range of one of the uncertain parameters.

Attention was focused on assessing the global sensitivity of the solution to the uncertain inputs, namely through the evaluation of Sobol indices. In particular, the analysis considered the average temperature in a region enclosing the track of the hurricane, as well as the heat flux in a circular region around its center.

Computations showed that for the conditions of the experiments, the variability in the mean temperature becomes dominated by the uncertainty in wind drag coefficient as the hurricane enters the control region. The uncertainty in wind drag coefficient was also found to be a dominant factor in the variability in the integral heat flux.

The analysis also indicated a complex model response, which featured the generation of skewed distribution functions with extended tails. These were observed for both local quantities, including the SST, as well as integral QoIs, including the regionally averaged SST and the integral heat flux around the center of the hurricane. This points to the need for generally conducting a systematic sampling of the random parameter space with sufficient resolution to capture key statistical features of the solution.

In follow-up work, we will focus on extending the present methodology by incorporating adaptive sampling schemes, which hold the promise of minimizing the computational cost required for an adequate representation of uncertainty in selected QoIs. In addition, we will also focus on exploiting PC representations as model surrogates, particularly in a Bayesian framework for data assimilation and parameter inference.

**Acknowledgement** This research was supported by the Office of Naval Research, Award N00014-101-0498.

## References

1. Babuška, I., Nobile, F., Tempone, R.: A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.* **45**(3), 1005–1034 (2007)
2. Berveiller, M., Sudret, B., Lemaire, M.: Stochastic finite element: a non intrusive approach by regression. *Eur. J. Comput. Mech.* **15**, 81–92 (2006)
3. Bleck, R.: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Model.* **4**, 55–88 (2002)
4. Chassignet, E., Smith, L., Halliwell, G., Bleck, R.: North Atlantic simulation with the Hybrid Coordinate Ocean Model (HYCOM): impact of the vertical coordinate choice, reference density, and thermobaricity. *J. Phys. Oceanogr.* **33**, 2504–2526 (2003)
5. Chassignet, E.P., Hurlburt, H.E., Smedstad, O.M., Halliwell, G.R., Hogan, P.J., Wallcraft, A.J., Baraille, R., Bleck, R.: The HYCOM (hybrid coordinate ocean model) data assimilative system. *J. Mar. Syst.* **65**(1–4), 60–83 (2007)
6. Crestaux, T., Maitre, O.L., Martinez, J.M.: Polynomial chaos expansion for sensitivity analysis. *Reliab. Eng. Syst. Saf.* **94**(7), 1161–1172 (2009, Special Issue on Sensitivity Analysis)
7. Debusschere, B., Najm, H., Pébay, P., Knio, O., Ghanem, R., Le Maître, O.: Numerical challenges in the use of polynomial chaos representations for stochastic processes. *J. Sci. Comput.* **26**(2), 698–719 (2004)

8. Donelan, M.A., Haus, B.K., Reul, N., Plant, W.J., Stiassnie, M., Graber, H.C., Brown, O.B., Saltzman, E.S.: On the limiting aerodynamic roughness of the ocean in very strong winds. *Geophys. Res. Lett.* **31**(L18306), 1–5 (2004). doi:[10.1029/2004GL019460](https://doi.org/10.1029/2004GL019460)
9. French, J.R., Drennan, W.M., Zhang, J.A., Black, P.G.: Turbulent fluxes in the hurricane boundary layer. Part I: momentum flux. *J. Atmos. Sci.* **64**, 1089–1102 (2007). doi:[10.1175/JAS3887.1](https://doi.org/10.1175/JAS3887.1)
10. Gerstner, T., Griebel, M.: Numerical integration using sparse grids. *Numer. Algorithms* **18**, 209–232 (1998)
11. Ghanem, R., Spanos, P.: A spectral stochastic finite element formulation for reliability analysis. *J. Eng. Mech. ASCE* **117**, 2351–2372 (1991)
12. Ghanem, R., Spanos, P.: *Stochastic Finite Elements: A Spectral Approach*, 2nd edn. Dover, New York (2002)
13. Haidvogel, D.B., Arango, H.G., Budgell, W., Cornuelle, B.D., Curchitser, E., Lorenzo, E.D., Fennel, K., Geyer, W., Hermann, A.J., Lanerolle, L., Levin, J., McWilliams, J.C., Miller, A.J., Moore, A.M., Powell, T., Shchepetkin, A.F., Sherwood, C., Signell, R.P., Warner, J.C., Wilkin, J.: Ocean forecasting in terrain-following coordinates: formulation and skill assessment of the regional ocean modeling system. *J. Comput. Phys.* **227**(7), 3595–3624 (2008). doi:[10.1016/j.jcp.2007.06.016](https://doi.org/10.1016/j.jcp.2007.06.016). <http://www.sciencedirect.com/science/article/pii/S0021999107002549>. Predicting weather, climate and extreme events
14. Halliwell, G.: Evaluation of the vertical coordinate and vertical mixing algorithms in the Hybrid-Coordinate Ocean Model (HYCOM). *Ocean Model.* **7**, 285–322 (2004). doi:[10.1016/j.ocemod.2003.10.002](https://doi.org/10.1016/j.ocemod.2003.10.002)
15. Halliwell, G., Shay, L., Brewster, J., Teague, W.: Evaluation and sensitivity analysis of an ocean model response to Hurricane Ivan. *Mon. Weather Rev.* **139**, 921–945 (2011). doi:[10.1175/2010MWR3104.1](https://doi.org/10.1175/2010MWR3104.1)
16. Hodur, R.: The naval research laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Weather Rev.* **125**(7), 1414–1430 (1997). doi:[10.1175/1520-0493\(1997\)125<1414:TNRLSC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1414:TNRLSC>2.0.CO;2)
17. Homma, T., Saltelli, A.: Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* **52**(1), 1–17 (1996)
18. Kallenberg, O.: *Foundations of Modern Probability*, 2nd edn. Springer, Berlin (2002)
19. Kara, A.B., Rochford, P.A., Hurlburt, H.E.: Efficient and accurate bulk parameterizations of air–sea fluxes for use in general circulation models. *J. Atmos. Ocean. Technol.* **17**(10), 1421–1438 (2000). doi:[10.1175/1520-0426](https://doi.org/10.1175/1520-0426)
20. Large, W.G., McWilliams, J.C., Doney, S.C.: Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.* **32**, 363–403 (1994)
21. Large, W.G., Pond, S.: Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr.* **11**, 324–336 (1981). doi:[10.1175/1520-0485](https://doi.org/10.1175/1520-0485)
22. Le Maître, O., Knio, O.: *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*. Scientific Computation. Springer, Berlin (2010)
23. Le Maître, O., Knio, O., Najm, H., Ghanem, R.: A stochastic projection method for fluid flow. I. Basic formulation. *J. Comput. Phys.* **173**, 481–511 (2001)
24. Le Maître, O., Najm, H., Pébay, P., Ghanem, R., Knio, O.: Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM J. Sci. Comput.* **29**(2), 864–889 (2007)
25. Le Maître, O., Reagan, M., Najm, H., Ghanem, R., Knio, O.: A stochastic projection method for fluid flow. II. Random process. *J. Comput. Phys.* **181**, 9–44 (2002)
26. Li, X.J., Chao, Y., McWilliams, J.C., Fu, L.L.: A comparison of two vertical-mixing schemes in a Pacific Ocean general circulation model. *J. Climate* **14**(7), 1377–1398 (2001). doi:[10.1175/1520-0442\(2001\)014<1377:ACOTVM>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<1377:ACOTVM>2.0.CO;2)
27. Marshall, J., Adcroft, A., Hill, C., Perelman, L., Heisey, C.: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.* **102**, 5753–5766 (1997)
28. Mathelin, L., Hussaini, M.: A stochastic collocation algorithm for uncertainty analysis. Tech. Rep. NASA/CR-2003-212153, NASA Langley Research Center (2003)
29. Munk, W., Wunsch, C.: Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res., Part 1, Oceanogr. Res. Pap.* **45**(12), 1977–2010 (1998). doi:[10.1016/S0967-0637\(98\)00070-3](https://doi.org/10.1016/S0967-0637(98)00070-3). <http://www.sciencedirect.com/science/article/pii/S0967063798000703>
30. Najm, H.: Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annu. Rev. Fluid Mech.* **41**, 35–52 (2009)
31. Najm, H., Debusschere, B., Marzouk, Y., Widmer, S., Le Maître, O.: Uncertainty quantification in chemical systems. *Int. J. Numer. Methods Eng.* **80**(6), 789–814 (2009)
32. Petras, K.: On the Smolyak cubature error for analytic functions. *Adv. Comput. Math.* **12**, 71–93 (2000)
33. Petras, K.: Fast calculation in the Smolyak algorithm. *Numer. Algorithms* **26**, 93–109 (2001)
34. Petras, K.: Smolyak cubature of given polynomial degree with few nodes for increasing dimension. *Numer. Math.* **93**(4), 729–753 (2003). doi:[10.1007/s002110200401](https://doi.org/10.1007/s002110200401)
35. Phenix, B., Dinero, J., Tatang, M., Tester, J., Howard, J., McRae, G.: Incorporation of parametric uncertainty into complex kinetic mechanisms: application to hydrogen oxidation in supercritical water. *Combust. Flame* **112**, 132–146 (1998)
36. Powell, M.D., Vickery, P.J., Reinhold, T.A.: Reduced drag coefficient for high wind speeds in tropical cyclones. *Nature* **422**, 279–283 (2003). doi:[10.1038/nature01481](https://doi.org/10.1038/nature01481). Wind measurement and wind stress calculation in (observed) hurricane conditions
37. Prasad, T.G., Hogan, P.J.: Upper-ocean response to hurricane Ivan in a 1/25° nested Gulf of Mexico HYCOM. *J. Geophys. Res.* **112**, C04013 (2007)
38. Reagan, M., Najm, H., Ghanem, R., Knio, O.: Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection. *Combust. Flame* **132**, 545–555 (2003)
39. Reagan, M.T., Najm, H.N., Pbay, P.P., Knio, O.M., Ghanem, R.G.: Quantifying uncertainty in chemical systems modeling. *Int. J. Chem. Kinet.* **37**(6), 368–382 (2005)
40. Shchepetkin, A.F., McWilliams, J.C.: The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Model.* **9**(4), 3090 (2005). doi:[10.1016/j.ocemod.2004.07.001](https://doi.org/10.1016/j.ocemod.2004.07.001)
41. Smolyak, S.: Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR* **4**, 240–243 (1963)
42. Sobol', I.M.: Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exper.* **1**(4), 407–414 (1993)
43. Sobol', I.M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55**(1–3), 271–280 (2001)

44. Sudret, B.: Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.* **93**(7), 964–979 (2008). doi:[10.1016/j.ress.2007.04.002](https://doi.org/10.1016/j.ress.2007.04.002). <http://www.sciencedirect.com/science/article/B6V4T-4NKJ0BN-1/2/d3554add11e0ff4b9fd1b69c06cdb16>. Bayesian networks in dependability
45. Thacker, W., Srinivasan, A., Iskandarani, M., Knio, O., Henaff, M.L.: Propagating boundary uncertainties using polynomial expansions. *Ocean Model.* (2011). doi:[10.1016/j.ocemod.2011.11.011](https://doi.org/10.1016/j.ocemod.2011.11.011)
46. Williams, D.: *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge (1991)
47. Xiu, D., Hesthaven, J.: High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.* **27**(3), 1118–1139 (2005)
48. Xiu, D., Karniadakis, G.: The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **24**, 619–644 (2002)
49. Xiu, D., Karniadakis, G.: Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.* **187**, 137–167 (2003)
50. Zamudio, L., Hogan, P.J.: Nesting the Gulf of Mexico in Atlantic HYCOM: oceanographic processes generated by hurricane Ivan. *Ocean Model.* **21**(3–4), 106–125 (2008)