Polynomial Chaos–Based Bayesian Inference of K-Profile Parameterization in a General Circulation Model of the Tropical Pacific

IHAB SRAJ
Duke University, Durham, North Carolina, and King Abdullah University for Science and Technology, Thuwal, Saudi Arabia

SARAH E. ZEDLER
The University of Texas at Austin, Austin, Texas

OMAR M. KNIO
Duke University, Durham, North Carolina, and King Abdullah University for Science and Technology, Thuwal, Saudi Arabia

CHARLES S. JACKSON
The University of Texas at Austin, Austin, Texas

IBRAHIM HOTEIT
King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

(Manuscript received 10 November 2015, in final form 15 August 2016)

ABSTRACT
The authors present a polynomial chaos (PC)–based Bayesian inference method for quantifying the uncertainties of the K-profile parameterization (KPP) within the MIT general circulation model (MITgcm) of the tropical Pacific. The inference of the uncertain parameters is based on a Markov chain Monte Carlo (MCMC) scheme that utilizes a newly formulated test statistic taking into account the different components representing the structures of turbulent mixing on both daily and seasonal time scales in addition to the data quality, and filters for the effects of parameter perturbations over those as a result of changes in the wind. To avoid the prohibitive computational cost of integrating the MITgcm model at each MCMC iteration, a surrogate model for the test statistic using the PC method is built. Because of the noise in the model predictions, a basis-pursuit-denoising (BPDN) compressed sensing approach is employed to determine the PC coefficients of a representative surrogate model. The PC surrogate is then used to evaluate the test statistic in the MCMC step for sampling the posterior of the uncertain parameters. Results of the posteriors indicate good agreement with the default values for two parameters of the KPP model, namely the critical bulk and gradient Richardson numbers; while the posteriors of the remaining parameters were barely informative.

1. Introduction
The present work seeks to calibrate the model parameters of the K-profile parameterization (KPP) model (Large et al. 1994, 1997) as implemented in the ocean model MIT general circulation model (MITgcm; Ferreira and Marshall 2006; Marshall et al. 1997a) of the tropical Pacific. The KPP model relies on a number of parameters whose default values are set based on a combination of theory, laboratory experiments, and atmospheric–oceanic boundary layer observations (Large et al. 1994, 1997). Our goal here is to quantify the uncertainties in these parameters where the ocean resolves the chaotic behavior of fluid dynamic models. The chaos in the response of the deterministic MITgcm model to the perturbation of these...
parameters leads to internal noise that in turn results in low signal-to-noise challenges, as will be discussed in more detail below.

An inverse modeling approach is adopted for the objective stated above in which a set of temperature, salinity, and horizontal current measurements are used to estimate the KPP parameters. Specifically, we employ a Bayesian approach to inverse problems that provides complete posterior statistics and not just a single value for the quantity of interest (Tarantola 2005). Traditionally, local model-data misfit of short-term turbulent mixing events is used to construct a cost function and then Bayesian inference is employed for the estimation of the uncertain parameters (Zedler et al. 2012; Sraj et al. 2014b). Here, instead, we use a test statistic for KPP parameter estimation that was introduced in Wagman et al. (2014) and in Zedler et al. (2016, manuscript submitted to Ocean Modell.). This statistic seeks to formulate a total cost function of different components representing the structures of turbulent mixing on both daily and multiyear time scales, takes data quality into account, and filters for the effects of parameter perturbations over those resulting from changes in the wind. At both time scales, the model and data are filtered before taking differences to capture the observations that are relevant to testing mixing physics parameters.

The end result of the Bayesian inference formulation is a multidimensional posterior (Malinverno 2002; Sraj et al. 2013; Olson et al. 2012) that can be directly sampled via a Markov chain Monte Carlo (MCMC) scheme. This, however, requires a prohibitive number of simulations of the forward model—one for every proposed set of parameters of the Markov chain. This practice renders Bayesian methods computationally prohibitive for large-scale models such as the MITgcm, where one model evaluation takes 22 h in computing time using 256 processors. To overcome this issue, we construct a surrogate model that approximates the forward model and can be used in the sampling MCMC. More precisely, we use the polynomial chaos (PC) method to construct the surrogate model from an ensemble of MITgcm model runs (Marzouk et al. 2007; Marzouk and Najm 2009). This approach further offers additional advantages such as computing model output sensitivities and additional statistics (Le Maitre and Knio 2010).

The PC method has been extensively investigated in the literature, and its suitability for large-scale models has been recently demonstrated in various settings. Alexanderian et al. (2012) implemented a sparse spectral projection PC approach to propagate parametric uncertainties of three KPP parameters in addition to the wind drag coefficient during a hurricane event. The study demonstrated the possibility of building a representative surrogate model for a realistic ocean model; however, the inverse problem was not tackled. Winokur et al. (2013) followed up on Alexanderian et al.’s (2012) work and implemented an adaptive strategy to design sparse ensembles of oceanic simulations for the purpose of constructing a PC surrogate with even less computational effort. Sraj et al. (2013) extended the previous work and combined a spectral projection PC approach with Bayesian inference to estimate the parameterized wind drag coefficient using temperature data collected during a typhoon event. The same problem was also solved using a gradient-based search method (Sraj et al. 2014a). Mattern et al. (2012) have similarly exploited the virtues of such polynomial expansions for examining the response of ecosystem models to finite perturbations of their uncertain parameters. Tsunami (Sraj et al. 2014b; Ge and Cheung 2011) and subsurface flow modeling (Elsheikh et al. 2014) were also investigated using similar PC approaches.

What is common in the aforementioned PC applications is that the processes studied occurred over short time scales of a few days, so that the internal noise within the model was small. This enabled a successful PC expansion construction using traditional spectral projection (Sraj et al. 2013; Reagan et al. 2003; Alexanderian et al. 2012). In this study, however, the major hurdle of constructing a PC surrogate model was the internal noise present in the MITgcm model due to the perturbation of the chosen KPP parameters, which was amplified over time by nonlinear interactions. As will be explained in section 4b(1), a spectral projection technique failed to construct a PC expansion that faithfully represents the model. Instead, we resort to a compressed sensing technique, namely, the basis-pursuit-denoising (BPDN) method (Peng et al. 2014), to determine the PC expansion coefficients. This technique first seeks to estimate the noise inherent in the signal and then solve approximately an optimization problem assuming sparsity in the PC expansion to determine the nonzero PC coefficients. The BPDN method offers an additional advantage in that a smaller number of model runs compared to the spectral projection method is required to determine the PC coefficients, as shown in section 4c. BPDN was recently employed to build a proxy model for an integral oil–gas plume model (Wang et al. 2016) and for an ocean model with initial and wind forcing uncertainties (Li et al. 2015). In the former, the model output was noisy as a result of the iterative solver used in the double-plume calculation (Socolofsky et al. 2008) and BPDN proved to be successful in building a representative PC surrogate that excludes the model output noise. In the latter, no noise was present in the model output; however, the model was unable to produce realistic simulations for prespecified
sets of parameters, as required by the spectral projection method. BPDN was therefore used as an alternative approach, as it does not have this requirement; a random sample of model parameters can be modeled instead (Doostan and Owhadi 2011).

We end the introduction section by asserting that the employed compressed sensing technique is not new but also is not a standard in large-scale applications. To our knowledge, BPDN has never been applied to a noisy, large-scale ocean model. Implementing this technique together with the choice of fitting a newly formulated test statistic (Wagman et al. 2014; Zedler et al. 2016, manuscript submitted to Ocean Modell.) for the purpose of inferring KPP model parameters represents the novelty of our current work.

The structure of the paper is as follows. Section 2 describes the MITgcm, presents the choice of the uncertain parameters, and describes the observation and cost function used in the estimation process. Section 3 introduces the Bayesian inference and its application to our specific problem. Section 4 discusses the PC method and presents several error studies to show the convergence of the constructed PC expansion. Section 5 presents the results of the inference of KPP parameters and section 6 summarizes our findings.

2. Model, uncertain parameters, and observations

a. MITgcm model

The MITgcm employed in this work is based on the primitive Navier–Stokes equations implemented in spherical coordinates with an implicit nonlinear free surface. The MITgcm implements the KPP turbulent mixing scheme (Adcroft 1995; Marshall et al. 1997a,b; Large et al. 1994) (see below) in a regional configuration based on that of Hoteit et al. (2008, 2010) for the simulation of oceanic flow. In particular, the domain chosen covers the region with latitudes from 26°S to 30°N and longitudes from 104°E to 70°W (Fig. 1). The time period of our model simulation is 2004–07. The initial and lateral boundary conditions are provided by the Ocean Comprehensible Atlas (OCCA) reanalysis that was developed for the 2004–07 time period using data assimilation in the MITgcm of available temperature and salinity ocean datasets (Forget 2010). The lateral boundary conditions for our model are implemented with a sponge layer (with a thickness of nine grid cells, and inner and outer boundary relaxation time scales of 20 and 1 days, respectively). For the lateral boundary conditions, the OCCA data assimilation product is interpolated at the model resolution of 1/8° and with a time step of 1 day. Therefore, at the initial time step, the boundary temperature and salinity conditions are approximately in equilibrium with the interior fields. Once our higher-resolution simulation starts, the velocity field quickly adjusts to the pressure gradient forces and establishes a realistic equatorial circulation. We note that our model runs on 256 processors and takes about 22 h for a single simulation. As described below in section 4, we needed to run the model 903 times, which required a total of about 5.5 million compute hours.

b. KPP model

In the ocean, turbulent mixing can ensue when there is net heat released to the atmosphere at the sea surface (i.e., at night), producing gravitationally unstable density inversions (convective mixing) and when there is sufficient vorticity-producing (on the x–z plane) vertical shear in the horizontal currents to overturn a nominally stratified water column (shear driven or Kelvin–Helmholz instability induced mixing). In general terms, the intensity of convective and shear-driven mixing depend on local water column properties and surface forcing conditions. Theoretically, the most vigorous turbulent mixing should occur when a weakly unstably stratified, strongly sheared flow is forced with strong winds and convection (i.e., at night). By contrast, the flow is most likely to be laminar when a strongly stably stratified, weakly sheared flow is forced with weak winds and large net heat going into the ocean (i.e., during the day). The KPP embodies these basic relationships, by making the intensity of mixing a function of the locally diagnosed properties of the water column that relate to the amenability to turbulent mixing (such as the bulk and gradient Richardson numbers), as well as the nonlocal surface wind stress and net heat flux forcing (through forcing parameters such as the friction velocity of wind and the Monin–Obukhov length). In the KPP, mixing is more intense under unstable convective surface forcing conditions. Readers interested in the details of the KPP are referred to Large et al. (1994) and Large and Gent (1999).

The KPP generates depth profiles of two quantities that are relevant for turbulent mixing, the eddy diffusivity/viscosity and a nonlocal term. The eddy diffusivity...
Table 1. List of KPP parameters to be estimated using Bayesian inference along with their MITgcm default values and assumed uniform prior.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter name</th>
<th>Symbol</th>
<th>Default value</th>
<th>Uniform prior [a, b]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>Critical bulk Richardson number</td>
<td>$R_i$</td>
<td>0.3</td>
<td>[0.1, 1.0]</td>
</tr>
<tr>
<td>$p_2$</td>
<td>Critical gradient Richardson number</td>
<td>$R_i^g$</td>
<td>0.7</td>
<td>[0.1, 1.0]</td>
</tr>
<tr>
<td>$p_3$</td>
<td>Structure function, unstable forcing, momentum</td>
<td>$\phi_{m,\text{unst}}$</td>
<td>16</td>
<td>[3.60, 331.06]</td>
</tr>
<tr>
<td>$p_4$</td>
<td>Structure function, unstable forcing, tracer</td>
<td>$\phi_{t,\text{unst}}$</td>
<td>16</td>
<td>[7.77, 67.02]</td>
</tr>
<tr>
<td>$p_5$</td>
<td>Nonlocal transport</td>
<td>$C^*$</td>
<td>10.0</td>
<td>[5.0, 15.0]</td>
</tr>
</tbody>
</table>

and viscosity at depth $z$ can be thought of as a scale of the intensity of the turbulent mixing there, with larger values indicating more vigorous turbulence. The role of the nonlocal term is to enhance the turbulent fluxes of temperature and salinity (but not the horizontal velocity components) under convective forcing conditions.

There are nine parameters in the KPP that pertain to convective or shear-driven mixing. A list of the five uncertain parameters in the KPP model under investigation in this work is presented in Table 1 along with the default values used in MITgcm for each parameter. The critical bulk and gradient Richardson ($R_i$ and $R_i^g$, respectively) numbers relate directly to local water column shear/stratification properties, with larger values generally making turbulent mixing more intense (for the same water column). Convective mixing parameter $\phi_{r,\text{unst}}$ and $\phi_{m,\text{unst}}$ depend directly on a combination of surface forcing and local water column shear/stratification considerations and are zero under stable forcing conditions (during the day). Increasing their value generally makes convection more vigorous (given the same water column properties and surface forcing). The nonlocal convective mixing parameter $C^*$ is proportional to the nonlocal convective term, so increasing it will make the turbulent fluxes for temperature and salinity larger. For all five parameters, we have assumed uniform priors where the corresponding bounds have been chosen to cover the default values with a window around them that is physically reasonable, as indicated in Table 1.

c. Observations and test statistic

The observational data for our experiment come from the TOGA TAO mooring array for the November 2003–November 2007 time period. The array consists of 77 moorings (Fig. 1) that are centered on the equator and that span the width of the tropical Pacific in the east–west direction, in the latitude range from 8°S to 8°N (McPhaden et al. 1998). The data used included measurements of temperature, salinity, and horizontal current components.

We use a combination of test statistics on these observations that have components that operate on daily (Wagman et al. 2014) and multiyear time scales (Zedler et al. 2016, manuscript submitted to Ocean Modell.). The test statistic for daily time scales $E(p)_R$ measures the model’s ability to reproduce the observed maximum lagged correlation between 2- and 6-day bandpassed wind stress and sea surface temperatures (Wagman et al. 2014). In particular, the test statistic includes a sum of errors at all $n$ buoy locations between the observed maximum correlation $R$ and the modeled correlation $r(p)$ at observed lag $L$ and modeled lag $l(p)$. The correlation test statistic takes on the form

$$E(p)_R = \frac{1}{2} \sum_{i=1}^{n} \left( \frac{(R_i - r_i(p))^2}{\sigma_i^2} + \frac{(L_i - l_i(p))^2}{\sigma_l^2} \right) \right)^{1/2},$$

where $\sigma_i^2$ and $\sigma_l^2$ reflect uncertainties in the comparison stemming from the number and quality of the observations as well as uncertainties in wind forcing as estimated from an ensemble of 20 experiments that measure the effects of different admixtures of three different wind products. Note that this test statistic assumes estimates of correlations and lags among buoys are independent.

There are four test statistics that operate on multiyear time scales, $E(p)_{t,\eta}$, which compares observed and modeled state variables ($\eta$) that have been filtered for structures relevant to boundary layer mixing, as represented by changes in the KPP parameters (Zedler et al. 2016, manuscript submitted to Ocean Modell.). The test statistic is calculated separately for each state variable, including comparisons for temperature, salinity, and zonal and meridional currents. The comparisons are made on 90-day averages over a 4-yr period, excluding the first year and a half:

Table 2. List of scaling factors used to weight cost function components. Here $uvel$ is east velocity, $vvel$ is north velocity, temp is temperature, salt is salinity, and $R$ is the correlation. In addition, $k_i$ is the estimated degrees of freedom.

<table>
<thead>
<tr>
<th>Component</th>
<th>$(S_i)_{k_i}$</th>
<th>$k_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{uvel}$</td>
<td>8.72</td>
<td>8</td>
</tr>
<tr>
<td>$S_{vvel}$</td>
<td>3.55</td>
<td>16</td>
</tr>
<tr>
<td>$S_{temp}$</td>
<td>9.06</td>
<td>8</td>
</tr>
<tr>
<td>$S_{salt}$</td>
<td>6.24</td>
<td>12</td>
</tr>
<tr>
<td>$S_{R}$</td>
<td>6.72</td>
<td>12</td>
</tr>
<tr>
<td>$S$</td>
<td>0.29</td>
<td>17</td>
</tr>
</tbody>
</table>
where E statistic represents an ensemble mean. The multiyear test statistic and their sum. Thespecifics of this strategy are presented in the following section. The total test statistic E(p) used to calibrate KPP parameters includes five components: one for the correlation on daily time scales E(p)R and four for the multiyear response for the state variables E(p)q. Each one of these and their sum is weighted using information about each statistic’s effective degrees of freedom:

$$E(p) = \sum_{q=1}^{S} (S_q) E(p)_q.$$  

Through this weighting process, we can partially account for the effect of dependencies within and among our test statistics on the calibration results.

3. Bayesian inference

a. Formalism

Let d = (d₁, ..., dₙ)ᵀ be a vector of observation data and p = (p₁, ..., pₙ)ᵀ be a vector of model parameters. We consider a forward model G that predicts the data as a function of the parameters such that

$$d \approx G(p).$$  

Let π(p) be the prior probability distribution of p (representing any a priori information on p), L(d | p) the likelihood function (the probability of obtaining d given p), and π(p | d) the posterior probability distribution of p (the probability of occurrence of p given d). In this case, Bayes’s rule governs this formulation:

$$π(p | d) \propto L(d | p) π(p).$$  

The expression of the likelihood function depends on the assumptions made on the errors e (discrepancies) between the model and the observations [e = d - G(p)]. It is often assumed that the errors eᵢ are independent and normally distributed with a covariance Σ. Traditionally, a metric E(p), called the cost function, is constructed from the sum over squared errors normalized by estimates of the variances:

$$E(p) = \frac{1}{2} e^T Σ^{-1} e.$$  

**Fig. 2.** Projection of the Smolyak sparse quadrature nodes corresponding to level 5 on a 2D plane.
In this case the likelihood function can be written as

\[
L(d | p) = \frac{1}{(2\pi)^{k_e/2} |\Sigma|^{1/2}} \exp[-E(p)]
\]

(8)

and the joint posterior in Eq. (6) is then expressed as

\[
\pi(p | d) \propto \frac{1}{(2\pi)^{k_e/2} |\Sigma|^{1/2}} \exp[-E(p)] \prod_{i=1}^{m} \pi(p_i).
\]

(9)

The prior of the parameters \(p_i\) is assumed non-informative, that is, having a uniform distribution such that \(\pi(p_i) = 1/(b_i - a_i)\), where \(a_i\) and \(b_i\) are the bounds of the prior indicated in Table 1 and \(k_e\) is the effective degrees of freedom.

To account for the unexplained variance between the data and model, we allow for the test statistic \(E\) to be rescaled by a precision parameter \(S\). Incorporating a precision parameter like \(S\) is common in statistical inference as a way of scaling model fit to data given the level of agreement of the model with the data (Jackson and Huerta 2016, manuscript submitted to Geosci. Model Dev. Discuss., hereafter JH; Jackson et al. 2004). The scaling parameter \(S\) is included in the calibration as an additional uncertain variable. Its value is dependent on the test statistic \(E\) in the above equations, where the joint posterior becomes

\[
\pi(p, S | d) \propto S^{k_e/2} \exp[-SE(p)] \pi(S) \prod_{i=1}^{m} \pi(p_i).
\]

(10)
The prior $\pi(S)$ follows a Gamma distribution (Wang and Zabaras 2005; Gelman et al. 2004; Gelman 2006; JH) that depends on scale and rate parameters $\alpha$ and $\beta$ as follows:

$$
\pi(S) = \begin{cases} 
\frac{\beta^\alpha}{\Gamma(\alpha)} S^{\alpha-1} \exp(-\beta S), & S > 0 \\
0, & \text{otherwise}
\end{cases} 
$$

(11)

We employ an empirical Bayesian approach to determine the effective degrees of freedom $k_e$, as well as the scale and rate parameters, $\alpha$ and $\beta$, needed to specify an informative gamma distribution prior for $S$. The empirical Bayes’s approach, as described by JH, makes use of an ensemble of “perfect” model experiments in which one substitutes modeled data for observations. The ensemble plays out the null hypothesis, which in our case shows the effects of uncertain initial conditions and internal variability on our test statistics. The effective degrees of freedom $k_e$ may be estimated from this ensemble of test statistics $\{E(p)\}$, using

$$
k_e = \frac{2\langle\{E(p)\}\rangle}{\text{var}(\{E(p)\})}. 
$$

(12)

Using this equation, we estimate that $k_e = 17$ for the combined test statistic $E(p)$ shown in Eq. (4). The effective degrees of freedom is also used to determine the mean $\langle S \rangle = \alpha/\beta$ for its prior distribution:

$$
\langle S \rangle = \frac{\sqrt{k_e}}{\sigma_{E(p)}}, 
$$

(13)

where $\sigma_{E(p)}$ is the standard deviation of $\{E(p)\}$ caused by uncertain initial conditions. We use the same Eq. (13) above to weight individual cost components shown in Eq. (4). The component weights $\langle S_i \rangle$ are fixed with values shown in Table 2. We use informative values for the scale and rate parameters, $\alpha = 18.18$ and $\beta = 72.02$, according to JH. Thus, the prior for $S$ has a mean value of $\alpha/\beta = 0.252$ and variance $\alpha/\beta^2 = 0.003$. Since $S$ is updated with information about $E(p)$, which includes the effects of model biases that were excluded in our perfect modeling experiments, the net effect of including $S$ in our calibration is to produce much flatter posteriors relative to solutions without it.

b. Sampling method

The described Bayesian formulation requires sampling the resulting posterior [Eq. (10)] to estimate the KPP parameters. The MCMC methods are convenient and popular sampling strategies that require a large number of posterior evaluations. In our case, each posterior evaluation requires single MITgcm simulation for a given set of KPP parameters to compute $E$, which is
computationally prohibitive. Thus, we seek to build a surrogate model for the quantity of interest (QoI) $E$, as described in the following section, and use the random-walk Metropolis MCMC algorithm (Metropolis et al. 1953) to accurately and efficiently sample the posterior. Since the scaling parameter $S$ is included in our test statistic as a scalar correction to the data covariance matrix, it is also included as a hyperparameter to be estimated in addition to the model parameters $p$. We assume the priors for $p$ and $S$ are independent. This implies that for each MCMC step we can use Gibbs sampling to iteratively generate a value of $p$ conditional on $S$ and a value of $S$ conditional on $p$ as follows.

1) We simulate $p$ conditional on $S$, applying a sampling algorithm for $p$, but for just one iteration:

$$\pi(p \mid S, d) \propto \exp[-SE(p)] \prod_{j=1}^{m} \pi(p_j).$$ (14)

2) We simulate $S$ conditional on $p$: for the informative gamma distribution, we have

$$\pi(S \mid p, d) \propto S^{k_r/2 + \alpha - 1} \exp[-S[E(p) + \beta]],$$ (15)

which results in a gamma distribution of parameters $k_r/2 + \alpha$ and $E(p) + \beta$.

The two steps are repeated until convergence.

4. Accelerating Bayesian inference

To reduce the cost of sampling the posterior, we rely on a surrogate model of the QoI that requires a much smaller ensemble of model runs (Malinverno 2002; Marzouk and Najm 2009). Here, we rely on polynomial chaos expansions for representing the QoIs, which, in addition can efficiently provide statistical properties, such as the mean, variance, and sensitivities (Crestaux et al. 2009).

As a result of the complexity of the MITgcm, constructing a surrogate for the different model outputs is not feasible. Instead, we construct a single surrogate for the test statistic $E$, which is the QoI in this case. This test statistic $E$ is computed from the outputs of the model runs required for the construction of the surrogate, as explained below. This practice simplifies the PC calculation where only one surrogate model would be constructed (Pratola et al. 2013) that can be sampled directly in the posterior of Eq. (10) [or Eqs. (14) and (15)].

a. Polynomial chaos

Polynomial chaos is an efficient method of representing stochastic processes for the purpose of quantifying the uncertainties in a specific system (Le Maitre...
and Knio 2010; Xiu and Tartakovsky 2004). This method is based on a probabilistic framework that represents the stochastic quantities of interest as truncated polynomial expansions (Ghanem and Spanos 1991). We briefly describe below the application of the PC method for the construction of a PC surrogate for the test statistic $E$.

We denote by $\xi = (\xi_1, \ldots, \xi_m)$ the canonical vector of random variables representing $m$ uncertain KPP parameters in the $[-1, +1]$ space. PC seeks to express the $E$ function of $\xi$ as a truncated polynomial expansion in the following form:

$$E(\xi) \approx \sum_{k=0}^{R} e_k \psi_k(\xi), \quad (16)$$

where $e_k$ are the polynomial coefficients to be determined and $\psi_k(\xi)$ are the multidimensional scaled Legendre polynomials. These polynomials form an orthogonal basis with respect to the following defined inner product:

$$\langle \psi_i, \psi_j \rangle = \int \psi_i(\xi) \psi_j(\xi) \rho(\xi) \, d\xi = \delta_{ij} \langle \psi_i^2 \rangle, \quad (17)$$

where $\delta_{ij}$ is the Kronecker delta and $\rho(\xi)$ is the underlying uniform probability distribution. We note that the required number of expansion terms for highest PC order $r$ can be determined as follows:

$$R + 1 = \frac{(m + r)!}{m! \, r!}. \quad (18)$$

b. Determination of PC coefficients

PC expansion coefficients $e_k$ can be determined using different methods. These methods are typically classified into two categories: intrusive methods (Galerkin) (Villegas et al. 2012) and nonintrusive methods (Berveiller et al. 2006; Blatman and Sudret 2011). In the former category, a modification of the forward model code is required where each model equation must be reformulated to apply the Galerkin projection method (Le Maître and Knio 2010; Ghanem and Spanos 1991). While in the latter category, an ensemble of model predictions of $E(\xi)$ for specific realizations of $\xi$ is used to determine $e_k$. The set of $\xi$ can be selected either at random (Peng et al. 2014)

---

1 Tensor products are for one-dimensional scaled Legendre polynomials.
or deterministically (Babuška et al. 2007; Xiu and Hesthaven 2005; Nobile et al. 2008). In the present paper, we adopt the nonintrusive approach since making modifications to the complex MITgcm model is not realistic. In particular, we employed two different nonintrusive methods, as indicated below.

1) NONINTRUSIVE SPECTRAL PROJECTION

We first applied the traditional nonintrusive spectral projection (NISP) method (Reagan et al. 2003; Constantine et al. 2012; Conrad and Marzouk 2013) that takes advantage of the orthogonality of the PC basis. The PC expansion coefficients can be determined as follows (Sraj et al. 2014b):

$$e_k = \frac{\langle E, \psi_k \rangle}{\langle \psi_k, \psi_k \rangle} = \frac{1}{\langle \psi_k, \psi_k \rangle} \int E \psi_k(\xi) \rho(\xi) d\xi.$$  \hspace{1cm} (19)

The stochastic integrals are computed numerically using appropriate quadratures$^2$ as follows:

$$\langle E, \psi_k \rangle = \sum_{q=1}^{Q} E(\xi_q) \psi_k(\xi_q) \omega_q,$$  \hspace{1cm} (20)

where $\xi_q$ is the canonical vector at the quadrature point $q$, $\omega_q$ is the corresponding weight, and $Q$ is the total number of nodes in the quadrature.

$^2$ The quadrature order should be consistent with the PC order.
In our present work, a quadrature was built based on a Smolyak sparse nested grid (Petras 2000, 2001, 2003; Gerstner and Griebel 2003; Smolyak 1963) to reduce the number of expensive deterministic MITgcm runs. For a PC expansion of order \( r = 5 \) (total number of terms in the truncated PC expansion \( R + 1 = 252 \)) and uncertain parameters \( m = 5 \), a total number of \( Q = 903 \) quadrature nodes were needed, corresponding to Smolyak level 5. A projection of the quadrature nodes is shown in Fig. 2 on a two-dimensional plane.

We therefore ran MITgcm 903 times as per the quadrature and calculated the test statistic from the different model outputs. The test statistic corresponding to the sparse quadrature is shown in Fig. 3 as a function of different parameter spaces, as indicated in each panel. The red dashed line in each panel represents the cases where the other parameters are set to the center of the corresponding priors (i.e., \( \xi = 0 \)). These figures also clearly show the uncertainty bounds in the test statistic due to the uncertainty in the input parameters. This is true for all five parameters.

The PC expansion coefficients are computed from the output of the 903 quadrature runs using Eq. (19). Figure 4 plots the spectrum of the normalized PC coefficients, \( e_k/e_0 \), in absolute value. The vertical lines separate the PC expansion terms into degrees, where \( r = 0, \ldots, 5 \). The spectrum shows clearly that the PC suffers from convergence issues as the NISP-estimated PC coefficients do not decay with further increasing PC order but instead grow. This can be attributed to the presence of internal noise in the model that is not tolerated by the NISP method and thus overfitting the model from the quadrature points with additional refinement of the PC order.

To further assess this convergence issue, we show in Fig. 5 the deterministic MITgcm realizations plotted on top of their PC expansion counterparts. The difference between the two sets of data confirms the inability of the NISP-estimated surrogate PC model to efficiently represent the QoI. In Fig. 6, we show the test statistic \( E \), corresponding to 89 MITgcm model runs, where we vary \( R_i \) only infinitesimally, as indicated in the plot. The test statistic value is highly sensitive to infinitesimal perturbations of \( R_i \). This likely results from internal noise in the model that is amplified over time, in part by nonlinear interactions in the model. This results in a nonsmoothness manifested as high-amplitude high-frequency noise.

As a conclusion, the construction of a converging PC expansion using the NISP method is not successful and the surrogate model is not representative of the QoI of MITgcm; therefore, it cannot be used for further analysis.

2) BASIS-PURSUIT DENOISING

In an attempt to find a suitable PC surrogate model for the test statistic \( E \), we resort to a different approach that tolerates noise in the model but also that is non-intrusive. Instead of using spectral projections, we employ a recent technique that uses compressed sensing (CS) for polynomial representations, which first estimates the noise in the model response and then approximates the model predictions using a PC representation that tolerates the corresponding noise level. Let \( \mathbf{e} = (e_0, \ldots, e_R) \) be a vector of the PC coefficients to be determined and let \( \mathbf{E} = [E(\xi_1), \ldots, E(\xi_Q)] \) be a vector of the forward model evaluations function of sampled \( \xi_e \). We also define \( \Psi \) as the matrix where each
row corresponds to the row vector of $R + 1$ PC basis functions evaluated at the sampled $\xi_q$. CS solves the problem,

$$E = \Psi e$$

by exploiting the approximate sparsity of the signal (i.e., the vector of PC coefficients $e$ that necessarily converge to zero). The sparsity is set by constraining the system and minimizing its energy, that is, its $l_1$ norm. CS seeks a solution with a minimum number of nonzero entries by solving the optimization problem:

$$E_{1,\delta} = \left\{ \arg \min_{e} \| e \|_1; \| E - \Psi e \|_2 = \delta \right\},$$

where $\delta$ is the noise estimated in the signal.

This $l_1$-minimization problem is referred to as basis pursuit (BP), when $\delta = 0$, and as BPDN when a noise $\delta$ in the system is assumed, as proposed in Donoho (2006). We adapt the latter approach since we acknowledge the existence of noise in the predicted QoI. We note that in Eq. (22) the constraint depends on selected sampled parameters $\xi_q$ and their corresponding $E(\xi_q)$ and not on a general sample of $\xi$ and $E(\xi)$. As a result, the coefficients $e$ may be chosen to fit the input realizations, and not accurately approximate the model. To avoid this situation, we determine the noise $\delta$ by cross validation, as discussed in Peng et al. (2014). To solve $E_{1,\delta}$, standard $l_1$-minimization solvers may be used. In this work we use the MATLAB package SPGL1 (Van den Berg and Friedlander 2007) based on the spectral projected gradient algorithm (Van den Berg and Friedlander 2009).

We note that BPDN is equivalent in principle to the lasso approach (Tibshirani 1996) with a key difference that the former seeks for a sparse solution while the

\[\text{FIG. 14. The 2D Smolyak sparse quadrature nodes corresponding to levels 1–4 in the canonical vector space.}\]
latter seeks for a balance between sparsity and fitting (Van den Berg and Friedlander 2009). This difference appears in the tuning parameters that enter the formulation, but since they are estimated from the data, one would get an equivalent model fit to the surface. For more details on the lasso technique, readers are referred to James et al. (2013).

Here, we applied the BPDN to determine the PC coefficients for the surrogate model of the test statistic. Instead of using a Monte Carlo sampling method to generate realizations, as in Peng et al. (2014), we take advantage of the previously simulated 903 MITgcm realizations and use them to solve the optimization problem. The resulting PC normalized coefficients spectrum \( \frac{e_k}{e_o} \) is plotted in Fig. 7 (in absolute value) up to order \( r = 5 \) to assess the decay of the PC expansion coefficients. The vertical lines indicate the PC terms for different polynomial orders \( r = 0, \ldots, 5 \). The spectrum shows clearly the decay in the PC coefficients with increasing PC order using BPDN as opposed to the NISP method. This decay, however, reaches a plateau eventually and further increasing the PC order would have no additional effect on the PC surrogate accuracy. We note that the BPDN method belongs to a class of compressed sensing techniques that are generally based on the use of a large basis set, namely whose cardinality can be much larger than the ensemble used to reconstruct the signal (i.e., \( R \gg Q \)). The denoising step is based on the introduction of a nuisance parameter, that is, determined as part of the cross-validation process and that enables it to handle noise in the underlying data.

To further check the effect of increasing the PC order \( r \) on the surrogate accuracy, the PC expansion is sampled to find the probability distribution function (PDF) of the test statistic \( E \) using different polynomial orders. The PDFs are shown in Fig. 8, where we observe that as the PC order is increased, the PDFs get closer to each other, indicating that the higher-order modes carry little energy and that a sufficiently large basis has consequently been selected for the BPDN reconstruction. In addition, in Fig. 9 we also show the cumulative distribution functions (CDFs) constructed using the full-model runs versus those obtained from the PC surrogates determined using BPDN with different bases. Figure 9 also indicates that the CDFs obtained using higher-order bases agree with each other and with the CDFs obtained from the full-model runs directly. We note that constructing CDFs using the independent model runs is not feasible, as those model runs have different weights and are not uniformly sampled; a much larger set of uniformly sampled runs would be required to extract the CDFs directly from the outputs, but in the present setting the corresponding computational cost is prohibitive.

To confirm the ability of the constructed PC expansion to represent the QoI, we again show the MITgcm realizations plotted on top of their PC counterparts in Fig. 10 (left). One interesting observation is the internal noise in MITgcm that appears in the corresponding test statistic, while PC gives a smooth function for \( E \). While the PC excludes the internal noise in the QoI, it clearly captures the mean of the signal. Figure 10 (right) shows the same but in a scatterplot where MITgcm realizations are plotted against their PC counterparts. To quantify

![Figure 15](image1.png)

**Fig. 15.** NRE results found using different numbers of nodes corresponding to different Smolyak levels of refinement.

![Figure 16](image2.png)

**Fig. 16.** Response curves of test statistic \( E \) as a function of different parameters (in canonical space). For each curve, the other four parameters are set to \( \xi_i = 0 \).
the agreement, we use another common error metric where we calculate the normalized relative error (NRE) between the 903 MITgcm simulations and their reconstruction using the built PC as follows:

$$NRE = \frac{\sum_{q=1}^{Q} \left| \sum_{k=10}^{R} \phi_{k} \psi_{k}(\xi_{q}) \right|^2}{\left( \sum_{q=1}^{Q} |E(\xi_{q})|^2 \right)^{1/2}}.$$  

The NRE is calculated and found to be \(\sim 3\%\), which is acceptable, suggesting that BPDN is successful in constructing a surrogate that is able to accurately simulate the QoI.

For further validation of our surrogate model, an independent set of 954 MITgcm runs was conducted simultaneously by varying the same set of KPP uncertain parameters in an identical model setup and using the same test statistic. The sample is shown in Fig. 11 as a 2D projection of the \(\xi_1 - \xi_2\) plane (representing \(R_i_c\) and \(R_i_g\)). The corresponding test statistic \(E\) is shown in Fig. 12 as a function of the different parameters’ spaces. The different plots in Fig. 12 show a functional trend, mainly for \(R_i_c\) and \(R_i_g\); however, the variation in \(E\) suggests an internal noise in the model outputs in addition to responding to other parameters.

The PC expansion is calculated for the different sample points and is shown in Fig. 13 (left) versus the MITgcm realization. The PC expansion appears to reproduce the mean of the deterministic model. Figure 13 (right) shows the same but in a scatterplot where MITgcm realizations are plotted against their PC counterparts. The NRE is calculated and found to be \(\sim 4.2\%\), which is also quite acceptable, indicating that BPDN is successful in producing a surrogate that is able to accurately simulate the QoI.

We end this section by stating that in the present setting the QoIs are seen to exhibit appreciable variation even when infinitesimal changes in model parameters are considered, as indicated in Fig. 6. The resulting nonsmoothness manifests itself as high-amplitude high-frequency noise superimposed on a milder variability. The compressed sensing was robust with this noise and was effective in capturing the underlying variability. However, the presence of the noise effectively precludes an application of a straightforward convergence analysis.

c. ** PC sensitivity to ensemble size**

Normally, a number \(N\) of random samples is required to compute the PC coefficients using BPDN; this number is much less than the number of the unknowns, that is, the size of the PC expansion \((R + 1 = 252)\) such that \(N \ll R + 1\). In the above results, we instead used an ensemble of \(Q = 903\) sparse quadrature MITgcm model runs corresponding to Smolyak level 5 (Fig. 2) to compute the PC coefficients using BPDN due to their availability (upon using NISP method) where in this case \(N = Q \gg R + 1\). Here, we explore the possibility of utilizing a small number of MITgcm realizations to build a faithful PC expansion surrogate. To create a smaller ensemble and avoid running new expensive MITgcm simulations, we consider lower levels of refinement of the Smolyak quadrature shown in Fig. 14 in the canonical vector space for levels 1, 2, 3, and 4 with total numbers of nodes of 11, 51, 151, and 391, respectively. These levels are nested and therefore the nodes in each level are a subset of the higher levels. Thus, a small number of model runs \(N\) can be extracted from the original \(Q = 903\) runs, as per the Smolyak levels, and used to construct different PC models using BPDN.
In Fig. 15, we show the NRE of the original 903 Smolyak runs computed using PC models built employing quadrature nodes from the different Smolyak levels. We observe that the error increases as $N$ decreases; however, the increase is not significant. In fact, even with level 2 (51 model runs) the NRE is around 5%. We note that in all the results below we used the PC expansion model constructed using the level-5 Smolyak runs with BPDN.

d. Statistical moments and sensitivity analysis

The PC expansion of $E$ formulated in Eq. (16) allows for us to compute few statistical moments easily from the PC coefficients because of the orthogonality of the basis used and to setting $\psi_0(\xi) = 1$. The mean of $E$ can then be shown to be equal to the constant coefficient of the PC expansion as follows:

$$\mu_E = \int E \rho(\xi) \, d\xi \approx \langle E, \psi_0 \rangle = e_0. \quad (24)$$

Similarly, the variance of $E$ can be computed as follows:

$$\sigma_E^2 = \int (E - \mu_E)^2 \rho(\xi) \, d\xi \approx \sum_{k=1}^{R} e_k^2 \langle \psi_k, \psi_k \rangle. \quad (25)$$

Using these equations, we computed the mean and the standard deviation to be $\mu_E = 325.65$ and $\sigma_E = 31.30$, respectively.

FIG. 18. Chains of parameters using MCMC samples. The MITgcm default values are indicated as horizontal lines.
To quantify the contribution of each of the five uncertain KPP parameters to the variance in the model output $E$, a global sensitivity analysis can be performed. For this purpose, we compute the total sensitivity index $S_i$, which is the ratio of the model variability by the $i$th uncertain KPP parameter $\xi_i$ to the total model variance [Eq. (25)]:

$$S_i = \frac{\sum_{k=1}^{K} e^2_k(\psi_k, \psi_k)}{\sum_{k=1}^{K} \sum_{j=1}^{R} e^2_k(\psi_k, \psi_k)}$$

$$K_j = \{k \in \{1, \ldots, R\}: \alpha^2_j > 0\}, \quad (26)$$

where $\alpha^2_j$ is the multi-index associated with the $k$th term and $i$th uncertain KPP parameter of the PC expansion (Le Maitre and Knio 2010; Crestaux et al. 2009; Sudret 2008).

The sensitivities of each of the uncertain KPP parameters are calculated and summarized in Table 3. It is notable that the test statistic is most sensitive to $R_i$ and then to $R_g$ and $\phi_{m,\text{unst}}$, as clearly reflected in the corresponding total sensitivities, $S_1 - S_3$, respectively.

**e. Response surfaces**

Another advantage of the PC expansion representation is the possibility of sampling the surrogate model with almost no computational cost, as it only requires polynomial evaluations. This allows us to construct a response surface for the $E$ function of the different uncertain KPP parameter $\xi_j$ in the prior range $[-1, +1]$, while the other parameters are set to $\xi_j = 0$. The plot shows strong dependence of $E$ on $R_i$, consistent with the sensitivity results shown earlier. Similarly, $E$ shows some variations in the function of $R_i$ and $\phi_{m,\text{unst}}$ but with a relatively less dependence compared to $R_i$. The remaining curves exhibit lines with a small slope, suggesting that $E$ depends only mildly on $\phi_{m,\text{unst}}$ and $C^*$. We also show the 2D response surfaces for $R_i$ versus $R_g$ in Fig. 17 (top), and $R_i$ versus $\phi_{m,\text{unst}}$ in Fig. 17 (bottom) (the other parameters are set to $\xi_k = 0$). The quadrature samples $E(\xi_j)$ are shown on top of the surfaces for comparison.

**5. KPP inference**

In this section, we report on the results of applying Bayesian inference to update the PDFs of the KPP parameters. To this end, we sample from the posterior distributions [Eqs. (14) and (15)] using a random-walk MCMC algorithm (Metropolis et al. 1953) that requires a large number of MITgcm simulations to compute $E$ for each sampled set of KPP parameters. To avoid such computational cost, we used the constructed PC surrogate discussed in section 4 to compute $E$ predictions instead and use them in Eqs. (14) and (15). We note that we ran the MCMC sampler $10^6$ times to attain convergence, as discussed below.

The outputs of the MCMC algorithm are the sample chains for the different KPP parameters in addition to the scaling parameter $S$. We show in Fig. 18 the trace plots of these parameters (chains versus MCMC iteration number) that clearly demonstrate well-mixed chains. This indicates that the distribution of the chains is expected to remain unchanged with further sampling and to converge to a stationary distribution. To further test for MCMC convergence, we computed the running mean for the different KPP parameters and plotted the results in Fig. 19. The plot shows that the running means have reached a steady state, indicating a good level of convergence among the chains. We also computed the empirical autocorrelation at lag $s$ $\gamma(s)/\gamma(0)$ for the different MCMC chains shown in
In Fig. 18, we also show the MITgcm default KPP parameters (from Table 1) as horizontal lines in each panel for comparison with the range of the chains. The chains of both $Ri_c$ and $Ri_g$ appear to be moving around their corresponding MITgcm default values, while for $\phi_{m,\text{unst}}$ and $\phi_{s,\text{unst}}$, the chains appear to cover a large part of the prior range but not moving around their corresponding MITgcm default values. The former observation indicates that the test statistic $E$ was informative and useful for inferring $Ri_c$ and $Ri_g$, while the latter is an indication of noninformative posteriors due to a lack of data for inferring $\phi_{m,\text{unst}}$ and $\phi_{s,\text{unst}}$. Regarding $C^*$, the chain appears to also cover a large part of the prior range; however, it is moving around the corresponding MITgcm default values, suggesting a wide posterior. Finally, the chain of the hyperparameter $S$ is well mixed with values ranging from $\sim 0.05$ to 0.12.

To confirm our conclusions from the chains, we estimated the marginalized posterior distributions from the MCMC chains of each parameter, discarding the first $2 \times 10^5$ burn-in iterations, using kernel density estimation (KDE) (Parzen 1962; Silverman 1986). We show these marginalized posterior PDFs in Fig. 21. Regarding $Ri_c$, its posterior PDF has a well-defined peak; however, it is skewed to the right with large spread. Because of the skewness in the PDF, we report the mean estimate to be $\sim 0.35$, which is close to the corresponding MITgcm default value of 0.3 (both values are shown as vertical
lines in the same PDF). For $\text{Ri}_s$, its posterior also has a well-defined peak of 0.68 that is very close to the corresponding MITgcm default value of 0.7 with little skewness to the left and a large spread. Regarding $\phi_{m,\text{unst}}$ and $\phi_{s,\text{unst}}$, their posterior PDFs have no clear shape and mostly flat consistent with our conclusions from their chains. The computed mean estimate has no usefulness. For $C^*$, its posterior PDF also has no clear shape and no clear peak but indicates that the mean value is in the upper range of the prior. The 95% intervals of high posterior probability are shown as shaded regions for the inferred parameters. Finally, the scaling parameter $S$ exhibits a Gaussian-like shape contrasted by the prior gamma distribution.

The posterior distributions are consistent with results of previous efforts to calibrate the KPP model (Large et al. 1994; Large and Gent 1999). They show that the default critical bulk and gradient Richardson numbers and the nonlocal convective parameter $C^*$ are within the range of acceptable parameter settings. In the absence of a detailed analysis of the role of $\phi_{m,\text{unst}}$ and $\phi_{s,\text{unst}}$ in our calibration, we can remark on possible explanations for their broad posterior distributions. It must be the case that either the eddy diffusivity and viscosity are weak functions of these parameters or that there are negative feedbacks in the model that ultimately limit their influence on the ocean state. Since under equivalent local ocean conditions the functional dependence of eddy diffusivity on $\phi_{s,\text{unst}}$ or eddy viscosity on $\phi_{m,\text{unst}}$ implies variation in their values by a factor of 2–3 over the range in our priors, we believe that the latter explanation is more likely [Large et al. (1994); see Eq. (B1)]. The fact that the posterior distributions are in line with previous research is encouraging for the application of uncertainty quantification methods, such as polynomial chaos expansion for sensitivity analysis. The question of whether data can be used to build more predictive models is still open.

### 6. Summary and conclusions

In this work, we have presented a polynomial chaos-based method for the inference of five KPP parameters. Below, we remark on the PC construction process and the KPP inference results.

The method relied on building a surrogate model for a newly developed test statistic instead of the model output. The advantage was to avoid building surrogates for several model outputs and for different time scales. The PC construction was implemented using two techniques. First a nonintrusive spectral projection method was used that required a quadrature of level 5 nodes ($Q = 903$ model runs). The computed PC expansion suffered from convergence issues as a result of the presence of internal noise in the predictions. This resulted in overfitting of the data with increasing levels of PC refinement. The resulting PC model was thus unsuitable and discarded. The second technique used was BPDN, which tolerates noise in the model and assumes sparsity in the PC expansion, requiring a smaller number of model runs. This technique proved to be successful, as it excludes the noise from the model when finding the PC coefficient that leads to a faithful PC surrogate. Several error metrics were computed to check the validity of the PC model.

As follow-up research, we are seeking to infer all nine KPP parameters in a calibration where the wind is treated as an adjustable parameter. The additional number of parameters dramatically increases the number of required expensive model runs. Thus, a different approach is needed where we will investigate an adaptive approach to PC (Winokur et al. 2013).

### Acknowledgments

This research made use of the resources of the Supercomputing Laboratory and computer clusters at King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia. IS, SZ, CJ, and IH are supported in part by KAUST Award CRG-1-560 2012-HOT-007; SZ and OK are supported in part by the Office of Advance Scientific Computing Research, U.S. Department of Energy, under Award DE-SC0008789.

### REFERENCES


