Drag Parameter Estimation Using Gradients and Hessian from a Polynomial Chaos Model Surrogate

IHAB SRAJ

Duke University, Durham, North Carolina

MOHAMED ISKANDARANI

University of Miami, Miami, Florida

W. CARLISLE THACKER

University of Miami, and Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida

ASHWANTH SRINIVASAN

University of Miami, Miami, Florida

OMAR M. KNIO

Duke University, Durham, North Carolina

(Manuscript received 18 March 2013, in final form 6 September 2013)

ABSTRACT

A variational inverse problem is solved using polynomial chaos expansions to infer several critical variables in the Hybrid Coordinate Ocean Model's (HYCOM's) wind drag parameterization. This alternative to the Bayesian inference approach in Sraj et al. avoids the complications of constructing the full posterior with Markov chain Monte Carlo sampling. It focuses instead on identifying the center and spread of the posterior distribution. The present approach leverages the polynomial chaos series to estimate, at very little extra cost, the gradients and Hessian of the cost function during minimization. The Hessian's inverse yields an estimate of the uncertainty in the solution when the latter's probability density is approximately Gaussian. The main computational burden is an ensemble of realizations to build the polynomial chaos expansion; no adjoint code or additional forward model runs are needed once the series is available. The ensuing optimal parameters are compared to those obtained in Sraj et al. where the full posterior distribution was constructed. The similarities and differences between the new methodology and a traditional adjoint-based calculation are discussed.

1. Introduction

This article is a follow-up to the parameter estimation problem presented in Sraj et al. (2013) where the posterior probability distributions of key variables in the wind drag parameterization at high wind speeds—the drag multiplicative factor, the saturation wind speed, and the drag slope after saturation—were inferred from airborne expendable bathythermograph (AXBT) temperature profiles collected during Typhoon Fanapi (2010). Sraj et al. (2013) used a Bayesian inference methodology to solve the inverse problem, and relied on Markov chain Monte Carlo (MCMC) to construct the full posterior distribution. The approach's efficiency hinged on a faithful polynomial chaos (PC) surrogate to circumvent the large computational cost associated with the MCMC sampling [each sample was the equivalent of a forward Hybrid Coordinate Ocean Model (HYCOM) run and 10^6 samples were used].

The goal of the present article is to point out that combining a PC surrogate and a variational approach to

Corresponding author address: Omar Knio, Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708.
E-mail: omar.knio@duke.edu

DOI: 10.1175/MWR-D-13-00087.1

the inverse problem in Sraj et al. (2013) will *easily* yield the center *and* spread of the posterior distribution,[1] will not incur the costs and complications of the MCMC step since the full posterior is not constructed, and will not require an adjoint code (only forward model runs are needed to build the surrogate). To this end, the inverse problem in Sraj et al. (2013) is first recast as the minimization of the log-likelihood cost function penalizing the misfit between predictions and observations; an optimization algorithm is then applied to obtain the solution. A major hurdle is in computing the gradients needed during optimization, and that usually necessitates the tedious development and application of an adjoint code. Here we completely bypass this hurdle by reusing the surrogate developed in Sraj et al. (2013) to compute the necessary gradients. The PC series also delivers the useful but hard-to-compute cost function's Hessian at very little extra computational cost. The Hessian can be used to enhance the robustness and performance of the minimization algorithm, and to provide an estimate of the spread of the posterior distribution around the optimal values. Once the surrogate is available, the minimization algorithm can proceed without the need for further model runs.[2] The methodology is applicable to a wide range of atmospheric and oceanic models, and is illustrated here for HYCOM, a full three-dimensional and complex ocean general circulation model.

The present work capitalizes on a key aspect of PC methods (Ghanem and Spanos 2002; Le Maître and Knio 2010) (and their close relatives known as stochastic collocation methods), namely, the availability of a series representation for the model response to uncertain parameters. This series can be efficiently manipulated for the purpose of statistical analysis, data assimilation, and inverse modeling. Marzouk et al. (2007) and Marzouk and Najm (2009) used PC series as efficient surrogates to speed up the construction of the posterior in Bayesian inference problems; their work directly influenced Sraj et al. (2013). PC expansions have been used in ensemble Kalman filter–based data assimilation systems to quickly and reliably update the covariance matrix (Saad and Ghanem 2009; Zeng and Zhang 2010; Li and Xiu 2009; Blanchard et al. 2010). They have also been combined with variational approaches to speed up different aspects of the solution procedure (Zabaras and Ganapathysubramanian 2008; Southward 2008; Pence et al. 2010). Zabaras and Ganapathysubramanian (2008) adopted stochastic collocation series to speed up inverse and optimal design problem using legacy codes, but relied on finite differencing to calculate the cost function gradients. Southward (2008) and Pence et al. (2010) were the first to differentiate the polynomial chaos series to update the search directions in a variational parameter estimation problem but they do not exploit the Hessian in their calculations. Mattern et al. (2012) used PC expansions to identify the optimal parameters of an ecosystem model by visually locating their cost function minimum; their methodology could be framed as a minimization algorithm albeit no gradients were actually computed. The present work exploits PC series to estimate the cost function gradient *and* Hessian by direct differentiation, and the Hessian is used to approximate the uncertainty in the optimal parameters; neither MCMC sampling nor adjoint codes are needed. It represents, to the authors' knowledge, a first application of this methodology to a full ocean general circulation model where actual observations are used.

The plan for the present article is as follows. Section 2 recasts the problem in Sraj et al. (2013) as a variational inverse problem whose iterative solution requires access to the gradient and Hessian of the cost function; the computations of these quantities from the PC surrogate are also described. Section 3 presents the variational solution to the inverse problem and compares it to the one obtained in Sraj et al. (2013). Finally, section 4 points out how the present streamlined approach relates to so-called adjoint methods, and how it fits into the familiar framework of data assimilation. For the sake of brevity we omit a lot of details concerning the drag parameters, the AXBT data, models, and the PC surrogate, and we refer the reader to Sraj et al. (2013), Winokur et al. (2013), and Alexanderian et al. (2012) for more in-depth discussions.

2. Problem formulation

As described by Sraj et al. (2013), the objective was to infer the probability density of the parameters α, $V_{\max}$, and m determining the drag coefficient from observations of temperature in the upper ocean. AXBT data provide numerous measurements of temperature T_i at location x_i and time t_i, where $i = 1, \ldots, N$ over 5 days during the passage of Typhoon Fanapi (2010). Sraj et al. (2013) described how to construct polynomial approximations to simulations $M_i = M_i\,(x_i, t_i, \alpha, V_{\max}, m)$ of those temperatures. Given a distribution of drag parameters, there is a distribution of differences $M_i - T_i$ that was assumed to be Gaussian for each i. The widths of these Gaussians might have been assumed to be the

[1] The center and spread are generally sufficient to gauge the value and uncertainty of the optimal parameters.

[2] The present results were obtained without a single additional HYCOM run.

same for all i, but to allow the information about the wind stress as manifested in the ocean's thermal structure to be treated differently on the different days during which the storm passes through the region, different variances σ_d^2 were allowed for different days $d = 1, \ldots, 5$; as in Sraj et al. (2013), σ_d^2 are treated as hyperparameters. The desired density for the drag parameters and for the σ_d^2 were constructed by drawing their values repeatedly from relatively uninformative prior distributions and evaluating the Gaussians via Bayes's theorem (Berger 1985; Epstein 1985; Berliner et al. 2003; Gelman et al. 2004; Bernardo and Smith 2007):

$$p(\alpha, V_{\max}, m, \{\sigma_d^2\} \mid T_i) \propto \prod_{i=1}^{N} \left\{ \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[\frac{-(M_i - T_i)^2}{2\sigma_i^2} \right] \right\} p(\alpha, V_{\max}, m, \{\sigma_d^2\}), \tag{1}$$

where $p(\alpha, V_{\max}, m, \{\sigma_d^2\})$ is the prior probability density function (pdf) for the parameters/variances and $p(\alpha, V_{\max}, m, \{\sigma_d^2\} \mid T_i)$ is the posterior density for the parameters/variances. Uniform priors were assumed for the drag parameters and Jeffreys priors for the variances σ_d^2.

The Monte Carlo step can be avoided if one is satisfied with only the mode of the posterior and an approximation to the spread around it: simply maximize the posterior density, or equivalently, minimize the negative of its logarithm:[3]

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

where J_d is the misfit cost for day d, the $\ln(\sigma_d^2)$ terms come from the normalization factors of the Gaussian likelihood functions and from the Jeffreys priors.[4] The expression for J_d is

$$J_d(\alpha, V_{\max}, m, \sigma_d^2) = \frac{1}{2\sigma_d^2} \sum_{i \in \mathcal{I}_d} [M_i - T_i]^2, \tag{3}$$

where $\mathcal{I}_d$ is the set of n_d indices of the observations from day d.

The minimization of the cost function [Eq. (2)] will be solved iteratively using an optimization algorithm that requires two iteration loops: an outer loop based on Newton's method to update the search directions, and an inner loop to find the cost function optimum along this direction using Brent's line-search algorithm (Brent 1973). The outer loop requires the gradient and the Hessian of the cost function (Lawless et al. 2005; Fletcher 1987; Avriel 1976), which are simple to obtain because of the polynomial nature of M_i (see below) and the explicit dependence on σ_d^2. Expressions for the components of the gradient are

$$\left[\frac{\partial \mathcal{J}}{\partial \alpha}, \frac{\partial \mathcal{J}}{\partial V_{\max}}, \frac{\partial \mathcal{J}}{\partial m} \right] = \sum_{d=1}^{5} \frac{1}{\sigma_d^2} \left\{ \sum_{i \in \mathcal{I}_d} (M_i - T_i) \left[\frac{\partial M_i}{\partial \alpha}, \frac{\partial M_i}{\partial V_{\max}}, \frac{\partial M_i}{\partial m} \right] \right\}, \quad \text{and} \tag{4}$$

$$\frac{\partial \mathcal{J}}{\partial \sigma_d^2} = -\frac{1}{2(\sigma_d^2)^2} \sum_{i \in \mathcal{I}_d} (M_i - T_i)^2 + \frac{n_d + 2}{2\sigma_d^2}. \tag{5}$$

Similar expressions for the elements of the Hessian matrix can easily be obtained by differentiating these; for brevity they are not shown.

We sketch below the PC surrogate's principal features since it is central to the gradient calculations. The PC surrogate is essentially a truncated polynomial series expansion of the following form:

$$M(\mathbf{x}, t, \alpha, V_{\max}, m) \doteq \sum_{k=0}^{P} \hat{M}_k(\mathbf{x}, t) \Psi_k(\alpha, V_{\max}, m), \tag{6}$$

where $\hat{M}_k$ are the series coefficients [see Winokur et al. (2013), Sraj et al. (2013), and Alexanderian et al. (2012) for how they are computed and how the series accuracy can be verified]; P is finite and depends on the truncation strategy adopted; and the functions $\Psi_k(\alpha, V_{\max}, m)$ form an orthogonal basis of an underlying $(\alpha, V_{\max}, m)$ probability space. For the present case of uniform

[3] The 2πs from the normalization of the likelihoods do not contribute to the maximum and can be ignored. Also, if the maximum is within the support of the uniform priors for the drag parameters, those priors also contribute only constant terms that can be neglected. However, the maximum may be on the boundary of that support.

[4] The Jeffreys prior for the variance of a Gaussian is inversely proportional to that variance: $p(\sigma_d^2) = 1/\sigma_d^2$ for $\sigma_d^2 > 0$ and 0, otherwise.

distributions, the basis functions are products[5] of univariate Legendre polynomials (Le Maître and Knio 2010). This series representation model is used to compute gradients necessary for the optimization algorithm; thus

$$\left[\frac{\partial M}{\partial \alpha}, \frac{\partial M}{\partial V_{\max}}, \frac{\partial M}{\partial m}\right] = \sum_{k=0}^{P} \hat{M}_k(\mathbf{x}, t)\left[\frac{\partial \Psi_k}{\partial \alpha}, \frac{\partial \Psi_k}{\partial V_{\max}}, \frac{\partial \Psi_k}{\partial m}\right]. \tag{7}$$

Since the basis functions are products of Legendre polynomials, analytic expressions for the derivatives can be easily obtained (Alexanderian et al. 2012). Once the coefficients $\hat{M}_k$ are available, all that is required is differentiating the basis functions and summing the series in Eq. (7) to evaluate the gradients in Eq. (4).

Note that elements of both the gradient and the Hessian of the cost function require differentiating the model counterparts of the data M_i. As these are polynomials of the drag parameters (Sraj et al. 2013; Alexanderian et al. 2012), their differentiation is straightforward. This is where using a polynomial approximation to numerical simulations is extremely valuable. Without them the gradient would have to be computed using an adjoint code constructed specifically for the model that produced the simulations, would have required significantly greater computational expense than is required to evaluate the polynomial-based gradient, and the Hessian would be even more expensive to obtain.

3. Results

The optimal drag parameters obtained by the variational formulation are presented and discussed in this section. These values are compared to those reported in Sraj et al. (2013). Posterior distributions for the drag parameters and variances using the two approaches are also presented and compared. The reader is referred to Sraj et al. (2013) and Winokur et al. (2013) for more details concerning the AXBT data and the surrogate construction.

The iterative minimization algorithm relied on a combination of Newton's method and on Brent iterative line searches. The l_2 norm of the cost function gradient was used as a stopping criterion with a tolerance of 10^{-5} for the outer iterations and 10^{-2} for the inner iterations. The initial guess was taken as the middle of the interval[6] over which the prior was assumed uniform for each parameter[7] (Sraj et al. 2013). A total of 16 outer iterations and 126 inner iterations were needed to reach the optimal solution. Figure 1 shows the evolution of the cost function decrease $|\Delta\mathcal{J}|$ between consecutive outer iterations. The first 11 outer iterations exhibit large $|\Delta\mathcal{J}|$ with commensurate decreases in the cost function itself; $|\Delta\mathcal{J}|$ decreases rapidly beyond the 12th iteration and reaches 10^{-10} at the end of the 16th iteration. The contour plots in Fig. 2 show the trajectory taken by the optimal parameters at the end of each outer iteration, and reveal narrow valleys along the two directions $V_{\max}$ and m.

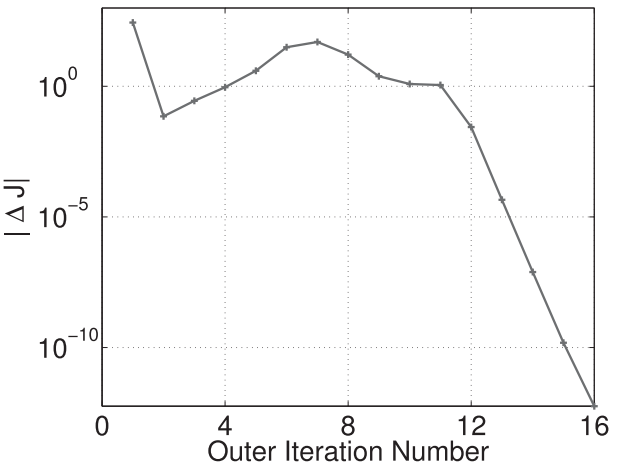

FIG. 1. The change in the cost function $|\Delta\mathcal{J}|$ between consecutive outer iterations.

The optimal values of α, $V_{\max}$, and m obtained at the end of the 16th iteration are: 1.0289, 34.0314 m s^{-1}, and -1.0195×10^{-5}, respectively. The optimal values of α and $V_{\max}$ are in agreement with the maximum-a-posteriori (MAP) values obtained using the MCMC approach (Sraj et al. 2013) (1.0267 and 34.0190, respectively).[8] The optimal m, however, is not in agreement with the MAP value of -0.4394×10^{-5} reported in Sraj et al. (2013). This is no cause for concern as the MCMC analysis showed the AXBT data to be uninformative with regard to m; a closer inspection of the present uncertainty in m (see below)

[5] The 3D basis function is the product of 1D basis functions; for example, $\Psi_k(\alpha, V_{\max}, m) = L_a(\alpha)L_b(V_{\max})L_c(m)$, where $L_a(\alpha)$, $L_b(V_{\max})$, and $L_c(m)$ are the usual Legendre polynomials of degree a, b, and c in the variable α, $V_{\max}$, and m, respectively. Note that the polynomial's argument can be mapped onto the standard interval $[-1, 1]$ on which Legendre polynomials are defined by a linear map.

[6] $0.4 \le \alpha \le 1.1, 20 \le V_{\max} \le 35$ m s^{-1}, and $-3.8 \times 10^{-5} \le m \le 0.0$.

[7] We have repeated the analysis starting from a different initial guess. The computations showed that the iterations converged to the same optimum. This is not surprising because in the present case the posterior distribution is unimodal.

[8] The MCMC was reran for 1 000 000 000 iterations to confirm the MAP values.

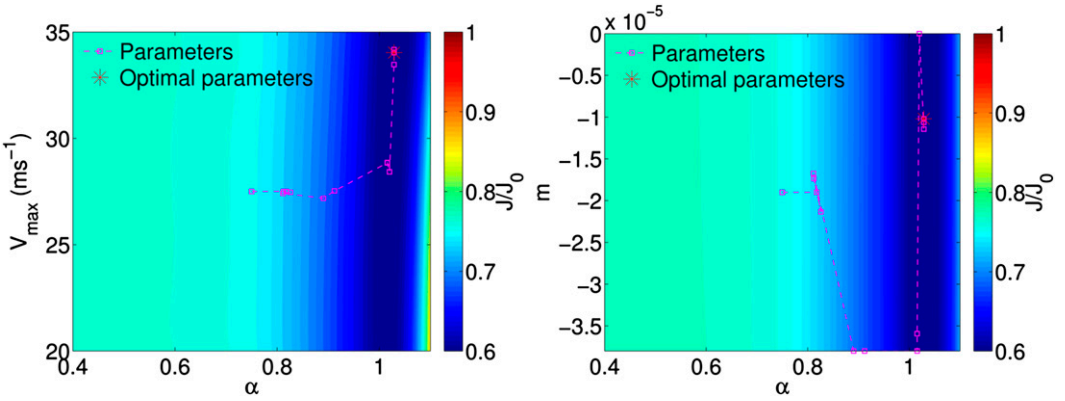

FIG. 2. Contour plots of normalized cost function $\mathcal{J}/\mathcal{J}_0$ along constant (left) $V_{\max}$ and (right) m surfaces. The dashed lines show the trajectory of the parameters at the end of each outer iteration, and reveal the last four iterations to be directed along the $V_{\max}$ and m directions.

corroborates the conclusion of the MCMC analysis. The optimal hyperparameters σ_d^2 are listed in Table 1 along with the MAP values; the comparison shows a perfect agreement between the results of the MCMC and variational approaches.

Although variational methods do not deal directly with the parameters' full posterior distributions, it is still possible to gauge the uncertainty in their estimate. This uncertainty gauge requires two key elements: an assumption that the parameters' pdf can be reasonably approximated with a Gaussian distribution, and access to the error covariance matrix. The assumption that the observation errors in the AXBT data are random and normally distributed, allows us to estimate this error covariance matrix as the inverse of the Hessian matrix of the cost function (Gejadze et al. 2011; Le Dimet et al. 2002; Thacker 1989). Furthermore, the square root of the diagonal elements of the error covariance matrix is a measure of the spread (single standard deviation) of the marginal posterior. The PC-provided Hessian is then all that is needed to construct the Gaussian posterior.

The computed spreads of the drag parameter were found to be: 0.0058, 1.2754, and 3.5214×10^{-5} for α, $V_{\max}$, and m, respectively, and their corresponding *presumed* Gaussian posteriors are shown in Fig. 3 (top) along with the posteriors obtained from the MCMC sampling. The following remarks can be made:

- The Gaussian α posterior matches quite well the MCMC posterior for which the assumption of Gaussian statistics holds.
- The Gaussian $V_{\max}$ posterior, however, is rather different from the MCMC prediction. First, it fails to capture the MCMC posterior's heavy tail. Second, the variational estimate of the spread is only half as much as the MCMC estimate (again, this could be due to the heavy tail of the MCMC $V_{\max}$ posterior). Third, the variational spread suggests that values outside the range of the original prior are possible.
- The Gaussian posterior for m is uninformative as the spread is large relative to the prior range, essentially mirroring the MCMC conclusion.

TABLE 1. Optimal parameters and hyperparameters and their spread calculated using variational and MCMC approaches.

	Method			
	Variational		MCMC	
Parameter	Optimal	Spread	MAP	Spread
A	1.0289	0.0058	1.0267	0.0064
$V_{\max}$	34.0314	1.2754	34.0190	2.4354
M	-1.0195×10^{-5}	3.5214×10^{-5}	-0.4394×10^{-5}	1.0824×10^{-5}
σ_1^2	0.6554	0.0637	0.6536	0.0655
σ_2^2	0.5712	0.0435	0.5699	0.0445
σ_3^2	0.5522	0.0407	0.5578	0.0418
σ_4^2	0.6684	0.0446	0.6742	0.0455
σ_5^2	0.9990	0.0686	1.0074	0.0702

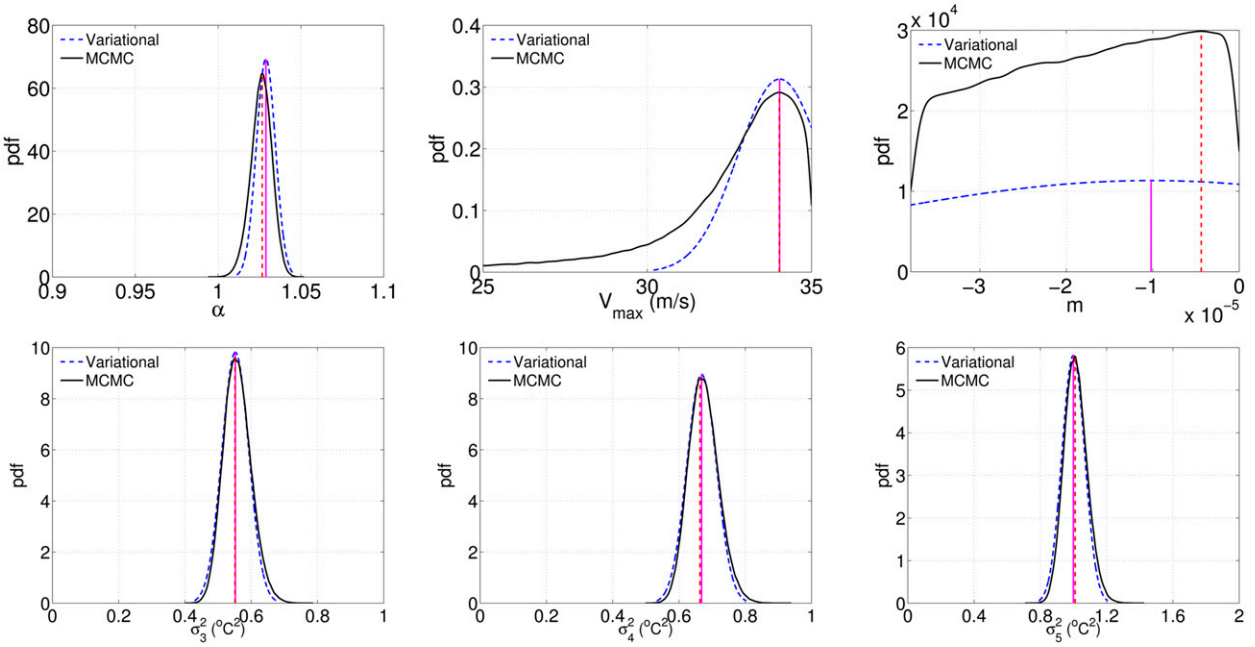

FIG. 3. Posterior probability distributions for (top) drag parameters and (bottom) variances σ_d^2 at selected days using variational method and MCMC from Sraj et al. (2013). The vertical lines correspond to the MAP values determined using MCMC and optimal parameters using the variational method.

- The σ_d^2 spreads match very well those obtained from the MCMC calculation; furthermore, their MCMC posterior is quite close to a Gaussian shape as shown in the bottom rows of Fig. 3.

In summary the present posteriors are, by and large, in agreement with those obtained after the MCMC sampling. The main discrepancies can be traced back to either non-Gaussian-like posteriors, or to a parameter's insensitivity to the data.

4. Discussion

This article's main message concerns the effective use of the PC surrogate to estimate the gradients of a numerical model. The success of this methodology in the context of the present inverse problem begs the question whether it could be useful in other problems where gradients are needed (e.g., data assimilation), and whether there are any efficiencies to be gained. Here we discuss qualitatively the similarities and differences between a traditional adjoint-based gradient calculation and the computations performed here.

Adjoint-based methods are considered efficient because the gradient calculation is independent of the number of control parameters (Talagrand and Courtier 1987; Courtier and Talagrand 1987, 1990; Thacker 1991). Each gradient calculation requires a single simulation of the forward model, and a single simulation of the adjoint model backward in time. The computational cost, however, can be substantial as the backward integrations can be equivalent to several (2–5) forward model runs (Baur and Strassen 1983; Griewank and Walther 2008). In the present experiment where the descent algorithm required 16 outer iterations and 126 inner iterations, the computational cost would have amounted to 158–206 HYCOM runs.

The computational cost of the PC calculations is dominated by the number of HYCOM simulations needed to determine the series coefficients. These coefficients are computed by evaluating multidimensional integrals in the (α, V_{max}, m) space through numerical quadrature. Each quadrature sample requires a HYCOM simulation and thus the cost scales linearly with the number of quadrature points Q. In turn Q depends on the type of quadrature, on the number of uncertain variables, and on the truncation strategy.[9] Since series truncation and quadrature sampling impact the approximation properties of the series and its computational cost, the present work relied on an

[9] The multidimensional form of Gauss-type quadrature would require $Q = (p + 1)^3$ for the three variable case; its cost grows exponentially with the number of variables. Smolyak (Smolyak 1963; Petras 2003; Gerstner and Griebel 2003) sparse quadrature offers an attractive alternative for high-dimensional integrals.

adaptive pseudospectral projection to build the series *and* to perform the quadrature, so that accuracy was guaranteed at a minimal computational cost (Constantine et al. 2012; Conrad and Marzouk 2013; Winokur et al. 2013). The adaptive procedure (see the appendix for additional details) required 67 samples only to construct the series to within a specified tolerance. No additional simulations were needed for the optimization proper, all that was required was to develop the differentiation formulas for the basis functions, and summing the appropriate series. Unlike adjoint calculations, however, the number of ensemble members strongly depends on the number of control parameters, and on the quadrature adopted. Furthermore, the surrogate delivers the model dependence on a subset of the variables.

A major drawback of traditional adjoints is the substantial upfront investments necessary in developing and maintaining adjoint codes for large general circulation models. Although automatic differentiation (AD) tools and adjoint compilers can help mitigate this cost (Giering and Kaminski 1998; Tber et al. 2007; Bischof et al. 1992; Zedler et al. 2012), a substantial amount of manual intervention is still necessary. Additionally, consistency between the forward model and its adjoint code dictates that modifications and upgrades to the forward model be propagated to the adjoint code as well, thus compounding the maintenance burden. This naturally leads one to explore approaches that overcome such burdens. For instance, Cao et al. (2007) and Altaf et al. (2013) have recently explored the use of a proper orthogonal decomposition/reduced model approach to determine an approximate adjoint. Similar to the present methodology, the approach in Altaf et al. (2013) is based on an ensemble of forward model runs, which effectively avoids the drawbacks of traditional adjoint techniques. On the other hand, fundamental differences in the construction exist, namely, regarding the selection of the ensemble, which in the present approach enables us to establish a global surrogate representation, and to estimate the spread around optimal parameters, namely, through the Hessian of the surrogate.

In summary the PC-based approach offers an attractive solution to parameter identification problems when the following issues are relevant: the problem requires a detailed exploration of a relatively low-dimensional parameter space, an adjoint model is not available for the complex forward model, the optimization solution requires access to the Hessian and/or to a global view of the cost function (e.g., to identify local minima), and more detailed information about the posterior distribution is required than just its center (e.g., spread).

Acknowledgments. This research was supported by the Office of Naval Research, Award N00014-101-0498, by the U.S. Department of Energy, Award DE-SC008789, and by the Gulf of Mexico Research Initiative Contracts SA1207GOMRI005 (CARTHE) and SA12GOMRI008 (DEEP-C).

APPENDIX

Truncation Strategy

Truncation strategy is concerned with the number of terms to retain in the series to guarantee its accuracy and on the series termination in a multidimensional space (akin to spherical harmonic truncation in spectral atmospheric models or polynomial truncation in finite-element models), that is, the degree of the highest polynomial retained along each uncertain dimension with terms like $\alpha^a V_{\max}^b m^c$. A rectangular (isotropic) truncation uses max $(a, b, c) \leq p$, where p is the maximum degree allowed in any one variable; in this case the total number of terms in the series is $P + 1 = (p + 1)^N$, where N is the number of uncertain variables. A total order truncation would enforce $a + b + c \leq p$ [akin to triangular truncation in finite elements; the triplets $(p, 0, 0)$ and $(0, p, 0)$ and $(0, 0, p)$ would be included for example but not $(1, p, 2)$]; the number of terms would then be $P + 1 = (N + p)!/N!p!$ (Le Maître and Knio 2010). An anistropic truncation would use a different maximum order along each direction because, for example, the lower-order terms are enough to describe the dependency of M on the independent variables (e.g., M varies linearly with m but like the sixth power of α). In the present instance an adaptive strategy was used to decide on the maximum order retained along each direction. The refinement criteria was an area-averaged SST that was monitored to determine which terms in the series were contributing most to its variance; once these terms were identified the series and sampling were refined along the most dominant direction. The adaptive algorithm required 6 refinement steps in the present study, resulting in 67 independent realizations with more sampling along the α parameter as a result of its larger impact on the SST variance. For more details on truncation strategies in the adaptive algorithm please see Gerstner and Griebel (2003), Conrad and Marzouk (2013), Winokur et al. (2013), Sraj et al. (2013), and Constantine et al. (2012).

REFERENCES

Alexanderian, A., J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. Thacker, and O. Knio, 2012: Global sensitivity analysis in an ocean general circulation model: A sparse spectral

projection approach. *Comput. Geosci.,* **16,** 757–778, doi:10.1007/s10596-012-9286-2.

Altaf, M. U., M. ElGharamti, A. W. Heemink, and I. Hoteit, 2013: A reduced adjoint approach to variational data assimilation. *Comput. Methods Appl. Mech. Eng.,* **254,** 1–13, doi:10.1016/j.cma.2012.10.003.

Avriel, M., 1976: *Nonlinear Programming: Analysis and Methods.* Prentice-Hall, 544 pp.

Baur, W., and V. Strassen, 1983: The complexity of partial derivatives. *Theor. Comput. Sci.,* **22** (3), 317–330, doi:10.1016/0304-3975(83)90110-X.

Berger, J., 1985: *Statistical Decision Theory and Bayesian Analysis.* 2nd ed. Springer-Verlag, 617 pp.

Berliner, L. M., R. F. Milliff, and C. K. Wikle, 2003: Bayesian hierarchical modeling of air-sea interaction. *J. Geophys. Res.,* **108,** 3104, doi:10.1029/2002JC001413.

Bernardo, J., and A. F. M. Smith, 2007: *Bayesian Theory.* 2nd ed. Wiley Series in Probability and Statistics, Wiley John & Sons, 640 pp.

Bischof, C., A. Carle, G. Corliss, A. Griewank, and P. Hovland, 1992: ADIFOR-generating derivative codes from Fortran programs. *Sci. Programm.,* **1** (1), 11–29.

Blanchard, E. D., A. Sandu, and C. Sandu, 2010: Polynomial chaos-based parameter estimation methods applied to a vehicle system. *Proc. Inst. Mech. Eng., Part K: J. Multibody Dyn.,* **224,** 59–81, doi:10.1243/14644193JMBD204.

Brent, R. P., 1973: *Algorithms for Minimization without Derivatives.* Courier Dover Publications, 195 pp.

Cao, Y., J. Zhu, I. M. Navon, and Z. Luo, 2007: A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *Int. J. Numer. Methods Fluids,* **53** (10), 1571–1583, doi:10.1002/fld.1365.

Conrad, P. R., and Y. M. Marzouk, 2013: Adaptive Smolyak pseudospectral approximations. *SIAM J. Sci. Comput.,* **35** (6), A2643–A2670, doi:10.1137/120890715.

Constantine, P., M. Eldred, and E. Phipps, 2012: Sparse pseudospectral approximation method. *Comput. Methods Appl. Mech. Eng.,* **229–232,** 1–12.

Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quart. J. Roy. Meteor. Soc.,* **113** (478), 1329–1347, doi:10.1002/qj.49711347813.

——, and ——, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus,* **42A** (5), 531–549, doi:10.1034/j.1600-0870.1990.t01-4-00004.x.

Epstein, E. S., 1985: *Statistical Inference and Prediction in Climatology: A Bayesian Approach.* Amer. Meteor. Soc., 199 pp.

Fletcher, R., 1987: *Practical Methods of Optimization.* 2nd ed. Wiley-Interscience, 436 pp.

Gejadze, I., G. Copeland, F.-X. L. Dimet, and V. Shutyaev, 2011: Computation of the analysis error covariance in variational data assimilation problems with nonlinear dynamics. *J. Comput. Phys.,* **230** (22), 7923–7943, doi:10.1016/j.jcp.2011.03.039.

Gelman, A., J. B. Carlin, H. L. Stern, and D. B. Rubin, 2004: *Bayesian Data Analysis.* 2nd ed. Chapman and Hall/CRC, 668 pp.

Gerstner, T., and M. Griebel, 2003: Dimension-adaptive tensor-product quadrature. *Computing,* **71** (1), 65–87, doi:10.1007/s00607-003-0015-5.

Ghanem, R., and P. Spanos, 2002: *Stochastic Finite Elements: A Spectral Approach.* 2nd ed. Dover, 222 pp.

Giering, R., and T. Kaminski, 1998: Recipes for adjoint code construction. *ACM Trans. Math. Software,* **24** (4), 437–474, doi:10.1145/293686.293695.

Griewank, A., and A. Walther, 2008: *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation.* 2nd ed. Society for Industrial and Applied Mathematics, 438 pp.

Lawless, A. S., S. Gratton, and N. K. Nichols, 2005: Approximate iterative methods for variational data assimilation. *Int. J. Numer. Methods Fluids,* **47** (10–11), 1129–1135, doi:10.1002/fld.851.

Le Dimet, F.-X., I. M. Navon, and D. N. Daescu, 2002: Second-order information in data assimilation. *Mon. Wea. Rev.,* **130,** 629–648.

Le Maître, O. P., and O. M. Knio, 2010: *Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics.* Springer-Verlag, 552 pp.

Li, J., and D. Xiu, 2009: A generalized polynomial chaos based ensemble Kalman filter with high accuracy. *J. Comput. Phys.,* **228** (15), 5454–5469, doi:10.1016/j.jcp.2009.04.029.

Marzouk, Y. M., and H. N. Najm, 2009: Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.,* **228** (6), 1862–1902, doi:10.1016/j.jcp.2008.11.024.

——, ——, and L. A. Rahn, 2007: Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.,* **224** (2), 560–586, doi:10.1016/j.jcp.2006.10.010.

Mattern, P., K. Fennel, and M. Dowd, 2012: Estimating time-dependent parameters for a biological ocean model using an emulator approach. *J. Mar. Syst.,* **96–97,** 32–47, doi:10.1016/j.jmarsys.2012.01.015.

Pence, B. L., H. K. Fathy, and J. L. Stein, 2010: A maximum likelihood approach to recursive polynomial chaos parameter estimation. *Proc. American Control Conference (ACC),* Baltimore, MD, American Automatic Control Council, 2144–2151.

Petras, K., 2003: Smolyak cubature of given polynomial degree with few nodes for increasing dimension. *Numer. Math.,* **93** (4), 729–753, doi:10.1007/s002110200401.

Saad, G., and R. Ghanem, 2009: Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter. *Water Resour. Res.,* **45,** W04417, doi:10.1029/2008WR007148.

Smolyak, S., 1963: Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR,* **4,** 240–243.

Southward, S. C., 2008: Real-time parameter ID using polynomial chaos expansions. *Proc. ASME 2007 Int. Mechanical Engineering Congress and Exposition, Vol. 9: Mechanical Systems and Control, Parts A, B, and C,* Seattle, WA, ASME, 1167–1173, doi:10.1115/IMECE2007-43745.

Sraj, I., and Coauthors, 2013: Bayesian inference of drag parameters using AXBT data from Typhoon Fanapi. *Mon. Wea. Rev.,* **141,** 2347–2366.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.,* **113** (478), 1311–1328, doi:10.1002/qj.49711347812.

Tber, M.-H., L. Hascoët, A. Vidard, and B. Dauvergne, 2007: Building the tangent and adjoint codes of the ocean general circulation model OPA with the automatic differentiation tool tapenade. Research Rep. 6372, INRIA, 28 pp.

Thacker, W. C., 1989: The role of the Hessian matrix in fitting models to measurements. *J. Geophys. Res.,* **94** (C5), 6177–6196.

——, 1991: Automatic differentiation from an oceanographer's perspective. *Automatic Differentiation of Algorithms: Theory,*

Implementation, and Application, A. Griewank and G. F. Corliss, Eds., SIAM, 191–201.

Winokur, J., P. Conrad, I. Sraj, O. Knio, A. Srinivasan, W. Thacker, Y. Marzouk, and M. Iskandarani, 2013: A priori testing of sparse adaptive polynomial chaos expansions using an ocean general circulation model database. *Comput. Geosci.,* **17** (6), 899–911, doi:10.1007/s10596-013-9361-3.

Zabaras, N. J., and B. Ganapathysubramanian, 2008: A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach. *J. Comput. Phys.,* **227** (9), 4697–4735, doi:10.1016/j.jcp.2008.01.019.

Zedler, S., G. Kanschat, R. Korty, and I. Hoteit, 2012: A new approach for the determination of the drag coefficient from the upper ocean response to a tropical cyclone: A feasibility study. *J. Oceanogr.,* **68** (2), 227–241.

Zeng, L., and D. Zhang, 2010: A stochastic collocation based Kalman filter for data assimilation. *Comput. Geosci.,* **14** (4), 721–744, doi:10.1007/s10596-010-9183-5.