

## On the Renormalization of the Covariance Operators

MAX YAREMCHUK AND MATTHEW CARRIER

*Naval Research Laboratory, Stennis Space Center, Mississippi*

(Manuscript received 17 June 2011, in final form 16 August 2011)

### ABSTRACT

Many background error correlation (BEC) models in data assimilation are formulated in terms of a smoothing operator  $\mathbf{B}$ , which simulates the action of the correlation matrix on a state vector normalized by respective BE variances. Under such formulation,  $\mathbf{B}$  has to have a unit diagonal and requires appropriate renormalization by rescaling. The exact computation of the rescaling factors (diagonal elements of  $\mathbf{B}$ ) is a computationally expensive procedure, which needs an efficient numerical approximation.

In this study approximate renormalization techniques based on the Monte Carlo (MC) and Hadamard matrix (HM) methods and on the analytic approximations derived under the assumption of the local homogeneity (LHA) of  $\mathbf{B}$  are compared using realistic BEC models designed for oceanographic applications. It is shown that although the accuracy of the MC and HM methods can be improved by additional smoothing, their computational cost remains significantly higher than the LHA method, which is shown to be effective even in the zeroth-order approximation. The next approximation improves the accuracy 1.5–2 times at a moderate increase of CPU time. A heuristic relationship for the smoothing scale in two and three dimensions is proposed for the first-order LHA approximation.

### 1. Introduction

Modeling of the background error correlation (BEC) by smoothing operators in variational data assimilation has recently gained considerable attention, primarily due to computational efficiency of their implementation and versatility in approximating anisotropic and inhomogeneous BECs (e.g., Xu 2005; Pannekoucke and Massart 2008; Mirouze and Weaver 2010). In the framework of this approach, the correlation matrix  $\mathbf{C}$  of the BE field is approximated by a positive function of the diffusion operator:

$$\mathbf{D} = \nabla \boldsymbol{\nu} \nabla, \quad (1)$$

where  $\boldsymbol{\nu}$  is the diffusion tensor, whose components may depend on the spatial coordinates  $\mathbf{x}$ . Attractive features of such models are their low computational cost and direct control of inhomogeneity and anisotropy by  $\boldsymbol{\nu}$  under the positive-definiteness constraint. Among the most popular functions of  $\mathbf{D}$  used in practice, are the exponential (yielding the Gaussian-shaped correlations)

and the inverse of a positive polynomial. Numerically, these functions are implemented by integrating the finite-difference diffusion equation on the model grid and have been used in many applications (e.g., Derber and Rosati 1989; Weaver et al. 2003; Di Lorenzo et al. 2007; Liu et al. 2009).

An important numerical issue arising with this type of covariance modeling is the necessity to find a rescaling transformation  $\mathbf{R}$  such that the diagonal elements of the covariance matrix are equal to the background error (BE) variances derived from the history of data assimilation into the numerical model. The direct way of finding  $\mathbf{R}$  is to compute the inverse square root of the diagonal elements of  $\mathbf{C}$ . The latter can be found by convolving  $\mathbf{C}$  with  $\delta$  functions in every point  $\mathbf{x}$  of the model grid and taking the results of these convolutions at the same points. This procedure, however, is practically unfeasible as the model grid size (the number of convolutions to be executed) may often exceed  $M \sim 10^6$ – $10^7$ .

In this study we compare the numerical efficiency of several methods of estimating the diagonal elements of a symmetric positive-definite matrix  $\mathbf{C}$  defined by its action on an arbitrary vector and propose an algorithm for estimating  $\mathbf{R}$  under the locally homogeneous approximation (i.e., slowly varying smoothing scales of  $\mathbf{C}$ ). The manuscript is organized as follows. In section 2, we briefly

*Corresponding author address:* Max Yaremchuk, Naval Research Laboratory, Bldg. 1009, Stennis Space Center, MS 39529.  
E-mail: max.yaremchuk@nrlssc.navy.mil

overview the diagonal estimation methods with an emphasis on the possibility to exploit the analytic formulas, available for the considered functional forms under the assumption of spatial homogeneity. In particular, an extension of the method of Purser et al. (2003) for estimating the Gaussian kernel diagonal is proposed for higher dimensions. In section 3, we describe the setup of numerical experiments with anisotropic inhomogeneous correlation model and document the results. Summary and conclusions complete the paper.

## 2. Renormalization methods

### a. Monte Carlo technique

This method originates from a large family of stochastic algorithms used for estimating elements and traces of extra-large matrices (e.g., Girard 1989; Hutchison 1989; Dong and Liu 1994). Weaver and Courtier (2001) were among the first to use this approach in geophysical applications for estimating the diagonal of the Gaussian-shaped BEC operators.

The underlying idea is to define an ensemble of  $K$  random vectors  $\mathbf{s}_k$  on the model grid and perform componentwise averaging of the products  $\tilde{\mathbf{s}} = \mathbf{C}\mathbf{s}$  according to the following formula:

$$\tilde{\mathbf{d}}(\mathbf{x}) = \overline{\mathbf{s} \odot \tilde{\mathbf{s}} \oslash \overline{\mathbf{s} \odot \mathbf{s}}}, \quad (2)$$

where the overbar denotes averaging over the ensemble, and  $\odot$  and  $\oslash$  stand for the componentwise multiplication and division of the vectors, respectively. Simple considerations show that when all the components of  $\mathbf{s}$  have identical  $\delta$ -correlated distributions with zero mean, contributions to  $\tilde{\mathbf{d}}$  from the off-diagonal elements tend to cancel out, and  $\tilde{\mathbf{d}}$  converges to  $\mathbf{d} = \text{diag}(\mathbf{C})$  as  $K \rightarrow \infty$ . More accurately, the squared relative approximation error:

$$\varepsilon^2(\mathbf{x}) = (\tilde{\mathbf{d}} - \mathbf{d})^2 / \mathbf{d}^2 \quad (3)$$

is inversely proportional to the ensemble size  $K$ . In other words, one may expect to achieve 10% accuracy at the expense of approximately 100 multiplications by  $\mathbf{C}$  if the first ensemble member gives a 100% error. This estimate may seem acceptable since in geophysical applications the BE variances are usually known with limited precision and approximating the diagonal with 5%–10% error seems satisfactory.

### b. The Hadamard matrix method

The Monte Carlo (MC) technique was developed further by Bekas et al. (2007), who noticed that the method may converge to  $\mathbf{d}$  in the *finite* number of iterations  $M$  if the ensemble vectors are mutually orthogonal. An easy

way to construct such an ensemble is to draw the vectors  $\mathbf{s}_k$  from the columns of the  $M \times M$  Hadamard matrix (HM), which span the model's state space (see appendix A for more details on the HM theory).

Although it is not proven yet that Hadamard matrices can be constructed for an arbitrary  $M$ , very efficient recursive algorithms for generating HM columns do exist for  $M$ s, whose factorization involves only prime numbers not exceeding 100. Since the exact convergence (which can be achieved at  $M$  iterations) is never needed in practice, it is not even necessary to draw  $\mathbf{s}_k$  from the Hadamard matrix, whose dimension exactly coincides with the state space dimension: if generation of  $\mathbf{s}_k$  is impossible because of some odd value of  $M$  (e.g., 1004), it can always be replaced by a slightly larger number (e.g., 1008), and  $\mathbf{s}_k$  can be defined as the first 1004 components of the 1008-dimensional Hadamard vectors. In the numerical experiments described in the next section it is shown that such modification does not affect the method's convergence during the first several hundred iterations.

### c. Locally homogeneous approximations

An alternative approach to the diagonal estimation employs a priori information on the structure of  $\mathbf{C}$ . Consider a homogeneous ( $\boldsymbol{\nu} = \text{const}$ ) case and assume that the coordinate axes are aligned along the eigenvectors of the diffusion tensor, whose (positive) eigenvalues are  $\lambda_i^2, i = 1, \dots, n$  and  $n$  is the number of dimensions of the physical space (1, 2, or 3). Then the matrix elements of the two types of BEC operators that are considered here can be written down explicitly (see appendix B) as

$$\mathbf{C}_1(\mathbf{x}, \mathbf{y}) = \exp(\mathbf{D}/2) = d \exp\left[\frac{-\rho^2}{2}\right], \quad (4)$$

$$\mathbf{C}_2(\mathbf{x}, \mathbf{y}) = (\mathbf{I} - \mathbf{D}/2m)^{-m} = d \frac{\bar{\rho}^s K_s(\bar{\rho})}{2^{s-1} \Gamma(s)}, \quad (5)$$

where

$$\rho = \sqrt{(\mathbf{x} - \mathbf{y})^T \boldsymbol{\nu}^{-1} (\mathbf{x} - \mathbf{y})}$$

is the distance between the correlated points, measured in terms of the smoothing scales  $\lambda_i$ ;  $d = (2\pi)^{-n/2} \Omega^{-1}$  are the (constant) diagonal elements of  $\mathbf{C}_{1,2}$ ;  $\Omega = \Pi \lambda_i = \sqrt{\det \boldsymbol{\nu}}$  is the diffusion volume element;  $\mathbf{I}$  is the identity operator;  $K$  and  $\Gamma$  denote the Bessel function of the second kind and the gamma function, respectively;  $m$  is a positive integer (an approximation order of  $\mathbf{C}_1$  by  $\mathbf{C}_2$ );  $s = m - n/2$ ; and  $\bar{\rho} = \sqrt{2m\rho}$ . The parameter  $m$  can also be interpreted as the number of "time steps" used in discrete-time integration of the corresponding diffusion equation by the implicit scheme (see appendix B).

When  $\nu$  varies in space, (4)–(5) are no longer valid, and the diagonal elements  $\mathbf{d}$  depend on  $\mathbf{x}$  and on the type of the BEC operator. However, if we assume that  $\nu$  is locally homogeneous (LH) (i.e., varies in space on a typical scale  $L$ , which is much larger than  $\lambda_i$ ) the diagonal elements  $\mathbf{d}(\mathbf{x})$  can be expanded in the powers of the small parameter  $\epsilon = \bar{\lambda}/L$ , where  $\bar{\lambda}$  is the mean eigenvalue of  $\sqrt{\nu}$ . The zeroth-order LH approximation term (LH0) is apparently

$$\mathbf{d}^0(\mathbf{x}) = (2\pi)^{-n/2} \Omega(\mathbf{x})^{-1} \tag{6}$$

because for infinitely slow variations of  $\nu$  ( $L \rightarrow \infty$ ) the normalization factors must converge to the above expression for the constant diagonal elements  $d$ . It is noteworthy that (6) was found to be useful even in the case of strong inhomogeneity  $\epsilon \geq 1$ . In particular, Mirouze and Weaver (2010) found that such an approximation provided 10% errors in a simplified 1D case.

The accuracy of (6) can formally be increased by considering the next term in the expansion of the diagonal elements of  $\mathbf{C}_{1,2}$ . The technique of such asymptotics has been well developed for the diagonal of the Gaussian kernel (4) in Riemannian spaces (e.g., Gusynin and Kushnir 1991; Avramidi 1999). More recently, the approach was considered by Purser (2008a,b) in the atmospheric data assimilation context. Application of this technique to the diffusion operator (1) in flat space yields the following asymptotic expression for the diagonal elements of  $\mathbf{C}_1$  in the local coordinate system, where  $\nu(\mathbf{x})$  is equal to the identity matrix and  $\mathbf{D}$  takes the form of the Laplacian operator  $\Delta$ :

$$\mathbf{C}_1(\mathbf{x}, \mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \left[ 1 - \frac{1}{2} \text{tr} \mathbf{h} - \frac{1}{12} \left( \frac{\Delta}{2} \text{tr} \mathbf{h} + \nabla \cdot \text{div} \mathbf{h} \right) \right] + O(\epsilon^5). \tag{7}$$

Here  $\mathbf{h}$  is a small ( $|\mathbf{h}| \sim \epsilon$ ) correction to  $\nu$  in the vicinity of  $\mathbf{x}$ . Note that terms in the round brackets have the order  $O(\epsilon^3)$  because each spatial differentiation adds an extra power of  $\epsilon$ .

The asymptotic estimate (7) involves second derivatives that tend to amplify errors in practical applications when  $\epsilon$  may not be small. Therefore, using (7) in its original form could be inaccurate even at a moderately small value of  $\epsilon$ . To increase the computational efficiency, it is also desirable to formulate the first-order approximation as a linear operator, acting on  $\mathbf{d}^0(\mathbf{x})$ . Keeping in mind that  $|\mathbf{h}| \sim \epsilon$ , we can utilize the following relationships:

$$\mathbf{d}^0(\mathbf{x}) = (2\pi)^{-n/2} \Omega(\mathbf{x})^{-1} \approx (2\pi)^{-n/2} \left( 1 - \frac{1}{2} \text{tr} \mathbf{h} \right), \tag{8}$$

$$\exp(\Delta/2) \approx \mathbf{I} + \frac{1}{2} \Delta, \tag{9}$$

and represent the second term in the round brackets of (7) as follows:

$$\nabla \cdot \text{div} \mathbf{h} = \frac{1}{n} \Delta \text{tr} \mathbf{h} + \nabla \cdot \text{div} \mathbf{h}', \tag{10}$$

where  $\mathbf{h}'$  is the traceless part of  $\mathbf{h}$ . On the other hand, if we neglect the divergence of  $\mathbf{h}'$ , (7) can be rewritten in the following form:

$$\mathbf{C}_1(\mathbf{x}, \mathbf{x}) \approx \frac{1}{(2\pi)^{n/2}} \left( 1 + \gamma_n \frac{\Delta}{2} \right) \left( 1 - \frac{1}{2} \text{tr} \mathbf{h} \right), \tag{11}$$

where

$$\gamma_n = \frac{1}{6} + \frac{1}{3n}. \tag{12}$$

Taking (8)–(9) into account and replacing  $\Delta$  by  $\mathbf{D}$ , we finally get the desired ansatz for the first-order approximation (LH1) of the diagonal:

$$\mathbf{d}^1 = \exp\left(\gamma_n \frac{\mathbf{D}}{2}\right) \mathbf{d}^0. \tag{13}$$

The relationship (13) was derived by Purser et al. (2003) for one-dimensional case ( $\gamma_1 = 0.5$ ) and tested by Mirouze and Weaver (2010), who reported a significant (2–4 times) improvement of the accuracy in 1D simulations.

It is likely that the estimate similar to (13) can also be obtained for  $\mathbf{C}_2$ , possibly with a different coefficient  $\tilde{\gamma}_n$ . We assume, however, that  $\tilde{\gamma}_n$  may not differ too much from  $\gamma_n$  given similarity in the shapes of the correlation functions (4)–(5) (Fig. 1). Furthermore, because of the approximate nature of (13), the best representation of  $\mathbf{d}(\mathbf{x})$  in realistic applications may be achieved with a value of  $\gamma_n$  significantly different from the one given by (12). For that reason, in the numerical experiments we adopt a more general form of (13), assuming

$$\mathbf{d}_1^1(\mathbf{x}) \approx \exp[\gamma \mathbf{D}/2] \mathbf{d}_1^0(\mathbf{x}); \quad \mathbf{d}_2^1(\mathbf{x}) \approx [\mathbf{I} - \gamma \mathbf{D}/4]^{-2} \mathbf{d}_2^0(\mathbf{x}), \tag{14}$$

and investigate the dependence of the respective approximation errors ( $\epsilon_{1,2}$ ) on the free parameter  $\gamma$ .

### 3. Numerical results

To assess the efficiency of the methods outlined in the previous section, two series of numerical experiments with realistically inhomogeneous BEC models are performed. In the first series we test the methods in the 2D

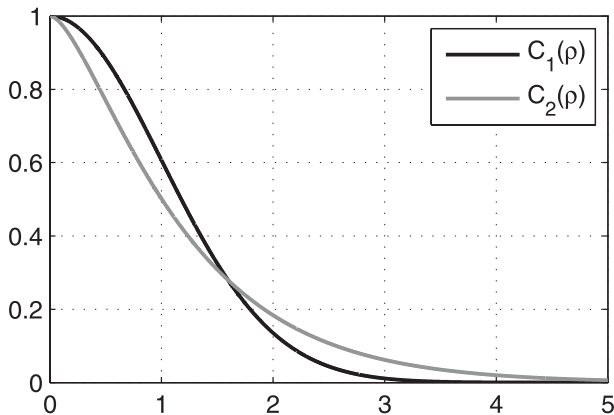


FIG. 1. Correlation functions corresponding to the homogeneous operators (4)–(5) with identical decorrelation scales.

case with the state vector having the dimension of several thousand. In the second series, the LH0 and LH1 techniques are examined in a realistic 3D setting with a state space dimension of  $\sim 10^6$ .

#### a. Experimental setting in 2D

The state space is described by scalar functions defined on the orthogonal curvilinear grid of the Navy Coastal Ocean Model (NCOM; Martin et al. 2009) set up in the Monterey Bay (Fig. 2). The number  $M$  of grid points (dimension of the state space) was 3438. A vector field  $\mathbf{v}(\mathbf{x})$  was used to generate the diffusion tensor as follows. The smaller principal axis  $\lambda_2$  of  $\sqrt{\mathbf{v}}$  is set to be orthogonal to  $\mathbf{v}$  with the corresponding “background” length scale  $\lambda_2 = 3\delta$ , where  $\delta(\mathbf{x})$  is the spatially varying grid step. The length of the larger axis  $\lambda_1$  is set to be equal to  $\max(1, \sqrt{|\mathbf{v}|/v})\lambda_2$ , where  $v$  is a prescribed threshold value of  $|\mathbf{v}|$ . If  $\mathbf{v}$  is a velocity field, then a structure like this simulates enhanced diffusive transport of model errors in the regions of strong currents on the background of isotropic error diffusion with the decorrelation scale  $\lambda_2$ .

In the 2D experiments, the vector field  $\mathbf{v}$  is generated by treating bottom topography  $h(\mathbf{x})$  (Fig. 2) as a streamfunction. The threshold value  $v$  was taken to be one-fifth of the rms variation of  $|\nabla h|$  over the domain.

All the experiments described in sections 3b–3e, are performed using the BEC models (4)–(5) with the parameters  $n = m = 2$ . Composite maps of five columns of the corresponding BEC operators are shown in Figs. 2a,b. The diffusion operator (1) is constrained to have a zero normal derivative at the open and rigid boundaries of the domain in both 2D and 3D experiments.

Numerically, the action of the Gaussian-shaped BEC operator  $\mathbf{C}_1$  on a state vector  $\mathbf{y}_0$  was evaluated by explicitly integrating the corresponding diffusion equation  $\mathbf{y}_t = \mathbf{D}/2\mathbf{y}$  for the virtual “time period” defined by  $\nu$

starting from the “initial condition”  $\mathbf{y}_0$ . The minimum number of “time steps” required for the scheme’s stability in such setting was 5256. The action of  $\mathbf{C}_2$  was computed by solving the system of equations  $(\mathbf{I} - \mathbf{D}/4)^2 \mathbf{y} = \mathbf{y}_0$  with a conjugate gradient method. The number of iterations, required for obtaining a solution, varied within 2000–2500. To make the shapes of the  $\mathbf{C}_1$  and  $\mathbf{C}_2$  compatible (Fig. 1), the diffusion tensor in  $\mathbf{C}_2$  was multiplied by  $8/\pi$  (Yaremchuk and Smith 2011).

The exact values  $\mathbf{d}(\mathbf{x})$  of the diagonal elements are shown in Fig. 2 (right panel). Their magnitude appears to be lower in the regions of “strong currents” (large  $\mathbf{v}$ ), as the corresponding  $\delta$  functions are dispersed over larger areas by diffusion. The  $\mathbf{d}(\mathbf{x})$  are higher near the boundaries because part of the domain available for dispersion is screened by the condition prescribing zero flux across either the open or rigid boundary.

#### b. Monte Carlo technique

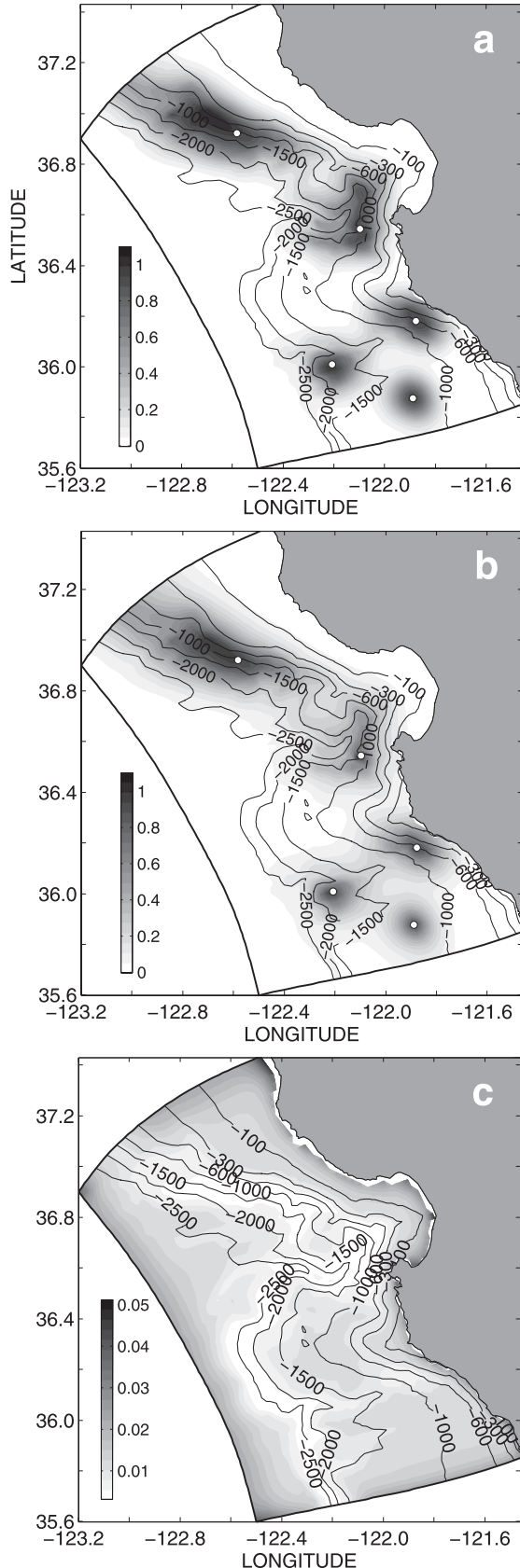
The MC method is implemented in two ways: in the first series of experiments, the components of  $\mathbf{s}_k$  are taken to be either 1 or  $-1$  with equal probability; in the second series they are drawn from the white noise on the interval  $[-1, 1]$ . The residual error  $\varepsilon$  is computed using (3). In both series the rates of reduction of  $\varepsilon$  with iteration  $k$  were similar and closely followed the  $\sqrt{k}$  law (thin solid line in Fig. 3c).

Figure 3a shows the distribution of  $\varepsilon(\mathbf{x})$  after 60 iterations of the MC method with the  $\mathbf{C}_2$  BEC model. The estimate is apparently affected by sampling noise, which can be identified by fine structures at scales close to the grid spacing.

To improve the accuracy, the MC estimates are low-pass filtered with the corresponding BEC operators at every iteration. To optimize the filter, the diffusion operator in  $\mathbf{C}$  is multiplied by the tunable parameter  $\gamma$ , which effectively reduced the mean decorrelation (smoothing) scale  $\kappa = \gamma^{-1/2}$  times. Figure 3b demonstrates the result of such optimal smoothing ( $\kappa_{\text{opt}} = 2.5$ ), which enabled almost fourfold reduction of the domain-averaged error  $\langle \varepsilon \rangle$  to 0.15. The right panel in Fig. 3 shows the evolution of  $\langle \varepsilon \rangle$  with iterations for the  $\mathbf{C}_2$  operator. The optimal smoothing scale reduction factor  $\kappa^{-1}$  appears to follow the law  $\kappa^{-1}(k) \sim k^{-1/3}$  (dashed line in Fig. 3c). It is remarkable that significant error reduction occurs even if the smoothing scale falls below the grid spacing  $\kappa^{-1} < 1/3$ .

#### c. The Hadamard matrix method

Experiments with the Hadamard scheme are also done in two series. Since the value of  $M = 3438$  is not divisible by 4, it is hard to find the HM of that dimension. Instead, in the first series of experiments,  $\mathbf{s}_k$  were specified as the first 3438 numbers taken from the columns of



the 3456-dimensional HM, which can be easily constructed from the 12-dimensional HMs.

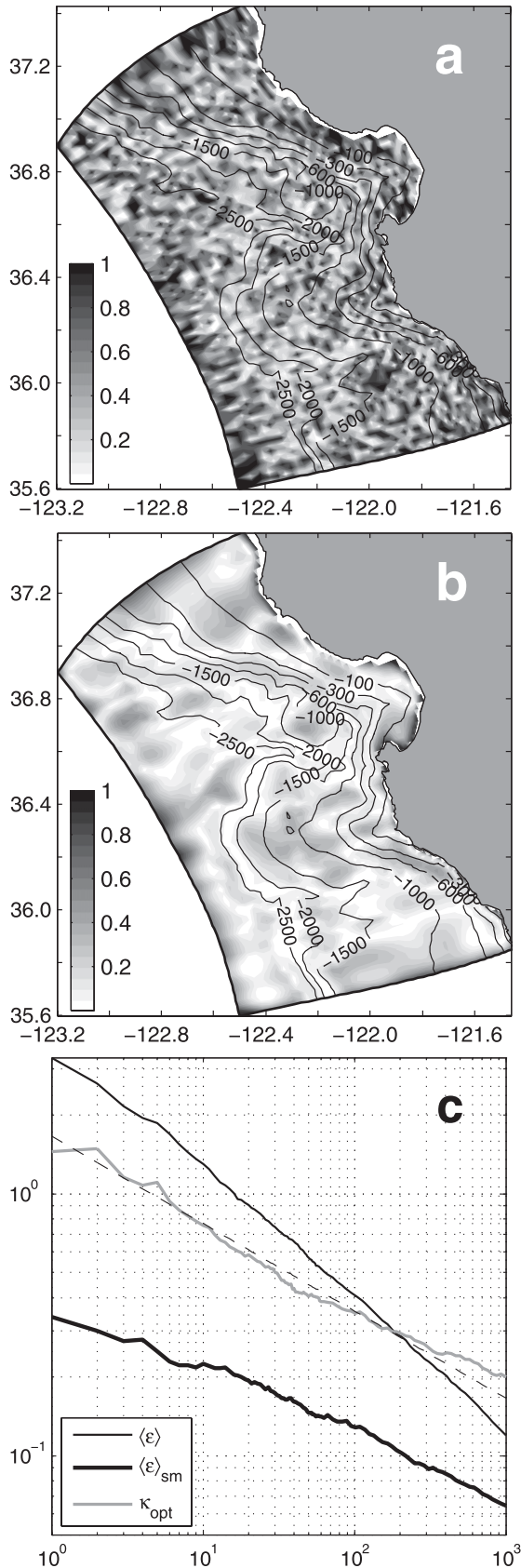
To check the impact of “nonexact” dimension on the convergence, we artificially increased  $M$  to 3456 by removing 18 land points in the domain. Experiments with the HM of exact dimension show that differences in convergence between the nonexact and exact experiments start to be visible only after several hundred iterations. After  $M$  iterations the error of the exact-HM method is reduced to the machine precision, while the error of the first series of experiments stumbled at approximately  $10^{-3}$  after 1500 iterations. This is consistent with the  $18/3438 \sim 0.5\%$  degree of nonorthogonality of  $\mathbf{s}_k$  drawn from the nonexact HM.

Similar to the MC method, the accuracy of HM estimates are improved significantly after smoothing. In addition it is found that the effect of smoothing can be enhanced if the computer mapping of the 2D model field on the 1D vector is randomized: Fig. 4a bears an apparent trace of columnwise numbering of the model field, which remains visible even after applying the algorithm, generating the HM columns. As a consequence, error distribution in Fig. 4a contains large scales in cross-shore direction, which tend to make the smoothing algorithm less effective (Fig. 4c, dashed curves). This drawback can be easily corrected by randomization of the above-mentioned map (the randomized HM and the RHM method), which provides an error pattern similar to Fig. 3a, but with somewhat lower value of  $\langle \epsilon \rangle$  clearly visible in Fig. 4c, where dashed lines show evolution of  $\langle \epsilon \rangle$  for the straight HM method before (upper line) and after smoothing, while the solid black lines show similar quantities for the RHM method.

Comparison with the MC method (gray curves in Fig. 4c) demonstrate a noticeable advantage of the HM technique (upper curves), which remains visible at higher iterations  $n > 100$  even after smoothing (lower curves). This advantage increases with iterations for two reasons: the HM method converges faster than  $k^{-1/2}$  by its nature, whereas the efficiency of smoothing (targeted at removing the small-scale error constituents) degrades as the signal-to-noise ratio of the diagonal estimates increases with  $k$ .

From the practical point of view, it is not reasonable to do more than several hundred iterations, as  $\langle \epsilon \rangle$  drops to

FIG. 2. Five columns of the BEC operators used in the study: (a) Gaussian-shaped correlations and (b) their approximation by the inverse of the second-order polynomial in  $\mathbf{D}$ . White circles denote locations of the diagonal elements of the corresponding correlation matrices. (c) The map of nonnormalized diagonal elements of  $\mathbf{C}_1$ . Depth contours are in m.



the value of a few percent (Fig. 3c), which is much smaller than the accuracy in the determination of the background error variances. We may therefore conclude that it is advantageous to use the RHM technique when  $k \approx 100$ . In this case (assuming that  $k \ll M$ ), utilization of the HM with exact dimension is not necessary, as it does not affect the convergence.

#### d. Asymptotic expansion method

Since the principal axes of the diffusion tensor at every point are defined by construction, computation of the zeroth-order approximation (6) to the normalization factors is not expensive. Near the boundaries, however, the factors described by (6) have to be adjusted by taking into account the geometric constraints imposed on the diffusion. In this study, this adjustment was computed for points located closer than  $3\lambda_1$  from the boundary. It is assumed that the boundary had negligible impact on the shape of the diffused  $\delta$  function (Fig. 5), so the normalization factor near the boundary was computed by dividing the reference factor (obtained by convolving the BEC operator with the  $\delta$  function in the “open sea”) by the respective integral over the land-free subdomain  $S'$  shown in Fig. 5:

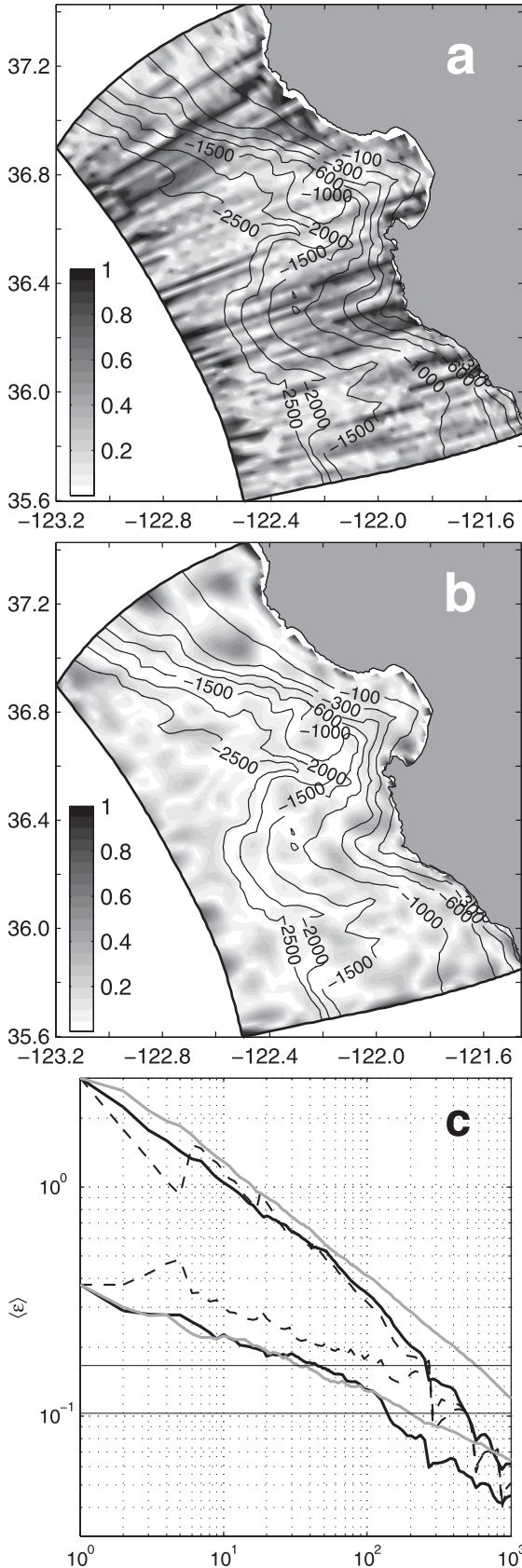
$$\mathbf{d}^0(\mathbf{x}) = \frac{1}{2\pi\Omega(\mathbf{x})} \left[ \int_{S'} \mathbf{C}(\mathbf{x}, \mathbf{y}) \delta(\mathbf{x} - \mathbf{y}) d\mathbf{y} \right]^{-1}. \quad (15)$$

Integrals in the rhs of (15) have to be taken numerically for all the near-boundary points  $\mathbf{x}$ . To speed up the computations, we adopted the LH assumption near the boundaries, and replaced the convolutions in (15) by the respective analytical functions (4)–(5) with the fixed value  $\nu(\mathbf{x})$ . It is necessary to note that the assumption underlying (15) is not exact for the zero normal gradient condition in use (e.g., see Mirouze and Weaver 2010). However, the errors, caused by neglecting distortions introduced by the zero-flux conditions are significantly smaller (3%–7%, see Fig. 5) than the accuracy of the LH assumption itself. These errors could be removed, for example, by introducing “transparent” conditions (Mirouze and Weaver 2010).

Figure 6 demonstrates horizontal distribution of the error  $\epsilon(\mathbf{x})$  obtained by approximating the diagonal elements of  $\mathbf{C}_1$  with (6) (i.e., the zeroth-order LH method, or LH0)

←

FIG. 3. (a) Error distribution after 60 iterations of the MC method, (b) its optimally smoothed version, (c) the respective dependences of the domain-averaged error  $\langle \epsilon \rangle$  and  $\kappa_{\text{opt}}^{-1}$  on the iteration number  $k$ . The dashed line is the approximation of  $\kappa_{\text{opt}}^{-1}$  by the  $k^{-1/3}$  law.



and with (13) (i.e., the first-order LH method, or LH1). Despite an apparent violation of the LH assumption in many regions (e.g.,  $\lambda_1$  changes from  $20\delta$  to the background value of  $3\delta$  at distances  $L \sim 5-6\delta < \lambda_1$  across the shelf break), the mean approximation error of the diagonal elements appears to be relatively small (19%) for the LH0 method, with most of the maxima confined to the regions of strong inhomogeneity (Fig. 6a). The next approximation (Fig. 6b) reduces  $\langle \varepsilon \rangle$  to 9%. Numerical experiments with the  $\mathbf{C}_2$  model have shown similar results (16% and 10% errors).

Another series of experiments are performed with the varying scaling parameter  $\gamma$  to find an optimal fit to  $\mathbf{d}$ . Computations were made for  $0 \leq \gamma \leq 1$ . The best result for the Gaussian BEC was obtained for  $\gamma_2 = 0.30$ , which is fairly consistent with the value ( $\gamma_2 = 0.33$ ) given by (12). In the case of  $\mathbf{C}_2$  operator, the optimal value is  $\gamma_2 = 0.24$ , still in a reasonable agreement with (12), given the strong inhomogeneity of  $\nu$  and deviation of the BEC operator from the Gaussian form. A somewhat smaller value of  $\gamma_2(\mathbf{C}_2)$  can be explained by the sharper shape of the respective correlation function at the origin (Fig. 1), which renders  $\mathbf{d}^0$  to be less dependent on the inhomogeneities in the distribution of  $\nu$ , and, therefore, requiring less smoothing in the next approximation.

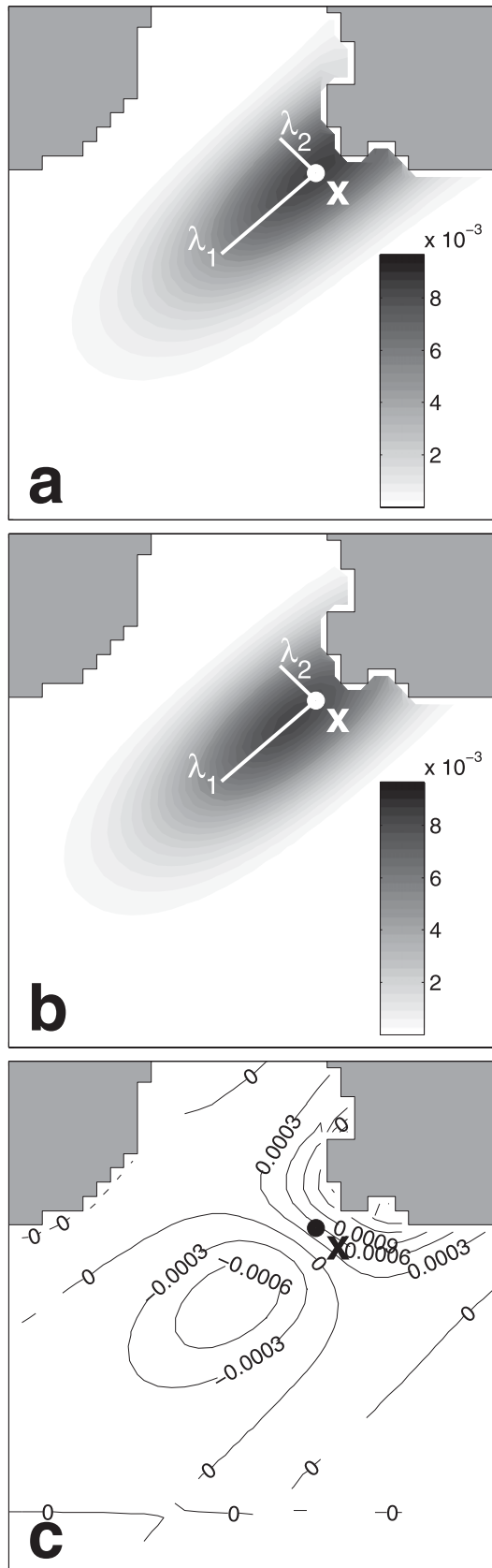
*e. Numerical efficiency*

Table 1 provides an overview of the performance for the tested methods. For comparison purposes we show CPU requirements by the smoothed MC and RHM methods after they achieve the accuracies of the LH0 and LH1 methods. It is seen that both MC and RHM methods are 300–1000 times more computationally expensive than the LH technique. In fact, for the 2D case considered, the computational cost of the stochastic methods is similar to that of the three-dimensional variational data assimilation (3DVAR) analysis itself.

Comparison of the spatial distributions of the approximation error  $\langle \varepsilon \rangle(\mathbf{x})$  favor the LH methods as well: they do show significantly less small-scale variations and may have a potential for further improvement.

Comparing Figs. 3b, 4b, and 6b shows that, in contrast to the MC and RHM methods, LH0 errors tend to increase

FIG. 4. (a) Error distribution after 60 iterations of the HM method. (b) As in (a), but for the smoothed RHM method. (c) Reduction of the domain-averaged error  $\langle \varepsilon \rangle$  with iterations for the MC (gray), RHM (solid), and HM (dashed) methods. The bottom graphs are obtained after optimal smoothing of the diagonal estimates. Thin horizontal lines show error levels provided by the LH0 ( $\langle \varepsilon \rangle = 0.17$ ) and LH1 methods ( $\langle \varepsilon \rangle = 0.10$ ).



in the regions of strong inhomogeneity, but they decrease substantially after smoothing by the LH1 algorithm. At the same time, the LH1 errors tend to have relatively higher values near the boundaries; the effect is less visible in the MC and RHM patterns (Figs. 3b and 4b). This feature can be partly attributed to certain inaccuracy in the algorithm for analytic estimation of the near-boundary elements (Fig. 5c).

Table 1 also shows that LH methods outperform both MC and HM techniques. Although considerable resources are required to compute near-boundary integrals for the adjustment factors (15), the overall CPU savings are quite remarkable.

#### f. LH experiments in the 3D setting

To check the performance of the LH0 and LH1 methods further, a larger 3D domain is taken from the NCOM setup in the Okinawa region south of Japan (Fig. 7), with horizontal resolution of 10 km and 45 vertical levels. The state vector dimension  $M$  (total number of the grid points) in this setting was 862 992.

Because of the large  $M$ , it is computationally unfeasible to directly compute all the diagonal elements of the BEC matrix. Therefore, accuracy checks are performed on a subset of 10 000 points, randomly distributed over the domain and the value of  $\langle \epsilon \rangle$  is estimated by averaging over these points.

The diffusion tensor is constructed in the way described in section 3a, but the generating field  $\mathbf{v}(\mathbf{x})$  is taken to be the horizontal velocity field from an NCOM run. The value of  $\lambda_3$  (in the vertical direction) is independent of horizontal coordinates, but varied in the vertical as  $3\delta_z$ , where  $\delta_z$  is the vertical grid step. Figure 7 illustrates spatial variability of the  $\mathbf{C}_1$  diagonal elements at  $z = 20$  m. The smallest values are observed in the regions of Kuroshio and the North Equatorial Current, where the largest velocities are observed, and the  $\Omega = \sqrt{\det \mathbf{v}}$  reaches the largest values [(6)]. To better test the algorithm, a relatively small threshold value  $\nu = 0.02 \text{ m s}^{-1}$  is prescribed, so that diffusion is anisotropic in more than 90% of the grid points.

Figure 8 demonstrates the accuracy of LH0 and LH1 methods in such setting: the LH0 method provides an

FIG. 5. Adjustment of the normalization factors near the boundary. (a) Map of the  $x$ th column of  $\mathbf{C}_1$  obtained numerically by integrating (15) over the real domain  $S'$  (the normalization factor is the map value at point  $\mathbf{x}$ ). (b) Approximation to (a) obtained by integrating  $\mathbf{C}_1$  in the infinite domain and renormalizing the result to have the same integral over  $S'$  as in (a). (c) The difference between (a) and its approximation (b).



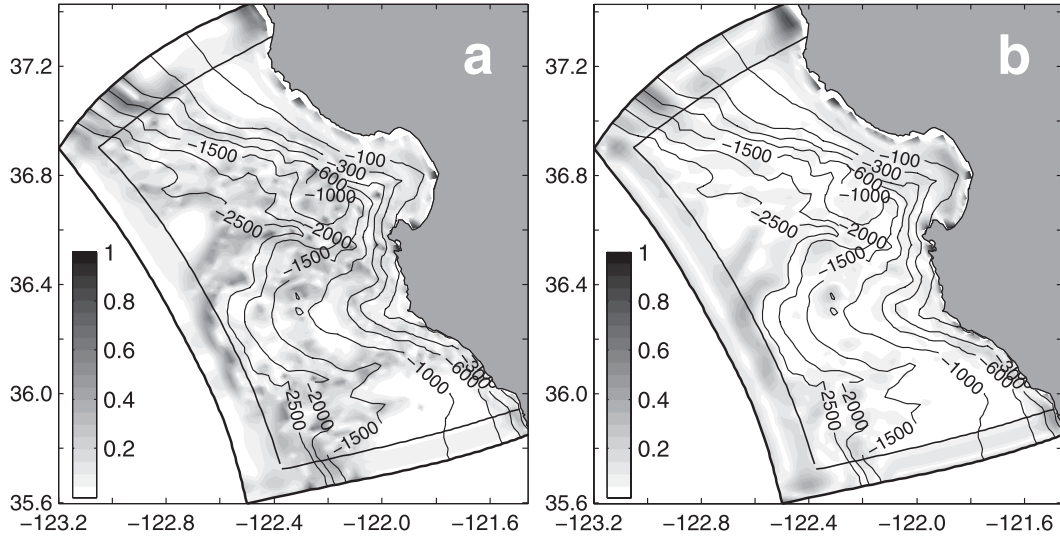


FIG. 6. Diagonal approximation errors under the (a) zeroth-order and (b) first-order LH methods for the Gaussian BEC model.

accuracy of 9%, which is further improved to 6% by the LH1 scheme. The major improvement occurs in the regions where points with highly anisotropic  $\nu$  neighbor isotropic points and reduce the diagonal elements in the latter. The effect is reflected by the negative bias of the scatterplot at high values of  $\mathbf{d}^0$ , which reach their maximum of 0.0237 in the points with isotropic  $\nu$  (Fig. 8a).

Figure 8c shows the dependence of approximation error  $\varepsilon$  on the value of  $\gamma_3$  for both correlation models. The best approximation is obtained at  $\gamma_3 = 0.26$ , a value somewhat lower than suggested by the heuristic formula ( $\gamma_3 = 5/18 = 0.28$ ). Similarly to the 2D case, the optimal value of  $\gamma_3(\mathbf{C}_2) = 0.21$  is less than  $\gamma_3(\mathbf{C}_1)$ , in agreement with more rapid off-diagonal decay of the  $\mathbf{C}_2$  matrix elements.

In general, it appears that the relationship (12) provides a reasonable guidance to the estimation of the smoothing parameter in the LH1 method. For the  $\mathbf{C}_1$  model, the operator acting on  $\mathbf{d}^0$  can be implemented by either reducing  $\gamma^{-1}$  times the number of time steps in the integration of the diffusion equation, or by  $\gamma^{-1/2}$ -fold reduction of the decorrelation radius. For the  $\mathbf{C}_2$  model only the second option is applicable: it would also reduce the number of iterations required for computing the action of the BEC operator.

TABLE 1. Relative CPU times required by the MC and RHM methods to achieve the accuracies ( $\varepsilon$ ) of the LH0 and LH1 methods (shown in parentheses).

	MC/LH0	MC/LH1	RHM/LH0	RHM/LH1
$\mathbf{C}_1$	755 (0.19)	1205 (0.09)	680 (0.19)	520 (0.09)
$\mathbf{C}_2$	780 (0.17)	490 (0.10)	850 (0.17)	330 (0.10)

#### 4. Summary and discussion

In this study we examined the computational efficiency of several techniques used for estimating the diagonal elements of the two types BEC operators: with the Gaussian-shaped kernel  $\mathbf{C}_1$  and with the kernel generated by the second-order polynomial approximation to  $\mathbf{C}_1$ . The considered techniques include the “stochastic” MC and HM methods, which retrieve  $\text{diag}(\mathbf{C})$  only from its action on a vector, and the “deterministic” scheme based on the analytic diagonal expansion under the assumption of local homogeneity of the diffusion tensor. The deterministic scheme was tested in two

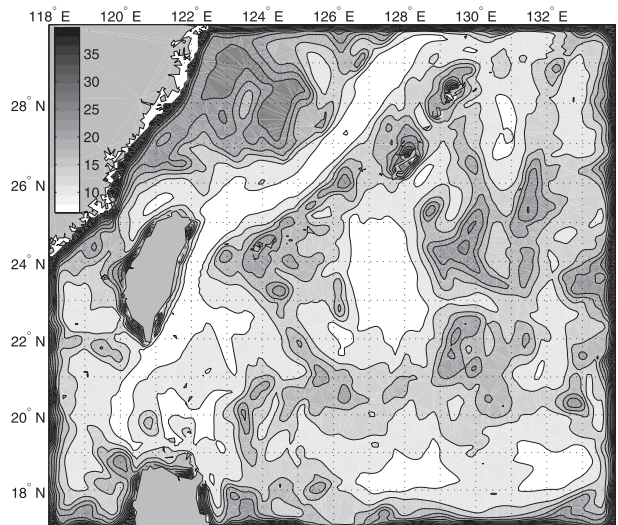
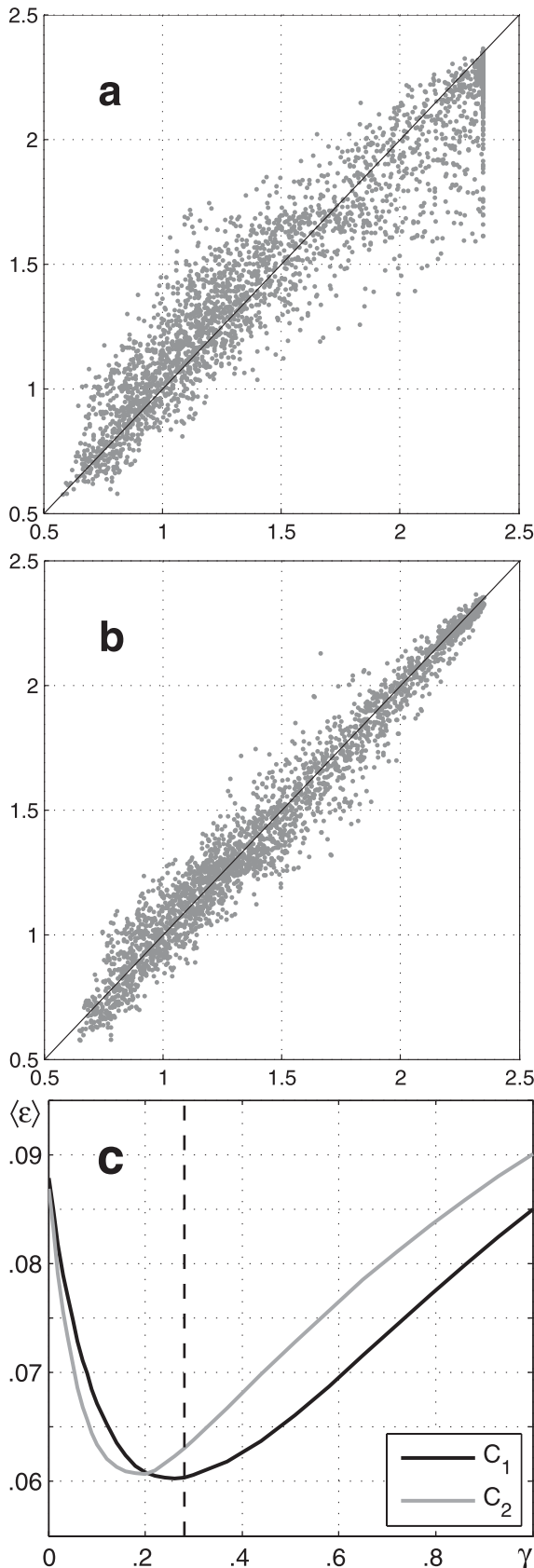


FIG. 7. Diagonal elements of  $\mathbf{C}_1$  in the Okinawa region at  $z = 20$  m. The actual values are multiplied by  $10^4$ .



regimes: the zeroth (LH0) and the first-order (LH1) approximations.

Numerical experiments conducted with realistic diffusion tensor models have shown that (i) HM technique proves to be superior in efficiency compared to the MC technique when accuracies of less than 10% ( $k > 100$ ) are required; (ii) both stochastic methods require 300–1000 times more CPU time to achieve the accuracy, compatible with the most efficient LH1 method; (iii) with the Gaussian model, the LH1 method demonstrates the best performance with the value of the smoothing parameter  $\gamma$  compatible with the one given by the relationship (12) derived from the asymptotic approximation of the Gaussian kernel diagonal.

In deriving the ansatz (13) for the LH1 model we followed the approach of Purser et al. (2003), who proposed to smooth the zeroth-order diagonal by the square root of the BEC operator in a one-dimensional case. Using the asymptotic technique for the heat kernel expansion, we obtained a formula for higher dimensions, and tested its validity by numerical experimentation.

It should be noted that the formal asymptotic expansion (7) is local by nature and tends to diverge in practical applications, where spatial variations of the diffusion tensor may occur at distances  $L$  comparable with the typical decorrelation scale  $\bar{\lambda}$ . To effectively immunize the expansion from the ill effects of the abrupt changes in  $\nu$ , we utilized a nonlocal empirical modification, still fully consistent with the original expansion in the limit  $\bar{\lambda}/L \rightarrow 0$ , but sufficiently robust with respect to the errors related to the high-order derivatives of  $\nu$ . A similar technique was developed by Purser (2008a,b), who used empirical saturation functions to stabilize the higher-order approximations of the heat kernel diagonal.

In general, results of our experiments show high computational efficiency of the LH1 scheme, whose total CPU requirements is just a fraction of the CPU time required by the convolution with BEC operator—a negligible amount compared to the cost of 3DVAR analysis.

A separate question, requiring further investigation, is the accurate treatment of the boundary conditions. In the present study we assumed that boundaries affect only the magnitude of the corresponding columns of  $\mathbf{C}$ , but not their structure. This approximation is only partly

FIG. 8. Scatterplots of the true diagonal elements of  $\mathbf{C}_1$  (vertical axis) vs their approximations by (a) LH0 and (b) LH1 algorithms. The actual values are multiplied by  $10^3$ . Near-boundary points are excluded. (c) Diagonal approximation errors as a function of  $\gamma$  for the  $\mathbf{C}_1$  (black) and  $\mathbf{C}_2$  (gray) BEC models. The dashed line shows the value of  $\gamma_3$  given by (12).

consistent with the zero normal flux conditions for  $\mathbf{D}$ , but can be avoided if one uses “transparent” boundary conditions (e.g., Mirouze and Weaver 2010), which do not require computation of the adjustment factors (15).

On the other hand, it might be beneficial to keep physical boundary conditions in the formulation of  $\mathbf{D}$ , as they are likely to bring more realism to the dynamics of the BE field. In the considered diffusion tensor model, anisotropic BE propagation is superimposed on the small-scale isotropic BE diffusion, which takes place at scales that are not well resolved by the grid (less than  $3\delta$ ). This may give some grounds to employ isotropic homogeneous model at distances  $d < 3\delta$  from the boundary, because the deterministic part of the BE transport associated with the boundary effects is not well resolved at these scales anyway. Such simplification greatly reduces the computational cost of the LH algorithm with zero-flux conditions because the region screened by the boundary is readily available from the land mask and needs no computation/rescaling at every point. In contrast to the 2D problem considered, such computations may be more costly in the 3D problems, as the number of adjustment factors  $n \sim N^{2/3}$  may be quite large even with a moderate grid size  $N \sim 10^6\text{--}10^7$ . However, our preliminary experiments with inhomogeneous BEC models with larger  $N$  have shown that estimation of the diagonal with the LH1 scheme still remains two orders in magnitude less expensive than the 3DVAR analysis.

Results of this study indicate that LH1 approximations to the BEC diagonal may serve as an efficient tool for renormalization of the correlation operators in variational data assimilation, as they are capable of providing 3%–10% accuracy in realistically inhomogeneous BEC models.

*Acknowledgments.* This study was supported by the Office of Naval Research (Program Element 0602435N).

## APPENDIX A

### Hadamard Matrices

By definition, a Hadamard matrix (HM) is a square matrix whose entries are either 1 or  $-1$  and whose columns are mutually orthogonal. The simplest way to construct HMs is the recursive Sylvester algorithm, which is based on the obvious property: if  $\mathbf{H}_M$  is an  $M \times M$  Hadamard matrix, then

$$\mathbf{H}_{2M} = \begin{bmatrix} \mathbf{H}_M & \mathbf{H}_M \\ \mathbf{H}_M & -\mathbf{H}_M \end{bmatrix}$$

is also HM. Starting from  $\mathbf{H}_2 = [1 \ 1; 1 \ -1]$ , the HMs with order  $M = 2^n$ ,  $n = 1, 2, \dots$  can be easily constructed. The

HMs with  $M = 12, 20$  were constructed “manually” more than a century ago. A more general HM construction algorithm, which employs the Galois fields theory was found in 1933. However, it is still unknown if HMs exist for all  $M = 4n$  where  $n$  is a positive integer.

Hadamard matrices are widely used in many branches of applied mathematics and statistics ([http://en.wikipedia.org/wiki/Hadamard\\_matrix](http://en.wikipedia.org/wiki/Hadamard_matrix)). In the present study we used the MatLab code that handles only the cases when  $M/12$  or  $M/20$  is a power of 2. Despite this restriction, the available values of  $M$  were sufficient for our purposes.

## APPENDIX B

### Correlation Modeling with Diffusion Operator

Consider a family of linear operators  $\mathbf{B}$  in the space  $\mathcal{H}$  of sufficiently smooth square-integrable functions  $f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ :

$$(\mathbf{I} - \mathbf{D}/2m)^m \mathbf{B}(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x}, \mathbf{x}'). \quad (\text{B1})$$

Assume that  $\nu = \text{const}$  in the definition of the diffusion operator (1) and perform the coordinate transformation  $\tilde{\mathbf{x}} = \nu^{-1/2}\mathbf{x}$ . In the new coordinate system the diffusion tensor  $\nu$  is the unity matrix,  $\mathbf{D}$  takes the form of the Laplacian operator  $\Delta$ , whereas the  $\delta$  function in the rhs of (B1) is rescaled as a result of the contraction  $\Omega = \sqrt{\det\nu}$  of the volume element:  $\delta(\mathbf{x}, \mathbf{x}') \rightarrow \Omega^{-1}\delta(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}')$ . Equation (B1) takes the following form:

$$(\mathbf{I} - \Delta/2m)^m \mathbf{B}(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}') = \frac{1}{\Omega} \delta(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}'). \quad (\text{B2})$$

After changing the coordinates in  $\mathcal{H}$  with the Fourier transform, both  $\Delta$  and the operator  $\mathbf{B}$  are diagonalized. The diagonal elements are given by

$$\mathbf{B}(\mathbf{k}) = \frac{1}{\Omega} (1 + \mathbf{k}^2/2m)^{-m}. \quad (\text{B3})$$

It is easy to note that by virtue of the textbook formula  $\exp x = \lim_{m \rightarrow \infty} (1 + x/m)^m$ , the Fourier representation (B3) of  $\mathbf{B}$  converges at large  $m$  to

$$\mathbf{B}_1(\mathbf{k}) = \frac{1}{\Omega} \exp(-\mathbf{k}^2/2), \quad (\text{B4})$$

the well-known expression for the Green function of the diffusion equation. Since the diagonals given by (B3)–(B4) are strictly positive, the respective operators in  $\mathcal{H}$  are positive definite, and can be interpreted as correlation operators. This property has been widely used for the background error covariance modeling in practical data assimilation.

Matrix elements of the operators  $\mathbf{B}_1 = \exp(\mathbf{D}/2)$  and  $\mathbf{B}_2 = (\mathbf{I} - \mathbf{D}/2m)^{-m}$  can be found via the inverse Fourier transformations of (B3)–(B4):

$$\begin{aligned} \mathbf{B}_1(\mathbf{r}) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathbf{B}_1(\mathbf{k}) \exp(i\mathbf{k}\mathbf{r}) d\mathbf{k} \\ &= \frac{1}{(2\pi)^{n/2} \Omega} \exp\left(-\frac{\mathbf{r}^2}{2}\right), \end{aligned} \quad (\text{B5})$$

$$\begin{aligned} \mathbf{B}_2(\mathbf{r}) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathbf{B}(\mathbf{k}) \exp(i\mathbf{k}\mathbf{r}) d\mathbf{k} \\ &= \frac{1}{(2\pi)^{n/2} \Omega} \frac{|\sqrt{2m\mathbf{r}}|^s K_s(|\sqrt{2m\mathbf{r}}|)}{2^{s-1}\Gamma(s)}, \end{aligned} \quad (\text{B6})$$

where  $s = m - n/2$  and  $\mathbf{r}^2 = (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)^T(\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)$  is the square of the distance between the correlated points  $\tilde{\mathbf{x}}_1$  and  $\tilde{\mathbf{x}}_2$  in the transformed coordinate system. Making a substitution  $\mathbf{r}^2 = (\mathbf{x}_1 - \mathbf{x}_2)^T \nu^{-1}(\mathbf{x}_1 - \mathbf{x}_2)$  in the rhs of (B5)–(B6), we obtain the matrix elements (4)–(5) in the original coordinate system.

From the viewpoint of numerical applications it is instructive to connect the discretized operator equation (B1) with various time integration techniques of the discretized diffusion equation. The numerical approximation of  $\mathbf{B}$  in (B1) is never calculated in practice because of the immense cost of such a computation. Instead, the result  $\hat{f}^m(\mathbf{x})$  of action by  $\mathbf{B}$  on a (discrete) model state vector  $\hat{f}^0(\mathbf{x})$  is calculated by solving the system of equations

$$(\mathbf{I} - \hat{\mathbf{D}}/2m)^m \hat{f}^m = \hat{f}^0, \quad (\text{B7})$$

where  $\hat{\mathbf{D}}$  denotes the discretized diffusion operator. If we now assume that  $\hat{f}^0(x)$  represents the “initial state” and prescribe the “time step”  $\delta t$  such that the “virtual ntegration time” is  $m\delta t = 1$ , the action of the correlation operator (B1) can be identified as a result of discrete-time integration of the diffusion equation  $\partial_t f = \mathbf{D}/2f$  with the implicit scheme:

$$\hat{f}^i - \hat{f}^{i-1} = \frac{1}{2} \delta t \hat{\mathbf{D}} \hat{f}^i, \quad i = 1, \dots, m \quad (\text{B8})$$

starting from the initial state  $\hat{f}^0$ .

Similarly, the action of  $\exp(\mathbf{D}/2)$  is never computed by convolving a state vector  $\hat{f}^0$  with the discretized kernel (B5), but rather by the discrete-time integration of the diffusion equation with the explicit numerical scheme:

$$\hat{f}^i - \hat{f}^{i-1} = \frac{1}{2} \delta t \hat{\mathbf{D}} \hat{f}^{i-1}, \quad i = 1, \dots, m. \quad (\text{B9})$$

In contrast to the implicit scheme (B8),  $\delta t$  in this numerical method is limited from above by the stability condition requiring that eigenvalues of the operator  $\mathbf{I} + \delta t \hat{\mathbf{D}}/2$  must be less than 1 in the absolute value. As a consequence, the minimum number of time steps  $m = 1/\delta t$  may be quite large, making the numerical implementation of  $\exp(\mathbf{D}/2)$  computationally impractical.

## REFERENCES

- Avramidi, I. G., 1999: Covariant techniques for computation of the heat kernel. *Rev. Math. Phys.*, **11**, 947–980.
- Bekas, C. F., E. Kokiopoulou, and Y. Saad, 2007: An estimator for the diagonal of a matrix. *Appl. Numer. Math.*, **57**, 1214–1229.
- Derber, J., and A. Rosati, 1989: A global ocean data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333–1347.
- Di Lorenzo, E., A. M. Moore, H. G. Arango, B. D. Cornuelle, A. J. Miller, B. S. Powell, B. S. Chua, and A. F. Bennett, 2007: Weak and strong constraint data assimilation in the Inverse Ocean Modelling System (ROMS): Development and application for a baroclinic coastal upwelling system. *Ocean Modell.*, **16**, 160–187.
- Dong, S.-J., and K.-F. Liu, 1994: Stochastic estimation with  $Z_2$  noise. *Phys. Lett.*, **328B**, 130–136.
- Girard, D. F., 1989: A fast Monte-Carlo cross-validation procedure for large least squares problems with noisy data. *Numer. Math.*, **56**, 1–23.
- Gusynin, V. P., and V. A. Kushnir, 1991: On-diagonal heat kernel expansion in covariant derivatives in curved space. *Class. Quantum Gravity*, **8**, 279–285.
- Hutchison, M. F., 1989: A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *J. Commun. Stat. Simul.*, **18**, 1059–1076.
- Liu, Y., J. Zhu, J. She, S. Zhuang, W. Fu, and J. Gao, 2009: Assimilating temperature and salinity profile observations using an anisotropic recursive filter in a coastal ocean model. *Ocean Modell.*, **30**, 75–87.
- Martin, P. J., C. N. Barron, L. F. Smedstad, T. J. Campbell, A. J. Wallcraft, R. C. Rhodes, C. Rowley, and T. L. Townsend, 2009: User’s manual for the Navy Coastal Ocean Model Version 4.0. NRL/MR/7320-09-9151, Naval Research Laboratory, Stennis Space Center, MS, 75 pp.
- Mirouze, I., and A. T. Weaver, 2010: Representation of correlation functions in variational data assimilation using an implicit diffusion operator. *Quart. J. Roy. Meteor. Soc.*, **136**, 1421–1443.
- Pannekoucke, O., and S. Massart, 2008: Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation. *Quart. J. Roy. Meteor. Soc.*, **134**, 1425–1438.
- Purser, R. J., 2008a: Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and the techniques of non-Euclidean geometry. Part I: Analytic solutions for symmetrical configurations and the validation of practical algorithms. NOAA/NCEP Office Note 456, 48 pp.
- , 2008b: Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and the techniques of non-Euclidean geometry. Part II: Riemannian geometry and

- the generic parametrix expansion method. NOAA/NCEP Office Note 457, 55 pp.
- , W. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131**, 1536–1548.
- Weaver, A., and P. Courtier, 2001: Correlation modeling on a sphere using a generalized diffusion equation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1815–1846.
- , J. Vialard, and D. L. T. Anderson, 2003: Three and four-dimensional variational assimilation with a general circulation model of the Tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks. *Mon. Wea. Rev.*, **131**, 1360–1378.
- Xu, Q., 2005: Representations of inverse covariances by differential operators. *Adv. Atmos. Sci.*, **22** (2), 181–198.
- Yaremchuk, M., and S. Smith, 2011: On the correlation functions associated with polynomials of the diffusion operator. *Quart. J. Roy. Meteor. Soc.*, **137**, 1927–1932.