Covariance Localization with the Diffusion-Based Correlation Models

Max Yaremchuk

Naval Research Laboratory, Stennis Space Center, Mississippi

Dmitry Nechaev

Department of Marine Science, University of Southern Mississippi, Hattiesburg, Mississippi

ABSTRACT

Improving the performance of ensemble filters applied to models with many state variables requires regularization of the covariance estimates by localizing the impact of observations on state variables. A covariance localization technique based on modeling of the sample covariance with polynomial functions of the diffusion operator (DL method) is presented. Performance of the technique is compared with the nonadaptive (NAL) and adaptive (AL) ensemble localization schemes in the framework of numerical experiments with synthetic covariance matrices in a realistically inhomogeneous setting. It is shown that the DL approach is comparable in accuracy with the AL method when the ensemble size is less than 100. With larger ensembles, the accuracy of the DL approach is limited by the local homogeneity assumption underlying the technique. Computationally, the DL method is comparable with the NAL technique if the ratio of the local decorrelation scale to the grid step is not too large.

1. Introduction

The problem of estimating the background error statistics is an important issue in the ensemble filtering and hybrid data assimilation algorithms that employ ensembles for error analysis and propagation. Increasing the accuracy in estimating the background error statistics remains a scientific and technical challenge, because the (co)variance estimates have to be drawn from a relatively small number of samples contaminated by the noise of diverse origin.

A particular type of background error covariance (BEC) estimation technique employs an ensemble of assimilations (e.g., Fisher 2003; Berre et al. 2006) to assess the covariance structure from the ensemble average. Because of computational limitations, ensemble size rarely exceeds 100 members in practice, thus limiting the accuracy of the straightforward averaging approach because of the significant level of sampling noise. The impact of sampling noise on the accuracy of the BEC estimates has been addressed by Houtekamer and Mitchell (1998) and

Hamill et al. (2001) and led to the development of the filtering techniques based on the Schur product of the sample correlations with the heuristic filters (localization operators). This approach tends to localize covariances in physical space and suppresses long-range correlations, whose accuracy is most affected by the sampling noise (e.g., Houtekamer and Mitchell 2001; Buehner 2005).

In the last decade, the localization techniques have been under rapid development in several directions with the major objective to relax the spatial homogeneity assumption underlying the original scheme. In particular, Fisher (2003), Deckmyn and Berre (2005), and Pannekoucke et al. (2007) utilized a wavelet approach to account for inhomogeneities in the covariance structure; Wu et al. (2002) and Purser et al. (2003) employed recursive filters to localize the covariances; Weaver and Courtier (2001), Pannekoucke and Massart (2008), and Weaver and Mirouze (2012) used a closely related diffusion operator approach; and Pannekoucke (2009) explored a hybrid scheme, featuring wavelet technique in combination with the diffusion method, while Anderson (2007) employed a sampling error approach to derive localization from multiple ensembles in the framework of the hierarchical ensemble filter technique. In the oil and gas exploration industry, anisotropic localization

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

functions were derived by combining the regions of sensitivity of the well data with prior geological models (e.g., Emerick and Reynolds 2011; Chen and Oliver 2010).

Another direction in the localization techniques was pioneered by Bishop and Hodyss (2007) who proposed to augment the original ensemble by including Schur cross products of the spatially smoothed ensemble members. Further development of this approach (Bishop and Hodyss 2009a,b; Bishop et al. 2011; Bishop and Hodyss 2011) demonstrated its flexibility in adapting the covariances to the 4D background flow structures, especially in the case of strongly inhomogeneous statistics. A certain disadvantage of the adaptive localization (AL) technique is a relatively high computational cost, associated with the necessity to operate with the expanded ensemble. A good review of the filtering/localization techniques was recently given by Berre and Desroziers (2010).

In this study we employ the numerical experimentation approach of Weaver and Mirouze (2012) who tested various approximations of the ensemble-generated covariance matrix by the exponent of the diffusion operator in an idealized configuration. The presented work considers four localization techniques applied to three different covariance models in a realistically inhomogeneous 2D setting. Our major focus is on comparing nonadaptive and adaptive localization methods with the techniques based on modeling sample covariance by polynomial functions of the diffusion operator. To make the comparison, we construct inhomogeneous covariance matrices **B**, generate the respective ensembles, and retrieve **B** from a limited number of ensemble members by the means of considered localization techniques. In the next section the four localization methods used are briefly overviewed. Methodology of the numerical experiments is described in section 3. In section 4, the localization methods are compared in terms of accuracy in approximating **B** for various ensemble sizes and their computational efficiency. The results are summarized and discussed in section 5.

2. Methods of covariance localization

a. Traditional scheme

Given an ensemble $\{\mathbf{x}_k\}/\sqrt{K-1} \in \mathbb{R}^N$ of K normalized error perturbations about the ensemble mean listed as columns of the $K \times N$ matrix **X**, their sample covariance **B** is estimated by

$$\mathbf{B} \equiv \text{cov}\{\mathbf{x}_k\} = \mathbf{X}\mathbf{X}^{\text{T}}. \quad (1)$$

In practice, the dimension of the model state N is much larger than K, and the sample estimate (1) always

contains spurious correlations at large distances. To increase the accuracy in approximation of the BEC matrix **B**, Houtekamer and Mitchell (1998) proposed to assign zero correlations to the components of **x** separated by distances larger than a certain prescribed value d (localization scale). Technically, such a "localized" covariance matrix $\mathbf{B}_\ell$ is obtained as the elementwise (Schur) product $\circ$ of the raw sample covariance **B** and the localization matrix $\mathbf{W}_d$, whose off-diagonal elements are set to zero if the distance between correlated points exceeds d:

$$\mathbf{B}_\ell = \mathbf{B} \circ \mathbf{W}_d. \quad (2)$$

This method simultaneously suppresses spurious ensemble correlations located far from the diagonal and shrinks the null space of **B**, whose "raw" dimension $N - K + 1$ is very large, and thus likely inconsistent with the rank of the true BEC matrix. A disadvantage of the technique is that it relies on a heuristic matrix $\mathbf{W}_d$, which does not explicitly take into account inhomogeneity and anisotropy of the background flow which affects the BEC evolution.

b. Adaptive methods

Recently, Bishop and Hodyss (2007, 2009a,b, 2011) developed a family of AL schemes. The idea is to compute **W** as the sample correlation matrix generated by Schur cross products $\hat{\mathbf{x}}_{ik}$ of the spatially smoothed (modulated) members of the original ensemble (e.g., Bishop and Hodyss 2009a, 2011):

$$\hat{\mathbf{x}}_{ij} = (\mathbf{S}\mathbf{x}_i) \circ (\mathbf{S}\mathbf{x}_j); \quad i = 1, \dots, K; \quad j = i, \dots, K, \quad (3)$$

where **S** is a suitably chosen smoothing operator while $J = K(K + 1)/2$ is the size of the modulated ensemble. Assuming that the columns of the $J \times N$ matrix $\hat{\mathbf{X}}$ list perturbations $\{\hat{\mathbf{x}}_{ij}\}$ of the modulated ensemble about their mean that are normalized to have unit variance and divided by $\sqrt{J-1}$, the adaptively localized BEC matrix is

$$\tilde{\mathbf{B}}_\ell^* = \mathbf{B} \circ \mathbf{W}^* \equiv \mathbf{B} \circ (\hat{\mathbf{X}}\hat{\mathbf{X}}^{\text{T}}). \quad (4)$$

To further increase stability and computational efficiency of the AL technique, Bishop and Hodyss (2011) supplemented the method with additional multiplication by $\mathbf{W}_d$:

$$\mathbf{B}_\ell^* = \mathbf{B} \circ \mathbf{W}^* \circ \mathbf{W}_d. \quad (5)$$

Recent experiments with this improved AL scheme have shown its good localization properties and reasonable numerical performance (Bishop and Hodyss 2011). A certain disadvantage of the method is the numerical cost:

apart from the necessity to smooth ensemble members, multiplication by $\mathbf{B}_\ell^*$ requires computing a convolution with a $KJN \times N$ matrix, whose columns are $\mathbf{x}_k \circ \hat{\mathbf{x}}_{ij} \circ \mathbf{w}_n$, where $\mathbf{w}_n$ are the columns of the square root of $\mathbf{W}_d$.

c. Modeling sample covariance

Another way of estimating the true covariance is to create its full-rank covariance model using the low-rank ensemble approximation (1). In recent years this approach, fueled by the developments in covariance modeling with the diffusion operator (e.g., Weaver and Courtier 2001; Xu 2005; Yaremchuk and Smith 2011; Yaremchuk and Sentchev 2012), has been studied by many authors (e.g., Belo Pereira and Berre 2006; Pannekoucke and Massart 2008; Pannekoucke 2009; Sato et al. 2009; Weaver and Mirouze 2012).

The idea of the approach is to parameterize the structure of the true BEC matrix by the diffusion tensor field $D^{\alpha\beta}(\mathbf{x})$, which defines the positive-definite diffusion operator $\mathbf{D} = -\nabla_\alpha D^{\alpha\beta}\nabla_\beta$.

To avoid confusion with notations, vectors and matrices in state space $\mathbb{R}^N$ are denoted by the boldface roman and boldface san serif fonts, respectively. In the 2D physical space $\mathbb{R}^2$ we adopt tensor notation, where vectors and matrices are boldface and italicized, Greek indices enumerate coordinates, take the values 1 and 2, and summation is assumed over repeating indices.

The operator $\mathbf{D}$ is used to construct the $\mathbf{B}$-approximating covariance model that is specified by a positive function F of $\mathbf{D}$ in order to meet the positive-definiteness property of $\mathbf{B}$. Furthermore, for computational reasons it is desirable that F could be computed recursively and at the same time it should invert the spectrum of $\mathbf{D}$ (i.e., the largest eigenvalues of $F\{\mathbf{D}\}$ should correspond to the smallest eigenvalues of $\mathbf{D}$). The latter requirement ensures the smoothing property of the BEC model, which is important in applications.

Among the functions satisfying these requirements are the exponent and its nth-order binomial (spline) approximations:

$$F_e\{\mathbf{D}\} = \exp\left(-\frac{\mathbf{D}}{2}\right), \tag{6}$$

$$F_n\{\mathbf{D}\} = \left(\mathbf{I} + \frac{\mathbf{D}}{2n}\right)^{-n}. \tag{7}$$

The functional forms in (6)–(7) are used to define the correlation matrix $\mathbf{C}$, which can be easily transformed into $\mathbf{B}$ by the renormalization formula $\mathbf{B} = \mathbf{VCV}$, where $\mathbf{V} = \mathrm{diag}(\mathbf{v})$, and $\mathbf{v} \in \mathbb{R}^N$ is the vector of rms error variances (square roots of the diagonal of $\mathbf{B}$). The elements $v(\mathbf{x})$ of $\mathbf{v}$ are relatively well known from the ensemble statistics as they suffer less from sampling errors than

ensemble estimates of the correlations. In its turn, the correlation matrix $\mathbf{C}$ can be obtained from $F\{\mathbf{D}\}$ by setting its diagonal elements to unity:

$$\mathbf{C} = \mathrm{diag}(\mathbf{f})^{-1/2} F\{\mathbf{D}\} \mathrm{diag}(\mathbf{f})^{-1/2}, \tag{8}$$

if a good approximation to the diagonal elements $\mathbf{f}$ of $F\{\mathbf{D}\}$ is available (Purser et al. 2003; Yaremchuk and Carrier 2012).

This study employs functions F_e and F_n for approximating the BEC matrix by selecting $D^{\alpha\beta}(\mathbf{x})$ in a way that the matrix $\mathbf{B} = \mathbf{VCV}$ given by (6)–(8) fits the structure of the sample covariance (1) for small distances and produces negligible correlations at large distances. The latter property is satisfied by the functions (6)–(7).

A standard method of finding $\mathbf{D}$ for the functional forms (6)–(7) is to use analytic relationships between the derivatives of $F\{\mathbf{D}\}$ in the vicinity of the diagonal (i.e., at small separations between correlated points) and the diffusion tensor (e.g., Belo Pereira and Berre 2006; Sato et al. 2009; Weaver and Mirouze 2012). These relationships are derived under the assumption that local decorrelation scales are much smaller than the typical scale of spatial variability of $\boldsymbol{D}$. In that case, the correlation matrix elements $\mathbf{C}(\mathbf{x}, \mathbf{y})$ are locally homogeneous (LH); that is, they depend only on the relative position $\mathbf{r} = \mathbf{x} - \mathbf{y}$ of the correlated points $\mathbf{x}$, $\mathbf{y}$, and can be written down explicitly (e.g., Yaremchuk and Smith 2011):

$$\mathbf{C}_e(\mathbf{r}) = \exp\left(-\frac{r^2}{2}\right), \tag{9}$$

$$\mathbf{C}_n(\mathbf{r}) = \frac{(\sqrt{2n}r)^{n-1} \mathcal{K}_{n-1}(\sqrt{2n}r)}{2^{n-2}(n-2)!}, \tag{10}$$

where

$$r^2 = r^\alpha D_{\alpha\beta}^{-1} r^\beta \tag{11}$$

is the squared distance measured in terms of the local decorrelation scales defined by the eigenvalues of $\boldsymbol{D}$ and $\mathcal{K}$ is the Bessel function of the second kind. Dependence of the correlation matrix elements on the distance r from the diagonal is shown in Fig. 1.

Direct differentiation of (9)–(10) at zero distance ($\mathbf{r} = 0$), yields the following relationships, useful for estimation of the diffusion tensor for the models (9)–(10), respectively:

$$D_{\alpha\beta}^{-1}(\mathbf{x}) = -[\nabla_\alpha \nabla_\beta \mathbf{C}_e], \tag{12}$$

$$D_{\alpha\beta}^{-1}(\mathbf{x}) = -\frac{n-2}{n}[\nabla_\alpha \nabla_\beta \mathbf{C}_n]. \tag{13}$$
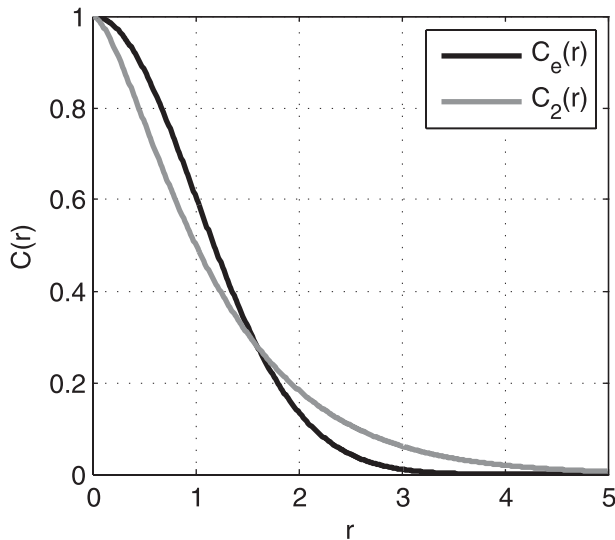
FIG. 1. Correlation functions of the Gaussian and second-order spline models described by (9)–(10).

Here square brackets denote extracting the diagonal values from a matrix. This approach requires **C** to be twice differentiable at the diagonal, which is not the case for spline models with $n < 3$. Expressions (12)–(13) were obtained in the 2D Cartesian coordinates by Weaver and Mirouze (2012). Similar relationships hold for an arbitrary correlation model satisfying the conditions of local homogeneity and appropriate differentiability of the correlation function at $r = 0$ (appendix A).

Taking into account the commutativity of the ensemble averaging and $\langle \ \rangle$ differentiation operators renders the rhs of (12)–(13) in the form involving correlations of the first derivatives of the ensemble members (see Belo Pereira and Berre 2006; Weaver and Mirouze 2012; appendix B):

$$[\nabla_\alpha \nabla_\beta \mathbf{C}] = \frac{\langle (\nabla_\alpha \mathbf{x}) \circ (\nabla_\beta \mathbf{x}) \rangle - (\nabla_\alpha \mathbf{v}) \circ (\nabla_\beta \mathbf{v})}{\mathbf{v} \circ \mathbf{v}}. \quad (14)$$

This expression together with relationships (12)–(13) is more convenient for numerical estimation of **D** via sample correlations because it is formulated in terms of the ensemble perturbations and does not involve second derivatives. Weaver and Mirouze (2012) have shown recently that the method is capable of delivering rms accuracies of 20%–80% in reconstructing $\mathbf{D}^{-1}$ in idealized 2D setting. The approach has a few drawbacks. First, the gradient computation tends to amplify sampling noise in the estimate of $\mathbf{D}^{-1}$. The inversion of $\mathbf{D}^{-1}$ is also prone to error amplification. For these reasons, the technique is often supplemented by additional smoothing (Raynaud et al. 2009; Berre and Desroziers 2010; Weaver and

Mirouze 2012). Second, the relationship (14) cannot be applied to the BEC models that are not differentiable at the diagonal, such as the second-order ($n = 2$) spline model (7) in 3D, which is characterized by the exponential correlation function.

An alternative approach is to estimate the diffusion tensor directly by minimizing the difference between the ensemble estimate of the correlations in the vicinity of the diagonal and its local analytic approximations (9)–(10). This approach is likely to be more robust, as it does not involve differentiation and matrix inversion and can be formulated as a least squares problem in the space of the unknown elements of **D**.

In the following sections we compare efficiency of the four localization methods: nonadaptive (section 2a), adaptive (section 2b), and the two described above methods of retrieving the diffusion tensor from the ensemble covariances. For brevity, we will refer to the latter two methods as "differential" and "integral" diffusion localization (DL) schemes.

To explore the efficiency, we adopt the following experimentation strategy: after specifying the "true" covariance matrices **B**, the respective ensembles are generated and then the obtained ensemble members are used to retrieve the approximate structure of **B** by a given localization method.

3. Methodology

a. Experimental setting

Numerical experiments with simulated ensembles were performed as follows. First, the true BEC matrix was specified together with the ensemble by selecting a variance distribution $\mathbf{v}(\mathbf{x})$ and a correlation model (6)–(7) in a real oceanic domain shown in Fig. 2. The variance distribution was chosen to simulate surface temperature variations in the northern Gulf of Mexico near the mouth of Mississippi. The true distribution of **D** (Fig. 2) was specified to mimic the background error dynamics driven by near-coastal topographically controlled circulation. We assumed that the corresponding background currents followed the depth contours and the larger eigenvector of **D** was oriented in that direction and was proportional to the magnitude of the local bathymetry gradient. In the regions where bottom slope was less than 20% of its rms value over the domain, the diffusion was set to be isotropic with the decorrelation scale of 15 km (see appendix C for more details).

Two BEC models used in the experiments were the Gaussian (6) and the second-order spline model (7). The corresponding true correlation matrices $\mathbf{C}_e$ and $\mathbf{C}_2$ were computed explicitly: first, all the columns of $F(\mathbf{D})$ were
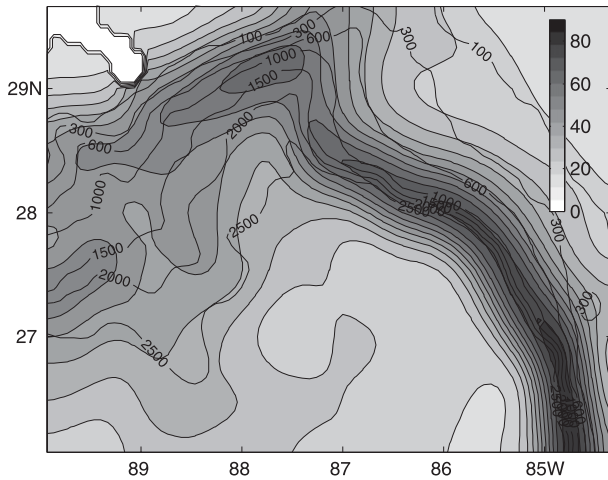
FIG. 2. True distribution of the longer principal axis of the diffusion tensor (km). Labeled contours show depth in meters.

computed as convolutions of the operators (9)–(10) with the δ functions located in every grid point of the domain. The resulting matrices were then renormalized by their diagonal elements using (8), and the true BEC matrices were then obtained by

$$\mathbf{B}_e = \mathbf{V}\mathbf{C}_e\mathbf{V}; \quad \mathbf{B}_2 = \mathbf{V}\mathbf{C}_2\mathbf{V}. \tag{15}$$

Sums of eight columns of $\mathbf{C}_e$ and $\mathbf{C}_2$ are shown in Fig. 3. The maximum anisotropy is observed in the southeast corner of the domain characterized by the steepest topography. The total number of matrix elements was $4603^2 \simeq 2 \times 10^7$.

The simulated ensembles $\mathbf{X}_e$ and $\mathbf{X}_m$ were generated by

$$\mathbf{X}_e = \mathbf{V}\mathbf{C}_e^{1/2}\mathbf{R}; \quad \mathbf{X}_2 = \mathbf{V}\mathbf{C}_2^{1/2}\mathbf{R}, \tag{16}$$

where $\mathbf{R}$ is the $K \times N$ matrix, whose columns are the random vectors with $N = 4603$ δ-correlated components evenly distributed with unit variance and the square root is defined by $\mathbf{C} = \mathbf{C}^{1/2}(\mathbf{C}^{1/2})^{\mathrm{T}}$. The value of K was 20 000.

The ensembles $\mathbf{X}_e$ and $\mathbf{X}_m$ were then used to estimate the true covariances $\mathbf{B}_e$ and $\mathbf{B}_2$ with the four localization techniques described in the previous section. The only exception is the differential method, which was not used with the spline model (7) because the corresponding correlation function (10) is not differentiable at the origin. In all the experiments the localization matrix $\mathbf{W}_d$ was Gaussian (9) with the isotropic diffusion tensor $\boldsymbol{D} = d^2\boldsymbol{I}$, where $\boldsymbol{I}$ is the 2×2 identity matrix and d is a tuning parameter defined in the next section.

Numerically, the action of $F_e\{\mathbf{D}\}$ on a state vector $\mathbf{x}$ was approximated by the recursive scheme:
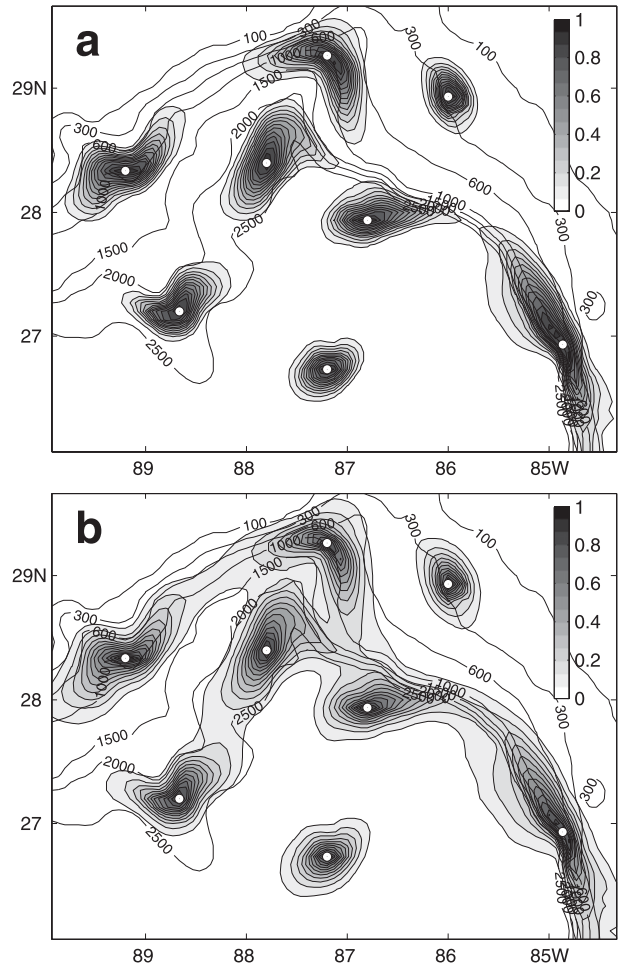

FIG. 3. True correlations for the (a) $\mathbf{C}_e$ and (b) $\mathbf{C}_2$ models plotted for eight different points. Locations of the points are shown by white circles.

$$\exp\left(-\frac{\mathbf{D}}{2}\right)\mathbf{x} \simeq \left(\mathbf{I} - \frac{\mathbf{D}}{2n}\right)^n \mathbf{x}, \tag{17}$$

which can be interpreted as "time integration" of the diffusion equation with the integration period defined by the maximum eigenvalue λ of $\mathbf{D}/2$ over the domain and the "time step" of λ/n. Similarly, $F_2\{\mathbf{D}\}\mathbf{x}$ was computed by iteratively solving the system of equations,

$$\left(\mathbf{I} + \frac{\mathbf{D}}{4}\right)^2 \mathbf{y} = \mathbf{x}, \tag{18}$$

with the minimum residual algorithm (Paige et al. 1995).

Computing the action of the operators $\mathbf{C}_e^{1/2}$ and $\mathbf{C}_2^{1/2}$, which appear in the relationships in (16) requires an algorithm for $F\{\mathbf{D}\}^{1/2}$, which was obtained by halving the number of time steps n in (17) and removing the square in the lhs of (18).

With the simulated ensembles in (16) at hand, the sample covariance matrices $\tilde{\mathbf{B}}_k$ were computed via (1) by varying the number of samples $\mathbf{x}_k$ randomly picked from these ensembles. Using the same samples, rms error variance fields $\tilde{\mathbf{v}}(\mathbf{x})$ and the correlation matrices $\tilde{\mathbf{C}}$ were also computed.

Given these ensemble statistics, the localized estimates of the true covariance matrix were computed with four localization techniques described in the previous section [(2), (5), and (9)–(14) for the DL estimates].

Technically, the DL estimates were obtained by fitting the diffusion tensor field to the structure of $\tilde{\mathbf{C}}$ with two techniques: the first one utilizes the approach based on differentiating the ensemble members [(12)–(14)], whereas the second one extracts $\boldsymbol{D}(\boldsymbol{x})$ from sample correlations $\tilde{\mathbf{C}}$ by minimization of the cost functions:

$$J(\boldsymbol{x}) = \int_\omega [\mathbf{C}(\boldsymbol{x} - \boldsymbol{y}) - \tilde{\mathbf{C}}(\boldsymbol{x}, \boldsymbol{y})]^2 \, d\boldsymbol{y} \to \min_{D(\boldsymbol{x})}, \quad (19)$$

where $\mathbf{C}$ is given by (9)–(10) and ω is a small vicinity of $\mathbf{x}$. Similar approach was tested in a less general formulation by Pannekoucke and Massart (2008) for the 2D Gaussian correlations. To minimize (19) we used the M1QN3 algorithm of Gilbert and Lemarechal (1989) that reduced the L_2 norm of the cost function gradient by three orders of magnitude in 3–6 iterations.

To distinguish between the two DL schemes, the corresponding estimates will be labeled by the superscripts $'$ and $°$ for the differential [(12)–(14)] and integral [(9)–(11), (19)] approaches, respectively.

After the diffusion tensor estimates were obtained using either the first or the second method, the localized estimates $\mathbf{C}'$ and $\mathbf{C}°$ of $\mathbf{C}$ were computed using (6)–(8). Equation (8) contains the diagonal elements of $F\{\mathbf{D}\}$, whose direct computation is numerically prohibitive in practice. For that reason, approximate formulas were used:

$$\mathbf{f} = (2\pi)^{-1} F\{\gamma\mathbf{D}\}\mathbf{d}, \quad (20)$$

where $\mathbf{d} = (\det\boldsymbol{D})^{-1/2}$ and $\gamma_e = 0.33$; $\gamma_2 = 0.28$ for the F_e and F_2 models, respectively (Yaremchuk and Carrier 2012).

Performance of the four localization techniques was measured in terms of the distance between the ensemble-estimated localized covariances $\mathbf{B}_\ell, \mathbf{B}_\ell^*, \mathbf{B}_\ell', \mathbf{B}_\ell°$ and the true covariance $\mathbf{B}$:

$$\rho(\mathbf{B}_\ell, \mathbf{B}) = \sqrt{\frac{|\mathbf{B}_\ell - \mathbf{B}|}{|\mathbf{B}|}}, \quad (21)$$

where $|\ |$ denotes the Frobenius norm. Relative distances between the respective correlation matrices were measured by the following relationship (Herdin et al. 2005):

$$\rho_1(\mathbf{C}_\ell, \boldsymbol{C}) = \sqrt{1 - \frac{\text{Tr}(\mathbf{C}_\ell\mathbf{C})}{|\mathbf{C}_\ell||\mathbf{C}|}}. \quad (22)$$

b. Numerical implementation

In addition to comparing the skills of the localization methods, their computational efficiencies are also compared. In practical applications, $\mathbf{B}_\ell$ and $\mathbf{B}_\ell^*$ are never computed directly, but represented in the "square root" form $\mathbf{B}_\ell = \mathbf{B}_\ell^{1/2}(\mathbf{B}_\ell^{1/2})^\text{T}$ to speed up computations. By virtue of the "square root theorem" (Bishop et al. 2011), $\mathbf{B}_\ell^{1/2}$ and $\mathbf{B}_\ell^{*1/2}$ are the $KN \times N$ and $KJN \times N$ matrices, whose columns are $\mathbf{x}_k \circ \mathbf{w}_n$ and $\mathbf{x}_k \circ \hat{\mathbf{x}}_{ij} \circ \mathbf{w}_n$, respectively (section 2b). The elements of localization matrix $\mathbf{W}_d$ were computed explicitly with the analytic equation (9). At distances exceeding several localization scales the elements were set to zero to avoid senseless multiplications by the tails of the Gaussian exponent. In the numerical experiments this "cutoff" distance was set to $3d$. The nonzero elements of the columns $\mathbf{w}_n$ of $\mathbf{W}_d^{1/2}$ were computed by reducing $\sqrt{2}$ times the localization scale in (11).

To explore the impact of the ensemble size on accuracy of the localization schemes, experiments were performed with five ensemble sizes: $k = 4, 10, 50, 200$, and 1000. The respective modulated ensembles (section 2b) were computed in a different manner for various k. For $k = 4$ and 10 both double and triple Schur products of the raw ensemble members were used, thus creating $J_4 = (4 \times 5)/2 + (4 \times 4 \times 5)/2 = 50$ and $J_{10} = (10 \times 11)/2 + (10 \times 10 \times 11)/2 = 605$ members. For $k = 50$ and 200 only the double products were used. The respective ensemble sizes were 1275 and 20 100. With $k = 1000$ only 20 000 randomly selected pairs were used to create $\{\mathbf{x}_j\}$. The smoothing operator $\mathbf{S}$ [(3)] was also isotropic Gaussian, but its scale d_s was different from d. Both d and d_s were optimized in every experiment to minimize the distance (21) from the true covariance.

The DL algorithms had additional specific features. Estimates of $\boldsymbol{D}'$ obtained from (12)–(14) were first smoothed with the scale of $l = 30$ km, then symmetrized and checked for the positive definiteness. In the case of a negative eigenvalue (a common situation for $k = 4$, 10), the tensor was discarded. The resulting gaps were filled with horizontal interpolation and smoothed again with the same scale.

When computing $\boldsymbol{D}°$, the lengths of principal axes and orientation of the larger axis were chosen as control parameters. This approach eliminated violation of positive definiteness and improved stability of the algorithm.
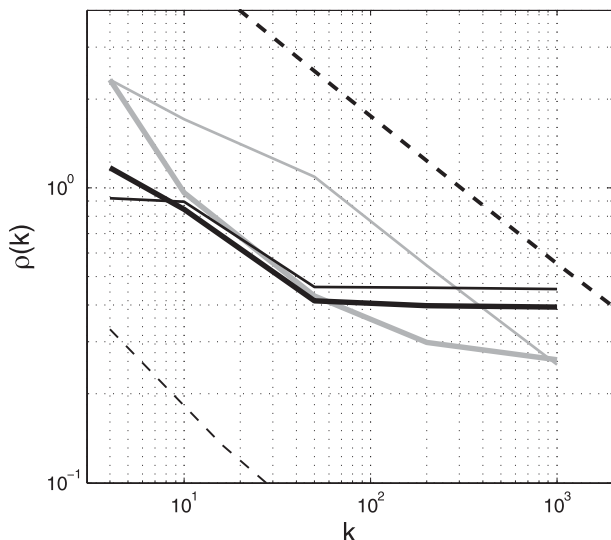
FIG. 4. Relative errors between the true covariance matrix (Gaussian model) and its ensemble estimates for various localization techniques as a function of the ensemble size k. Thick dashed line shows the error of the nonlocalized estimate $\mathbf{B}$ [(1)]. Thin dashed line is the error of the variance estimate. Errors of the NAL $\mathbf{B}_\ell$ (thin line) and AL $\mathbf{B}_\ell^*$ (thick line) methods are shown in gray. Solid black lines correspond to the differential $\mathbf{B}_\ell'$ (thin line) and integral $\mathbf{B}_\ell^\circ$ (thick line) DL methods.

The fitting domain ω was a square four grid steps in size. Tensor parameters were smoothed with the same scale as has been used in the computations of $\mathbf{D}'$.

4. Results

a. Skill comparison

Figure 4 compares skills [(21)] of the four localization techniques for the Gaussian covariance model as a function of the number of ensemble members k. The straight dashed lines provide errors for the raw variance and covariance estimates without localization. As expected, both $\rho(\mathbf{B})$ and $\rho([\mathbf{B}])$ closely follow the law $1/\sqrt{k}$ with the variance estimate $\rho([\mathbf{B}])$ being approximately 20 times more accurate than the estimate of the covariance.

For $k = 4$, the difference between $\rho(\mathbf{B}_\ell)$ and $\rho(\mathbf{B}_\ell^*)$ appears negligible because of the extremely large sampling errors, which cannot be reduced by updating the ensemble with modulated members. In the "practical" range of $10 < k < 500$, the adaptive scheme delivers a 2–3 times better estimate than the nonadaptive localization (NAL) technique, but this advantage disappears at $k > 500$ because of the increase of raw ensemble skill. This type of behavior has been also observed in the experiments where we kept both localization scale d and the smoothing scale d_s constant and equal to 100 km (i.e., did not optimize their values for a given k). In that

case the error curves converged at slightly larger $k \sim$ 1200–1500.

The DL schemes demonstrate a significantly better performance at $k < 20$, although $\rho(\mathbf{B}_\ell')$ is 20%–30% larger than $\rho(\mathbf{B}_\ell^\circ)$ starting from $n = 10$. Flattening of the curves for $\mathbf{B}_\ell', \mathbf{B}_\ell^\circ$ at large k can be explained by two factors. The first one is a certain inconsistency of the true covariance structure with the LH assumption used in the derivation of (9)–(14): Fig. 2 shows that the typical scale of variability of the diffusion tensor's axes is compatible with their magnitude throughout the domain, and in some places (e.g., steep bottom regions in the southwest) it is even smaller than the local decorrelation scales. The second factor is associated with the violation of the LH assumption in computing the normalization factors with (20). Although (20) is capable of approximating the diagonal elements at the error level of 5%–10%, its contribution to the asymptotic error of 0.4 (Fig. 4) is not negligible. Similar observations are reported in the idealized experiments of Weaver and Mirouze (2012).

Figure 5 shows the absolute difference between the eight columns of $\mathbf{C}_\ell'$, $\mathbf{C}_\ell^\circ$ and the respective columns of the true correlation matrix for the Gaussian model shown in Fig. 2a. It is seen that the difference is not zero even in the diagonal points (shown by black circles) where both correlation estimates are supposed to be equal to one by definition. This difference can be virtually embedded as an additional error in the variance estimate $\mathbf{V}$, which is primarily defined by the size of the ensemble. In the reported experiments this diagonal approximation error ranged within 5%–8%, and started to contribute significantly at $k > 30$ (i.e., when the variance estimation error falls below 10%; lower dashed line in Fig. 4). The impact of the diagonal approximation error is less visible when comparing covariance matrices in terms of (22), which is more sensitive to the errors in the off-diagonal elements (Fig. 6).

The degree of inhomogeneity of the true covariance can, in principle, be assessed from asymmetry of the local correlations derived from the ensemble when k is large enough to suppress sampling noise. When the LH assumption is satisfied with high accuracy, the correlation matrix elements satisfy (9)–(10), and therefore should be nearly invariant under the mirror transformations $\mathbf{r} \rightarrow -\mathbf{r}$ in the vicinity of the diagonal. We checked this property for the true correlation matrices and found relatively high degrees of asymmetry (0.24 and 0.28 for $\mathbf{C}_e$ and $\mathbf{C}_2$, respectively). In combination with 5%–8% diagonal errors, these figures may explain the asymptotic error level in approximating the true covariances by the DL schemes (Fig. 4).

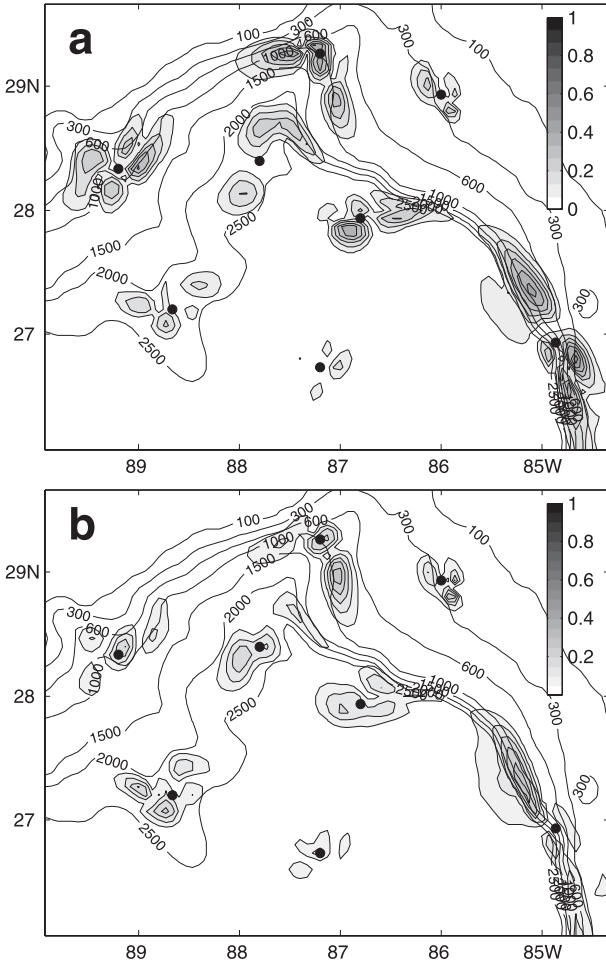Another feature observed in the experiments, is a persistently better performance of the DL methods at

FIG. 5. Absolute difference between eight columns of the true correlation matrix for the Gaussian model (Fig. 3a) and its DL approximations (a) $\mathbf{C}'_\ell$ and (b) $\mathbf{C}^\circ_\ell$ obtained with 50 ensemble members. Filled circles show locations of the diagonal elements.

FIG. 6. Relative errors ρ_1 between the true covariance matrix (spline model, Fig. 2b) and its ensemble estimates for various localization techniques as a function of the ensemble size k. Thick dashed line shows the error of the nonlocalized estimate $\mathbf{B}$ [(1)]. Thin dashed line is the error of the variance estimate. Errors of the NAL $\mathbf{B}_\ell$ (thin line) and AL $\mathbf{B}^*_\ell$ (thick line) methods are shown in gray. Solid black line gives the error of the integral $\mathbf{B}^\circ_\ell$ DL method.

smoothed and renormalized to have the same variance $\mathbf{V}$ as the original models $\mathbf{B}_e$ and $\mathbf{B}_2$.

Figure 8 demonstrates that in the case of $\mathbf{B}_n$ model the approximation errors of the DL schemes are still below the errors of the AL scheme when $n < 30$–40. Furthermore, the DL schemes keep being competitive in the entire range of the practical ensemble sizes (up to

small ensemble sizes k (Figs. 4 and 6). One may assume that this property could be attributed to the fact that the DL schemes have an a priori advantage because the structure of the true covariances is already embedded into the underlying diffusion models used for approximation. To check this, we generated an alternative true covariance matrix $\mathbf{B}_n$, which was far enough from both $\mathbf{B}_e$ and $\mathbf{B}_2$ to eliminate this advantage (Fig. 7).

To do this, we randomly picked 1000 members from each of the ensembles $\mathbf{X}_e$ and $\mathbf{X}_2$, and then generated additional 20 000 members using the adaptive technique described in section 2b. Pairs for Schur cross products were composed by randomly picking members from the two ensembles and never from one. The resulting 22 000-member ensemble was used to compute $\mathbf{B}_n$ with (1). After that the columns of $\mathbf{B}_n$ were additionally
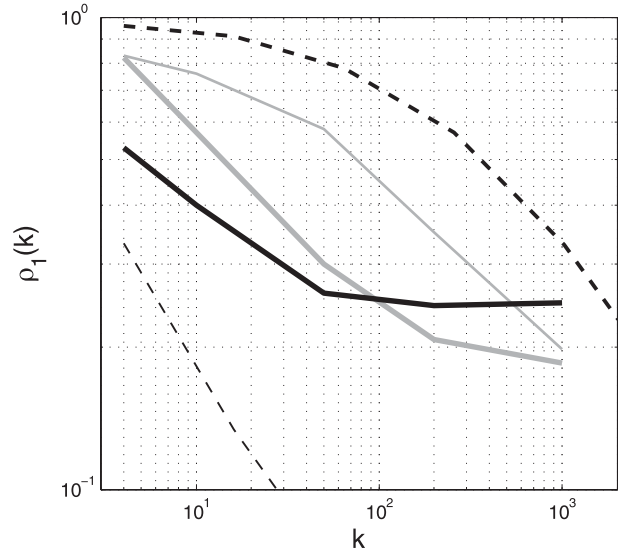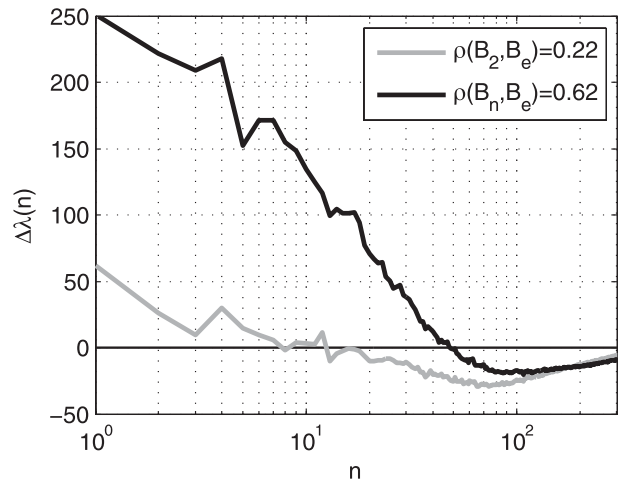

FIG. 7. Difference between the 300 largest eigenvalues of $\mathbf{B}_2$ and $\mathbf{B}_e$ (gray line) and of $\mathbf{B}_n$ and $\mathbf{B}_e$. (top right) Distances between the corresponding matrices are shown.
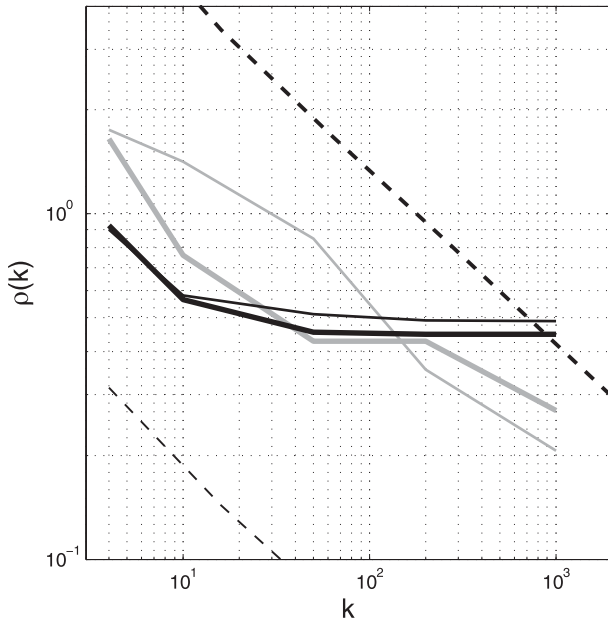
FIG. 8. As in Fig. 4, but for the true covariance $\mathbf{B}_n$.

$n = 150$–200). We therefore may assume that better performance at small ensemble sizes in an intrinsic property of the DL method, which could possibly be explained by its enhanced ability to better capture near-diagonal structure of the correlations. However, only experiments with real assimilation systems can confirm this hypothesis.

One can notice a relatively weak performance of the AL scheme (thick gray line in Fig. 8) as compared to the case of true covariance described by the $\mathbf{B}_e$ model (Fig. 4). Such a behavior can be explained by the fact that the smoothing scale d_s was the same as was used for generation of the modulated ensembles in the experiments with $\mathbf{B}_e$. In general, adjustment of the localization scales significantly improved the approximation accuracy of $\mathbf{B}_\ell$ and $\mathbf{B}_\ell^*$, especially at low k for the standard localization scheme whose optimal values of $d(k)$ changed in a wide range from $d(4) = 30$ to $d(1000) = 500$ km. For the adaptive scheme variations of d were significantly smaller: $d(4) = 100$ to $d(4) = 500$ km.

These figures shed some light on the role near-diagonal elements play in the overall structure of the considered covariance matrices. It appears that accurate estimation of these elements eliminates a larger portion of the error in approximation of the true covariance. To support this idea, we computed distances between the three considered covariances $\mathbf{B}_e$, $\mathbf{B}_2$, and $\mathbf{B}_n$ and their approximations obtained by setting to zero all the off-diagonal elements, located farther than a certain distance r (measured in physical space) from the diagonal. As expected, the major

portion of the error is eliminated when elements within the mean decorrelation scale are accounted for. This feature of the considered covariances partly explains the better skill of the DL schemes that are "more focused" on accurate representation of the near-diagonal structure of the covariance matrices. In addition, DL models are capable to deliver better smoothness away from the diagonal, which is essential for elimination the imbalance problems that may arise when prediction models are used with the resulting analysis (e.g., Kepert 2011).

b. Computational efficiency

In the previous section we have shown that DL schemes appear to be competitive in accuracy with both NAL and AL techniques when the number of ensemble members k is relatively small. When $k > 70 - 100$, the AL scheme provides better accuracy (Figs. 4–8), but the DL method may still remain competitive up to $k \sim 100$. On the other hand, it is much less computationally expensive, because it does not require generation of the costly modulated ensemble.

The cost of localization is defined by the multiplication of the square root of the localized covariance matrix by a state vector. In the case of the NAL scheme, this product involves $M \sim kNn_d$ multiplications, where n_d is the number of nonzero elements in the column of $\mathbf{W}_d^{1/2}$. For the AL scheme [(5)] this number is J times larger and may require significant computational resources.

The cost of implementing the DL schemes consists of two components: estimation of the diffusion tensor and multiplication by the square root of the localization matrix. The number of multiplications required to compute D' at a grid point is approximately proportional to $9k$, because local correlations have to be computed only in the nearest neighborhood of the diagonal and each computation involves k products of the ensemble members. Differentiation, inversion [(12)–(13)], and smoothing adds approximately 50 operations for a grid point thus giving the estimate of $M' \simeq (9k + 50)N$ for the overall cost of computing D'. The cost of multiplication by the square root of $\mathbf{B}_\ell'$ is proportional to Nn_*m, where $n_* = 9$ is the number of elements in the (2D) stencil of D', and $m \sim 10^2$ is the number of either "time steps" in case of $\mathbf{C}_e$ or the number of iterations in solving the respective linear system in the case of $\mathbf{C}_2$ localization model. This brings the estimate of the total number of operations to $M' \sim 9(k + m + 5)N$.

Computing $D°$ is somewhat more expensive than D' because it involves solving a minimization problem at every grid point. In the 2D case considered, estimation of $D°$ required approximately $25(k + 20n_o)$ operations per grid point where $n_o = 5$ is the average number of iterations required for convergence of the minimization

routine and 25 is the number of grid points occupied by the optimization subdomain ω [(19)].

Taking the typical value of $n_d = 49$ for the number of grid points in the localization stencil, the following estimates can be obtained:

$$M \simeq 50kN,$$

$$M^* \simeq 50kN\,J,$$

$$M' \simeq 50kN\left(0.2 + \frac{1}{k} + \frac{m}{5k}\right),$$

$$M^\circ \simeq 50kN\left(0.5 + \frac{2}{k} + \frac{m}{5k}\right).$$

Assuming that $k \gg 1$ and taking the NAL cost $M = 50kN$ as a benchmark, the following estimates of the (normalized by M) localization costs $\overline{M}$ can be obtained:

$$\overline{M}^* = J; \quad \overline{M}' = 0.2\left(1 + \frac{m}{k}\right); \quad \overline{M}^\circ = 0.2\left(2.5 + \frac{m}{k}\right). \tag{23}$$

In the reported experiments the typical value of m ranged between 120–180 for the Gaussian model and 150–300 for the spline model. Thus, for the ensemble size of $k = 50$ both DL models appear to be computationally competitive with the NAL technique ($\overline{M}'_e \sim 0.7 - 0.9$, $\overline{M}^\circ_e \sim 1.0 - 1.2$). Similar CPU time ratios were observed in the reported experiments. As is seen from (23) the computational advantage of the DL schemes improves with the growth of the ensemble size k, although their accuracy tends to stagnate (Figs. 4, 6, and 8).

5. Conclusions

Numerical experiments with the DL schemes in a realistically inhomogeneous 2D setting have shown their competitiveness with the NAL and AL methods in terms of accuracy within the range of ensemble sizes $k \sim$ 20–100 used in the data assimilation practice. For larger ensemble sizes the DL method does not give any error improvement as it reaches the limits imposed by the assumption of local homogeneity.

From the computational point of view, the DL method appears to be compatible with the NAL technique, which is in turn less expensive than the adaptive algorithms proposed by Bishop and Hodyss (2007, 2009a,b). Conducted experiments also indicate that the AL method is significantly more accurate than NAL in the case of strongly inhomogeneous covariances when the ensemble size is less than several hundred.

Comparison of the differential and integral DL schemes have shown that the differential method is 20%–50% less

computationally expensive, although it appears to be somewhat less robust and accurate when applied in realistically inhomogeneous environment. An advantage of the integral approach is that it can be utilized with correlation models that are not differential at the origin.

It should be also noted that the computational efficiency of the DL schemes strongly depends on the number of iterations m needed to compute the action of the localization operator on a state vector. This number is controlled by the ratio of the local decorrelation scale (length of the largest principal axis of **D**) to the grid step, which never exceeded 7 in the reported experiments. Therefore, the DL methods may lose computational efficiency when the model is capable to describe motions at scales well below those resolved by observations. This restriction can be bypassed if the covariances are localized on a grid compatible with the decorrelation scale, a technique suggested by Bishop et al. (2011) to speed up the localization algorithms.

The DL algorithms have enough room for further development along several directions. In particular, the degree of local inhomogeneity of the target covariance could possibly be assessed by monitoring dependence of spatial asymmetry of the local correlations on the number of ensemble members used for their evaluation. This information could then be blended in the cost function (19) to prevent overfitting sample correlations by the analytic model. Efficient higher-order approximations to the diagonal elements of $F\{\mathbf{D}\}$ could also be thought out to improve the accuracy in estimating the DL correlation matrix. Finally, the overall accuracy of the DL covariance estimates could also be improved through their renormalization by the optimally filtered (e.g., Raynaud et al. 2009; Berre and Desroziers 2010) diagonal elements of $\mathbf{f} \circ \mathbf{v}$. This approach can simultaneously reduce sampling errors in the variance field $\mathbf{v}$ estimates and errors associated with the LH assumption in computing the diagonal elements of $F\{\mathbf{D}\}$.

One should also keep in mind that ensembles encountered in large geophysical DA problems are likely to have more complicated structure than the simulated ensembles described by (16). In particular, real-life ensembles are often biased and they do not normally demonstrate $k^{-1/2}$ error scaling for realistic ensemble sizes. On the other hand, the "true" covariance matrices are never known and can hardly be computed for real applications in the nearest future. As a consequence, the only way to compare localization techniques is to estimate their forecast skill and computational efficiency within the real DA problems. Presented results give only an indication that further studies of the DL methods are worth pursuing as they seem to be competitive with other localization techniques. A definite answer could

be given only by the aforementioned experiments with real ensembles, which will be the subject of our future research.

APPENDIX A

Differentiation of the Correlation Functions

To simplify the notations, denote derivatives of a correlation function $C(r)$ by the subscript r and the inverse of the diffusion tensor by $R_{\alpha\beta}$. The second derivative of a correlation function $C(r)$ is

$$\mathbf{\nabla}_\alpha \mathbf{\nabla}_\beta C(r) = C_{rr}(\mathbf{\nabla}_\alpha r)(\mathbf{\nabla}_\beta r) + C_r \mathbf{\nabla}_\alpha \mathbf{\nabla}_\beta r. \quad (A1)$$

Taking the first and second derivatives of (11) under the assumption of local homogeneity yields

$$\mathbf{\nabla}_\alpha r = \frac{1}{r} R_{\alpha\beta} r^\beta; \quad \mathbf{\nabla}_\alpha \mathbf{\nabla}_\beta r = \frac{1}{r}[R_{\alpha\beta} - \Gamma_{\alpha\beta}], \quad (A2)$$

where

$$\Gamma_{\alpha\beta} = \frac{1}{r^2} R_{\alpha\mu} r^\mu R_{\beta\nu} r^\nu$$

is bounded at $r \to 0$.

After substituting (A2) into (A1) and rearranging the terms, (A1) takes the form:

$$\mathbf{\nabla}_\alpha \mathbf{\nabla}_\beta C = \frac{C_r}{r} R_{\alpha\beta} + \left(C_{rr} - \frac{C_r}{r}\right) \Gamma_{\alpha\beta}. \quad (A3)$$

Substitution of the expression in the rhs of (9) into (A3) and taking the limit $r \to 0$ yields (12). Similar operation with the rhs of (10) shows that the second term in the rhs of (A3) is zero if $n > 2$, whereas the first term is equal to $n/(2 - n)$. Note that constraint $n > 2$ is imposed by the condition of differentiability of the correlation function (10) at $r = 0$.

More generally, by using Fourier representation of the covariance function [e.g., Eq. (11) in Yaremchuk and Smith (2011)] it is easy to show that the relationship

$$\lim_{r\to 0}\left(C_{rr} - \frac{C_r}{r}\right) = 0 \quad (A4)$$

holds for arbitrary correlation functions twice differentiable at $r = 0$ and satisfying the local homogeneity condition. Therefore, the differential method that is based on the relationship

$$R_{\alpha\beta} = \left(\lim_{r\to 0} C_r/r\right)^{-1} \mathbf{\nabla}_\alpha \mathbf{\nabla}_\beta C \quad (A5)$$

could be applied to a much broader class of correlation models than those described by (6)–(7).

APPENDIX B

Estimating Second Derivatives of the Correlation Function from Ensemble Perturbations

By definition, the tensor of second derivatives of the BEC matrix $\mathbf{B}(\mathbf{x}, \mathbf{y}) \equiv \mathbf{B}_{xy}$ can be represented in two ways:

$$\nabla^x_\alpha \nabla^y_\beta \mathbf{B}_{xy} = \langle (\nabla^x_\alpha \mathbf{x}_x)(\nabla^y_\beta \mathbf{x}_y) \rangle = \nabla^x_\alpha \nabla^y_\beta (\mathbf{V}_x \mathbf{C}_{xy} \mathbf{V}_y), \quad (B1)$$

where bold italicized superscripts denote the variables of differentiation and the subscripts enumerate the corresponding coordinates in physical space. The rhs of (B1) can be rewritten as

$$\begin{aligned}\nabla^x_\alpha \nabla^y_\beta (\mathbf{V}_x \mathbf{C}_{xy} \mathbf{V}_y) &= (\nabla^x_\alpha \mathbf{V}_x)(\nabla^y_\beta \mathbf{V}_y)\mathbf{C}_{xy} \\ &+ \mathbf{V}_y (\nabla^x_\alpha \mathbf{V}_x) \cdot \nabla^y_\beta \mathbf{C}_{xy} \\ &+ \mathbf{V}_x (\nabla^y_\beta \mathbf{V}_y)\nabla^x_\alpha \mathbf{C}_{xy} + \mathbf{V}_x \mathbf{V}_y \nabla^x_\alpha \nabla^y_\beta \mathbf{C}_{xy}\end{aligned}$$
$$(B2)$$

Taking the value of (B2) at the diagonal ($\mathbf{x} = \mathbf{y}$) under the assumption of local homogeneity $\mathbf{C}_{xy} = \mathbf{C}_{x-y}$ implies that $\mathbf{y} = \mathbf{x}$ in all the expressions involving $\mathbf{V}$ and its derivatives and $\nabla^x_\alpha \nabla^y_\beta \mathbf{C}_{x-y} = -\nabla^r_\alpha \nabla^r_\beta \mathbf{C}_r = -[\mathbf{\nabla}_\alpha \mathbf{\nabla}_\beta \mathbf{C}]$. Assuming that the correlation function is differentiable at $\mathbf{r} = 0$ also implies that its gradients at $\mathbf{r} = 0$ are zero and, therefore, two middle terms in the rhs of (B2) vanish. After taking into account the right equality in (B1) and the definition $\mathbf{C}_{xx} = 1$, (B2) transforms into

$$\langle (\mathbf{\nabla}_\alpha \mathbf{x}) \circ (\mathbf{\nabla}_\beta \mathbf{x}) \rangle = (\mathbf{\nabla}_\alpha \mathbf{v}) \circ (\mathbf{\nabla}_\beta \mathbf{v}) - \mathbf{v} \circ \mathbf{v} \circ [\mathbf{\nabla}_\alpha \mathbf{\nabla}_\beta \mathbf{C}], \quad (B3)$$

which yields (14) after rearrangement of the terms.

APPENDIX C

Diffusion Tensor Model

Numerically, the diffusion operator is defined by

$$\mathbf{D} = (\nu \tilde{\nabla})^{\mathrm{T}}(\nu \tilde{\nabla}),$$

where $\tilde{\nabla}$ is the first-order finite-difference representation of the gradient in 2D and ν is the square root of the local diffusion tensor ($\nu^T \nu = D$) represented by

$$\nu = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\gamma & \sin\gamma \\ -\sin\gamma & \cos\gamma \end{pmatrix} a_0. \qquad \text{(C1)}$$

Here $a_0 = 15$ km is the background decorrelation scale, aa_0 is the square root of the larger eigenvalue of D, and γ is the direction of the eigenvector, corresponding to this eigenvalue. The larger principal axis of D is aligned along the depth $h(x, y)$ contours and its magnitude is proportional to the bottom slope $s = (h_x^2 + h_y^2)^{1/2}$. Specifically, the parameters a and γ are defined by

$$a = \theta(s - s_c)\left(\sqrt{s/s_c} - 1\right) + 1, \qquad \text{(C2)}$$

$$\gamma = \theta(s - s_c)\tan^{-1}(-h_x/h_y), \qquad \text{(C3)}$$

where θ stands for the step function. With this definition, the diffusion is isotropic ($\nu = a_0 I$) when the slope is below the critical value s_c, which is chosen to be $s = 0.0003$. In this case, only 20% of points in the domain were characterized by isotropic diffusion.

REFERENCES

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D,* **230,** 99–111.

Belo Pereira, M., and L. Berre, 2006: The use of ensemble approach to study the background error covariances in a global NWP model. *Mon. Wea. Rev.,* **134,** 2466–2489.

Berre, L., and G. Desroziers, 2010: Filtering of background error variances and correlations by local spatial averaging: A review. *Mon. Wea. Rev.,* **138,** 3693–3720.

——, E. Stefanesku, and M. Belo Pereira, 2006: The representation of the analysis effect in three error simulation techniques. *Tellus,* **58A,** 196–209.

Bishop, C. H., and D. Hodyss, 2007: Flow adaptive moderation of spurious ensemble correlations and its use in ensemble based data assimilation. *Quart. J. Roy. Meteor. Soc.,* **133,** 2029–2044.

——, and ——, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus,* **61,** 84–96.

——, and ——, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus,* **61,** 97–111.

——, and ——, 2011: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Wea. Rev.,* **139,** 1241–1255.

——, ——, P. Steinle, H. Sims, A. M. Clayton, A. C. Lorenc, D. M. Barker, and M. Buehner, 2011: Efficient ensemble covariance localization in variational data assimilation. *Mon. Wea. Rev.,* **139,** 573–580.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.,* **131,** 1013–1043.

Chen, Y., and D. S. Oliver, 2010: Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Comput. Geosci.,* **14,** 579–601.

Deckmyn, A., and L. Berre, 2005: A wavelet approach to representing background error covariances in a limited-area model. *Mon. Wea. Rev.,* **133,** 1279–1294.

Emerick, A., and A. Reynolds, 2011: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Comput. Geosci.,* **15,** 251–269.

Fisher, M., 2003: Background error covariance modelling. *Proc. ECMWF Seminar on Recent Developments in Data Assimilation,* Reading, United Kingdom, ECMWF, 45–63. [Available online at http://www.ecmwf.int/publications/library/ecpublications/_pdf/seminar/2003/sem2003_fisher.pdf.]

Gilbert, J. Ch., and C. Lemarechal, 1989: Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Program.,* **45,** 407–435.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129,** 2776–2790.

Herdin, M., N. Czink, H. Özcelik, and E. Bonek, 2005: Correlation matrix distance, a meaningful measure for evaluation of nonstationary MIMO channels. *Proc. 61st IEEE Vehicular Technology Conf.,* Stockholm, Sweden, Institute of Electrical and Electronics Engineers, 136–140.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 796–811.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123–137.

Kepert, J. D., 2011: Balance-aware covariance localisation for atmospheric and oceanic ensemble Kalman filters. *Comput. Geosci.,* **15,** 239–250.

Paige, C., B. Parlett, and H. van der Vorst, 1995: Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numer. Linear Algebra Appl.,* **2,** 115–134.

Pannekoucke, O., 2009: Heterogeneous correlation modeling based on the wavelet diagonal assumption and on the diffusion operator. *Mon. Wea. Rev.,* **137,** 2995–3012.

——, 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.

——, L. Berre, and G. Desroziers, 2007: Filtering properties of wavelets for local background error correlations. *Quart. J. Roy. Meteor. Soc.,* **133,** 363–379.

Purser, R. J., 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.

Raynaud, L., L. Berre, and G. Desroziers, 2009: Objective filtering of the ensemble-based error variances. *Quart. J. Roy. Meteor. Soc.,* **135,** 1003–1014.

Sato, Y., M. S. F. V. De Pondeca, R. J. Purser, and D. F. Parrish, 2009: Ensemble-based background error covariance implementations using spatial recursive filters in NCEP's grid-point statistical interpolation system. NCEP Office Note 459, Camp Springs, MD, 20 pp. [Available online at http://www.emc.ncep.noaa.gov/officenotes/newernotes/on459.pdf.]

Weaver, A. T., and P. Courtier, 2001: Correlation modeling on a sphere using a generalized diffusion equation. *Quart. J. Roy. Meteor. Soc.,* **127,** 1815–1846.

——, and I. Mirouze, 2012: On the diffusion equation and its application to isotropic and anisotropic correlation modeling in variational assimilation. *Quart. J. Roy. Meteor. Soc.,* doi:10.1002/qj.1955, in press.

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.,* **130,** 2905–2916.

Xu, Q., 2005: Representations of inverse covariances by differential operators. *Adv. Atmos. Sci.,* **22,** 181–198.

Yaremchuk, M., and S. Smith, 2011: On the correlation functions associated with the polynomials of the diffusion operator. *Quart. J. Roy. Meteor. Soc.,* **137,** 1927–1932.

——, and M. Carrier, 2012: On the renormalization of the covariance operators. *Mon. Wea. Rev.,* **140,** 637–649.

——, and A. Sentchev, 2012: Multi-scale correlation functions associated with polynomials of the diffusion operator. *Quart. J. Roy. Meteor. Soc.,* **138,** 1948–1953, doi:10.1002/qj.1896.