

Chapter 8

Background error correlation modeling with diffusion operators

Max Yaremchuk, Matthew Carrier, Scott Smith and Gregg Jacobs

Abstract Many background error correlation (BEC) models in data assimilation are formulated in terms of a positive-definite smoothing operator \mathbf{B} that is employed to simulate the action of correlation matrix on a vector in state space. In this chapter, a general procedure for constructing a BEC model as a rational function of the diffusion operator \mathbf{D} is presented and analytic expressions for the respective correlation functions in the homogeneous case are obtained. It is shown that this class of BEC models can describe multi-scale stochastic fields whose characteristic scales can be expressed in terms of the polynomial coefficients of the model. In particular, the connection between the inverse binomial model and the well-known Gaussian model $\mathbf{B}_g = \exp \mathbf{D}$ is established and the relationships between the respective decorrelation scales are derived.

By its definition, the BEC operator 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, therefore an efficient numerical approximation is needed. Under the assumption of local homogeneity of \mathbf{D} , a heuristic method for computing the diagonal elements of \mathbf{B} is proposed. It is shown that the method is sufficiently accurate for realistic applications, and requires 10^2 times less computational resources than other methods of diagonal estimation that do not take into account prior information on the structure of \mathbf{B} .

M. Yaremchuk (✉) · M. Carrier · S. Smith · G. Jacobs
Naval Research Laboratory, Stennis Space Center, MS 39529, USA
e-mail: max.yaremchuk@nrlssc.navy.mil

S. K. Park and L. Xu (eds.), *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol II)*, DOI 10.1007/978-3-642-35088-7_8,
©Springer-Verlag Berlin Heidelberg 2013

8.1 Introduction

In recent years, heuristic background error correlation (BEC) modelling has become an area of active research in geophysical data assimilation. Of particular interest are the BEC models constructed with positive functions of the diffusion operator,

$$\mathbf{D} = \nabla \nu \nabla \quad (8.1)$$

where ν is the spatially varying positive-definite diffusion tensor. This type of BEC model is attractive for several reasons: a) it guarantees positive definiteness of the resulting correlation functions (CFs), b) it is computationally inexpensive in most practical applications, and c) it allows straightforward control of inhomogeneity and anisotropy via the diffusion tensor. In the traditional approach of correlation modeling where spatial correlations are specified by prescribed analytical functions, care should be taken to maintain positive definiteness of the respective correlation operator, especially in anisotropic and/or inhomogeneous cases [1, 2].

Among the most popular operators \mathbf{B} used in practical BEC modeling are those using the exponential and the inverse binomial functions of \mathbf{D} :

$$\mathbf{B}_g = \exp(a^2 \mathbf{D}); \quad \mathbf{B}_m = \left(\mathbf{I} - \frac{a^2 \mathbf{D}}{m} \right)^{-m} \quad (8.2)$$

where \mathbf{I} is the identity operator, a is a scaling parameter and m is a positive integer. Since \mathbf{D} has a non-positive spectrum whose larger eigenvalues correspond to the smaller-scale eigenvectors, the operators \mathbf{B}_g and \mathbf{B}_m are positive-definite and suppress small-scale variability. Both types of BEC models (8.2) are extensively used in geophysical applications. Numerically, they are implemented by integration of the diffusion equation using either explicit (in the case of \mathbf{B}_g [3, 4, 5]) or implicit (in the case of \mathbf{B}_m [6, 7]) integration schemes.

A disadvantage of the BEC models (8.2) is that there is a limited freedom in the shape of local CFs, which have either the shape of the Gaussian bell (\mathbf{B}_g) or provide its m th-order strictly positive approximations (\mathbf{B}_m) [8, 9]. In order to allow negative correlations, one has to consider operators generated by the arbitrary polynomials in \mathbf{D} . The quadratic polynomial case was studied recently by Hristopulos and Elogne [10, 11] and Yaremchuk and Smith [9], who obtained analytic representations of the CFs and derived relationships between the polynomial coefficients and the spectral parameters of \mathbf{B} in the homogeneous case.

In a more realistic inhomogeneous setting, the diffusion tensor varies in space, making analytic methods inapplicable. Nevertheless, they can still give a reasonable guidance for quick estimation of the diagonal elements of \mathbf{B} (normalization factors), whose values are crucial for constructing the BEC models. The importance of accurately computing $\text{diag} \mathbf{B}$ is evident from the fact that the operators \mathbf{B} under consideration are formulated numerically as multiplication algorithms by the matrices, whose elements are not explicitly known. On the other hand, since the BEC

operator \mathbf{C} is represented numerically by the correlation matrix, it must have a unit diagonal and, therefore, knowledge of the diagonal elements of \mathbf{B} is required for renormalization:

$$\mathbf{C} = (\text{diag}\mathbf{B})^{-1/2}\mathbf{B}(\text{diag}\mathbf{B})^{-1/2} \quad (8.3)$$

Equation (8.3) shows that the considered BEC models involve two separate algorithms: one for computing the action of \mathbf{B} and another for estimating the normalization factors $(\text{diag}\mathbf{B})^{-1/2}$.

Purser with coauthors [12, 13, 14] were among the first to employ analytic methods for estimating the normalization factors for the Gaussian operator \mathbf{B}_g in geophysical applications. Somewhat earlier, an asymptotic technique was developed for estimating the diagonal of the Gaussian kernel in Riemannian spaces to study quantum effects in general relativity (e.g., [15],[16]). These ideas can be utilized to derive a useful algorithm for estimating the normalization factors.

In this chapter, we first give an overview of the recent developments in constructing the \mathbf{D} -operator BEC models, and illustrate their major features with the examples in the homogeneous case $v = \text{const.}$ In particular, in section 2.2, the relationships between the scaling parameters for the Gaussian model and its m th-order approximation (8.2) are obtained and the respective CFs are given. In section 2.3 the inverse binomial model is extended to an arbitrary polynomial of \mathbf{D} : Expressions for the CFs and normalization factors are derived, and relationships are established between the structure of the BEC spectrum and the polynomial coefficients. In section 3, after a brief overview of the diagonal estimation methods, a heuristic formula for computing $\text{diag}\mathbf{B}_g$ is derived (section 3.2) and then tested numerically against other methods in a set of realistic oceanographic applications (sections 3.3-3.5). Results of similar tests with the \mathbf{B}_m model are also presented. Summary and discussion of the prospects for the \mathbf{D} -operator BEC modeling complete the chapter.

8.2 Diffusion operator and covariance modeling

The convenience of the diffusion operator (8.1) for constructing the BEC models can be explained by the non-negative spectrum of $-\mathbf{D}$: An operator that is generated by a positive rational function F of $-\mathbf{D}$ whose eigenvalues tend to zero at large wavenumbers, is positive-definite and has a smoothing property, i.e. tends to suppress high-frequency components of the solution. In this section we consider two types of such functions: Those that are generated by the m th-order binomials (Section 2.2) and the others by the inverse of a positive polynomial (Section 2.3). To allow analytical treatment, anisotropic homogeneous case in the boundless domain is considered.

The benefit of analytical consideration is its ability to reveal local correlation structure and therefore provide a reasonable guidance to construction of more general operators \mathbf{B} . In addition, as it has been shown recently, good approximations to $\text{diag}\mathbf{B}$ can be obtained by using analytical results obtained with the homogeneous

versions of \mathbf{B} (e.g., [12, 17, 18]). Therefore, analytical formulas describing homogeneous BEC operators are of significant practical interest. The analytical results may facilitate practical design of the cost functions in variational data assimilation problems, because they give explicit relationships between the shape of the local CFs and the structure of the corresponding BEC operator.

8.2.1 Correlation functions and normalization

Consider an anisotropic, homogeneous diffusion operator (8.1) in \mathbb{R}^n , $n = 1, \dots, 3$, with $\mathbf{x} \in \mathbb{R}^n$ representing points in the physical space. By using the coordinate transformation $\mathbf{x}' = \nu^{-1/2}\mathbf{x}$, the problem can be reduced to considering isotropic operators of the form

$$\mathbf{B} = F(-\Delta), \quad (8.4)$$

where Δ is the Laplacian (e.g., [8, 10]) and F is an arbitrary positive function. In the case of an inhomogeneous diffusion ($\nu \neq \text{const}$) the global transformation cannot be found. Transformations of this type, however, can be used locally for constructing \mathbf{B} and the normalization factors (Section 3). All of the formulas that are written below are assumed to be in the transformed coordinates \mathbf{x}' with primes omitted to simplify the notation.

The operator (8.4) is diagonalized with the Fourier transform, and the diagonal elements are $B(\mathbf{k}) = F(\mathbf{k}^2)$ where \mathbf{k} is the Fourier coordinate (wavenumber). Because of homogeneity, the matrix elements of \mathbf{B} in the \mathbf{x} -representation depend only on the distance $r = |\mathbf{x}|$ from the diagonal. They can be computed by applying the inverse Fourier transform to $B(\mathbf{k})$:

$$B^n(\mathbf{x}) = (2\pi)^{-n} \int_{\mathbb{R}^n} B(\mathbf{k}) \exp(-i\mathbf{k}\mathbf{x}) d\mathbf{k}. \quad (8.5)$$

By integrating over the directions in \mathbb{R}^n (Appendix 1), (8.5) can be reduced to

$$B^n(r) = (2\pi)^{-n/2} \int_0^\infty B(k) k^{n-1} (kr)^s J_{-s}(kr) dk \quad (8.6)$$

where $k \equiv |\mathbf{k}|$, J denotes the Bessel function of the first kind, and $s = 1 - n/2$. The respective matrix elements of the correlation operator (CFs) are obtained by normalization:

$$C^n(r) = B^n(r)/B^n(0) \quad (8.7)$$

In practical applications, the diffusion operator is not homogeneous, and the analytic representations (8.6–8.7) cannot be obtained. However, the action of \mathbf{B} on a state vector can be computed numerically at a relatively low cost. The major problem with such modelling is the efficient estimation of the diagonal elements

$$\mathbf{B}^n(\mathbf{x}, \mathbf{x}) \equiv \int_{\mathbb{R}^n} \mathbf{B}^n(\mathbf{x}, \mathbf{y}) \delta(\mathbf{x} - \mathbf{y}) d\mathbf{y} \quad (8.8)$$

which are necessary to rescale \mathbf{B} to have its diagonal elements equal to unity. In practice, the rescaling factors $N^n(\mathbf{x})$ are defined as reciprocals of $\mathbf{B}^n(\mathbf{x}, \mathbf{x})$.

Computing the integral (8.8) numerically is expensive, because the convolutions with the δ -functions have to be performed at all points \mathbf{x} of the numerical grid. However, reasonable approximations [12, 18] for $N^n(\mathbf{x})$ can be obtained by using asymptotic expansions of (8.8) under the assumption of weak inhomogeneity (see Section 3).

8.2.2 The Gaussian model and its binomial approximations

The Gaussian-shaped correlation model is widely used in geophysical applications. Numerically, it is implemented by approximating $\exp(a^2 \mathbf{D}/2)$ with the binomial:

$$\mathbf{B}_g(\mathbf{D}) = \exp\left(\frac{a^2 \mathbf{D}}{2}\right) \approx \left(\mathbf{I} + \frac{a^2 \mathbf{D}}{2m}\right)^m, \quad (8.9)$$

where m is a large positive integer. This numerical approach is often referred to as "integration of the diffusion equation" and has been used in practice for several decades [3, 4, 5, 7]. There is, however, a certain disadvantage associated with the numerical stability of the integration: The number of "integration time steps" m has to be large enough for the eigenvalues of the binomial operator in the rhs of (8.9) to be less than 1 in the absolute value. This constraint may limit m from below by a large value, which can make the computation rather expensive.

Another option is to use a different approximation in (8.9):

$$\mathbf{B}_m(\mathbf{D}) = \left(\mathbf{I} - \frac{a^2 \mathbf{D}}{2m}\right)^{-m}. \quad (8.10)$$

The eigenvalues of the operator in the rhs of (8.10) do not exceed 1, and the "integration procedure" is unconditionally stable. This approach is often referred to as "implicit integration of the diffusion equation" (see Appendix 2). and has been used in many practical applications as well [6, 7, 19].

In the Fourier representation both models (8.9) and (8.10) approximate the same Gaussian function of k :

$$\mathbf{B}_e^n(k) = \left[1 - \frac{a^2 k^2}{2m}\right]^m \approx \exp\left(-\frac{a^2 k^2}{2}\right) \quad (8.11)$$

$$\mathbf{B}_m^n(k) = \left[1 + \frac{a^2 k^2}{2m}\right]^{-m} \approx \exp\left(-\frac{a^2 k^2}{2}\right) \quad (8.12)$$

Since the value of m in (8.11) is fairly large in practice, the resulting CF is hardly distinguishable from a Gaussian-shaped curve with a half-width a .

Substituting (8.12) into (8.6), integrating over k , and normalizing the result by $\mathbf{B}_m^n(0)$ yields the CFs of the Matern family [20] enumerated by $s = m - n/2$ and scaled by $a_* = a/\sqrt{2m}$:

$$C_m^n(\rho) = \frac{\rho^s K_s(\rho)}{2^{s-1} \Gamma(s)}, \quad (8.13)$$

where $\rho = r/a_*$, Γ is the gamma-function and K stands for the modified Bessel function of the second kind [22]. The respective normalization factors are

$$N_m^n = \frac{\sqrt{\pi} \Gamma(m)}{\Gamma(m-1/2)} \omega_n a_*^n \quad (8.14)$$

where $\omega_1 = 2$, $\omega_2 = 2\pi$, and $\omega_3 = 4\pi$. In the limiting case of $m \rightarrow \infty$, the CFs (8.13) take the Gaussian form:

$$C_\infty^n = \exp(-r^2/2a^2); \quad n = 1, \dots \quad (8.15)$$

Consecutive approximations of the Gaussian CF by (8.13) are shown in Figure 1. It is remarkable that when $m = 1$, the CFs (8.13) have singularities at $\rho = 0$ in both two and three dimensions (see also Table 1). This means that *in the continuous case* the first-order approximations become invalid when $n > 1$. Numerically, however, the CFs do exist for $n > 1$, but their decorrelation scale is limited by the grid size δ (the corresponding CF is shown by the dotted line in the left panel of Fig. 1). This occurs because the numerical analogue of the δ -function is never singular, but has a finite amplitude inversely proportional to the volume of a grid cell, therefore, resulting in a finite value of the convolution (8.8) even if it is infinite in the continuous case. After normalization by that finite value, the CF is 1 at $r = 0$, but its effective decorrelation scale remains proportional to the local grid size.

The left panel in Figure 1 shows that low-order binomial approximations (8.13) underestimate the decorrelation scale a of the target Gaussian function. This unpleasant property can be corrected by optimizing the value of a in (8.10) to obtain the best fit with the Gaussian CF. Since the Gaussian and its approximating functions are both positive and have similar shapes, a reasonable optimization criterion is to set their integral decorrelation scales equal to each other:

$$\int_0^\infty C_m^n(\rho) dr \equiv \frac{a_{opt}}{\sqrt{2m}} \int_0^\infty C_m^n(y) dy = \int_0^\infty \exp(-\frac{r^2}{2a^2}) dr = \frac{\sqrt{\pi} a}{\sqrt{2}}. \quad (8.16)$$

Expression (8.16) shows that $a_{opt} = \xi_m^n a$, where the rescaling coefficient ξ_m^n is defined as:

$$\xi_m^n = \sqrt{\pi m} \left[\int_0^\infty C_m^n(y) dy \right]^{-1} = \frac{\Gamma(s)}{\Gamma(s+1/2)} \sqrt{m}. \quad (8.17)$$

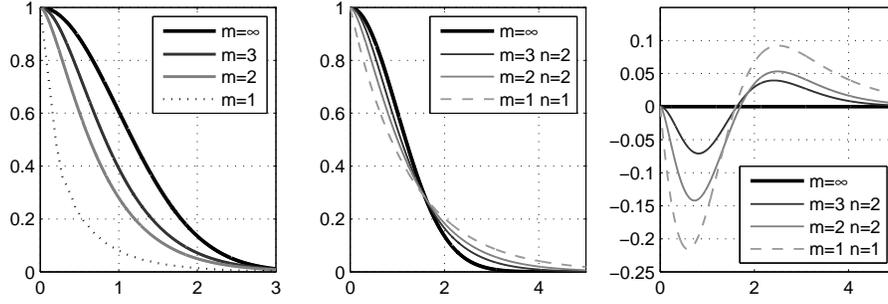


Fig. 8.1 Left: Binomial approximations (8.13) of the Gaussian CF in two dimensions ($n = 2$). The CF for $m = 1$ is shown by the dotted line for the numerical realization with the grid step $\delta = a/4$. Middle: Same approximations, but with optimally adjusted correlation radii for various combinations of m and n . Right: Differences between the Gaussian CF and its approximations shown in the middle panel. The horizontal axes are scaled by a .

The values of ξ_m^n for $m, n < 4$ and their respective approximation errors

$$e_m^n = \int_0^\infty |C_m^n - C_\infty| dr / \left[\int_0^\infty |C_\infty| dr \right]$$

are assembled in Table 1.

Table 8.1 Correlation functions associated with the power approximations (8.10) of the Gaussian CF in n dimensions. The CFs for $n = 1$ and 3 are rewritten in terms of elementary functions for convenience. The correlation radius adjustment coefficients ξ_m^n are shown below the formulas together with the corresponding relative errors e_m^n in approximation of the Gaussian CF (bold numbers).

	$n = 1$	$n = 2$	$n = 3$
$m = 1$	$\frac{\exp(-\rho)}{\sqrt{\pi}}$ 0.33	$K_0(\rho)$ —	$\frac{\exp(-\rho)}{\rho}$ —
$m = 2$	$\frac{(1+\rho)\exp(-\rho)}{\sqrt{\pi/2}}$ 0.13	$\frac{\rho K_1(\rho)}{\sqrt{8/\pi}}$ 0.19	$\frac{\exp(-\rho)}{\sqrt{2\pi}}$ 0.33
$m = 3$	$\frac{(1+\rho+\rho^2/3)\exp(-\rho)}{\sqrt{27\pi/8}}$ 0.08	$\frac{\rho^2 K_2(\rho)/2}{\sqrt{16/3\pi}}$ 0.10	$\frac{(1+\rho)\exp(-\rho)}{\sqrt{3\pi/4}}$ 0.13

The coefficients ξ_m^n along with relationship (8.12) provide an expression for estimating the scaling parameter in the binomial model (8.10) which approximates the Gaussian-shaped CF with a given radius a :

$$a_{binom} = \xi_m^n a / \sqrt{2m} \quad (8.18)$$

8.2.3 The inverse polynomial model

A certain disadvantage of the binomial models (8.9) and (8.10) is their inability to represent oscillating CFs whose spectra may have multiple maxima. This issue can be overcome by considering the BEC models of the form.

$$\mathbf{B} = \left[\mathbf{I} + \sum_{j=1}^J a_j \mathbf{D}^j \right]^{-1} \quad (8.19)$$

Here a_j are the real numbers, constrained by the positive definiteness requirement of \mathbf{B} . In the Fourier representation, the operator (8.19) acts as multiplication by the inverse of the polynomial in k^2 , and the positive-definiteness property translates into the requirement that the spectral polynomial

$$B^{-1}(k^2) = 1 + \sum_{j=1}^J a_j (-k^2)^j \quad (8.20)$$

to be positive for all $k^2 > 0$. This constraint is equivalent to the statement that the rhs of (8.20) must not have real positive roots. Therefore, $B^{-1}(k^2)$ can also be represented in the form

$$B^{-1}(k^2) = \frac{1}{Z} \prod_{m=1}^M (k^2 + z_m^2)(k^2 + \bar{z}_m^2), \quad (8.21)$$

where $M = J/2$,

$$Z = \prod_m |z_m^2|^2, \quad (8.22)$$

the overline denotes the complex conjugate, and $z_m = a_m + ib_m$ are arbitrary complex numbers with $\text{Im}(z_m^2) \neq 0$. In its general form, the polynomial (8.21) is additionally multiplied by the product of the arbitrary number of real negative roots ($b_m = 0$). The ensuing analysis of (8.21) will be simplified by omitting the product (summation) limits over m and assuming there are no real negative or multiple roots. The latter requirement is not restrictive in practice, because location of the roots is never known exactly, and the BEC spectrum can always be well approximated by (8.21) [21].

It is instructive to note that the polynomial (8.21) can also be rewritten as

$$B^{-1}(k^2) = \frac{1}{Z} \prod_{m=1}^M (a_m^2 + (k - b_m)^2)(a_m^2 + (k + b_m)^2), \quad (8.23)$$

Compared to the spectral representation (8.20), representation (8.23) has the advantage that its free parameters are not constrained by the positive-definiteness require-

ment, and they have a sensible meaning of the scales (b^{-1}) and "energies" (a^{-1}) of the modes forming the spectrum.

Using equation (8.6), the matrix elements of \mathbf{B} can now be written as

$$B^n(r) = \frac{Zr^{-s}}{(2\pi)^{\frac{n}{2}}} \int_0^\infty \frac{k^{s+1} J_s(kr) dk}{\prod_m (k^2 + z_m^2)(k^2 + \bar{z}_m^2)}, \quad (8.24)$$

where $s = n/2 - 1$. The integral in (8.24) can be taken by decomposing

$$B(k) = \frac{Z}{\prod_m (k^2 + z_m^2)(k^2 + \bar{z}_m^2)} \quad (8.25)$$

into elementary fractions:

$$B(k) = \sum_m \left[\frac{q_m}{k^2 + z_m^2} + \frac{\bar{q}_m}{k^2 + \bar{z}_m^2} \right], \quad (8.26)$$

where

$$q_m = \frac{Z}{(\bar{z}_m^2 - z_m^2) \prod_{j \neq m} (z_m^2 - z_j^2)(z_m^2 - \bar{z}_j^2)} \quad (8.27)$$

After substitution of (8.26) into (8.24), the integral is reduced to the sum of Hankel-Nicholson type integrals [22] and can be taken explicitly, yielding

$$B^n(r) = \frac{2r^{2-n}}{(2\pi)^{\frac{n}{2}}} \sum_m \langle q_m \rho_m^s K_s(\rho_m) \rangle \quad (8.28)$$

where $\rho_m = z_m r$, and angular brackets denote taking the real part (cf. eq. (8.13)).

The corresponding correlation functions $C^n(r)$ are obtained through normalizing (8.28) by $B^n(0)$. The first three values at $r = 0$ are

$$B^1(0) = \sum_m \langle q_m \bar{z}_m \rangle |z_m|^{-2} \quad (8.29)$$

$$B^2(0) = -\frac{1}{\pi} \sum_m \langle q_m \log z_m \rangle \quad (8.30)$$

$$B^3(0) = -\frac{1}{2\pi} \sum_m \langle q_m z_m \rangle \quad (8.31)$$

The normalization factors can be found by integrating $C^n(r)$ over \mathbb{R}^n :

$$N^n = \frac{2}{B^n(0)} \sum_m \frac{\langle q_m \bar{z}_m^2 \rangle}{|z_m|^4} \quad (8.32)$$

Relationships (8.28)-(8.32) provide analytical expressions for the CFs and the normalization factors.

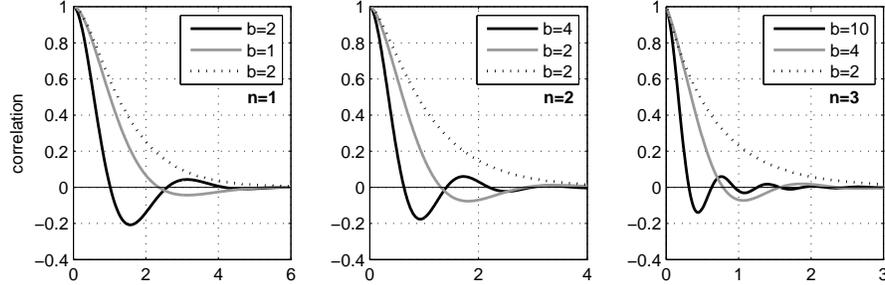


Fig. 8.2 Two-parameter CFs corresponding to the inverse BEC (8.21) with $a = 1, M = 1$. The horizontal axis is scaled by a . Dotted lines show CFs corresponding to the special case with two negative roots $k_1^2 = -a, k_2^2 = -b$ not described by the spectral polynomial (8.21).

In the important case of the quadratic polynomial ($M = 1$) the BEC model is defined by two parameters a, b (Figure 2). Expressions for the respective CFs in 1- and 3-dimensional cases can be rewritten in terms of the elementary functions [9, 21]

$$C^1(a, b, r) = \frac{\sqrt{a^2 + b^2}}{b} \exp(-ar) \cos(br - \arctan \frac{a}{b}) \quad (8.33)$$

$$C^3(a, b, r) = \exp(-ar) \frac{\sin(br)}{br} \quad (8.34)$$

and the normalization factors are given by

$$N^1 = \frac{4a}{a^2 + b^2}; \quad N^2 = \frac{8\pi ab}{2(a^2 + b^2)^2 \arctan(b/a)}; \quad N^3 = \frac{8\pi a}{(a^2 + b^2)^2} \quad (8.35)$$

In practical applications, a BEC model is often constructed by fitting the spectral (8.25) or correlation (8.28) functions to those derived from experimental data. These functions are characterized by $2m$ parameters which give enough freedom for approximating complex spectra. The approximation procedure can be formulated as a least squares problem in $2m$ dimensions, which may be rather difficult to solve due to the non-linearity of B with respect to the fitting parameters a_m and b_m . Therefore, it is useful to have guidance on how the BEC model parameters are related to the scales and amplitudes of the physical modes that contribute to the experimental spectrum (Fig. 3).

The contribution of the m th mode to the spectrum can be assessed by integrating the right hand side of (8.26):

$$E_m = \int_0^{\infty} \left[\frac{q_m}{k^2 + z_m^2} + \frac{\bar{q}_m}{k^2 + \bar{z}_m^2} \right] dk = \frac{\pi \langle q_m \bar{z}_m \rangle}{|z_m|^2} \quad (8.36)$$

In the limit when distances $|b_l - b_m|$ between the spectral peaks of \mathbf{B} are much larger than their half-widths a_m , (i.e. $a_m/b_m \ll 0$ in particular), equation (8.36) can be simplified using the asymptotic approximations

$$z_m \approx ib_m; \quad q_m \approx \frac{b_m^3}{4ia_m\Pi_m}; \quad \Pi_m \equiv \prod_{j \neq m} (1 - b_m^2/b_j^2)^2$$

to yield

$$E_m \approx \frac{\pi b_m^2}{4a_m\Pi_m}. \quad (8.37)$$

Asymptotic values of the spectral density at the peaks are respectively

$$B(b_m) \approx \frac{b_m^2}{4a_m^2\Pi_m} = \frac{E_m}{\pi a_m}, \quad (8.38)$$

i.e. the peak amplitudes are inversely proportional to b_m^2 and to the square of the mode scale a_m^{-1} . Expressions (8.36)–(8.38) can be useful in generating the first guess values for z_m to initialize an iterative procedure of approximating experimental data.

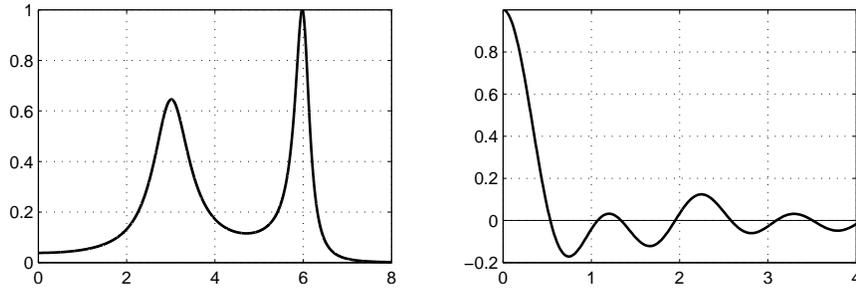


Fig. 8.3 An example of the normalized spectrum (left) and the respective correlation function (right) for the fourth-order polynomial (8.26) in two dimensions. ($M = 2$; $z_1 = .5 + 3i$; $z_2 = .2 + 6i$).

After the model parameters are established, the action of \mathbf{B}^{-1} can be computed recursively (cf. equations (8.21)–(8.22):

$$\mathbf{B}^{-1} = \prod_m [\mathbf{I} - |z_m^2|^{-2} \mathbf{D}(2\langle z_m^2 \rangle \mathbf{I} - \mathbf{D})] \quad (8.39)$$

The inverse BEC model (8.39) can then be employed to compute either the action of \mathbf{B} with an iterative inversion algorithm or to directly compute the gradient of a 3dVar cost function involving the quadratic form $\mathbf{x}^T \mathbf{B}^{-1} \mathbf{x}$, where \mathbf{x} is the state vector.

The above analysis gives an insight on the shape of the local CFs and provides a direct connection between the scales described by \mathbf{B} and the polynomial coefficients of the considered BEC models (8.9), (8.10), (8.25) or (8.39). The second important ingredient in constructing the BEC operator \mathbf{C} (eq. (8.3)) is estimating the diagonal elements of \mathbf{B} , which is a more technical but equally important problem.

8.3 Diagonal estimation

8.3.1 Stochastic methods

In the last few decades a large family of stochastic algorithms were developed for estimating elements and traces of extra-large matrices emerging from numerical solutions of the PDEs in applied physics (e.g., [23, 24, 25]). Weaver and Courtier [27] were among the first to use this approach in geophysical applications for estimating the diagonal of the operator (8.9).

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 $\tilde{\mathbf{s}} = \mathbf{B}\mathbf{s}$ according to the formula:

$$\tilde{\mathbf{d}}(\mathbf{x}) = \overline{\mathbf{s} \odot \tilde{\mathbf{s}} \oslash \overline{\mathbf{s} \odot \mathbf{s}}}, \quad (8.40)$$

where the overline denotes averaging over the ensemble and \odot , \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 δ -correlated distributions with zero mean, the contributions to $\tilde{\mathbf{d}}$ from the off-diagonal elements tend to cancel out, and $\tilde{\mathbf{d}}$ converges to $\mathbf{d} = \text{diag}\mathbf{B}$ as $K \rightarrow \infty$. More accurately, the squared relative approximation error

$$\varepsilon^2(\mathbf{x}) = (\tilde{\mathbf{d}} - \mathbf{d})^2 / \mathbf{d}^2 \quad (8.41)$$

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{B} 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.

The above described Monte-Carlo (MC) technique was developed further by Bekas et al [26], who noticed that the method may converge to \mathbf{d} in the *finite* number of iterations that equals to the matrix dimension N 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 $N \times N$ Hadamard matrix (HM), which span the model's state space (see Appendix 3).

In the numerical experiments below we use MC and HM techniques as testbeds for the diagonal estimation methods which can be derived from analytical considerations and are exploit prior knowledge of the structure of \mathbf{B} .

8.3.2 Locally homogeneous approximations

Consider homogeneous ($v = \text{const}$) operators (8.2) with $a^2 = 1/2$ and assume that the coordinate axes are aligned along the eigenvectors of the diffusion tensor, whose (positive) eigenvalues are $\lambda_i^2, i = 1, \dots, n$. Then the matrix elements of $\mathbf{B}_{g,m}$ can be written down explicitly as

$$\mathbf{B}_g(\mathbf{x}, \mathbf{y}) = \exp(\mathbf{D}/2) = d \exp\left[\frac{-\rho^2}{2}\right] \quad (8.42)$$

$$\mathbf{B}_m(\mathbf{x}, \mathbf{y}) = (\mathbf{I} - \mathbf{D}/2m)^{-m} = d \frac{\bar{\rho}^s K_s(\bar{\rho})}{2^{s-1} \Gamma(s)} \quad (8.43)$$

where

$$\rho = \sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{v}^{-1} (\mathbf{x} - \mathbf{y})}$$

is the distance between the correlated points (measured in terms of the smoothing scales λ_i), $d = (2\pi)^{-n/2} \Omega^{-1}$ are the (constant) diagonal elements of $\mathbf{B}_{g,m}$, $\Omega = \Pi \lambda_i = \sqrt{\det v}$ is the diffusion volume element, and $\bar{\rho} = \sqrt{2m\rho}$.

When v varies in space, (8.42-8.43) are no longer valid, and the diagonal elements \mathbf{d} depend on \mathbf{x} and the type of the \mathbf{B} operator. However, if we assume that v is locally homogeneous (LH), i.e. varies in space on a typical scale L which is much larger than λ_i , the diagonal elements $\mathbf{d}(\mathbf{x})$ can be expanded in the powers of the small parameter $\varepsilon = \bar{\lambda}/L$, where $\bar{\lambda}$ is the mean eigenvalue of \sqrt{v} . The zeroth-order LH approximation term (LH0) is apparently

$$\mathbf{d}^0(\mathbf{x}) = (2\pi)^{-n/2} \Omega(\mathbf{x})^{-1} \quad (8.44)$$

because for infinitely slow variations of v ($L \rightarrow \infty$), the normalization factors must converge to the above expression for the constant diagonal elements d . It is noteworthy that the formula (8.44) is found to be useful even in the case of strong inhomogeneity $\varepsilon \geq 1$. In particular, numerical experiments of Mirouze and Weaver [17] have shown that such an approximation provided 10% errors in a simplified 1d case.

The accuracy of (8.44) can formally be increased by considering the next term in the expansion of the diagonal elements of $\mathbf{B}_{g,m}$. The technique of such asymptotics has been well developed for the diagonal of the Gaussian kernel (8.42) in Riemannian spaces (e.g., [15, 16]). More recently, the approach was considered by Purser [13, 14] in the atmospheric data assimilation context. The application of this tech-

nique to the diffusion operator (8.1) in flat space yields the following asymptotic expression for the diagonal elements of \mathbf{B}_g in the local coordinate system where $\mathbf{v}(\mathbf{x})$ is equal to the identity matrix, and \mathbf{D} takes the form of the Laplacian operator:

$$\mathbf{B}_g(\mathbf{x}, \mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \left[1 - \frac{1}{2} \text{tr} \mathbf{h} - \frac{1}{12} \left(\frac{\Delta}{2} \text{tr} \mathbf{h} + \nabla \cdot \text{div} \mathbf{h} \right) \right] + O(\varepsilon^5) \quad (8.45)$$

Here \mathbf{h} is a small ($|\mathbf{h}| \sim \varepsilon$) correction to \mathbf{v} within the vicinity of \mathbf{x} . Note that the terms in the parentheses have the order $O(\varepsilon^3)$, because each spatial differentiation adds an extra power of ε .

The asymptotic estimate (8.45) involves second derivatives which tend to amplify errors in practical applications when ε may not be small. Therefore, using (8.45) in its original form could be inaccurate even at a moderately small value of ε . To increase the computational efficiency, it is also desirable to formulate the first-order approximation as a linear operator, which acts on $\mathbf{d}^0(\mathbf{x})$. Keeping in mind that $|\mathbf{h}| \sim \varepsilon$, and utilizing the relationships:

$$\mathbf{d}^0(\mathbf{x}) = (2\pi)^{-n/2} \Omega(\mathbf{x})^{-1} \approx (2\pi)^{-n/2} \left(1 - \frac{1}{2} \text{tr} \mathbf{h} \right) \quad (8.46)$$

$$\exp(\Delta/2) \approx \mathbf{I} + \frac{1}{2} \Delta, \quad (8.47)$$

the second term in the parentheses of (8.45) can be represented as follows:

$$\nabla \cdot \text{div} \mathbf{h} = \frac{1}{n} \Delta \text{tr} \mathbf{h} + \nabla \cdot \text{div} \mathbf{h}' \quad (8.48)$$

where \mathbf{h}' is the traceless part of \mathbf{h} . On the other hand, if the divergence of \mathbf{h}' is neglected, the equation (8.45) can be rewritten in the form

$$\mathbf{B}_g(\mathbf{x}, \mathbf{x}) \approx \frac{1}{(2\pi)^{n/2}} \left(1 + \gamma_n \frac{\Delta}{2} \right) \left(1 - \frac{1}{2} \text{tr} \mathbf{h} \right) \quad (8.49)$$

where

$$\gamma_n = \frac{1}{6} + \frac{1}{3n}. \quad (8.50)$$

Taking (8.46-8.47) into account and replacing Δ by \mathbf{D} , the desired ansatz for the first-order approximation (LH1) of the diagonal elements is obtained:

$$\mathbf{d}_g^1 = \exp \left(\gamma_n \frac{\mathbf{D}}{2} \right) \mathbf{d}_g^0 \quad (8.51)$$

The relationship (8.51) was derived by Purser et al. [12] for the one-dimensional case ($\gamma_1 = 0.5$) and tested by Mirouze and Weaver [17], who reported a significant (2-4 times) improvement of the accuracy in 1d simulations.

An estimate similar to (8.51) can also be obtained for \mathbf{B}_m , possibly with a different coefficient $\tilde{\gamma}_n$. It is assumed, however, that $\tilde{\gamma}_n$ may not differ too much from γ_n given similarity in the shapes (Fig. 1) of the correlation functions (8.42)-(8.43). Furthermore, because of the approximate nature of (8.51), the best representation of $\mathbf{d}(\mathbf{x})$ in realistic applications may be achieved with a value of γ_n that is significantly different from the one given by (8.50). For this reason, a more general form of (8.51) was adopted in the numerical experiments, assuming

$$\mathbf{d}_g^1(\mathbf{x}) \approx \exp[\gamma \mathbf{D}/2] \mathbf{d}_g^0(\mathbf{x}); \quad \mathbf{d}_2^1(\mathbf{x}) \approx [\mathbf{I} - \gamma \mathbf{D}/4]^{-2} \mathbf{d}_2^0(\mathbf{x}) \quad (8.52)$$

for the Gaussian model and its second-order ($m = 2$) spline approximation (10).

The following experiments investigate the dependence of the respective approximation errors $\langle \varepsilon_{g,2} \rangle$ on the free parameter γ .

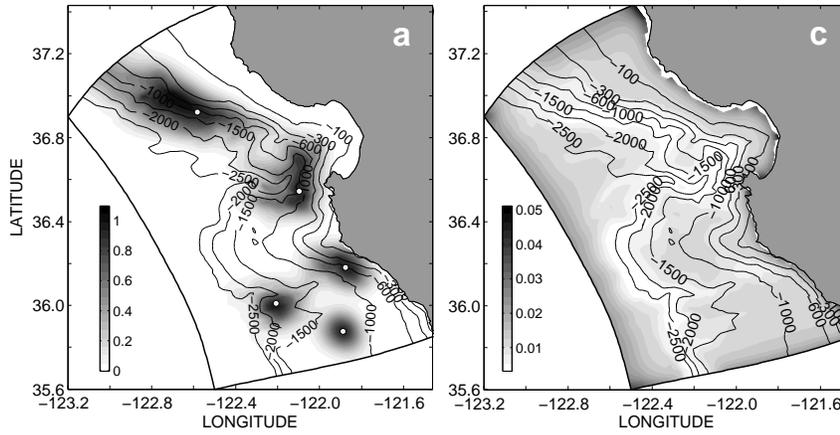


Fig. 8.4 Left: A composite map of five columns of the \mathbf{B}_g operator. White circles denote locations of the diagonal elements of the corresponding correlation matrices. Right panel shows the map of the non-normalized diagonal elements of \mathbf{B}_g . Depth contours are in meters.

8.3.3 Numerical results

To assess the efficiency of the methods outlined in Sections 3.1-3.2, two series of numerical experiments with realistically inhomogeneous BEC models are performed. In the first series the methods were tested in the 2d case with the state vector having a 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 $N \sim 10^6$.

8.3.3.1 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) [28] set up in the Monterey Bay (Fig. 4). The number N of grid points (dimension of the state space) was 3,438. A vector field $\mathbf{u}(\mathbf{x})$ was used to generate the diffusion tensor as follows. The smaller principal axis λ_2 of $\sqrt{\mathbf{v}}$ is set to be orthogonal to \mathbf{u} 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 λ_1 is set to be equal to $\max(1, \sqrt{|\mathbf{u}|/u})\lambda_2$, where u is a prescribed threshold value of $|\mathbf{u}|$. If \mathbf{u} 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 λ_2 .

In the 2d experiments, the vector field \mathbf{u} is generated by treating bottom topography $h(\mathbf{x})$ (Fig. 4) as a stream function. 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 the next two sections are performed using the BEC models (8.42-8.43) with the parameters $n = m = 2$. A composite map of five columns of \mathbf{B}_g is shown in Figure 4a. The diffusion operator (1) is constrained to have zero normal derivative at the open and rigid boundaries of the domain in both 2d and 3d experiments.

Numerically, the action of \mathbf{B}_g 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 v , starting from the "initial condition" \mathbf{y}_0 . The minimum number of "time steps" required for the scheme's stability in such a setting was 5,256. The action of \mathbf{B}_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 2,000-2,500. To make the shapes of the \mathbf{B}_g and \mathbf{B}_2 compatible (Fig. 1), the diffusion tensor in \mathbf{B}_2 was multiplied by $8/\pi$ (see Table 1).

The exact values $\mathbf{d}(\mathbf{x})$ of the diagonal elements are shown in Figure 4b. Their magnitude appears to be lower in the regions of "strong currents" (large \mathbf{u}), as the corresponding δ -functions are dispersed over larger areas by diffusion. $\mathbf{d}(\mathbf{x})$ are higher near the boundaries because part of the domain available for dispersion is screened by the condition that prescribes zero flux across either open or rigid boundaries.

8.3.3.2 Statistical methods

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 ε is computed using equation (8.41). In both series, the rates of reduction of ε with iteration k are similar and closely follow the \sqrt{k} law (upper gray line in Fig. 5a).

To improve the accuracy, the MC estimates are low-pass filtered with the corresponding \mathbf{B} -operators at every iteration (Fig. 5b). To optimize the filter, the diffusion operators in $\mathbf{B}_{g,2}$ are multiplied by the tunable parameter γ , which effectively reduced the mean decorrelation (smoothing) scale $\gamma^{-1/2}$ times. The lower lines in Figure 5a demonstrate the result of such optimal smoothing: this procedure resulted in an almost four-fold reduction of the domain-averaged error $\langle \varepsilon \rangle$ to 0.1 after performing 60 iterations (averaging over 60 ensemble members).

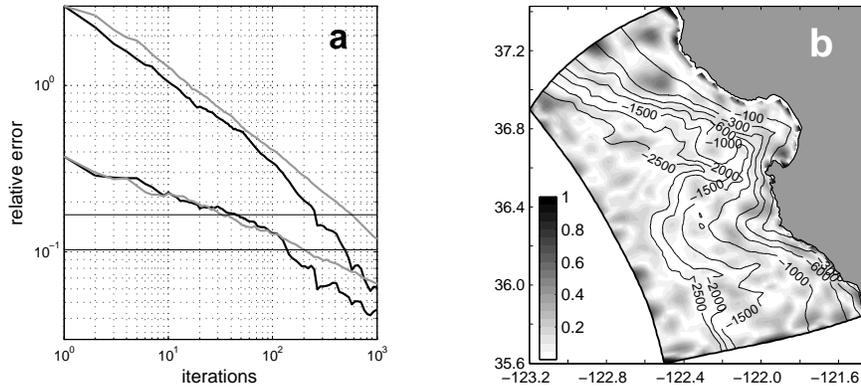


Fig. 8.5 a) reduction of the domain-averaged diagonal estimation error $\langle \varepsilon \rangle$ with iterations for the HM (black) and MC (gray) methods for the \mathbf{B}_2 model. The lower curves are obtained after optimal smoothing of the estimates. The thin horizontal lines show the error levels that are provided by the asymptotic zeroth- ($\langle \varepsilon \rangle = 0.17$) and first-order ($\langle \varepsilon \rangle = 0.10$) methods which do not require iterative schemes. b) Horizontal distribution of $\varepsilon(\mathbf{B}_2)$ after 60 iterations of the HM method with smoothing.

Experiments with the HM method (black curves in Fig 5a) show that horizontal smoothing significantly improves the accuracy of the estimates, especially after the first few dozens of iterations. Comparison with the MC method (gray curves in Fig. 5a) demonstrates a noticeable advantage of the HM technique (black curves), which remains visible at higher iterations $k > 100$ even after smoothing (lower curves). This advantage increases with increasing 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 the iteration number k .

From the practical point of view, it is not reasonable to do more than several hundred iterations, as $\langle \varepsilon \rangle$ drops to the value of a few per cent (Fig. 5a), which is much smaller than the accuracy in the determination of the background error variances. It can therefore be concluded that it is advantageous to use the HM technique, when making more than a hundred iterations is computationally affordable.

8.3.3.3 Asymptotic expansion method

Since the principal axes of the diffusion tensor at every point are defined by construction, computation of the zeroth-order approximation (8.44) to the normalization factors is not expensive. Near the boundaries, however, the factors described by (8.44) have to be adjusted by taking into account the geometric constraints imposed on the diffusion. This adjustment was computed for points located closer than $3\lambda_1$ from the boundary and it was assumed that the boundary had negligible impact on the shape of the diffused δ -function [18].

Figure 6 demonstrates the horizontal distribution of the error $\varepsilon(\mathbf{x})$ obtained by approximating the diagonal elements of \mathbf{B}_g with (8.44) (zeroth-order LH method, or LH0) and with (8.51), (the first-order LH method LH1). Despite an apparent violation of the LH assumption in many regions (e.g., λ_1 changes from 20δ to the background value of 3δ 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{B}_2 model have shown similar results (16% and 10% errors).

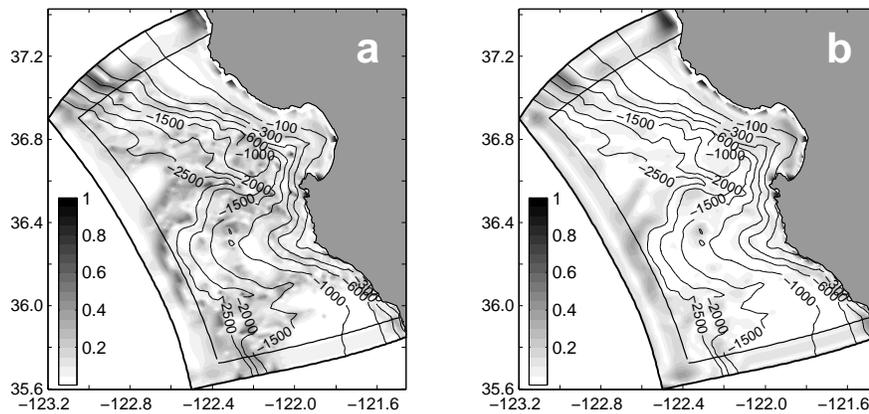


Fig. 8.6 Diagonal approximation errors under the zeroth-order (a), and first-order (b) LH methods for the \mathbf{B}_g model. The thin black line inside the boundaries shows the domain of error averaging.

Another series of experiments are performed with the varying scaling parameter γ to find an optimal fit to \mathbf{d} . Computations were made for $0 \leq \gamma \leq 1$. The best result for \mathbf{B}_g was obtained for $\gamma_2 = 0.30$ which is fairly consistent with the value ($\gamma_2 = 0.33$) given by (8.50). In the case of the \mathbf{B}_2 operator, the optimal value is $\gamma_2 = 0.24$, still in a reasonable agreement with (8.50), given the strong inhomogeneity of v and deviation of the \mathbf{B}_2 operator from the Gaussian form. A somewhat

smaller value of $\gamma_2(\mathbf{B}_2)$ 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 v , and, therefore, requires less smoothing in the next approximation.

8.3.4 Numerical efficiency

Table 2 provides an overview of the performance for the tested methods. For comparison purposes we show CPU requirements by the smoothed MC and HM methods after they achieve the accuracies of the LH0 and LH1 methods. It is seen that both MC and HM 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 diagonal estimation method is similar to the cost of the 3dvar analysis itself, which required several hundred iterations. The remarkable CPU saving are due to the fact that the LH methods explicitly take into account information on the local structure of \mathbf{B} which can be derived by analytical methods. Comparison of the

Table 8.2 Relative CPU times required by the MC and HM methods to achieve the accuracies $\langle \varepsilon \rangle$ of the LH0 and LH1 methods (shown in brackets).

	MC/LH0	MC/LH1	HM/LH0	HM/LH1
\mathbf{B}_g	755 (.19)	1205 (.09)	680 (.19)	520 (.09)
\mathbf{B}_2	780 (.17)	490 (.10)	850 (.17)	330 (.10)

spatial distributions of the approximation error $\langle \varepsilon \rangle(\mathbf{x})$ (Fig. 5b, 6b) favor the LH methods as well: They show significantly less small-scale variations and may have a potential for further improvement. Comparing Fig. 5b and 6b also shows that, in contrast to the statistical methods, LH0 errors tend to increase in the regions of strong inhomogeneity, but 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 HM pattern (Fig. 5b). This feature can be partly attributed to certain inaccuracy in estimation of the near-boundary elements. However, there is certainly room for further improvement with the issue.

8.3.5 LH experiments in 3d setting

To check the performance of the LH0 and LH1 methods further, a larger 3d NCOM domain is set up in the Okinawa region (Fig. 7) with horizontal resolution of 10 km and 45 vertical levels. The state vector dimension N (total number of the grid points) in this setting was 862,992.

Because of the large N , 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 $\langle \varepsilon \rangle$ is estimated by averaging over these points.

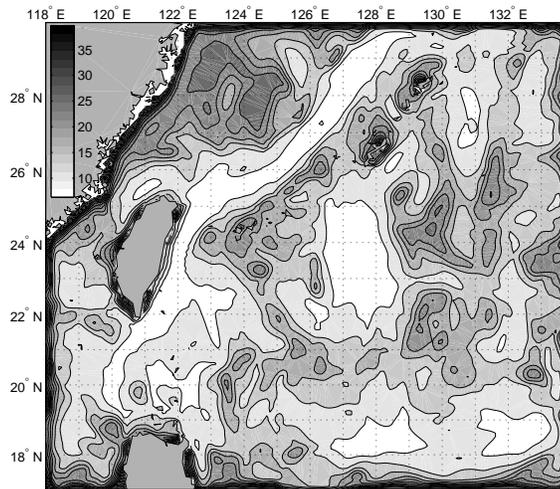


Fig. 8.7 Diagonal elements of \mathbf{B}_g in the Okinawa region at $z = 20m$. The actual values are multiplied by 10^4 .

The diffusion tensor is constructed in the same way that is described in Section 3.1, but the generating field $\mathbf{u}(\mathbf{x})$ is taken to be the horizontal velocity field from an NCOM run. The value of λ_3 (in the vertical direction) is independent of horizontal coordinates, but varies in the vertical as $3\delta_z$, where δ_z is the vertical grid step. Figure 7 illustrates spatial variability of the \mathbf{B}_g diagonal elements at $z=20$ m. The smallest values are observed in the regions of the Kuroshio and the North Equatorial Current, where the largest velocities are observed, and the $\Omega = \sqrt{\det \mathbf{v}}$ reaches its largest values (eq. 8.44). To better test the algorithm, a relatively small threshold value of $v=0.02$ m/s is prescribed, so that diffusion is anisotropic in more than 90% of the grid points.

major issues associated with this type of models: construction of a positive-definite smoothing operator \mathbf{B} as a rational function of \mathbf{D} and the estimation of $\text{diag}\mathbf{B}$.

In section 2 analytic relationships between the polynomial coefficients of \mathbf{B} and the parameters controlling the shape of correlation functions were derived. Although only homogeneous operators in boundless domains were considered, these relationships provide reasonable guidance to constructing more realistic BEC operators, especially in cases when the typical scale of variability of the diffusion tensor is much larger than the local decorrelation scale ρ_c and/or most of the observations are separated from the boundaries by distances, exceeding ρ_c . In a similar way, weak inhomogeneity can be introduced by variable coefficients $z_m(\mathbf{x})$, and the local CF shapes can be assessed using (8.13), (8.28-8.31).

Similar issues have been recently studied by many authors (e.g., [8, 10, 11, 17]). In particular, analytic formulas analogous to (8.33)-(8.35), were derived in somewhat different setting by Hristopulos and Elogne [10, 11] whom considered quadratic polynomials of similar structure. Xu [8] analyzed Taylor expansions of $\exp\mathbf{B}$ and obtained recursive relations for the polynomial coefficients associated with an arbitrary CF. Mirouze and Weaver [17] demonstrated a possibility to generate oscillating CFs using higher-order polynomials in one dimension.

Relationships (8.28-8.32) generalize these results for the polynomial model of an arbitrary order M . We assume, however, that the inverse quadratic model ($M = 1$) is of major practical interest for two reasons. First, the BEC operators that are encountered in GFD applications are rarely homogeneous and observational statistics are usually insufficient to capture the details of the spatial variability of the CFs. Therefore, experimental estimates of the BECs are either limited to low-rank ensemble estimates or have to rely on the very rough assumption of homogeneity. Needless to say, that in the latter case the structure of a sample CF should be elaborated with sufficiently low detailization and be well accounted for by a two-parameter BEC model (Fig. 2). The second reason is that the use of higher-order polynomials considerably degrades the conditioning of the linear systems that are being solved in the assimilation process and, therefore, may require sophisticated preconditioners.

The second equally important aspect of the d-operator BEC modeling is the computational efficiency of estimating the diagonal elements of \mathbf{B} . Two types of the BEC operators were considered: with the Gaussian-shaped kernel \mathbf{B}_g and with the kernel generated by the second-order binomial approximation to \mathbf{B}_g . The tested techniques include the "stochastic" MC and HM methods, which retrieve $\text{diag}\mathbf{B}$ iteratively from its action on a sequence of model state vectors, 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 regimes: the zeroth (LH0) and the first-order (LH1) approximations.

Numerical experiments conducted with realistic diffusion tensor models show that: a) the HM technique proves to be superior in efficiency compared to the MC technique when accuracies of less than 10% error ($k > 100$) are required; b) both stochastic methods require 300-1000 times more CPU time to achieve the accuracy, compatible with the most efficient LH1 method; c) with the Gaussian model, the

LH1 method demonstrates the best performance with the value of the smoothing parameter γ compatible with the one given by the relationship (8.50) derived from the asymptotic approximation of the Gaussian kernel diagonal. In deriving the ansatz (8.51) for the LH1 model, we followed the approach of Purser et al [12], whom proposed to smooth the zeroth-order diagonal by the square-root of the BEC operator in the 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 (8.45) 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 v , we utilized a non-local empirical modification, still fully consistent with the original expansion in the limit $\bar{\lambda}/L \rightarrow 0$, but sufficiently robust with respect to the numerical errors related to the high-order derivatives of v . A similar technique was developed by Purser [12, 13], who used empirical saturation functions to stabilize higher-order approximations of the \mathbf{B}_g .

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 the BEC operator – a negligible amount compared to the cost of a 3dVar analysis. Therefore, 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 accuracy to 3-10% error in realistically inhomogeneous BEC models.

A separate question, that requires 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{B} , but not their structure. This approximation is only partly consistent with the zero normal flux conditions for d , but can be avoided if one uses "transparent" boundary conditions (e.g. [17]) which do not require computation of the adjustment factors. On the other hand, it might be beneficial to keep physical (no-flux) boundary conditions in the formulation of d , as they are likely to bring more realism to the dynamics of the BE field.

Another important issue is parameterization of $v(\mathbf{x})$ using the background fields and their statistics. In the simple diffusion tensor model used in the experiments, anisotropic BE propagation is governed by the background velocity field and superimposed on the small-scale isotropic BE diffusion, which takes place at scales that are not well resolved by the grid (less than 3δ). More sophisticated parameterizations of $v(\mathbf{x})$ are surely possible and require further studies. In particular, recent studies have shown that since $v(\mathbf{x})$ has only $n(n+1)/2$ independent components, it can be estimated from ensembles of moderate ($\sim 100n$) size with reasonable accuracy [29, 30, 31, 32]. Finally, the considered BEC models could also be effectively used for adaptive/flow-dependent covariance localization [33, 34], which is an issue of crucial importance in improving the forecast skill of the state-of-the-art data assimilation systems.

Acknowledgements This study was supported by the Office of Naval Research (Program element 0602435N) as part of the project "Exploring error covariances in variational data assimilation".

Appendix 1

Let θ be the angle between \mathbf{x} and \mathbf{k} in \mathbb{R}^n and $n > 2$. Then the integral (8.5) can be rewritten in spherical coordinates as

$$B^n(r) = (2\pi)^{-n} \int_0^\infty B(k) \int_{\Omega_{n-1}} \exp(-ikr \cos \theta) k^{n-1} dk d\Omega_{n-1}, \quad (8.53)$$

where $d\Omega_{n-1}$ is the element of the surface area of the unit sphere. Since $\cos \theta$ changes symmetrically within the limits of integration, the imaginary part of the exponent vanishes. Furthermore, using the identity $d\Omega_{n-1} = d\Omega_{n-2} \cdot \sin^{n-2} \theta d\theta$, the integral (8.53) can be rewritten as

$$B^n(r) = (2\pi)^{-n} \int_0^\infty B(k) k^{n-1} dk \int_{\Omega_{n-2}} d\Omega_{n-2} \int_0^\pi \cos(kr \cos \theta) \sin^{n-2} \theta d\theta \quad (8.54)$$

Integration over θ and substitution of the formula for the surface of $(n-2)$ -dimensional unit sphere into (8.54) yields (8.6).

The general relationship (8.6) also holds for $n = 1, 2$ although these cases require a special (less complicated) treatment.

Appendix 2

In practice, the matrix elements of the operator (8.10) are never calculated explicitly due to the immense cost of such a computation. Instead, the result $\hat{\mathbf{x}}^m(\mathbf{x})$ of the action by \mathbf{B} on a (discrete) model state vector $\hat{\mathbf{x}}^0(\mathbf{x})$ is calculated by solving the linear system of equations

$$\left(\mathbf{I} - \hat{\mathbf{D}}/2m \right)^m \hat{\mathbf{x}}^m = \hat{\mathbf{x}}^0, \quad (8.55)$$

where $\hat{\mathbf{D}}$ denotes the discretized diffusion operator. If $\hat{\mathbf{x}}^0(\mathbf{x})$ represents the "initial state" and the "time step" δt is prescribed such that the "integration time" is $m\delta t = 1$, then action of the operator (8.55) can be identified as a result of a discrete-time integration of the diffusion equation $\partial_t \mathbf{x} = \mathbf{D}/2\mathbf{x}$ with the implicit scheme

$$\hat{\mathbf{x}}^i - \hat{\mathbf{x}}^{i-1} = \frac{1}{2} \delta t \hat{\mathbf{D}} \hat{\mathbf{x}}^i, \quad i = 1, \dots, m \quad (8.56)$$

starting from the initial state $\hat{\mathbf{x}}^0$. Here i denotes the time step number.

Similarly, the action of $\exp(\mathbf{D}/2)$ is never computed by convolving a state vector $\hat{\mathbf{x}}^0$ with the discretized kernel (8.42), but rather by the discrete-time integration of the diffusion equation with the explicit numerical scheme

$$\hat{\mathbf{x}}^i - \hat{\mathbf{x}}^{i-1} = \frac{1}{2} \delta t \hat{\mathbf{D}} \hat{\mathbf{x}}^{i-1}, \quad i = 1, \dots, m \quad (8.57)$$

such that

$$\hat{\mathbf{x}}^m = \left(\mathbf{I} + \hat{\mathbf{D}}/2m \right)^m \mathbf{x}^0 \quad (8.58)$$

in correspondence with the asymptotic relation (8.9) for the Gaussian kernel \mathbf{B}_g .

Appendix 3

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 H_N is an $N \times N$ Hadamard matrix, then

$$H_{2N} = \begin{bmatrix} H_N & H_N \\ H_N & -H_N \end{bmatrix}$$

is also an HM. Starting from $H_2 = [1 \ 1; 1 \ -1]$, the HMs with order $N = 2^n$, $n = 1, 2, \dots$ can be easily constructed. HMs with $N = 12, 20$ were constructed "manually" more than a century ago. A more general HM construction algorithm, which employs the Galois fields theory, was found in 1933. In the present study we used the MatLab software that only handles the cases when $M/12$, or $M/20$ is a power of 2. Despite this restriction, the available values of M were sufficient for purposes of this chapter.

References

1. Gaspari, G., S. E. Cohn, J. Guo and S. Pawson, Construction and application of covariance functions with variable length-fields. *Q. J. R. Meteorol. Soc.*, **132**, 1815, (2006).
2. Gregori, P., E. Porcu, J. Mateu and Z. Sasvari, On potentially negative space time covariances obtained as sum of products of marginal ones, *Ann. Inst. Stat. Math.*, **60**, 865, (2008).
3. Derber, J., and A. Rosati, A global oceanic data assimilation system, *J. Phys. Oceanogr.*, **19**, 1333 (1989).
4. Egbert, G. D., A. F. Bennett, and M. G. G. Foreman, Topex/Poseidon tides estimated using a global inverse model, *J. Geophys. Res.*, **99**, 24821, (1994).

5. Weaver, A. T., J. Vialard, and D. L. T. Anderson, 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, (2003).
6. Ngodock, H., E., B.S. Chua, and A. F. Bennett, 2000: Generalized inversion of a reduced gravity primitive equation ocean model and tropical atmosphere ocean data, *Mon. Wea. Rev.*, **128**, 1757, (2000).
7. 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, Weak and strong constraint data assimilation in the Inverse Ocean Modelling System (ROMS): development and application for a baroclinic coastal upwelling system, *Ocean Modelling*, **16**, 160, (2007).
8. Xu, Q., Representations of inverse covariances by differential operators, *Adv. Atm. Sci.*, **22**(2), 181, (2005).
9. Yaremchuk, M. and S. Smith, On the correlation functions associated with polynomials of the diffusion operator, *Q. J. R. Met. Soc.*, **137**, 1927-1932, (2011).
10. Hristopoulos D. T. and S. N. Elogne, 2007: Analytic properties and covariance functions of a new class of generalized Gibbs random fields. *IEEE Trans. Inform. Theory* **53**, 4467-4679.
11. Hristopoulos D. T. and S. N. Elogne, 2009. Computationally efficient spatial interpolators based on Spartan spatial random fields. *IEEE Trans. Signal Processing*, **57**, 3475-3487.
12. 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.
13. 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*, 48pp.
14. Purser, R. J., 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 parametrized expansion method, *NOAA/NCEP Office Note 457*, 55pp.
15. Gusynin, V.P., and V. A. Kushnir, On-diagonal heat kernel expansion in covariant derivatives in curved space, *Class. Quantum Gravity*, **8**, 279-285, (1991).
16. Avramidi, I. G., Covariant techniques for computation of the heat kernel, *Reviews in Math. Physics*, **11**, 947-980, (1999).
17. Mirouze, I., and A. Weaver, Representation of the correlation functions in variational data assimilation using an implicit diffusion operator, *Q. J. R. Met. Soc.*, **136**, 1421, (2010).
18. Yaremchuk, M. and M. Carrier, On the renormalization of the covariance operators *Mon. Wea. Rev.*, **140**, 639-647, (2012).
19. Carrier, M., and H. Ngodock, Background error correlation model based on the implicit solution of a diffusion equation, *Ocean Modelling*, **35**, 45-53, (2010).
20. Stein, M. L., *Interpolation of spatial data. Some theory for kriging*. Springer, New York, (1999).
21. Yaremchuk, M. and A. Sentchev, Multi-scale correlation functions associated with the polynomials of the diffusion operator. *Q. J. R. Met. Soc.*, **138**, (2012) (in press).
22. Abramowitz, M. and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs and mathematical tables*. Dover Publications, New York, (1972).
23. Girard, D. F., 1987: Un algorithme simple et rapide pour la validation croisee generalisee sur des problemes de grande taille, *RR 669-M, Grenoble, France: Informatique et Mathematiques Appliquees de Grenoble*.
24. Dong, S-J., and K-F. Liu, 1994: Stochastic estimation with Z_2 noise, *Phys. Lett. B*, **328**, 130-136.
25. Hutchison, M.F., 1989: A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines, *J. Commun. Stat. Simulations*, **18**, 1059-1076.
26. Bekas, C.F., E. Kokiopoulou, and Y. Saad, 2007: An estimator for the diagonal of a matrix, *Appl. Num. Math.*, **57**, 1214-1229.

27. Weaver, A. and P. Courtier, Correlation modelling on the sphere using a generalized diffusion equation, *Q. J. R. Meteorol. Soc.*, **127**, 1815, (2001).
28. Martin, P., C. Barron, L. Smedstad, A. Wallcraft, R. Rhodes, T. Campbell, and C. Rowley, Software Design Description for the Navy Coastal Ocean Model (NCOM) Ver. 4.0. *Naval Research Laboratory Report NRL/MR/7320-08-9149*, 151 pp, (2008).
29. Belo-Pereira, M. and L. Berre, The use of ensemble approach to study the background error covariances in a global NWP model, *Mon. Wea. Rev.*, **134**, 2466, (2006).
30. Pannekoucke, O. and S. Massart, Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation, *Q. J. R. Met. Soc.*, **134**, 1425, (2008).
31. Pannekoucke, O. and L. Berre, and G. Desroziers, Background-error correlation length-scale estimates and their sampling statistics, *Q. J. R. Met. Soc.*, **134**, 497, (2008).
32. Berre, L. and G. Desroziers, Filtering of background error variances and correlations by local spatial averaging: a review, *Mon. Wea. Rev.*, **138**, 3693, (2010).
33. Bishop, C. and D. Hodyss, Flow adaptive moderation of spurious ensemble correlations and its use in ensemble data assimilation *Q. J. R. Met. Soc.*, **133**, 2029, (2007).
34. Bishop, C. and D. Hodyss, Adaptive ensemble covariance localization in ensemble 4D-VAR estimation, *Mon. Wea. Rev.*, **139**, 1241, (2011).