

On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation

A. T. Weaver* and I. Mirouze

CERFACS/SUC URA 1875, Toulouse, France

*Correspondence to: A. T. Weaver, CERFACS, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France.
E-mail: anthony.weaver@cerfacs.fr

Differential operators derived from the explicit or implicit solution of a diffusion equation are widely used for modelling background-error correlations in geophysical applications of variational data assimilation. Key theoretical results underpinning the diffusion method are reviewed. Solutions to the isotropic diffusion problem on both the spherical space \mathbb{S}^2 and the d -dimensional Euclidean space \mathbb{R}^d are considered first. In \mathbb{R}^d the correlation functions implied by explicit diffusion are approximately Gaussian, whereas those implied by implicit diffusion belong to the larger class of Matérn functions which contains the Gaussian function as a limiting case. The Daley length-scale, defined as $D = \sqrt{-d/\nabla^2 c(r)}|_{r=0}$ where ∇^2 is the d -dimensional Laplacian operator and $r = |\mathbf{r}|$ is Euclidean distance, is used as a standard parameter for comparing the different isotropic functions $c(r)$. Diffusion on \mathbb{S}^2 is shown to be well approximated by diffusion on \mathbb{R}^2 for length-scales of interest. As a result, fundamental parameters that define the correlation model on \mathbb{S}^2 can be specified using more convenient expressions available on \mathbb{R}^2 .

Anisotropic Gaussian or Matérn correlation functions on \mathbb{R}^d can be represented by a diffusion operator furnished with a symmetric and positive-definite diffusion tensor. For anisotropic functions $c(\mathbf{r})$, the tensor $D = -(\nabla \nabla^T c(\mathbf{r})|_{r=0})^{-1}$ where ∇ is the d -dimensional gradient operator, is a natural generalization of the (square of) the Daley length-scale for characterizing the spatial scales of the function. Relationships between this tensor, which we call the Daley tensor, and the diffusion tensor of the explicit and implicit diffusion operators are established. Methods to estimate the elements of the local Daley tensor from a sample of simulated background errors are presented and compared in an idealized experiment with spatially varying covariance parameters. Since the number of independent parameters needed to specify the local diffusion tensor is of the order of the total number of grid points N , sampling errors are inherently much smaller than those involved in the order N^2 estimation problem of the full correlation function. While the correlation models presented in this paper are general, the discussion is slanted to their application to background-error correlation modelling in ocean data assimilation. Copyright © 2012 Royal Meteorological Society

Key Words: correlation functions; covariance modelling; background error; ocean data assimilation; diffusion tensor; ensemble estimation

Received 30 June 2011; Revised 13 March 2012; Accepted 20 March 2012; Published online in Wiley Online Library 17 May 2012

Citation: Weaver AT, Mirouze I. 2013. On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Q. J. R. Meteorol. Soc.* **139**: 242–260. DOI:10.1002/qj.1955

1. Introduction

Various methods have been proposed for modelling background-error correlations in geophysical applications of variational data assimilation (VDA) (see Bannister, 2008, for example, for a thorough review of methods used in atmospheric VDA). In ocean VDA, background-error correlation models based on the diffusion equation are popular. The method has its origins in the work of Derber and Rosati (1989), who proposed the use of an iterative Laplacian grid-point filter in order to approximate a Gaussian correlation operator. Egbert *et al.* (1994) described a close variant of the algorithm in which the Laplacian grid-point filter could be interpreted as a pseudo-time-step integration of a diffusion equation with an explicit scheme. Weaver and Courtier (2001) (hereafter WC01) described the algorithm in more detail and proposed various extensions to account for more general correlation functions than the quasi-Gaussian of the original Derber and Rosati (1989) algorithm. Correlation models based on explicit diffusion methods have been used in various VDA systems in oceanography (Weaver *et al.*, 2003; Di Lorenzo *et al.*, 2007; Muccino *et al.*, 2008; Daget *et al.*, 2009; Kurapov *et al.*, 2009; Moore *et al.*, 2011), meteorology (Bennett *et al.* 1996), and atmospheric chemistry (Geer *et al.*, 2006; Elbern *et al.*, 2010).

An explicit diffusion scheme is appealing because of its simplicity, but can be expensive if many iterations are required to keep the scheme numerically stable. This can occur when the local diffusion scale is 'large' relative to the local grid size. To keep the explicit scheme affordable, the correlation length-scales must be bounded even if statistics or physical considerations suggest that larger values would be more appropriate. This limitation can be overcome by reformulating the diffusion model using an implicit scheme which has the advantage of being unconditionally stable.

One-dimensional (1D) implicit diffusion operators have been used for representing temporal and vertical correlation functions (Bennett *et al.*, 1997; Chua and Bennett, 2001; Ngodock, 2005) and products of 1D implicit diffusion operators have been used for constructing two-dimensional (2D) and three-dimensional (3D) correlation models (Chua and Bennett, 2001; Zaron *et al.*, 2009). The correlation kernels associated with the 1D implicit diffusion operator belong to the family of M th-order autoregressive (AR) functions where M is the number of implicit iterations (Mirouze and Weaver, 2010; hereafter MW10). As discussed by MW10, the 1D implicit diffusion operator is closely linked to the recursive filter (Lorenc, 1992; Hayden and Purser, 1995), which has been developed extensively in meteorology for constructing correlation models in multiple dimensions (Wu *et al.*, 2002; Purser *et al.*, 2003a, 2003b; Liu *et al.*, 2007). The recursive filter has also been employed in ocean data assimilation systems (Martin *et al.*, 2007; Dobricic and Pinardi, 2008; Liu *et al.*, 2009).

The 1D implicit diffusion approach for constructing 2D and 3D correlation models can be convenient for computational reasons, but has limitations. For example, with few iterations, the product of 1D implicit diffusion operators produces a well-known spurious anisotropic response (Purser *et al.*, 2003a). Unphysical features can also appear near complex boundaries, such as coastlines or islands in an ocean model, where correlation functions cannot always be reasonably represented by a product of

separable functions of the model's coordinates. Correlation models based on 2D or 3D implicit diffusion operators can overcome these limitations but are more difficult to implement since they involve the solution of a large linear system (matrices of dimension $\mathcal{O}(10^6 \times 10^6)$ or larger in VDA). Some progress in the development of this approach has been made by Weaver and Ricci (2004) and Massart *et al.* (2012), who used sparse matrix methods to solve a 2D implicit diffusion problem directly, and by Carrier and Ngodock (2010) and S. Gratton (2011, personal communication), who used iterative methods based on conjugate gradient or multi-grid to approximate the solution of a 2D or 3D implicit diffusion problem.

Multidimensional implicit diffusion correlation operators can be interpreted in terms of smoothing norm splines, which were introduced to atmospheric data assimilation by Wahba and Wendelberger (1982) and Wahba (1982), and discussed within an oceanographic context by McIntosh (1990). In the norm spline approach, the background term of the cost function in VDA is formulated in terms of a linear combination of weighted derivative operators that penalize explicitly the amplitude and curvature of the solution. When the weighting coefficients are given by binomial coefficients, the inverse of the background-error correlation operator implied by the norm spline can be expressed as the inverse of an implicit diffusion operator. The direct penalty approach was popular in some of the early studies of four-dimensional VDA (Thacker, 1988; Sheinbaum and Anderson, 1990) but generally leads to a poorly conditioned minimization problem (Lorenc *et al.*, 2000). Effective preconditioning techniques for VDA require access to the background-error covariance operator itself. An interesting exception is the recent study of Yaremchuk *et al.* (2011), who propose a variational formulation in which the inverse of the background-error covariance is modelled directly using the inverse of a low-order (two-iteration) 3D implicit diffusion operator. No apparent conditioning problems were reported in their examples from an ocean VDA system.

The present paper has a dual purpose: first, to provide a review of the diffusion equation as a basis for constructing anisotropic and inhomogeneous correlation models for data assimilation; and second, to illustrate how fundamental parameters that control spatial smoothness properties of these models can be estimated using ensemble methods. Section 2 brings together key results from data assimilation and geostatistics on the isotropic diffusion problem. Diffusion is considered both on the sphere and in the d -dimensional Euclidean space. Analytical expressions for the isotropic correlation functions implied by appropriately normalized explicit and implicit diffusion in these spaces are presented and compared. The Daley length-scale is used as a standard parameter for comparing the different functions, and expressions relating it to the parameters of the diffusion-model are established.

The results from section 2 provide the foundation for building anisotropic correlation models with the diffusion equation. This is discussed in sections 3 and 4. The Daley tensor is introduced, which is defined as the negative inverse of the tensor of second derivatives of the correlation function evaluated at zero distance (the Hessian tensor). The Daley tensor is an anisotropic generalization of the Daley length-scale. Expressions relating the Daley tensor to the diffusion tensor of the diffusion models are given. Section 4 discusses techniques for estimating the Daley

tensor from statistics of a sample of simulated errors such as those that would be available from an ensemble data assimilation system. Idealized experiments are then presented to compare the effectiveness of two of the estimation techniques. Conclusions are given in section 5. Appendix A provides a derivation of the relationship between the Daley and diffusion tensors for the correlation functions represented by the implicit diffusion equation in \mathbb{R}^2 . Appendix B provides a derivation of the key formulae for estimating the Daley tensor.

2. Isotropic diffusion

Coordinate systems of global atmospheric and ocean models refer to the spherical-shell geometry of the atmosphere and ocean. From a mathematical perspective, this leads naturally to consideration of 2D ‘horizontal’ correlation functions on the spherical space \mathbb{S}^2 . The product of a 2D correlation function on \mathbb{S}^2 and a 1D correlation function on the bounded subset of the Euclidean space \mathbb{R}^1 is commonly used to construct 3D correlation functions on the spherical-shell subspace of \mathbb{R}^3 that defines the model domain. This approach of separating the horizontal and vertical correlation functions is usually justified by the fact that the global atmospheric and ocean circulations are characterized by scales that are much larger in the horizontal direction (along geopotential surfaces) than in the vertical direction (perpendicular to geopotential surfaces). In the remainder of this section, the correlation functions that can be represented by isotropic diffusion on \mathbb{S}^2 and the general Euclidean space \mathbb{R}^d are described. Table 1 provides a brief description of the main symbols used in this section.

2.1. Explicit diffusion on \mathbb{S}^2

Consider the 2D diffusion equation applied to the scalar field $\eta(\lambda, \phi, s)$:

$$\frac{\partial \eta}{\partial s} - \kappa \nabla^2 \eta = 0, \tag{1}$$

where $\kappa > 0$ is a diffusion coefficient, and

$$\nabla^2 = \frac{1}{a^2 \cos \phi} \frac{\partial}{\partial \phi} \left(\cos \phi \frac{\partial}{\partial \phi} \right) + \frac{1}{a^2 \cos^2 \phi} \frac{\partial^2}{\partial \lambda^2}$$

is the Laplacian operator in geographical coordinates (λ, ϕ) , λ denoting longitude ($0 \leq \lambda \leq 2\pi$), ϕ latitude ($-\pi/2 \leq \phi \leq \pi/2$), and a the radius of the sphere (the Earth’s radius in our case). In the context of this paper, s is to be interpreted as a dimensionless pseudo-time coordinate. The diffusion coefficient then has physical units of length squared. The solution of Eq. (1) on \mathbb{S}^2 can be interpreted as a covariance operator (e.g. see WC01). Let

$$\eta(\lambda, \phi, 0) = \gamma^s \tilde{\eta}(\lambda, \phi) \tag{2}$$

denote the initial condition, where γ^s is a normalization constant. The solution at some $s > 0$ can be expressed as the integral operator $C^s : \tilde{\eta} \mapsto \eta(s)$,

$$\eta(\lambda, \phi, s) = \int_{\mathbb{S}^2} c^s(\theta) \tilde{\eta}(\lambda', \phi') a^2 \cos \phi' d\lambda' d\phi', \tag{3}$$

Table 1. A list of the main generic symbols used in section 2. The specification of the superscripts α and β is summarized in the bottom table. A quantity in \mathbb{R}^d is supplemented with a subscript d if it depends explicitly on the dimension of the space; otherwise it is omitted.

Symbol	Description
C^α, C^β	Correlation operators on \mathbb{R}^d and \mathbb{S}^2
$C^\alpha(\mathbf{x}, \mathbf{x}')$	General correlation function on \mathbb{R}^d
\mathbf{x}	Vector of Cartesian coordinates
$c_d^\alpha(r)$	Isotropic correlation function on \mathbb{R}^d
r	Euclidean distance
$\hat{c}_d^\alpha(\hat{\mathbf{x}})$	Fourier transform of $c_d^\alpha(r)$
$\hat{\mathbf{x}}$	Vector of spectral wave numbers
$c^\beta(\theta)$	Isotropic correlation function on \mathbb{S}^2
θ	Angular separation
c_n^β	Legendre coefficients for $c^\beta(\theta)$
n	Total wave number
D^α, D^β	Daley length-scale of $c_d^\alpha(r)$ and $c^\beta(\theta)$
$\gamma_d^\alpha, \gamma^\beta$	Normalization constants on \mathbb{R}^d and \mathbb{S}^2

Superscript	Description	
α	g	Regular diffusion on \mathbb{R}^d
	w	Implicit diffusion on \mathbb{R}^d
β	s	Regular diffusion on \mathbb{S}^2
	h	Implicit diffusion on \mathbb{S}^2

where $c^s(\theta)$ is an *isotropic* function that depends on the angular separation θ , $0 \leq \theta \leq \pi$, between points (λ, ϕ) and (λ', ϕ') on the sphere:

$$\cos \theta = \cos \phi \cos \phi' \cos(\lambda - \lambda') + \sin \phi \sin \phi'. \tag{4}$$

The normalization constant γ^s in Eq. (2) has been absorbed into the function $c^s(\theta)$ which has the specific form

$$c^s(\theta) = \sum_{n=0}^{\infty} c_n^s P_n^0(\cos \theta), \tag{5}$$

where

$$c_n^s = \frac{\gamma^s}{4\pi a^2} \sqrt{2n+1} \exp\left(-\frac{\kappa s}{a^2} n(n+1)\right), \tag{6}$$

n being the total wave number, and $P_n^0(\cos \theta)$ the Legendre polynomials, normalized such that $P_n^0(1) = \sqrt{2n+1}$, following the usual convention in meteorology (Courtier *et al.*, 1998). All isotropic covariance functions on \mathbb{S}^2 can be expressed, as in Eq. (5), as an expansion in terms of the Legendre polynomials (Weber and Talkner, 1993; Theorem 2.11 of Gaspari and Cohn, 1999). They are positive-definite functions on \mathbb{S}^2 if the spectral coefficients are positive, which is clearly the case for all of the coefficients c_n^s . Equation (3) is thus a valid covariance operator on \mathbb{S}^2 .

The covariance function is readily transformed into a correlation function ($c^s(0) = 1$) by defining the normalization constant as

$$\gamma^s = 4\pi a^2 \left(\sum_{n=0}^{\infty} (2n+1) \exp\left(-\frac{\kappa s}{a^2} n(n+1)\right) \right)^{-1}. \tag{7}$$

The fundamental parameter controlling the shape of the correlation function is the *product* κs in Eq. (6). To define

the length-scale of $c^s(\theta)$, we use the standard definition from Daley (1991, p. 110), the geometrical interpretation of which is discussed by Pannekoucke *et al.* (2008). For $c^s(\theta)$, the Daley length-scale reads

$$D^s = \sqrt{-\frac{2}{\nabla^2 c^s|_{\theta=0}}} = a \sqrt{\frac{2}{\sum_{n=0}^{\infty} n(n+1)\sqrt{2n+1} c_n^s}} \quad (8)$$

Equations (6) and (8) provide a relationship between κ_s and D^s that allows us to control the correlation shape (length-scale).

Now consider a discretized version of Eq. (1) in which the first-order derivative is approximated using a forward-Euler (*explicit*) scheme. This yields

$$\eta(\lambda, \phi, s_m) = \eta(\lambda, \phi, s_{m-1}) + \kappa \Delta s \nabla^2 \eta(\lambda, \phi, s_{m-1}), \quad (9)$$

where m is a positive integer, $\Delta s = s_m - s_{m-1}$ is the step size, and ∇^2 is understood to be the Laplacian operator in discretized form. For convenience, we can assume that $s_m = m$ so that the step size $\Delta s = 1$. This parameter can thus be ignored hereafter without loss of generality. Repeated applications of Eq. (9) on the interval $0 < m \leq M$ leads to the linear operator

$$\eta(\lambda, \phi, M) = (1 + \kappa \nabla^2)^M \eta(\lambda, \phi, 0), \quad (10)$$

where $\eta(\lambda, \phi, 0)$ is given by Eq. (2). For clarity we let

$$\kappa = L^2, \quad (11)$$

to emphasize that the coefficient is positive and can be interpreted as the square of a scale parameter.

The key idea is that, on a numerical grid, the effect of the *integral* correlation operator (3) on an arbitrary scalar field $\tilde{\eta}(\lambda, \phi)$ can be approximated by applying a discretized *differential* operator (10). This is the essence of the original Derber and Rosati (1989) scheme. The parameter κs of the correlation function $c^s(\theta)$ can be related to the parameters M and κ of Eq. (10) by noticing that $\kappa s = \kappa M = ML^2$. In practice, it is customary to prescribe the Daley length-scale (D^s). Given D^s , the product ML^2 can be determined by a non-trivial inversion of Eq. (8). This has been done by trial and error for the illustrative examples presented in this paper. To determine M and L^2 from the product ML^2 , we have an additional requirement that M must be sufficiently large (L^2 sufficiently small) in order to maintain the numerical stability of the explicit scheme. Provided M is not too ‘large’, applying the discretized operator (10) is an efficient way of evaluating the integral operator (3). What defines an acceptable value of M will depend on the application.

To represent a larger family of correlation functions than Eqs (5) and (6), WC01 proposed a *generalized* diffusion model in which the scaled Laplacian in Eq. (1) is replaced by a linear combination of powers of scaled Laplacians:

$$-\kappa \nabla^2 \mapsto \sum_{p=1}^P \kappa_p (-\nabla^2)^p, \quad (12)$$

where the diffusion coefficients $\kappa_p > 0$ can be related to a general set of scale parameters L_p via the equation

$$\kappa_p = L_p^{2p}, \quad p = 1, \dots, P. \quad (13)$$

The resulting correlation functions have the same basic form as Eq. (5) but with the c_n^s given by

$$c_n^s = \frac{\gamma^s}{4\pi a^2} \sqrt{2n+1} \exp\left(-\sum_{p=1}^P \frac{\kappa_p s}{a^{2p}} (n(n+1))^p\right), \quad (14)$$

and the appropriate modification to γ^s to produce a unit-amplitude function. Equation (6) is a special case of Eq. (14) with $P = 1$. Unlike the standard diffusion model, the generalized diffusion model can be used to represent correlation functions that change sign, as illustrated in Figure 1 of WC01. This is an appealing feature if there is compelling evidence of negative correlations in the error fields, although representing them with powers of Laplacian operators would clearly increase the cost of the correlation model.

2.2. Explicit diffusion on \mathbb{R}^d

Now consider the diffusion equation (1) on the d -dimensional Euclidean space \mathbb{R}^d , where ∇^2 now represents the Laplacian operator in Cartesian coordinates $\mathbf{x} = (x_1, \dots, x_d)$. While our particular interest concerns the spaces $\mathbb{R}^1, \mathbb{R}^2$ and \mathbb{R}^3 , it is easier to consider them as special cases of the general diffusion problem in \mathbb{R}^d . The initial condition of the diffusion problem can be written as

$$\eta(\mathbf{x}, 0) = \gamma_d^g \tilde{\eta}(\mathbf{x}), \quad (15)$$

where $\gamma_d^g > 0$ is a normalization constant and $\tilde{\eta}(\mathbf{x})$ is assumed to be bounded at infinity. Using the Fourier transform (FT), the solution at ‘time’ $s > 0$ can be written as a convolution operator $\mathcal{C}_d^g : \tilde{\eta} \mapsto \eta(s)$:

$$\eta(\mathbf{x}, s) = \int_{\mathbb{R}^d} C(\mathbf{x}, \mathbf{x}') \tilde{\eta}(\mathbf{x}') d\mathbf{x}', \quad (16)$$

where $C(\mathbf{x}, \mathbf{x}') = c^g(r)$ is the Gaussian function

$$c^g(r) = \frac{\gamma_d^g}{(4\pi\kappa s)^{d/2}} e^{-r^2/4\kappa s}, \quad (17)$$

$r = |\mathbf{x} - \mathbf{x}'|$ being the Euclidean distance between points \mathbf{x} and \mathbf{x}' on \mathbb{R}^d . Setting the normalization factor to

$$\gamma_d^g = (4\pi\kappa s)^{d/2} \quad (18)$$

ensures that $c^g(0) = 1$.

The Daley length-scale for any twice differentiable, isotropic correlation function $c(r)$ in d dimensions is given by

$$D = \sqrt{\frac{d}{\text{tr}(-\nabla \nabla^T c|_{r=0})}} = \sqrt{-\frac{d}{\nabla^2 c|_{r=0}}}, \quad (19)$$

where $\nabla \nabla^T$ is the outer product of the d -dimensional gradient operator $\nabla = (\partial/\partial x_1 \dots \partial/\partial x_d)^T$ and its transpose. The

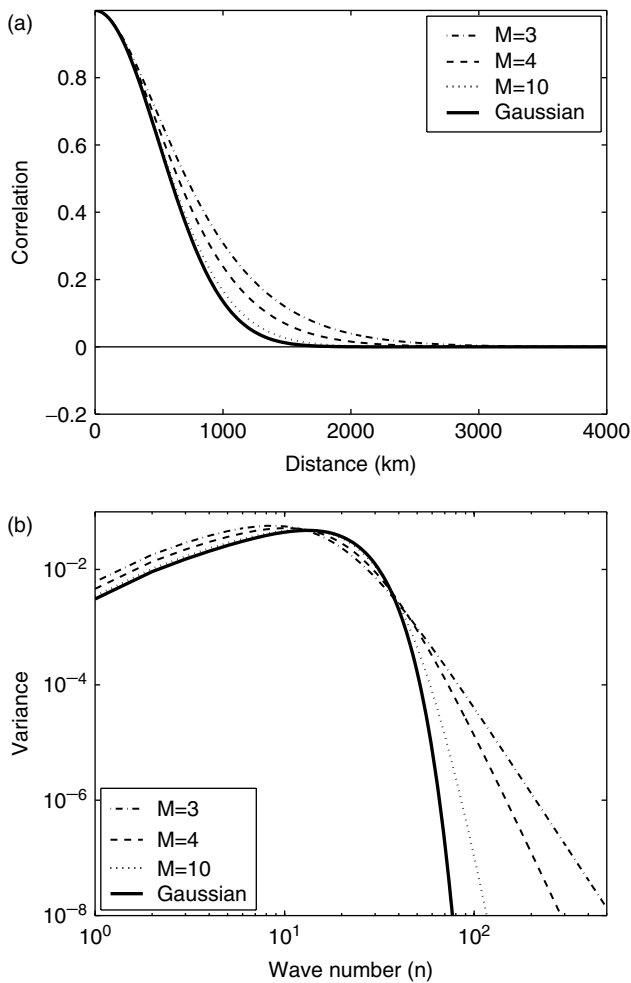


Figure 1. The grid-point values $c^h(r(\theta))$ (upper panel) and the variance-power spectra $\sqrt{2n+1}c_n^h$ (lower panel) of sample correlation functions generated with Eqs (5) and (30) using different values of M . The scale parameters have been set to $L = 353$ km, $L = 250$ km and $L = 125$ km for the functions corresponding to $M = 3$ (dashed-dotted curves), $M = 4$ (dashed curves) and $M = 10$ (dotted curves), respectively, in order to achieve a common Daley length-scale of $D^h = 500$ km (see Eq. (8)). The Gaussian correlation function $c^g(r(\theta))$ (Eq. (24)) with $D^g = 500$ km is shown for reference (thick solid curves). The correlation functions in the upper panel are plotted as a function of chordal distance (Eq. (23)). A spectral truncation at $n = 500$ has been used. The lower panel is plotted on a log-log scale.

quantity within the trace operator is the correlation Hessian tensor (Chorti and Hristopoulos, 2008). The Hessian tensor plays a fundamental role in characterizing the anisotropic correlation functions described later in this paper (sections 3 and 4). For the d -dimensional Gaussian function, the Daley length-scale is

$$D^g \approx \sqrt{2\kappa s}. \tag{20}$$

In terms of D^g , the normalization factor is

$$\gamma_d^g = (2\pi)^{d/2} (D^g)^d. \tag{21}$$

As before, we can approximate Eq. (16) with a differential operator based on a discretization of the diffusion equation using an M -step explicit scheme. In terms of the parameters M and L^2 of the explicit diffusion operator, Eqs (20) and (21) become

$$D^g = \sqrt{2ML} \tag{22}$$

and $\gamma_d^g = (4M\pi)^{d/2} L^d$.

Let us consider now the interpretation of the Gaussian function on \mathbb{S}^2 . First, since \mathbb{S}^2 is embedded in \mathbb{R}^3 , a valid isotropic correlation function on \mathbb{S}^2 can always be constructed from a valid isotropic correlation function in \mathbb{R}^3 by restricting $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{x}' = (x'_1, x'_2, x'_3)$ to be points on the sphere. Expressing these points in geographical coordinates $\mathbf{x} = (a \cos \phi \cos \lambda, a \cos \phi \sin \lambda, a \sin \phi)$ and $\mathbf{x}' = (a \cos \phi' \cos \lambda', a \cos \phi' \sin \lambda', a \sin \phi')$ leads to the chordal distance measure

$$r = r(\theta) = a\sqrt{2(1 - \cos \theta)}, \quad 0 \leq \theta \leq \pi, \tag{23}$$

where $\cos \theta$ is given by Eq. (4). The Gaussian correlation function on \mathbb{R}^3 confined to the subspace \mathbb{S}^2 is thus

$$c^g(r(\theta)) = e^{-(r(\theta))^2/2(D^g)^2} = e^{-a^2(1-\cos \theta)/(D^g)^2}.$$

From Eq. (23) we notice that r depends only on $\cos \theta$, or alternatively θ , and that $\cos \theta = 1 - r^2/2a^2$, where $0 \leq r \leq 2a$. We also recall that all isotropic correlation functions on \mathbb{S}^2 can be expressed as a Legendre expansion that depends only on $\cos \theta$ (Eq. (5)). It is then possible to represent any isotropic correlation function on \mathbb{S}^2 as a function of either r or θ .*

In particular, consider the representation of the Gaussian on \mathbb{S}^2 in terms of the Legendre polynomials. As shown in WC01:

$$c^g(\theta) = \sum_{n=0}^{\infty} c_n^g P_n^0(\cos \theta)', \tag{24}$$

where

$$c_n^g = \tilde{\gamma}^g \sqrt{2n+1} \frac{I_{n+1/2}(\omega)}{I_{1/2}(\omega)}$$

and $\tilde{\gamma}^g = \frac{e^{-\omega} \sinh(\omega)}{\omega}$,

$I_{n+1/2}(\omega)$ denoting the modified Bessel function of fractional order $n + 1/2$, and $\omega = (a/D^g)^2$. In view of the results on \mathbb{R}^d , one might expect that the correlation kernel $c^h(\theta)$ implied by diffusion on \mathbb{S}^2 (Eq. (5)) is similar to the Gaussian correlation function (24) on \mathbb{S}^2 . Indeed, for a given length-scale D^g , it is possible to find a corresponding parameter κs in Eq. (6) such that the difference between $c^h(\theta)$ and $c^g(\theta)$ is ‘small’ (Roberts and Ursell, 1960; Hartman and Watson, 1974). In particular, consider the scales of interest in atmospheric and ocean data assimilation for which $\omega \gg 1$. Matching the $n = 0$ coefficients c_0^h and c_0^g of the Legendre polynomials and noting that

$$\tilde{\gamma}^g \approx \frac{(D^g)^2}{2a^2}$$

for large ω , we obtain the approximation to the normalization factor

$$\gamma^g \approx 2\pi (D^g)^2. \tag{25}$$

*Isotropic correlation functions on \mathbb{S}^2 will be written explicitly as a function of $r(\theta)$ whenever the context requires an interpretation in terms of chordal distance.

Now matching the $n = 1$ coefficients c_1^s and c_1^g and using (25) leads to the approximation

$$\kappa s \approx \frac{(D^g)^2}{2}. \tag{26}$$

WC01 illustrate the excellent agreement between $c^s(\theta)$ and $c^g(\theta)$, particularly for the large scales, for a given length-scale D^g and with κs approximated according to Eq. (26) (see their Figure A1).

Equations (25) and (26) are none other than those derived earlier for the diffusion problem in \mathbb{R}^2 (cf. Eqs (21) and (20) with $d = 2$). In other words, for length-scales small compared to the radius of the Earth, we obtain the somewhat intuitive result that diffusion on the sphere (the C^s operator) is well approximated by diffusion on the 2D Cartesian plane (the C_2^g operator). For calibrating the correlation model, it is then possible to employ the simple expressions (26) and (25) for the length-scale and normalization factor in place of the more complicated expressions (8) and (7).

There are two main drawbacks with the generalized explicit diffusion model of WC01. First, the correlation functions that can be represented by the model have limited flexibility in the spectral domain, especially at high wave numbers where their decay rates are at least as fast as that of the Gaussian function. In data assimilation, this can result in excessive smoothing of small-scale features in the analysis (Purser *et al.*, 2003b). Second, the explicit scheme is subject to a stability criterion that depends on the ratio of the length-scale and grid size, raised to the power of $2P$. As a result, many iterations may be required when the length-scale is large compared with the grid resolution. With the variable coefficient and anisotropic versions of the model discussed later, the computational cost of the algorithm can be especially high. A diffusion model based on an *implicit* formulation can overcome these limitations, as described next.

2.3. Implicit diffusion on \mathbb{S}^2

Consider again the diffusion equation (1) but this time discretized using a backward-Euler (*implicit*) scheme:

$$\eta(\lambda, \phi, s_m) = \eta(\lambda, \phi, s_{m-1}) + \kappa \Delta s \nabla^2 \eta(\lambda, \phi, s_m), \tag{27}$$

where, as in Eq. (10), we can assume $s_m = m$ and hence $\Delta s = 1$, and interpret κ as the square of a scale parameter (Eq. (11)). Rearranging Eq. (27) and applying it repeatedly on the interval $0 < m \leq M$ leads to the ‘reverse-time’ or inverse diffusion operator

$$(1 - \kappa \nabla^2)^M \eta(\lambda, \phi, M) = \eta(\lambda, \phi, 0). \tag{28}$$

Equation (28) can be interpreted as a *roughening* operator as opposed to the diffusion operator itself, which is a *smoothing* operator.

Following Eq. (2), we define the initial condition as

$$\eta(\lambda, \phi, 0) = \gamma^h \tilde{\eta}(\lambda, \phi) \tag{29}$$

where γ^h is a normalization constant. Weaver and Ricci (2004) show that the differential operator $(C^h)^{-1} : \eta(M) \mapsto \tilde{\eta}$ is the *inverse* of a correlation operator $C^h : \tilde{\eta} \mapsto \eta(M)$, where the latter is given by an integral

equation of the form (3), with isotropic correlation function $c^h(\theta)$ of the Legendre form (5) as its kernel. The spectral coefficients of $c^h(\theta)$ are strictly positive and given by

$$c_n^h = \frac{\gamma^h}{4\pi a^2} \sqrt{2n+1} \left(1 + \frac{L^2}{a^2} n(n+1)\right)^{-M}. \tag{30}$$

The normalization factor is

$$\gamma^h = 4\pi a^2 \left(\sum_{n=0}^{\infty} (2n+1) \left(1 + \frac{L^2}{a^2} n(n+1)\right)^{-M}\right)^{-1} \tag{31}$$

and the Daley length-scale is given by Eq. (8) with c_n^s replaced by c_n^h .

In the explicit diffusion model, the only free parameter was the *product* $\kappa s_M = ML^2$ which controls the spatial scale of the quasi-Gaussian correlation kernel (Eq. (6) with $s = s_M$). The implicit diffusion model, on the other hand, allows for greater control of the shape characteristics of the associated correlation kernels since *both* L^2 and M are free parameters. Numerically, this extra flexibility is reflected by the important property of *unconditional stability* of the implicit scheme. In the limiting case of $M \rightarrow \infty$, with ML^2 held fixed, the spectral coefficients (30) reduce to those of the quasi-Gaussian solution which is the only correlation function that can be represented by solving the diffusion equation explicitly.

The upper panel in Figure 1 displays correlation functions $c^h(r(\theta))$ for different values of M and a constant Daley length-scale (500 km). The values are plotted as a function of chordal distance $r(\theta)$. The Gaussian function $c^g(r(\theta))$ is also shown for reference. Increasing the value of M decreases the ‘fatness’ of the tail of $c^h(r(\theta))$, with the Gaussian providing the upper limit as $M \rightarrow \infty$. The total variance of $c^h(r(\theta))$ and $c^g(r(\theta))$ is given by their value at the origin, which is equal to one. The coefficients $\sqrt{2n+1} c_n^h$ and $\sqrt{2n+1} c_n^g$ give the contribution of each wave number n to the total variance of $c^h(r(\theta))$ and $c^g(r(\theta))$, respectively, and thus define the variance-power spectra. The lower panel in Figure 1 shows a log-log plot of this spectra as a function of n . Here we see that the increased fatness in correlation shape for low values of M is associated with higher variance and a reduced damping rate in the small scales, slightly less variance in the intermediate scales, and increased variance in the large scales.

As with the generalized diffusion equation, a linear combination of powers of scaled Laplacian operators (12) can be introduced in Eq. (28) to yield a larger family of correlation functions, but at extra cost. The spectral coefficients of this larger family are given by

$$c_n^h = \frac{\gamma^h}{4\pi a^2} \sqrt{2n+1} \left(1 + \sum_{p=1}^P \left(\frac{L_p^2}{a^2}\right)^p (n(n+1))^p\right)^{-M}, \tag{32}$$

with γ^h modified accordingly so that $c^h(0) = 1$. The smoothing spline functions introduced by Wahba (1982) correspond to the special case of Eqs (5) and (32) for which $M = 2$.

Increasing the degree P of the polynomial of the Laplacian leads to correlation functions that oscillate about the zero axis. This is illustrated in the upper panel of Figure 2,

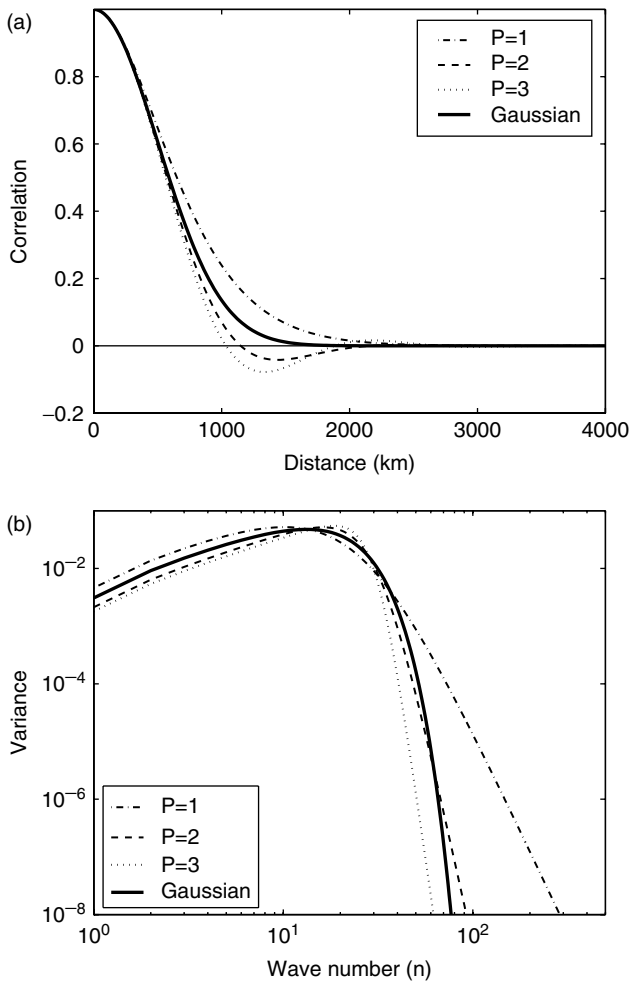


Figure 2. As Figure 1 but for sample correlation functions generated with Eqs (5) and (32) with a fixed value of $M = 4$ and different values of P . The scale parameters have been adjusted to yield a common Daley length-scale of 500 km: $L_1 = 250$ km with $P = 1$ (dashed-dotted curves), $L_1 = 0$ and $L_2 = 206$ km with $P = 2$ (dashed curves), $L_1 = L_2 = 0$, $L_3 = 209$ km with $P = 3$ (dotted curves).

where the generalized $c^h(r(\theta))$ are displayed with different values of P but a fixed value of $M = 4$. The amplitude of the negative lobes increases with increasing value of P . In spectral space, the negative lobes are associated with a decrease in variance in the large scales and an increase in variance in the intermediate scales. Increasing the value P also leads to a steepening of the decay rate of the variance in the smaller scales.

A straightforward variant of Eq. (32) that can be used to enhance the oscillations while maintaining a gradual spectral decay rate at high wave numbers is

$$c_n^h = \frac{\gamma^h}{4\pi a^2} \sqrt{2n+1} \left(1 + \sum_{p=1}^P \rho_p \left(\frac{L_p^2}{a^2} \right)^p (n(n+1))^p \right)^{-M}, \quad (33)$$

where ρ_p is a dimensionless coefficient that can take on both negative and positive values. This is equivalent to redefining the diffusion coefficients (13) as $\kappa_p = \rho_p L^p$. Equation (33) yields positive coefficients by restricting M to be even. Examples are shown in Figure 3 for the case $P = 2$ and $M = 2$, and a single scale parameter $L_1 = L_2 = L$. Here

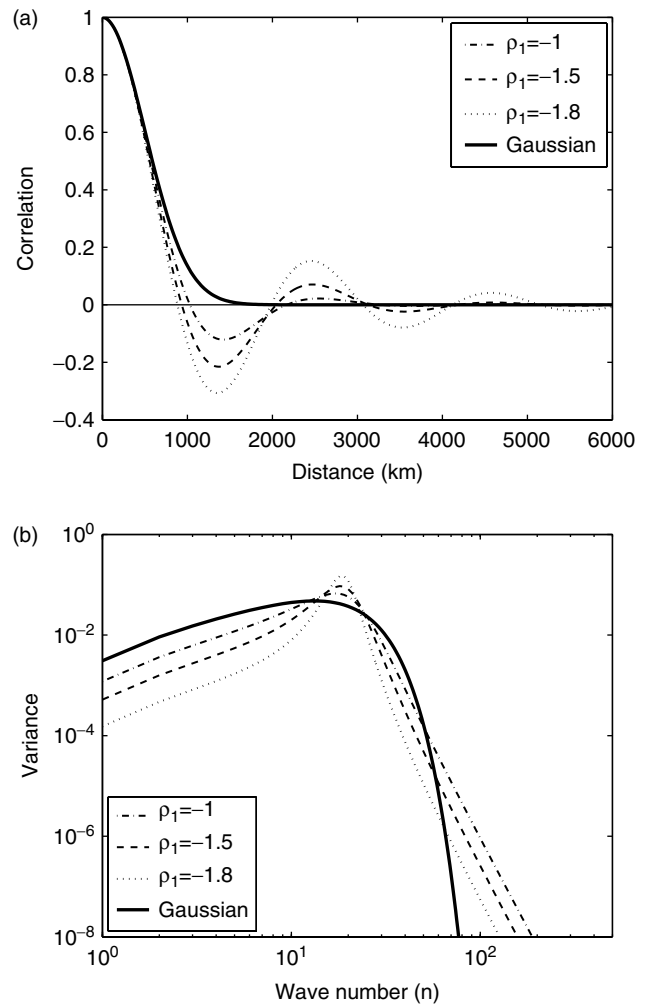


Figure 3. As Figure 1 but for sample correlation functions generated with Eqs (5) and (33) with a fixed value of $M = 2$, $P = 2$ and $\rho_2 = 1$, and different values of ρ_1 . A single scale parameter $L_1 = L_2 = L$ has been used and adjusted to yield a common Daley length-scale of 500 km: $L = 308$ km with $\rho_1 = -1$ (dashed-dotted curves), $L = 326$ km with $\rho_1 = -1.5$ (dashed curves), and $L = 340$ km with $\rho_1 = -1.8$ (dotted curves).

ρ_2 has been set to one and negative values have been used for ρ_1 . Increasing the magnitude of ρ_1 results in a significant increase in the amplitude of the oscillations and a much sharper spectral peak at intermediate scales. Notice that by setting $\rho_1 = 2$ we recover the non-oscillatory correlation function governed by Eq. (30) with $M = 4$, which is displayed in Figure 1 (dashed curves).

On a numerical grid, C^h can be approximated by a discrete operator that solves the linear system (28)–(29) for a given right-hand side $\tilde{\eta}(\lambda, \phi)$. We refer to C^h as an *implicit* diffusion correlation operator. Although the cost of each iteration of an implicit diffusion operator will generally increase relative to that of the explicit scheme, the total cost of the implicit algorithm can easily decrease through the possibility of performing significantly fewer iterations.

2.4. Implicit diffusion on \mathbb{R}^d

The starting point is the following general fractional differential operator $(C_d^w)^{-1} : \psi \mapsto \tilde{\psi}$ (Whittle, 1954, 1963; Guttorp and Gneiting, 2006):

$$(\gamma_d^w)^{-1} (1 - L^2 \nabla^2)^{v+d/2} \psi(\mathbf{x}) = \tilde{\psi}(\mathbf{x}), \quad (34)$$

where $\tilde{\psi}(\mathbf{x}) \in \mathbb{R}^d$ is assumed to be bounded at infinity, $\nu > 0$ is a smoothness parameter, and $\gamma_d^w > 0$ is a normalization constant that depends on the dimension of the space. The FT of Eq. (34) gives the relation

$$\widehat{\psi}(\hat{\mathbf{x}}) = \widehat{c}_d^w(\hat{r}) \widehat{\tilde{\psi}}(\hat{\mathbf{x}}), \tag{35}$$

where $\widehat{\psi}(\hat{\mathbf{x}})$ and $\widehat{\tilde{\psi}}(\hat{\mathbf{x}})$ denote the FTs of $\psi(\mathbf{x})$ and $\tilde{\psi}(\mathbf{x})$, respectively, and

$$\widehat{c}_d^w(\hat{r}) = \frac{\gamma_d^w}{(1 + L^2 \hat{r}^2)^{\nu+d/2}}, \tag{36}$$

$\hat{r} = |\hat{\mathbf{x}}|$ being the norm of the vector of spectral wave numbers associated with \mathbf{x} (see also Yaglom, 1987, p. 363, Eq. (4.130); Stein, 1999, p. 49, Eq. (32); or Gneiting *et al.*, 2009, p. 16, Eq. (20)). Setting

$$\gamma_d^w = 2^d \pi^{d/2} \frac{\Gamma(\nu + d/2)}{\Gamma(\nu)} L^d, \tag{37}$$

where $\Gamma(\nu)$ denotes the Gamma function, and applying the inverse FT to Eq. (35) leads to an integral solution of the general form (16), where $C(\mathbf{x}, \mathbf{x}') = c^w(r)$ is a unit-amplitude isotropic function given by

$$c^w(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{L}\right)^\nu K_\nu\left(\frac{r}{L}\right), \tag{38}$$

$K_\nu(r/L)$ denoting the modified Bessel function of the second kind of order ν , and $r = |\mathbf{x} - \mathbf{x}'|$. Since $\widehat{c}_d^w(\hat{r})$ is strictly positive, $c^w(r)$ is a valid correlation function in \mathbb{R}^d (Bochner's theorem; see Theorem 2.10 in Gaspari and Cohn, 1999). Notice that the power spectrum $\widehat{c}_d^w(\hat{r})$ depends on d but the correlation function itself $c^w(r)$ is independent of d .

Equation (38) is a class of correlation function well known in the geostatistical literature as the Whittle–Matérn or Matérn family (Gneiting, 1999; Stein, 1999; Guttorp and Gneiting, 2006). The link between this correlation family and the fractional differential operator (34) is attributed to Whittle (1954, 1963). Of particular interest here is the subclass of Matérn functions that correspond to (positive) integer values of the parameter $M = \nu + d/2$. For this subclass, the inverse correlation operator $(C_d^w)^{-1}$ has a greatly simplified representation for numerical applications and can be interpreted as an M -step implicitly formulated diffusion operator (MW10), where (cf. Eqs (15) and (16))

$$\begin{aligned} \tilde{\psi}(\mathbf{x}) &\mapsto \eta(\mathbf{x}, 0) = \gamma_d^w \tilde{\eta}(\mathbf{x}), \\ \psi(\mathbf{x}) &\mapsto \eta(\mathbf{x}, M). \end{aligned}$$

The correlation kernels and their associated FT are given by

$$c_d^w(r) = \frac{2^{1-M+d/2}}{\Gamma(M-d/2)} \left(\frac{r}{L}\right)^{M-d/2} K_{M-d/2}\left(\frac{r}{L}\right), \tag{39}$$

and

$$\widehat{c}_d^w(\hat{r}) = \frac{\gamma_d^w}{(1 + L^2 \hat{r}^2)^M}.$$

Equation (39) yields valid correlation functions if $M > d - 1$ ($\nu > 0$; Guttorp and Gneiting, 2006). Notice

also that in contrast to the full Matérn family, the implicit-diffusion kernels depend on d (which has been made explicit by adding the subscript d in $c_d^w(r)$) but their normalized power spectrum $\widehat{c}_d^w(\hat{r})/\widehat{c}_d^w(0)$ is independent of d . For odd values of d , Eq. (39) reduces to a polynomial of order $M - (d + 1)/2$ times an exponential function; this is the well-known class of AR functions.

Of relevance here are the spaces \mathbb{R}^1 , \mathbb{R}^2 and \mathbb{R}^3 . The implicit-diffusion kernels on these spaces can be written explicitly as

$$c_1^w(r) = \sum_{j=0}^{M-1} \beta_{j,M} \left(\frac{r}{L}\right)^j e^{-r/L}, \tag{40}$$

$$c_2^w(r) = \frac{2^{2-M}}{(M-2)!} \left(\frac{r}{L}\right)^{M-1} K_{M-1}\left(\frac{r}{L}\right), \tag{41}$$

$$\text{and } c_3^w(r) = \sum_{j=0}^{M-2} \beta_{j,M-1} \left(\frac{r}{L}\right)^j e^{-r/L}, \tag{42}$$

where

$$\beta_{j,M} = \frac{2^j (M-1)! (2M-j-2)!}{j! (M-j-1)! (2M-2)!}.$$

From Eq. (37), the expressions for the normalization constants become

$$\gamma_1^w = \frac{2^{2M-1} [(M-1)!]^2}{(2M-2)!} L,$$

$$\gamma_2^w = 4\pi (M-1) L^2 \tag{43}$$

$$\text{and } \gamma_3^w = \frac{2^{2M-1} \pi [(M-2)!]^2 (M-1)}{(2M-4)!} L^3.$$

Using Eq. (19), the Daley length-scale of the implicit diffusion kernels in \mathbb{R}^d can be evaluated as

$$D_d^w = \sqrt{2M - d - 2} L. \tag{44}$$

Equation (44) is derived in MW10 for $d = 1$ and in Appendix A for $d = 2$. The generalization to $d > 2$ follows by noting that the correlation functions associated with odd d all have the form (40) with $M \mapsto M - (d - 1)/2$, while those functions with even d all have the form (41) with $M \mapsto M - (d - 2)/2$. Equation (44) imposes further restrictions on the choice of M where now we require

$$M > \begin{cases} (d+1)/2 & \text{if } d \text{ odd,} \\ (d+2)/2 & \text{if } d \text{ even.} \end{cases}$$

In \mathbb{R}^2 , for example, we require $M > 2$. Finally, even values of M are more convenient than odd values of M since they greatly simplify the derivation of a 'square-root' factor of the diffusion operator, which is important for estimating normalization factors and for preconditioning in variational assimilation (WC01).

The explicit diffusion kernels are the limiting case of the implicit diffusion kernels as $M \rightarrow \infty$ with D_d^w fixed. This is easily deduced from Eqs (36), (37) and (44) where, for $M = \nu + d/2$ large, $D_d^w \mapsto D^g$ (Eq. (22)), $\gamma_d^w \mapsto \gamma_d^g$ (Eq. (21)), and $\widehat{c}_d^w(\hat{r}) \mapsto \gamma_d^g \widehat{c}^g(\hat{r})$, where $\widehat{c}^g(\hat{r}) = \exp(-\hat{r}^2 (D^g)^2 / 2)$ is the FT of the d -dimensional Gaussian function (17). Based on the similarity of the

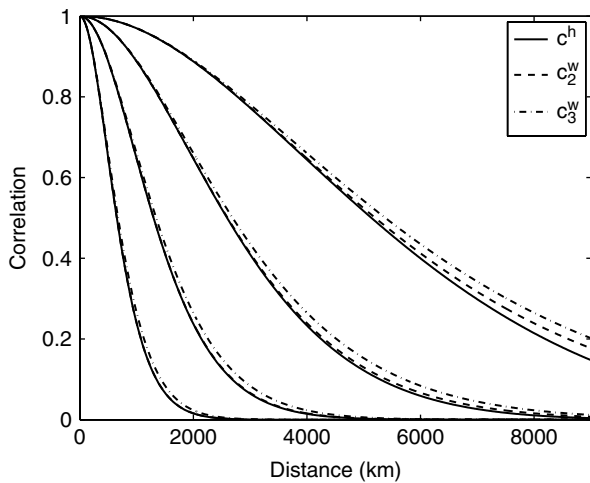


Figure 4. A comparison of the correlation functions $c^h(r(\theta))$ (Eqs (5) and (30)), $c_2^w(r(\theta))$ (Eq. (41)) and $c_3^w(r(\theta))$ (Eq. (42)) for $M = 4$. The four sets of curves correspond to Daley length-scales of 500 km, 1000 km, 2000 km and 4000 km (curves from left to right, respectively). Correlations are plotted as a function of chordal distance. A spectral truncation at $n = 500$ has been used for $c^h(r(\theta))$.

explicit diffusion kernels on \mathbb{S}^2 and \mathbb{R}^2 when $(D^g/a)^2 \ll 1$, one would expect similar agreement between the implicit diffusion kernels on \mathbb{S}^2 and \mathbb{R}^2 when $(D_2^w/a)^2 \ll 1$.

Figure 4 shows the correlation function $c^h(r(\theta))$ for $M = 4$ plotted as a function of chordal distance $r(\theta)$ for four different Daley length-scales D^h (solid curves). For comparison, the correlation functions $c_2^w(r(\theta))$ for $M = 4$ are also shown (dashed curves). The Daley scales D_2^w have been set to D^h and then the corresponding L have been computed from Eq. (44). As expected, the curves are virtually indistinguishable for length-scales of primary interest (< 1000 km). Only when the length-scale exceeds 2000 km do noticeable differences appear, and these mainly occur in the tail of the function. In other words, the differential operator $(\mathcal{C}^h)^{-1}$ on \mathbb{S}^2 can be well approximated locally by the differential operator $(\mathcal{C}_2^w)^{-1}$ acting on the tangent plane \mathbb{R}^2 . The simple relations (43) and (44) can then be used in place of the spectral expansions (31) and (8) to provide a good approximation of the normalization factor and length-scale. This is convenient especially with grid-point ocean models where spectral expansions cannot be readily computed due to the presence of complex land boundaries.

It is important to stress, however, that $c_2^w(r(\theta))$ itself is not a valid correlation function on \mathbb{S}^2 . A valid correlation function on \mathbb{S}^2 from the Matérn family is the AR function $c_3^w(r(\theta))$. For example, Gaspari and Cohn (1999) discuss the second-order AR (SOAR) function on \mathbb{S}^2 (see their Eq. (2.36)). Figure 4 shows the fourth-order AR function for different length-scales (dashed-dotted curves). The differences between $c_3^w(r(\theta))$ and $c^h(r(\theta))$ are larger than those between $c_2^w(r(\theta))$ and $c^h(r(\theta))$ but still quite small for length-scales less than 1000 km.

A more general set of correlation functions on \mathbb{R}^d can be modelled using a linear combination of implicit diffusion operators or a generalized implicit diffusion operator constructed from the inverse of a polynomial of Laplacian operators raised to the power of M (MW10; Yaremchuk and Smith, 2011; or see Purser *et al.*, 2003b, for related approaches involving the recursive filter). The correlation

functions generated by the first approach are described by a linear combination of Matérn functions where the weighting coefficients for each function are specified such that the combined function is positive definite. Gregori *et al.* (2008) provide general conditions on the model parameters for achieving this. MW10 provide an example in \mathbb{R}^1 in which two SOAR functions are combined to produce a correlation function with negative lobes.

The second approach is analogous to the one outlined in section 2.3 for the problem on \mathbb{S}^2 (Eqs (32) and (33)). Hristopulos (2003), Hristopulos and Elogne (2007) and Yaremchuk and Smith (2011) have studied extensively the special case $M = 1$ and $P = 2$ on \mathbb{R}^d for which the parameter settings $\kappa_1 = \rho L^2$ and $\kappa_2 = L^4$ with $\rho < 0$ and satisfying $\rho^2 < 4$ yield a family of positive-definite, oscillatory functions such as those illustrated in Figure 3 on \mathbb{S}^2 . With all of these approaches, however, the advantages of increasing the flexibility in the correlation model have to be carefully measured against the increase in computational cost that results from the need to solve additional or more complicated large linear systems, and the difficulty of having to estimate additional parameters.

3. Anisotropic diffusion

Isotropic correlation models are commonly used in data assimilation algorithms because of their simplicity and computational convenience. There is no reason, however, to expect actual background-error correlations to be isotropic in geophysical fluids such as the ocean. On the contrary, one would expect them to be strongly anisotropic, particularly near coastlines, bathymetry, or ocean fronts. General anisotropic correlation models allow for preferential stretching or shrinking of the correlation functions along arbitrary directions. With a diffusion-based correlation model this can be done using a diffusion tensor, as outlined in this section. To fix the concepts and definitions, we focus mainly on the homogeneous and anisotropic problem. Methods for estimating the parameters of a general inhomogeneous and anisotropic diffusion model are described in section 4.

3.1. Homogeneity and anisotropy

Consider the 2D diffusion equation on \mathbb{R}^2 ,

$$\frac{\partial \eta}{\partial s} - \nabla \cdot \kappa \nabla \eta = 0, \quad (45)$$

where $\kappa \in \mathbb{R}^2 \times \mathbb{R}^2$ is an anisotropic, but constant *diffusion tensor*

$$\kappa = \begin{pmatrix} \kappa_{xx} & \kappa_{xy} \\ \kappa_{yx} & \kappa_{yy} \end{pmatrix} \quad (46)$$

which is assumed to be symmetric ($\kappa_{yx} = \kappa_{xy}$) and positive definite ($\kappa_{xx}\kappa_{yy} > \kappa_{xy}^2$) so that κ is guaranteed to be invertible. The diagonal terms of the tensor determine the strength of the diffusion in the coordinate directions x and y , while the off-diagonal elements allow the principal axes of the diffusion to be rotated relative to x and y .

The solution of Eq. (45) is a straightforward extension of the solution to the isotropic problem (Pannekoucke and Massart, 2008; Pannekoucke, 2009). Given the initial

condition $\eta(x, y, 0) = \gamma_2^g \tilde{\eta}(x, y)$, the solution can be expressed as Eq. (16) (with $d = 2$) where the kernel is given by the Gaussian function

$$c^g(\tilde{r}) = \frac{\gamma_2^g}{4\pi s |\kappa|^{1/2}} e^{-\tilde{r}^2/4s}, \quad (47)$$

$|\kappa|$ denoting the determinant of κ , and \tilde{r} the non-dimensional distance measure

$$\tilde{r} = \sqrt{(\mathbf{x} - \mathbf{x}')^T \kappa^{-1} (\mathbf{x} - \mathbf{x}')}, \quad (48)$$

with $\mathbf{x} = (x, y)^T$. From this definition, κ can also be interpreted as the *aspect tensor* of the Gaussian function (47) (Purser *et al.*, 2003b). The elements of κ have physical units of length squared. Setting

$$\gamma_2^g = 4\pi s |\kappa|^{1/2}$$

ensures that $c^g(0) = 1$.

For a homogeneous and at least twice-differentiable correlation function, we can define the Hessian tensor (Swerling, 1962; Hristopulos, 2002; Chorti and Hristopulos, 2008), which for the 2D Gaussian function is

$$\mathbf{H}^g = -\nabla \nabla^T c^g|_{\tilde{r}=0}, \quad (49)$$

where $\nabla \nabla^T$ is the outer product of the 2D gradient operator $\nabla = (\partial/\partial x \ \partial/\partial y)^T$ and its transpose. The correlation Hessian tensor is of interest here since it is a quantity that can be estimated from sample statistics of background error (see section 4). Following the basic procedure described in Appendix A, it is straightforward to verify that

$$\mathbf{H}^g = \frac{1}{2s} \kappa^{-1}. \quad (50)$$

In the isotropic case, $\kappa = \kappa \mathbf{I}$ and hence $\mathbf{H}^g = (D^g)^{-2} \mathbf{I}$ where $(D^g)^2 = 2\kappa s$ is the square of the Daley length-scale. The inverse of the tensor (49)

$$\mathbf{D}^g := (\mathbf{H}^g)^{-1} \quad (51)$$

can thus be considered as a generalization of the (square of the) Daley length-scale to the anisotropic case. We will thus refer to this quantity as a *Daley tensor*.

For the 3D diffusion equation, the diffusion tensor $\kappa \in \mathbb{R}^3 \times \mathbb{R}^3$ contains six independent elements:

$$\kappa = \begin{pmatrix} \kappa_{xx} & \kappa_{xy} & \kappa_{xz} \\ \kappa_{yx} & \kappa_{yy} & \kappa_{yz} \\ \kappa_{zx} & \kappa_{zy} & \kappa_{zz} \end{pmatrix} \quad (52)$$

where $\kappa_{yx} = \kappa_{xy}$, $\kappa_{zx} = \kappa_{xz}$ and $\kappa_{zy} = \kappa_{yz}$. In direct analogy with the 2D problem, the integral solution involves a 3D Gaussian kernel with aspect tensor given by (52), and normalization constant given by $\gamma_3^g = (4\pi s)^{3/2} |\kappa|^{1/2}$ (cf. Eq. (18)). The relationships (49)–(51) hold for the 3D problem with ∇ now interpreted as the 3D gradient operator.

To approximate a 2D or 3D anisotropic and homogeneous Gaussian correlation operator numerically, we can solve Eq. (45) with an explicit scheme,

$$\eta(\mathbf{x}, M) = \gamma_d^g (1 + \nabla \cdot \kappa \nabla)^M \tilde{\eta}(\mathbf{x}),$$

where from Eqs (50) and (51)

$$\kappa = \frac{1}{2M} \mathbf{D}^g, \quad (53)$$

and the operator $1 + \nabla \cdot \kappa \nabla$ is understood to be in discrete form. If the non-diagonal tensor elements of κ are zero, which can always be achieved by rotating the model coordinates to be aligned with the principal axes of the ellipse or ellipsoid implied by Eq. (48) (see, for example, Xu, 2005), then the 2D or 3D Gaussian operator can be replaced by a product of 1D Gaussian operators acting independently along each direction x , y and z . Ignoring boundary conditions, each 1D Gaussian operator can in turn be approximated by a 1D diffusion operator discretized using an M -step explicit scheme.

Extending these results to the d -dimensional implicit case, we can define a set of anisotropic and homogeneous Matérn correlation operators, with $\nu = M - d/2$, as solutions to the following linear system (cf. Eq. (34)):

$$(\gamma_d^w)^{-1} (1 - \nabla \cdot \kappa \nabla)^M \eta(\mathbf{x}, M) = \tilde{\eta}(\mathbf{x}),$$

where

$$\gamma_d^w = 2^d \pi^{d/2} \frac{\Gamma(M)}{\Gamma(M - d/2)} |\kappa|^{1/2}. \quad (54)$$

The associated correlation functions are given by

$$c_d^w(\tilde{r}) = \frac{2^{1-M+d/2}}{\Gamma(M-d/2)} \tilde{r}^{M-d/2} K_{M-d/2}(\tilde{r}), \quad (55)$$

with \tilde{r} defined by Eq. (48). As for the Gaussian, we can derive the following relationships between the Hessian tensor of $c_d^w(\tilde{r})$ and diffusion tensor κ (see Appendix A):

$$\left. \begin{aligned} \mathbf{H}_d^w &= -\nabla \nabla^T c_d^w|_{\tilde{r}=0}, \\ \mathbf{D}_d^w &:= (\mathbf{H}_d^w)^{-1}, \\ \kappa &= \frac{1}{2M - d - 2} \mathbf{D}_d^w. \end{aligned} \right\} \quad (56)$$

The solution described in section 2.4 corresponds to the isotropic case $\kappa = L^2 \mathbf{I}$ with $L^2 = (D_d^w)^2 / (2M - d - 2)$.

3.2. Inhomogeneity and anisotropy

Analytical expressions for the correlation kernels of the anisotropic diffusion operators in \mathbb{R}^d with spatially varying diffusion tensors $\kappa(\mathbf{x})$ are not known in general. Paciorek and Schervish (2006) describe a family of anisotropic and inhomogeneous correlation functions that generalize the standard isotropic and homogeneous Gaussian and Matérn family. These correlation functions have the form

$$C^g(\mathbf{x}, \mathbf{x}') = \beta(\mathbf{x}, \mathbf{x}') \exp(-\tilde{r}^2/2) \quad (57)$$

for the Gaussian-like function, and

$$C_d^w(\mathbf{x}, \mathbf{x}') = \beta(\mathbf{x}, \mathbf{x}') \times \frac{2^{1-M+d/2}}{\Gamma(M-d/2)} \tilde{r}^{M-d/2} K_{M-d/2}(\tilde{r}) \quad (58)$$

for the Matérn-like functions ($\nu = M - d/2$) where

$$\tilde{r} = \sqrt{(\mathbf{x} - \mathbf{x}')^T \left(\frac{\mathbf{A}(\mathbf{x}) + \mathbf{A}(\mathbf{x}')}{2} \right)^{-1} (\mathbf{x} - \mathbf{x}')}$$

and

$$\beta(\mathbf{x}, \mathbf{x}') = |\mathbf{A}(\mathbf{x})|^{1/4} |\mathbf{A}(\mathbf{x}')|^{1/4} \times \left| \frac{1}{2} (\mathbf{A}(\mathbf{x}) + \mathbf{A}(\mathbf{x}')) \right|^{-1/2},$$

$\mathbf{A}(\mathbf{x})$ and $\mathbf{A}(\mathbf{x}')$ denoting the (symmetric and positive definite) aspect tensors at points \mathbf{x} and \mathbf{x}' , respectively. Equations (57) and (58) with $\mathbf{A}(\mathbf{x}) \approx 2s\kappa(\mathbf{x})$ for the Gaussian-like function and $\mathbf{A}(\mathbf{x}) \approx \kappa(\mathbf{x})$ for the Matérn-like functions can be considered as the approximate kernels of the explicit and implicit forms of the anisotropic diffusion operator when the diffusion tensors $\kappa(\mathbf{x})$ vary slowly and smoothly in space. This is illustrated in MW10 who provide examples in 1D comparing a two-step implicit-diffusion kernel and an inhomogeneous version of the SOAR function for different spatial distributions of the length-scale parameter.

4. Specifying the anisotropic tensor

The elements κ_{xz} , κ_{zx} , κ_{yz} and κ_{zy} of the 3D diffusion tensor account for anisotropy between the horizontal and vertical directions. The importance of these terms compared to the diagonal terms is related to the choice of vertical coordinate in the correlation model. In an ocean model, for example, a natural vertical coordinate is a hybrid coordinate involving a standard geopotential (z) coordinate in unstratified regions such as the mixed layer, an isopycnal (ρ) coordinate in strongly stratified regions, and a terrain-following (s) coordinate near the ocean bottom, the latter being particularly important in shallow coastal regions (Haidvogel and Beckmann, 1999). In this hybrid coordinate system, the flow is more naturally decoupled into ‘horizontal’ and ‘vertical’ processes. If the same coordinate system is adopted for a background-error correlation model then it is reasonable to assume, at least from a physical viewpoint, that the non-diagonal tensor elements κ_{xz} , κ_{zx} , κ_{yz} and κ_{zy} , and possibly κ_{xy} and κ_{yx} , are small and can be neglected. However, anisotropy in background-error correlations can also arise from the assimilation of data, especially when the data coverage is irregular. In general, the relative importance of the diagonal and non-diagonal terms of the tensor can only be determined after a thorough diagnostic study involving, for instance, the direct estimation of the elements of the Daley tensor.

Many ocean models used for global- and basin-scale circulation studies employ a z coordinate. WC01 illustrated how a standard isopycnal diffusion tensor used to parametrize mixing of unresolved processes in a z -coordinate ocean model could also be used to transform the coordinates of a background-error correlation model formulated as an explicit 3D diffusion operator. An analogous coordinate transformation was proposed within the framework of Optimal Interpolation by Balmaseda *et al.* (2008). While the isopycnal correlation model has appealing features, the implementation based on the explicit scheme proposed by WC01 is too expensive

for routine applications since a prohibitively high number of iterations is required to maintain numerical stability in regions of strong isopycnal gradients. Moreover, the specification of the Daley length-scales must be performed in a z -coordinate space, which makes estimating them more difficult in a z -coordinate model. In the remainder of this section we explore alternative methods for defining anisotropic and inhomogeneous correlations, which involve estimating the Daley tensor directly in the model coordinate system.

4.1. Ensemble estimation methods

Given an estimate of the Daley tensor, the anisotropic response of the explicit diffusion operator can be calibrated using Eq. (53), which relates the Daley tensor of the Gaussian function to the diffusion tensor. Alternatively, the anisotropic response of the implicit diffusion operator can be calibrated using the third expression in (56), which relates the Daley tensors of the $\nu = M - d/2$ Matérn functions to the diffusion tensor. Several authors have proposed methods for estimating the Daley tensor using perturbations from an ensemble of model states (Belo Pereira and Berre, 2006; Pannekoucke and Massart, 2008; Pannekoucke, 2009; Sato *et al.*, 2009). The basic procedure is outlined below. Two of the methods will then be compared in idealized experiments using the diffusion equation. For simplicity, we focus on the 2D case.

Assume that an ensemble of N_e model states is available and that the distribution of these states about their mean is a good approximation of the *true* probability distribution function (pdf) of the model-state (background) error ϵ . In variational assimilation, this pdf is assumed to be Gaussian and thus fully described by its mean ($E[\epsilon(\mathbf{x})] = 0$) and covariance function

$$B(\mathbf{x}, \mathbf{x}') = E[\epsilon(\mathbf{x})\epsilon(\mathbf{x}')], \quad (59)$$

where $E[\cdot]$ denotes the expectation operator. The associated correlation function $C(\mathbf{x}, \mathbf{x}')$ can be determined from the factorization

$$C(\mathbf{x}, \mathbf{x}') = \frac{B(\mathbf{x}, \mathbf{x}')}{\sigma(\mathbf{x})\sigma(\mathbf{x}')}, \quad (60)$$

where

$$\sigma(\mathbf{x}) = \sqrt{E[\epsilon(\mathbf{x})^2]}$$

is the standard deviation of ϵ at \mathbf{x} . Assuming that $C(\mathbf{x}, \mathbf{x}')$ is at least twice differentiable then we can define the symmetric tensor

$$\mathbf{T}(\mathbf{x}, \mathbf{x}') = -\nabla \nabla'^T C(\mathbf{x}, \mathbf{x}'),$$

where $\nabla = (\partial/\partial x \ \partial/\partial y)^T$ and $\nabla' = (\partial/\partial x' \ \partial/\partial y')^T$. The local correlation Hessian tensor is the value of \mathbf{T} at $\mathbf{x} = \mathbf{x}'$. Assume further that, in a neighbourhood of \mathbf{x} , $C(\mathbf{x}, \mathbf{x}')$ can be well approximated by a homogeneous function $\tilde{C}(\mathbf{r})$ where $\mathbf{r} = \mathbf{x} - \mathbf{x}' = (x - x', y - y')^T = (\tilde{x}, \tilde{y})^T$. Letting $\tilde{\nabla} = (\partial/\partial \tilde{x} \ \partial/\partial \tilde{y})^T$, then we can define

$$\tilde{\mathbf{T}}(\mathbf{r}) = -\tilde{\nabla} \tilde{\nabla}^T \tilde{C}(\mathbf{r}), \quad (61)$$

such that $\tilde{\mathbf{T}}(\mathbf{0}) \approx \mathbf{T}(\mathbf{x}, \mathbf{x})$ (see Appendix B).

Let $\mathbf{H}(\mathbf{x}) = \mathbf{T}(\mathbf{x}, \mathbf{x})$ denote the local correlation Hessian tensor at \mathbf{x} . Pannekoucke and Massart (2008) and Pannekoucke (2009) assume a Gaussian form for the correlation function and then invert this function at each point \mathbf{x} to estimate $\mathbf{H}(\mathbf{x})$ in terms of sample correlation estimates with neighbouring points. Belo Pereira and Berre (2006) propose an alternative method for estimating $\mathbf{H}(\mathbf{x})$, which does not require a prior assumption on the functional form of the correlations. Their method leads to the expression

$$\widehat{\mathbf{H}}(\mathbf{x}) = \frac{\overline{\nabla\epsilon(\mathbf{x}) (\nabla\epsilon(\mathbf{x}))^T} - \nabla\widehat{\sigma}(\mathbf{x}) (\nabla\widehat{\sigma}(\mathbf{x}))^T}{(\widehat{\sigma}(\mathbf{x}))^2}, \quad (62)$$

where

$$\overline{\nabla\epsilon(\mathbf{x}) (\nabla\epsilon(\mathbf{x}))^T} = \frac{1}{N_e - 1} \sum_{l=1}^{N_e} \nabla\epsilon_l(\mathbf{x}) (\nabla\epsilon_l(\mathbf{x}))^T$$

and $(\widehat{\sigma}(\mathbf{x}))^2 = \overline{(\epsilon(\mathbf{x}))^2} = \frac{1}{N_e - 1} \sum_{l=1}^{N_e} (\epsilon_l(\mathbf{x}))^2$.

The gradient terms can be estimated numerically using finite differences. When the correlation function is strictly homogeneous, $\widehat{\mathbf{H}}$ is equivalent to the constant tensor $\widetilde{\mathbf{T}}(\mathbf{0})$ (Eq. (61)) if the sampling operator is replaced by the expectation operator (see Appendix B). Note that the derivation of Eq. (62) is based on the rather general assumptions of differentiability and local homogeneity of the correlation function. A specific assumption about the actual form of the correlation function is implied only when $\widehat{\mathbf{H}}$ is employed with a particular diffusion operator (explicit or M -step implicit scheme). Hristopulos (2002) and Chorti and Hristopulos (2008) describe a related approach in geostatistics which involves estimating the aspect ratios and orientation angle required to transform the local covariance Hessian tensor into isotropic form.

Multidimensional Gaussian correlation operators can be applied efficiently using a combination of 1D recursive filters or 1D diffusion operators (Purser *et al.*, 2003a, 2003b; MW10). For anisotropic Gaussian operators, the so-called triad (hexad) algorithm (Purser *et al.*, 2003b; Purser, 2005) allows one to determine from the aspect tensor of the 2D (3D) Gaussian function, the three (six) generalized grid-lines along which the 1D filters should be applied. Within that framework, various flow-dependent formulations of the aspect tensor have been proposed (De Pondeva *et al.*, 2006; Liu *et al.*, 2007, 2009; Sato *et al.*, 2009). Of particular interest here is the hybrid formulation of Sato *et al.* (2009) where the inverse of the aspect tensor of the Gaussian function is defined as a linear combination of a ‘conventional term’ based on a quasi-isotropic, static formulation ($\mathbf{A}_{\text{iso}}^{-1}$) and an ‘ensemble term’ formed from the sample covariance of the gradient of the ensemble-generated perturbations, normalized by the sample variance of the perturbations:

$$\widehat{\mathbf{A}}^{-1}(\mathbf{x}) = \alpha \mathbf{A}_{\text{iso}}^{-1}(\mathbf{x}) + \beta \widetilde{\mathbf{H}}(\mathbf{x}) \quad (63)$$

where

$$\widetilde{\mathbf{H}}(\mathbf{x}) = \frac{\overline{\nabla\epsilon(\mathbf{x}) (\nabla\epsilon(\mathbf{x}))^T}}{(\widehat{\sigma}(\mathbf{x}))^2}, \quad (64)$$

and α and β are weighting coefficients. Sato *et al.* (2009) provide a heuristic derivation of Eqs (63)–(64). They are equivalent to Eq. (62) when $\alpha = 0$ and $\beta = 1$, and when the standard deviations are constant. As with the hybrid covariance formulations that involve combining static and ensemble-based expressions of the *full* covariance matrix (Wang *et al.*, 2008), the static term in the aspect tensor is intended to give the estimate more robustness especially when the ensemble size is small, although accounting for it requires extra parameters that must be tuned empirically.

Finally, with small ensemble sizes it can be advantageous to apply a local spatial averaging or filtering operator to the estimated variances and covariances in order to reduce the effects of sampling error (see, for example, Berre and Desroziers (2010) for a thorough review of recent work in this area). Letting F denote a particular filtering operator then the expressions for the filtered estimate of the inverse tensor are given by Eqs (62) and (63)–(64) with

$$\overline{\nabla\epsilon(\mathbf{x}) (\nabla\epsilon(\mathbf{x}))^T} \mapsto F\left(\overline{\nabla\epsilon(\mathbf{x}) (\nabla\epsilon(\mathbf{x}))^T}\right),$$

$$(\widehat{\sigma}(\mathbf{x}))^2 \mapsto F((\widehat{\sigma}(\mathbf{x}))^2),$$

and $\widehat{\sigma}(\mathbf{x}) \mapsto \sqrt{F((\widehat{\sigma}(\mathbf{x}))^2)}$.

4.2. Numerical experiments

In this section we perform idealized experiments to evaluate and compare the effectiveness of Eq. (62) and Eqs (63)–(64) for estimating the parameters of an anisotropic tensor. For simplicity, we focus on the 2D anisotropic diffusion problem and the solution algorithm based on the explicit scheme. Furthermore, for the tensor estimated using Eqs (63)–(64), only the special case $\alpha = 0$ and $\beta = 1$ is considered.

The experimental design is as follows. First, we define the ‘true’ covariance matrix of the problem as

$$\mathbf{B} = \boldsymbol{\Sigma} \boldsymbol{\Gamma}^{1/2} \mathbf{L} \boldsymbol{\Gamma}^{1/2} \boldsymbol{\Sigma},$$

where \mathbf{L} is the M -step explicit diffusion operator $(1 + \nabla \cdot \kappa \nabla)^M$ discretized using a standard centred finite-difference scheme on a uniform grid, $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{1/2} \boldsymbol{\Gamma}^{1/2}$ is a diagonal matrix of normalization factors, and $\boldsymbol{\Sigma}$ is a diagonal matrix of standard deviations σ . With constant parameters and ignoring the influence of boundaries, \mathbf{B} defines a Gaussian covariance matrix.

Next, a sample of N_e spatially uncorrelated random vectors $\widehat{\boldsymbol{\epsilon}}_l, l = 1, \dots, N_e$, are produced on the grid, where the distribution of each $\widehat{\boldsymbol{\epsilon}}_l$ is taken to be Gaussian with $E[\widehat{\boldsymbol{\epsilon}}_l] = \mathbf{0}$ and $E[\widehat{\boldsymbol{\epsilon}}_l \widehat{\boldsymbol{\epsilon}}_l^T] = \mathbf{I}$. Each vector $\widehat{\boldsymbol{\epsilon}}_l$ is then transformed into a new vector $\boldsymbol{\epsilon}_l$ such that $E[\boldsymbol{\epsilon}_l \boldsymbol{\epsilon}_l^T] = \mathbf{B}$. This is done using the ‘square-root’ of the \mathbf{B} -operator,

$$\boldsymbol{\epsilon}_l = \boldsymbol{\Sigma} \boldsymbol{\Gamma}^{1/2} \mathbf{L}^{1/2} \widehat{\boldsymbol{\epsilon}}_l,$$

where $\mathbf{L} = \mathbf{L}^{1/2} \mathbf{L}^{1/2}$, the exponent 1/2 implying $M/2$ iterations of the explicit diffusion operator (with M taken to be even). The sample covariance matrix constructed from the ensemble of $\boldsymbol{\epsilon}_l$ vectors provides an estimate of the true covariance matrix:

$$\mathbf{B} \approx \frac{1}{N_e - 1} \sum_{l=1}^{N_e} \boldsymbol{\epsilon}_l' \boldsymbol{\epsilon}_l'^T, \quad (65)$$

where $\boldsymbol{\epsilon}_l' = \boldsymbol{\epsilon}_l - \frac{1}{N_e} \sum_{k=1}^{N_e} \boldsymbol{\epsilon}_k$.

Our interest here is not to try to reconstruct the full covariance matrix from (65) but rather, with the help of Eq. (62) or Eq. (64), to try to reconstruct the anisotropic tensor used in \mathbf{L} to generate the ϵ_l . Indeed, the sampling errors resulting from estimating the local anisotropic tensor can be expected to be much smaller than those resulting from estimating the full covariance matrix. For the 2D problem, the tensor estimation requires, at each grid point, sample estimates of the standard deviation of ϵ and of the three independent tensor elements involving the gradient of ϵ , i.e. a total of $4N$ elements where N is the total number of grid points. This is much smaller than the $(N^2 + N)/2$ independent elements required to determine the full covariance matrix.

The numerical experiments are performed in a square domain on a 2D grid $\mathbf{x}_{i,j} = (x_i, y_j)$ where $x_i = i\Delta x$, $i = 1, \dots, \sqrt{N}$, and $y_j = j\Delta y$, $j = 1, \dots, \sqrt{N}$. Here, $\Delta x = \Delta y = 1$ unit and $N = 200 \times 60$, and thus the effective size of the \mathbf{B} matrix is $(1.2 \times 10^4) \times (1.2 \times 10^4)$. Neumann boundary conditions are employed at the solid walls located at the domain edges. As a result, the implied correlation function near the boundary is slightly modified from the target Gaussian (MW10).

At each grid point $\mathbf{x}_{i,j}$, the ‘true’ diffusion tensor κ for \mathbf{L} is defined according to (cf. Eq. (53))

$$\kappa_{i,j} = \frac{1}{2M} \mathbf{D}_{i,j}, \quad (66)$$

where $\mathbf{D}_{i,j}$ is the local Daley tensor which is formulated as

$$\mathbf{D}_{i,j} = \mathbf{R}_{i,j} \bar{\mathbf{D}}_{i,j} \mathbf{R}_{i,j}^{-1},$$

$\bar{\mathbf{D}}_{i,j}$ being a diagonal matrix and $\mathbf{R}_{i,j}$ a rotation matrix ($\mathbf{R}_{i,j}^{-1} = \mathbf{R}_{i,j}^T$). The elements $(D_{xx})_{i,j}$, $(D_{yy})_{i,j}$ and $(D_{xy})_{i,j} = (D_{yx})_{i,j}$ of $\mathbf{D}_{i,j}$ are thus determined by the diagonal elements $(\bar{D}_{xx})_{i,j}$ and $(\bar{D}_{yy})_{i,j}$ of $\bar{\mathbf{D}}_{i,j}$ and by the rotation angle $\theta_{i,j}$ of $\mathbf{R}_{i,j}$. For the experiments described here, $\theta_{i,j} = \theta$ is constant, while the parameters $(\bar{D}_{xx})_{i,j}$ and $(\bar{D}_{yy})_{i,j}$ are specified as a simple oscillatory function of the spatial coordinates $\mathbf{x}_{i,j}$:

$$\begin{aligned} (\bar{D}_{xx})_{i,j} &= A_1 f(\mathbf{x}_{i,j}) + B_1, \\ (\bar{D}_{yy})_{i,j} &= -A_1 f(\mathbf{x}_{i,j}) + B_1, \end{aligned}$$

where

$$f(\mathbf{x}_{i,j}) = \cos\left(\frac{2\pi x_i}{X}\right) \cos\left(\frac{2\pi y_j}{Y}\right),$$

$A_1 = \frac{1}{2}(\bar{D}_{\max} - \bar{D}_{\min})$, $B_1 = \frac{1}{2}(\bar{D}_{\max} + \bar{D}_{\min})$, and $X = Y = 20$. Similarly, the variances are specified as

$$\sigma_{i,j}^2 = A_2 f(\mathbf{x}_{i,j}) + B_2,$$

where $A_2 = \frac{1}{2}(\sigma_{\max}^2 - \sigma_{\min}^2)$ and $B_2 = \frac{1}{2}(\sigma_{\max}^2 + \sigma_{\min}^2)$. Experiments are performed with different values of the parameters θ , \bar{D}_{\min} , \bar{D}_{\max} , σ_{\min}^2 and σ_{\max}^2 (see Table 2).

The normalization factors $\gamma_{i,j}$ of the diagonal matrix $\mathbf{\Gamma}$ are approximately given by the expression

$$\gamma_{i,j} \approx 2\pi |\mathbf{D}_{i,j}|^{1/2}. \quad (67)$$

This approximation was used by Pannekoucke and Massart (2008), for example, and is reasonable if the

diffusion tensor varies in space on a scale much larger than the local correlation scale and with a proper treatment of the boundary conditions (MW10). The factors can be estimated to a higher accuracy using more refined analytical approximations (Purser *et al.*, 2003b; Purser, 2008a, 2008b; MW10; Yaremchuk and Carrier, 2012) or randomization methods (WC01; Yaremchuk and Carrier, 2012). They can also be computed exactly using the δ -function method (WC01; MW10). In this idealized study, we employ the exact normalization method in order to avoid introducing a bias in the ensemble perturbations and thus complicating the interpretation of the results. In practice, however, the exact computation is generally not affordable and hence the representation of covariances using the diffusion equation will also be affected by approximations in the normalization factors. The errors that can result from using the approximate expression (67) are illustrated in the experiments below.

The estimation of the tensor via the statistical relationships (62) and (64) is achieved using centred finite-differences. Estimates of the first derivatives of the error and its standard deviation produce values at the interface of the grid cells, i.e. at the half-integer points $(i + 1/2, j)$ for the x -component and $(i, j + 1/2)$ for the y -component. The sample variance of these quantities is computed directly at these points to evaluate the numerator in the expressions for the diagonal elements of the tensor. The off-diagonal elements involve estimates of the cross-product of the x - and y -components of the derivatives. This requires interpolation of one of the component derivatives to the point where the other component derivative is defined. To estimate the cross-product at $(i + 1/2, j)$, the x -component derivative that is defined there is multiplied with an estimate of the y -component derivative obtained by averaging its values from the four surrounding points $(i, j + 1/2)$, $(i + 1, j + 1/2)$, $(i, j - 1/2)$ and $(i + 1, j - 1/2)$, and vice-versa for estimating the cross-product at $(i, j + 1/2)$. To compute the denominator in the expressions for the tensor elements, the sample variance of the error is interpolated from (i, j) points to $(i + 1/2, j)$ or $(i, j + 1/2)$ points. In order to use the estimated tensor elements in the diffusion equation, the elements are first averaged to the (i, j) points and then the off-diagonal elements averaged to force symmetry. The estimated tensor is then inverted at each point and used with the relation (66) to define the diffusion tensor at each point. Finally, interpolation is used to define the values of the tensor elements at the half-integer points $(i + 1/2, j)$ or $(i, j + 1/2)$ where they are required with the centred-difference formulation of $\nabla \cdot \kappa \nabla$.

Table 2 summarizes the results from several experiments with different parameter settings $\mathcal{P} = (\theta, \bar{D}_{\min}, \bar{D}_{\max}, \sigma_{\min}, \sigma_{\max})$. Three cases are considered. In the first case, the principal axes of the anisotropic correlations are aligned with the grid-lines, and the variance is constant: $\mathcal{P}_1 = (0, 3, 6, 1, 1)$. The second case extends the first case by allowing the variances to vary in space: $\mathcal{P}_2 = (0, 3, 6, 1, 5)$. Finally, the third case extends the second case by rotating the principal axes of the anisotropic correlations relative to the grid-lines: $\mathcal{P}_3 = (\pi/4, 3, 6, 1, 5)$. The quality of the estimation is measured in terms of the domain-averaged bias and root-mean-square error (RMSE) of the estimates of the elements (H_{xx}, H_{yy}, H_{xy}) . (The results for the estimates of H_{yx} are not given since they are almost identical to those of H_{xy}). For reference, the domain-averaged RMS

Table 2. Bias and RMSE of the estimates of the correlation Hessian tensor elements (H_{xx}, H_{yy}, H_{xy}) using expressions (62) (\hat{H}) and (64) (\tilde{H}). The second column lists the parameter settings in the ‘true’ covariance model and the third column indicates the choice of ensemble size (N_e) and spatial filtering scale (N_{avg}) used in the estimation process. The RMS of the true values of (H_{xx}, H_{yy}, H_{xy}) are $(5.3, 5.3, 0) \times 10^{-2}$ when $\theta = 0$, and $(5.0, 5.0, 1.7) \times 10^{-2}$ when $\theta = \pi/4$.

Exp	$(\theta, D_{min}, D_{max}, \sigma_{min}, \sigma_{max})$	(N_e, N_{avg})	Method	(H_{xx}, H_{yy}, H_{xy})	
				Bias $\times 10^{-2}$	RMSE $\times 10^{-2}$
1	(0, 3, 6, 1, 1)	(100, 0)	\hat{H}	(0.06, 0.20, 0.0)	(1.2, 1.6, 0.46)
2	(0, 3, 6, 1, 1)	(100, 0)	\tilde{H}	(-0.01, 0.15, 0.0)	(1.2, 1.6, 0.47)
3	(0, 3, 6, 1, 1)	(10, 0)	\hat{H}	(0.33, 0.39, 0.07)	(4.6, 4.8, 2.1)
4	(0, 3, 6, 1, 1)	(10, 0)	\tilde{H}	(0.92, 0.99, 0.10)	(4.9, 5.2, 2.2)
5	(0, 3, 6, 1, 5)	(100, 0)	\hat{H}	(-0.07, -0.22, 0.0)	(1.2, 1.6, 0.47)
6	(0, 3, 6, 1, 5)	(100, 0)	\tilde{H}	(1.2, 1.1, -0.01)	(2.5, 2.8, 1.2)
7	$(\pi/4, 3, 6, 1, 5)$	(100, 0)	\hat{H}	(-0.22, -0.40, 0.01)	(1.4, 1.7, 0.85)
8	$(\pi/4, 3, 6, 1, 5)$	(100, 0)	\tilde{H}	(1.1, 0.88, 0.0)	(2.6, 2.7, 1.4)
9	$(\pi/4, 3, 6, 1, 5)$	(10, 0)	\hat{H}	(0.17, 0.10, -0.02)	(4.0, 4.5, 2.3)
10	$(\pi/4, 3, 6, 1, 5)$	(10, 1)	\hat{H}	(0.33, 0.28, 1.9)	(3.3, 3.6, 1.9)
11	$(\pi/4, 3, 6, 1, 5)$	(10, 3)	\hat{H}	(0.51, 0.45, -0.36)	(2.0, 2.2, 2.8)
12	$(\pi/4, 3, 6, 1, 5)$	(100, 1)	\hat{H}	(0.02, 0.14, -0.07)	(1.2, 1.4, 0.86)
13	$(\pi/4, 3, 6, 1, 5)$	(100, 3)	\hat{H}	(0.31, 0.19, -0.32)	(0.98, 1.0, 1.1)

of the true values of (H_{xx}, H_{yy}, H_{xy}) are $(5.3, 5.3, 0) \times 10^{-2}$ when $\theta = 0$, and $(5.0, 5.0, 1.7) \times 10^{-2}$ when $\theta = \pi/4$.

With \mathcal{P}_1 , \hat{H} (Eq. (62)) and \tilde{H} (Eq. (64)) produce similar results with a relatively large ensemble ($N_e = 100$) as one might expect since the true variances are constant (Exps 1–2 in Table 2). Interestingly, however, \tilde{H} is noticeably more accurate than \hat{H} with a small ensemble size ($N_e = 10$; Exps 3–4). When the variances are spatially varying (\mathcal{P}_2 and \mathcal{P}_3), the errors for \tilde{H} become significantly larger, whereas those for \hat{H} are similar to the constant variance case (Exps 5–8). This illustrates the importance of the second term in Eq. (62).

Local spatial filtering is beneficial for reducing the RMSE especially when the ensemble size is small ($N_e = 10$; Exps 9–11). With the raw ensemble estimates ($N_{avg} = 0$), the RMSE is comparable to the RMS of the true signal (Exps 3–4, 9). Here, a very simple filtering procedure has been used in which the estimate at $x_{i,j}$ is obtained by averaging estimates at points within N_{avg} grid points of (i, j) where in the examples considered $N_{avg} = 1$ or 3. This increases the size of the averaging sample at each point to $N_{eff} = (2N_{avg} + 1)^2 \times N_e$, except near the boundary where fewer points are used in the averaging process. While increasing the value of N_{avg} reduces the RMSE, it does so at the expense of increasing the bias in the estimates. With a larger ensemble ($N_e = 100$), good results are obtained when a ‘light’ filtering is applied ($N_{avg} = 1$), with both the bias and RMSE being reduced relative to the no filtering case for all but the off-diagonal elements which are very slightly degraded (Exps 7, 12–13). The filter in this example is very simple and the choice of filtering scale is somewhat ad hoc. More sophisticated (objective) filters could be expected to perform better as discussed by Raynaud *et al.* (2009) and Berre and Desroziers (2010), and recently by Raynaud and Pannekoucke (2012) within the context of filters based on diffusion.

The correlations obtained using the ‘true’ tensor with the parameter settings \mathcal{P}_3 are illustrated at selected points in Figure 5(a). Figure 5(b) shows the corresponding correlations estimated directly from the sample covariance matrix (Eq. (65)) with a 100-member ensemble. Sampling errors are large and manifest themselves as spurious non-local correlations. In contrast, the diffusion-based

correlation model is localized by construction. The correlations resulting from estimating the diffusion tensor from the 100-member ensemble are shown in Figure 5(c). The estimated correlations are in good agreement with the true correlations and notably capture prominent anisotropic features such as the rotation of the principal axes relative to the grid lines. The third correlation pattern from the left boundary is computed with respect to a point that is located midway between maximum and minimum values of D_{xx} , D_{yy} and σ^2 where the spatial derivative of these parameters is maximum and thus where one would expect the local homogeneous assumption to be least valid. At this location the estimated errors are largest and up to 20% (Figure 5d). The breakdown of homogeneity also affects the accuracy of the approximate expression (67) for the normalization factors. This can be seen in Figure 5 parts (e) and (f), which show the estimated correlations and associated error when approximate normalization factors from Eq. (67) are used in place of the exact factors that were used to produce Figure 5(c). The amplitude of the error now reaches 50% for the third correlation pattern (the colour bar is truncated at 30%) and is noticeably larger for the other correlation patterns as well. Finally, Figure 5 parts (g) and (h) show the correlations and associated errors obtained using the tensor estimated with $N_e = 10$ combined with local spatial averaging. While the correlations are not as accurate as those with $N_e = 100$, they are still reasonable approximations. The maximum error for the third correlation pattern is approximately 25% and reaches 36% when the approximate normalization factors are used (not shown).

5. Summary and conclusions

Accounting for general background-error correlations effectively and efficiently is a considerable challenge in geophysical data assimilation. In VDA, general background-error correlation models can be defined using differential operators constructed numerically from the explicit or implicit solution of a diffusion equation. Theoretical results underpinning the diffusion approach to correlation modelling were reviewed in this paper. First, the isotropic, constant-coefficient diffusion problem was considered both

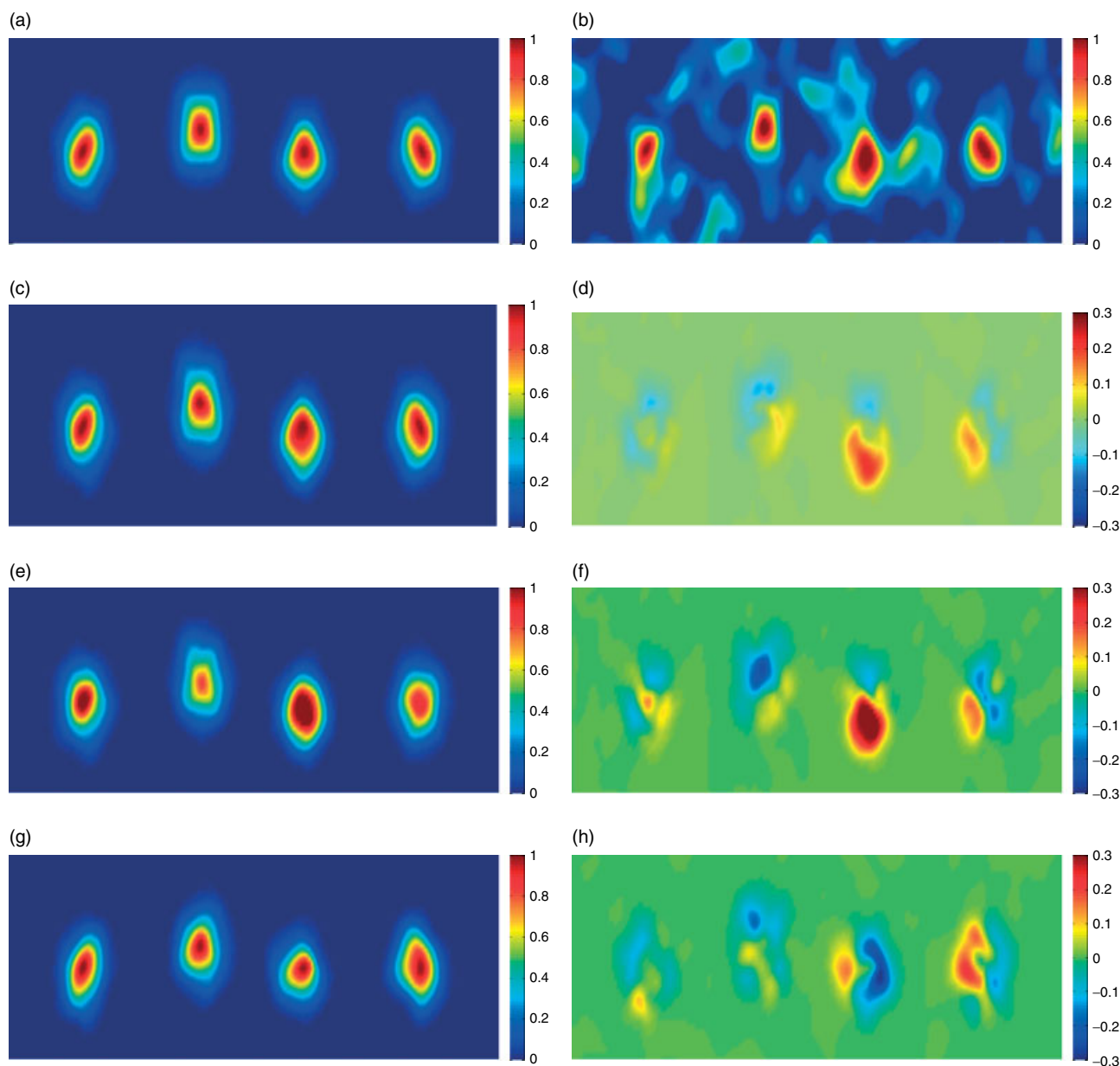


Figure 5. (a) The ‘true’ correlations at selected points in the domain. (b) The correlations estimated directly from the sample covariance matrix (Eq. (65)) with 100 ensemble members. The correlations produced using the diffusion equation with the diffusion tensor estimated with (c) 100 members, no local spatial averaging, and exact normalization; (e) 100 members, no local spatial averaging, and normalization factors approximated using Eq. (67); and (g) 10 members, local spatial averaging with a 7×7 grid-point window, and exact normalization. The differences between the correlations obtained using the estimated diffusion tensor and the true correlations are illustrated in panels (d), (f) and (h).

on the sphere and in the d -dimensional Euclidean space \mathbb{R}^d . The covariance functions (kernels) of the integral solution operators implied by explicit and implicit diffusion in these spaces were identified. The solutions on the sphere were shown to be well approximated by the solutions on \mathbb{R}^2 for scales of interest in ocean and meteorological data assimilation. Expressions relating the diffusion model parameters to the parameters that control the length-scale and amplitude (normalization factor) of the covariance function were also given. These results provided the basis for constructing more general correlation operators via anisotropic diffusion, which was the focus of the second part of the paper.

Anisotropic diffusion was considered in \mathbb{R}^d . The anisotropic diffusion problem is characterized by a diffusion tensor that controls the direction of the covariance response, as well as its scale and amplitude. Solutions to the

anisotropic, constant-tensor diffusion problem are integral operators that involve covariance kernels with the same basic form as those of the isotropic, constant-coefficient problem. With the explicit scheme, these functions are approximately Gaussian, whereas with the implicit scheme they are members of the larger Matérn family (e.g., in \mathbb{R}^3 they are AR functions). For the anisotropic functions, distance is defined by a norm whose metric is given by the inverse of the diffusion tensor. This metric can in turn be related to the correlation Hessian tensor which is defined by the tensor of second-derivatives of the correlation function evaluated at zero separation. The importance of this tensor is that it can be related to quantities that can be estimated directly from ensemble statistics. The inverse of the correlation Hessian tensor was referred to as the Daley tensor in this paper in view of its close connection to the conventional Daley length-scale in the isotropic case.

Ensemble data assimilation methods can be used to provide flow-dependent estimates of the background-error covariances. In realistic applications, the number of independent background-error covariances that need to be estimated is huge and the number of ensemble members that can be affordably run is very limited. Methods are then required to synthesize the ensemble-covariance information to avoid manipulating huge covariance matrices, on the one hand, and to reduce the effects of sampling error, on the other.

The correlation information in the ensemble can be synthesized using a diffusion model with an anisotropic and spatially varying tensor.[†] Procedures for estimating the local Hessian tensor (which in turn can be related to the diffusion tensor) from ensemble perturbations were described and compared in idealized numerical experiments. The method of Belo Pereira and Berre (2006), which assumes local homogeneity of the correlation function but accounts for spatially varying variances, was shown to work particularly well, and is well suited for the automated computations required in a cycled ensemble data assimilation system. Local spatial filtering of the tensor was critical with small ensemble sizes (order 10), but the raw ensemble with 100 members gave good results without spatial filtering in our example. In general, a carefully designed objective filter would be beneficial in order to maximize the signal-to-noise ratio of the ensemble-estimates of the tensor elements in a similar way that it has been shown to be beneficial to the ensemble estimation of background-error variances (Raynaud *et al.*, 2009; Berre and Desroziers, 2010).

In realistic applications, the numerical stability condition associated with explicit diffusion schemes can severely limit their computational efficiency. In particular, many iterations are likely to be needed with general anisotropic and inhomogeneous diffusion models that employ ensemble-estimated tensors. Implicit diffusion schemes are more robust but require solving a large linear system for which efficient methods that are well-suited to massively parallel machines are required. This important practical aspect of the problem was not addressed in this paper and should be the subject of further research.

Acknowledgements

Financial support from the French National Research Agency (ANR) COSINUS programme (VODA project, no. ANR-08-COSI-016), the RTRA STAE foundation (ADTAO project), the European Framework Programme 7 (COMBINE project, GA 226520), and the French LEFE-ASSIM programme is gratefully acknowledged. This work benefited from discussions with Loik Berre, Serge Gratton, Sébastien Massart, Olivier Pannekoucke and Andrea Piacentini. The anonymous reviewers and the Associate Editor Martin Leutbecher provided many helpful remarks for improving the presentation of the paper.

[†]It is common in VDA to use multivariate balance operators to transform the background-error variables into a set of new variables whose cross-covariances can be effectively neglected. The background-error covariances of the transformed variables can then be treated as univariate functions and represented via a diffusion model.

Appendix A

The Daley tensor of the 2D implicit-diffusion kernels

In this appendix we show that the expression for the Daley tensor of the 2D implicit-diffusion kernels $c_2^w(\tilde{r})$ (Eq. (55) with $d = 2$) is related to the diffusion tensor by Eq. (56). For clarity of notation the subscript and superscript of $c_2^w(\tilde{r})$ will be dropped hereafter. The relationships between the Daley and diffusion tensors for the implicit-diffusion kernels in higher dimensions and for the Gaussian function (Eqs (49)–(51)) are straightforward to verify following the basic procedure outlined here.

From the chain rule, the three independent elements of the outer product in the first equation of (56) can be written as

$$\left. \begin{aligned} \frac{\partial^2 c}{\partial x^2} &= \frac{\partial^2 c}{\partial \tilde{r}^2} \left(\frac{\partial \tilde{r}}{\partial x} \right)^2 + \frac{\partial c}{\partial \tilde{r}} \left(\frac{\partial^2 \tilde{r}}{\partial x^2} \right), \\ \frac{\partial^2 c}{\partial y^2} &= \frac{\partial^2 c}{\partial \tilde{r}^2} \left(\frac{\partial \tilde{r}}{\partial y} \right)^2 + \frac{\partial c}{\partial \tilde{r}} \left(\frac{\partial^2 \tilde{r}}{\partial y^2} \right), \\ \frac{\partial^2 c}{\partial x \partial y} &= \frac{\partial^2 c}{\partial \tilde{r}^2} \left(\frac{\partial \tilde{r}}{\partial x} \right) \left(\frac{\partial \tilde{r}}{\partial y} \right) + \frac{\partial c}{\partial \tilde{r}} \left(\frac{\partial^2 \tilde{r}}{\partial x \partial y} \right). \end{aligned} \right\} \quad (68)$$

Expressing Eq. (55) with $d = 2$ as

$$c(\tilde{r}) = \alpha_M \tilde{r}^{M-1} K_{M-1}(\tilde{r}), \quad (69)$$

where $\alpha_M = 2^{2-M}/(M-2)!$ and $M > 2$, and using the following recurrence relation for the modified Bessel functions of the second kind of integer order n (Eq. 9.6.26 of Abramowitz and Stegun, 1970),

$$\frac{\partial K_n}{\partial \tilde{r}} = -\frac{n}{\tilde{r}} K_n - K_{n-1},$$

where $K_n = K_n(\tilde{r})$, allows us to write

$$\left. \begin{aligned} \frac{\partial c}{\partial \tilde{r}} &= -\alpha_M \tilde{r}^{M-1} K_{M-2}, \\ \frac{\partial^2 c}{\partial \tilde{r}^2} &= -\alpha_M (\tilde{r}^{M-2} K_{M-2} - \tilde{r}^{M-1} K_{M-3}). \end{aligned} \right\} \quad (70)$$

The inverse of the symmetric diffusion tensor (46) can be written as

$$\kappa^{-1} = \begin{pmatrix} \kappa_{xx}^{-1} & -\tau \kappa_{xy}^{-1} \\ -\tau \kappa_{xy}^{-1} & \kappa_{yy}^{-1} \end{pmatrix},$$

where $\tau = 1/(\mu - 1)$ and $\mu = \kappa_{xx}\kappa_{yy}/\kappa_{xy}^2$. In expanded form the nondimensional distance measure (48) then reads

$$\begin{aligned} \tilde{r}^2 &= \kappa_{xx}^{-1} (x - x')^2 + \kappa_{yy}^{-1} (y - y')^2 \\ &\quad - 2\tau \kappa_{xy}^{-1} (x - x') (y - y'). \end{aligned} \quad (71)$$

From Eq. (71) we can derive the following relations

$$\left. \begin{aligned} \frac{\partial \tilde{r}}{\partial x} &= \tilde{r}^{-1} X, \\ \frac{\partial \tilde{r}}{\partial y} &= \tilde{r}^{-1} Y, \\ \frac{\partial^2 \tilde{r}}{\partial x^2} &= -\tilde{r}^{-3} X^2 + \tilde{r}^{-1} \kappa_{xx}^{-1}, \\ \frac{\partial^2 \tilde{r}}{\partial y^2} &= -\tilde{r}^{-3} Y^2 + \tilde{r}^{-1} \kappa_{yy}^{-1}, \\ \frac{\partial^2 \tilde{r}}{\partial x \partial y} &= -\tilde{r}^{-3} X Y - \tilde{r}^{-1} \tau \kappa_{xy}^{-1}, \end{aligned} \right\} \quad (72)$$

where

$$X = \kappa_{xx}^{-1} (x - x') - \tau \kappa_{xy}^{-1} (y - y'),$$

$$Y = \kappa_{yy}^{-1} (y - y') - \tau \kappa_{xy}^{-1} (x - x').$$

Substituting Eqs (70) and (72) in Eq. (68) yields

$$\frac{\partial^2 c}{\partial x^2} = \alpha_M (\tilde{r}^{M-3} K_{M-3} X^2 - \tilde{r}^{M-2} K_{M-2} \kappa_{xx}^{-1}),$$

$$\frac{\partial^2 c}{\partial y^2} = \alpha_M (\tilde{r}^{M-3} K_{M-3} Y^2 - \tilde{r}^{M-2} K_{M-2} \kappa_{yy}^{-1}),$$

$$\frac{\partial^2 c}{\partial x \partial y} = \alpha_M (\tilde{r}^{M-3} K_{M-3} X Y + \tilde{r}^{M-2} K_{M-2} \tau \kappa_{xy}^{-1}).$$

Since $c(0) = 1$, for all allowable M , we have from Eq. (69) the general relation

$$\tilde{r}^n K_n(\tilde{r}) \Big|_{\tilde{r}=0} = \frac{1}{\alpha_{n+1}}.$$

Hence

$$\frac{\partial^2 c}{\partial x^2} \Big|_{\tilde{r}=0} = \frac{\alpha_M}{\alpha_{M-2}} X^2 \Big|_{\tilde{r}=0} - \frac{\alpha_M}{\alpha_{M-1}} \kappa_{xx}^{-1} \Big|_{\tilde{r}=0}, \quad (73)$$

$$\frac{\partial^2 c}{\partial y^2} \Big|_{\tilde{r}=0} = \frac{\alpha_M}{\alpha_{M-2}} Y^2 \Big|_{\tilde{r}=0} - \frac{\alpha_M}{\alpha_{M-1}} \kappa_{yy}^{-1} \Big|_{\tilde{r}=0}, \quad (74)$$

$$\frac{\partial^2 c}{\partial x \partial y} \Big|_{\tilde{r}=0} = \frac{\alpha_M}{\alpha_{M-2}} X Y \Big|_{\tilde{r}=0} + \frac{\alpha_M}{\alpha_{M-1}} \tau \kappa_{xy}^{-1} \Big|_{\tilde{r}=0}.$$

The first term on the right-hand side of each of the above equations vanishes since $X = Y = 0$ at $\tilde{r} = 0$, while the common coefficient of the second term in each equation is

$$\frac{\alpha_M}{\alpha_{M-1}} = \frac{1}{2M - 4}.$$

Thus we obtain the relationship governed by (56) with $d = 2$.

In the isotropic case, $\kappa_{xx} = \kappa_{yy} = L^2$ and $\tau \kappa_{xy}^{-1} = 0$. Equations (73) and (74) can then be averaged and inverted to yield the standard definition of the square of the 2D Daley length-scale involving the Laplacian operator (Eq. (44) with $d = 2$).

Appendix B

Estimating the Hessian tensor from an ensemble of simulated errors

In this appendix we provide a derivation of Eq. (62) for the special case of a homogeneous correlation function. The derivation is similar to the one given in the Appendix of Belo Pereira and Berre (2006) except for notational changes and greater emphasis here on some of the underlying assumptions.

The starting point is the general expression (59) for the covariance function $B(\mathbf{x}, \mathbf{x}')$ of the ensemble of model-state errors. We consider here the 2D case where $\mathbf{x} = (x, y)^T$ and $\mathbf{x}' = (x', y')^T$, and assume that $B(\mathbf{x}, \mathbf{x}')$ is at least twice differentiable. We can express the covariance function of

the derivatives of the ensemble errors as follows (Swerling, 1962; Daley, 1991, p. 156):

$$B_{xx'}(\mathbf{x}, \mathbf{x}') = E \left[\left(\frac{\partial \epsilon(\mathbf{x})}{\partial x} \right) \left(\frac{\partial \epsilon(\mathbf{x}')}{\partial x'} \right) \right] = \frac{\partial^2 B(\mathbf{x}, \mathbf{x}')}{\partial x \partial x'},$$

$$B_{yy'}(\mathbf{x}, \mathbf{x}') = E \left[\left(\frac{\partial \epsilon(\mathbf{x})}{\partial y} \right) \left(\frac{\partial \epsilon(\mathbf{x}')}{\partial y'} \right) \right] = \frac{\partial^2 B(\mathbf{x}, \mathbf{x}')}{\partial y \partial y'},$$

$$B_{xy'}(\mathbf{x}, \mathbf{x}') = E \left[\left(\frac{\partial \epsilon(\mathbf{x})}{\partial x} \right) \left(\frac{\partial \epsilon(\mathbf{x}')}{\partial y'} \right) \right] = \frac{\partial^2 B(\mathbf{x}, \mathbf{x}')}{\partial x \partial y'},$$

$$B_{yx'}(\mathbf{x}, \mathbf{x}') = E \left[\left(\frac{\partial \epsilon(\mathbf{x})}{\partial y} \right) \left(\frac{\partial \epsilon(\mathbf{x}')}{\partial x'} \right) \right] = \frac{\partial^2 B(\mathbf{x}, \mathbf{x}')}{\partial y \partial x'}.$$

Using Eq. (60) the derivatives on the right-hand side of the above equations can be evaluated in terms of the standard deviations $\sigma(\mathbf{x})$ and correlation function $C(\mathbf{x}, \mathbf{x}')$. Focusing on the first of these equations this yields

$$B_{xx'}(\mathbf{x}, \mathbf{x}') = \left(\frac{\partial \sigma(\mathbf{x})}{\partial x} \right) \left(\frac{\partial \sigma(\mathbf{x}')}{\partial x'} \right) C(\mathbf{x}, \mathbf{x}')$$

$$+ \sigma(\mathbf{x}) \left(\frac{\partial \sigma(\mathbf{x}')}{\partial x'} \right) \left(\frac{\partial C(\mathbf{x}, \mathbf{x}')}{\partial x} \right)$$

$$+ \sigma(\mathbf{x}') \left(\frac{\partial \sigma(\mathbf{x})}{\partial x} \right) \left(\frac{\partial C(\mathbf{x}, \mathbf{x}')}{\partial x'} \right)$$

$$+ \sigma(\mathbf{x}) \sigma(\mathbf{x}') \frac{\partial^2 C(\mathbf{x}, \mathbf{x}')}{\partial x \partial x'}.$$

Under the assumption of homogeneous correlations, we can write $C(\mathbf{x}, \mathbf{x}') = c(\mathbf{r})$ where $\mathbf{r} = \mathbf{x} - \mathbf{x}' = (x - x', y - y')^T$ (Gaspari and Cohn, 1999). Using the chain rule, the derivatives of C with respect to x, x', y and y' can be rewritten in terms of derivatives of c with respect to $\tilde{x} = x - x'$ and $\tilde{y} = y - y'$. For $B_{xx'}(\mathbf{x}, \mathbf{x}')$ this gives

$$B_{xx'}(\mathbf{x}, \mathbf{x}') = \left(\frac{\partial \sigma(\mathbf{x})}{\partial x} \right) \left(\frac{\partial \sigma(\mathbf{x}')}{\partial x'} \right) c(\mathbf{r})$$

$$+ \sigma(\mathbf{x}) \left(\frac{\partial \sigma(\mathbf{x}')}{\partial x'} \right) \left(\frac{\partial c(\mathbf{r})}{\partial \tilde{x}} \right)$$

$$- \sigma(\mathbf{x}') \left(\frac{\partial \sigma(\mathbf{x})}{\partial x} \right) \left(\frac{\partial c(\mathbf{r})}{\partial \tilde{x}} \right)$$

$$- \sigma(\mathbf{x}) \sigma(\mathbf{x}') \frac{\partial^2 c(\mathbf{r})}{\partial \tilde{x}^2}.$$

Evaluating the above equation at $\mathbf{x} = \mathbf{x}'$ ($\mathbf{r} = \mathbf{0}$), and noting that $\partial c(\mathbf{r}) / \partial \tilde{x} \Big|_{\mathbf{r}=\mathbf{0}} = 0$ since $c(\mathbf{r})$ is maximum at $\mathbf{r} = \mathbf{0}$ and that $c(\mathbf{0}) = 1$, yields

$$B_{xx}(\mathbf{x}, \mathbf{x}) = \left(\frac{\partial \sigma(\mathbf{x})}{\partial x} \right)^2 - (\sigma(\mathbf{x}))^2 \frac{\partial^2 c(\mathbf{r})}{\partial \tilde{x}^2} \Big|_{\mathbf{r}=\mathbf{0}},$$

which can be rearranged as

$$- \frac{\partial^2 c(\mathbf{r})}{\partial \tilde{x}^2} \Big|_{\mathbf{r}=\mathbf{0}} = \frac{1}{\sigma^2} \left(B_{xx}(\mathbf{x}, \mathbf{x}) - \left(\frac{\partial \sigma}{\partial x} \right)^2 \right), \quad (75)$$

where $\sigma = \sigma(\mathbf{x})$. A similar analysis for the other covariance functions yields

$$-\frac{\partial^2 c(\mathbf{r})}{\partial \tilde{y}^2} \Big|_{\mathbf{r}=\mathbf{0}} = \frac{1}{\sigma^2} \left(B_{yy}(\mathbf{x}, \mathbf{x}) - \left(\frac{\partial \sigma}{\partial y} \right)^2 \right), \quad (76)$$

$$-\frac{\partial^2 c(\mathbf{r})}{\partial \tilde{x} \partial \tilde{y}} \Big|_{\mathbf{r}=\mathbf{0}} = \frac{1}{\sigma^2} \left(B_{xy}(\mathbf{x}, \mathbf{x}) - \left(\frac{\partial \sigma}{\partial x} \right) \left(\frac{\partial \sigma}{\partial y} \right) \right), \quad (77)$$

$$-\frac{\partial^2 c(\mathbf{r})}{\partial \tilde{y} \partial \tilde{x}} \Big|_{\mathbf{r}=\mathbf{0}} = \frac{1}{\sigma^2} \left(B_{yx}(\mathbf{x}, \mathbf{x}) - \left(\frac{\partial \sigma}{\partial y} \right) \left(\frac{\partial \sigma}{\partial x} \right) \right). \quad (78)$$

In tensor notation, the left-hand side of Eqs (75)–(78) is equivalent to Eq. (61) evaluated at $\mathbf{r} = \mathbf{0}$, while the right-hand side of the equations can be identified with the right-hand side of Eq. (62).

References

- Balmaseda MA, Vidard A, Anderson DLT. 2008. The ECMWF Ocean Analysis System: ORA-S3. *Mon. Weather Rev.* **136**: 3018–3034.
- Bannister RN. 2008. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **134**: 1971–1996.
- Belo Pereira M, Berre L. 2006. The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Weather Rev.* **134**: 2466–2489.
- Bennett AF, Chua BS, Leslie LM. 1996. Generalized inversion of a global numerical weather prediction model. *Meteorol. Atmos. Phys.* **60**: 165–178.
- Bennett AF, Chua BS, Leslie LM. 1997. Generalized inversion of a global numerical weather prediction model. II: analysis and implementation. *Meteorol. Atmos. Phys.* **62**: 129–140.
- Berre L, Desroziers G. 2010. Filtering of background error variances and correlations by local spatial averaging: a review. *Mon. Weather Rev.* **138**: 3693–3720.
- Carrier MJ, Ngodock H. 2010. Background-error correlation model based on the implicit solution of a diffusion equation. *Ocean Model.* **35**: 45–53.
- Chorti A, Hristopulos DT. 2008. Non-parametric identification of anisotropic (elliptic) correlations in spatially distributed data sets. *IEEE Trans. Signal Process.* **56**: 4738–4751.
- Chua BS, Bennett AF. 2001. An inverse ocean modeling system. *Ocean Model.* **3**: 137–165.
- Daget N, Weaver AT, Balmaseda MA. 2009. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteorol. Soc.* **135**: 1071–1094.
- Daley R. 1991. *Atmospheric Data Analysis*. Cambridge University Press: Cambridge, UK.
- De Ponte MSFV, Purser RJ, Parrish DF, Derber J. 2006. ‘Comparison of strategies for the specification of anisotropies in the covariances of a three-dimensional atmospheric data assimilation system’. Office Note 452, National Centers for Environmental Prediction: Camp Springs, MD.
- Derber JC, Rosati A. 1989. A global oceanic data assimilation system. *J. Phys. Oceanogr.* **19**: 1333–1347.
- Di Lorenzo E, Moore AM, Arango HG, Cornuelle BD, Miller AJ, Powell B, Chua BS, Bennett AF. 2007. Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): Development and application for a baroclinic coastal upwelling system. *Ocean Model.* **16**: 160–187.
- Dobricic S, Pinardi N. 2008. An oceanographic three-dimensional variational data assimilation scheme. *Ocean Model.* **22**: 89–105.
- Egbert G, Bennett A, Foreman M. 1994. Topex/Poseidon tides estimated using a global inverse model. *J. Geophys. Res.* **99**: 24821–24852.
- Elbern H, Schwinger J, Botchorishvili R. 2010. Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: system configuration. *J. Geophys. Res.* **115**: D06302, DOI: 10.1029/2009JD011953.
- Gaspari G, Cohn S. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**: 723–757.
- Geer AJ, Lahoz WA, Bekki S, Bormann N, Errera Q, Eskes HJ, Fonteyn D, Jackson DR, Juckes MN, Massart S, Peuch V-H, Rharmili S, Segers A. 2006. The ASSET intercomparison of ozone analyses: method and first results. *Atmos. Chem. Phys.* **6**: 5445–5474.
- Gneiting T. 1999. Correlation functions for atmospheric data analysis. *Q. J. R. Meteorol. Soc.* **125**: 2449–2464.
- Gneiting T, Kleiber W, Schlather M. 2009. ‘Matérn cross-covariance functions for multivariate random fields’. Technical Report No. 549, University of Washington, Department of Statistics. <http://www.stat.washington.edu/research/reports/2009/tr549.pdf>
- Gregori P, Porcu E, Mateu J, Sasvári Z. 2008. On potentially negative space time covariances obtained as sum of products of marginal ones. *Ann. Inst. Stat. Math.* **60**: 865–882.
- Guttorp P, Gneiting T. 2006. Miscellaneous studies in the history of probability and statistics XLIX: on the Matérn correlation family. *Biometrika* **93**: 989–995.
- Haidvogel DB, Beckmann A. 1999. *Numerical Ocean Circulation Modeling*. Imperial College Press: London.
- Hartman P, Watson GS. 1974. ‘Normal’ distribution functions on spheres and the modified Bessel functions. *Ann. Prob.* **2**: 593–607.
- Hayden CM, Purser RJ. 1995. Recursive filter objective analysis of meteorological fields: applications to NESDIS operational processing. *J. Appl. Meteorol.* **34**: 3–15.
- Hristopulos DT. 2002. New anisotropic covariance models and estimation of anisotropic parameters based on the covariance tensor identity. *Stoch. Environ. Res. Risk Assess.* **16**: 43–62.
- Hristopulos DT. 2003. Spartan Gibbs random field models for geostatistical applications. *SIAM J. Sci. Comput.* **24**: 2125–2162.
- Hristopulos DT, Elogne SN. 2007. Analytic properties and covariance functions of a new class of generalized Gibbs random fields. *IEEE Trans. Inform. Theory.* **53**: 4467–4679.
- Kurapov AL, Egbert GD, Allen JS, Miller RN. 2009. Representer-based analyses in the coastal upwelling system. *Dyn. Atmos. Oceans* **48**: 198–218.
- Liu H, Xue M, Purser RJ, Parrish DF. 2007. Retrieval of moisture from simulated GPS slant-path water vapor observations using 3DVAR with anisotropic recursive filter. *Mon. Weather Rev.* **135**: 1506–1521.
- Liu Y, Zhu J, She J, Zhuang S, Fu W, Gao J. 2009. Assimilating temperature and salinity profile observations using an anisotropic recursive filter in a coastal ocean model. *Ocean Model.* **30**: 75–87.
- Lorenc AC. 1992. Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.* **118**: 569–591.
- Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW. 2000. The Met Office global three-dimensional variational data assimilation system. *Q. J. R. Meteorol. Soc.* **126**: 2991–3012.
- Martin MJ, Hines A, Bell MJ. 2007. Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. *Q. J. R. Meteorol. Soc.* **133**: 981–995.
- Massart S, Piacentini P, Pannekouck O. 2012. Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses. *Q. J. R. Meteorol. Soc.*, DOI: 10.1002/qj.971.
- McIntosh P. 1990. Oceanographic data interpolation: objective analysis and splines. *J. Geophys. Res.* **95**: 13529–13541.
- Mirouze I, Weaver AT. 2010. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Q. J. R. Meteorol. Soc.* **136**: 1421–1443.
- Moore AM, Arango HG, Broquet G, Powell BS, Weaver AT, Zavala-Garay J. 2011. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. I: System overview and formulation. *Prog. Oceanogr.* **91**: 50–73.
- Muccino JC, Arango HG, Bennett AF, Chua BS, Cornuelle BD, Di Lorenzo E, Egbert GD, Haidvogel D, Levin JC, Luo H, Miller AJ, Moore AM, Zaron ED. 2008. The Inverse Ocean Modelling System. Part II: applications. *J. Atmos. Ocean. Tech.* **25**: 1623–1637.
- Ngodock HE. 2005. Efficient implementation of covariance multiplication for data assimilation with the representer method. *Ocean Model.* **8**: 237–251.
- Paciorek CJ, Schervish MJ. 2006. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17**: 483–506.
- Pannekoucke O. 2009. Heterogeneous correlation modeling based on the wavelet diagonal assumption and on the diffusion operator. *Mon. Weather Rev.* **131**: 2995–3012.
- Pannekoucke O, Massart S. 2008. Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation. *Q. J. R. Meteorol. Soc.* **134**: 1425–1438.
- Pannekoucke O, Berre L, Desroziers G. 2008. Background-error correlation length-scale estimates and their sampling statistics. *Q. J. R. Meteorol. Soc.* **134**: 497–508.
- Purser RJ. 2005. ‘A geometrical approach to the synthesis of smooth anisotropic covariance operators for data assimilation’. Office Note 447, National Centers for Environmental Prediction: Camp Springs, MD.

- Purser RJ. 2008a. 'Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry. Part I: Analytic solutions for symmetrical configurations and the validation of practical algorithms'. Office Note 456, National Centers for Environmental Prediction: Camp Springs, MD.
- Purser RJ. 2008b. 'Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry. Part II: Riemannian geometry and the generic parametric expansion method'. Office Note 457, National Centers for Environmental Prediction: Camp Springs, MD.
- Purser RJ, Wu WS, Parrish DF, Roberts NM. 2003a. Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Weather Rev.* **131**: 1524–1535.
- Purser RJ, Wu WS, Parrish DF, Roberts NM. 2003b. Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Weather Rev.* **131**: 1536–1548.
- Raynaud L, Pannekoucke O. 2012. Heterogeneous filtering of ensemble-based background-error variances. *Q. J. R. Meteorol. Soc.*, DOI:10.1002/qj.1890.
- Raynaud L, Berre L, Desroziers G. 2009. Objective filtering of ensemble-based background-error variances. *Q. J. R. Meteorol. Soc.* **134**: 1003–1014.
- Roberts PH, Ursell HD. 1960. Random walk on a sphere and on a Riemannian manifold. *Philos. Trans. R. Soc. London A* **252**: 317–356.
- Sato Y, De Ponte MSFV, Purser RJ, Parrish DF. 2009. 'Ensemble-based background error covariance implementations using spatial recursive filters in NCEP's grid-point statistical interpolation system'. Office Note 459, National Centers for Environmental Prediction: Camp Springs, MD.
- Sheinbaum J, Anderson DLT. 1990. Variational assimilation of XBT data. Part II: Sensitivity studies and use of smoothing constraints. *J. Phys. Oceanogr.* **20**: 689–704.
- Stein ML. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer: New-York.
- Swerling P. 1962. Statistical properties of the contours of random surfaces. *IRE Trans. Inform. Theory* **IT-8**: 315–321.
- Thacker WC. 1988. Fitting models to inadequate data by enforcing spatial and temporal smoothness. *J. Geophys. Res.* **93**: 10655–10665.
- Wahba G. 1982. 'Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data'. In *Multivariate Approximation Theory II*, Schempp W, Zeller K (eds). Birkhäuser: Basel; 407–429.
- Wahba G, Wendelberger J. 1980. Some new mathematical methods for variational objective analysis using splines and cross-validation. *Mon. Weather Rev.* **108**: 36–57.
- Wang X, Barker DM, Synder C, Hamill TM. 2008. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Weather Rev.* **136**: 5116–5131.
- Weaver AT, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **127**: 1815–1846.
- Weaver AT, Ricci S. 2004. 'Constructing a background-error correlation model using generalized diffusion operators'. In *ECMWF Proceedings of the Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*. ECMWF: Reading, UK; 327–339.
- Weaver AT, Vialard J, Anderson DLT. 2003. Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific Ocean. Part 1: Formulation, internal diagnostics and consistency checks. *Mon. Weather Rev.* **131**: 1360–1378.
- Weber RO, Talkner P. 1993. Some remarks on spatial correlation function models. *Mon. Weather Rev.* **121**: 2611–2617.
- Whittle P. 1954. On stationary processes in the plane. *Biometrika* **41**: 434–449.
- Whittle P. 1963. Stochastic processes in several dimensions. *Bull. Inst. Int. Statist.* **40**: 974–994.
- Wu WS, Purser RJ, Parrish DF. 2002. Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Weather Rev.* **130**: 2905–2916.
- Xu Q. 2005. Representations of inverse covariances by differential operators. *Adv. Atmos. Sci.* **22**: 181–198.
- Yaglom AM. 1987. *Correlation Theory of Stationary and Related Random Functions. I: Basic Results*. Springer: New York.
- Yaremchuk M, Carrier M. 2012. On the renormalization of the covariance operators. *Mon. Weather Rev.* **140**: 637–649.
- Yaremchuk M, Smith S. 2011. On the correlation functions associated with polynomials of the diffusion operator. *Q. J. R. Meteorol. Soc.* **137**: 1927–1932.
- Yaremchuk M, Nechaev D, Pan C. 2011. A hybrid background-error covariance model for assimilating glider data into a coastal ocean model. *Mon. Weather Rev.* **139**: 1879–1890.
- Zaron ED, Chavanne C, Egbert GD, Flament P. 2009. Baroclinic tidal generation in the Kauai Channel inferred from high-frequency radio Doppler current meters. *Dyn. Atmos. Oceans* **48**: 93–120.