Constructing a background-error correlation model using generalized diffusion operators

Anthony T. Weaver and Sophie Ricci

CERFACS
42 avenue Gaspard Coriolis
31057 Toulouse Cedex 1
France
weaver@cerfacs.fr, ricci@cerfacs.fr

1 Introduction

Correlation models are at the heart of background-error covariance matrices and are fundamental for determining how observational information is spread spatially (Fisher 2003). The true correlations of background error are not known but can be estimated from statistics of observed-minus-background fields (Hollingsworth and Lönnberg 1986; Lönnberg and Hollingsworth 1986) or a suitable proxy for background error such as model forecast differences (Parrish and Derber 1992; Rabier *et al.* 1998; Fisher 2003). From the available estimates of the statistics of background error, correlation models attempt to parametrise key features of the shape and spectrum of the correlation functions. Correlation models are embedded within correlation operators. Applying a correlation operator to a given field involves solving an integral equation over the model domain where the kernel of the operator is a correlation function (here parametrised by a model). Application of the background-error correlation operator is generally the most expensive step in a three-dimensional (3D) variational analysis. The numerical efficiency of the correlation operator is critical. Therefore, not all correlation models (and hence operators) are suitable for large dimensional problems such as those encountered in atmospheric or ocean data assimilation.

This paper gives an overview of the theory and numerical implementation of the generalized diffusion approach for defining correlation models in variational data assimilation. Fisher (2003) and Derber *et al.* (2003) describe alternative correlation models based on spectral and recursive filtering techniques. Some of the similarities between these different methods will be discussed in this paper. The basic algorithm detailed in this paper consists of numerically integrating a generalized diffusion equation (GDE) in order to provide an efficient way of accomplishing the smoothing action of a correlation operator. The choice of numerical integration scheme is shown to be important in determining the class of correlation functions that can be represented by the GDE. A much larger class of correlation functions can be modelled with a time-implicit scheme than with a time-explicit scheme due to the property of unconditional stability of an implicit scheme. Analytical expressions for the correlation functions are derived for the two-dimensional (2D) isotropic case on the sphere. A formal connection between the time-implicit GDE and smoothing splines is also established. Techniques for adapting the GDE to account for inhomogeneous, anisotropic and non-separable correlation functions are discussed within the context of ocean data assimilation. Finally, some of the important issues involved in the numerical implementation of the GDE are highlighted. Numerical examples are presented from an ocean variational assimilation system.

2 A one-dimensional example

The following one-dimensional (1D) problem provides a simple framework for interpreting the basic procedure for constructing correlation operators using a generalized diffusion equation (GDE). Consider the classical diffusion equation

$$\frac{\partial \eta}{\partial t} - \kappa \frac{\partial^2 \eta}{\partial z^2} = 0 \tag{1}$$

where κ is a constant diffusion coefficient, and $\eta(z,t)$ is an arbitrary scalar field (e.g., temperature) defined on the infinite line such that $\eta(z,t)$ vanishes as $z \to \pm\infty$. The solution of (1) at the end of a (pseudo-)time interval $0 \le t \le T$ is given by the integral equation

$$\eta(z,T) = \frac{1}{\sqrt{4\pi\kappa T}} \int_{z'} e^{-(z-z')^2/4\kappa T}\, \eta(z',0)\, dz' \tag{2}$$

where $\eta(z,0)$ is the initial condition. Equation (2) shows that $\eta(z,T)$ is the result of a convolution of $\eta(z,0)$ with the Gaussian covariance function $(\sqrt{4\pi\kappa T})^{-1} \exp(-z^2/4\kappa T)$ where the product $2\kappa T$ can be interpreted as the square of the length scale of the Gaussian function (Daley 1991) and the constant coefficient $(\sqrt{4\pi\kappa T})^{-1}$ (with physical dimensions of inverse length) as the "variance". A correlation function is a covariance function with unit amplitude. The Gaussian covariance operator can thus be transformed into a Gaussian *correlation* operator by post-multiplying $\eta(z,T)$ by $\sqrt{4\pi\kappa T}$.

The key idea is that, on a discrete grid, we can perform the action of a Gaussian correlation operator by iterating a discretized version of the differential equation (1) from an initial condition $\eta(z,0)$, and normalizing the result as above. This is a computationally efficient way of evaluating the convolution integral in (2), and is the essence of the diffusion (Laplacian-filter) algorithm for constructing 2D and 3D correlation operators, originally proposed by Derber and Rosati (1989), described in more detail by Egbert *et al.* (1994), and later extended by Weaver and Courtier (2001) (hereafter referred to as WC01).

3 An isotropic correlation model: generalized diffusion and smoothing splines

The theoretical basis for employing a GDE to represent the action of a 2D correlation operator on the sphere and a 1D correlation operator in the vertical is described in detail in WC01. The purpose of this section is to outline the theory from a slightly different angle in order to expose a larger class of correlation functions than those derived by WC01. For simplicity, we restrict the initial discussion to the homogeneous and isotropic case. Inhomogeneous and anisotropic extensions will be considered in the next section.

Consider the following 2D differential operator defined on the sphere:

$$\widehat{\eta}(\lambda,\phi) = \left(\sum_{p=0}^{P} \alpha_p \left(-\nabla^2\right)^p \right)^M \eta(\lambda,\phi) \tag{3}$$

where $\eta(\lambda,\phi)$ and $\widehat{\eta}(\lambda,\phi)$ are scalar fields defined on the spherical domain of radius a, λ is longitude ($0 \le \lambda \le 2\pi$), ϕ is latitude ($-\pi/2 \le \phi \le \pi/2$), and ∇^2 is the Laplacian operator in spherical coordinates:

$$\nabla^2 \equiv \frac{1}{a^2 \cos\phi} \left\{ \frac{1}{\cos\phi} \frac{\partial^2}{\partial \lambda^2} + \frac{\partial}{\partial \phi} \left(\cos\phi \frac{\partial}{\partial \phi} \right) \right\}. \tag{4}$$

The weighting coefficients α_p, $p = 0, \ldots, P$, are assumed to be non-negative with $\alpha_0 = 1$, and M is a positive integer which we take to be even for convenience. *Roughening* operators similar to (3) are at the heart

of smoothing splines on the sphere (Wahba 1981; 1982). In spline algorithms, smoothing is achieved by minimizing the sum of an observation term (J_o) and a quadratic norm J_s (analogous to the background term J_b in variational assimilation) formulated as an explicit penalty on small-scale variability in the solutions. Some examples of the use of spline algorithms in meteorological and oceanographic data assimilation are given in Wahba and Wendelberger (1982), McIntosh (1990), Sheinbaum and Anderson (1990) and Brasseur *et al.* (1996). A penalty term using (3) would be of the form

$$\begin{aligned} J_s &= \int_{\lambda=0}^{2\pi}\int_{\phi=-\pi/2}^{\pi/2} \eta(\lambda,\phi)\left(\sum_{p=0}^{P}\alpha_p\left(-\nabla^2\right)^p\right)^M \eta(\lambda,\phi)\; a^2\cos\phi\, d\lambda\, d\phi \\ &= \int_{\lambda=0}^{2\pi}\int_{\phi=-\pi/2}^{\pi/2} \left(\left(\sum_{p=0}^{P}\alpha_p\left(-\nabla^2\right)^p\right)^{M/2}\eta(\lambda,\phi)\right)^2 a^2\cos\phi\, d\lambda\, d\phi \end{aligned} \tag{5}$$

where the *square-root* factorization in the last expression follows from the self-adjointness of the Laplacian operator in the sense of the scalar product $\langle \eta_1,\eta_2\rangle = \int\int \eta_1\eta_2\, a^2\cos\phi\, d\lambda\, d\phi$. The norm spline considered by Wahba (1982) is a special case of (5) with $M=2$.

Before formally establishing that the inverse of (3) is a valid (positive definite) covariance operator on the sphere, we first illustrate how (3) can be cast within the framework of the GDE

$$\frac{\partial\eta}{\partial t} + \sum_{p=1}^{P}\kappa_p\left(-\nabla^2\right)^p\eta = 0 \tag{6}$$

where the diffusion coefficients κ_p, $p=1,\ldots,P$, are assumed to be non-negative. Consider a numerical solution of (6) over a pseudo-time interval $t_0=0$ to $t_M = M\Delta t = T$ from an initial condition $\eta(t_0)$, where Δt is the time step and M the total number of time levels. WC01 considered a semi-discrete version of (6) in which the time discretisation is evaluated using an *explicit* forward scheme. Other discretisation schemes are of course possible and as we shall see the choice of scheme is important for determining the class of correlation functions that can be represented by the GDE. Here we consider a classical *implicit* scheme for which the Laplacian terms are evaluated at time step $t_m = m\Delta t$ rather than $t_{m-1} = (m-1)\Delta t$ as in an explicit scheme, where $m=1,\ldots,M$. Over one time step, the implicit form of (6) is

$$\frac{\eta(t_m)-\eta(t_{m-1})}{\Delta t} + \sum_{p=1}^{P}\kappa_p\left(-\nabla^2\right)^p\eta(t_m) = 0 \tag{7}$$

which can be rearranged to give

$$\eta(t_m) = \left(1+\sum_{p=1}^{P}\kappa_p\Delta t\;\left(-\nabla^2\right)^p\right)^{-1}\eta(t_{m-1}) \tag{8}$$

where the term within large brackets is understood to be a discrete matrix operator. Successive applications of (8) from $m=1$ to $m=M$ yields

$$\eta(\lambda,\phi,T) = \left(1+\sum_{p=1}^{P}\kappa_p\Delta t\;\left(-\nabla^2\right)^p\right)^{-M}\eta(\lambda,\phi,0) \tag{9}$$

where $\eta(\lambda,\phi,0)=\eta(t_0)$ and $\eta(\lambda,\phi,T)=\eta(t_M)$. The inverse of (9) is

$$\eta(\lambda,\phi,0) = \left(1+\sum_{p=1}^{P}\kappa_p\Delta t\;\left(-\nabla^2\right)^p\right)^{M}\eta(\lambda,\phi,T). \tag{10}$$

We can match (10) to (3) by identifying $\widehat{\eta}(\lambda,\phi)$ with the initial condition $\eta(\lambda,\phi,0)$, $\eta(\lambda,\phi)$ with the final condition $\eta(\lambda,\phi,T)$, and α_p with the product $\kappa_p \Delta t$ where $\alpha_0 = \kappa_0 \Delta t \equiv 1$. Application of the inverse of the differential operator (3) is thus equivalent to performing M iterations of a GDE with a time-implicit discretisation scheme.

We now set out to establish that the inverse of (3) is a valid covariance operator on the sphere and to derive the specific form of the isotropic covariance functions that this operator can represent. On the sphere, $\eta(\lambda,\phi)$ can be expanded as

$$\eta(\lambda,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \eta_n^m \, Y_n^m(\lambda,\phi) \tag{11}$$

where m is the zonal wavenumber, n is the total wavenumber, $Y_n^m(\lambda,\phi)$ are the spherical harmonics (normalized as defined in WC01), and η_n^m are spectral expansion coefficients. A similar expansion exists for $\widehat{\eta}(\lambda,\phi)$. Since $Y_n^m(\lambda,\phi)$ are the eigenvectors of the Laplacian operator on the sphere, with $-n(n+1)/a^2$ the associated eigenvalues, the spectral coefficients of $\eta(\lambda,\phi)$ and $\widehat{\eta}(\lambda,\phi)$ are related by

$$\widehat{\eta}_n^m = \left(\sum_{p=0}^{P} \alpha_p \left(\frac{n(n+1)}{a^2} \right)^p \right)^M \eta_n^m = \left(1 + \sum_{p=1}^{P} \alpha_p \left(\frac{n(n+1)}{a^2} \right)^p \right)^M \eta_n^m. \tag{12}$$

From the orthogonality of $Y_n^m(\lambda,\phi)$ and the Addition Theorem, it is then straightforward to show (e.g., following the standard procedure laid out in WC01) that the inverse of (3) has an integral representation of the form

$$\eta(\lambda,\phi) = \frac{1}{4\pi a^2} \int_{\lambda'=0}^{2\pi} \int_{\phi'=-\pi/2}^{\pi/2} h(\theta\,;\, M,\, P,\, \alpha_1,\ldots,\alpha_P)\, \widehat{\eta}(\lambda',\phi') \;\; a^2 \cos\phi' \, d\lambda' \, d\phi' \tag{13}$$

where

$$h(\theta\,;\, M,\, P,\, \alpha_1,\ldots,\alpha_P) = \sum_{n=0}^{\infty} \underbrace{\sqrt{2n+1} \left(1 + \sum_{p=1}^{P} \alpha_p \left(\frac{n(n+1)}{a^2} \right)^p \right)^{-M}}_{h_n} P_n^0(\cos\theta), \tag{14}$$

$P_n^0(\cos\theta)$ being the Legendre polynomials, and θ the angular separation (great circle distance) between the points (λ,ϕ) and (λ',ϕ') on the sphere:

$$\cos\theta = \cos\phi \cos\phi' \cos(\lambda - \lambda') + \sin\phi \sin\phi'. \tag{15}$$

Since the coefficients h_n of the Legendre polynomials in (14) are positive, the kernel $h(\theta)$ of the integral operator (13) is the representation of an isotropic covariance function (e.g., see Gaspari and Cohn (1999) for a thorough discussion on the theory of correlation functions). Equations (13) and (14) thus define a valid covariance operator on the sphere. It is readily transformed into a valid correlation operator by multiplying $\eta(\lambda,\phi)$ by the normalization constant $4\pi a^2/h(0)$.

It is interesting to note that by writing $\alpha_p = \kappa_p \Delta t = \frac{1}{M}(\kappa_p T)$ and letting $M \to \infty$ while keeping $\kappa_p T$ fixed, we obtain from (12) an exponential relationship between spectral coefficients

$$\widehat{\eta}_n^m = \exp\left\{ \sum_{p=1}^{P} \kappa_p T \left(\frac{n(n+1)}{a^2} \right)^p \right\} \eta_n^m. \tag{16}$$

Equation (16) can be derived directly from the GDE (Eq. 6), with $\widehat{\eta}_n^m$ and η_n^m defined to be the spectral coefficients of $\eta(\lambda,\phi,t)$ at $t = 0$ and $t = T$, respectively. As shown by WC01, (16) leads to the family of correlation

functions

$$f(\theta\,;\,P,\,\kappa_1 T,...,\kappa_P T) = \sum_{n=0}^{\infty} \underbrace{\sqrt{2n+1}\exp\left\{-\sum_{p=1}^{P}\kappa_p T\left(\frac{n(n+1)}{a^2}\right)^p\right\}}_{f_n} P_n^0(\cos\theta). \tag{17}$$

In (17) the free parameters of the correlation operator are the parameters $\kappa_p T$ $(= M\alpha_p)$ and the maximum number of Laplacians P. In (14) there is an additional parameter M that can be used to give further control over the spectrum and shape of the correlation functions. Equations (14) and (17) provide an obvious relationship between the GDE and spectral approaches (Courtier *et al.* 1998; Rabier *et al.* 1998; Fisher 2003) for modelling isotropic correlation functions on the sphere. In the GDE, the spectral coefficients of the isotropic correlation functions have a specific, yet flexible, functional form. In practice, the GDE is integrated in grid-point space so the evaluation of spectral expansions of the form (14) and (17) is never actually needed. In the spectral approach, however, the evaluation of correlation integrals is done directly in spectral space with truncated spectral expansions.

Figure 1 shows examples of the grid-point representation (left panel) and variance-power spectrum (right panel) of correlation functions computed using (14) and the analytical solution of the GDE (17). The spectral expansions have been evaluated numerically with a truncation at total wavenumber 212. The solid curves in Fig. 1 correspond to the approximately Gaussian correlation function derived from the analytical solution of the classical diffusion equation (Eq. 17 with $P = 1$), while the dotted and dashed curves correspond to analytical solutions of the GDE for $P = 2$ and $P = 3$. The dashed-dotted curves correspond to solutions of (14) with $P = 1, M = 10$ (dashed one-dotted curves), and $P = 2, M = 2$ (dashed three-dotted curves). For these fixed values of P (and M), the correlation parameters $\kappa_p T$ (α_p) have been tuned so that the length scale, L, of the correlation functions is 500 km in all examples, where $L^2 = -f\left(\nabla^2 f\right)^{-1}|_{\theta=0}$ (see WC01). Generally speaking, the effect of increasing P (the number of Laplacians) is to increase the amplitude of the negative lobes in the correlation function and to sharpen its spectral decay at high wavenumbers, while the effect of decreasing M (the number of implicit time-steps) is to increase the "fatness" of the tail of the correlation function and to slow its spectral decay at high wavenumbers. Figure 1 illustrates that the GDE provides a sufficiently flexible framework to represent correlation functions with very different characteristics. This attractive property of the GDE can enable it to capture key features (e.g., spectral decay rates) of observed estimates of the auto-correlations of different geophysical fields (Julien and Thiébaux 1975; Hollingsworth and Lönnberg 1986; Stammer 1997; Wilke *et al.* 1999).

From a numerical viewpoint, the extra degree of freedom M arises from the property of *unconditional stability* of the time-implicit scheme. In contrast, a time-explicit scheme (as considered in WC01) is conditionally stable; M is determined as a function of the free parameters $\kappa_p T = M\,\alpha_p$ and in practice is chosen small enough to leave the scheme stable. The stability requirement can be particularly penalising on the computational efficiency of the method in some cases, for instance, when the length scale is large compared to the grid size. So, not only does the implicit approach allow us to extend the class of correlation functions as outlined above, it also allows, in general, for a significant reduction in the number of iterations of the generalized diffusion operator. The computational cost of the method is mainly determined by the efficiency of the implicit solver. This point will be discussed further in section (5).

So far we have described a general 2D statistical model for representing horizontal correlation functions on the sphere. By analogy, the inverse of a 1D operator of the form

$$\widehat{\eta}(z) = \left(\sum_{q=0}^{Q}\beta_q\left(-\frac{\partial^2}{\partial z^2}\right)^q\right)^N \eta(z) \tag{18}$$

can be used, with appropriate boundary conditions, to define a general statistical model for representing vertical correlation functions, where z denotes the vertical coordinate. The parameters q, Q, β_q and N in (18) are

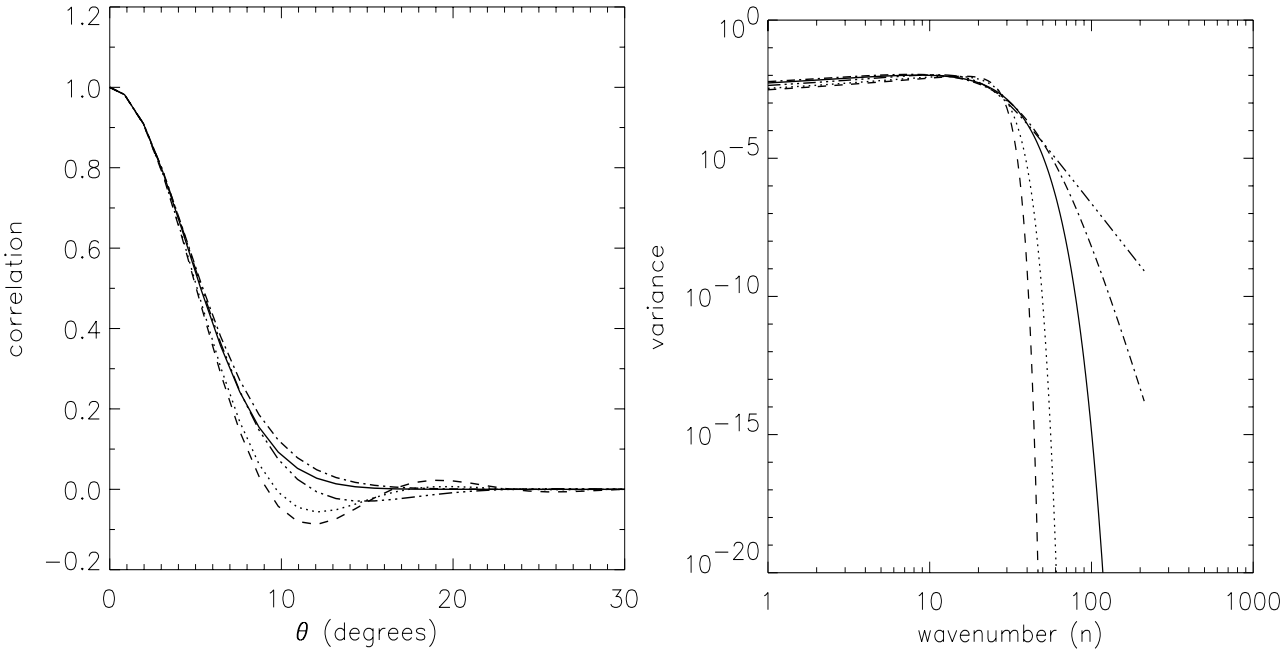

Figure 1: The grid-point values (left panel) and the variance-power spectrum (right panel) of sample correlation functions generated using (14) and (17). The solid, dotted and dashed curves have been generated using (17) with $P = 1$, $P = 2$ and $P = 3$, respectively. The dashed one-dotted and dashed three-dotted curves have been generated using (14) with $P = 1, M = 10$ and $P = 2, M = 2$, respectively. The values of the coefficients α_p in (14) and $\kappa_p T$ in (17) have been tuned to give a common length scale of 500 km.

analogous to p, P, α_p and M in (3). The inverse of a 3D correlation operator can then be derived by combining the operators (3) and (18),

$$\widehat{\eta}(\lambda,\phi,z) = \left(\sum_{r=0}^{R} \beta_r \left(-\frac{\partial^2}{\partial z^2}\right)^r\right)^N \left(\sum_{p=0}^{P} \alpha_p \left(-\nabla^2\right)^p\right)^M \eta(\lambda,\phi,z) \tag{19}$$

where the associated 3D correlation function is given by the product of a 1D and 2D correlation function. Computing numerically the action of the inverse of the operator (19) requires solving first a 1D implicit diffusion equation and then a 2D implicit diffusion equation. To represent correlation functions of the form (17), the GDE can be solved using either an implicit scheme with large M (and $\kappa_p T$ fixed) or, as in WC01, using an explicit scheme.

Another useful feature of the implicit GDE is that it gives immediate access to an inverse correlation operator (Eq. 19) with explicit control of the shape and spectrum of the associated correlation function (Eq.14). An inverse correlation operator for the background error is usually not needed in practice when the standard preconditioning transformation involving the square-root of the background-error covariance matrix is employed (Courtier 1997; Derber and Bouttier 1999; Fisher 2003). Nevertheless, the inverse operator may be needed if alternative preconditioners are employed or if the initial estimate of the control vector is not known *a priori* and must therefore be computed using the inverse of the preconditioning transformation. Note that direct application of the inverse correlation operator (19) is generally much simpler than performing the matrix inversion (9) of the correlation operator itself.

To summarize, in this section we have illustrated how we can transform the problem of solving a general 3D correlation integral equation into an equivalent, but generally computationally much more efficient problem of

solving a product of GDEs using a time-implicit scheme. In the next two sections, we discuss some generalizations to account for inhomogeneous, anisotropic and non-separable correlation functions and outline some of the important numerical aspects of the method.

4 Inhomogeneity, anisotropy and non-separability

In the ocean or atmosphere, correlation structures are generally more complex than those permitted by a homogeneous, isotropic and separable correlation model as presented in the previous section. In this section we outline how the GDE can be adapted to account for inhomogeneous, anisotropic and non-separable correlations.

Consider the 2D differential operator (3). To account for spatial variations in the length scale of the correlation function, it is sufficient to replace $\alpha_p \nabla^2$ by $\nabla \cdot A_p \nabla$, where $A_p = A_p(\lambda, \phi)$ is an inhomogeneous function of horizontal position, and ∇ and $\nabla\cdot$ are the gradient and divergence operators, respectively. With reference to the GDE (6), this is equivalent to using inhomogeneous diffusion coefficients; i.e., replacing $\kappa_p \Delta t \nabla^2$ by $\Delta t \nabla \cdot K_p \nabla$ where $K_p = K_p(\lambda, \phi)$. For example, this feature would allow for small length scales to be used in boundary current regions such as the Gulf Stream while keeping larger length scales in the equatorial regions.

As discussed by WC01, anisotropic variations in the correlation functions can be accounted for by generalizing the diffusion coefficient K_p to a diffusion tensor $\mathbf{K}_p$. The diagonal elements, K_p^λ and K_p^ϕ, of the tensor can be adjusted relative to one another to allow the coordinates of the correlation model to be stretched or shrunk in one of the directions, thereby transforming circular correlation surfaces into elliptical ones. This is a useful feature to include near the equator where zonal scales are typically greater than meridional scales. More complicated anisotropic correlations can be produced by introducing off-diagonal terms in the diffusion tensor through a rotation of the coordinates of the correlation model (WC01).

Since horizontal scales in the ocean (and atmosphere) are generally much larger than vertical scales, it is convenient to keep a general separation of the correlation model into a horizontal and vertical component as in (19). This does not mean to say that the 3D correlation functions constructed with the GDE must be defined to be the product of strictly separable functions of the horizontal and vertical coordinate. For example, it may be desirable to use a non-separable formulation in which the horizontal (vertical) length scale is a function of the vertical (horizontal) coordinate. It may also be desirable to define the horizontal and vertical coordinates of the correlation model to be different from those of the ocean model. In an appropriately defined coordinate system, the separability assumption may not be a particularly restrictive one. For example, an isopycnal coordinate defined with respect to the background state would be a natural (flow-dependent) 2D coordinate in the tropical thermocline, whereas a terrain-following 2D coordinate would be appropriate near coastlines or near the ocean bottom. Coordinate transformations of these type can be handled within the tensorial formalism of the GDE and in some cases can be implemented straightforwardly by exploiting existing anisotropic tensors in ocean model diffusion parametrisations (Griffies *et al.* 1998; Madec *et al.* 1998).

5 Numerical aspects

The 3D correlation operator based on the GDE is formulated in grid-point space. The complete numerical representation of the operator is given by the symmetric product

$$\begin{aligned} \mathbf{C} &= \mathbf{\Lambda} \mathbf{L}^{1/2} \mathbf{W}^{-1} (\mathbf{L}^{1/2})^T \mathbf{\Lambda} \\ &= \left(\mathbf{\Lambda} \mathbf{L}^{1/2} \mathbf{W}^{-1/2} \right) \left(\mathbf{\Lambda} \mathbf{L}^{1/2} \mathbf{W}^{-1/2} \right)^T \end{aligned} \tag{20}$$

$$= \mathbf{C}^{1/2}(\mathbf{C}^{1/2})^T \tag{21}$$

where $\mathbf{L} = \mathbf{L}_h \mathbf{L}_v$ is the product of the horizontal ($\mathbf{L}_h$) matrix operator (9) and the vertical ($\mathbf{L}_v$) matrix operator defined by the inverse of the discrete version of (18). The factor $\mathbf{L}^{1/2} = \mathbf{L}_h^{1/2} \mathbf{L}_v^{1/2}$ corresponds to $M/2$ and $N/2$ iterations of $\mathbf{L}_h$ and $\mathbf{L}_v$, respectively. $\mathbf{W}$ is a diagonal matrix of volume elements that define the weights in the discrete analogue of the scalar product $\langle \eta_1, \eta_2 \rangle = \int\int\int \eta_1 \eta_2 \, a^2 \cos\phi \, d\lambda \, d\phi \, dz$. It enters naturally into the (symmetric) expression for $\mathbf{C}$ since $\mathbf{L}$ is self-adjoint with respect to this scalar product ($\mathbf{L} = \mathbf{W}^{-1}\mathbf{L}^T\mathbf{W}$). This property was already used in the 2D case to derive the square-root factorization in (5). It is worth remarking that even if the discrete representation of the Laplacian operator is not exactly self-adjoint (which may arise, for example, if the discretised form of a diffusion tensor K_p is not symmetric), the symmetric attribute of a correlation matrix can still be enforced practically by formulating $\mathbf{C}$, as in (20), as a product of square-root factors $\mathbf{C}^{1/2}$ and $(\mathbf{C}^{1/2})^T$. The matrix $\mathbf{\Lambda}$ in (20) is diagonal and contains normalization factors to ensure that the diagonal elements of $\mathbf{C}$ are equal to unity. These normalization factors are equal to the inverse of the intrinsic "standard deviations" of $\mathbf{L}^{1/2}\mathbf{W}^{-1}(\mathbf{L}^{1/2})^T$. The computation of these factors is an important aspect of the algorithm and is discussed shortly.

Figure 2 shows an example from the variational assimilation system for the OPA model (Weaver *et al.* 2003) of a correlation pattern generated using an implicit version of the GDE. The correlation pattern has been produced by applying (20) to a unit impulse located at a grid-point on the equator in the tropical Pacific. A bi-Laplacian version of the GDE ($P = 2$ with $\alpha_1 = 0$) with $M = 4$ iterations has been used in this example. The value of α_2 has been chosen to give an isotropic length scale of 4°. The diagonal elements of the diffusion tensor have been set to $K_2^\lambda = 4\alpha_2$ and $K_2^\phi = \alpha_2/4$ to give a locally anisotropic response at the equator. The negative lobes are clearly visible in Fig. 2, right panel, which shows a 1D representation at the equator of the correlation pattern in the left panel. Figure 2 gives an excellent match to the correlation function predicted from (14). In this example, a direct solver designed for sparse, multi-diagonal, symmetric matrices (Duff 2002) was used to produce an efficient inversion of (10). Iterative techniques such as conjugate gradient could also have been used to invert (19) (approximately) and these may be better suited for larger and denser versions of the GDE matrix (e.g., arising from the use of non-diagonal diffusion tensors) than the one considered in this example.

In an ocean model, the application of a correlation operator is complicated by the presence of continental boundaries. Boundary conditions can be imposed directly within a finite-difference representation of a Laplacian operator using a land-ocean mask array. This is a standard technique to account for complex boundaries in ocean models while keeping the symmetry of the finite-difference expression of the Laplacian (e.g., Madec *et al.* 1998). The application of the boundary condition will generally result in large changes in the amplitude of the GDE-generated covariance structures over short distances near the boundary. This is not a significant problem provided the amplitude can be estimated accurately and the normalization factors modified accordingly to ensure that the resulting covariance function has unit amplitude.

For the homogeneous and isotropic formulation of the GDE, the normalization matrix is simply a constant multiple of the identity matrix $\mathbf{I}$ (e.g., $\mathbf{\Lambda} = \sqrt{4\pi a^2/h(0)}\,\mathbf{I}$ for the 2D GDE in (13)). For anisotropic and inhomogeneous versions of the GDE, the normalization factors are no longer constant and a specific algorithm is required to compute them. Let Λ_i denote the i-th diagonal element of $\mathbf{\Lambda}$. Letting $\mathbf{v}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$, where the non-zero element is defined at the i-th grid point, then it follows from (20) that

$$\Lambda_i = \left(\widehat{\mathbf{v}}_i^T \widehat{\mathbf{v}}_i\right)^{-1/2} \tag{22}$$

where

$$\widehat{\mathbf{v}}_i = \mathbf{W}^{1/2}\mathbf{L}^{1/2}\mathbf{v}_i. \tag{23}$$

Each element of $\mathbf{\Lambda}$ can thus be computed exactly by applying the square root of the GDE operator to a unit impulse at each grid point. For a model with $O(10^5 - 10^6)$ independent grid points, the entire computation

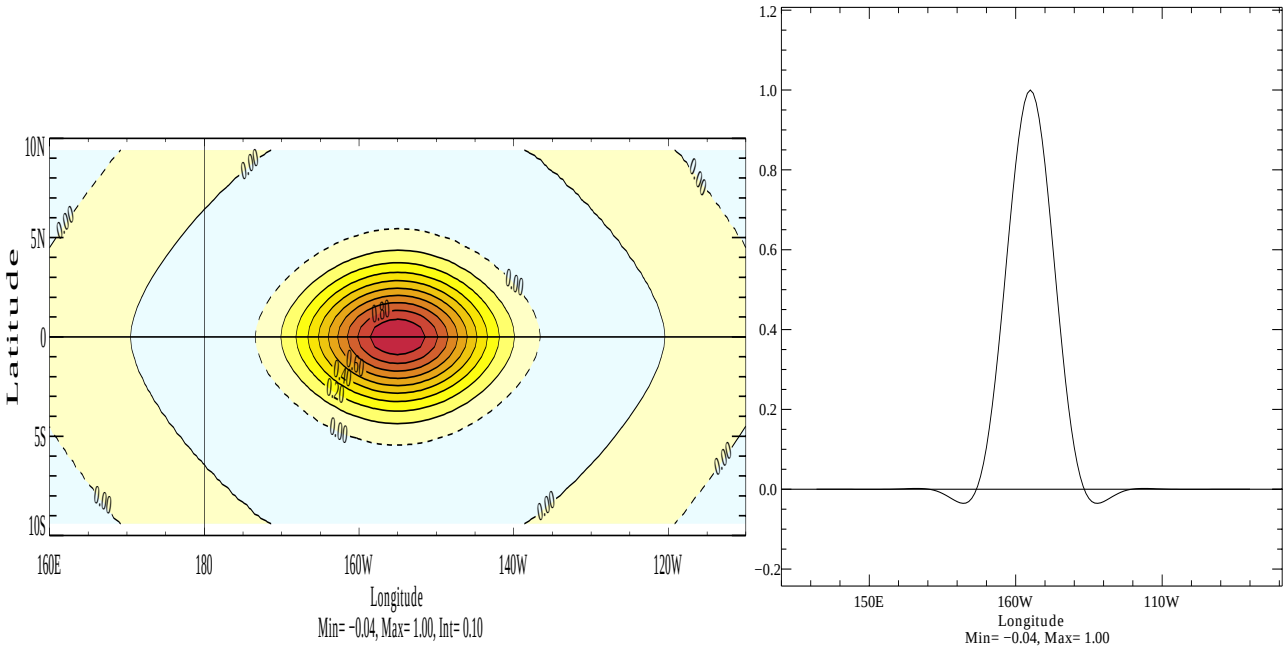

Figure 2: Left panel: horizontal section of a correlation function generated using an implicit version of the GDE with $P = 2$ ($\alpha_1 = 0$) and $M = 4$. The value of α_2 has been chosen to give an isotropic length scale of 4°. The correlations have been made anisotropic by stretching (shrinking) them zonally (meridionally) by a factor of 2. Right panel: the amplitude (vertical axis) of the correlation function as a function of longitude at the equator.

is clearly expensive. This may not be such a critical issue if the parameters of the correlation model are held constant from one assimilation cycle to the next since these factors would only need to be computed once, on the first assimilation cycle. However, if the parameters are varied between cycles, which would be the case for example with flow-dependent formulations of the diffusion tensor, then the normalization factors would have to be recomputed at the start of each assimilation cycle. To compute these factors on each cycle using the above method would lead to an enormous increase in the overall cost of the assimilation algorithm.

A practical algorithm for estimating the normalization factors is randomization (Fisher and Courtier 1995; Andersson 2003). The randomization algorithm is also based on several applications of the square-root operator but generally far fewer applications than required by the exact method. Let $\mathbf{v}_r$ denote a random vector with the property that $E[\mathbf{v}_r] = 0$ and $E[\mathbf{v}_r \mathbf{v}_r^T] = \mathbf{I}$, and define the square-root operator

$$\widehat{\mathbf{v}}_r = \mathbf{L}^{1/2} \mathbf{W}^{-1/2} \mathbf{v}_r \tag{24}$$

so that $E[\widehat{\mathbf{v}}_r \widehat{\mathbf{v}}_r^T] = \mathbf{L}^{1/2} \mathbf{W}^{-1} (\mathbf{L}^{1/2})^T$. Therefore, given an ensemble of R random vectors

$$\Lambda_i \approx \left(\mathrm{diag}_i \left\{ \frac{1}{R-1} \sum_{r=0}^{R} \widehat{\mathbf{v}}_r \widehat{\mathbf{v}}_r^T \right\} \right)^{-1/2} \tag{25}$$

where the randomization error is proportional to $1/\sqrt{R}$. The effect of the ensemble size on the accuracy of the correlations is illustrated in Fig. 3. Figure 3a shows an example of a "correct" correlation pattern for which the normalization factors have been computed exactly. Figures 3b and c show that the pattern produced using randomization estimates of the normalization factors for ensemble sizes of $R = 100$ and $R = 1000$; Figs 3d and e show their difference from the "correct" correlation pattern in Fig. 3a. For $R = 100$, there is noticeable distortion of the correlation pattern, with a maximum error of 0.14. Randomization errors of this magnitude

may lead to a small source of noise in the analysis (Fig. 3d). It will also implicitly modify the effective background-error standard deviations used in the analysis, but this effect may be relatively unimportant given our uncertainty of the true statistics of background error. For $R = 1000$, the distortion of the correlation pattern from the randomization error is hardly visible by eye (the maximum error is 0.04). Purser *et al.* (2003b) describe another method for estimating the normalization factors which also makes use of the square-root operator. Their method provides a good estimate of the normalization factors providing the scale parameters (K_p) are smoothly and slowly varying functions of the model's spatial coordinates.

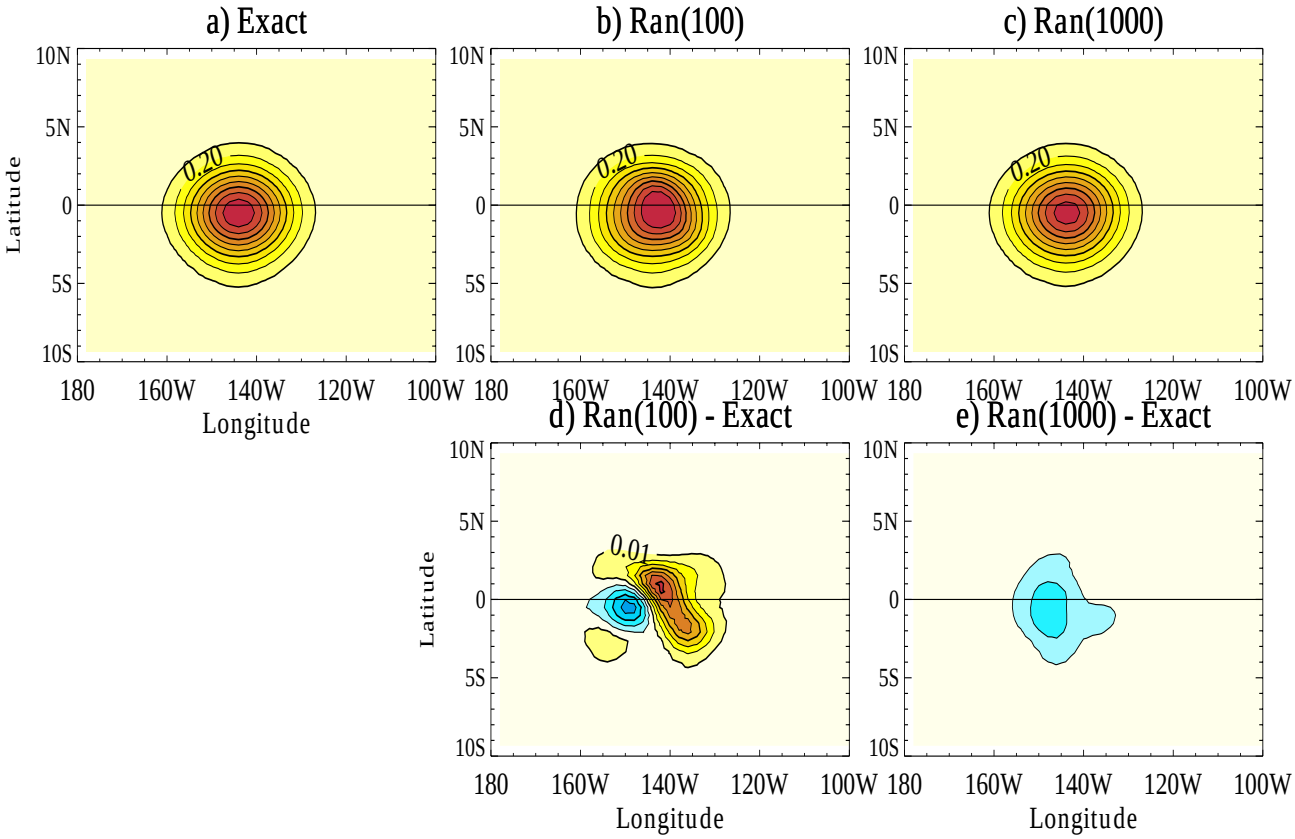

Figure 3: a) An example of a time-explicit GDE-generated correlation function where the normalization factors have been computed exactly. The same correlation function but with the normalization factors estimated from randomization with b) $R = 100$ and c) $R = 1000$. Panels d) and e) show the respective difference of the correlations in panels b) and c) with the exact correlations in panel a). The contour interval is 0.1 in a)–c) and 0.02 in d) and e).

Finally, it is worth pointing out the similarity between the time-implicit GDE and the recursive filter (Lorenc 1992; Purser *et al.* 2003a,b; Wu *et al.* 2003; Derber *et al.* 2003) approaches to correlation modelling. The connection between these methods is evident by considering the numerical solution of the inverse of the 1D equation (18) on an infinite domain in which the matrix representation of the second-derivative operator is decomposed into a symmetric Cholesky factorization $\mathbf{G}\mathbf{G}^T$ where $\mathbf{G}$ ($\mathbf{G}^T$) is a lower (upper) triangular matrix (Golub and Van Loan 1989). As discussed by Purser *et al.* (2003b), inverting equation (18) can then be obtained through a sequence of forward-elimination and backward-substitution steps analogous to the forward and backward iterations of the recursive filter. Supplied with appropriate boundary conditions on a finite domain, the recursive filter can in turn be matched to the forward-backward solution of the Langevin equations used in the representer method to evaluate a 1D (temporal) covariance operator with an exponential correlation function (Bennett *et al.* 1996; Chua and Bennett 2001; Ngodock 2003).

6 Conclusions

Generalized diffusion operators provide a well-based theoretical and practical framework for constructing general correlation models for data assimilation. In addition, they have the appealing feature of being conceptually straightforward, providing a "physically" intuitive way of interpreting the smoothing action of a correlation operator. Time-explicit diffusion schemes are much simpler to implement than time-implicit schemes. However, they have the disadvantage of restricting the class of correlation functions that the GDE can represent to those with high wavenumber variance spectra that decay at least as fast as the Gaussian. In addition, explicit schemes can be computationally expensive since a large number of iterations may be required to keep them stable. What is the most efficient method for solving a time-implicit GDE is an open question. Direct solvers are ideal for relatively small problems such as the one considered in the example in this paper, but become memory intensive and cumbersome for larger problems and on complicated grids. Iterative techniques could be promising provided good preconditioners can be found.

Acknowledgements

We would like to thank Andrea Piacentini and Luc Giraud for their help in implementing the matrix solver for the implicit diffusion equation. A. Weaver would like to thank Grace Wahba for pointing out her article on vector splines which helped him clarify the connection between diffusion- and spline-based correlation models.

References

Andersson, E., 2003: Modelling the temporal evolution of innovation statistics. This volume, pp 153–164.

Bennett, A. F., Chua, B. S. and Leslie, L. M., 1996: Generalized inversion of a global numerical weather prediction model. *Meteorol. Atms. Phys.*, **62**, 129–150.

Brasseur, P., Beckers, J. M., Brankart, J. M. and Schoenauen, R., 1996: Seasonal temperature and salinity fields in the Mediterranean Sea: climatological and analyses of an historical data set. *Deep Sea. Res.*, **43**, 159–192.

Chua, B. S. and Bennett, A. F., 2001: An inverse ocean modeling system. *Ocean Modelling*, **3**, 137–165.

Courtier, P., 1997: Dual formulation of four dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, **123**, 2449–2462.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljević, D, Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. *Q. J. R. Meteorol. Soc.*, **124**, 1783–1808.

Daley, R., 1991: *Atmospheric data analysis.* Cambridge atmospheric and space sciences series, Cambridge University Press, 457 pp.

Derber, J. and Rosati, A., 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333–1347.

Derber, J. and Bouttier, F. 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195–221.

Derber, J. C., Purser, R. J., Wu, W.-S., Treadon, R., Pondeca, M., Parrish, D. and Kleist, D.: Flow dependent

J_b in a grid-point 3D-Var. This volume, pp 125–134.

Duff, I, 2002: MA57 - A new code for the solution of sparse symmetric indefinite systems. RAL Technical Report, RAL-TR-2002-024

Egbert, G. D., Bennett, A. F. and Foreman, M. G. G., 1994: Topex/Poseidon tides estimated using a global inverse model. *J. Geophys. Res.*, **99**, 24,821–24,852.

Fisher, M. and Courtier, P., 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Tech. Memo. No. 220.

Fisher, M., 2003: Background error covariance modelling. This volume, pp 45–63.

Gaspari, G. and Cohn, S., 1999: Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**, 723–757.

Golub, G. H. and Van Loan, C. F., 1989: *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland.

Griffies, S. M., Gnanadesikan, A., Pacanowski, R. C., Larichev, V. D., Dukowicz, J. K. and Smith, R. D., 1998: Isoneutral diffusion in a *z*-coordinate ocean model. *J. Phys. Oceanogr.*, **28**, 805–830.

Hollingsworth, A. and Lönnberg, P., 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field *Tellus*, **38A**, 3–15.

Julian, P. R. and Thiébaux, H. J., 1975: On some properties of correlation functions used in optimum interpolation schemes. *Mon. Wea. Rev.*, **103**, 605–616.

Lönnberg, P. and Hollingsworth, A., 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus*, **38A**, 137–161.

Lorenc, A. C., 1992: Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.*, **118**, 569–591.

Madec, G., Delecluse, P., Imbard, M. and Levy, C., 1998: OPA 8.1 Ocean General Circulation Model Reference Manual. Technical note no. 11, LODYC/IPSL, Paris, France.

McIntosh, P. C., 1990: Oceanographic data interpolation: objective analysis and splines. *J. Geophys. Res.*, **95**, 13529–13541.

Ngodock, H. E., 2003: Efficient implementation of covariance multiplication for data assimilation with the representer method. *To appear in Ocean Modelling*.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.

Purser, R. J., Wu, W.-S., Parrish, D. F. and Roberts, N. M., 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524–1535.

Purser, R. J., Wu, W.-S., Parrish, D. F. and Roberts, N. M., 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: spatially inhomogeneous and anisotropic covariances. *Mon. Wea. Rev.*, **131**, 1536–1548.

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A. and Bouttier, F., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part II: Structure functions. *Q. J. R. Meteorol. Soc.*, **124**, 1809–1829.

Sheinbaum, J. and Anderson, D. L. T., 1990: Variational assimilation of XBT data. Part II: sensitivity studies and use of smoothing constraints. *J. Phys. Oceanogr.*, **20**, 689–704.

Stammer, D., 1997: Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements. *J. Phys. Oceanogr.*, **27**, 1743–1768.

Wahba, G. and Wendelberger, J., 1980: Some new mathematical methods for variational objective analysis using splines and cross-validation. *Mon. Wea. Rev.*, **108**, 36–57.

Wahba, G., 1981: Spline interpolation and smoothing on the sphere. *SIAM. J. Scientific and Statistical Computing*, **2**, No. 1, 5–16.

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*, W. Schempp and K. Zeller, eds., Birkhauser, pp 407–429.

Weaver, A. T. and Courtier, P., 2001: Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.*, **127**, 1815–1846.

Weaver, A. T., Vialard, J. and Anderson, D. L. T., 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. Wea. Rev.*, **131**, 1360–1378.

Wilke, C. K., Milliff, R. F. and Large, W. G., 1999: Surface wind variability on spatial scales from 1 to 1000km observed during TOGA COARE, *J. Atmos. Sci.*, **56**, 2222–2231.

Wu, W.-S., Purser, R. J. and Parrish, D. F., 2003: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.