TECHNICAL MEMORANDUM

306

Correlation modelling on the sphere using a generalized diffusion equation

A. T. Weaver and P. Courtier

Research Department

June 2000

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Correlation modelling on the sphere using a generalized diffusion equation

A. T. Weaver[1] and P. Courtier[2]

[1] *Present affiliation: Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, Toulouse, France*
[2] *Present affiliation: Météo-France, Paris, France*

Abstract

An important element of a data assimilation system is the statistical model used for representing the correlations of background error. This paper describes a practical algorithm that can be used to model a large class of two- and three-dimensional, univariate correlation functions on the sphere. Application of the algorithm involves a numerical integration of a generalized diffusion-type equation (GDE). The GDE is formed by replacing the Laplacian operator in the classical diffusion equation by a weighted linear combination of powers of the Laplacian. The general solution of the GDE defines, after appropriate normalization, a correlation operator on the sphere, which can be interpreted as a convolution integral, the kernel of which is an isotropic correlation function. The free parameters controlling the shape and length scale of the correlation function are the products $\kappa_p T$, $p = 1, 2, ...$, where $(-1)^p \kappa_p$ is a weighting ("diffusion") coefficient ($\kappa_p > 0$) attached to the Laplacian with exponent p, and T is the total integration "time". The algorithm can be viewed as an extension of the iterative Laplacian grid-point smoother of Derber and Rosati (1989), the latter following from a numerical integration of the classical diffusion equation (i.e., a special case of the GDE with $\kappa_p = 0$ for all $p > 1$), for which the correlation function can be shown to be well approximated by a Gaussian with length scale equal to $\sqrt{2\kappa_1 T}$.

The Laplacian-based correlation model is particularly well suited for ocean models as it can be easily generalized to account for complex boundaries imposed by coastlines. Furthermore, a one-dimensional analogue of the GDE can be used to model a family of vertical correlation functions, which in combination with the two-dimensional GDE, forms the basis of a three-dimensional, (generally) non-separable correlation model. Generalizations to account for anisotropic correlations are also possible by stretching and/or rotating the computational coordinates via a symmetric "diffusion" tensor. Examples are presented from a variational assimilation system currently under development for the OPA ocean general circulation model.

1 Introduction

A central task in the development of a statistical (e.g., variational) data assimilation system is the estimation and representation of the background state error covariances. In practice,

these covariances must be modelled as approximations to the true covariances of background error. In general, this is done by parameterizing the covariances using smoothing functions or filters, and balance relationships which employ such simplifying assumptions as isotropy, homogeneity, geostrophy etc. (Daley 1991). The number of adjustable parameters in the covariance model, such as the standard deviations and correlation length scales of the background error, is usually quite small compared to the number of individual covariances that actually need to be specified (typically on the order of 10^{11} or greater for global applications with general circulation models of the atmosphere or ocean). These parameters can be determined by fitting the covariance models to available statistical information on the background errors, obtained, for example, from statistics of the observed-minus-background field (Hollingsworth and Lönnberg 1986; Lönnberg and Hollingsworth 1986) or of some other proxy for background error.

Ideally, the covariance models should be made flexible enough to capture the main characteristics of the available estimates of the background error covariances. This paper focusses on a particular aspect of covariance modelling, namely that of deriving efficient and flexible algorithms for representing two-dimensional (2D) and three-dimensional (3D), univariate correlations on the sphere, with a special emphasis on the applicability of these algorithms for variational data assimilation systems with ocean models.

The background error covariance models employed in variational assimilation systems with atmospheric models commonly rely on a representation of correlation functions in terms of a spherical-harmonic expansion. Such a formulation is used operationally at the European Centre for Medium-Range Weather Forecasts (ECMWF) (Courtier *et al.* 1998; Rabier *et al.* 2000), the National Centers for Environmental Prediction (NCEP) (Parrish and Derber 1992), the Canadian Meteorological Center (Gauthier *et al.* 1999) and the United Kingdom Meteorological Office (UKMO) (Lorenc 1997). The free parameters in the correlation model are the spectral variances of the correlation function, which are generally specified from statistics of forecast differences valid at the same time (Rabier *et al.* 1998). Spectrally-based correlation models, however, are not very practical for ocean models since the lateral boundary conditions imposed by the coastlines are difficult to handle within a global spectral basis function expansion. Correlation models formulated directly in physical (grid-point) space are generally easier to adapt to bounded-domains and hence better suited for the ocean assimilation problem. For example, Lorenc (1992, 1997) and Parrish *et al.* (1997) have developed correlation models based on recursive grid-point filters, one variant of which is applied routinely in the UKMO real-time global ocean forecasting system (Bell *et al.* 2000). The recursive filter is very efficient and can accommodate geographical variations in correlation length scale, but has limited flexibility in the shape of the correlation function and is difficult to make isotropic (Lorenc 1997).

Derber and Rosati (1989) proposed the use of an iterative Laplacian grid-point filter (smoother) in order to build a 2D isotropic correlation function that approximates a Gaussian. The Derber and Rosati scheme is used operationally at NCEP for producing ocean initial conditions for seasonal climate forecasting (Behringer *et al.* 1998). A close variant of the algorithm has been described in some detail by Egbert *et al.* (1994) and Bennett *et al.* (1997), who interpret the Laplacian filter in terms of a time-step integration of a diffusion equation. The key idea is that the integral solution of the diffusion equation defines a covariance operator. In this paper, we build on these ideas to construct 2D and 3D univariate correlation models that

are both numerically efficient and sufficiently general to support a number of desirable features: correlation functions with different shapes (not just Gaussian), geographically variable length scales, horizontal/vertical non-separability, and 3D anisotropy.

The paper is organized as follows. Section 2 presents the variational assimilation problem. A general, multivariate formulation of the background error covariance matrix is also described in order to isolate clearly the univariate, correlation component, the specification of which is the subject of this paper. In section 3, the main features of the 2D "Gaussian" correlation model of Derber and Rosati are described. Particular attention is given to the interpretation of the model in terms of the spherical harmonics (ignoring boundaries). A straightforward extension of this correlation model is then proposed as a means of generating a general class of 2D isotropic correlation functions on the sphere. A one-dimensional (1D) version of the model is also introduced for representing a class of vertical correlation functions. Several issues related to the practical implementation of the algorithm are discussed in section 4. A 3D correlation model is also developed. In section 5, the correlation structures are illustrated in several examples from a variational assimilation system (Weaver and Vialard 1999) for the OPA Ocean General Circulation Model (OGCM) of the Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC) (Madec *et al.* 1999). In particular, the impact of ocean boundaries and various extensions to account for 2D and 3D anisotropy are discussed and illustrated. Concluding remarks are made in section 6. An appendix provides some mathematical details and a numerical illustration of a matching procedure that can be used to relate the exact solution of the diffusion equation on the sphere to a Gaussian covariance operator.

The notation used in the paper closely follows that of Ide *et al.* (1997).

2 Formulation of the problem

2.1 Variational assimilation and its dual formulation

Following Courtier *et al.* (1998), the variational formulation of three-dimensional assimilation (3D-Var) is introduced. In its incremental formulation (Courtier *et al.* 1994), 3D-Var attempts to minimize the following cost function,

$$J(\delta\mathbf{x}) = \underbrace{\frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x}}_{J_b} + \underbrace{\frac{1}{2}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})}_{J_o} \tag{1}$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$, $\mathbf{x}$ being the model state vector and $\mathbf{x}^b$ its background estimate. The observation vector, $\mathbf{y}^o$, is contained within the the innovation vector

$$\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b) \tag{2}$$

where $\mathbf{H}$ is a suitable linear approximation of the observation operator $H(\cdot)$ in the vicinity of $\mathbf{x}^b$. The matrices $\mathbf{B}$ and $\mathbf{R}$ contain the assumed covariances of background and observation error, respectively. In practice, these matrices are defined as approximations to the true error covariance matrices

$$\mathbf{B} \equiv E\left[\left(\mathbf{x}^b - E[\mathbf{x}]\right)\left(\mathbf{x}^b - E[\mathbf{x}]\right)^T\right] \tag{3}$$

$$\mathbf{R} \equiv E\left[(\mathbf{y}^o - H(E[\mathbf{x}]))\,(\mathbf{y}^o - H(E[\mathbf{x}]))^T\right] \tag{4}$$

where $E[\cdot]$ denotes mathematical expectation.

The observation term (J_o) measures the fit between the model increment in observation space and the innovation vector. The background term (J_b) penalizes the size of the increment vector (i.e., measures the fit to the background state). At the minimum, the resulting analysis increment $\delta\mathbf{x}^a$ is added to $\mathbf{x}^b$ in order to provide the analysis $\mathbf{x}^a$;

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a. \tag{5}$$

Following Courtier (1997) and Derber and Bouttier (1999), the cost function (1) is rewritten in terms of a new variable, $\mathbf{v}$, defined by

$$\mathbf{v} = \mathbf{B}^{-T/2}\,\delta\mathbf{x} \tag{6}$$

where $\mathbf{B}^{1/2}$ is taken to be any "square-root" matrix such that $\mathbf{B} = \mathbf{B}^{T/2}\mathbf{B}^{1/2}$. This leads to the equivalent analysis problem of minimizing

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}) \tag{7}$$

with

$$\delta\mathbf{x} = \mathbf{B}^{T/2}\mathbf{v}. \tag{8}$$

A gradient descent method is used to iterate to the minumum of J. On each iteration, the gradient of J with respect to $\mathbf{v}$ is required, which from (7) and (8) is given by

$$\nabla_{\mathbf{v}}J = \nabla_{\mathbf{v}}J_b + \nabla_{\mathbf{v}}J_o = \mathbf{v} + \mathbf{B}^{1/2}\,\nabla_{\delta\mathbf{x}}J_o \tag{9}$$

where

$$\nabla_{\delta\mathbf{x}}J_o = \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}). \tag{10}$$

At the start of minimization $\mathbf{v} = \delta\mathbf{x} = 0$, providing the initial guess is taken to be $\mathbf{x} = \mathbf{x}^b$. As a result, only the transformation (8) from $\mathbf{v}$ to $\delta\mathbf{x}$ (involving $\mathbf{B}^{T/2}$), and the associated adjoint transformation in (9) from $\nabla_{\delta\mathbf{x}}J_o$ to $\nabla_{\mathbf{v}}J_o$ (involving $\mathbf{B}^{1/2}$) are actually required on each iteration. This formulation thus allows us to circumvent the explicit specification of the inverse matrix $\mathbf{B}^{-1}$, while at the same time providing a generally efficient preconditioner for the minimization (Lorenc 1988).

The three-dimensional Physical-space Statistical Analysis System (3D-PSAS) (Cohn *et al.* 1997) provides an alternative algorithm for solving the incremental, variational analysis problem (Courtier 1997). 3D-PSAS involves the iterative minimization of a cost function of the form

$$F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{w} - \mathbf{w}^T\mathbf{d} \tag{11}$$

with

$$\delta \mathbf{x} = \mathbf{B}\mathbf{H}^T \mathbf{w}, \tag{12}$$

$\mathbf{w}$ being an increment defined in the dual of observation space. As with 3D-Var, 3D-PSAS may be implemented using an algorithm for applying $\mathbf{B}^{1/2}$ and $\mathbf{B}^{T/2}$ to a given vector; the inverse operators are not required providing $\mathbf{w} = 0$ ($\mathbf{x} = \mathbf{x}^b$) is the first guess for minimization. Strictly speaking, only $\mathbf{B}$ is required for 3D-PSAS. Nevertheless, the definition of $\mathbf{B}$ through a square-root factorization is a practical way of ensuring both its symmetry and positive definiteness (Gaspari and Cohn 1999).

It is worth remarking that, although we focus the discussion in this paper on $\mathbf{B}$ and the associated 3D variational analysis problem, the basic techniques we develop will be equally applicable for representing model and/or boundary-forcing error covariances in weak-constraint four-dimensional variational assimilation (4D-Var) or its dual formulation (4D-PSAS) (Egbert *et al.* 1994; Courtier 1997).

2.2 General formulation of the background error covariance matrix

Because of its size, $\mathbf{B}$ can be neither estimated completely nor stored explicitly. (Recall that $\mathbf{B}$ is an $N \times N$ matrix where N, the dimension of the model state vector, is typically greater than 10^6.) We are therefore forced to model $\mathbf{B}$ as an operator. Specifically, we wish to define $\mathbf{B}$ as a sequence of operators, from which a factorization of the form $\mathbf{B}^{T/2}\,\mathbf{B}^{1/2}$ can be easily deduced. For convenience we continue to use matrix notation, although it is important to bear in mind that the matrices are never explicitly computed in practice. By definition, $\mathbf{B}$ is a symmetric and positive definite matrix. The diagonal and off-diagonal elements of $\mathbf{B}$ correspond to the background state error variances and covariances respectively. The covariances can be split into two contributions: block-diagonal elements that represent the auto-covariances between grid-points corresponding to a particular model variable (i.e., the univariate component of $\mathbf{B}$), and the remaining elements that represent the cross-covariances between grid-points corresponding to different model variables (i.e., the multivariate component of $\mathbf{B}$).

Recently, Derber and Bouttier (1999) have proposed a compact and powerful formulation of the background error covariance matrix which allows separating the univariate and multivariate components of $\mathbf{B}$ into distinct operators. It is useful to outline their formulation in order to put the present work in the context of a fully multivariate covariance operator. In their formulation, the model variables are first partioned into balanced and unbalanced components, except for one variable which is taken in totality (i.e., for this variable there is no distinction between balanced and unbalanced). This (total) variable is then used to establish linear balances with the remaining variables. This procedure can be represented through a linear balance operator, $\mathbf{K}_b'$, acting on the unbalanced variables. As the model increment is defined as the sum of the balanced and unbalanced components, the full balance operator is $\mathbf{K}_b = \mathbf{K}_b' + \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. In terms of the balance operator, $\mathbf{B}$ takes the form

$$\mathbf{B} = \mathbf{K}_b\,\mathbf{B}_u\,\mathbf{K}_b^T \tag{13}$$

where $\mathbf{B}_u$ is the error covariance matrix for the unbalanced variables, which is taken to have

a block-diagonal structure (i.e., the cross-covariances between the unbalanced variables are assumed to be negligible).

By definition, $\mathbf{B}_u$ can be factored as

$$\mathbf{B}_u = \mathbf{\Sigma}\,\mathbf{C}\,\mathbf{\Sigma} \tag{14}$$

where $\mathbf{\Sigma}$ is a diagonal matrix of background error standard deviations and $\mathbf{C}$ a symmetric matrix of background error correlations for the unbalanced variables. Thus, from (13) and (14), we can represent the transformation in (8) by the sequence of operators

$$\delta\mathbf{x} = \mathbf{K}_b\,\mathbf{\Sigma}\,\mathbf{C}^{T/2}\,\mathbf{v} \tag{15}$$

where $\mathbf{C}^{1/2}$ is defined such that $\mathbf{C} = \mathbf{C}^{T/2}\mathbf{C}^{1/2}$. As $\mathbf{C}$ is block-diagonal, the operator $\mathbf{C}^{T/2}\,\mathbf{v}$ in (15) can in turn be split into individual operators, $\mathbf{C}_\alpha^{T/2}\,\mathbf{v}_\alpha$ that act independently on the different variable components, $\mathbf{v}_\alpha$, of $\mathbf{v}$.

The purpose of this paper is to focus on the design of univariate correlation operators ($\mathbf{C}_\alpha = \mathbf{C}_\alpha^{T/2}\mathbf{C}_\alpha^{1/2}$) that are both efficient for large state vector assimilation problems and well adapted to oceanographic applications in complex-boundary domains. Deriving an appropriate multivariate balance operator ($\mathbf{K}_b$) and estimating the free statistical parameters in the covariance model (e.g., $\mathbf{\Sigma}$) are equally important issues but are beyond the scope of this paper.

3 Correlation modelling on the sphere

3.1 Modelling a Gaussian correlation function using the diffusion equation

The following 1D problem provides a simple framework for interpreting the basic procedure for constructing correlation operators on the sphere, which we develop subsequently in sections 3b-c. The starting point is the 1D diffusion equation

$$\frac{\partial\eta}{\partial t} - \kappa\frac{\partial^2\eta}{\partial z^2} = 0 \tag{16}$$

where κ is a spatially homogeneous diffusion coefficient. We consider solutions to (16) on the real line $I\!R$ such that $\eta(z,t)$ vanishes as $z \to \pm\infty$. It is convenient to represent the general solution symbolically by an integral operator $\mathcal{L}$;

$$\eta(z,0) \xrightarrow{\mathcal{L}} \eta(z,T) \tag{17}$$

where $\eta(z,T)$ is the result of integrating (16) over a time interval $0 \le t \le T$ with $\eta(z,0)$ as initial condition.

A simple interpretation of the above operator can be given by considering the explicit solution of (16). The Fourier transform of (16) yields

$$\frac{\partial\widehat{\eta}(\widehat{z},t)}{\partial t} = -\kappa\widehat{z}^2\widehat{\eta}(\widehat{z},t) \tag{18}$$

where $\widehat{\eta}(\widehat{z},t)$ denotes the Fourier transform of $\eta(z,t)$ and $\widehat{z}$ is the variable in Fourier space. Equation (18) is readily integrated in time to give

$$\widehat{\eta}(\widehat{z},T) = \widehat{\eta}(\widehat{z},0)\, e^{-\kappa T \widehat{z}^2}. \tag{19}$$

In (19) we recognize the product of $\widehat{\eta}(\widehat{z},0)$ with a Gaussian function. As the inverse Fourier transform of a Gaussian is also a Gaussian, $\eta(z,T)$ is the result of the convolution of $\eta(z,0)$ with a Gaussian function;

$$\eta(z,T) = \frac{1}{\sqrt{4\pi\kappa T}} \int_{z'} e^{-(z-z')^2/4\kappa T}\, \eta(z',0)\, dz' \tag{20}$$

where the product $2\kappa T$ can be interpreted as the square of the length scale of the Gaussian function.

The Gaussian in (20) is homogeneous and isotropic as it depends only on Euclidean distance $r = |z - z'|$. It is well known that such a function defines a valid (positive definite) covariance function on $I\!R$. This follows from the fact that its Fourier transform (the exponential in (19)) is everywhere positive (e.g., see Gaspari and Cohn 1999). The integral equation (20) thus defines a covariance operator, which provides an interpretation of the operator $\mathcal{L}$.

By definition, a correlation function has unit variance; i.e., its value at the origin, $r = 0$, equals one. The Gaussian covariance operator $\mathcal{L}$ can be easily transformed into a Gaussian correlation operator $\mathcal{C}_\eta$ by post-multiplying $\eta(z,T)$ by the constant factor $\sqrt{4\pi\kappa T}$, which can be represented symbolically by

$$\eta(z,0) \xrightarrow{\mathcal{C}_\eta} \sqrt{4\pi\kappa T}\ \eta(z,T). \tag{21}$$

On a discrete grid, the action of $\mathcal{C}_\eta$ may be effectuated in one of two ways: either by evaluating the convolution integral in (20) directly using numerical quadrature, or by the generally more efficient technique of iterating a discretized (e.g., finite difference) version of the differential equation (16) and normalizing the result as in (21). The latter is the essence of the Derber and Rosati scheme (see also Egbert *et al.* (1994)). In the next section, we present the theoretical basis of the scheme for application on the sphere.

3.2 Interpretation on the sphere

In 3D Euclidean space ($I\!R^3$), the solution of the diffusion equation leads to a convolution integral with a Gaussian function that depends only on Euclidean distance $r = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}$ between points (x,y,z) and (x',y',z') in $I\!R^3$. This is an obvious extension of the 1D example in the previous section. Any homogeneous and isotropic correlation function in $I\!R^3$, such as the 3D Gaussian, can be readily transformed into a valid isotropic correlation function on the spherical space S^2 if Euclidean distance is substituted by *chordal* distance

$$r = 2a \sin\theta/2 = 2a(1 - \cos\theta)^{1/2} \tag{22}$$

where θ, $0 \leq \theta \leq \pi$, is the angular separation (great circle distance) between points on the sphere of radius a (Weber and Talkner 1993; Gaspari and Cohn 1999).[1] One might expect, therefore, that the solution of the diffusion equation on the sphere leads to a convolution integral involving the Gaussian function $e^{-r^2/2L^2}$ where r is given by (22) and $L^2 = 2\kappa T$ as in the 1D example. This turns out not to be the case as shown below; the solution has a covariance interpretation but the actual covariance function is not Gaussian, even though numerically it can be very closely matched to a Gaussian as discussed by Hartman and Watson (1974). This feature is investigated in more detail in the appendix.

Consider the diffusion equation

$$\frac{\partial \eta}{\partial t} - \kappa \nabla^2 \eta = 0 \tag{23}$$

where ∇^2 is the Laplacian operator defined on the sphere Σ. When the domain over Σ contains boundaries (e.g., coastlines in the case of an ocean model), we must supplement (23) with boundary conditions. For the moment we consider Σ to be free of boundaries. Boundary related issues will be addressed in section 5b.

A scalar field $\eta(\lambda, \phi, t)$ may be expanded as

$$\eta(\lambda, \phi, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \eta_n^m(t) Y_n^m(\lambda, \phi) \tag{24}$$

where $Y_n^m(\lambda, \phi)$ are the spherical harmonics, λ is longitude ($0 \leq \lambda \leq 2\pi$) and ϕ is latitude ($-\pi/2 \leq \phi \leq \pi/2$). The $Y_n^m(\lambda, \phi)$ are orthogonal and normalized following the usual convention in meteorology,

$$\frac{1}{4\pi a^2} \int_\Sigma Y_n^m(\lambda, \phi) Y_{n'}^{m'*}(\lambda, \phi)\, d\Sigma = \delta_{nn'} \delta_{mm'} \tag{25}$$

where $*$ denotes complex conjugate, $\delta_{nn'}$ is the Kroneker delta, and $d\Sigma = a^2 \cos\phi \, d\lambda \, d\phi$. Since the spherical harmonics are the eigenvectors of the Laplacian operator on the sphere, with $-n(n+1)/a^2$ the associated eigenvalues, the diffusion equation (23) becomes

$$\frac{d\eta_n^m}{dt} = -\kappa \frac{n(n+1)}{a^2} \eta_n^m, \tag{26}$$

which over the interval $[0, T]$ readily integrates as

$$\eta_n^m(T) = \eta_n^m(0) e^{-\kappa T n(n+1)/a^2}. \tag{27}$$

The spectral coefficients $\eta_n^m(0)$ can be determined by multiplying (24) by $Y_n^{m*}(\lambda, \phi)$ and applying the orthonormality condition (25);

$$\eta_n^m(0) = \frac{1}{4\pi a^2} \int_\Sigma \eta(\lambda, \phi, 0)\, Y_n^{m*}(\lambda, \phi)\, d\Sigma. \tag{28}$$

[1] As pointed out by Weber and Talkner, the Gaussian $e^{-(a\theta)^2/2L^2}$, whose argument $-a\theta$ follows from a first-order development of (22) for small θ, is not necessarily a valid (positive definite) correlation function on the sphere even though it has often been used in meteorology and oceanography.

Substituting (28) into (27), and (27) into (24), and applying the Addition Theorem for the spherical harmonics (Arfken 1966, pg. 450), the solution to (23) can be written as

$$\eta(\lambda,\phi,T) = \frac{1}{4\pi a^2}\int_{\Sigma'} f(\theta;\kappa T)\,\eta(\lambda',\phi',0)\,d\Sigma' \tag{29}$$

where

$$f(\theta;\kappa T) = \sum_{n=0}^{\infty} f_n P_n^0(\cos\theta) = \sum_{n=0}^{\infty} \sqrt{2n+1}\,e^{-\kappa T n(n+1)/a^2} P_n^0(\cos\theta), \tag{30}$$

θ being the angular separation between the points (λ,ϕ) and (λ',ϕ') on the sphere, and P_n^0 the Legendre polynomials. As the coefficients f_n of the Legendre polynomials are positive, the integral kernel $f(\theta;\kappa T)$ is the *representation* of an isotropic covariance function (Courtier *et al.* 1998). Equation (29) thus defines a covariance operator and can be interpreted as a convolution of the scalar field $\eta(\lambda,\phi,0)$ with the covariance function $f(\theta;\kappa T)$.

The value of the covariance function at the origin ($\theta = 0$) gives the variance at any point;

$$f(0;\kappa T) = \sum_{n=0}^{\infty} f_n\sqrt{2n+1} = \sum_{n=0}^{\infty} (2n+1)\; e^{-\kappa T n(n+1)/a^2} \tag{31}$$

where we have used the fact that $P_n^0(1) = \sqrt{2n+1}$ as implied by the normalization in (25). Each coefficient $f_n\sqrt{2n+1}$ gives the contribution to the total variance from a given total wavenumber n, and hence provides the variance power spectrum of $f(\theta;\kappa T)$. Furthermore, $f(\theta;\kappa T)/f(0;\kappa T)$ defines the associated correlation function.

The length scale of the correlation function can be defined following Daley (1991);

$$\begin{aligned} L^2 &= -2\frac{f(0;\kappa T)}{\nabla^2 f(0;\kappa T)} \\ &= 2a^2\frac{\sum_{n=0}^{\infty}(2n+1)\,e^{-\kappa T\, n(n+1)/a^2}}{\sum_{n=0}^{\infty} n\,(n+1)\,(2n+1)\,e^{-\kappa T\, n(n+1)/a^2}} \end{aligned} \tag{32}$$

which shows that, as in the 1D example, κT is the parameter controlling the length scale of the correlation function. Moreover, as shown in the appendix, for scales relevant in meteorology and oceanography ($L \ll a$), matching (30) to a Gaussian leads to

$$L^2 \approx 2\kappa T \tag{33}$$

in direct analogy with the 1D example.

3.3 Extension to a larger class of isotropic correlation functions on the sphere

The correlation model can be extended to represent a larger class of correlation functions by considering the solution of a more general partial differential equation,

$$\frac{\partial\eta}{\partial t} + \sum_{p=1}^{P} \kappa_p \left(-\nabla^2\right)^p \eta = 0 \tag{34}$$

where κ_p, $p = 1, ..., P$, are non-negative weighting coefficients. The classical diffusion equation (23) is a special case of (34) with $\kappa_p = 0$ for all $p > 1$. Thus, we will refer to (34) hereafter as a generalized "diffusion" equation (GDE). The general solution of (34) is of the form (29),

$$\eta(\lambda, \phi, T) = \frac{1}{4\pi a^2} \int_{\Sigma'} f(\theta; \kappa_1 T, ..., \kappa_P T)\, \eta(\lambda', \phi', 0)\, d\Sigma' \tag{35}$$

with integral kernel given by

$$f(\theta; \kappa_1 T, ..., \kappa_P T) = \sum_{n=0}^{\infty} \sqrt{2n+1}\, \exp\left\{ -\sum_{p=1}^{P} \kappa_p T \left(n(n+1)/a^2 \right)^p \right\} P_n^0(\cos\theta). \tag{36}$$

Equation (36) defines a family of isotropic covariance functions on the sphere; the associated correlation functions are $f(\theta; \kappa_1 T, ..., \kappa_P T)/f(0; \kappa_1 T, ..., \kappa_P T)$. The length scale of the correlation function can be defined in the usual way;

$$\begin{aligned} L^2 &= -2 \frac{f(0; \kappa_1 T, ..., \kappa_P T)}{\nabla^2 f(0; \kappa_1 T, ..., \kappa_P T)} \\ &= 2a^2 \frac{\sum_{n=0}^{\infty} (2n+1) \exp\left\{ -\sum_{p=1}^{P} \kappa_p T \left(n(n+1)/a^2 \right)^p \right\}}{\sum_{n=0}^{\infty} n\,(n+1)\,(2n+1) \exp\left\{ -\sum_{p=1}^{P} \kappa_p T \left(n(n+1)/a^2 \right)^p \right\}}. \end{aligned} \tag{37}$$

The free parameters controlling the spectrum (shape) and length scale of the correlation function are the sequence of products $\kappa_p T$, $p = 1, ..., P$, where P is defined such that $\kappa_P \neq 0$ and $\kappa_p = 0$ for all $p > P$. For the special case $P = 1$, (36) and (37) reduce to (30) and (32) with $\kappa_1 = \kappa$.

Figure 1 illustrates the variance power spectrum (upper panel) and grid-point representation (lower panel) of four different correlation functions generated numerically using (36) with a truncation at total wavenumber 106. The solid curves in Fig. 1 correspond to the near-Gaussian ($P = 1$) from Fig. 7 in the appendix. The dashed and dashed-dotted curves correspond to $P = 2$ and $P = 3$ respectively, with $\kappa_p = 0$ for all $p < P$ in both cases. The dotted curves also correspond to $P = 2$ but with non-zero values defined for both κ_1 and κ_2. For each correlation function the (non-zero) products $\kappa_p T$ have been tuned according to (37) in order to give a length scale of 500 km as in the Gaussian example in Fig. 7.

The effect of including higher order terms in the GDE (while holding the length scale fixed) is clear. From the spectrum, there is a progressive decrease in the variance at large scales (wavenumbers 1 to 14), increase in the variance at intermediate scales (wavenumbers 15 and 30), and sharper drop-off in the variance at small scales (wavenumbers greater than 30). In grid-point space, the correlation functions are very similar over $0 \leq r < L$. The higher order terms in the GDE, however, increase the rate at which the correlation function drops off to zero over $L < r < 2L$, and increase the amplitude of the negative-positive oscillations in the correlation function for $r > 2L$, with these oscillations tending to zero at large separation distances.

3.4 Modelling vertical correlation functions

In direct analogy to (34), a generalized 1D diffusion equation can be used as the basis for modelling a class of 1D correlation functions. Consider the equation

$$\frac{\partial \eta}{\partial t} + \sum_{r=1}^{R} \kappa_r^v \left(-\frac{\partial^2}{\partial z^2} \right)^r \eta = 0 \tag{38}$$

where R and κ_r^v are analogous to P and κ_p in (34). The general solution of (38) on IR can be easily obtained using the Fourier transform, following the example in section 3a. In particular, the solution can be expressed as a convolution integral, the integral kernel of which is a covariance function given by the inverse Fourier transform

$$h(z; \kappa_1^v T_v, ..., \kappa_R^v T_v) = \frac{1}{2\pi} \int_{\widehat{z}} \exp \left\{ -\sum_{r=1}^{R} \kappa_r^v T_v \widehat{z}^{2r} \right\} e^{-\hat{i}\, \widehat{z} z}\, d\widehat{z} \tag{39}$$

where T_v is the total integration time and $\hat{i} = \sqrt{-1}$. As with $f(\theta; \kappa_1 T, ..., \kappa_P T)$, the sequence of products of the diffusion coefficients (κ_r^v) with the integration time (T_v) controls the shape and length scale of $h(z; \kappa_1^v T_v, ..., \kappa_R^v T_v)$. For $R = 1$, (39) reduces analytically to the Gaussian in (20). For $R > 1$, with $\kappa_R \neq 0$, there is no obvious closed analytical form of the covariance functions; however, for given values of $\kappa_r^v T_v$, (39) can be solved numerically to yield a family of covariance functions similar to those in Fig. 1. Taking z as the vertical coordinate in the model, then (38), together with appropriate top and bottom boundary conditions, provides a statistical model that may be used for representing the vertical correlations in $\mathbf{C}_\alpha$, as discussed further in sections 4c and 5d.

4 Practical implementation

In the previous section, we showed that the integral solution of the diffusion equation on the sphere could be interpreted as a covariance operator. In particular, we showed that the solution could be expressed as a convolution of the initial condition with an isotropic covariance function that could be closely matched to a Gaussian. The possibility of representing other isotropic covariance functions by solving a more general differential equation was also discussed. In this section, we discuss the practical implementation of the algorithm within the design of a univariate correlation operator. Some generalizations of the algorithm will be discussed and illustrated in section 5 with reference to specific examples.

4.1 Discrete representation of the generalized diffusion equation within a horizontal, univariate correlation operator

Consider the following semi-discrete representation of the GDE (Eq. (34)), based on an explicit forward differencing of the temporal derivative;

$$\eta(t_m) = \eta(t_{m-1}) - \sum_{p=1}^{P} \kappa_p \Delta t \left(-\nabla^2 \right)^p \eta(t_{m-1}) \tag{40}$$

where $\Delta t = t_m/m$ is the step size and m is a positive integer. The total integration time is $T = t_M$, M being the total number of integration steps. The sequence of products $\kappa_p T = \kappa_p M \Delta t$, $p = 1, ..., P$, determines the spectrum of the covariance function from (36), and the length scale of the covariance function from (37). Once these parameters are specified, $\kappa_p \Delta t$ in (40) is set to $\kappa_p T/M$ with the value of M chosen large enough so as to maintain the numerical stability of the forward differencing scheme. For the diffusion equation ($P = 1$), the stability condition is roughly $\kappa_1 \Delta t/e^2 < 1/4$ where e is the minimum grid spacing. In this case, the product $\kappa_1 \Delta t$ can be determined directly from the length scale criterion (33), $\kappa_1 \Delta t = L^2/2M$. For a given L, this leads to the requirement that $M > 2(L/e)^2$.

Assuming that the geographical coordinates (λ, ϕ) are continuous functions of a general set of orthogonal curvilinear coordinates (i, j), the Laplacian in (40) can be written in the general form

$$\nabla^2 \eta = \frac{1}{e_1 e_2} \left(\frac{\partial}{\partial i} \left(\frac{e_2}{e_1} \frac{\partial \eta}{\partial i} \right) + \frac{\partial}{\partial j} \left(\frac{e_1}{e_2} \frac{\partial \eta}{\partial j} \right) \right) \tag{41}$$

where $e_1 = e_1(\lambda(i,j), \phi(i,j)) \equiv \partial s_1/\partial i$ and $e_2 = e_2(\lambda(i,j), \phi(i,j)) \equiv \partial s_2/\partial j$ are metric coefficients that define the curvilinear distance elements $(ds_1, ds_2) = (e_1 di, e_2 dj)$ in the (i, j) coordinate system. Equation (41) is self-adjoint for the scalar product $\langle \eta_1, \eta_2 \rangle = \int \eta_1 \eta_2 \, ds_1 ds_2$; i.e., $\langle \nabla^2 \eta_1, \eta_2 \rangle = \langle \eta_1, \nabla^2 \eta_2 \rangle$. The discrete analogue of this scalar product is $\langle \eta_1, \eta_2 \rangle = \eta_1^T \mathbf{W} \eta_2$ where $\mathbf{W}$ is a diagonal matrix of local area elements $\Delta s_1 \Delta s_2$ (and η_1 and η_2 are to be interpreted now as finite-dimensional vectors). Letting $\mathbf{D}$ denote the matrix representation of the discretized Laplacian in (41), and $\mathbf{D}^*$ its corresponding adjoint, then from the self-adjointness property of $\mathbf{D}$ and the general definition of the adjoint operator, $\langle \mathbf{D}\eta_1, \eta_2 \rangle = \langle \eta_1, \mathbf{D}^* \eta_2 \rangle$, it follows that $\mathbf{D} = \mathbf{D}^* = \mathbf{W}^{-1} \mathbf{D}^T \mathbf{W}$.

The discrete operator, $\mathbf{L}_P$, that defines the convolution integral (35), can be represented by an integration of (40) over M steps;

$$\eta(t_M) = \mathbf{L}_P \, \eta(t_0) \tag{42}$$

where

$$\mathbf{L}_P = \left[\mathbf{I} - \sum_{p=1}^{P} \kappa_p \Delta t \, (-\mathbf{D})^p \right]^M . \tag{43}$$

From the self-adjointness property of $\mathbf{D}$, we can factor $\mathbf{L}_P$ as

$$\begin{aligned}
\mathbf{L}_P &= \left[\mathbf{I} - \sum_{p=1}^{P} \kappa_p \Delta t \, (-\mathbf{D})^p \right]^{M/2} \left[\mathbf{I} - \sum_{p=1}^{P} \kappa_p \Delta t \, (-\mathbf{D})^p \right]^{M/2} \\
&= \mathbf{L}_P^{1/2} \, \mathbf{L}_P^{1/2} \\
&= \left[\mathbf{I} - \sum_{p=1}^{P} \kappa_p \Delta t \left(-\mathbf{W}^{-1} \mathbf{D}^T \mathbf{W} \right)^p \right]^{M/2} \left[\mathbf{I} - \sum_{p=1}^{P} \kappa_p \Delta t \, (-\mathbf{D})^p \right]^{M/2} \\
&= \mathbf{W}^{-1} \left[\mathbf{I} - \sum_{p=1}^{P} \kappa_p \Delta t \left(-\mathbf{D}^T \right)^p \right]^{M/2} \cdot \mathbf{W} \left[\mathbf{I} - \sum_{p=1}^{P} \kappa_p \Delta t \, (-\mathbf{D})^p \right]^{M/2}
\end{aligned}$$

$$= \mathbf{W}^{-1}\,\mathbf{L}_P^{T/2}\,\mathbf{W}\,\mathbf{L}_P^{1/2}$$
$$= \mathbf{L}_P^{1/2}\,\mathbf{W}^{-1}\,\mathbf{L}_P^{T/2}\,\mathbf{W}. \tag{44}$$

$\mathbf{L}_P$ is *self-adjoint* with respect to the metric $\mathbf{W}$; i.e., $\mathbf{L}_P$ maps a field from the primal space (e.g., model space) into itself. The action of $\mathbf{L}_P$ thus conserves the physical units of the input field. $\mathbf{L}_P$ can be transformed into a *symmetric* covariance matrix by multiplying it either to the left by $\mathbf{W}$ or to the right by $\mathbf{W}^{-1}$, as evident from the factorization (44). We may define the covariance matrix either way, but its interpretation is somewhat clearer by following the latter. Therefore, without loss of generality, we define the covariance matrix as

$$\mathbf{L}_P\,\mathbf{W}^{-1} = \mathbf{L}_P^{1/2}\,\mathbf{W}^{-1}\,\mathbf{L}_P^{T/2}. \tag{45}$$

The associated covariance operator then corresponds to an integration of the difference equation (40) from an initial condition scaled at each grid point by the inverse of the corresponding local area element. This is consistent with the general definition of a covariance operator, which represents a mapping from the dual space into the primal space and hence does not conserve the physical units of the input field (Tarantola 1987).

A matrix, $\mathbf{\Lambda}$, of normalization coefficients is required to convert (45) into a correlation matrix; i.e., $\mathbf{\Lambda}$ ensures that the standard error deviations effectively used in $\mathbf{B}_u$ (Eq. (14)) are indeed those which are specified through $\mathbf{\Sigma}$. In general, the elements of $\mathbf{\Lambda}$ will be spatially dependent (see section 5b) in which case $\mathbf{\Lambda}$, like $\mathbf{\Sigma}$, must be introduced on either side of the filter covariance matrix in order to maintain symmetry (cf. the scalar normalization in (21)). Thus, the auto-correlation matrix associated with an arbitrary scalar field α has the form

$$\mathbf{C}_\alpha = \mathbf{\Lambda}\,\mathbf{L}_P^{1/2}\,\mathbf{W}^{-1}\,\mathbf{L}_P^{T/2}\,\mathbf{\Lambda} \tag{46}$$
$$= \left(\mathbf{W}^{-1/2}\,\mathbf{L}_P^{T/2}\,\mathbf{\Lambda}\right)^T \left(\mathbf{W}^{-1/2}\,\mathbf{L}_P^{T/2}\,\mathbf{\Lambda}\right)$$
$$= \mathbf{C}_\alpha^{T/2}\,\mathbf{C}_\alpha^{1/2}. \tag{47}$$

The procedure for computing the normalization coefficients is discussed in some detail in the next section.

The presence of $\mathbf{W}$ in the general matrix expression for $\mathbf{C}_\alpha$ illustrates explicitly the grid-dependence of the correlation operator and hence of the scalar product of the background term (J_b) in (1) whose metric depends on $\mathbf{C}_\alpha^{-1}$. This scalar product defines a weighted sum of increments over the model domain and thus naturally should depend on the model grid.

Lorenc (1997) and Parrish *et al.* (1997) discuss a similar procedure for designing a 2D horizontal covariance operator using a 1D recursive filter. The 2D covariance operator is formed from a matrix product of the 1D filter, applied successively along orthogonal grid directions (cf. $\mathbf{L}_P^{1/2}$ in (44)), with its adjoint (cf. $\mathbf{W}^{-1}\mathbf{L}_P^{T/2}\mathbf{W}$ in (44)). The resulting covariance operator, however, does not give an isotropic response as illustrated by Lorenc (1997). This contrasts the Laplacian-based filter ($\mathbf{L}_P$) which is isotropic by construction. (A numerical validation of this result is provided by Fig. 2 (upper panel) of section 5a). Furthermore, the Laplacian can be made invariant in any orthogonal curvilinear coordinate system, a feature that is particularly useful for models with distorted grids, such as global ocean models that have the north pole singularity rotated onto a land point (Madec and Imbard 1996).

4.2 The normalization matrix

As discussed in the previous section, a diagonal normalization matrix $\mathbf{\Lambda}$ is required to convert the covariance matrix $\mathbf{L}_P\mathbf{W}^{-1}$ into a correlation matrix of the form (47).[2] The diagonal elements of $\mathbf{\Lambda}$ are the inverse of the square root of the corresponding diagonal elements of $\mathbf{L}_P\mathbf{W}^{-1}$; i.e., the inverse of the intrinsic standard deviations (square root of the variances) of $\mathbf{L}_P\mathbf{W}^{-1}$. Note that the variances of $\mathbf{L}_P\mathbf{W}^{-1}$ have physical dimensions of inverse length squared, in contrast to the variances of $\mathbf{B}_u$ which have physical dimensions equal to the square of the physical units of the background variables.

For the isotropic filter $\mathbf{L}_P$, applied in a boundary-free domain, $\mathbf{\Lambda}$ is simply a constant multiple of the identity matrix where the constant can be determined analytically from the variance of the covariance functions in (36); $\mathbf{\Lambda} = (1/\sqrt{t})\,\mathbf{I}$ where $t = f(0; \kappa_1 T, ..., \kappa_P T)/4\pi a^2$, the extra factor $4\pi a^2$ coming from the normalization in (35). For the $\mathbf{L}_1$-filter, $t \approx g(0;\gamma)/4\pi a^2 \approx 1/2\pi L^2$ (cf. Eq. (77) in the appendix).

In section 5, we discuss more general implementations of the filter (e.g., applications in bounded domains (section 5b) and extensions to account for anisotropic correlations (sections 5c-d)) for which the true variances are no longer well approximated at all grid-points by a spatially homogeneous constant. To compute the true variances at each grid-point requires a specific algorithm. Two algorithms are proposed below, one of which computes the variances exactly, the other approximately.

Let t_l denote the l-th diagonal element of the matrix $\mathbf{L}_P\mathbf{W}^{-1}$; i.e., the intrinsic variance of the covariance filter at the l-th model grid-point. Since, in practice, $\mathbf{L}_P\mathbf{W}^{-1}$ is available as an operator, the element t_l can be computed by applying $\mathbf{L}_P\mathbf{W}^{-1}$ to the vector $\mathbf{e}_l = (0, ..., 0, 1, 0, ..., 0)^T$, the entry equal to one being located at the l-th grid-point. The value of the filtered field at the l-th grid-point gives the filter variance at that point; i.e., $t_l = \mathbf{e}_l^T\,\mathbf{L}_P\,\mathbf{W}^{-1}\,\mathbf{e}_l$ with $1/\sqrt{t_l}$ defining the l-th diagonal element of $\mathbf{\Lambda}$. This algorithm is computationally expensive as it requires, in general, as many applications of the covariance filter as there are model grid-points. The algorithm is, however, well suited for parallel computers as the filter variance at the different points can each be computed on a separate processor.

Rather than computing the variances exactly using the above algorithm, a cheaper alternative is to use a randomization method (Fisher and Courtier 1995; Andersson *et al.* 2000) to estimate the variances approximately. Consider the transformation $\tilde{\mathbf{v}} = \mathbf{L}_P^{1/2}\mathbf{W}^{-1/2}\mathbf{v}$ where $\mathbf{v}$ is a Gaussian random variable, drawn from a population having zero mean and unit variance; i.e., $E[\mathbf{v}] = 0$ and $E[\mathbf{v}\mathbf{v}^T] = \mathbf{I}$. The transformed random variable verifies $E[\tilde{\mathbf{v}}] = 0$ and $E[\tilde{\mathbf{v}}\tilde{\mathbf{v}}^T] = \mathbf{L}_P^{1/2}\mathbf{W}^{-1}\mathbf{L}_P^{T/2} = \mathbf{L}_P\mathbf{W}^{-1}$. Therefore, from an ensemble of Q random vectors $\tilde{\mathbf{v}}_q$, $q = 1, ..., Q$, the variances of the filter covariance matrix can be estimated from the diagonal elements of

$$\mathbf{L}_P\mathbf{W}^{-1} \approx \frac{1}{Q}\sum_{q=1}^{Q}\tilde{\mathbf{v}}_q\tilde{\mathbf{v}}_q^T = \frac{1}{Q}\sum_{q=1}^{Q}\left(\mathbf{L}_P^{1/2}\mathbf{W}^{-1/2}\mathbf{v}_q\right)\left(\mathbf{L}_P^{1/2}\mathbf{W}^{-1/2}\mathbf{v}_q\right)^T. \tag{48}$$

The variances converge to the true variances when Q is large. For Gaussian statistics, it can

[2] $\mathbf{L}_P$ is a scale-dependent filter, with the property that it conserves the spatial mean of the input field. For this reason, the diagonal elements of the covariance matrix $\mathbf{L}_P\mathbf{W}^{-1}$ will not, in general, be equal to one.

be shown (Barlow 1989, pg. 89) that the standard error in the estimated standard deviations (inverse of the normalization factors) is $1/\sqrt{2Q}$ (e.g., 10% for an ensemble of 50).

4.3 A three-dimensional correlation model

In a multi-level ocean or atmosphere model, the univariate correlation matrix $\mathbf{C}_\alpha$ must account for error correlations in the vertical as well as the horizontal. Three-dimensional correlation models are often assumed to be separable in that they are constructed as a product of a horizontal and vertical correlation function, which depend only on horizontal and vertical separation respectively (Daley 1991). Specifically, the function formed from the product of the covariance functions (36) and (39), normalized by their respective variances, corresponds to a special class of separable, 3D correlation functions, which may be modelled by a combined integration of the 1D and 2D GDEs.

We assume that the vertical direction is normal to the geopotential surfaces on the sphere. Furthermore, we assume that the vertical distance z is a continuous function of a generalized vertical coordinate k, orthogonal to the curvilinear coordinate surfaces (i, j) on the sphere, so that the vertical second-derivative operator in (38) can be expressed as

$$\frac{\partial^2 \eta}{\partial z^2} = \frac{1}{e_3}\left(\frac{\partial}{\partial k}\left(\frac{1}{e_3}\frac{\partial \eta}{\partial k}\right)\right) \tag{49}$$

where $e_3 = \partial s_3/\partial k = \partial z/\partial k$ is the metric coefficient defining the vertical distance element $ds_3 = e_3 dk$ in this coordinate system. In what follows, we will use the subscript/superscript v to distinguish the parameters and operators associated with the vertical correlation model from those associated with the horizontal correlation model of section 4a, which for notational consistency will hereafter be denoted with a subscript/superscript h. Let $\mathbf{D}^v$ be the matrix representation of a discretized version of (49). The matrix operator, $\mathbf{L}_R^v$, that integrates the 1D (vertical) GDE from $t = 0$ to $t = T_v = M_v \Delta t_v$ can thus be expressed as

$$\mathbf{L}_R^v = \left[\mathbf{I} - \sum_{r=1}^{R} \kappa_r^v \Delta t_v \, (-\mathbf{D}^v)^r\right]^{M_v}. \tag{50}$$

The symmetric vertical covariance matrix associated with (50) is then $\mathbf{L}_R^v \mathbf{W}_v^{-1}$ where $\mathbf{W}_v$ is a diagonal matrix of vertical grid elements Δs_3 (cf. Eq. (45)).

The convolution of an arbitrary scalar input field with the combined horizontal-vertical covariance function may be achieved by a two-step procedure: (i) for each model level k, integrate the 2D GDE (34) from $t = 0$ to $t = T_h$ with the field to be convolved taken as initial condition; (ii) for each horizontal grid point (i, j), integrate the vertical GDE (38) from $t = 0$ to $t = T_v$ using the result from (i) as initial condition. Noting that the horizontal and vertical operators commute, we can express this 3D covariance operator as

$$\begin{aligned} \mathbf{L}_R^v \mathbf{W}_v^{-1} \mathbf{L}_P^h \mathbf{W}_h^{-1} &= \mathbf{L}_R^{v\,1/2} \mathbf{W}_v^{-1} \mathbf{L}_R^{v\,T/2} \mathbf{L}_P^{h\,1/2} \mathbf{W}_h^{-1} \mathbf{L}_P^{h\,T/2} \\ &= \mathbf{L}_R^{v\,1/2} \mathbf{L}_P^{h\,1/2} \mathbf{W}^{-1} \mathbf{L}_P^{h\,T/2} \mathbf{L}_R^{v\,T/2} \end{aligned} \tag{51}$$

where $\mathbf{W} = \mathbf{W}_h \mathbf{W}_v$ is a diagonal matrix of volume elements $\Delta s_1 \Delta s_2 \Delta s_3$. Thus, the 3D correlation matrix has the same form as (46) with $\mathbf{L}_P$ replaced by the product $\mathbf{L}_R^v \mathbf{L}_P^h$. Note

that the elements of the normalization matrix $\mathbf{\Lambda}$ must now have physical dimensions of volume in order to render the correlation matrix dimensionally consistent.

In summary, the square-root factors of the three-dimensional correlation matrix take the form

$$\mathbf{C}_\alpha^{T/2} = \mathbf{\Lambda}\,\mathbf{L}_R^{v\,1/2}\,\mathbf{L}_P^{h\,1/2}\,\mathbf{W}^{-1/2}, \tag{52}$$

$$\mathbf{C}_\alpha^{1/2} = \mathbf{W}^{-1/2}\,\mathbf{L}_P^{h\,T/2}\,\mathbf{L}_R^{v\,T/2}\,\mathbf{\Lambda}. \tag{53}$$

In words, the operator $\mathbf{C}_\alpha^{T/2}$, which in 3D-Var is needed for transforming the increment from minimization space into model space (Eq. (15)), involves the following sequence of operations:

1. multiply each element of the input vector by the inverse of the square root of its associated volume element;

2. perform $M_h/2$ iterations with the horizontal filter $\mathbf{L}_P^h$;

3. perform $M_v/2$ iterations with the vertical filter $\mathbf{L}_R^v$;

4. multiply each element of the filtered vector by its corresponding normalization factor.

Likewise, the "adjoint" operator $\mathbf{C}_\alpha^{1/2}$, which in 3D-Var is needed for defining $\mathbf{B}^{1/2}$ in the gradient transformation (Eq. (9)), involves the sequence of operations 1 through 4 applied in reverse order, with the horizontal and vertical filters in operations 2 and 3 replaced by their respective adjoints.

5 Illustrations

The univariate $\mathbf{C}_\alpha$-operator described above has been incorporated within a variational assimilation system (Weaver and Vialard 1999) being developed for the OPA OGCM (Madec *et al.* 1999). In this section, this system will be used to illustrate some of the basic features of the correlation model, including generalizations that account for anisotropic correlations. In all examples, the horizontal and/or vertical $\mathbf{L}_1$-filter is used.

The discrete differential operators used in the correlation model are based on similar operators used to solve the equations of the ocean model. These operators are formulated in orthogonal curvilinear coordinates and discretized on an Arakawa C-grid using second-order accurate, centred finite differences. The examples presented here have been conducted with a tropical Pacific version of the model. The configuration extends from 120°E to 70°W and from 30°S to 30°N. The entire model domain is enclosed by solid boundaries, with realistic coastlines imposed along the eastern and western boundaries, and artificial walls along the southern and northern boundaries. The zonal resolution is 1° and the meridional resolution varies smoothly from 0.5° at the equator to 2° at the northern and southern boundaries. The basic configuration has been adapted from Vialard and Delecluse (1998). The interested reader should refer to that paper and references therein for a more complete description of the model.

5.1 Isotropic correlations

The auto-correlations implied by the $\mathbf{L}_1^h$-filter are illustrated in Fig. 2 for a point (l) located on the equator at 160° W (i.e., in the central Pacific, far away from boundaries). The auto-correlation field has been obtained by applying (46) with $P = 1$ to the unit vector $\mathbf{e}_l$ defined above. Assuming that the variance of background errors is spatially homogeneous (i.e., assuming $\mathbf{\Sigma} = \sigma_b \mathbf{I}$ with σ_b constant), then Fig. 2 corresponds to the analysis increment, up to a proportionality constant, that would be obtained from a single observation. This is easy to see from (12) noting that for a single observation, $\mathbf{w}$ becomes a scalar and $\mathbf{H}^T$ a column vector equal to $\mathbf{e}_l$, where the value of one in the l-th entry of $\mathbf{e}_l$ corresponds to the observation location. The correlation length scale is $L = 4°$ in the upper panel and $L = 1°$ in the lower panel. The upper panel is a numerical validation of the isotropic response of the filter for a value of L large relative to the grid resolution. The lower panel illustrates how the degree of isotropy is degraded when L is comparable to the grid resolution. In this latter example, L is equal to the zonal resolution and approximately twice the meridional resolution. Relative to the correlations computed directly from the Gaussian function, those computed in the lower panel using the diffusion model are everywhere underestimated, particularly in the zonal direction where the resolution is coarsest.

5.2 Boundary considerations

In a bounded domain the Laplacian must be supplemented with boundary conditions. For the C-grid representation of the discrete $\mathbf{L}_1^h$-filter, this requires specifying the value of the filtered variable's first derivatives normal to boundary points. For the higher-order $\mathbf{L}_P^h$-filters ($P > 1$), additional boundary conditions are required; in general, the first, third, ..., and $(2P-1)^{\text{th}}$ derivatives normal to the boundary must be specified. These boundary conditions can be viewed as free parameters of the correlation model.

Figure 3 illustrates the auto-correlations generated by the $\mathbf{L}_1^h$-filter at three points adjacent to the coastlines, assuming vanishing first derivatives normal to boundary points (in the ocean model this boundary condition is analogous to the "no-flux" condition on tracer fields or the "free-slip" condition on the tangential component of the velocity field). The parameters of the correlation model correspond to those used for the $\mathbf{L}_1^h$-filter in the upper panel of Fig. 2. The spatial structure of the correlations near boundary regions is determined primarily by the geometry of the coastlines; the exact nature of the boundary condition has a much weaker influence. For example, the correlations in Fig. 3 are relatively robust to changes in boundary condition from a no-flux to a dissipative flux form.

The primary effect of changing the nature of the boundary condition is to modify the intrinsic variances (t_l) of the filter, and hence the normalization factors ($1/\sqrt{t_l}$), in the vicinity of the boundary. Figure 4 shows the difference between the true normalization factors, $1/\sqrt{t_l}$, computed using the exact procedure described in section 4b, and the constant normalization factor, $\sqrt{2\pi}L$, estimated from the analytical solution of the diffusion equation in a boundary-free domain with spatially homogeneous diffusion coefficient. The values of the normalization factors match the analytical estimate very closely except in regions closely confined to the boundaries where they are smaller by up to a factor of 2.5; i.e., the filter variances are greater

by up to a factor of 6. However, changing from a no-flux boundary condition to a stronger, dissipative flux condition results in a reduction of the filter variances, as one might expect intuitively. In terms of reducing the cost of the filter, the no-flux condition is preferable as it imposes a less stringent constraint on the size of the time step needed to maintain the numerical stability of the forward-differencing scheme.

5.3 Anisotropic correlations: coordinate stretching

There is substantial observational evidence to suggest that correlation scales near the equator are longer in the zonal direction than in the meridional direction (e.g., Meyers *et al.* 1988). Such anisotropy can be accounted for in the diffusion equation by stretching the zonal coordinates in the Laplacian operator, as discussed in Derber and Rosati (1989) and Behringer *et al.* (1998). With reference to the GDE in (34), this can be achieved by replacing ∇^2 with $\nabla \cdot \mathcal{R}\nabla$ where $\mathcal{R}$ is a diagonal, second-rank tensor of nondimensional "diffusion" coefficients r_i and r_j,

$$\mathcal{R} = \begin{pmatrix} r_i & 0 \\ 0 & r_j \end{pmatrix}, \tag{54}$$

∇ is the gradient operator

$$\nabla = \left(\frac{1}{e_1}\frac{\partial}{\partial i}, \frac{1}{e_2}\frac{\partial}{\partial j} \right)^T, \tag{55}$$

and $\nabla \cdot \mathbf{c}$ is the divergence operator

$$\nabla \cdot \mathbf{c} = \frac{1}{e_1 e_2}\left(\frac{\partial(e_2 c_1)}{\partial i} + \frac{\partial(e_1 c_2)}{\partial j} \right), \tag{56}$$

with $\mathbf{c} = (c_1, c_2)^T$. The tensor elements r_i and r_j act, respectively, on the i and j components of the gradient. For the isotropic case, $r_i = r_j \equiv 1$. Note that in the ocean model used in the examples here, there is a one-to-one correspondence between the computational coordinates i and j and the geographical coordinates λ and ϕ; i.e., $\lambda = \lambda(i)$ and $\phi = \phi(j)$.

Figure 5 illustrates the auto-correlations generated at three different latitudes by an anistropic version of the $\mathbf{L}_1^h$-filter (cf. Fig. 1 in Behringer *et al.* (1998)). The zonal and meridional length scales have been defined to be 8° and 2°, respectively, at the equator and 4° in both directions at $|\phi| \geq 20°$ N/S, with a linear transition between these values over $20° \text{ S} < \phi < 20° \text{ N}$. This has been achieved by setting $\kappa_1 T$ to $L^2/2$ (as in Fig. 2, upper panel), and varying the tensor elements linearly from $r_i = 4$ and $r_j = 1/4$ at the equator to $r_i = r_j = 1$ at $|\phi| \geq 20°$ N/S. More generally, the r_i and r_j can be made inhomogeneous functions of both spatial coordinates (i, j) in order to allow for fully global variations in the length scale.

5.4 Three-dimensional flow-dependent correlations: coordinate rotation

The surface over which the horizontal correlations are effectively computed depends on the choice of vertical coordinate in the correlation model. In section 4c, the direction of the vertical

coordinate was assumed to be normal to the geopotential (z-)surfaces, implying that the horizontal correlation function (36) is isotropic along these surfaces. Other choices are of course possible. For example, in the ECMWF 3D-Var (Courtier *et al.* 1998), the horizontal surfaces are defined with respect to the hybrid η-coordinate (a pressure-based, terrain-following vertical coordinate) used in the ECMWF atmospheric model. In general, the vertical coordinate of the correlation model can be independent of the vertical coordinate used in the actual atmosphere or ocean model. For example, Daley and Barker (1999) suggest the use of potential temperature (isentropic) coordinates as a means of introducing more flow dependency into the background error correlation model of an atmospheric assimilation system. The analogy in an ocean assimilation system is to define the background error correlations with respect to a potential density (isopycnal) coordinate (e.g., as proposed by Gavart and DeMey (1997) in the context of Empirical Orthogonal Functions). In this section, we illustrate that a transformation from, for example, geopotential to isopycnal coordinates is a natural extension of the correlation model based on the GDE. For simplicity, we consider only the $\mathbf{L}_1$-filter, although in principle the analysis may be extended to accommodate the higher-order filters.

In the 3D $\mathbf{L}_1$-filter, the length scale of the horizontal Gaussian function is determined from $L = \sqrt{2\kappa_1^h T_h}$ and the depth scale of the vertical Gaussian function from $D = \sqrt{2\kappa_1^v T_v}$. In other words, for a given correlation scale, we are free to adjust both the diffusion coefficient and the integration time, providing their product remains constant. (Note that for the higher-order filters, the product also controls the shape of the correlation function). We can exploit this fact to derive an alternative form of the $\mathbf{L}_1$-filter, which will simplify the transformation to isopycnal coordinates.

First, for given scales D and L, let us define identical time integration parameters for the vertical and horizontal $\mathbf{L}_1$-filters; i.e., we set $M_v = M_h$ and $\Delta t_v = \Delta t_h$ (and hence $T_v = T_h$), and redefine the vertical diffusion coefficient to be $\widetilde{\kappa}_1^v = \kappa_1^h D^2/L^2$ so that $D = \sqrt{2\widetilde{\kappa}_1^v T_h}$. From (43) and (50), we can express the 3D $\mathbf{L}_1$-filter as

$$\begin{aligned}
\mathbf{L}_1 &= \mathbf{L}_1^v \mathbf{L}_1^h \\
&= \left[\mathbf{I} + \kappa_1^v \Delta t_v \, \mathbf{D}^v\right]^{M_v} \left[\mathbf{I} + \kappa_1^h \Delta t_h \, \mathbf{D}^h\right]^{M_h} \qquad (57) \\
&= \left[\left(\mathbf{I} + \widetilde{\kappa}_1^v \Delta t_h \, \mathbf{D}^v\right)\left(\mathbf{I} + \kappa_1^h \Delta t_h \, \mathbf{D}^h\right)\right]^{M_h} \\
&\approx \left[\mathbf{I} + \kappa_1^h \Delta t_h \, \mathbf{D}^h + \widetilde{\kappa}_1^v \Delta t_h \, \mathbf{D}^v\right]^{M_h} \\
&= \left[\mathbf{I} + \kappa_1^h \Delta t_h \left(\mathbf{D}^h + \frac{\widetilde{\kappa}_1^v}{\kappa_1^h} \mathbf{D}^v\right)\right]^{M_h} \qquad (58)
\end{aligned}$$

where we have neglected the term of $\mathrm{O}\left(\Delta t_h^2\right)$. (This term can be made small by choosing M_h sufficiently large, which we are free to do but at the expense of increasing the cost of the filter).

Equation (58) is the matrix representation of a 3D Laplacian diffusion operator acting on $0 \leq t \leq T_h$. The 3D correlation function associated with this operator is horizontally/vertically anisotropic by virtue of the relative scaling factor $\widetilde{\kappa}_1^v/\kappa_1^h = D^2/L^2$ between the vertical and horizontal second-derivative operators, where $D \ll L$ for meteorological and oceanographic scales of interest. The relevance of this alternative form of the $\mathbf{L}_1$-filter is that we are now in a position to exploit a classical transformation from geopotential to isopycnal coordinates

(Redi 1982), as commonly employed in z-coordinate OGCMs for parameterizing lateral mixing along isopycnal surfaces (e.g., Pacanowski 1996; Madec *et al.* 1999). First, as in the previous section, we can express the 3D Laplacian in a more general tensorial form, $\nabla \cdot \mathcal{R} \nabla$, where

$$\mathcal{R} = \begin{pmatrix} r_i & 0 & 0 \\ 0 & r_j & 0 \\ 0 & 0 & r_k \end{pmatrix}, \tag{59}$$

$$\nabla = \left(\frac{1}{e_1}\frac{\partial}{\partial i}, \frac{1}{e_2}\frac{\partial}{\partial j}, \frac{1}{e_3}\frac{\partial}{\partial k} \right)^T \tag{60}$$

and

$$\nabla \cdot \mathbf{c} = \frac{1}{e_1 e_2}\left(\frac{\partial (e_2 c_1)}{\partial i} + \frac{\partial (e_1 c_2)}{\partial j} \right) + \frac{1}{e_3}\frac{\partial c_3}{\partial k} \tag{61}$$

with $\mathbf{c} = (c_1, c_2, c_3)^T$. The tensor elements r_i, r_j and r_k can accommodate coordinate stretching in the horizontal and vertical planes and can be varied with each grid point (i, j, k) in order to permit geographical variations of the length and depth scales. Defining the length scales to be a function of depth or the depth scale to be a function of horizontal position is one way of introducing non-separability into the correlation model. Equation (58) can be seen as a special case with $r_i = r_j = 1$ and $r_k = \tilde{\kappa}_1^v / \kappa_1^h$.

Now, we make the assumption that $\mathcal{R}$ is diagonal in an isopycnal coordinate system, rather than in a z-coordinate system as considered above. This requires a simple reinterpretation of the tensor elements in $\mathcal{R}$; r_i and r_j are assumed to be referenced to the "horizontal" plane defined by a constant isopycnal surface, while r_k is assumed to be referenced to the "vertical" plane normal to this surface. The representation of the isopycnal $\mathcal{R}$ in z-coordinates (our computational coordinate system) can then be obtained by a coordinate rotation, as described in Redi (1982) for the horizontally isotropic case $r_i = r_j$ (see also Griffies *et al.* (1998)). The extension to the horizontally anisotropic case $r_i \neq r_j$ is straightforward and treated below.

At the beginning of an assimilation cycle, we assume the availability of a statically stable background potential density surface, ρ^b, with respect to which we can define the "horizontal" surfaces of the correlation model. In a z-coordinate OGCM, ρ^b is computed diagnostically from the equation of state, given the background potential temperature field $T_\theta^b(i, j, k)$ and background salinity field $S^b(i, j, k)$ (these being the prognostic density variables in the OGCM), and the local ocean depth $z(k)$ as input (i.e., $\rho^b = \rho^b(T_\theta^b(i, j, k), S^b(i, j, k), z(k))$). Implicitly, in defining ρ^b as our coordinate, we are assuming that the background errors are isotropic, or weakly anisotropic in the sense of the diagonal tensor (59) with $r_i \neq r_j$, along the ρ^b surfaces. Such an hypothesis can only be verified objectively by computing the actual statistics of the observation-minus-background field (e.g., as in Hollingsworth and Lönnberg (1986)). Nevertheless, the use of ρ^b as a coordinate has a certain appeal physically, which makes it an attractive possibility providing the background density state is reasonably accurate. If not, then there may be little virtue in using a ρ^b-coordinate. In such a case, a z-coordinate may be a wiser choice.

The diagonal tensor $\mathcal{R}$ in the isopycnal coordinate system transforms as $\widetilde{\mathcal{R}} = \mathbf{S}\mathcal{R}\mathbf{S}^T$ in the z-coordinate system, where $\mathbf{S}$ is a rotation matrix. The elements of $\mathbf{S}$ depend on the components and magnitude of the isopycnal slope vector $(-a_1, -a_2, 0)$ where

$$a_1 = \frac{e_3}{e_1}\left(\frac{\partial\rho^b}{\partial i}\right)\left(\frac{\partial\rho^b}{\partial k}\right)^{-1} \qquad a_2 = \frac{e_3}{e_2}\left(\frac{\partial\rho^b}{\partial j}\right)\left(\frac{\partial\rho^b}{\partial k}\right)^{-1}. \tag{62}$$

(For details, we refer the reader to Redi (1982) or Griffies *et al.* (1998)). Since the slopes a_1 and a_2 are generally less than 10^{-2} in the ocean and since $r_k \sim \mathrm{O}(D^2/L^2) \ll 1$ while $r_i, r_j \sim \mathrm{O}(1)$, the full expression for $\widetilde{\mathcal{R}}$ can be simplified appreciably. In particular, for the horizontally anisotropic case ($r_i \neq r_j$), the expression for the isopycnal tensor approximates as

$$\widetilde{\mathcal{R}} \approx \begin{pmatrix} \dfrac{r_i a_2^2 + r_j a_1^2}{a_1^2 + a_2^2} & \dfrac{a_1 a_2}{a_1^2 + a_2^2}(r_j - r_i) & -r_j a_1 \\ \dfrac{a_1 a_2}{a_1^2 + a_2^2}(r_j - r_i) & \dfrac{r_i a_1^2 + r_j a_2^2}{a_1^2 + a_2^2} & -r_j a_2 \\ -r_j a_1 & -r_j a_2 & r_k + r_j\left(a_1^2 + a_2^2\right) \end{pmatrix}. \tag{63}$$

For the horizontally isotropic (HI) case ($r_i = r_j$), (63) reduces to the familiar "small slope" approximation to $\widetilde{\mathcal{R}}$ commonly used in an isopycnal mixing parameterization (cf. Eq. (3) in Griffies *et al.* (1998));

$$\widetilde{\mathcal{R}}_{\mathrm{HI}} \approx \begin{pmatrix} r_i & 0 & -r_i a_1 \\ 0 & r_i & -r_i a_2 \\ -r_i a_1 & -r_i a_2 & r_k + r_i\left(a_1^2 + a_2^2\right) \end{pmatrix}. \tag{64}$$

Figure 6 shows an example of the auto-correlation fields generated by a z-coordinate and a ρ^b-coordinate version of the 3D $\mathbf{L}_1$-filter. The upper panel shows a meridional-vertical section of a typical background temperature field through 110° W in the eastern Pacific basin. The corresponding potential density profile is very similar (i.e., the isopycnals are only weakly modified by salinity). The background field actually corresponds to the analysis at the end of a one-month 4D-Var analysis of *in situ* temperature observations, as described in Weaver and Vialard (1999). The longitude 110° W coincides with a particular section of the Tropical Atmosphere-Ocean array of moored buoys, the temperature measurements from which were assimilated in the 4D-Var experiment. In other words, the background profile at 110° W is already quite close to observations so there may be good merit in building this information into the background error correlation model.

The middle panel in Fig. 6 shows the auto-correlation field generated by the separable, z-coordinate, 3D $\mathbf{L}_1$-filter (Eq. (57)) for a point located at 8° N and at a depth of 55 m. The depth scale and length scale have been set to constant values of 20 m and 4°, respectively. As the separable filter acts along geopotential surfaces, the resulting auto-correlation field is insensitive to the background profile, in particular to the strong meridional gradient in the temperature field between 4° N and 9° N which is associated with the North Equatorial Counter Current. The bottom panel shows the auto-correlation field obtained using the alternative form of the 3D $\mathbf{L}_1$-filter (Eq. (58)) in conjunction with the isopycnal tensor (64). As expected, the

correlations are strongest along isopycnal surfaces and fall off rapidly across isopycnal surfaces. In a 3D analysis, these correlations would essentially define the surfaces over which to smooth an observation located at the correlation point. Thus, a correlation model based on the lower panel would clearly be less destructive to the background density profile than a correlation model based on the middle panel.

The correlations in the middle panel were computed using 10 iterations of the vertical filter ($\mathbf{L}_1^v$) and 250 iterations of the horizontal filter ($\mathbf{L}_1^h$). The correlations in the bottom panel were computed using 8000 iterations of the 3D $\mathbf{L}_1$-filter, which was roughly the number required for numerical stability. This represents a substantial increase (roughly a factor of 30) in cost relative to the separable filter, suggesting that alternative (e.g., semi-implicit) time discretization schemes should be explored as a possibility for reducing the cost of the isopycnal filter. Furthermore, as the isopycnal correlation model is dependent on the background state, in a cycling assimilation system the normalization factors would need to be recomputed at the start of each assimilation cycle, which could only be done feasibly using, for example, the randomization method described in section 4b. In a z-coordinate correlation model, the normalization factors must be computed only at the beginning of the first assimilation cycle (providing the correlation parameters remain time-invariant). In Fig. 6, the normalization factors were computed using the randomization method with a 100-member ensemble, which explains why the variance at the correlation point is not exactly equal to one.

6 Concluding remarks

A practical algorithm for modelling a large class of 2D and 3D univariate correlations functions on the sphere has been described. The theoretical basis of the algorithm lies in considering the solution of a generalized diffusion equation formed by replacing the Laplacian operator in the classical diffusion equation by a weighted linear combination of powers of the Laplacian operator, the weights being analogous to diffusion coefficients. The integral solution of this equation can be interpreted as a covariance operator on the sphere; the kernel of the covariance operator is an isotropic covariance function and has an explicit representation in terms of the zonal spherical harmonics. The fundamental parameters of the correlation model are the sequence of weighting coefficients multiplied by the total integration time of the generalized diffusion equation. The practical implementation of the algorithm is iterative and forms a class of Laplacian-based grid-point filters, one example of which is the basis of the well-known Derber and Rosati scheme (Derber and Rosati 1989).

The most important features of the algorithm are summarized below.

- The algorithm is easily generalized to account for complex boundary domains and hence is particularly well suited for modelling correlation functions in ocean data assimilation applications.

- The shape of the correlation function can be controlled by adjusting the relative weights of the different terms in the generalized diffusion equation. In general, the more weight given to the higher order Laplacian terms, the more oscillatory the correlation function. For

the simplest case in which a zero-weight is given to all but the first-order Laplacian term, the correlation function is well approximated by a Gaussian (Derber and Rosati 1989).

- The algorithm can be used to represent a class of 1D correlation functions in the vertical as well as 2D correlation functions over the sphere. This provides the basis of a general 3D correlation operator on the sphere.
- Geographical variations in the length scale and shape of the correlation function can be accounted for by defining the diffusion coefficients to be a function of the model's spatial coordinates.
- The correlation functions can be made anisotropic by stretching and/or rotating the computational coordinate system via a symmetric tensor, as commonly applied in classical 3D diffusion models. In an ocean model, the rotation of the "horizontal" correlation surfaces from geopotential to isopycnal coordinates was one particularly attractive possibility illustrated in this paper.
- The algorithm can be implemented efficiently in most cases. The numerical cost of the algorithm is largely determined by the minimum number of iterations needed to maintain the filter stability, which in turn depends on the specific parameter choices of the correlation model.

Acknowledgements

This work was initiated at LODYC in Paris. Much of the practical implementation was carried out by the first author during a year spent at ECMWF, which was organized as part of a collaboration between the French MERCATOR project and the seasonal forecasting project at ECMWF. In particular, he would like to thank David Anderson, Jean-Claude André, Philippe Courtier, Pascale Delecluse and Anthony Hollingsworth for making these arrangements possible. David Anderson and Jérôme Vialard provided many useful suggestions for improving an early draft of the manuscript. We are also grateful to Andrea Piacentini for his insightful remarks on some of the numerical aspects of the algorithm.

Appendix: Matching the solution of the heat equation on the sphere to a Gaussian covariance operator

Consider the function

$$g(\theta;\gamma) = \sum_{n=0}^{\infty} g_n P_n^0(\cos\theta) = \sum_{n=0}^{\infty} \sqrt{2n+1}\,\frac{I_{n+\frac{1}{2}}(\gamma)}{I_{\frac{1}{2}}(\gamma)}\,P_n^0(\cos\theta) \tag{65}$$

where

$$\gamma = \frac{a^2}{L^2}, \tag{66}$$

L being a length scale to be interpreted shortly. $I_{n+\frac{1}{2}}(\gamma)$ is the modified Bessel function of fractional order $n+\frac{1}{2}$ and argument γ. Equation (65) is a covariance function by virtue of the positivity of the $I_{n+\frac{1}{2}}(\gamma)$ for $\gamma > 0$. Using the expansion (Abramowitz and Stegun 1964, Formula 10.2.36)

$$e^{\gamma\cos\theta} = \sum_{n=0}^{\infty} \sqrt{2n+1}\sqrt{\frac{\pi}{2\gamma}} I_{n+\frac{1}{2}}(\gamma) P_n^0(\cos\theta), \tag{67}$$

(65) can be written in the alternative form

$$g(\theta;\gamma) = \sqrt{\frac{2\gamma}{\pi}} \frac{1}{I_{\frac{1}{2}}(\gamma)} e^{\gamma\cos\theta} = \frac{\gamma}{\sinh\gamma} e^{\gamma\cos\theta}. \tag{68}$$

The length scale of $g(\theta,\gamma)$ may be defined following Daley (1991);

$$L^2 = -2\frac{g(0;\gamma)}{\nabla^2 g(0;\gamma)}. \tag{69}$$

The denominator in (69) can be evaluated by applying the Laplacian in spherical coordinates,

$$\nabla^2 = \frac{1}{a^2\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial}{\partial\phi}\right) + \frac{1}{a^2\cos^2\phi}\frac{\partial^2}{\partial\lambda^2}, \tag{70}$$

to (68) with θ, the angular separation between points (λ,ϕ) and (λ',ϕ'), defined according to the great circle distance formula

$$\cos\theta = \cos\phi\,\cos\phi'\cos(\lambda-\lambda') + \sin\phi\,\sin\phi'. \tag{71}$$

This leads to

$$L^2 = \frac{a^2}{\gamma}; \tag{72}$$

thus the interpretation of L as the length scale of the covariance function represented by $g(\theta,\gamma)$. In terms of chordal distance r (Eq. (22)), (68) can be written as

$$g(r(\theta);\gamma) = \sqrt{\frac{2\gamma}{\pi}} \frac{e^{\gamma}}{I_{\frac{1}{2}}(\gamma)} e^{-r^2/2L^2} = \frac{\gamma e^{\gamma}}{\sinh\gamma} e^{-r^2/2L^2}. \tag{73}$$

Hartman and Watson (1974) remark that (30) may be closely matched to (65) by equating their coefficients on the P_1^0 Legendre polynomial,

$$e^{-2\kappa T/a^2} = \frac{I_{\frac{3}{2}}(\gamma)}{I_{\frac{1}{2}}(\gamma)} = \frac{\cosh\gamma}{\sinh\gamma} - \frac{1}{\gamma}, \tag{74}$$

the coefficients for $n=0$ being already equal to one. For a length scale small compared to the radius of the earth ($\gamma \gg 1$), a first order development of the above equality leads to

$$\kappa T \approx \frac{L^2}{2} \tag{75}$$

as in the 1D example. Furthermore, from (73) the variance at any point is

$$g(0;\gamma) = \frac{\gamma e^{\gamma}}{\sinh \gamma} \approx 2\gamma. \tag{76}$$

The covariance operator in (29) can thus be approximated by

$$\eta(\lambda, \phi, T) \approx \frac{1}{2\pi L^2} \int_{\Sigma'} e^{-r^2/2L^2}\, \eta(\lambda', \phi', 0)\, d\Sigma', \tag{77}$$

with L^2 given by (75).

The g_n in (65) can be evaluated from the modified spherical functions of the first kind using, for example, the IMSL library. Figure 7, upper panel, depicts the variance power spectrum of the correlation functions $f(\theta, \kappa T)/f(0, \kappa T)$ and $g(\theta, \gamma)/g(0, \gamma)$, as well as their difference. The lower panel depicts their grid-point values for a length scale of 500 km. A truncation at wavenumber 106 has been used. The agreement is, as expected, excellent, particularly for the large scales. The length scale of the truncated function $g(\theta, \gamma)$ is 500.00 km while it is 500.26 km for $f(\theta, \kappa T)$. The maximum difference in grid-point space is 10^{-4}.

References

Abramowitz, M. and Stegun, I., 1964: *Handbook of mathematical functions.* Dover publications, Inc., New York.

Andersson, E., Fisher, M., Munro, R. and McNally, A., 2000: Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme and the explanation of a case of poor convergence, *Q. J. R. Meteorol. Soc.*, **126**, 1455–1472.

Arfken, G., 1966: *Mathematical Methods for Physicists.* Academic Press Inc.

Barlow, R. J., 1989: *Statistics: a guide to the use of statistical methods in the physical sciences.* John Wiley & Son Ltd.

Behringer, D., Ji, M. and Leetma, A., 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Mon. Wea. Rev.*, **126**, 1013–1021.

Bell, M. J., Forbes, R. M. and Hines, A., 2000: Assessment of the FOAM global data assimilation system for real-time ocean forecasting. *J. Mar. Sys.*, **25**, 1–22.

Bennett, A. F., Chua, B. S. and Leslie, L. M., 1997: Generalized inversion of a global numerical weather prediction model, II: Analysis and implementation. *Meteorol. Atmos. Phys.*, **62**, 129–140.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M., and Lamich, D., 1997: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, 2913–2926.

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.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367–1388.

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

Daley, R. and Barker, E., 1999: *The NAVDAS Source Book.* Naval Research Laboratory, Monterey California.

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.

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.

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

Gauthier, P., Charette, C., Fillion, L., Koclas, P. and Laroche, S., 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: the global analysis. *Atmosphere-Ocean*, **37**, 103–156.

Gavart, M. and DeMey, P., 1997: Isopycnal EOFs in the Azores Current Region: a statistical tool for dynamical analysis and data assimilation. *J. Phys. Oceanogr.*, **27**, 2146–2157.

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.

Hartman, P. and Watson, G. S., 1974: "Normal" distribution functions on spheres and the modified Bessel functions. *Annals of Probability*, **2**, 593–607.

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**, 111–136.

Ide, K., Courtier, P., Ghil, M. and Lorenc, A. C., 1997: Unified notation for data assimilation: operational, sequential and variational. *J. Met. Soc. Japan*, **75**, 181–189.

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., 1988: Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.*, **114**, 205–240.

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

Lorenc, A. C., 1997: Development of an operational variational assimilation scheme. *J. Met. Soc. Japan*, **75**, 339–346.

Madec, G. and Imbard, M., 1996: A global ocean mesh to overcome the North Pole singularity. *Clim. Dyn.*, **12**, 381–388.

Madec, G., Delecluse, P., Imbard, M. and Levy, C., 1999: OPA, release 8.1, Ocean General Circulation Model reference manual. Internal report, LODYC/IPSL, France.

Meyers, G., Phillips, H., Smith, N. and Sprintall, J., 1991: Space and time scales for optimal interpolation of temperature- tropical Pacific Ocean. *Prog. Oceanogr.*, **28**, 189–218.

Pacanowski, R. C., 1996: MOM 2 documentation, user's guide and reference manual. GFDL Ocean Tech. Rep. 3.1, Geophysical Fluid Dynamics Laboratory/NOAA.

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

Parrish, D. F., Derber, J. C., Purser, R. J., Wu, W.-S. and Pu, Z.-X., 1997: The NCEP global analysis system: Recent improvements and future plans. *J. Met. Soc. Japan*, **75**, 359–365.

Redi, H. R., 1982: Oceanic isopycnal mixing by coordinate rotation. *J. Phys. Oceanogr.*, **12**, 1154–1158.

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.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A., 2000: The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126**, 1143–1170.

Tarantola, A., 1987: *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier.

Vialard, J. and Delecluse, P., 1998: An OGCM study for the TOGA decade. Part I: role of salinity in the physics of the western Pacific fresh pool. *J. Phys. Oceanogr.*, **28**, 1071–1088.

Weaver, A. T. and Vialard, J., 1999: Development of an ocean incremental 4D-Var scheme for seasonal prediction. In Proceedings of the Third WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography. 7-11 June 1999, Quebec City, Canada.

Weber, R. O. and Talkner, P., 1993: Some remarks on spatial correlation function models. *Mon. Wea. Rev.*, **121**, 2611–2617.

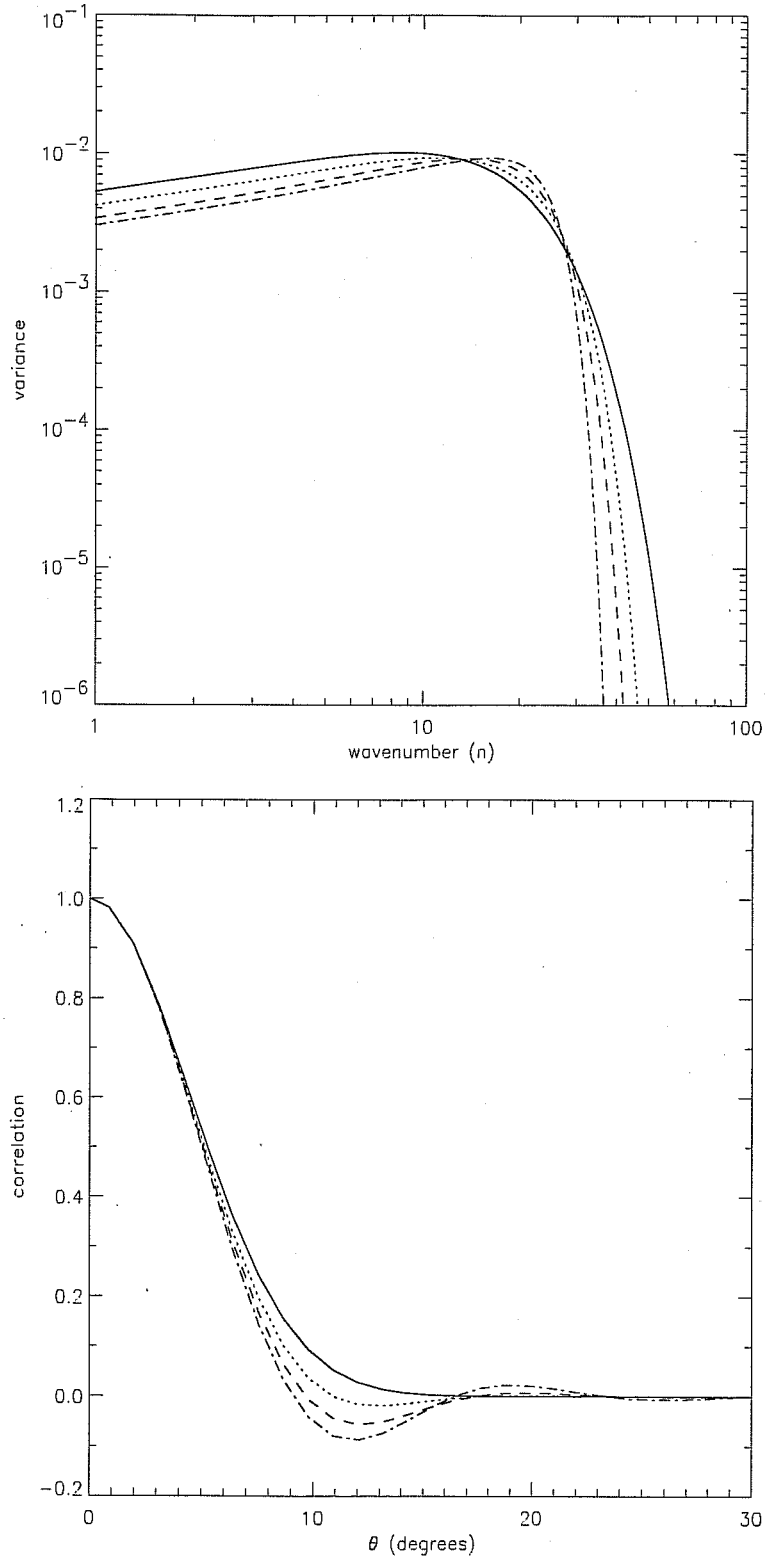

Figure 1. The variance power spectrum (upper panel) and grid-point values (lower panel) for four different correlations functions $f(\theta;\kappa_1 T,...,\kappa_P T)/f(0;\kappa_1 T,...,\kappa_P T)$ generated from Eq. (36). For each correlation function, the diffusion parameters $\kappa_p T$ have been tuned to give a length scale of $L = 500$ km. The solid curves correspond to the near-Gaussian function in Fig. 7 ($P = 1$ and $\kappa_1 T = 3.08 \times 10^{-3}$). The other curves correspond to parameter choices $P = 2$, $\kappa_1 = 0$ and $\kappa_2 T = 3.02 \times 10^{-6}$ (dashed); $P = 2$, $\kappa_1 T = 1.54 \times 10^{-3}$ and $\kappa_2 T = 1.39 \times 10^{-6}$ (dotted); and $P = 3$, $\kappa_1 = \kappa_2 = 0$ and $\kappa_3 T = 3.78 \times 10^{-9}$ (dashed-dotted).

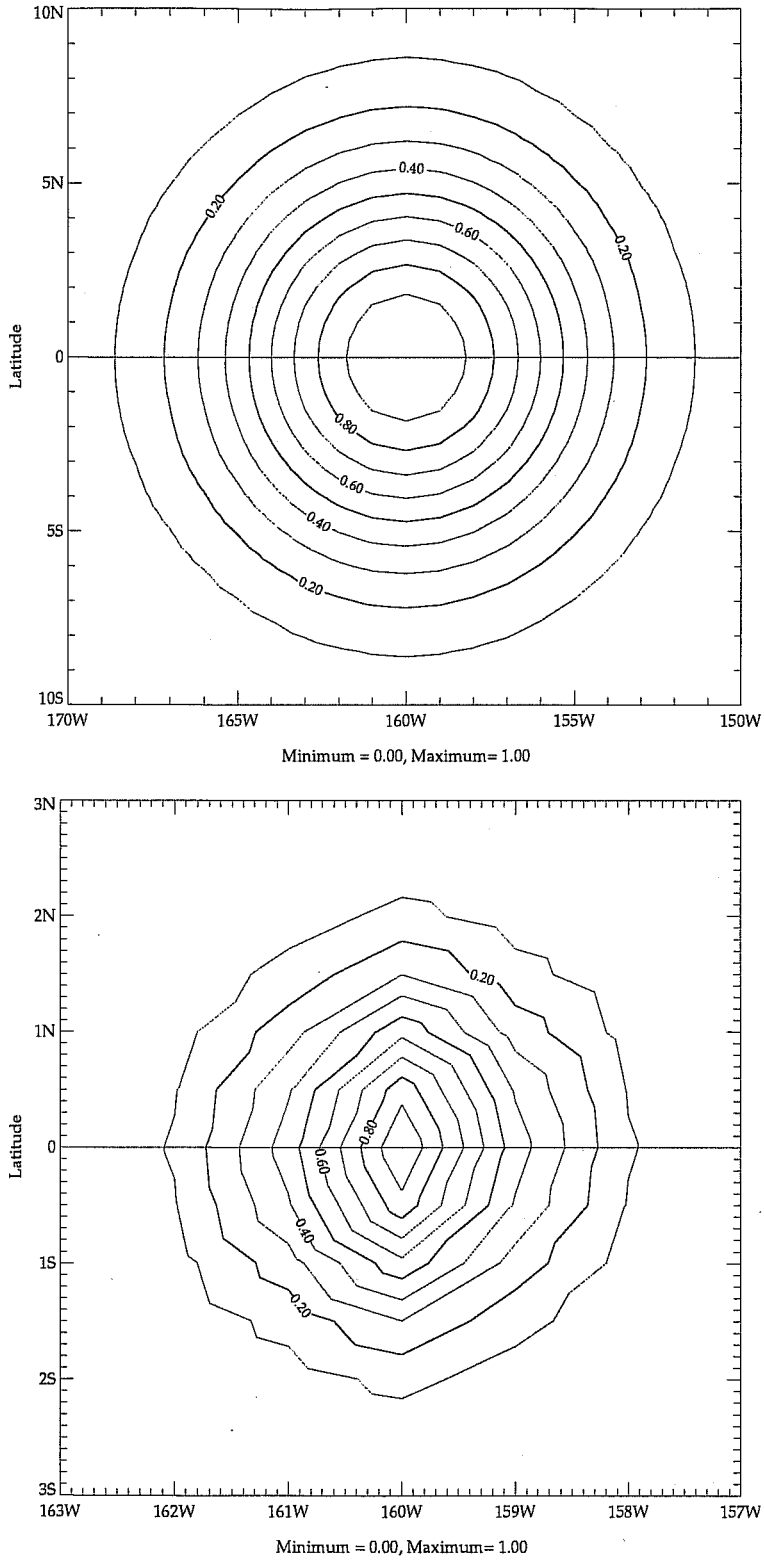

Figure 2. The auto-correlation field generated by the $\mathbf{L}_1^h$-filter (i.e., by applying (46) with $P = 1$ to the vector $\mathbf{e}_I = (0, ..., 0, 1, 0, ..., 0)^T$). The correlation length scale is 4° (upper panel) and 1° (lower panel). The zonal resolution is 1° and the meridional resolution is approximately 0.5°. The horizontal and vertical axes are in degrees longitude and latitude respectively.

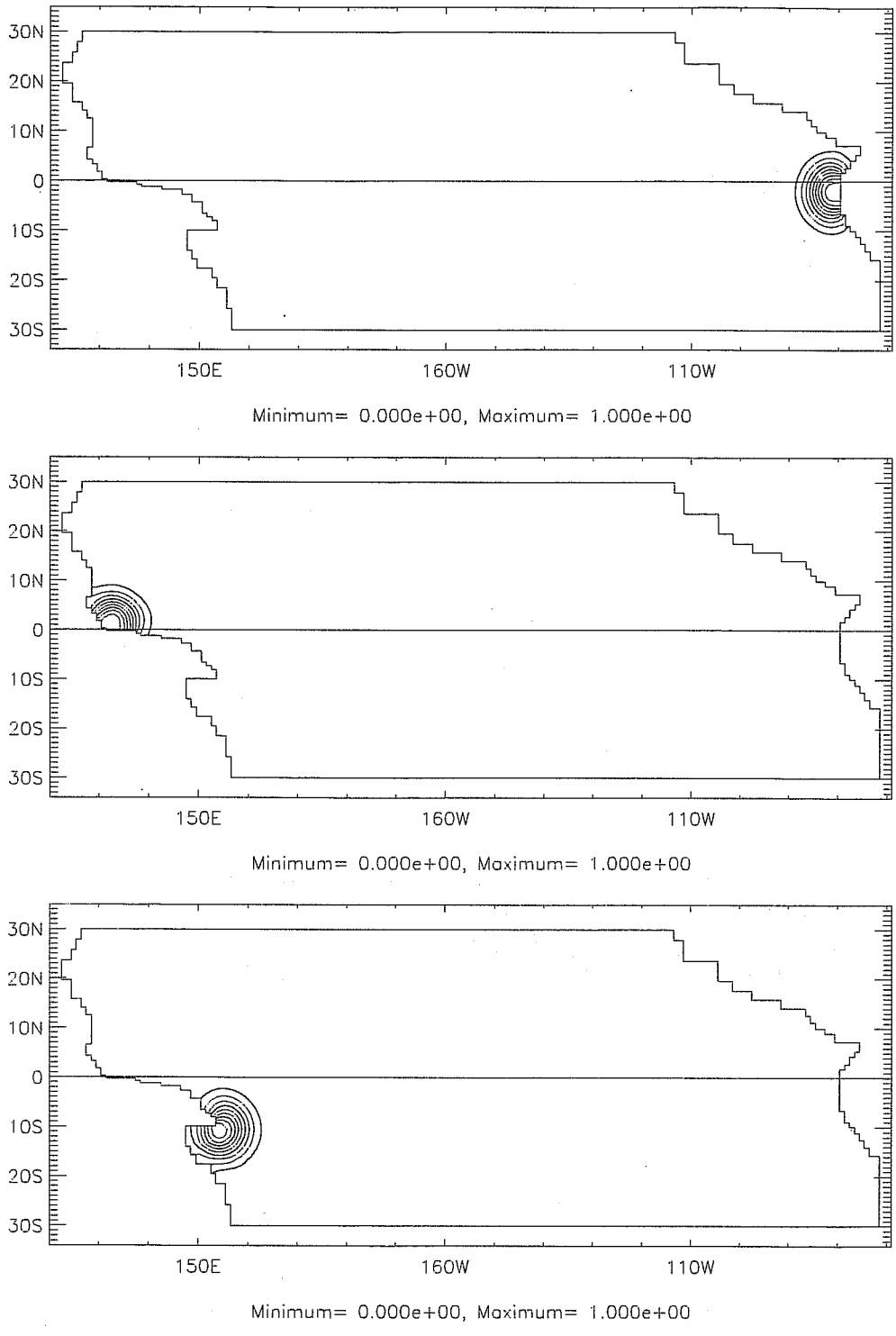

Figure 3. The auto-correlation field generated by the $\mathbf{L}_1^h$-filter at three grid-points adjacent to the continental boundaries. The domain of the ocean model covers the tropical Pacific basin and is closed along all boundaries. The contour interval is 0.1.

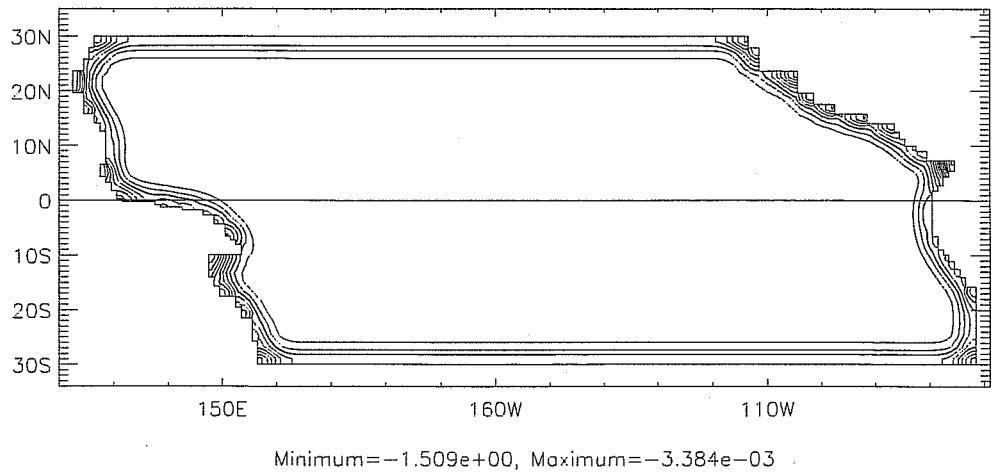

Figure 4. The relative difference $(1 - \sqrt{t_l/t})$ between the true normalization factors $(1/\sqrt{t_l})$ for the $\mathbf{L}_1^h$-filter at tracer grid-points in the upper-most level in the model and the constant normalization factor $(t = \sqrt{2\pi}L)$ estimated from the analytical solution of the diffusion equation in a boundary-free domain with spatially homogeneous diffusion coefficient. The boundary conditions used for the $\mathbf{L}_1^h$-filter assume vanishing normal derivatives ("no flux") at the boundaries. The contour interval is 0.1.

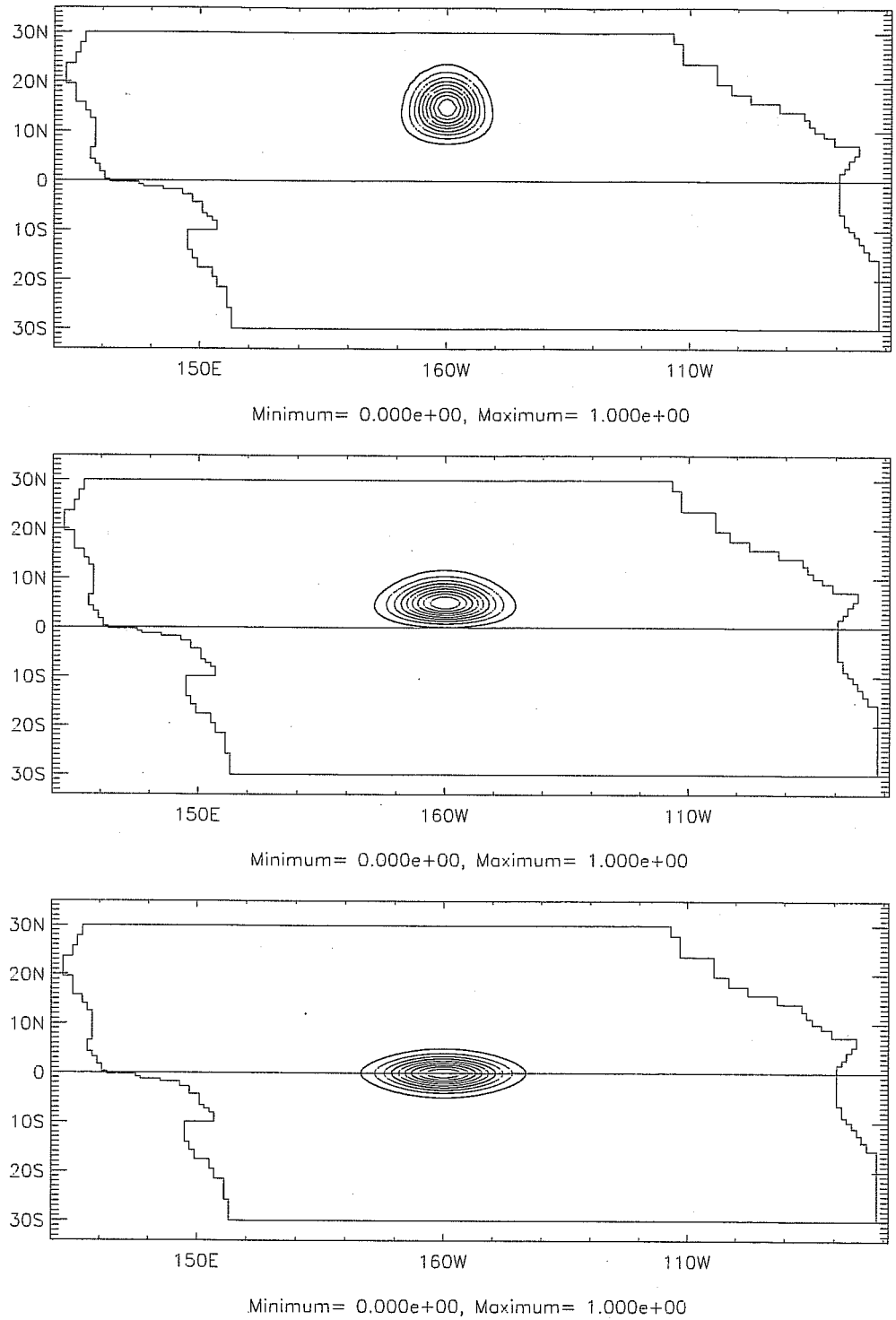

Figure 5. The auto-correlation field generated at three different latitudes ($\phi = 0°$, 5° N and 15° N) by an anisotropic version of the $\mathbf{L}_1^h$-filter. The contour interval is 0.1.

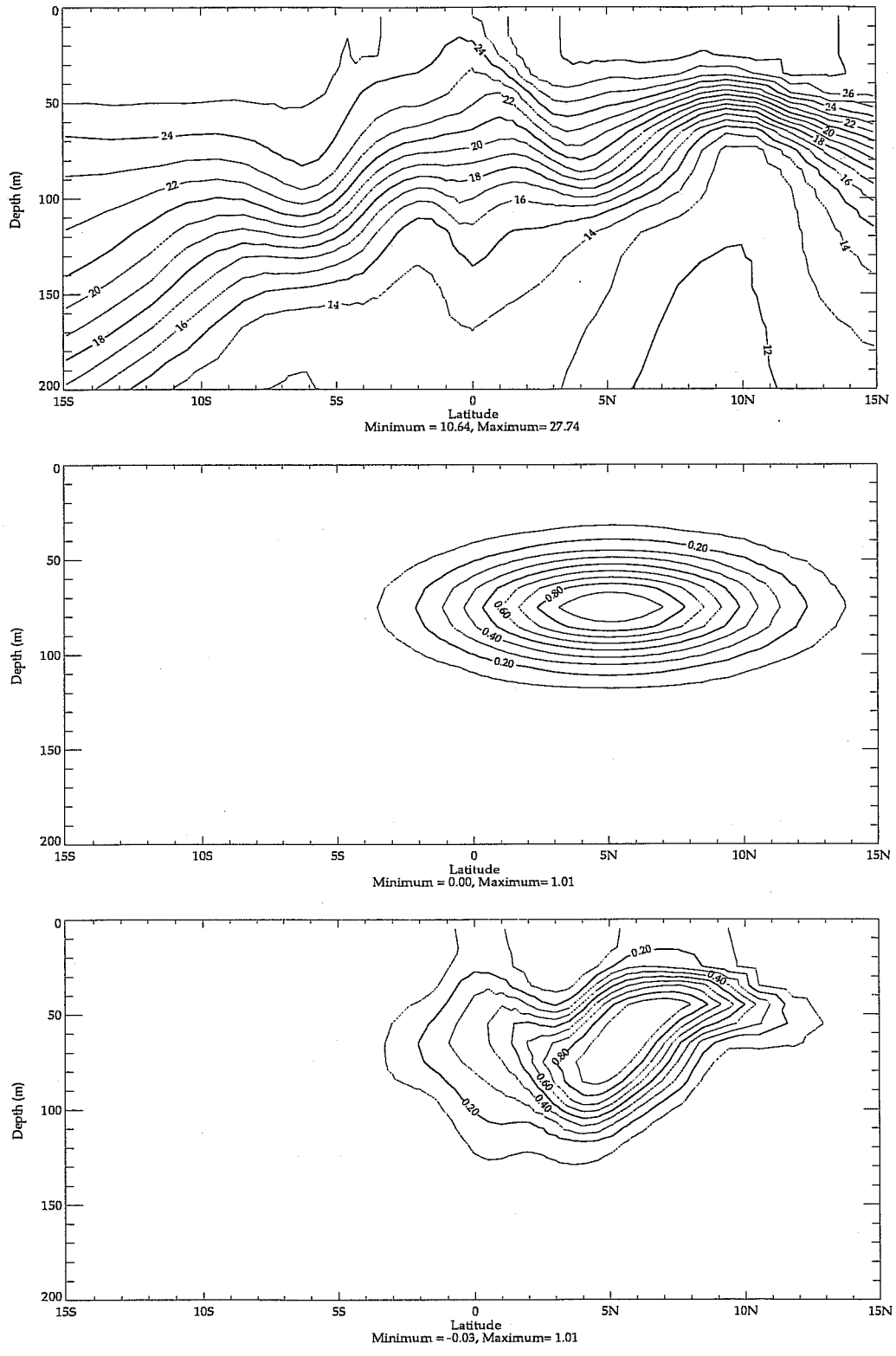

Figure 6. The upper panel illustrates a meridional section of a typical background potential temperature field in the eastern tropical Pacific (110° W). The middle panel shows the auto-correlation field at a depth of 55 m and latitude 8° N generated by a separable form of the 3D $\mathbf{L}_1$-filter (Eq. (57)) defined with respect to the geopotential coordinate system. The lower panel shows the corresponding auto-correlation field obtained using the 3D $\mathbf{L}_1$-filter (Eq. (58)) defined with respect to an isopycnal coordinate system based on the background isopycnal surfaces associated with the upper panel.

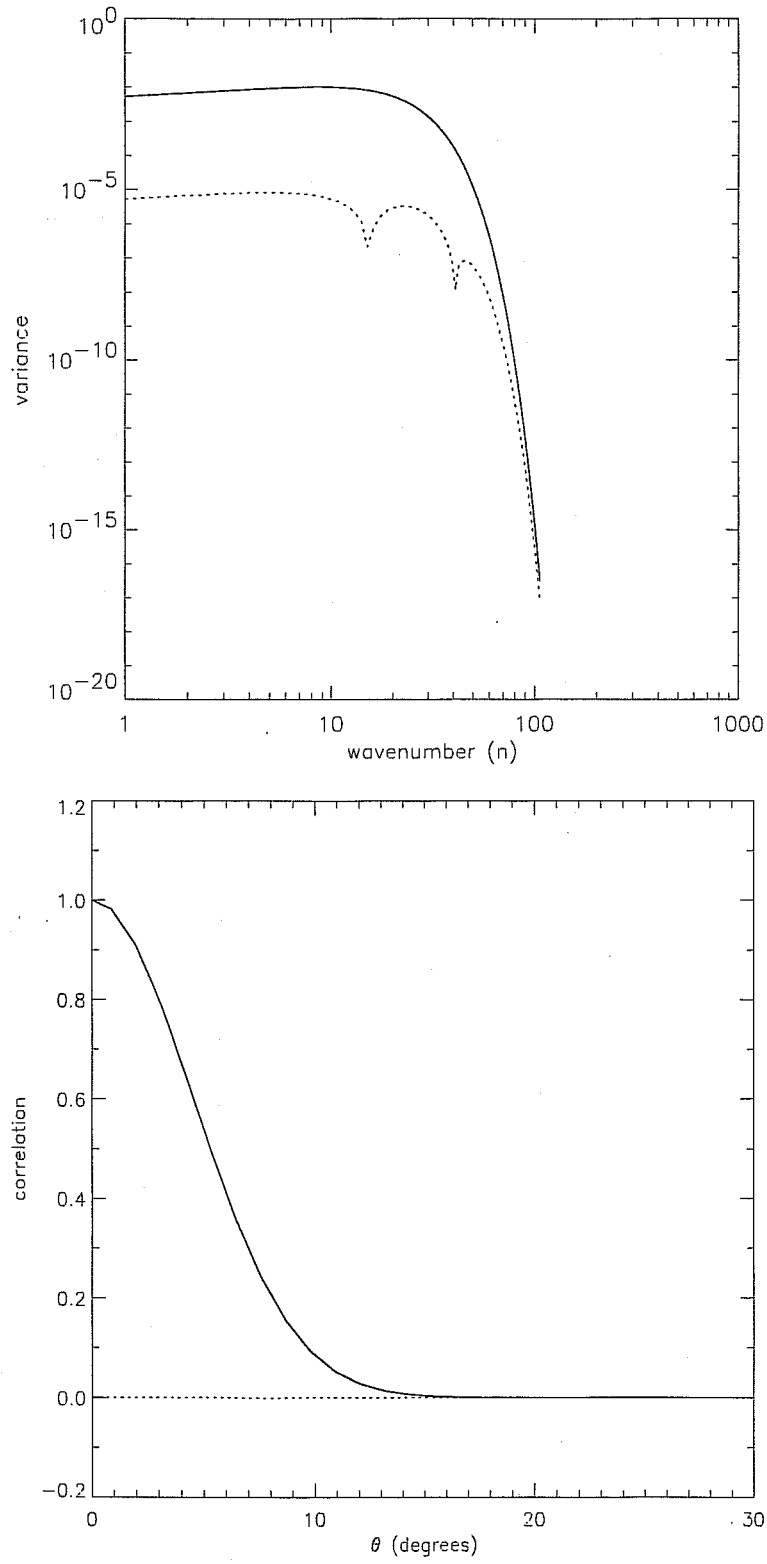

Figure 7. The variance power spectrum (upper panel) and grid-point values (lower panel) of the isotropic correlation functions $g(\theta,\gamma)/g(0,\gamma)$ (solid curve) and $f(\theta,\kappa T)/f(0,\kappa T)$ (dashed curve). The curves are indistinguishable; the dotted curve is their difference.