Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially Inhomogeneous and Anisotropic General Covariances

R. JAMES PURSER AND WAN-SHU WU

Science Applications International Corporation, Beltsville, Maryland

DAVID F. PARRISH

National Centers for Environmental Prediction/Environmental Modeling Center, Washington, D.C.

NIGEL M. ROBERTS

Joint Centre for Mesoscale Meteorology, University of Reading, Reading, and Met Office, Bracknell, United Kingdom

ABSTRACT

In this second part of a two-part study of recursive filter techniques applied to the synthesis of covariances in a variational analysis, methods by which non-Gaussian shapes and spatial inhomogeneities and anisotropies for the covariances may be introduced in a well-controlled way are examined. These methods permit an analysis scheme to possess covariance structures with adaptive variations of amplitude, scale, profile shape, and degrees of local anisotropy, all as functions of geographical location and altitude.

First, it is shown how a wider and more useful variety of covariance shapes than just the Gaussian may be obtained by the positive superposition of Gaussian components of different scales, or by further combinations of these operators with the application of Laplacian operators in order for the products to possess negative sidelobes in their radial profiles.

Then it is shown how the techniques of recursive filters may be generalized to admit the construction of covariances whose characteristic scales relative to the grid become adaptive to geographical location, while preserving the necessary properties of self-adjointness and positivity. Special attention is paid to the problems of amplitude control for these spatially inhomogeneous filters and an estimate for the kernel amplitude is proposed based upon an asymptotic analysis of the problem.

Finally, a further generalization of the filters that enables fully anisotropic and geographically adaptive covariances to be constructed in a computationally efficient way is discussed.

1. Introduction

In the first part of this study (Purser et al. 2003, henceforth referred to as Part I) we focused on the numerical methods that could be applied efficiently to the task of generating spatially homogeneous and horizontally isotropic smoothing kernels on a regular grid. We showed how carefully constructed high-order quasi-Gaussian recursive filters could overcome some of the well-known deficiencies of the simpler first-order recursive filters used in empirical data analysis schemes by Purser and McQuigg (1982), Hayden and Purser (1995), and discussed by Lorenc (1997). The most serious problem with the first-order filters is that, unless iterated many times, they show a pronounced spurious anisotropy

aligned with the directions of the numerical grid lines. The high-order filters of Part I produce more closely circular contours of amplitude without the need for numerous costly iterations. In addition, we showed how to improve the treatment of boundaries and noted that Fujita and Purser (2001) used a generalization of the cyclic boundary treatment as the basis of one of the efficient parallelization strategies they tested in a related context.

The emphasis of this second part of the study is the generalization and refinement of the methods of Part I to enable a far richer variety of covariance operators to be generated with controlled geographic variations. Section 2 deals with specific proposals for the construction of non-Gaussian parameterized families of covariance models based on linear superposition of the quasi-Gaussian "building blocks" that the recursive filters of Part I provide. There are many reasons that Gaussian forms are convenient in this respect, and some of these are discussed in Part I. Gaspari and Cohn (1998, 1999)

Corresponding author address: R. James Purser, NOAA/NCEP, W/NP2 Rm. 207, WWBG, 5200 Auth Rd., Camp Springs, MD 20746-4304.
E-mail: jim.purser@noaa.gov

have successfully used efficient compact-support approximations to Gaussian-shaped covariances in their implementation of the physical-space statistical analysis system (PSAS) described in da Silva et al. (1995). Some of the techniques for synthesizing more general covariances from quasi-Gaussian building blocks apply equally to these direct applications of compactly supported covariance models. One covariance family that we have found to be extremely convenient to use and beneficial in applications comprises bell-shaped distributions with significantly fatter tails than the Gaussian. We discuss the efficient construction of approximations to these fat-tailed distributions that allow a broader dynamical range of scales in the analysis increments to be assimilated.

A variation of the characteristic horizontal scale of the filters allows analysis schemes to adapt to the geographic variations in the density and quality of the recent observational data, which clearly are reflected in the statistical characteristics of the errors of the forecast background field—the starting point for the new analysis. Also, for global, and other large-scale, domains, it is not possible to maintain a uniform resolution everywhere with an orthogonal coordinate grid, owing to the intrinsic curvature of the earth. Thus, even were we to require a covariance of perfectly uniform characteristic scale in units of true distance, this would necessarily translate to a requirement for a filter of *varying* scale in the grid units that are relevant to the numerical construction of the filter. These matters are dealt with in section 3 and a proposal for the approximation of the filter's amplitude in the case of inhomogeneous filters is provided in the appendix.

A further important generalization of the covariances involves deliberately relaxing the artificial constraint of local isotropy. The advantages in a variational analysis of *not* insisting on isotropic covariances are probably obvious to any practitioner and are discussed in the recent article by Otte et al. (2001). Studies of the covariances implied by four-dimensional variational methods (e.g., Thépaut et al. 1996) provide objective confirmation of the need for such covariances. When the analysis is performed in a favorably distorted coordinate system, such as the geostrophic momentum coordinates used in the study of Desroziers (1997), an anisotropy in the geographical metric occurs automatically even while the filter remains "isotropic" relative to the (distorted) grid. In a somewhat analogous way, a meteorological analysis carried out in an isentropic coordinate framework (e.g., Shapiro and Hastings 1973; Benjamin 1989) automatically provides enhanced vertical resolution in zones of above average static stability where it is very often beneficial to clarify the more intricate structures of frontal zones or inversion layers. However, it is also desirable to be able to acquire a more precise control over the form and degree of anisotropy assigned to the covariance kernels than one obtains by relying on the fortuitous properties of certain predefined coordinate choices. Section 4 describes a new approach to the construction of general anisotropic covariances, which we call the "hexad" algorithm. At any given location, the three-dimensional covariance is synthesized from combinations of six recursive filters that act along oblique lines of the analysis grid (in two dimensions there is an analogous "triad" algorithm). The symmetric "aspect tensor" prescribing the local second moment structure of the desired covariance can be shown to have a unique hexad (or triad) representation by line filters on the given grid. We note that the application of anisotropic filtering is analogous to a diffusion process generalized to posses a diffusivity in the form of an anisotropic symmetric tensor. The aspect tensor for a homogeneous diffusion process with diffusivity D acting for a duration, $t,$ is then just twice the product of D and t. This explicitly diffusive approach is discussed by Weaver and Courtier (2001) in their extension of the earlier work of Derber and Rosati (1989), and is another valid way of considering the problem of generating anisotropic Gaussian smoothers. The main advantage of employing recursive filters is that they tend to require significantly fewer iterations to achieve the desired Gaussian form. Our concluding remarks are given in section 5.

2. Synthesis of a covariance in terms of Gaussians

Let $G(\mathbf{x}_1, \mathbf{x}_2)$ denote the kernel of a (quasi-) Gaussian recursive filter operator of the kind defined by Eq. (3.23) of Part I. For a univariate covariance constructed directly from this Gaussian, the amplitude (i.e., the "variance") may be made to vary geographically while preserving the important property of self-adjointness by modulating the otherwise homogeneous filter before and after application:

$$B(\mathbf{x}_1, \mathbf{x}_2) = w^{1/2}(\mathbf{x}_1)G(\mathbf{x}_1, \mathbf{x}_2)w^{1/2}(\mathbf{x}_2), \qquad (2.1)$$

where $w(\mathbf{x})$ is the total effective weight at $\mathbf{x}$ applied to the G. However, it has been recognized that objective analysis using the Gaussian shape to model the covariance severely hampers the ability of the analysis to assimilate the smallest scales of significant background error. In adverse configurations of the data, the problem is apt to manifest itself in excessive and damaging extrapolation effects at the edge of isolated data voids where the analysis strives to fit the surrounding more densely distributed data smoothly. Lorenc (1981) provides an illustration of this effect with idealized data. The small-scale analysis increments are inadvertently inhibited when, as with a Gaussian model, the presumed power spectrum values at moderately large wavenumbers become much smaller than the values that the data and experience indicate to be appropriate. Recall that the power spectrum for a spatially homogeneous covariance model is simply the Fourier transform of that covariance. For example, the two-dimen-

sional Gaussian function of unit integrated weight, and scale parameter, a,

$$G_a(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2\pi a^2} \exp\left[-\frac{1}{2}\left(\frac{\mathbf{x}_1 - \mathbf{x}_2}{a}\right)^2\right], \quad (2.2)$$

is associated with the power spectrum,

$$\hat{G}_a(k, l) = \iint G_a(\mathbf{x}, 0) \exp(-i\mathbf{k} \cdot \mathbf{x}) \, dx \, dy$$

$$= \exp\left[-\frac{a^2}{2}(k^2 + l^2)\right], \quad (2.3)$$

which is itself of Gaussian form and which, therefore, does possess a rapidly diminishing tail in the spectral region of total wavenumbers exceeding a few times the "characteristic wavenumber," $(1/a)$. Clearly, the ideal remedy would be the representation of the actual covariance operator (whatever that is); at least it should be possible, within the space of isotropic covariances, to find one possessing more appropriate tail characteristics, although this almost certainly necessitates exploring options other than the Gaussian family. In principle, however, by the methods of Laplace transforms, it should be possible to synthesize almost *any* practical isotropic covariance profile as a superposition of Gaussian components, as is noted by Schoenberg (1938) and more recently discussed by Gneiting (1999):

$$\hat{B}(k, l) = \int_0^\infty w\left(\frac{a^2}{2}\right) \exp\left[-\frac{a^2}{2}(k^2 + l^2)\right] d\left(\frac{a^2}{2}\right). \quad (2.4)$$

In practice, we would wish to approximate the general superposition (2.4) by a discrete approximation involving many Gaussian constituents. That is,

$$\hat{B}(k, l) = \sum_p^M w_p \exp\left[-\frac{a_p}{2}(k^2 + l^2)\right], \quad (2.5)$$

where $M \approx 5$ typically. Given that the individual Gaussian components are positive definite, then $\hat{B}$ is

also, provided the weights w_p are all non-negative (in general, a positive superposition of positive definite operators is itself positive definite). We would also like to confine ourselves to constructions generalizing this type so that, in the spatial domain, the combination can be expressed in a way that does not jeopardize the self-adjointness that the iterative solution algorithms depend on. The appropriate form in the spatial domain is therefore

$$B(\mathbf{x}_1, \mathbf{x}_2) = \sum_p^M w_p^{1/2}(\mathbf{x}_1) G_{a_p}(\mathbf{x}_1, \mathbf{x}_2) w_p^{1/2}(\mathbf{x}_2). \quad (2.6)$$

Note that, although in practice we should expect the weights w_p to vary smoothly in space, the self-adjointness and positive definiteness of this construction hold regardless. The profile of weights must be regarded as samples, at the discrete scales selected, of an underlying *continuous* weight profile. The *density* of selected scales per "octave" (the term we shall use for a change by a factor of 2 in a scale or a wavenumber) is something we must determine according to the smoothness of the continuous weight profile in the log-scale domain. With an adequate scale resolution, it then becomes possible to change the overall scale progressively across the extent of a large domain without appreciably altering (except by linear contraction or expansion) the intended *shape* of the covariance profile.

One of the simplest general families of scale profiles accommodating the requirement that the sampled weights w_p all be non-negative is what we shall call the "hyper-Gaussian" family of functions. For the two-dimensional isotropic hyper-Gaussian functions normalized to have unit integrals, a generic member of this family is characterized by a scale parameter, σ, and a shape parameter, γ. One may regard each member function as being a continuous superposition of isotropic Gaussians of horizontal scales $\exp(s)$ according to a weighting profile in s that is itself of a Gaussian form, centered on $\log(\sigma)$ and with a dispersion parameter in s of γ:

$$H_{\sigma,\gamma}(\mathbf{x}) = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\gamma}} \exp\left[-\frac{1}{2}\frac{(s - \log\sigma)^2}{\gamma}\right] \frac{1}{2\pi \exp(2s)} \exp\left[-\frac{1}{2}\left(\frac{x}{\exp(s)}\right)^2\right] ds. \quad (2.7)$$

Note that, in the limiting case, $\gamma = 0$, this model reverts to pure Gaussian form.

Figure 1a illustrates the radial profile of the correlation implied by this covariance model for $\gamma = 0, 0.1, 0.2,$ and 0.3. The parameter γ provides control over what we may refer to, following statistical parlance, as the "kurtosis" of the distribution, which is a quality of the response shape sensitive primarily to the nondimensionalized fourth moment, as defined below. Suppose

we define unidirectional moments of an isotropic distribution, $H(\mathbf{x})$:

$$\mu_m = \iint H(\mathbf{x}) x^m \, dx \, dy. \quad (2.8)$$

Thus, for the normalized Gaussian, G_a, we have the moments, $\mu_0 = 1$, $\mu_2 = a^2$, and $\mu_4 = 3a^4$. Since the construction (2.7) is a linear superposition of Gaussians,

and each moment is a linear functional of the distribution, we find that, for the hyper-Gaussian with shape parameter γ, the corresponding moments are $\mu_0 = 1$, $\mu_2 = \sigma^2 \exp(2\gamma)$ and $\mu_4 = 3\sigma^4 \exp(8\gamma)$. Then, if we adopt the definition of kurtosis to be the nondimensional quantity,

$$\kappa = \frac{\mu_0 \mu_4}{(\mu_2)^2}, \qquad (2.9)$$

we find that the kurtosis for the hyper-Gaussian of shape parameter γ is

$$\kappa_\gamma = 3 \exp(4\gamma).$$

As a generic shape parameter, the kurtosis has its limitations. In particular, it is generally not appropriate to define the kurtosis of a distribution that has regions of negative values, for example, covariances with negative sidelobes such as those defined below.

A family of covariances whose profiles possess negative sidelobes can be generated by a very similar superposition. We do this by replacing the Gaussian basis by the corresponding functions obtained by taking the negative Laplacian of each Gaussian. The resulting covariances,

$$H'_{\sigma,\gamma}(\mathbf{x}) = -\nabla^2 H_{\sigma,\gamma}(\mathbf{x}), \qquad (2.10)$$

have the correlation profiles depicted in Fig. 1b. In terms of the filtering operations represented by these profiles, the application of the Laplacian operator will require some extra cost.

A further consequence of the superposition property is that the power spectra of the hyper-Gaussian, and its negative Laplacian, are expressible as simple integrals:

$$\hat{H}_{\sigma,\gamma}(\mathbf{k}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\gamma}} \exp\left[-\frac{1}{2} \frac{(s - \log\sigma)^2}{\gamma}\right] \exp\left\{-\frac{1}{2}[\exp(s)\mathbf{k}]^2\right\} ds \quad \text{and} \qquad (2.11a)$$

$$\hat{H}'_{\sigma,\gamma}(\mathbf{k}) = \mathbf{k}^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\gamma}} \exp\left[-\frac{1}{2} \frac{(s - \log\sigma)^2}{\gamma}\right] \exp\left\{-\frac{1}{2}[\exp(s)\mathbf{k}]^2\right\} ds. \qquad (2.11b)$$

These power spectrum families are shown, for the same range of shape parameters as before, in Figs. 2a (for $\hat{H}$) and 2b (for $\hat{H}'$). Note the dramatic effect on the power at small scales (large wavenumber) that results from even small positive values of the shape parameter, γ.

The natural question that now arises is how fine a resolution in the log scale, s, is required to adequately represent these covariance models by the approximations that replace the integral representations (with respect to s) by discrete summations. This can be answered by observing how far the discrete approximations' power spectra depart from the exact integral representations' power spectra. In practice, we find that, for $\gamma = 0.3$, about three scales per octave appears to be adequate. For smaller γ it is prudent to increase this density of discrete scales. In the context of a multigrid construction (discussed in Part I), it is clearly convenient numerically to have an integer number of discrete smoothing scales of the basic Gaussians in each octave, or, in other words, to have the same whole number of smoothing scales per grid of the multigrid hierarchy.

3. Inhomogeneous generalizations

In this section we treat cases in which the grid remains orthogonal and smooth in terms of its resolution, but not necessarily uniform or without curvature. At the same time, we treat the case in which the filter remains locally isotropic, but whose smoothing scale is permitted to vary geographically. Polar grids, such as plane polars or global latitude and longitude grids, possess special rotational symmetries that can be exploited in the case of the spatially homogeneous smoothing filters that respect those symmetries. But polar grids also present unique difficulties involving the polar singularities themselves, which then require special corrective measures to be applied to the filters. We pay attention to these problems in this section and suggest some of the remedies that are possible by exploiting the analogy between recursive filtering and a general diffusion process. In this regard, the generalized recursive filters can be regarded as accelerated numerical solvers for approximating the action of inhomogeneous diffusion. These methods can therefore be considered consistent alternatives to the direct application of diffusion operators proposed by Derber and Rosati (1989) and recently generalized by Weaver and Courtier (2001).

a. Inhomogeneities of grids or filter scales

One should not be led to believe that our construction of quasi-Gaussian filters is necessarily restricted to perfectly uniform Cartesian grids. On a smoothly varying nonuniform grid in one dimension, the tridiagonal discretization of the differential operator d^2/dx^2, and polynomials of the discretization by which the requisite powers of d^2/dx^2 are approximated, still lead to banded matrices that, as shown below, can be rendered symmetric by a similarity transformation with a diagonal matrix
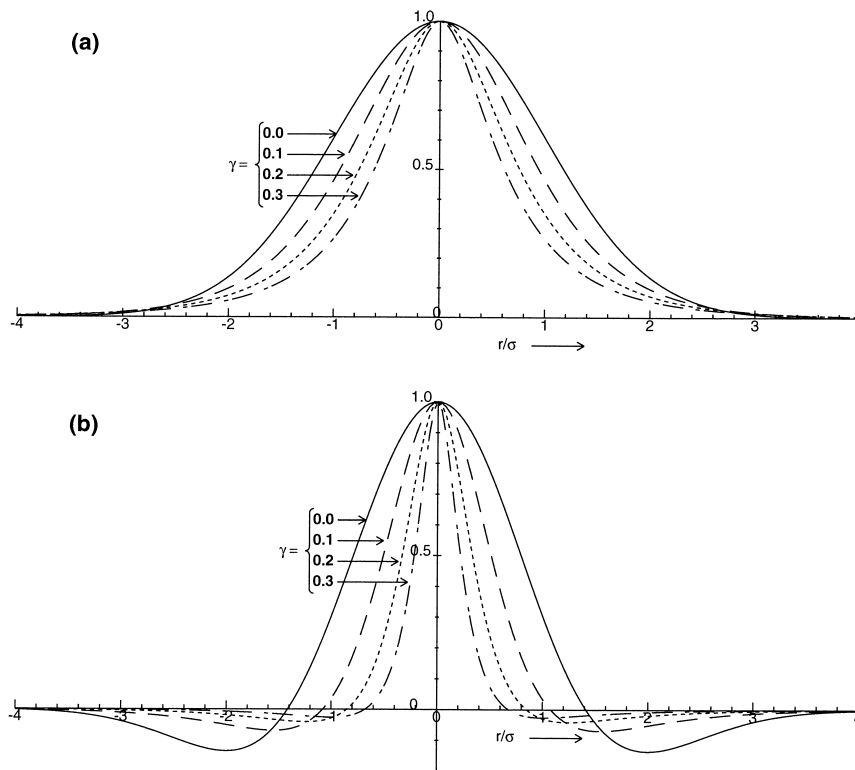
FIG. 1. (a) Cross-sectional profiles of the fat-tailed "hyper-Gaussian" covariance models defined in section 5 for a range of shape parameters γ. (b) The result of applying the negative Laplacian, and renormalization of amplitude, to these hyper-Gaussian functions.

related to the metrical properties of the grid. A power, i, of the simplest tridiagonal representation of a second derivative operator results in a band matrix of $2i + 1$ diagonals. Also, we can generalize the conditions of homogeneity of the smoothing scales to incorporate the effects of a scale that can vary smoothly across the grid, again, without invalidating the property of self-adjointness. However, this additional generalization requires that, in all appearances of the operator, $[-(a^2/2)d^2/dx^2]$, in the counterpart to the polynomial (3.8) of Part I, a form of the second derivative factor is substituted that is self-adjoint even when a is a function of x. Of the qualifying possibilities, the one that is most convenient in practice and that leads to a substance-conserving filter is the one most closely identified with the operation of a diffusive process:

$$-\frac{d}{dx}\frac{a^2(x)}{2}\frac{d}{dx}. \tag{3.1}$$

The operator, (3.1), would be appropriate when the grid lines along which x varies are all parallel but, in a general orthogonal curvilinear grid, this is no longer true. The final generalization we add in this section is the accommodation of grids with converging or diverging grid lines. We do this by including a metric term, τ, whose reciprocal is the density of x-grid lines so that τ itself may be thought of as the line or area measure

(according to whether the grid is two- or three-dimensional) of the interface orthogonal to the grid line and attributed to it in finite-difference operations. Using partial derivatives to emphasize the implied multidimensionality, the operator we need to generalize (3.1) is

$$-\frac{1}{\tau}\frac{\partial}{\partial x}\tau\frac{a^2(x)}{2}\frac{\partial}{\partial x}, \tag{3.2}$$

which is self-adjoint in the sense of an inner product defined

$$(s, t) \equiv \int s(x)t(x)\tau(x)\,dx. \tag{3.3}$$

Let x_i be the main grid coordinates for integers i and let the intermediate staggered grid of points such as $x_{i+1/2}$ be a smooth interpolation from it. Likewise, by smooth interpolation, we assume a and τ to be available at the main and staggered grids. Define

$$\delta x_i = x_{i+1/2} - x_{i-1/2}, \tag{3.4}$$

$$\delta x_{i+1/2} = x_{i+1} - x_i. \tag{3.5}$$

Redefining the scale parameter $\sigma = a/\delta x$ as in Part I, the simplest consistent discretization of the operator (3.2) is
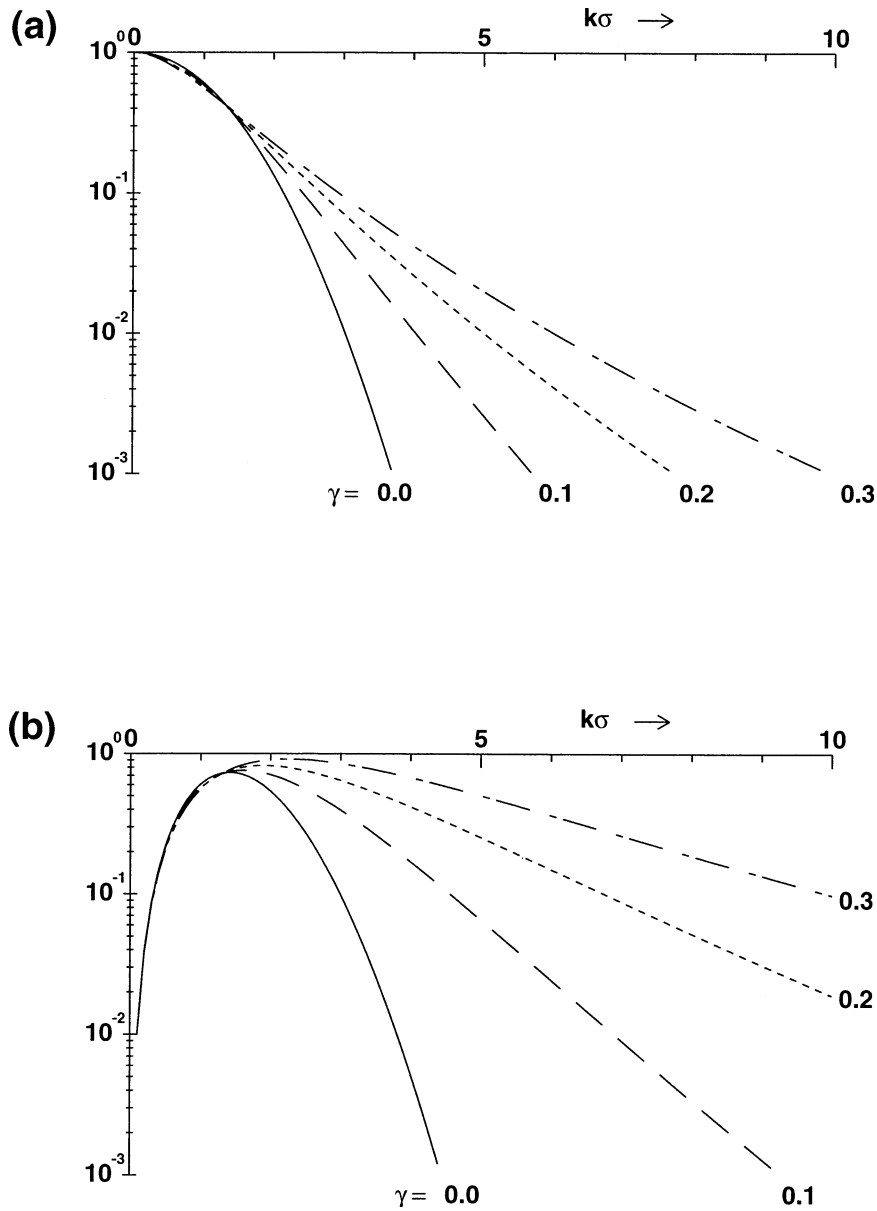
(a)

(b)

FIG. 2. Power spectra in log-linear coordinates for the covariances depicted in Fig. 1. (a) The hyper-Gaussians and (b) the negative Laplacians of the hyper-Gaussians.

$$\left[-\frac{1}{\tau} \frac{\partial}{\partial x} \tau \frac{a^2(x)}{2} \frac{\partial s}{\partial x} \right]_i$$

$$\approx \frac{1}{2\tau_i \delta x_i} \left[\frac{\tau_{i-1/2} a_{i-1/2}^2 (\delta s_{i-1/2})}{\delta x_{i-1/2}} - \frac{\tau_{i+1/2} a_{i+1/2}^2 (\delta s_{i+1/2})}{\delta x_{i+1/2}} \right]$$

$$\equiv \frac{1}{2\nu_i} [\nu_{i-1/2} \sigma_{i-1/2}^2 (-s_{i-1} + s_i)$$

$$+ \nu_{i+1/2} \sigma_{i+1/2}^2 (s_i - s_{i+1})], \qquad (3.6)$$

where

$$\nu_i = \tau_i \delta x_i \qquad (3.7)$$

defines the local grid cell measure (area or volume). We can relate this operator to a tridiagonal matrix, **K**, that serves to generalize the K of Eq. (3.1) of Part I. A representative row, i, of **K**, is defined by

$$K_{i,i-1} = -\frac{(\nu\sigma^2)_{i-1/2}}{\sqrt{(\nu\sigma^2)_i (\nu\sigma^2)_{i-1}}}, \qquad (3.8a)$$

$$K_{i,i} = \frac{(\nu\sigma^2)_{i-1/2} + (\nu\sigma^2)_{i+1/2}}{(\nu\sigma^2)_i}, \qquad (3.8b)$$

$$K_{i,i+1} = -\frac{(\nu\sigma^2)_{i+1/2}}{\sqrt{(\nu\sigma^2)_i (\nu\sigma^2)_{i+1}}}, \qquad (3.8c)$$

in terms of which, the finite-difference operator of (3.6) is obtained:

$$\left[-\frac{1}{\tau}\frac{\partial}{\partial x}\tau\frac{a^2(x)}{2}\frac{\partial s}{\partial x}\right]_i \approx \frac{1}{2}\left(\frac{1}{\sqrt{\boldsymbol{\nu}}}\boldsymbol{\sigma}\mathbf{K}\boldsymbol{\sigma}\sqrt{\boldsymbol{\nu}}\mathbf{s}\right)_i, \quad (3.9)$$

where $\boldsymbol{\sigma}$ and $\sqrt{\boldsymbol{\nu}}$ are the diagonal matrices formed from the values σ_i and $\sqrt{\boldsymbol{\nu}_i}$. The components of the matrix $\mathbf{K}$ obey the approximation

$$[K_{i,i-1}, K_i, K_{i,i+1}] \approx [-1, 2, -1] \qquad (3.10)$$

very closely when a, δ_x, and τ are all smooth and slowly varying in x, tending to these values in the limiting case of constant a, δx, and τ. Having found a consistent self-adjoint, but low-order-accurate numerical approximation to the appropriate second derivative, the suggested refinement of accuracy available through the use of the coefficients $b_{1,j}$ defined in Part I is

$$-\frac{1}{\tau}\frac{\partial}{\partial x}\tau\frac{a^2(x)}{2}\frac{\partial}{\partial x} \approx \frac{1}{2}\frac{1}{\sqrt{\boldsymbol{\nu}}}\boldsymbol{\sigma}\left(\sum_{j=1}^{n} b_{1,j}\mathbf{K}^j\right)\boldsymbol{\sigma}\sqrt{\boldsymbol{\nu}}. \quad (3.11)$$

Taking the Taylor series for the exponential of this operator, but truncating all the terms composing matrices of half-bandwidth exceeding n (as discussed in Part I, n is typically four to obtain results closely approximating the Gaussian shape), we obtaining the sought-for generalization of $\mathcal{D}^*_{(n)}$ in Eq. (3.12) of Part I:

$$\mathcal{D}^*_{(n)} \approx \frac{1}{\sqrt{\boldsymbol{\nu}}}\exp\left(\frac{\boldsymbol{\sigma}}{2}\sum_{j=1}^{n} b_{1,j}\mathbf{K}^j\boldsymbol{\sigma}\right)\sqrt{\boldsymbol{\nu}}. \quad (3.12)$$

While the coefficient-finding method of appendix A of Part I is no longer applicable in the general inhomogeneous case, Cholesky factorization is still possible, since at least the matrix sandwiched between diagonals $1/\sqrt{\boldsymbol{\nu}}$ and $\sqrt{\boldsymbol{\nu}}$ of (3.12) remains symmetric. This factorization provides the means to construct the associated advancing and backing recursive filters. However, these filters now have coefficients varying in space and so are slightly more complicated to apply. Also, the method of setting end conditions described in appendix B of Part I can no longer accurately simulate the indefinite continuation of the grid beyond a boundary in general. In practice, we apply those techniques as if the smoothing parameters at the boundary point in question are the values everywhere and we then find that the imperfections that result are often barely noticeable.

In order to control the amplitude of the covariance synthesized from inhomogeneous filters of the kind we have described we need to estimate, for each point in the domain, the amplitude of the result of applying the sequence of basic filters to a unit impulse located at this *same* point. The homogeneous case conforms to the Gaussian model. The Gaussian model with constant scale parameter a for one direction corresponds to diffusion in this direction with "diffusivity" D for "time" t, when

$$2Dt = a^2 \qquad (3.13)$$

and the impulse-response value of the result is just $(4\pi Dt)^{-1/2}$. In more than one dimension, the diffusivity D generalizes to a tensor (Weaver and Courtier 2001) and the appropriate generalization of the impulse response, $|4\pi Dt|^{-1/2}$, involves the determinant of D. What we have referred to as the aspect tensor is simply the tensorial generalization to these higher dimensions of the combination $2Dt$ of (3.13). But inhomogeneity of scale, which we may interpret as inhomogeneity of the effective diffusivity in the diffusion analog of our filters, leads to impulse-response functions that differ slightly from the profile calculated on the basis of the Gaussian model. An asymptotic analysis of this difference, which is outlined in the appendix, provides us with a valuable practical refinement to the Gaussian amplitude approximation. It emerges from this analysis that, to a good approximation, the impulse response at a given point of simulated inhomogeneous diffusion acting for time t is the same as the amplitude obtained by diffusing for duration t with an alternative *homogeneous* diffusion process whose constant diffusivity $\overline{D}$ is a local weighted average of D. The appropriate weighted average can be obtained by applying to D the original diffusion process, but for only half the usual time, $t/2$, and evaluating the result of the smoothed diffusivity field $\overline{D}$ at the point under consideration. The refinement is valid to first order in the magnitude of the modulation of D, so it is actually sufficient (and more practical) to simply smooth the single field of $|4\pi Dt|$. Note that the diffusion operation acting for duration $t/2$ is equivalent to the application of the "square root" (in the convolution or linear operator sense) of the total filter. (Note that, for homogeneous filters, the square root filter is expressible, by way of the convolution theorem, as the inverse transform of the square root of the Fourier transform of the original filter.) In many practical applications of the recursive filter method to data analysis, this square root filter, or at least a very good approximation to it, is already available owing to the manner in which the total filter is synthesized from simpler components. Thus, the amplitude refinement imposes no significant extra computational burden in typical cases where, once constructed, the same smoothing operator is subsequently applied over and over again.

The generalizations of the recursive filters we have described in this subsection work on a wide variety of grids provided each grid itself contains no singularity. But this restriction unfortunately precludes the use of the methods on a polar grid in the immediate vicinity of the pole. In order to treat such a case, the next subsection discusses some of the special techniques that can be brought to bear.

b. Polar grids

We shall only treat in detail the special case of filters with homogeneous filtering scale, a. On a plane-polar

grid or on a global grid of latitudes and longitudes, the recursive filter method can be adapted in conjunction with discrete Fourier transforms applied azimuthally or longitudinally to data, providing that the longitudes are uniformly spaced and of a number factorable into small primes, as required for the efficient application of the FFT algorithm (e.g., Press et al. 1992). Fourier transformation separates the two-dimensional smoothing problem into independent one-dimensional filtering operations in the radial or meridional directions applied to the zonal Fourier coefficients.

Suppose θ and λ are latitude and longitude, respectively. From initial data $\chi(\theta, \lambda)$ we may apply the zonal transform:

$$\hat{\chi}(\theta, m) = \int \chi(\theta, \lambda) \exp(-2\pi im\lambda) \, d\lambda. \quad (3.14)$$

The idea is to redefine the operator $\mathcal{D}^*_{(n)}$ of Eq. (3.9) of Part I in a fully two-dimensional way as the polynomial of $-(a^2/2)\nabla^2$ instead of $-(a^2/2)d^2/dx^2$. But we shall see that, through the process of Fourier transformation, the action of this operator is reduced to a set of one-dimensional operators acting on the various wavenumber components separately. For constant a, the operator corresponding to, $-(a^2/2)\nabla^2$, takes the following form in the semispectral domain (latitude and zonal wavenumber):

$$-\frac{a^2}{2}\nabla^2 = \frac{a^2m^2}{2\tau^2} - \frac{a^2}{2R^2\tau} \frac{\partial}{\partial\theta} \tau \frac{\partial}{\partial\theta}, \quad (3.15)$$

where the metric term, τ, is now defined for an earth of radius R simply as

$$\tau(\theta) = R\cos(\theta). \quad (3.16)$$

Note that, for each separate zonal wavenumber m, this operator has its simplest numerical representation in the discrete latitude grid as a tridiagonal matrix, but high-order corrective terms can be added using the method involving coefficients $b_{1,j}$ described in section 2a of Part I. As in the context of a nonuniform one-dimensional grid discussed above, this matrix operator is transformable to symmetric form through a similarity relation. In fact, the second term of the operator on the right of (3.15) applied to a scalar s may be identified exactly with either side of (3.9) when we use the substitutions

$$\nu_i \equiv \tau_i R\delta\theta_i \quad \text{and} \quad (3.17a)$$

$$\sigma_i \equiv a/(R\delta\theta_i), \quad (3.17b)$$

together with the same $\mathbf{K}$ defined as in (3.8a)–(3.8c). Following the previous example, the Taylor series of the exponential function of the discrete matrix representation of $-(a^2/2)\nabla^2$ is taken, and again, only component matrices of half-bandwidth not exceeding n are retained, in order to obtain the representation, at each wavenumber m, of $\mathcal{D}_{(n)}$. By analogy to (3.12), the numerical approximation $\mathcal{D}^*_{(n,m)}$ to the wavenumber m com-

ponents of this nth-order operator, $\mathcal{D}_{(n)}$, for the polar grid may be expressed:

$$\mathcal{D}^*_{(n,m)} \approx \frac{1}{\sqrt{\boldsymbol{\nu}}} \exp\left(\frac{a^2m^2}{2\tau^2} + \frac{\boldsymbol{\sigma}}{2} \sum_{j=1}^{n} b_{1,j}\mathbf{K}^j\boldsymbol{\sigma}\right)\sqrt{\boldsymbol{\nu}}. \quad (3.18)$$

From the derived Cholesky factors, the coefficients of the advancing and backing basic filters are then extracted separately for each wavenumber component.

These filters are more expensive to apply globally than the doubly recursive filters of section 3a because they require a zonal FFT to be applied to the input data at each latitude and an inverse FFT applied to the final output data. But they do provide a satisfactory solution to the "polar problem" in the case of homogeneous smoothing scale, a. In an earlier phase of this study, two of us (RJP and NMR) investigated filters of this semispectral form for a global analysis and devised methods for constructing hybrid filters in which only the polar caps are treated by the FFT, the data elsewhere being dealt with by the methods of section 3a. A discussion of a proposed extension to this technique to accommodate geographically inhomogeneous scale can be found in Purser et al. (2001), but perhaps a more relevant generalization is the accommodation of fully anisotropic functions, discussed below.

4. Anisotropic covariances

In addition to spatial inhomogeneity, we would like to be able to stretch the shape of a local representative contour surface of the covariance function into the form of an ellipsoid (or an ellipse, if in two dimensions). Except in the unnatural special cases where the principal axes for the stretching exactly coincide with the coordinate grid directions, we cannot achieve the desired stretching without including nonstandard grid lines among the set of directions along which recursive smoothing operators apply. In three dimensions, the description of a general linear stretching involves six independent components of the symmetric aspect tensor defining the spatial second moments. The essentially additive property of second moments under composition by spatially unbiased filters (which is an exact result in the case of spatially homogeneous smoothers, and a good approximation in most other cases) allows the six independent aspect tensor components to be resolved into a hexad of generalized grid lines and their associated one-dimensional second moments of dispersion. A special convention for choosing this hexad, which we will briefly describe, ensures that this resolution of the aspect tensor is essentially unique.

a. Definition of a feasible hexad

On a grid represented by any 3-vector of integers, the directions of a feasible hexad are collectively the set

TABLE 1. Components of the hexads **g** and **g′** discussed in the examples of section 4.

p	$(\mathbf{g}_p)^{\mathrm{T}}$	$(\mathbf{g}_p')^{\mathrm{T}}$
1	(1, −1, 0)	(1, 1, −1)
2	(0, 0, 1)	(0, 0, 1)
3	(1, 0, −1)	(0, 1, −1)
4	(0, −1, 0)	(1, 0, −1)
5	(1, 0, 0)	(0, 1, 0)
6	(0, 1, −1)	(−1, 0, 0)

generated by the integer displacements, or "generators," $\mathbf{g}_p$, for $p = 1, \ldots 6$, where the triple product

$$[\mathbf{g}_1, \mathbf{g}_1, \mathbf{g}_1] = 1, \quad \text{and} \tag{4.1}$$

$$\mathbf{g}_2 = \mathbf{g}_5 - \mathbf{g}_3, \tag{4.2a}$$

$$\mathbf{g}_4 = \mathbf{g}_1 - \mathbf{g}_5, \quad \text{and} \tag{4.2b}$$

$$\mathbf{g}_6 = \mathbf{g}_3 - \mathbf{g}_1. \tag{4.2c}$$

In the context of such a hexad it requires only linear analysis to resolve the aspect tensor components into the additive components associated with each of the six generators of a given hexad. For example, suppose the intended aspect tensor at a particular grid location is given by some symmetrix matrix, **A**, and, adopting and extending the notation established in Part I, let the six grid-relative smoothing scales associated with the given hexad of generators, $\mathbf{g}_p$, be denoted σ_p. A *unit* value for such a smoothing scale assigned to a direction defined by a generator **g** would result in a contribution to the aspect tensor of $\mathbf{gg}^{\mathrm{T}}$, and therefore, by the additive property of the contributions from the sequence of six smoothing operations with scales σ_p, we shall find that the final aspect tensor obtained at the end of this sequence will be

$$\mathbf{A} = \sum_p^6 \sigma_p^2 \mathbf{g}_p (\mathbf{g}_p)^{\mathrm{T}}. \tag{4.3}$$

Formally, since there are six independent components of **A**, this tensor equation can be rearranged as a set of six linear equations for the six unknowns σ_p^2 forced by the "right hand side" vector, **A**, of the independent components of **A** through the action of the appropriate 6 × 6 matrix constructed from the corresponding components of the terms $\mathbf{g}_p \mathbf{g}_p^{\mathrm{T}}$. For example, suppose we are given the hexad whose generators are defined by the components of the set **g** listed in Table 1, and an aspect tensor:

$$\mathbf{A} \equiv \begin{bmatrix} 3 & 1 & -2 \\ 1 & 3 & -2 \\ -2 & -2 & 4 \end{bmatrix}. \tag{4.4}$$

Putting the independent components of **A** into the 6-vector, **A**, according to

$$\mathbf{A}_1 = \mathbf{A}_{1,1}, \tag{4.5a}$$

$$\mathbf{A}_2 = \mathbf{A}_{2,2}, \tag{4.5b}$$

$$\mathbf{A}_3 = \mathbf{A}_{3,3}, \tag{4.5c}$$

$$\mathbf{A}_4 = \mathbf{A}_{2,3}, \tag{4.5d}$$

$$\mathbf{A}_5 = \mathbf{A}_{1,3}, \quad \text{and} \tag{4.5e}$$

$$\mathbf{A}_6 = \mathbf{A}_{1,2}, \tag{4.5f}$$

we resolve the aspect tensor into this hexad's smoothing quantities, σ_p^2, through the following matrix equation:

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_3^2 \\ \sigma_4^2 \\ \sigma_5^2 \\ \sigma_6^2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 4 \\ -2 \\ -2 \\ 1 \end{bmatrix}. \tag{4.6}$$

Each column of the square matrix in (4.6) comes from the six independent components, taken in the same order as in (4.5a)–(4.5f), of the tensor $\mathbf{g}_p \mathbf{g}_p^{\mathrm{T}}$ for the corresponding generator. For example, the independent components of

$$\mathbf{g}_1 \mathbf{g}_1^{\mathrm{T}} \equiv \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.7}$$

are gathered into the vector $[1, 1, 0, 0, 0, -1]^{\mathrm{T}}$ forming the first column of the matrix in (4.6). For this example, the solution of (4.6) is

$$\{\sigma_p^2\} = [-1, 0, 2, 2, 2, 2]. \tag{4.8}$$

However, for the hexad to have validity as a smoother, all six of these projected components σ_p^2 must also be non-negative, which is evidently not the case for the hexad of generators in the present example, since the first component $\sigma_1^2 < 0$. Although the proof is beyond the scope of this paper, it can be shown that for every positive-definite aspect tensor there is always, and only, one way to associate positive projected components with the directions of a feasible hexad. The following subsection describes the procedure by which the required hexad is found.

b. The hexad algorithm

Geometrically, any feasible hexad of generators, together with their antipodes, constitute a set of 12 points whose convex hull forms a linearly transformed *cuboctahedron,* two examples of which are shown in Fig. 3. [A cuboctahedron is the semiregular polyhedron formed by slicing off the corners of a cube between its edge midpoints to leave a boundary of eight equilateral triangles and six squares; e.g., see Coxeter (1973, 18–19).] In seeking the valid hexad for a given aspect tensor,
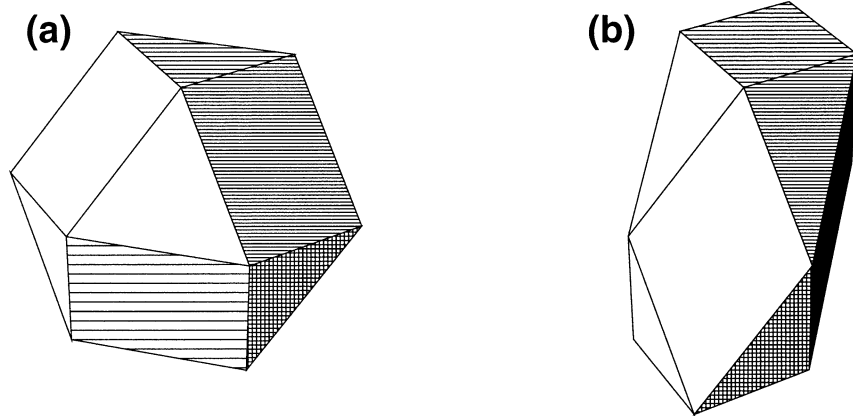
FIG. 3. Geometric depiction of an iterative step in the hexad algorithm, by which one skewed cuboctahedron in the grid has one antipodal pair of its 12 vertices (defining one of the active generalized grid lines) replaced by another pair, forming another skewed cuboctahedron. The (a) before and (b) after pictures show that the topological configurations remain equivalent although disjoint regions of the space of aspect tensors are made accessible by positive smoothing confined to the two respective hexads of generalized grid lines. Generically, a given aspect tensor can be resolved into positive line-smoothing operations associated with only one hexad of this form.

when an invalid trial hexad (Fig. 3a, say) projects a *negative* component of this tensor onto one of its six diameters, we step closer (e.g., Fig. 3b) to the sought-after valid hexad by replacing only the generator associated with the offending line with a generator of the only possible alternative line that the defining hexad rules permit. The stepwise replacement of one trial hexad configuration by another therefore has a nice geometrical interpretation, which is illustrated in Fig. 3 by the mutation from panel a to b. We see that the pair of vertices at the extreme right and left of Fig. 3a are absent in the configuration of Fig. 3b, while the extreme upper and lower vertices shown in Fig. 3b are absent in Fig. 3a. If the vertex at the extreme right of Fig. 3a had the designation, $\mathbf{g}_1$, with other vertices assigned so as to conform to the rules, (4.1) and (4.2), then one way of assigning labels to the vertices of the configuration in Fig. 3b, also consistent with these rules, is to define the new generators (primed) as follows:

$$\mathbf{g}_1' = \mathbf{g}_3 - \mathbf{g}_4, \tag{4.9a}$$

$$\mathbf{g}_2' = \mathbf{g}_2, \tag{4.9b}$$

$$\mathbf{g}_3' = \mathbf{g}_6, \tag{4.9c}$$

$$\mathbf{g}_4' = \mathbf{g}_3, \tag{4.9d}$$

$$\mathbf{g}_5' = -\mathbf{g}_4, \quad \text{and} \tag{4.9e}$$

$$\mathbf{g}_6' = -\mathbf{g}_5. \tag{4.9f}$$

Corresponding rules are easily found for the other five transition possibilities. In the case of the numerical example introduced in the last subsection, where it was indeed found that generator $\mathbf{g}_1$ led to a corresponding σ_1^2 with negative value, the replacement of $\mathbf{g}_1$ by the alternative, $\mathbf{g}_1'$, in accordance with these prescribed rules

leads to the new hexad defined in the right column of Table 1. In this case, when we resolve the aspect tensor into the components σ_p^2 we obtain

$$\{\sigma_p^2\} = [1, 1, 1, 1, 1, 1], \tag{4.10}$$

whose components, all being positive, confirm that the new hexad in this example *is* a valid one. Note that Fig. 3 may be regarded as a pictorial representation of our one-step hexad iteration viewed approximately from the direction of the vector, (1, 1, 1).

In practice, a short chain of such iterations will usually suffice to locate the given aspect tensor's unique valid hexad (and there is an analogous "triad algorithm" for the two-dimensional case). The resolution of a given aspect tensor into its equivalent hexad is carried out at every analysis grid point and the six non-negative values of σ_p^2 that are obtained at each point can then be used to prescribe the coefficients of the recursive filter that produces this degree of dispersion along the generalized grid line that the corresponding hexad generator defines. Since the generators at one grid point are typically largely the same ones at the neighboring grid points, it becomes possible to link together the smoothing step in each particular generalized grid direction to form continuous nontrivial line segments. It then becomes desirable to define a procedure for smoothing along such segments in an efficient way that avoids the application of the filter along a line in one direction interfering with the application along another intersecting line. This matter is addressed in the following section.

c. The "chromatic hexad"

Each grid point belongs to six distinct line segments along which filters are required to act. It is clearly un-

desirable to have recursive filter operators acting concurrently on two or more of the line segments dictated by the hexad algorithm when the line segments in question intersect at some shared grid point, because the outcome of the first of the filters to reach this shared point will then interfere with the action of the other filters that follow. However, it is equally undesirable to restrict the algorithm to a purely sequential process when state-of-the-art mainframe computers are now massively parallel. Fortunately, it is possible to "color code" each of the generators of the grid and, hence, the various line segments on which the filters operate, so that every hexad consists of six different "colors" from a total palette of seven. Since no two segments of the same color can ever intersect, this means that, as long as each color is dealt with sequentially by the filtering algorithm, no conflicts can arise. To see how the coloring assignment comes about, we associate the three components (g_x, g_y, g_z) of a given generator **g**, with the corresponding binary digits $(\hat{g}_x, \hat{g}_y, \hat{g}_z)$, where each $\hat{g}$ is 0 or 1 according to whether the corresponding g is even or odd. It is easy to show that the generators of a feasible hexad cannot all be even, which allows us to assign seven colors to the combinations. Moreover, we find that the colors assigned to the generators of any valid hexad are all different; we conveniently assign the hexad itself the color missing from its generators. A useful variant of the iterative procedure, the "chromatic hexad" algorithm supplies, for a given aspect tensor, the correct hexad, the corresponding six smoothing coefficients, and the assigned colors to allow efficient scheduling of the smoothing operations on a parallel computer.

Preliminary results obtained with the anisotropic covariances generated with the hexad algorithm are given in Parrish and Purser (1998). Further applications and refinements of the method will be described in future publications.

5. Discussion

We have extended the methods of Part I to show how recursive filters can be used to construct covariances with shapes more general than purely Gaussian. The problem of efficiently accommodating approximately isotropic but spatially inhomogeneous covariance functions in a variational analysis has also been solved using recursive numerical filters. This approach may be regarded as an alternative to that of Gaspari and Cohn (1998, 1999), who use direct calculation of the covariances but achieve numerical efficiency by restricting the approximations to Gaussian forms to having compact support. In our case, the covariances are never explicitly computed; instead, their effects as convolution operators are represented, through a sequence of applications of carefully designed recursive filters operating along the various lines of the appropriately chosen computational grids. In a regional analysis, there is no reason not to use the grid of the intended numerical prediction model. In a global context, where the usual latitude and longitude grid possesses polar singularities, we may either adopt the special methods for polar grids discussed in section 3b or, by invoking additional interpolations, cover the global domain in overlapping maps, each of which being furnished by an appropriate Cartesian grid. For example, we can adopt square Cartesian grids embedded in the respective polar stereographic projections for the polar cap regions and a Mercator grid elsewhere, in order to preserve the property of local isotropy, and use the multi-Gaussian methods of synthesis to provide the necessary control over the horizontal scale (needed to compensate for the map-scaling factor, if nothing else). Experiments reveal no evidence that the analysis results are significantly degraded by adopting grids that differ from the model grid, as long as the conversions between them are by high-order accurate interpolations. However, this synthetic method results in non-Gaussian covariances, even when Gaussians are preferred, and, since we must account for the cost of the additional grid-to-grid interpolations, it can be more expensive than adopting the special procedures of section 3b. For a global analysis, the user must choose the method best adapted to his or her requirements. These matters are dealt with in greater detail in Wu et al. (2002), which successfully applies these recursive filter methods to a global analysis of real meteorological data.

An additional development that we have described in section 4 is the further generalization of the covariance operators to accommodate fully anisotropic effects. Recent approaches to three-dimensional data assimilation where it is *not* assumed that the covariances must be locally isotropic have been reported by Desroziers (1997) and by Riishøjgaard (1998), and objective statistical methods for estimating the parameters of anisotropy from the data themselves are suggested by the work of Dee and da Silva (1999) and Purser and Parrish (2003). We expect that anisotropic covariances will play a much more significant role in variational data assimilation in the future.

APPENDIX

Amplitude Estimation for Inhomogeneous Quasi-Gaussian Filters

The control of amplitude (variance) of the covariance filter is quite straightforward when the filter is spatially homogeneous but, in the inhomogeneous case, the response function is no longer simply a Gaussian and an error is therefore incurred when amplitudes are estimated purely on the basis of the Gaussian formula. Fortunately, when the modulation of the filter's smoothing scale occurs slowly and smoothly across the domain, it becomes possible to improve upon the Gaussian amplitude formula by taking into account the local variation of the smoothing parameters through the application of an asymptotic analysis. We present an outline of this method as it applies to ''first order'' perturbations of scale in one dimension and we employ the diffusion model to represent the overall effect of the filter. However, we will find that this theory generalizes to diffusion in higher dimensions, making this work relevant also to the approaches of Derber and Rosati (1989) and Weaver and Courtier (2001).

With a ''diffusivity'' $D(x)$, the outcome of the spatially inhomogeneous filter is identified with the application for ''time'' t of the diffusion equation:

$$\dot{\psi} \equiv \frac{d\psi}{dt} = \frac{d}{dx}D\frac{d\psi}{dx} \equiv (D\psi')'. \quad (A.1)$$

The local variation of D is described by expanding it as a series about the origin:

$$D(x) = D_0 + D_1x + D_2x^2 + \dots . \quad (A.2)$$

It is convenient to write the evolving solution in the following form:

$$\psi(x, t) = \exp[g(x, t)], \quad (A.3)$$

where g is expressed as a series,

$$g(x, t) = g_0(t) + g_1(t)x + g_2(t)x^2 + \dots . \quad (A.4)$$

For uniform $D = D_0$, the solution that starts with a unit impulse at $x = 0$ and $t = 0$ has, at future time,

$$g_0 = -\frac{1}{2}\log t - \frac{1}{2}\log(4\pi D_0) \quad \text{and} \quad (A.5)$$

$$g_2 = -\frac{1}{4D_0 t}, \quad (A.6)$$

with other coefficients g_k vanishing. We seek to determine the principal effect on the amplitude term, g_0, of small variations in D associated with nonvanishing coefficients D_k for $k > 0$.

Equating powers of x in the evolution equation for g implied by (A.1) and (A.3),

$$\dot{g} = (Dg')' + D(g')^2, \quad (A.7)$$

we obtain, after some algebra,

$$\hat{g}_k = \sum_{j=0}^{k+1} \left[(k + 1)h_{k+2-j} + \sum_{i=0}^{k+1-j} h_i h_{k+2-j-i} \right] D_j, \quad (A.8)$$

where

$$h_k \equiv kg_k. \quad (A.9)$$

By symmetry, the terms D_k with odd k cannot affect the amplitude to first order in their magnitudes. Thus, we are able to consider expansions in only even powers of x for D and hence for g in order to obtain the first-order effect of inhomogeneity on amplitude. Expanding (A.8) for the first few even powers of x,

$$\dot{g}_0 = 2g_2D_0, \quad (A.10)$$

$$\dot{g}_2 = 12g_4D_0 + 6g_2D_2 + 4g_2g_2D_0, \quad \text{and} \quad (A.11)$$

$$\dot{g}_4 = 30g_6D_0 + 20g_4D_2 + 10g_2D_4 + 16g_2g_4D_0 + 4g_2g_2D_2. \quad (A.12)$$

We may further expand g_2, g_4, etc., in powers of t starting with terms in t^{-1}; for example,

$$g_2(t) \equiv g_{2,-1}t^{-1} + g_{2,0} + g_{2,1}t + \dots . \quad (A.13)$$

From terms t^{-2} in (A.11),

$$g_{2,-1} = -\frac{1}{4D_0}, \quad (A.14)$$

and from corresponding terms in (A.12),

$$g_{4,-1} = \frac{D_2}{12D_0^2}. \quad (A.15)$$

With these substitutions, terms t^{-1} in (A.11) imply

$$g_{2,0} = -\frac{D_2}{4D_0}. \quad (A.16)$$

Then, integrating (A.10), we obtain

$$g_0(t) = -\frac{1}{2}\log t + g_{0,0} - \frac{1}{2}D_2t + O(t^2). \quad (A.17)$$

As in the case for uniform diffusivity, the normalization of the initial unit impulse requires that

$$g_{0,0} = -\frac{1}{2}\log(4\pi D_0), \quad (A.18)$$

but if we absorb the first-order perturbation to g_0 in the form of an ''effective'' Gaussian model's diffusivity, $\overline{D}$, that is, by equating

$$-\frac{1}{2}\log(4\pi\overline{D}) \equiv -\frac{1}{2}\log(4\pi D_0) - \frac{1}{2}D_2t + O(t^2), \quad (A.19)$$

then,

$$\overline{D} \approx D_0\exp(D_2t), \quad (A.20)$$

$$\equiv D_0 + D_2D_0t + O(t^2). \quad (A.21)$$

Owing to the properties

$$\int \psi(x, t/2) \, dx = 1 \quad \text{and} \qquad (A.22)$$

$$\int \psi(x, t/2)x^2 \, dx \approx D_0 t, \qquad (A.23)$$

we may employ the approximation

$$\overline{D} \approx \int \psi(x, t/2)D(x) \, dx \qquad (A.24)$$

to acquire a serviceable and robust effective diffusivity, $\overline{D}$, in the Gaussian amplitude formula valid at $x = 0$, which will largely compensate for the errors caused by the inhomogeneity. But, generalizing this result to other locations x is equivalent to applying the ''square root'' filter to the field $D(x)$, for which an adequate non-self-adjoint representation is available as one-half of the general construction of self-adjoint filters that we have described.

In higher dimensions the Gaussian amplitude factor now comes from the determinant $|D|$ of the tensorial diffusivity, D (which is half the aspect tensor when the diffusion acts for a unit time). Therefore, the first-order correction can still be found by smoothing $|D|$ with the square root filter, or a good approximation to it. In practice, the complete quasi-Gaussian smoother is conveniently synthesized in two halves to ensure self-adjointness; the first half being a sequence of the recursive filters applied in each of the necessary directions, the second half being the reverse sequence of the adjoints of each of these basic filters. A good enough approximation to the square root filter we require for the amplitude refinements, where the filter is *not* required to be self-adjoint, is then simply the first ''half'' of the complete self-adjoint smoothing filter.

REFERENCES

Benjamin, S. G., 1989: An isentropic meso-α-scale analysis system and its sensitivity to aircraft and surface observations. *Mon. Wea. Rev.,* **117,** 1586–1603.

Coxeter, H. S. M., 1973: *Regular Polytopes.* Dover, 321 pp.

da Silva, A., J. Pfaendtner, J. Guo, M. Sienkiewicz, and S. Cohn, 1995: Assessing the effects of data selection with DAO's physical-space statistical analysis system. Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, WMO Tech. Rep. TD 651, 273–278.

Dee, D. P., and A. M. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.,* **127,** 1822–1834.

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

Desroziers, G., 1997: A coordinate change for data assimilation in spherical geometry of frontal structure. *Mon. Wea. Rev.,* **125,** 3030–3038.

Fujita, T., and R. J. Purser, 2001: Parallel implementation of compact numerical schemes. NOAA/NCEP Office Note 434, 34 pp. [Available from NCEP, 5200 Auth Rd., Camp Springs, MD 20746.]

Gaspari, G., and S. E. Cohn, 1998: Construction of correlation functions in two and three dimensions. Office Note Series on Global Modeling and Data Assimilation, DAO Office Note 96-03R1, DAO, GSFC, 53 pp.

——, and ——, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.

Gneiting, T., 1999: Correlation functions for atmospheric data analysis. *Quart. J. Roy. Meteor. Soc.,* **125,** 2449–2464.

Hayden, C. M., and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteor.,* **34,** 3–15.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **109,** 701–721.

——, 1997: Development of an operational variational assimilation scheme. *J. Meteor. Soc. Japan,* **75,** 339–346.

Otte, T. L., N. L. Seaman, and D. R. Stauffer, 2001: A heuristic study on the importance of anisotropic error distributions in data assimilation. *Mon. Wea. Rev.,* **129,** 766–783.

Parrish, D. F., and R. J. Purser, 1998: Anisotropic covariances in 3DVAR: Application to hurricane Doppler radar observations. *Proc. HIRLAM Workshop on Variational Analysis,* Toulouse, France, Météo-France, 57–65. [Available from Met Eireann, Glasnevin Hill, Dublin 9, Ireland.]

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in Fortran 77.* 2d ed. Cambridge University Press, 933 pp.

Purser, R. J., and R. McQuigg, 1982: A successive correction analysis scheme using recursive numerical filters. Met Office Tech. Note 154, British Meteorological Office, 17 pp.

——, and D. F. Parrish, 2003: A Bayesian technique for estimating continuously varying statistical parameters of a variational assimilation. *Meteor. Atmos. Phys.,* **82,** 209–226.

——, W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2001: Numerical aspects of the application of recursive filters to variational statistical analysis. NOAA/NCEP Office Note 431, 34 pp. [Available from NCEP, 5200 Auth Rd., Camp Springs, MD 20746.]

——, ——, ——, and ——, 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.

Riishøjgaard, L.-P., 1998: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus,* **50A,** 42–57.

Schoenberg, I. J., 1938: Metric spaces and completely monotone functions. *Ann. Math.,* **39,** 811–841.

Shapiro, M. A., and J. T. Hastings, 1973: Objective cross section analysis by Hermite polynomial interpolation on isentropic surfaces. *J. Appl. Meteor.,* **12,** 753–762.

Thépaut, J.-N., P. Courtier, G. Belaud, and G. Lemaître, 1996: Dynamical structure functions in a four-dimensional variational assimilation: A case study. *Quart. J. Roy. Meteor. Soc.,* **122,** 535–561.

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

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