# DRAFT

U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road
Camp Springs, MD 20746-4304

**Office Note XXX**

# NUMERICAL ASPECTS OF THE APPLICATION OF RECURSIVE FILTERS TO VARIATIONAL ANALYSIS WITH SPATIALLY INHOMOGENEOUS COVARIANCES

R. James Purser*

Wan-Shu Wu†
General Sciences Corporation, Beltsville, Maryland

David F. Parrish‡
Environmental Modeling Center

Nigel M. Roberts
Joint Centre for Mesoscale Meteorology, University of Reading, U.K., and United Kingdom Meteorological Office

January 2001



* email: jpurser@ncep.noaa.gov
† email: wanshu.wu@noaa.gov
‡ email: dparrish@ncep.noaa.gov

**Abstract**

We describe the application of efficient numerical recursive filters to the task of convolving a spatial distribution of 'forcing' terms with a quasi-Gaussian self-adjoint smoothing kernel. In the context of variational analysis, this smoothing operation is interpreted to be a either a covariance function of background error, or a contributing component to a covariance function of non-Gaussian profile formed by the superposition of a number of such quasi-Gaussian smoothing operators. A superposition of positively-weighted quasi-Gaussian smoothers enables a useful range of covariance profiles to be synthesized which, in their idealized univariate and spatially homogeneous forms, imply power spectra that exhibit tails substantially fatter than those corresponding to the single Gaussian of approximately equivalent width. As a consequence, these synthetic covariances are more suitable statistical representations of background error than single Gaussians in the typical situations where a broad dynamical range of scales contribute significantly to this error.

A further expansion of the potential range of synthetic covariances is achieved by combinations that also involve the negative-Laplacians of the basic component smoothers, thus enabling the negatively-correlated side-lobes of the covariances typical of some background errors to be more faithfully modelled. The methods we describe are not restricted to the production of spatially-homogeneous covariances; by spatially modulating either the superposition weights or the digital filtering coefficients themselves, it becomes possible to synthesize operators consistent with the properties of covariances which display adaptive variations of amplitude, scale and profile shape across the geographical domain. This is clearly desirable when the background itself derives from earlier data whose spatial distribution exhibits marked inhomogeneities of density or quality, and it is probably desirable also in the case of varying synoptic regimes within the domain.

Among the computational aspects of the recursive filters, we treat the problems of periodic and nonperiodic boundary conditions and an approach to achieving efficient parallelization.

## 1. Introduction

There are many methods available for objectively analyzing the meteorological data required to initialize a numerical weather prediction model (for example, see Daley 1991). Those methods based on formal statistical principles (e.g., Gandin 1963, Lorenc 1986, Parrish and Derber 1992, Lorenc 1997, Courtier et al. 1998), which permit proper account to be taken of multivariate aspects of the problem, have now largely superseded the overtly empirical methods of 'successive corrections' (Bergthorssen and Döös 1955, Cressman, 1959; Barnes, 1964). Nevertheless, for specialized applications, the empirical methods continue to enjoy the advantages of greater computational efficiency and the ability to adapt more flexibly to the typically large inhomogeneities of density and quality of the available data. While the high efficiency of empirical methods becomes progressively less of a critical factor as available computational power continues to increase, adaptivity remains a factor of considerable importance in circumstances where the day-to-day variability of data quality and quantity are hard to predict before-hand, such as occurs in the processing of satellite sounding data. In this context Hayden and Purser (1995), following up on the work of Purser and McQuigg (1982), developed a numerically efficient and spatially adaptive analysis scheme using spatial smoothers. Each spatial smoother was built up of more basic numerical operators consisting of rather simple recursive filters acting unidirectionally upon the gridded data residuals.

The numerical efficiency of these basic operators can also be turned to advantage within a statistical analysis scheme, specifically in the synthesis of the effective covariance-convolution operators needed by the descent algorithms of the large-scale linear (or, at worst, weakly nonlinear) solvers involved. The Statistical Spectral Interpolation (SSI) of the National Centers for Environmental Prediction (NCEP) is an example of an analysis scheme in which the spectral representation of the background error covariance is employed directly (Parrish and Derber 1992). Methods of this type are inherently limited in their ability to deal conveniently with geographical inhomogeneities. One motivation of the present study was to develop the tool of recursive filters to allow the operational three-dimensional variational analysis (3DVAR) scheme to accommodate spatial inhomogeneities in the background covariance. Even inhomogeneous covariances must preserve the symmetry of self-adjointness upon which the success of many iterative solution algorithms depend. While this was not a property required by the smoothers employed in the empirical (nonstatistical) analysis of Hayden and Purser (1995), it can be engineered without great difficulty, as this paper will demonstrate.

A brief review of the ideas that underlie 3DVAR is given in section 2 in order to clarify the points at which the recursive filter plays a part. In section 3 we set forth the relevant theory pertaining to the construction of basic recursive filters capable of being forged into convolution operators reasonably representing the qualities desired by modelled covariance-convolutions within an adaptive analysis scheme with a uniform cartesian grid and with homogeneous covariances. Like the Gaussian covariances of Derber and Rosati (1989), which are obtained by multiple iterations of a diffusion operator, the basic recursive filters are crafted to produce approximately Gaussian smoothing kernels (but in fewer numerical operations than are typical in the explicit diffusion method). Some of the technicalities discussed in this section are treated in greater detail in the appendices. Section 4 treats some more general cases where the grid spacing may vary smoothly (but still assuming that it remains orthogonal) and where the spatial scale of the covariance function may smoothly vary (but still assuming local isotropy of the scale parameter). A discussion of the special case of polar grids is provided, with suggested approaches to overcome the difficulties associated with the polar coordinate singularity. Section 5 deals with specific proposals for the construction of non-Gaussian parameterized families of covariance models based on linear superposition of the the quasi-Gaussian 'building blocks' that the recursive filters provide. 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. In the concluding section we touch on the problems of managing the potential for spatial adaptivity achieved by the techniques presented here. We also offer some thoughts on the prospects of extending the freedom to provide inhomogeneous but essentially isotropic covariances to more general constructions, available through refinements of the filtering technique, by which even this restriction of local isotropy may be profitably relaxed.

## 2. 3DVAR

In this section we attempt to follow the notation of Ide et al. (1997), writing the abstract vector representing the atmospheric state as $\mathbf{x}$, with 'background' and 'analysis' versions of this indicated by subscripts, that is: $\mathbf{x}_b$ and $\mathbf{x}_a$. The component of error in the background is denoted $\epsilon_b$:

$$\mathbf{x}_b = \mathbf{x} + \epsilon_b. \tag{2.1}$$

The observational data are collected into another abstract vector $\mathbf{y}_o$ whose components are related to the state vector $\mathbf{x}$ through the application of a generalized, possibly nonlinear interpolation operator $\mathcal{H}$ together with an effective measurement error $\epsilon_0$:

$$\mathbf{y}_o = \mathcal{H}(\mathbf{x}) + \epsilon_o. \tag{2.2}$$

The statistical characteristics of the errors $\epsilon_b$ and $\epsilon_o$ are quite difficult to describe in complete detail owing to numerous complicating factors. For example, the statistics of these quantities are not really well described by normal distributions (otherwise there would be no need to pay special attention to quality control and nonlinear balancing); the error characteristics of the background are usually dependent upon the geographical location, the season and the synoptic situation; the observation errors are frequently biased or contain components of mutual correlation that defy simple description. Nevertheless, even a partial accounting for the statistical behavior of errors, in the form of a relatively simple statistical analysis scheme, can provide a valuable objective way to make the information from new observations available to a numerical forecasting system. Among the common simplifying assumptions, we normally assume unbiasedness:

$$\begin{aligned} \langle \epsilon_b \rangle &= 0, && (2.3) \\ \langle \epsilon_o \rangle &= 0. && (2.4) \end{aligned}$$

The covariance, $\mathbf{R} \equiv \langle \epsilon_o \epsilon_o^T \rangle$, of observational error is assumed to be diagonal, equivalent (for normal statistics, at least) to assuming the observational errors are statistically independent. The corresponding covariance, $\mathbf{B} \equiv \langle \epsilon_b \epsilon_b^T \rangle$, of background error, however, is *never* assumed to be diagonal in its representation based on the state-space constructed from the gridded value components; the characteristically smooth form in space of background errors implies that neighboring points have errors, in fields of the same type, that are strongly positively correlated. Although the principles of variational analysis can accommodate strong nonlinearities if required, it is often numerically convenient to exploit the typically weak nonlinearity of $\mathcal{H}$ by approximating the effects on $\mathcal{H}(\mathbf{x})$ of small increments of $\mathbf{x}$, using the linearization $\mathcal{H}$:

$$\mathbf{H} d\mathbf{x} = d\mathcal{H}(\mathbf{x}). \tag{2.5}$$

The 'primal' variational principle in 3DVAR seeks the minimum over $\mathbf{x}$ of the penalty function $\mathcal{L}_1(\mathbf{x})$ defined by:

$$2\mathcal{L}_1(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x})), \tag{2.6}$$

which may be justified by minimum-variance arguments or, more generally, by considerations of Bayesian estimation when nonlinearities become significant (for example, see Lorenc 1986). The solution, $\mathbf{x} = \mathbf{x}_a$, then obeys:

$$\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) = \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x})), \tag{2.7}$$

and hence, must define an increment of the form,

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{B}\mathbf{H}^T\mathbf{f}, \tag{2.8}$$

where

$$\mathbf{f} = \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x})). \tag{2.9}$$

The inherent smoothness of the background field errors $\epsilon_b$, and hence that of the covariance $\mathbf{B}$ of these errors, is therefore imprinted on the analysis increments themselves.

The dimensionality of $\mathbf{f}$ (the number of independent data) is typically much smaller than the dimensionality of $\mathbf{x}$ (the number of gridded state values). If we neglect the effects of nonlinearity (which can be accommodated easily by appropriate refinements) we find that, instead of solving for $\mathbf{x}$ directly, we can instead first solve for the smaller vector $\mathbf{f}$ in the implied 'dual' variational principle that minimizes $\mathcal{L}_2(\mathbf{f})$ defined by:

$$2\mathcal{L}_2(\mathbf{f}) = \mathbf{f}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{f} - 2\mathbf{f}^T\mathbf{d}, \tag{2.10}$$

where $\mathbf{d}$ is the 'innovation':

$$\mathbf{d} = \mathbf{y}_o - \mathcal{H}(\mathbf{x}_b). \tag{2.11}$$

This duality is discussed by Courtier (1997), who shows that the primal and dual forms imply essentially identical condition numbers for the alternative large-scale symmetric-matrix linear inversion problems they imply when only the most basic preconditioning strategies are employed in each case. More ingeneous strategies of preconditioning based on data clustering, as recently proposed by Cohn et al (1998) and Daley and Barker (2000), would seem to favor the adoption of the dual form, but consideration of some of the nonlinear aspects of the problems make the more direct estimation of analysis increments via the primal form more attractive. Regardless of which form of 3DVAR is adopted, given that the sizes of the symmetric 'system matrices' are too large to admit direct solution in either case, one must rely on iterative methods, such as conjugate gradient or quasi-Newton solvers, to converge towards a practical solution. Then one finds that the most costly part of each iterative step of such a solution algorithm is the operation of multiplying some grid-space vector $\mathbf{v}$ by the covariance matrix $\mathbf{B}$ (or at least, a sequence of operations whose net effect and cost is equivalent to performing this operation). This effort has to be expended precisely once per iteration whether treating the primal or the dual form of the problem. Even one matrix-vector multiplication of this form is prohibitively expensive to perform every iteration if it is computed explicitly with a full matrix having the dimensionality of the gridded state $\mathbf{x}$.

The difficulty is tackled by progressively reducing an operation of the form $\mathbf{Bv}$ into smaller, less costly factors. In the first step, the multivariate structure of $\mathbf{B}$ is broken apart by the judicious selection of a set of nonstandard analysis variables for which the contributions from

**B** naturally separate out. For example, a single variable representing the quasi-geostrophically balanced combination of mass- and rotational wind-fields can be attributed a univariate spatial covariance for its background errors quite independently of the corresponding spatial covariance for the residual unbalanced rotational wind component. Meanwhile, the divergent wind field can be treated independently of either. Further steps in the program of reducing the operator **B** might be, next, to carry out a crude separation of a few additive components of the operator on the basis of their characteristic spatial scales. If this can be done to render the resulting operator components into Gaussian forms, then, in the absence of anisotropies obliquely oriented with respect to the grid, the Gaussians themselves may be factored into the three respective coordinate directions. Finally, along each single dimension, a final computational economy may be gained by employing a spatially recursive filter, carefully constructed to mimic the required Gaussian convolution operator, but at a fraction of the still considerable cost of applying directly the explicit Gaussian convolution operator itself. It is the objective of the following sections to reveal precisely how such a recursive filter may be fabricated and applied. However, it is appropriate to reiterate that the virtue of a recursive filter used in this way derives merely from its inherent computational efficiency, which, owing to the unique factorization properties of multidimensional Gaussians, can *only* be exploited in two or three dimensions when the effective convolution kernels are of approximately Gaussian form. We do *not* wish to imply that the Gaussian form is inherently desirable in data assimilation. On the contrary, careful investigation of the spatial profiles of forecast background error (Thiébaux 1976; Thiébaux et al. 1986; Hollingsworth and Lönnberg 1986) reveal covariance functions that cannot be reconciled with the Gaussian shape alone. But, by treating the two- or three-dimensional quasi-Gaussian filter combination as a relatively cheap 'building block', a far larger range of possible profile shapes becomes accessible, by the superposition of appropriately weighted combinations of quasi-Gaussians of different sizes and by the application of the negative-Laplacian operator to such components in order to induce the negatively-correlated side-lobes characteristic of some components of background error. Thus, the motivating consideration for using recursive filters in this context is predominantly that of computational efficiency together with the recognition that more much general forms become available through the exploitation of superposition.

## 3. Homogeneous recursive filtering theory

### (*a*) *Quasi-Gaussian recursive filters in one dimension*

Let $K/\delta x^2$ denote the finite difference operator:

$$K(\psi)_i/\delta x^2 = -(\psi_{i-1} - 2\psi_i + \psi_{i+1})/\delta x^2, \tag{3.1}$$

approximating the differential operator, $-d^2/dx^2$, on a line-grid of uniform spacing $\delta x$. The spectral representation of the operator at wavenumber $k$ (wavelength $2\pi/k$) is

$$\hat{K}(k) = \left(2\sin\left(\frac{k\delta x}{2}\right)\right)^2 .$$

Inverting this relationship, we obtain a formula for $k^2$ in terms of, $\hat{K}$:

$$k^2 = \frac{4}{\delta x^2}\left(\arcsin\left(\frac{\hat{K}^{1/2}}{2}\right)\right)^2 .$$

Clearly, the same formula relates operator $-d^2/dx^2$ to operator $K$; in fact, the algebraic manipulations we set forth here can be regarded as an application of the 'calculus of operators' (Dahlquist and Björck 1974, p. 311). Using the standard expansion:

$$\arcsin(z) = \sum_{i=0}^{\infty} \gamma_i z^{2i+1}, \quad |z| < 1 \tag{3.2}$$

where

$$\gamma_i = \frac{1}{(2i+1)}.\frac{(2i-1)!!}{(2i)!!} \equiv 1, \qquad \frac{1}{3}.\frac{1}{2}, \qquad \frac{1}{5}.\frac{1.3}{2.4}, \qquad \frac{1}{7}.\frac{1.3.5}{2.4.6}, \qquad \ldots \tag{3.3}$$

we may obtain a power expansion for $k^2\delta x^2$, and thence, the expansions for the term, $(k^2\delta x^2)^i$:

$$(k^2\delta x^2)^i = \sum_{j\geq i} b_{i,j}(\hat{K})^j. \tag{3.4}$$

The coefficients, $b_{i,j}$, which are all positive and rational, are listed in Table 1 for $j \leq 6$.

TABLE 1. COEFFICIENTS $b_{i,j}$ FOR QUASI-GAUSSIAN FILTERS UP TO DEGREE SIX

| $i$ | $j=1$ | $j=2$ | $j=3$ | $j=4$ | $j=5$ | $j=6$ |
|---|---|---|---|---|---|---|
| 1 | 1 | $\frac{1}{12}$ | $\frac{1}{90}$ | $\frac{1}{560}$ | $\frac{1}{3150}$ | $\frac{1}{16632}$ |
| 2 | | 1 | $\frac{1}{6}$ | $\frac{7}{240}$ | $\frac{41}{7560}$ | $\frac{479}{453600}$ |
| 3 | | | 1 | $\frac{1}{4}$ | $\frac{13}{240}$ | $\frac{695}{60480}$ |
| 4 | | | | 1 | $\frac{1}{3}$ | $\frac{31}{360}$ |
| 5 | | | | | 1 | $\frac{5}{12}$ |
| 6 | | | | | | 1 |

Consider the differential operator, $\mathcal{D}_{(n)}$,

$$\mathcal{D}_{(n)} = 1 - \frac{a^2}{2}\frac{d^2}{dx^2} + \frac{a^4}{2!4}\frac{d^4}{dx^4}\cdots + \frac{1}{n!}\left(-\frac{a^2}{2}\frac{d^2}{dx^2}\right)^n, \tag{3.5}$$

whose spectral representation is:

$$\hat{\mathcal{D}}_{(n)} \equiv 1 + \frac{a^2k^2}{2} + \frac{a^4k^4}{2!4}\cdots + \frac{1}{n!}\left(\frac{a^2k^2}{2}\right)^n$$

or, with $\sigma = a/\delta x$,

$$\hat{\mathcal{D}}_{(n)} = 1 + \sigma^2\left(\frac{k^2\delta x^2}{2}\right) + \frac{\sigma^4}{2!}\left(\frac{k^2\delta x^2}{2}\right)^2 \cdots + \frac{\sigma^{2n}}{n!}\left(\frac{k^2\delta x^2}{2}\right)^n. \tag{3.6}$$

Since,

$$\lim_{n\to\infty} \hat{\mathcal{D}}_{(n)} = \exp\left(\frac{a^2k^2}{2}\right), \tag{3.7}$$

the substitution of each power series (3.4) up to degree $n$ for the powers of $k^2$ into (3.6) gives us a way of approximating this exponential function in terms of $\hat{K}$. Correspondingly, there is a finite difference operator, $\mathcal{D}^*_{(n)}$, composed of the $n$th-degree expansion of $K$ implied by this approximation:

$$\mathcal{D}^*_{(n)} = 1 + b_{1,1}\frac{\sigma^2}{2}K + \left[b_{1,2}\left(\frac{\sigma^2}{2}\right) + \frac{b_{2,2}}{2!}\left(\frac{\sigma^2}{2}\right)^2\right]K^2 \ldots + \left[\sum_{j=1}^{n}\frac{b_{j,n}}{j!}\left(\frac{\sigma^2}{2}\right)^j\right]K^n. \tag{3.8}$$

That is,

$$\mathcal{D}^*_{(n)} = \exp\left(\frac{\sigma^2}{2}\sum_{j=1}^{\infty} b_{1,j}K^j\right) + \mathcal{O}(K^{n+1}). \tag{3.9}$$

Note that, owing to the positivity of all the coefficients $b_{i,j}$, this operator is positive definite and therefore possesses a well-defined inverse. Note also that, for $\sigma \gg 1$, the only coefficients in (3.8) that remain significant are the 'diagonal' ones, $b_{i,i} = 1$, yielding simply the truncated Taylor series for the exponential function of $\sigma^2 K/2$. Shortly, we shall examine the practical impact of the off-diagonal components, $b_{i,j},\ j > i$, but first we describe the process of extracting from the above algebraic developments a practical class of smoothing filters.

The reciprocal of the function $\exp(a^2k^2/2)$ in (3.7) is a Gaussian function in $k$ and is the Fourier transform of a convolution operator (on the line $x$) whose kernel is also of Gaussian form. Provided we can find a practical way to invert the operator equation,

$$\mathcal{D}^*_{(n)}\mathbf{s} = \mathbf{p}, \tag{3.10}$$

for a given input distribution, $\mathbf{p}$, the resulting output, $\mathbf{s}$, will be an approximation to the convolution of $\mathbf{p}$ by the Gaussian function whose spectral transform is the reciprocal of the right-hand-side of (3.7). The approximation, $(\mathcal{D}^*_{(n)})^{-1}$, to this convolution is what we refer to as a 'quasi-Gaussian filter'. The common centered second-moment of operator $\mathcal{D}_{(n)}$ and its approximation, $\mathcal{D}^*_{(n)}$, is exactly $-a^2$, so $a$ is a convenient measure of the intrinsic distance scale of the smoothing filter implied by the inversion of (3.10). A useful fact is that the square of the intrinsic scale of the composition of sequential smoothing filters is the sum of squares of the scales of the individual components. Also, as a consequence of the statisticians' 'central limit theorem' (e.g., Wilks 1995) applied to convolutions in general, the effective convolution-kernel of such a composition of several identical filter-factors resembles a Gaussian more closely than does the representative factor. Thus, provided it becomes feasible to invert (3.10), we possess the means to convolve a gridded input distribution with a smooth quasi-Gaussian kernel, at least in one dimension.

As a matrix, $\mathcal{D}^*_{(n)}$ is banded and, for an infinite domain, symmetric. Conventionally, the linear inversion of a system such as (3.10) might be effected by employing an LU factorization (Dahlquist and Björck 1974) of $\mathcal{D}_{(n)}$,

$$\mathcal{D}_{(n)} \equiv \mathcal{A}\mathcal{B}, \tag{3.11}$$

with lower-triangular band matrix, $\mathcal{A}$ and upper-triangular band matrix $\mathcal{B}$, allowing the solution to proceed as two steps of recursive substitution. On an infinite grid, the same principle pertains, but with the guaranteed simplification of: (i) a translational symmetry ensuring that every row of $\mathcal{A}$ is identical (allowing for the trivial translation) and every row of $\mathcal{B}$ is identical; (ii) ordinary matrix symmetry by which we can ensure that $\mathcal{B}$ is simply the transpose of $\mathcal{A}$. In this case, the LU decomposition of $\mathcal{D}_{(n)}$ is also of the symmetric, or Cholesky type (Dahlquist and Björck 1974).

In the two stages of solution,

$$\begin{aligned} \mathcal{A}\mathbf{q} &= \mathbf{p}, && (3.12)\\ \mathcal{B}\mathbf{s} &= \mathbf{q}, && (3.13) \end{aligned}$$

the explicit recursions of the back-substitutions are the following basic recursive filters:

$$\begin{aligned} q_i &= \beta p_i + \sum_{j=1}^{n} \alpha_j q_{i-j}, && (3.14)\\ s_i &= \beta q_i + \sum_{j=1}^{n} \alpha_j s_{i+j}, && (3.15) \end{aligned}$$

which are conveniently referred to as the 'advancing' and 'backing' steps respectively since, in the first, index $i$ must be treated in increasing order while, in the second, it must be treated in decreasing order, in order that the terms on the right are already available at each step. Note that the correspondences between notations of (3.12), (3.13) and of (3.14), (3.15) are:

$$\begin{aligned} A_{i,i} = B_{i,i} &= 1/\beta, && (3.16)\\ A_{i,i-j} = B_{i,i+j} &= -\alpha_j/\beta,\ j \in [1, n]. && (3.17) \end{aligned}$$

Defining the total 'substance' implied by the distribution $\mathbf{p}$ to be $\sum_i \delta x\, p_i$, the operator $\mathcal{D}^*_{(n)}$, and hence its inverse, preserve this quantity. By symmetry, the factor operators, $\mathcal{A}$ and $\mathcal{B}$ must therefore also preserve substance, implying that:

$$\beta = 1 - \sum_{j=1}^{n} \alpha_j. \qquad (3.18)$$

The task of distilling the coefficients $\alpha_j$ from the parameters defining $\mathcal{D}^*_{(n)}$ is somewhat technical and is relegated to Appendix A. Filters may be constructed at different orders, $n$, and with or without the refinements implied by the off-diagonal coefficients, $B_{i,j}$, for $j > i$. For $n = 1$ the filter response comprises back-to-back decreasing exponential functions, which Fig. 1a shows (dashed curve) in comparison to the Gaussian function (solid curve) of the same width of one grid unit. Better approximations to the Gaussian are obtained after application of the second-order filters, as shown in Fig. 1b, and fourth-order filters, shown in Fig. 1c, for the case of the filter with only the diagonal coefficients $b$ (short-dashed curves) and with all the $b$-coefficients (long-dashed curves). We see that the advantage of keeping all the coefficients is greater at higher order, where they make the resulting filter response a significantly better approximation to the intended Gaussian function. However, the alternative treatments of the $b$ coefficients are virtually indistinguishable at smoothing scales of a few grid units, as the truncation errors of the component numerical derivative operators become insignificant in comparison to the error resulting from the finite truncation of the series for the Gaussian employed in the construction of the filter operator. The cost of applying the filters with or without the off-diagonal $b$ coefficients is the same; therefore, we always adopt the more accurate formulation that includes the off-diagonal coefficients.

We have described the idealized case of operators acting on data extending indefinitely in both directions. In practice, we are confronted with geometrical constraints, either in the form of definite lateral boundaries to the domain, or as periodic conditions appropriate to a cyclic domain. Fortunately, it is possible to generalize the application of the advancing and backing recursive filters to both of these situations. Appendix B treats the case of lateral boundaries and shows how the effect of a continuation of the domain to infinity can be simulated by the imposition of appropriate 'turning' conditions at the transition between the advancing and backing stages. Appendix C treats the case of periodic boundary conditions. In both of these special cases the main part of the filtering algorithm and the basic filter coefficients employed are the same as in the case of the infinite domain. By a generalization of the treatment used in the cyclic case, one may efficiently distribute the recursions across multiple processors of a massively parallel computer, as we describe in appendix D.

### (*b*) *Quasi-Gaussian filters in two dimensions*

Let $x$ and $y$ be horizontal Cartesian coordinates, $k$ and $l$ the associated wavenumber components. Then in two dimensions, we can exploit the factoring property of isotropic Gaussians:

$$\exp\left(-\frac{a^2\rho^2}{2}\right) = \exp\left(-\frac{a^2k^2}{2}\right)\exp\left(-\frac{a^2l^2}{2}\right), \tag{3.19}$$

where $\rho = (k^2 + l^2)^{1/2}$ is the total wavenumber. In terms of basic one-dimensional Gaussian smoothing filters, $\mathcal{D}^{(x)}_{(\infty)}$ and $\mathcal{D}^{(y)}_{(\infty)}$, operating in the $x$ and $y$ directions, a two-dimensional *isotropic* filter, $G_a$, also of Gaussian form, results from the successive application of the one-dimensional factors, $\mathcal{D}^{(x)}_{(\infty)}$ and $\mathcal{D}^{(y)}_{(\infty)}$. For example, an input field, $\chi$, is smoothed to produce the output field, $\psi$, by the convolution:

$$\psi(\mathbf{x}_1) = \int\int G_a(\mathbf{x}_1, \mathbf{x}_2)\chi(\mathbf{x}_2)dx_2\, dy_2 \equiv G_a \star \chi, \tag{3.20}$$

where

$$G_a = \mathcal{D}^{(y)}_{(\infty)} \star \mathcal{D}^{(x)}_{(\infty)}$$

The crucial significance of the Gaussian form for the one dimensional filters is that this form is the *only* shape which, upon combination by convolution in the $x$ and $y$ directions, produces an *isotropic* product filter. In order to generalize our filters to alternative shapes, while preserving two dimensional isotropy, we shall always attempt to base the construction of the more general filters on the 'building blocks' supplied by the quasi-Gaussian products of the approximations, $\mathcal{D}^{(x)}_{(\infty)}$ and $\mathcal{D}^{(y)}_{(\infty)}$, to the true Gaussian smoothers. But we must first establish what is the minimum order of such a filter that will preserve the isotropy of the product combination, at least to a degree that ensures that any residual anisotropies are not obtrusively obvious.

Fig. 2 depicts the results obtained by smoothing a delta function placed at the center of a square grid. Fig. 2a shows the result of a single application of the first-order filter, $\mathcal{D}_{(1)}$, in the $x$ and $y$ directions. This result is clearly neither smooth nor even approximately isotropic. Figs. 2b and 2c show the results obtained by using the filters of orders two and four. We see that the appearance of isotropy is not adequately attained until the order exceeds two, but the fourth-order filter shown in Fig. 2c seems to provide an excellent approximation to the isotropic Gaussian. For applications in data assimilation, it is usually worth the cost of applying a filter of at least fourth-order if the filter is to be applied only once in each of the orthogonal grid directions. For a roughly equivalent cost, one may also apply the simple first-order filter four times in succession (but with a scale only a half as large in each instance); the result is shown in Fig. 2d, but is clearly inferior to the use of the single fourth-order filter.

Very often, the physical variables of interest in an analysis are derivatives of the variables it is convenient to base the covariance model on. For example, covariances of the steamfunction or velocity potential (scalars) are often more convenient to handle than the derived covariances among velocity components at two locations. Since we may wish to employ the results of our filters as building blocks of such differentiated covariances, it is as well to examine the derivatives of fields analogous to those of Fig. 2. In order to permit any departures from isotropy to stand out more clearly, we take the Laplacian of the result of smoothing the delta function. Fig. 3 shows three such results (with a slightly smaller scale than was used in Fig. 2), involving single applications (in $x$ and $y$) of $\mathcal{D}_{(n)}$ with $n$ being 2, 4 and 6 in panels (a), (b) and (c). Even more so than in Fig. 2, we see that it is not until we adopt at least about fourth-order filtering that we obtain an acceptable degree of isotropy. For reference, the 'right answer' obtained using the Laplacian of the true Gaussian, $G_a$, is shown in Fig. 3d.

### (c) *Numerical robustness and multigrid refinements*

A recognized problem with high-order recursive filters (e.g., Otnes and Enochson, 1972) is their susceptibility to numerical noise, especially as the filtering scale becomes significantly larger than the grid scale. A natural remedy, in cases where the grid dimensions permit it, is to employ a 'multigrid' strategy. A general discussion of such methods can be found in Brandt (1977). Essentially, the field to be smoothed at a certain filtering scale is first transferred (by adjoint-interpolation from the initial fine grid) to a grid whose coarseness is comparable with, but still sufficiently resolves, this smoothness scale. The smoothing is performed by the high-order recursive filter on the generally somewhat coarser grid, now without risk of numerical noise, and at a numerical cost that is usually significantly less than the cost of the equivalent operation applied to the original fine grid. The resulting smooth field is finally interpolated back to the fine grid. The implied operator representing this combination of steps remains self-adjoint and, provided the order of accuracy of the interpolations is large enough, no discernable hint of roughness appears in the resulting smooth output.

The simplest multigrid structure is one in which the spacing in the successive grids doubles. Then, except for the possible overlaps (which are desirable in the case of bounded domains in order to preserve the same centered interpolation operators everywhere), each coarse grid is a subset of its finer predecessor. For cyclic domains, this simplification obviously works only when the periodic grid dimensions are divisible by powers of two. For bounded domains, the judicious use of overlaps enables one to adopt the scale-doubling arrangement without numerical restrictions on the grid dimensions. Interpolation is assumed to occur only between adjacent grids of the hierarchy. Interpolation from a coarse grid to the next finer grid in two dimensions is accomplished by passing through an intermediate stage in which one dimension is 'fine' and the other is 'coarse'. In this way, only one-dimensional interpolation operators need be considered. Assuming each coarse grid overlaps the next finer grid by a sufficient margin, all interpolations can be performed using the same standard centered formula. Table 2 lists the coefficients for mid-point interpolation from a uniform grid at (even) orders of accuracy between two and twelve. Experience suggests that sixth-order interpolations are adequate for most purposes.

TABLE 2. COEFFICIENTS $w_j$ FOR UNIFORM GRID MID-POINT INTERPOLATION AT ORDERS OF ACCURACY, $n$, UP TO 12.

| $n$ | $\gamma$ | $\gamma w_{1/2}$ | $\gamma w_{3/2}$ | $\gamma w_{5/2}$ | $\gamma w_{7/2}$ | $\gamma w_{9/2}$ | $\gamma w_{11/2}$ |
|---|---|---|---|---|---|---|---|
| 2 | 2 | 1 | | | | | |
| 4 | 16 | 9 | $-1$ | | | | |
| 6 | 256 | 150 | $-25$ | 3 | | | |
| 8 | 2048 | 1225 | $-245$ | 49 | $-5$ | | |
| 10 | 65536 | 39690 | $-8820$ | 2268 | $-405$ | 35 | |
| 12 | 524228 | 320166 | $-76230$ | 22869 | $-5445$ | 847 | $-63$ |

If a single quasi-Gaussian smoothing is to be performed, then it might seem unduly complicated and, perhaps, inefficient to perform the interpolations step-wise through the intervening hierarchy of grid scales when a single grid-to-grid interpolation would suffice. However, for the purposes of simulating a background covariance operator, the simple Gaussian form is inappropriate and the more robust and versatile covariance operators are those synthesized from *several* Gaussians drawn from a range of charactersitic scales. In this context of multi-Gaussian synthesis, the value of the multi-grid approach becomes more strikingly evident, for it not only avoids the risk of numerical noise, but also enables a broad and numerous spectrum of component Gaussian filter building blocks to be combined together in an efficient synthesis that admits considerable control of the combination's amplitude, shape and overall scale. We shall return to a more detailed discussion of this topic in section 5.

## 4. 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 which can be exploited in the case of the spatially homogeneous smoothing filters which 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.

### (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 can be rendered symmetric by a similarity transformation with a diagonal matrix related to the metrical properties of the grid. 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 (3.5) of this operator, a form of the second derivative factor is substituted which is self-adjoint even when $a$ is a function of $x$. Of the qualifying possibilities, the one that is most convenient in practice and which leads to a substance-conserving filter, is the one most closely identified with the operation of a diffusive process:

$$-\frac{d}{dx}\frac{a^2(x)}{2}\frac{d}{dx}. \tag{4.1}$$

The operator, (4.1), 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, $\tau$, whose reciprocal is the density of $x$-grid lines so that $\tau$ 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 multi-dimensionality, the operator we need to generalize (4.1) is:

$$-\frac{1}{\tau}\frac{\partial}{\partial x}\tau\frac{a^2(x)}{2}\frac{\partial}{\partial x}, \tag{4.2}$$

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

$$(s,t) \equiv \int s(x)t(x)\tau(x)dx. \tag{4.3}$$

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 $\tau$ to be available at the main and staggered grids. Define

$$\delta x_i = x_{i+1/2} - x_{i-1/2}, \tag{4.4}$$

$$\delta x_{i+1/2} = x_{i+1} - x_i. \tag{4.5}$$

Recalling that $\sigma = a/\delta x$, the simplest consistent discretization of the operator (4.2) is

$$\begin{aligned}\left(-\frac{1}{\tau}\frac{\partial}{\partial x}\tau\frac{a^2(x)}{2}\frac{\partial s}{\partial x}\right)_i &\approx \frac{1}{2\tau_i\delta x_i}\left[\tau_{i-1/2}a^2_{i-1/2}(\delta s_{i-1/2})/\delta x_{i-1/2} - \tau_{i+1/2}a^2_{i+1/2}(\delta s_{i+1/2})/\delta x_{i+1/2}\right] \\ &\equiv \frac{1}{2\nu_i}\left[\nu_{i-1/2}\sigma^2_{i-1/2}(-s_{i-1}+s_i) + \nu_{i+1/2}\sigma^2_{i+1/2}(s_i - s_{i+1})\right],\end{aligned} \tag{4.6}$$

where

$$\nu_i = \tau_i\delta x_i, \tag{4.7}$$

defines the local grid cell measure (area or volume). We can relate this operator to a tridiagonal matrix, $\mathbf{K}$, that serves to generalize the $K$ of (3.1). A representative row, $i$, of $\mathbf{K}$, is defined by:

$$K_{i,i-1} = \frac{(\nu\sigma^2)_{i-1/2}}{\sqrt{(\nu\sigma^2)_i(\nu\sigma^2)_{i-1}}}, \tag{4.8}$$

$$K_{i,i} = \frac{(\nu\sigma^2)_{i-1/2} + (\nu\sigma^2)_{i+1/2}}{(\nu\sigma^2)_i}, \tag{4.9}$$

$$K_{i,i+1} = \frac{(\nu\sigma^2)_{i+1/2}}{\sqrt{(\nu\sigma^2)_i(\nu\sigma^2)_{i+1}}}, \tag{4.10}$$

in terms of which, the finite difference operator of (4.6) is obtained:

$$\left(-\frac{1}{\tau}\frac{\partial}{\partial x}\tau\frac{a^2(x)}{2}\frac{\partial s}{\partial x}\right)_i \approx \frac{1}{2}\left(\frac{1}{\sqrt{\boldsymbol{\nu}}}\boldsymbol{\sigma}\mathbf{K}\boldsymbol{\sigma}\sqrt{\boldsymbol{\nu}}\,\mathbf{s}\right)_i \tag{4.11}$$

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

$$[K_{i,i-1}, K_i, K_{i,i+1}] \approx [-1, 2, -1] \tag{4.12}$$

very closely when $a$, $\delta x$ and $\tau$ are all smooth and slowly varying in $x$, tending to these values in the limiting case of constant $a$, $\delta x$ and $\tau$. 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 $b_{1,j}$ is:

$$-\frac{1}{\tau}\frac{\partial}{\partial x}\tau\frac{a^2(x)}{2}\frac{\partial}{\partial x} \approx \frac{1}{2}\frac{1}{\sqrt{\boldsymbol{\nu}}}\boldsymbol{\sigma}(\sum_{j=1}^{n} b_{1,j}\mathbf{K}^j)\boldsymbol{\sigma}\sqrt{\boldsymbol{\nu}} \tag{4.13}$$

Taking the exponential of this operator, but truncating all the terms comprising matrices of half-bandwidth exceeding $n$, we obtaining the sought-for generalization of $\mathcal{D}^*_{(n)}$ in (3.8):

$$\mathcal{D}^*_{(n)} \approx \frac{1}{\sqrt{\boldsymbol{\nu}}}\exp\left(\frac{\boldsymbol{\sigma}}{2}\sum_{j=1}^{n} b_{1,j}\mathbf{K}^j\boldsymbol{\sigma}\right)\sqrt{\boldsymbol{\nu}}. \tag{4.14}$$

While the coefficient-finding method of Appendix A 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 (4.14) 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 can no longer accurately simulate the indefinite continuation of the grid beyond a boundary in general, but the imperfections that result are often barely noticeable in practice.

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 for 'time' $t$ when,

$$2Dt = a^2 \tag{4.15}$$

and the impulse-response value of the result is just $(4\pi Dt)^{-1/2}$. In more than one dimension, the diffusivity generalizes to a tensor and the appropriate generalization of the impulse-response, $|4\pi Dt|^{-1/2}$ involves the determinant of $D$. But inhomogeneity of scale, which we may interpret as inhomogeneity of the effective 'diffusivity' $D$ in the diffusion analogue of our filters, leads to impulse-response functions which differ slightly from the profile calculated on the basis of the Gaussian model. An asymptotic analysis of this difference, which is outlined in Appendix E, 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 $\bar{D}$ 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 $\bar{D}$ 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 sense) of the total 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 burden in typical cases where, once constructed, it is subsequently applied to data numerous times.

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 first treat 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 Fourier transforms applied azimuthally or longitudinally to data (providing that the longitudes are uniformly spaced and of a highly-factorable number such as is required for the application of the 'Fast Fourier Transform' [FFT] algorithm). Fourier transformation invoked in the azimuthal or longitudinal direction separates the two-dimensional smoothing problem into independent one-dimensional filtering problems in the radial or latitudinal direction for the smoothing of the zonal Fourier coefficients. Suppose $\theta$ and $\lambda$ are latitude and longitude respectively. From initial data $\chi(\theta, \lambda)$ we may apply the zonal transform:

$$\hat{\chi}(\theta, m) = \int \chi(\theta, \lambda) \exp(-2\pi i m \lambda) d\lambda. \tag{4.16}$$

The idea is to redefine the operator $\mathcal{D}^*_{(n)}$ of (3.8) in a fully two-dimensional way as the polynomial of $-a^2\nabla^2$ instead of $-a^2\, d^2/dx^2$. 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, $-a^2\nabla^2$, takes the form in the semi-spectral domain (latitude and zonal wavenumber):

$$-a^2\nabla^2 = \frac{a^2m^2}{\tau^2} - \frac{a^2}{R^2\tau}\frac{\partial}{\partial\theta}\tau\frac{\partial}{\partial\theta}, \tag{4.17}$$

where the metric term, $\tau$, is now defined for an earth of radius $R$ simply as:

$$\tau(\theta) = R\cos(\theta). \tag{4.18}$$

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 coefficients $b_{1,j}$ in the way described in Section 4a if the half-bandwidth $n$ of the final operators is anticipated to be larger than one. As in the case discussed in section 4a of a nonuniform one-dimensional grid, this is trivially transformable (through a similarity relation) to symmetric form. As we might expect, the exponential function of the discrete matrix representation of $-(a^2/2)\nabla^2$ is taken, but only component matrices of half-bandwidth not exceeding $n$ are retained, in order to obtain the representation, at each wavenumber $m$, of $\mathcal{D}_{(n)}$. From the derived Cholesky factors, the coefficients of the advancing and backing basic filters are then quite easily extracted.

These filters are more expensive to apply globally than the doubly-recursive filters of Section 3 because they require a zonal Fourier transform to be applied to the input data at each latitude and an inverse transform applied to the final output data. But they do provide a satisfactory solution to the 'polar problem' in the case of homogeneous smoothing scale, $a$. Moreover, they serve to regularize the zonal aspect of the filtering which, in the vicinity of the poles where grid meridians cluster tightly together, presents numerical conditioning difficulties for the purely recursive filters, as discussed in Section 3c. In an earlier phase of this study, two of us (RJP and NMR) investigated filters of this semi-spectral form for a global analysis and devised methods for constructing hybrid filters in which only the polar caps are treated by the Fourier transform method, the data elsewhere being dealt with by the methods of Section 4a.

The main difficulty in constructing a hybrid filter of this kind is achieving a satisfactory blending of the alternative methods at the outer reaches of each designated polar cap. Let us suppose first that we can isolate each hemisphere (we discuss below how this can be done) so that we effectively have only one pole to deal with. We shall order the latitude grid points increasing towards the pole and denote by $i_c$ the first latitude grid point of the polar cap where Fourier transformation may be applied. In order to preserve self-adjointness, it is convenient to split the filtering in the zonal direction outside the polar cap (that is, $i < i_c$) into two equal self-adjoint halves, say $(\mathcal{D}_X)^{-1/2}$. Each half can then sandwich the self-adjoint meridional filter, say $(\mathcal{D}_Y)^{-1}$. Thus, neglecting the poles the output, $\mathbf{s}$, is obtained through the input, $\mathbf{p}$, according to:

$$\mathcal{D}_X^{1/2}\mathcal{D}_Y\mathcal{D}_X^{1/2}\mathbf{s} = \mathbf{p}. \tag{4.19}$$

Within the polar caps, we cannot assume separability into the two grid directions (except via the zonal Fourier transforms), so the symbolic application of the filtering there is expressed as the inversion of a single, but two-dimensional operator:

$$\mathcal{D}_F \mathbf{s} = \mathbf{p}. \tag{4.20}$$

While it is clearly not possible to convert the polar filter (4.20) into the form (4.19) with a one-dimensional factor $D_Y$, we can trivially perform Fourier transforms on (4.19) to put it into the same form as the operator implied by (4.20); the filter $D_Y$ is unaffected by zonal Fourier transformation while $D_X^{1/2}$ separates, for each wavenumber $m$, into a diagonal operator in latitude. Thus, within the polar cap, we may employ the Fourier transforms to separate the components for each wavenumber and define a new operator,

$$D_{Y'} = (\mathcal{D}_X)^{-1/2} \mathcal{D}_F (\mathcal{D}_X)^{-1/2} \tag{4.21}$$

so that the polar filtering, though still involving a fully two-dimensional operator ($D_{Y'}$ still requires Fourier transformation for its efficient implementation), formally more closely resembles the construction of (4.19). The polar filtering can now be implemented:

$$\mathcal{D}_X^{1/2} \mathcal{D}_{Y'} \mathcal{D}_X^{1/2} \mathbf{s} = \mathbf{p}. \tag{4.22}$$

where the difference between the new operator $D_{Y'}$ and $D_Y$ is not expected to be very large near the edge of the polar cap. For each wavenumber $m$, these operators may be blended smoothly just inside the polar cap, i.e., for each $m$, the matrix components of the blended result, operator $\hat{\mathcal{D}}_{Y''}$, might be given,

$$(\hat{\mathcal{D}}_{Y''})_{i,j} = (\hat{\mathcal{D}}_Y)_{i,j}(1 - W_{i,j}) + (\hat{\mathcal{D}}_{Y'})_{i,j} W_{i,j} \tag{4.23}$$

where blending weights satisfy:

$$W_{i,j} = 0, i, j < i_c, \tag{4.24}$$

together with the condition that preserves self-adjointness:

$$W_{i,j} = W_{j,i}, \tag{4.25}$$

and where these weights increase smoothly (in $(i, j)$) to unity inside the cap.

Having arranged for the latitude grid point ordering to increase towards the poles, we shall find that the L-U decomposition of the hybrid operator $\mathcal{D}_{Y''}$ outside the polar caps remains unchanged from the L-U decomposition of $\mathcal{D}_Y$, there being no communication between the separate meridians. Inside the caps, the partial separability is maintained only in the sense of between different wavenumbers so, within the caps, we revert to zonal Fourier transformation in order to complete the L-U decomposition at the 'bottom right' corner of each wavenumber's matrix. This technique, though it involves significant algorithmic complexity, does produce satisfactory self-adjoint filtering results at an acceptable computational cost over a hemisphere. The input data in each hemisphere is treated using a different factorization of $\mathcal{D}_Y$ — both cases being of the Cholesky type, but differing through the consequences of reversing the index ordering. Since the alternative products are, in both cases, equivalent, no discontinuity should appear at the 'join' along the equator in the output.

It is possible to carry out an analogous factorization of the operators even when the filtering scale varies with position, but the Fourier transformation within the polar caps is no longer so directly useful, except as a way to improve the conditioning of the problem; the irreduceably two-dimensional character of the operator, $\mathcal{D}_F$, in the polar caps leads to a full matrix representation which Fourier transformation now fails to diagonalize. Although the numerical problem is therefore larger, the calculations are still rendered manageable provided the polar caps are small in geographical extent and the meridional resolution is suitably reduced, as in a 'skipped' or 'reduced' latitude-longitude grid.

We have described a suite of numerical techniques by which a locally isotropic quasi-Gaussian smoothing filter can be efficiently applied. The Gaussian profile is not necessarily best for all applications to problems in data assimilation where the broader range of dynamical scales offered by 'fat-tailed' distributions is often more desirable. The next section discusses a further generalization to the method where, using additive combinations of pseudo-Gaussians, we are able to synthesize locally isotropic filters with a more general profile.

## 5. Synthesis of a covariance in terms of Gaussians

At this point we can proceed almost directly to the construction of an analysis scheme using a covariance model based on a quasi-Gaussian profile of appropriately chosen constant scale. The amplitude (that is, the 'variance') by which the covariance may vary geographically can be controlled, while preserving the important property of self-adjointness, by modulating the otherwise homogeneous filter before and after application:

$$B(\mathbf{x}_1, \mathbf{x}_2) = w^{1/2}(\mathbf{x}_1) G(\mathbf{x}_1, \mathbf{x}_2) w^{1/2}(\mathbf{x}_2), \tag{5.1}$$

where $w(\mathbf{x})$ is the total effective weight at $\mathbf{x}$ applied to the Gaussian filter, $G$. 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 sigificant 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-dimensional Gaussian function of unit integrated weight, and scale parameter, $a$:

$$G_a(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2\pi a^2} \exp\left[-\frac{1}{2}\left(\frac{\mathbf{x}_1 - \mathbf{x}_2}{a}\right)^2\right], \tag{5.2}$$

is associated with the power spectrum,

$$\hat{G}_a(k, l) = \int\int G_a(\mathbf{x}, 0) \exp(-i\mathbf{k}\cdot\mathbf{x}) dx\, dy = \exp\left[-\frac{a^2}{2}(k^2 + l^2)\right], \tag{5.3}$$

which is itself of Gaussian form, and which therefore does possess a rapidly diminishing tail in the spectral region of total wave numbers 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):

$$\hat{B}(k,l) = \int_0^\infty w\left(\frac{a^2}{2}\right) \exp\left[-\frac{a^2}{2}(k^2+l^2)\right] d\left(\frac{a^2}{2}\right). \tag{5.4}$$

In practice, we would wish to approximate the general superposition (5.4) by a discrete approximation involving many Gaussian constituents. That is:

$$\hat{B}(k,l) = \sum_p w_p \exp\left[-\frac{a_p}{2}(k^2+l^2)\right]. \tag{5.5}$$

We would also like to confine ourselves to constructions of this type that result in all the weights, $w_p$, being non-negative so that, in the spatial domain, the combination can be expressed in a way that allows gradual regional variations without jeopardizing the self-adjointness that the iterative solution algorithms depend on. The appropriate form in the physical domain is therefore:

$$B(\mathbf{x}_1,\mathbf{x}_2) = \sum_p w_p^{1/2}(\mathbf{x}_1) G_{a_p}(\mathbf{x}_1,\mathbf{x}_2) w_p^{1/2}(\mathbf{x}_2). \tag{5.6}$$

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 two 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 'hyperGaussian' family of functions. For the two-dimensional isotropic hyperGaussian functions normalized to have unit integrals, a generic member of this family is characterized by a scale parameter, $\sigma$, and a shape parameter, $\gamma$. 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 $\gamma$:

$$H_{\sigma,\gamma}(\mathbf{x}) = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\gamma}} \exp\left[-\frac{1}{2}\frac{(s-\log\sigma)^2}{\gamma}\right] \frac{1}{2\pi\exp(2s)} \exp\left[-\frac{1}{2}\left(\frac{\mathbf{x}}{\exp(s)}\right)^2\right] ds. \tag{5.7}$$

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

Fig. 4a illustrates the radial profile of the correlation implied by this covariance model for $\gamma = 0$, 0.1, 0.2 and 0.3. The parameter $\gamma$ provides control over what we may refer to, following statistical parlance, as the 'kurtosis' of the distribution. Suppose we define unidirectional moments of an isotropic distribution, $H(\mathbf{x})$:

$$\mu_m = \int \int H(\mathbf{x}) x^m \, dx \, dy. \tag{5.8}$$

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, (5.7) is a linear superposition of Gaussians, and each moment is a linear functional of the distribution, we find that, for the hyperGaussian with shape parameter $\gamma$, 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:

$$\kappa = \frac{\mu_0 \mu_4}{(\mu_2)^2}, \tag{5.9}$$

we find that the kurtosis for the hyperGaussian of shape parameter $\gamma$ 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,

$$H'_{\sigma,\gamma}(\mathbf{x}) = -\nabla^2 H_{\sigma,\gamma}(\mathbf{x}), \tag{5.10}$$

have the correlation profiles depicted in Fig. 4b. 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:

$$\hat{H}_{\sigma,\gamma}(\mathbf{k}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\gamma}} \exp\left[-\frac{1}{2}\frac{(s - \log\sigma)^2}{\gamma}\right] \exp\left[-\frac{1}{2}\left(\exp(s)\mathbf{k}\right)^2\right] ds. \tag{5.11}$$

$$\hat{H}'_{\sigma,\gamma}(\mathbf{k}) = \mathbf{k}^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\gamma}} \exp\left[-\frac{1}{2}\frac{(s - \log\sigma)^2}{\gamma}\right] \exp\left[-\frac{1}{2}\left(\exp(s)\mathbf{k}\right)^2\right] ds. \tag{5.12}$$

These power spectrum families are shown, for the same range of shape parameters as before, in Figs. 5a (for $\hat{H}$) and 5b (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, $\gamma$.

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 $\gamma$ it is prudent to increase this density of discrete scales. In the context of the multigrid construction, 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.

## 6. Discussion

The problem of efficiently accommodating approximately isotropic but spatially inhomogeneous covariance functions in a variational analysis has been solved using recursive numerical filters. The covariances are never explicitly computed; instead, it is their effects as convolution operators that 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 4 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 Mercator grid elsewhere, in order to preserve the property of local isotropy, and use the multi-Gaussian methods of synthesis of Section 5 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 other grids different from the model grid, provided the final conversions from one to the other are by high-order accurate interpolations. In the case of the conversions from the polar stereographic grid to the global latitude and longitude grid, the high-order interpolations proceed using the so-called 'cascade' method (Purser and Leslie, 1991), in which the two-dimensional problem is efficiently split into separate one-dimensional operations. 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 4b. For a global analysis, the user must choose the method best adapted to his or her requirements.

An additional development that we expect to report on soon is the further generalization of the covariance operators to accommodate fully anisotropic effects in both two and three dimensions. 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 (2000). Algorithms exist for extending the recursive filter method in this way in a regional domain and their application to variational data analysis looks very promising.

Acknowledgments

The authors would like to thank Drs. John Derber, Dezso Devenyi, and Andrew Lorenc for many helpful discussions. We also thank Prof. Eugenia Kalnay and Drs. Stephen Lord and Roger Daley for their encouragement and support.

## Appendix A

*Obtaining filter coefficients for a given scale*

In section 3, $\mathcal{D}^*_{(n)}$ was defined as a real-coefficient monic polynomial of $K$, so we may use this polynomial's roots $\kappa_p$ (whose complex members come in conjugate pairs) to perform the operator factorization,

$$\mathcal{D}^*_{(n)} = \prod_{p=1}^{n} \left(1 - \frac{\hat{K}}{\kappa_p}\right). \tag{A.1}$$

In terms of the shift operator, $Z$ defined:

$$Z(\psi)_i = \psi_{i+1}, \tag{A.2}$$

we have

$$-K\delta x^2 = (Z - 2 + Z^{-1}), \tag{A.3}$$

and each factor in (A.1) is expressible:

$$1 - \frac{K}{\kappa_p} = \frac{(Z - 2\omega_p + Z^{-1})}{2 - \omega_p}$$

where

$$\omega_p = 1 - \kappa_p \delta x^2/2.$$

Therefore, the smaller of the two possible roots,

$$\zeta_p = \left(\omega_p + (\omega_p^2 - 1)^{1/2}\right)^{\pm 1} \tag{A.4}$$

of the quadratic,

$$z^2 - 2\omega_p z + 1$$

allows the factorization of the term,

$$1 - \frac{K}{\kappa_p} = \left(\frac{Z - \zeta_p}{1 - \zeta_p}\right)\left(\frac{Z^{-1} - \zeta_p}{1 - \zeta_p}\right). \tag{A.5}$$

The operation,

$$\psi = \left(\frac{Z^{-1} - \zeta_p}{1 - \zeta_p}\right)\chi$$

describes the application of the right-moving, complex coefficient, first-order recursive filter,

$$\psi_i = (1 - \zeta_p)\chi_i + \zeta_p \psi_{i-1},$$

whose stability is guaranteed by the property $|\zeta_p| < 1$ [the other root of (A.4) is its reciprocal]. Likewise, the inverse of the other operator factor on the right of (A.5) describes the operation of a left-moving stable recursive filter. By a similar decomposition for all the $p \in \{1, \dots, n\}$, we deduce that the operators of (3.12) and (3.13), comprising the right-moving and left-moving high-order filters, are formally:

$$\mathcal{A} = \prod_{p=1}^{n} \left( \frac{Z^{-1} - \zeta_p}{1 - \zeta_p} \right), \tag{A.6}$$

$$\mathcal{B} = \prod_{p=1}^{n} \left( \frac{Z - \zeta_p}{1 - \zeta_p} \right), \tag{A.7}$$

whose operator kernels are real by virtue of the occurrences of the complex $\kappa_p$, and hence of the $\zeta_p$, in conjugate pairs, and whose coefficients are constructible by the explicit convolution products of $Z^{-1}$ and $Z$ prescribed by (A.6) and (A.7).

We summarize the practical steps required in order to obtain these coefficients. First, one locates the complex roots, $\kappa_p$, of the real-coefficient polynomial in $K$ that defines $\mathcal{D}^*_{(n)}$. Then, the corresponding quantities $\omega$ and, in each case, the smaller of the two roots $\zeta_p$ are obtained. Then the convolution polynomial (A.6) is constructed using these complex (in general) $\zeta_p$. Finally, we invoke (3.14) and (3.15) to get the algorithmically convenient coefficients $\alpha_i$ and $\gamma$ of this filtering scheme.

However, the complex arithmetic can be avoided by an alternative iterative method of obtaining the convolution kernels of $\mathcal{A}$ and its adjoint $\mathcal{B}$. According to (3.11), these are the rows of the Cholesky factors of an infinite symmetric band matric representation of $\mathcal{D}^*_{(n)}$, but in practice, given sufficiently large *finite* symmetric matrix having the same generic rows as $\mathcal{D}^*_{(n)}$, one finds that the last rows of its Cholesky lower-triangular factor will be numerically indistinguishable, both from each other, and from the generic row of the Cholesky factor of the infinite matrix representation of $\mathcal{D}^*_{(n)}$. By an adaptation of the regular Cholesky algorithm,

$$\mathcal{A}_{i,j} = (\mathcal{D}^*_{i,j} - \sum_{k<j} \mathcal{A}_{i,k}\mathcal{A}_{j,k})/A_{j,j}, \quad j < i, \tag{A.8}$$

$$\mathcal{A}_{i,i} = (\mathcal{D}^*_{i,i} - \sum_{k<i} \mathcal{A}_{i,k}\mathcal{A}_{i,k})^{1/2},, \tag{A.9}$$

one can compute successive rows until, within round-off, the subsequent differences from row to row are negligible. This iterative approach is equally valid and never fails in practice to reproduce the same numerical coefficients as is given by the formal root-finding procedure.

## Appendix B

### *Non-periodic end conditions*

Consider a mutually adjoint pair of $n$th-order basic recursive filters $\mathcal{A}$ ('advancing') and $\mathcal{B}$ ('backing') on a uniform grid and with constant coefficients such that, if $\mathbf{q} = \mathcal{A}\mathbf{p}$, and $\mathbf{r} = \mathcal{B}\mathbf{p}$, then, for a generic grid point $i$,

$$q_i = \beta p_i + \sum\nolimits_{j=1}^{n} \alpha_j q_{i-j}, \tag{B.1}$$

$$r_i = \beta p_i + \sum\nolimits_{j=1}^{n} \alpha_j q_{i+j}. \tag{B.2}$$

On an infinite grid, assuming all data remain bounded at infinity, one can easily verify for the symmetrized output, $\mathbf{s}$:

$$\mathbf{s} = \mathcal{B}\mathcal{A}\mathbf{p} \equiv \mathcal{A}\mathcal{B}\mathbf{p}. \tag{B.3}$$

We ask, given a *finite* domain, $i \in [1, N]$, and assuming this interval contains the support of input $\mathbf{p}$: how does one 'prime' those values of $\mathbf{s}$ just inside the boundary at $i = N$ in order to enable the backing filter, $\mathcal{B}$ when applied to $\mathbf{q} = \mathcal{A}\mathbf{p}$, to simulate implicitly the effect of a continuation of the gridded values of $\mathbf{q}$ beyond this boundary? The solution of this puzzle is found by exploiting the commutativity (B.3). Let $\hat{\mathbf{s}}_j$ denote the $n$-vector of $(s_{j+1-n}), \ldots, s_j)^T$. Then $\hat{\mathbf{s}}_N$ are the last $n$ values of $\mathbf{s}$ belonging to the actual domain while $\hat{\mathbf{s}}_{N+n}$ is the vector of $n$ values one would have obtained just beyond, in the grid were continued. Define a lower-triangular $n \times n$ matrix, $\mathbf{L}$ with elements,

$$\begin{aligned} L_{i,i} &= 1, && \text{(B.4)} \\ L_{i+j,i} &= -\alpha_j, && \text{(B.5)} \end{aligned}$$

and an upper-triangular $n \times n$ matrix, $\mathbf{U}$, with elements,

$$U_{i,i+j} = \alpha_{N-j}. \tag{B.6}$$

Then, since $\hat{\mathbf{r}}_{N+n} = \mathbf{0}$, it must follow from $\mathbf{s} = \mathcal{A}\mathbf{r}$, that

$$\mathbf{L}\hat{\mathbf{s}}_{N+n} = \mathbf{U}\hat{\mathbf{s}}_N, \tag{B.7}$$

and, from $\mathbf{s} = \mathcal{B}\mathbf{q}$, that

$$\mathbf{L}^T\hat{\mathbf{s}}_N = \mathbf{U}^T\hat{\mathbf{s}}_{N+n} + \hat{\mathbf{q}}_N\beta. \tag{B.8}$$

Thus, eliminating $\hat{\mathbf{s}}_{N+n}$, we obtain the turning conditions that prime the backing filter:

$$(\mathbf{L}^T - \mathbf{U}^T\mathbf{L}^{-1}\mathbf{U})\hat{\mathbf{s}}_N = \hat{\mathbf{q}}_N\beta. \tag{B.9}$$

For the simplest filter pair, $n = 1$, $\alpha_1 = \alpha$, this formula reduces to

$$(1 - \alpha^2)s_N = \beta q_N. \tag{B.10}$$

## Appendix C

### *Periodic end conditions*

Preserving the notation of Appendix B, we consider the problem of defining the correct priming vector $\hat{\mathbf{q}}_0$ for the advancing filter on a cyclic domain with period $N$, such that the values obtained are consistent with the wrap-around condition, $\hat{\mathbf{q}}_N = \hat{\mathbf{q}}_0$. In the recursion (3.14), the sensitivity of $\hat{\mathbf{q}}_N$ to $\hat{\mathbf{q}}_{N-1}$, given that input element $p_N$ is unchanged, is expressed:

$$\mathbf{d}\hat{\mathbf{q}}_N = \mathbf{T}\mathbf{d}\hat{\mathbf{q}}_0$$

where

$$\mathbf{T} = \begin{pmatrix} 0, & 1, & 0, & \dots, & 0 \\ 0, & 0, & 1, & \dots, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_n, & \alpha_{n-1}, & \alpha_{n-2}, & \dots, & \alpha_1 \end{pmatrix}.$$

Applying the chain-rule, we deduce that the sensitivity of $\hat{\mathbf{q}}_N$ to $\hat{\mathbf{q}}_0$ when the intervening input elements, $p_1, \dots, p_N$, remain unchanged, is:

$$\mathbf{d}\hat{\mathbf{q}}_N = \mathbf{T}^N \mathbf{d}\hat{\mathbf{q}}_0, \tag{C.1}$$

or, for some linear n-vector function $\hat{\mathbf{h}}$ of the $p_1, \dots, p_N$,

$$\hat{\mathbf{q}}_N = \mathbf{T}^N \hat{\mathbf{q}}_0 + \hat{\mathbf{h}}(\mathbf{p}). \tag{C.2}$$

The identification of $\hat{\mathbf{q}}_0$ and $\hat{\mathbf{q}}_N$ therefore requires that

$$\hat{\mathbf{q}}_0 = (\mathbf{I} - \mathbf{T}^N)^{-1} \hat{\mathbf{h}}, \tag{C.3}$$

To solve this, a preliminary advancing recursive sweep primed using $\hat{\mathbf{q}}_0^* = \mathbf{0}$ provides the vector $\hat{\mathbf{h}} = \hat{\mathbf{q}}_N^*$ obtained as the last $n$ output values of this inconsistent solution. Thus, from (C.3), the proper consistent priming conditions are derived, and a second sweep completes the result for the advancing filter.

A similar procedure is used for the backing filter on the cyclic domain, so the overall cost is double that of the recursive filter on a nonperiodic domain of $N$ points. For this reason, when performance is critical, it may be preferable in practice to employ a generous overlap and the nonperiodic version of the filter instead, as was done for the global assimilation experiments of section 5. However, as we find in Appendix D, the extra overhead of computing the proper cyclic condition is a factor for consideration only for serial processing; when the domain is divided into segments, whether it is periodic or not, the recursions *always* need to be repeated to achieve inter-segment consistency.

## Appendix D

*Parallel and distributed processing*

Divide the domain of $N$ points into $M$ consecutive segments, $\mathbf{p}^{(1)}, \dots \mathbf{p}^{(M)}$ (for input data) with segment $J$ comprising $N_J$ points (and therefore $\sum_{J=1}^{M} N_J = N$) and consider each segment to be assigned to a separate processor of a parallel computer. Adapting the notation of the earlier appendices, the problem we must then address for the advancing filter $\mathcal{A}$ is how most efficiently to achieve consistency between the last $n$ output values $\hat{\mathbf{q}}_{N_J}^{(J)}$ of segment $J$ and the priming vector $\hat{\mathbf{q}}_0^{(J+1)}$ in the next segment. As in the serial treatment of the cyclic case discussed in appendix C, we deduce that

$$\hat{\mathbf{q}}_{N_J}^{(J)} = T^{(N_J)} \hat{\mathbf{q}}_0^{(J)} + \hat{\mathbf{h}}^{(J)}, \tag{D.1}$$

so that, by concurrent computations in all $M$ segments, the vectors $\hat{\mathbf{h}}^{(J)}$ of each is obtained by running the advancing filters in each segment with null priming vectors. The consistency condition:

$$\hat{\mathbf{q}}_0^{(J)} = \mathbf{T}^{N_J}\hat{\mathbf{q}}_{N_{J-1}}^{(N_{J-1})} + \hat{\mathbf{h}}^{(J)}, \tag{D.2}$$

provides row $J$ of a block-matrix expression:

$$\begin{pmatrix} \mathbf{I}, & 0, & \dots, & 0, & 0 \\ -\mathbf{T}^{N_1}, & \mathbf{I}, & \dots, & 0, & 0 \\ 0, & \ddots & \ddots & \mathbf{I}, & 0 \\ 0, & 0, & \dots, & -\mathbf{T}^{N_{M-1}}, & \mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{q}}_0^{(1)} \\ \hat{\mathbf{q}}_0^{(2)} \\ \vdots \\ \hat{\mathbf{q}}_0^{(M)} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{h}}^{(1)} \\ \hat{\mathbf{h}}^{(2)} \\ \vdots \\ \hat{\mathbf{h}}^{(M)} \end{pmatrix}. \tag{D.3}$$

This 'reconciliation equation' is one that can also be distributed when several parallel lines of data are collected together, as in a typical application in two or three dimensions. Therefore, the wall-clock time associated with the reconciliation is usually not very significant compared to the regular recursions the precede and follow it. For a cyclic domain, the only change is that the block-bidiagonal matrix of (D.3) is replaced by the corresponding *cyclic* block-bidiagonal, with the block, $-\mathbf{T}^{N_M}$ occupying the top-right location. The final step is, of course, the properly primed recursions performed concurrently within all processors.

## Appendix E

*Amplitude estimation for inhomogeneous quasi-Gaussian filters*

The control of amplitude (variance) of the covariance filter is quite straight-forward 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. Thus, with a 'diffusivity' $D(x)$, the outcome of the spatially inhomogeneous filter is identified with the application for 'time' $t$ of the diffusion equation:

$$\dot{\psi} \equiv \frac{d\psi}{dt} = \frac{d}{dx}D\frac{d\psi}{dx} \equiv (D\psi\prime)\prime. \tag{E.1}$$

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

$$D(x) = D_0 + D_1 x + D_2 x^2 + \dots \quad . \tag{E.2}$$

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

$$\psi(x,t) = \exp(g(x,t)), \tag{E.3}$$

where $g$ is expressed as a series:

$$g(x,t) = g_0(t) + g_1(t)x + g_2(t)x^2 + \dots \quad . \tag{E.4}$$

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

$$g_0 = -\frac{1}{2}\log t - \frac{1}{2}\log(4\pi D_0), \tag{E.5}$$

$$g_2 = -\frac{1}{4D_0 t}, \tag{E.6}$$

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 (E.1) and (E.3):

$$\dot{g} = (Dg\prime)\prime + D(g\prime)^2, \tag{E.7}$$

we obtain, after some albegra:

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

where

$$h_k \equiv k g_k. \tag{E.9}$$

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 (E.8) for the first few even power of $x$:

$$\dot{g}_0 = 2g_2 D_0, \tag{E.10}$$

$$\dot{g}_2 = 12g_4 D_0 + 6g_2 D_2 + 4g_2 g_2 D_0, \tag{E.11}$$

$$\dot{g}_4 = 30g_6 D_0 + 20g_4 D_2 + 10g_2 D_4 + 16g_2 g_4 D_0 + 4g_2 g_2 D_2. \tag{E.12}$$

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

$$g_2(t) \equiv g_{2,-1}t^{-1} + g_{2,0} + g_{2,1}t + \ldots \qquad . \tag{E.13}$$

From terms $t^{-2}$ in (E.11):

$$g_{2,-1} = -\frac{1}{4D_0}, \tag{E.14}$$

and from corresponding terms in (E.12):

$$g_{4,-1} = \frac{D_2}{12D_0^2}. \tag{E.15}$$

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

$$g_{2,0} = -\frac{D_2}{4D_0}. \tag{E.16}$$

Then, integrating (E.10), we obtain:

$$g_0(t) = -\frac{1}{2}\log t + g_{0,0} - \frac{1}{2}D_2 t + \mathcal{O}(t^2). \tag{E.17}$$

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

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

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

$$-\frac{1}{2}\log(4\pi\bar{D}) \equiv -\frac{1}{2}\log(4\pi D_0) - \frac{1}{2}D_2 t + \mathcal{O}(t^2), \tag{E.19}$$

then,

$$\begin{aligned} \bar{D} &\approx D_0 \exp(D_2 t), && \text{(E.20)} \\ &\equiv D_0 + D_2 D_0 t + \mathcal{O}(t^2). && \text{(E.21)} \end{aligned}$$

Owing to the properties,

$$\begin{aligned} \int \psi(x, t/2)dx &= 1, && \text{(E.22)} \\ \int \psi(x, t/2)x^2 dx &\approx D_0 t, && \text{(E.23)} \end{aligned}$$

we may employ the approximation,

$$\bar{D} \approx \int \psi(x, t/2)D(x)dx \tag{E.24}$$

to acquire a serviceable and robust effective diffusivity, $\bar{D}$, 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, $D$. The first-order correction can still be found by smoothing $|D|$ with the square-root filter, since, to a sufficient approximation, the filter is factorable into the individual dimensions yielding the one-dimensional filters we have just analyzed and the determinant is simply the product of the diffusivities in each of the component dimensions. Moreover, this generalization remains valid even for anisotropic filters (to which we intend to devote a future article), since even they can be constructed from uni-directional diffusion operators, but acting now in mutually oblique directions.

## References

Barnes, S. L. 1964 A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3,** 396–409.

Bergthorssen, P., and B. Döös 1955 Numerical weather map analysis. *Tellus*, **7,** 329–340.

Brandt, A. 1977 Multilevel adaptive solutions of boundary value problems. *Math. Comp.*, **31,** 333–390.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich 1998 Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.*, **126,** 2737–3052.

Courtier, P. 1997 Dual formulations of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **123,** 2449–2461.

Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124,** 1783–1807.

Cressman, G. P. 1959 An operational analysis scheme. *Mon. Wea. Rev.*, **87,** 367–374.

Dahlquist, G., and A. Björck 1974 *Numerical Methods*, Prentice Hall, 573pp.

Daley, R. A. 1991 *Atmospheric Data Assimilation.* Cambridge University Press, 457pp.

Daley, R., and E. Barker 2000 *NAVDAS 2000 Source Book* Naval Research Laboratory, Monterey, CA 93943-5502, NRL Publiciation NRL/PU/7530–00-418, 153 pp.

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.

Gandin, L. S. 1963 *Objective Analysis of Meteorological Fields*, Leningrad, Gidromet; (Jerusalem, Isreal Program for Scientific Translations; 1965, 242pp.)

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.

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

Lorenc, A. C. 1981 A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **101,** 701–721.

Lorenc, A. C. 1986 Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112,** 1177–1194.

Lorenc, A. C. 1997 Development of an operational variational Assimilation Scheme. *Journal of the Meteorological Society of Japan*, **75,** No. 1B, 339–346.

Otnes, R. K., and L. Enochson 1972 *Digital Time Series Analysis.* Wiley, 467pp.

Parrish, D. F., and J. C. Derber 1992 The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120,** 1747–1763.

Purser, R. J., and R. McQuigg 1982 A successive correction analysis scheme using recursive numerical filters. Met. O 11 Tech. Note, No. 154, British Meteorological Office, 17pp.

Purser, R. J., and L. M. Leslie | 1991 | An efficient interpolation procedure for high-order three-dimensional semi-Lagrangian models. *Mon. Wea. Rev.*, **119,** 2492–2498.

Purser, R. J., and D. F. Parrish | 2000 | A Bayesian technique for estimating continuously varying statistical parameters of a variational assimilation. NOAA/NCEP Office Note 429. 28pp.

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.

Thiébaux, H. J. | 1976 | Anisotropic correlation functions for objective analysis. *Mon. Wea. Rev.*, **104,** 994–1002.

Thiébaux, H. J., H. L. Mitchell, and D. W. Shantz | 1986 | Horizontal structure of hemispheric forecast error correlations for geopotential and temperature. *Mon. Wea. Rev.*, **114,** 1048–1066.

Wilks, D. S. | 1995 | *Statistical Methods in the Atmospheric Sciences: an Introduction.* Academic Press, 467pp.

Figure 1. Comparison of one-dimensional applications of recursive filters approximating a Gaussian (shown solid). Dashed curves show filter approximations: (a) order $n = 1$; (b) $n = 2$ with (long dashes) and without (short dashes) the off-diagonal $b$ coefficient refinements; (c) $n = 4$ with and without off-diagonal $b$ coefficients.

Figure 2. Sequential application of quasi-Gaussian recursive filters of order $n$ in two dimensions. (a) $n=1$; (b) $n=2$; (c) $n=4$; (d) four applications of filters with $n=1$ with scale parameter adjusted to make the result comparable with the other single-pass filters. Contours are shown at multiples of odd integers.

Figure 3. Negative-Laplacian applied to quasi-Gaussian recursive filters with: (a) $n=2$; (b) $n=4$; (c) $n=6$. (d) corresponding contours for the exact Gaussian. Contours as in Fig. 3 and with negative contours shown as broken curves.

Figure 4. (a) Cross-section profiles of the fat-tailed 'hyper-Gaussian' covariance models defined in Section 5 for a range of shape parameters $\gamma$. (b) the result of applying the negative-Laplacian, and renormalization of amplitude, to these hyper-Gaussian functions.

Figure 5. Power spectra in log-linear coordinates for the covariances depicted in Fig. 4. (a) the hyper-Gaussians; (b) the negative-Laplacians of the hyper-Gaussians.