Covariance modelling in a grid-point analysis

R. James Purser[1], Manuel S. F. V. de Pondeca[1], David F. Parrish[2] and Dezső Dévényi[3]

[1]*SAIC at* [2]*NOAA/NCEP/EMC*
[3]*CIRES/NOAA/ESRL*
USA
Jim.Purser@noaa.gov

ABSTRACT

A brief survey is provided of some of the numerical techniques developed, or under active investigation at the National Centers for Environmental Prediction (NCEP) directed at improving the modelling of background error covariance operators used in the operational global and regional data assimilation systems.

1 Introduction

Some of the filtering techniques that are being used, or are intended for future use, as the covariance generators in NCEP's Grid-point Statistical Interpolation (GSI) method of three-dimensional variational assimilation (3Dvar) are described. The first three topics discussed: the recursive filter; the Triad/Hexad methods; and their "blended" refinement, are more or less completed developments. The remaining sections will be devoted to pending research developments that include: a new normalization strategy for inhomogeneous filters based on ideas in differential geometry; multi-scale covariance synthesis facilitated by multigrid techniques; Analysis error characterization employing multiscale filters, and their possible application to preconditioning; and, finally, an approach to treating covariances possessing sharp geographical gradients. While the focus is on applications to 3Dvar, it is noted that equivalent techniques remain relevant in the context of weak-constraint 4Dvar, where it is desirable to have a smooth time-space covariance of the model error component.

2 The Recursive Filter

The recursive filter (e.g., Purser and McQuigg 1982, Hayden and Purser 1995) is an efficient numerical technique for simulating diffusion along a grid line. Covariance operators can be constructed with it (Purser 1983, Lorenc 1992), as is now done in the GSI at NCEP.

The idea is to take a finite-degree Taylor series approximation to the exponential function, e.g., truncating at fourth degree:

$$\exp(z) \approx 1 + z + z^2/2 + z^3/6 + z^4/24. \tag{1}$$

Then apply this function to an effective Diffusivity$\times$negative-Laplacian$\times$Duration:

$$S = I + (-tD\nabla^2) + (-tD\nabla^2)^2/2 + (-tD\nabla^2)^3/6 + (-tD\nabla^2)^4/24. \tag{2}$$

In finite difference terms in one dimension, this is a symmetric band matrix operator. Since the isotropic diffusion operator in N dimensions factorizes into separate orthogonal one-dimensional diffusion operators, we will find that the inversion of S applied to an input field is well approximated by the sequence of N one-dimensional

linear inversions applied consecutively along each of these orthogonal directions. Any given factor, say S_1 of S, acting only along the x_1 direction, is a symmetric band matrix possessing a band Cholesky factorization,

$$S_1 = L_1 L_1^T, \tag{3}$$

where L_1 is a lower triangular band matrix, L_1^T its transpose. This allows the inversion of each part of the linear system approximating

$$\exp(-tD\nabla^2)\psi = \chi, \tag{4}$$

projecting into the x_i direction to be carried out in two recursive steps,

$$\phi_i = L_i^{-1}\chi_{i-1}, \tag{5}$$

$$\chi_i = L_i^{-T}\phi_i. \tag{6}$$

The first of these is a series of recursive "back-substitution" steps progressing forward along the x_i grid; the second is the adjoint of this "recursive filter" retrogressing along the same grid. But the final result, $\psi \equiv \chi_N$, after all N directions have been processed in this manner, is an approximation to the application of the *inverse* of the operator in (4) applied to the original $\chi_0 \equiv \chi$:

$$\psi = \exp(+tD\nabla^2)\chi, \tag{7}$$

which is formally a way of writing the solution to the duration-t diffusion equation applied to χ. There are minor technicalities involving boundary conditions, but these are most conveniently swept away by assuming mirror-reflection conditions at boundaries (which may be placed slightly beyond the boundaries of the physical domain of interest in order to minimize numerical artifacts).

Difficulties can arise with recursive filters if the degree of the exponential-approximating polynomial is too high and the diffusivity×duration too large compared to the grid spacing. But such a polynomial can be factored into real-coefficient polynomials of degree not exceeding two, and this factored filter solves the conditioning problem.

It is convenient to choose the duration to be $t = 1/2$ since this makes the "aspect tensor" (normalized centered second-moment) of the response function identical to the diffusivity in the idealized case of spatially homogeneous smoothing in the absence of boundaries.

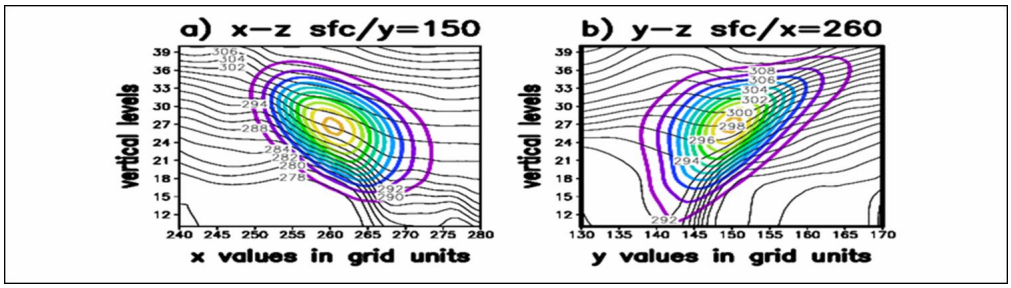

Figure 1: A pair of vertical cross-sections showing the response function resulting from a point source of an anisotropic covariance. Following Desroziers (1997), the anisotropy is designed to make the response function look horizontally isotropic and approximately homogeneous in geostrophic momentum space. For a point source inside a baroclinic zone, these physical-space cross-sections display significant anisotropy.

The idea of using a diffusive process to generate a covariance contribution is made explicit in the assimilation techniques of Derber and Rosati (1989) and Egbert et al. (1994), which are extended to the anisotropic diffusion on the sphere by Weaver and Courtier (2001). Figure 1 shows a pair of vertical cross-sections of a covariance whose anisotropy is guided by a variant of the semi-geostrophic transformation method of Desroziers (1997). Robust and efficient techniques enabling the production and accurate amplitude modulation of such covariances are described in outline in the following three sections.

3 Triad and Hexad methods

The general diffusion operation in N dimensions can be formally broken down into N line operations applied sequentially. (One convenient choice of the N *directions* is given by the columns of a Cholesky factor of the aspect tensor. The *amount* of smoothing in this jth direction must then be such as to make the orthogonal projection of this line-smoother's aspect tensor into the jth coordinate direction equal to the square of the jth diagonal element of the Cholesky factor.) The generally oblique line-oriented smoothers for each of the given directions in such a decomposition can still take the form of a pair of recursive operations along each of the set of such parallel lines threading the grid. However, with a generally oblique orientation to these lines, the progression of the filters would need to be interrupted at every single step to bring the target points of the filter back to actual grid points. (A scheme like this resembles a semi-Lagrangian advection scheme at each stage j, except with "time" replaced by the spatial coordinate directions, x_j.)

An attractive alternative, to avoid the massive amount of interpolations that such a "Cholesky filter" would entail, is to seek a larger set of filtering directions which, when oblique, still constrains all the lines to remain generalized grid lines (such as 45-degree lines of a square grid). Three directions in two dimensions (2D) and six directions in three dimensions (3D) suffice. This is the essential idea behind the "Triad" and "Hexad" algorithms (and their generalization to higher dimensions) as described by Purser et al. (2003).

The best choice for triads involves three alignments whose smallest grid steps, or "generators", form the smallest possible grid triangle. Equivalently, triad directions lie parallel to the diameters of the smallest possible grid-hexagons (a pair of which are illustrated in Fig. 2).

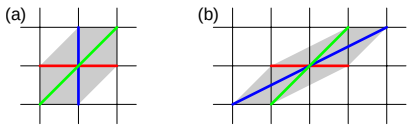

Figure 2: Two possible triads in the lattice of generators of a square grid. The triads share two of the three directions (horizontal, and 45 degree), but no other triad possesses both of these particular directions.

The "projection" of the given aspect tensor onto the respective 1D aspect tensors, or "spreads", of the line operators is essentially linear, since normalized second moments add under pure convolution (sequential filters), and approximately so even when the inhomogeneities render the convolutions "impure". But in general, the three spreads resulting from the projection of the desired aspect tensor into any given triad are likely to leave one of these spreads negative, which cannot correspond to any sensible smoothing or diffusing operator.

The *Triad Algorithm* responds to the occurrence of negative spread in a trial triad by replacing the offending spread's direction by that given by the alternative diameter that will allow a minimal grid-hexagon to appear (see Fig. 2).The new spread (from a new projection) is then always non-negative, although one of the other two spreads will typically have become negative through this transition; in this case, the process is iterated until a valid solution (three spreads non-negative) is found. Such a solution always exists and is unique for any given aspect tensor. This algorithm is applied at each geographic gridpoint.

The *Hexad algorithm* abides by the same general principle, but now six directions are needed to provide the necessary degrees of freedom. Geometry dictates that these directions correspond to those of the vertex-diameters of the smallest skewed "cuboctahedron" (a figure with 12 vertices, 6 quadrilateral faces and 8 triangular faces, as illustrated by the pair of "adjacent hexads" in Fig. 3.)

Again, lattice geometry forces unique rules for the transition between one hexad and its alternative, once five of the generators (and hence, cuboctahedron diameters) are nominated to be held fixed. The generic transition is exemplified in Fig. 3.

The triads or hexads and their accompanying spreads will clearly change with geographical location in the

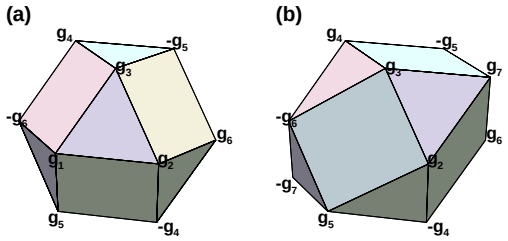

Figure 3: Two possible hexads in the lattice of generators of a cubic grid. The hexads share five generators, $g_1 \ldots g_5$, but the hexad of panel (a) possesses g_1 while that of panel (b) possesses g_7 instead. No other hexads possess all five of $g_1 \ldots g_5$, so any transition is unambiguous (discounting vertex permutations). Further details are given in Purser (2005)

usual case of an inhomogeneous covariance. Sequencing the line operations involved among the various triads or hexads so that there is no mutual interference of the filters whose lines of operations intersect is therefore a nontrivial problem when these operations are multi-processed.

The solution is provided by an idea from abstract algebra. The generators of a lattice can be "colored" using the non-vanishing elements of a suitable "Galois Field". The colors form a periodic pattern (the period being some prime number) over the generator lattice. For example, every generator of the basic triad is a different color supplied by the non-null elements of GF(4), the Galois field of four elements. (Each Galois field has a null element that plays the role of "coordinate origin" in some sense.) In three dimensions the simplest relevant Galois field is GF(8), which provides seven "colors", of which each hexad enjoys six. (Although it is only the "additive" properties of these Galois fields that we use for coloring the triad and hexad generators, their multiplicative properties suggest the structure of the most elegant and efficient transition rules in the coded algorithms.)

Since the generators of each triad or hexad are now assigned different "colors", then, to avoid a numerical clash of intersecting line filters at grid points, while allowing operations to proceed in parallel, it is sufficient to do all line operations of one color, at one session, and complete the session before another color's operations are begun.

Early experiments with these methods revealed slightly ugly numerical artifacts in the synthesized covariances at places where triads or hexads changed. Using numerical schemes designed to exaggerate these "dislocations" we show an array of nine impulses inhomogeneously smoothed by the basic triad algorithm in Fig.4, where we see that, even distributing the smoothing over four successive diffusing iterations (panel b), does not entirely eliminate the effect. The cure involves enlarging the allowed set of smoothing directions at each point so that a single triad or hexad can be replaced by an equivalent symmetric superposition of triads or hexads that form a "ball" in the linear aspect tensor space, with the ball centered on the original aspect tensor "point" in this space. Again we invoke the (approximate) linearity of spreads to justify this remedial measure, in which the amount of the fragment of the ball falling within the province of each triad or hexad is "weighed" and collected to the centroid of this fragment. It can be shown that, by capping the radius of such a ball, the number of triads or hexads intersected is limited to a number that implies no more than a total of four smoothing directions for such "blended triads" and no more than 13 directions for "blended hexads". The ball is regarded as a blending kernel that effectively mixes (symmetrically) a neighborhood (in aspect tensor space) of the smoothing operators around the intended aspect tensor value. By such a blending the cause of unsightly dislocations in the covariance field, which can be traced to an excessive rate of change of filter coefficients near the transition boundaries, is effectively removed. Fig. 5 shows the results obtained with the blended triads, where now, even a single iteration shows no unpleasant numerical artifacts. In the blended algorithms, Galois sequencing for conflict-free parallel algorithms are still possible, though the respective fields need to be larger; in 2D we now find that the eight non-null elements of GF(9) can be paired up to provide the needed four colors for the blended triads; in 3D, the 26 non-null elements of GF(27) pair up to provide the 13 colors required to segregate the 13

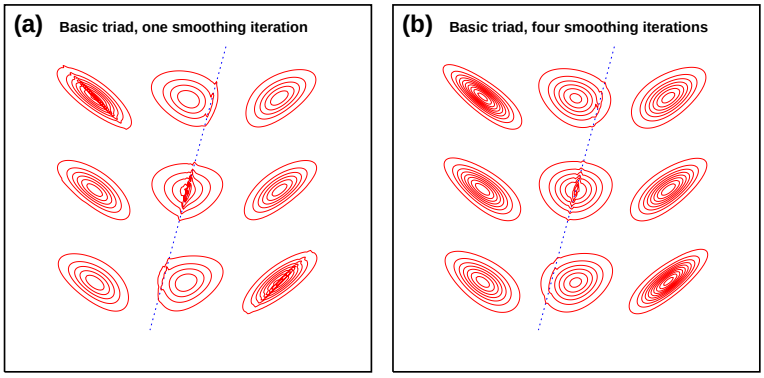

Figure 4: Examples of "dislocations" occurring along the lines of transition between two triads' domains when smoothing by the basic triad (but with filter numerics selected to exaggerate the visual impact of the effect). Even after distributing the smoothing by four sequential smoothing operations, the dislocation remains visible.

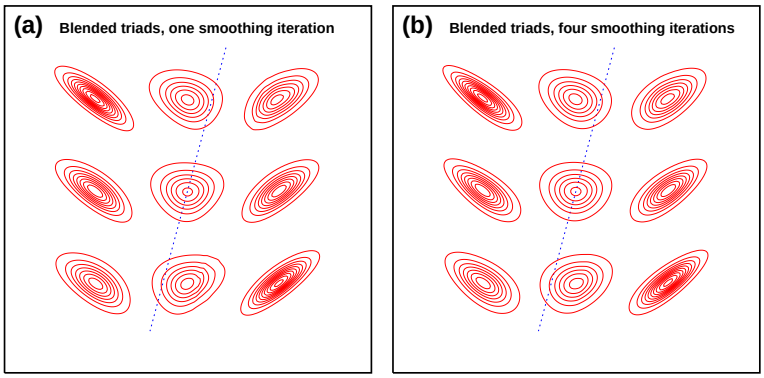

Figure 5: Examples showing the results of applying the "blended triads" method to perform the same smoothing operations as in Fig. 4. The dislocations incurred by the basic triad method no longer intrude.

directions required by each blending of hexads. The full details of these technique are given in Purser (2005).

4 The problem of normalization

A homogeneous diffusive process in Euclidean geometry and without boundaries is easy to normalize whether or not it is isotropic – the final function has a standard Gaussian form and the amplitude, for aspect tensor A, will be $|2\pi A|^{-N/2}$ in N dimensions. Weaver and Courtier (2001) address the isotropic problem in the spherical case. The difficulty occurs when the diffusivity varies, or the domain is significantly non-Euclidean. At present we are using a randomized-trace Monte Carlo method similar to those described by Rabier and Courtier (1992), Weaver and Courtier (2001), Weaver and Ricci (2004), but the cost is substantial if we want to adapt to a different covariance from one analysis time to another. We are therefore exploring other potentially more efficient methods of approximation based on the small-t asymptotics of the inhomogeneous diffusion equation.

The case of spatially varying anisotropic diffusivity can be mapped into an equivalent problem of *uniform* and *isotropic* diffusivity in a non-Euclidean (Riemannian) space, albeit with a spatial "capacitance" function of position appearing in the transformed diffusion problem. (Capacitance may be thought of as a varying gauge of metal forming a thin smoothly curved shell in the 2D heat-flow analogy.) But what is gained by such an algebraic manoeuvre?

First, we note that, since we are not simulating "real" diffusion, but merely taking advantage of a convenient numerical way to generate self-adjoint quasi-Gaussian covariance contributions, we are at liberty to choose the capacitance in the *untransformed* problem that anticipates a uniformity of this quantity once transformed. The original aspect tensor effectively assumes the role of contravariant metric tensor, g^{ij}, in the transformed problem in which diffusivity is unity isotropically. What we gain is that the amplitude can now only depend upon the distribution of (intrinsic) *curvature* of the transformed space – a second derivative of the metric when examined in local (Riemann) "normal coordinates". In a way, we have simplified the factors upon which the amplitude of the filtering kernel can depend, having identified the only possible source of the amplitude variation to be the curvature of the transformed space. The transformed diffusion equation is now *always* of the standard form, expressed in tensor notation:

$$\frac{\partial P}{\partial t} = \frac{1}{g^{\frac{1}{2}}} \frac{\partial}{\partial x^i} g^{\frac{1}{2}} g^{ij} \frac{\partial P}{\partial x^j}. \tag{8}$$

Note that some choices of aspect tensor distribution, such as those determined by a mapping to a deformed domain geometrically identical to the physical one, but in which covariances regain their horizontally isotropic appearance (like the semi-geostrophic space suggested by Desroziers 1997) will actually solve the normalization immediately by recovering an effectively "flat" transformed space. More generally, it is necessary to deal with the effects of the space's intrinsic curvature.

Our problem, the asymptotic approximation to the "heat kernel" in Riemannian geometry is one in which mathematicians have taken a lively interest for a variety of reasons and there is a considerable literature on this fascinating subject (McKean, 1970; Gilkey 1984; Rosenberg 1997, Grigor'yan and Noguchi 1998, Lafferty and Lebanon 2005). The next section discusses some ideas that can possibly improve the computational efficiency and reliability of the inhomogeneous filter amplitude estimation.

5 Differential geometry and the "Parametrix" expansion

In the general case, the transformation to a Riemannian geometry might not seem to have made the amplitude estimation problem any easier. However, it does seem to simplify the extraction of a useful asymptotic expansion for the "amplitude quotient", that is, the ratio of the actual diffusion filter amplitude divided by the corresponding Gaussian formula for it. The approach is the "Parametrix method". Normal coordinates are constructed to be "as Cartesian as possible locally"; axes are orthogonal at the local origin, through which straight-line radials in the normal coordinates are actually geodesics of the Riemann geometry and distances are true along them. The idea behind the parametrix method is to represent the evolving solution of the diffusion problem as the corresponding Euclidean solution in the normal coordinates, multiplied by a modulating function that is smooth in time t and in space in the neighborhood of the origin.

Then, for sufficiently small t the modulating function can be expanded (asymptotically) in a series of powers of the normal coordinates and in t. The only part of the solution we shall end up being interested in is the part at the local origin and at $t = 1/2$, and the terms which contribute to the successive powers of t in this evaluation can be shown to be polynomials in the curvature tensor and its covariant derivatives. The recursive process by which successive terms are developed resembles a more complicated version of the "method of Frobenius" used to find power series solutions of ordinary differential equations.

While, in principle, this process can be continued without limit, in practice it rapidly leads to extremely complicated algebra. The successive approximations, typical of many asymptotic series, do not necessarily converge and, when they do converge, do not necessarily converge to the true solution, even for the smallest t. Nevertheless, provided the original aspect tensor varies sufficiently smoothly and gradually, the approximate solutions obtained by this approach should be adequate. Most of the algebraic operations involved can actually be reduced to forms that lend themselves to mechanization. We give idealized examples of the results obtained from the asymptotic method but only give a fragmentary sketch of the steps required for the general case. (Details

will be supplied in a pair of forthcoming NCEP office notes treating this entire topic.)

In the special 2D case where there is axisymmetry in the curvature about the local origin, the curved-space diffusion equation is:

$$\frac{\partial P}{\partial t} = \frac{1}{g^{\frac{1}{2}}} \frac{\partial}{\partial r} g^{\frac{1}{2}} \frac{\partial P}{\partial r}, \tag{9}$$

where P is the quantity diffused, g is a transverse metric term, r the radial distance. The solution of (9) is related to the solution to the Euclidean diffusion equation,

$$\frac{\partial P_0}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} r \frac{\partial P_0}{\partial r}, \tag{10}$$

which is:

$$P_0 = \frac{1}{4\pi t} \exp\left(-\frac{r^2}{4t}\right), \tag{11}$$

by assuming

$$P(t,r) = (1 + R(t,r)) P_0(t,r). \tag{12}$$

The radial distribution of g can be related to the profile of Gaussian curvature, κ, and leads to a forcing function for the asymptotic problem which we can put into the form,

$$\mathscr{G} = Dr^2 + Er^4 + \ldots, \tag{13}$$

Then, after some algebra, we find that the modulating function R must everywhere satisfy a balance among three sub-expressions, conveniently written:

$$0 = \mathscr{R} + \mathscr{G} + \mathscr{Q}, \tag{14}$$

where $\mathscr{R}$ is a combination linear in R and not involving any of the curvature terms in $\mathscr{G}$:

$$\mathscr{R} = t\left(\frac{\partial R}{\partial t} - \frac{1}{r}\frac{\partial R}{\partial r} - \frac{\partial^2 R}{\partial r^2}\right) + r\frac{\partial R}{\partial r}, \tag{15}$$

while $\mathscr{Q}$ is a term separately linear in $\mathscr{R}$ and in $\mathscr{G}$:

$$\mathscr{Q} = \mathscr{G}\left(R - \frac{2t}{r}\frac{\partial R}{\partial r}\right). \tag{16}$$

When we come to expand R in powers:

$$R(t,r) \approx \sum_{s,p \geq 0} R_{s,p} t^s r^p, \tag{17}$$

we notice that the expression for $\mathscr{R}$, while *inhomogeneous* separately in its effects on the time and space exponents, s and p, is *homogeneous* in its effect upon the "slab index" defined, $h = 2s + p$. This fact suggests the asymptotic procedure of iteratively solving (14) at successive integer values of this h.

The first nontrivial result of this procedure, for $h = 2$, can be shown not to involve the $\mathscr{Q}$ term, but to give the simpler balance between the $\mathscr{R}$ and $\mathscr{G}$ terms at this degree that consequently imply,

$$R^{(2)}(t,r) = -2Dt - \frac{1}{2}Dr^2, \tag{18}$$

where the parenthetic superscript on R indicates the maximum degree h of the contributing terms. The quantity D can be shown to be proportional to the central Gaussian curvature, $K = \kappa(0)$:

$$D = -\frac{K}{6}, \tag{19}$$

so that, evaluating the amplitude quotient, $A = R(\frac{1}{2}, 0)$, at the nominated standard time $t = 1/2$, one obtains the first order amplitude quotient (relative to the planar Gaussian formula):

$$A^{(1)} = 1 + \frac{K}{6}. \tag{20}$$

The parenthetic superscript on A is interpreted slightly differently; it now denotes the maximum contributing degree of K or its derivatives (but *not* of h, as we had earlier for R). It is found that, upon working through the solution corresponding to the next non-trivial slab index, $h = 4$ (where we *do* now need to involve $\mathscr{Q}$), and expanding E in terms of K and the Laplacian, at the center, of the curvature, κ:

$$E = \frac{-K^2}{90} - \frac{\nabla^2 \kappa}{40}, \tag{21}$$

the next order of approximation produces:

$$A^{(2)} = 1 + \frac{K}{6} + \frac{1}{60}(K^2 + \nabla^2 \kappa), \tag{22}$$

and so on.

For the idealized case of uniform curvature in 2D and 3D the asymptotic solutions, drastically simplified by the absence of any of the derivatives of curvature, nevertheless reveal some features that are instructive about the general case.

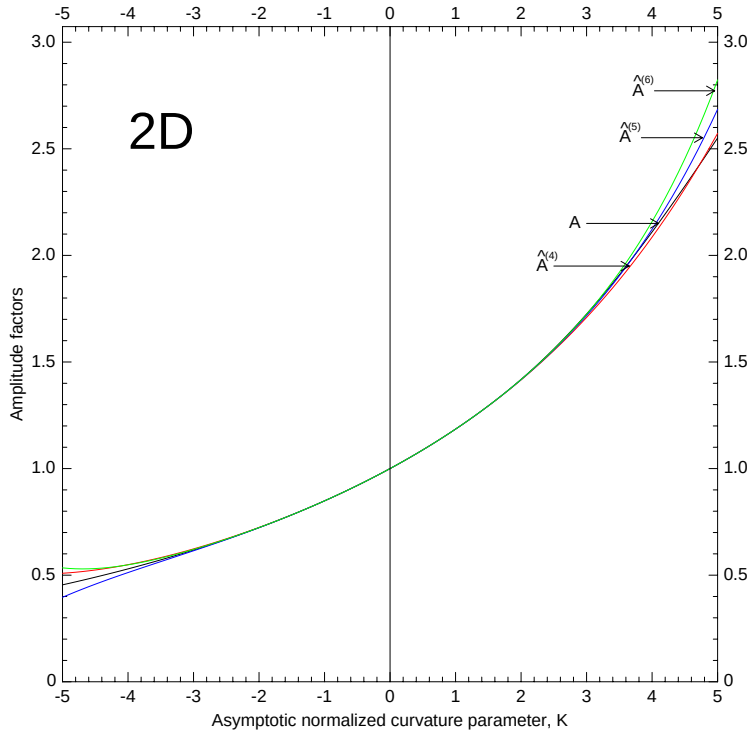

Figure 6: Comparison between true 2D amplitude quotient, $A(K)$, and a few of the small-K asymptotic expansions to it, truncated to the degree indicated by the parenthetic superscript in each case, for the geometries (hyperbolic and spherical) of constant curvature. The series is a divergent one, though very accurate in the first few terms even for the relatively large K shown at the limbs of this graph.

In the 2D case, depicted in Fig. 6, the series expansion is asymptotic in the typical sense, being divergent at any K when continued to ever larger truncations. In contrast, the 3D case actually exhibits a *convergent* expansion, moreover, to the simple analytic function,

$$\hat{A}(K) = \exp(K/2). \tag{23}$$

Also, for negative K, corresponding to a hyperbolic geometry, this is also the true solution, $\hat{A}(K) = A(K)$, $K \leq 0$. However, the same *convergent* asymptotic expansion gives an erroneous limit for positive K (the "3-sphere"), as the graph of Fig. 7 confirms. (The behavior of the asymptotics for odd- and even-dimensioned spaces alternates, with the 2D and 3D cases providing the exemplars for the general pattern.) In general Riemannian

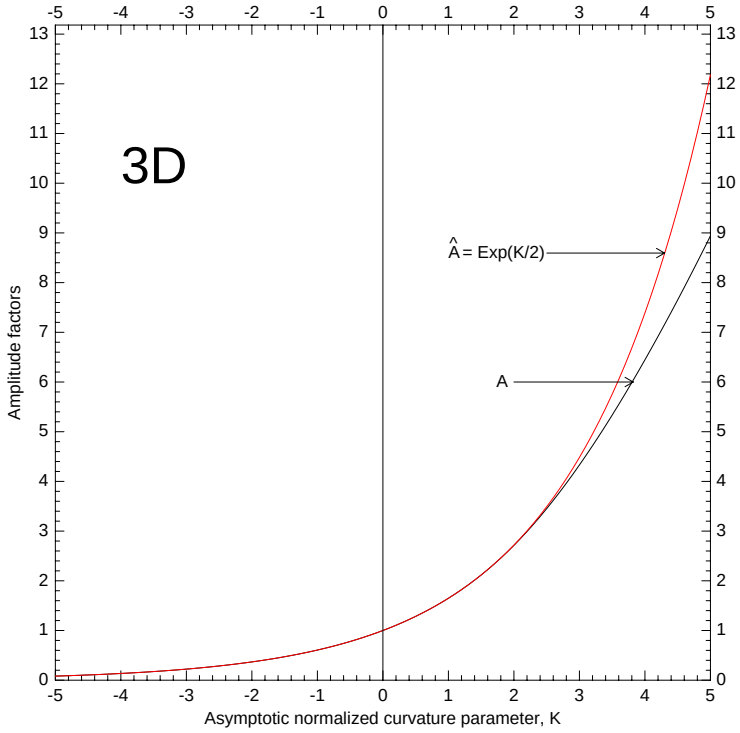

Figure 7: Comparison between true 3D amplitude quotient, $A(K)$, and the convergent limit of the asymptotic expansion approximating it. This limit converges to the true solution only at negative K.

geometry, the division of the terms that supply the constraints for the parametrix expansion still produces three kinds of sub-expressions, $\mathscr{G}$, $\mathscr{R}$ and $\mathscr{Q}$, which must balance at each successive slab index. Naturally we must generalize the slab index; in 2D it is generalized to $h = 2s + p + q$ and in 3D to $h = 2s + p + q + r$, where the terms of R are correspondingly generalized (in 3D):

$$R(t,x,y,z) = R_{s,p,q,r} t^s x^p y^q z^r. \tag{24}$$

The expansions lead to each asymptotic term being a polynomial primarily in the tensorial quantities that constitute the Taylor series coefficients of the (covariant) metric tensor in the normal coordinates; but by a theorem of Cartan (see Berger et al. 1971) relating these tensors to universal polynomials in the Riemann curvature tensor, $R^i{}_{jkl}$, and its covariant derivatives, it is then possible (in principle) to express the desired coefficients for the approximation to the amplitude quotient also in terms of a finite combination of the Riemann tensor and its first few covariant derivatives. In 3D all the curvature is expressible in terms of the simpler symmetric "Ricci" tensor (e.g., Synge and Schild 1949, Kreyszig 1991), defined as the contraction of the Riemann tensor:

$$R_{jl} = R^i{}_{jil}, \tag{25}$$

for in this case,

$$R_{ijkl} = g_{ik}R_{jl} + g_{jl}R_{ik} - g_{il}R_{jk} - g_{jk}R_{il} - (g_{ik}g_{jl} - g_{il}g_{jk})R/2, \tag{26}$$

where the "Ricci scalar",

$$R = R^i_i, \tag{27}$$

is the contraction (i.e., the trace) of the Ricci tensor itself. In 2D, curvature is further restricted to the Gaussian curvature, κ, which equals one half (minus one half, in some notations) of the Ricci scalar.

Using a result from Gilkey (1984), we find that the N-dimensional generalization of the second order amplitude quotient approximation, (22) is

$$A^{(2)} = 1 + \frac{R}{12} + \frac{1}{1440}\left(2R_{ijkl}R^{ijkl} - 2R_{ij}R^{ij} + 5R^2 + 12\nabla^2 R\right), \tag{28}$$

which, by (26), specializes for dimensions $N \leq 3$, to:

$$A^{(2)} = 1 + \frac{R}{12} + \frac{1}{480}\left(2R_{ij}R^{ij} + R^2 + 4\nabla^2 R\right). \tag{29}$$

The prospect of employing such a method, even at this low order of expansion, shows some promise of being able to improve efficiency of the amplitude estimation. Being an asymptotic method, we expect it to fail for values of the curvatures or their derivatives exceeding about $\mathcal{O}(1)$. However, we can use the computed diagnostics of these quantities, even before applying any amplitude adjustment, to assess whether we are within the parameter range for which such an adjustment can be expected to be valid; if the values are diagnosed to be excessive, we can take remedial measures by smoothing or reducing the variability of the aspect tensor components that we begin with.

6 Covariance synthesis by multifilter superposition in a multigrid framework

Although discussion up to this point has focused on the technicalities involved in the construction of Gaussian contributions to the covariance, we should not ignore the fact that the simple quasi-Gaussian that results from a *single* application of a diffusion-simulating filter is probably *not* a very realistic model of background uncertainty by itself.

If we can afford to construct a superposition of such quasi-Gaussians, each at a different scale and relative amplitude, then we can sculpt a more promising and realistic covariance shape. (Actually, it is better to use this approach to construct the asymmetrical *square-root* of the covariance, as this way, it becomes much easier to ensure the final result is positive definite.)

The computational expense of applying the recursive filters repeatedly, but at different scales, could be greatly mitigated by adopting a multigrid strategy. Then, filtering at the coarse scales is done (cheaply) on only a correspondingly coarse grid. The contribution is interpolated to the next finer grid and added to another quasi-Gaussian contribution appropriate to this scale, and so on. Of course, the quantity being smoothed must also be carried consistently *up* this hierarchy in the first place, using the adjoint of the grid-refinement interpolators.

7 Analysis error estimation and characterization; Preconditioning

In classical data assimilation theory, we understand the analysis precision to be simply the sum of the precisions of background and measurements and, if observations were spatially homogeneous, we could ascribe to them a local error power spectrum and obtain the corresponding analysis power spectrum directly.

If, with realistic observation distributions, we artificially assume this homogeneity locally, then spatially adaptive filters with profiles shaped by the multigrid method discussed in the previous section, would be a useful way to characterize the resulting errors. As discussed by Lorenc (1992), the appropriate localized spectral distribution of the precision for a scatter of point observations of the analyzed variable should be flat ("white noise"), at least over the range of wavenumbers resolved. More generally, the spectral form of vertically and

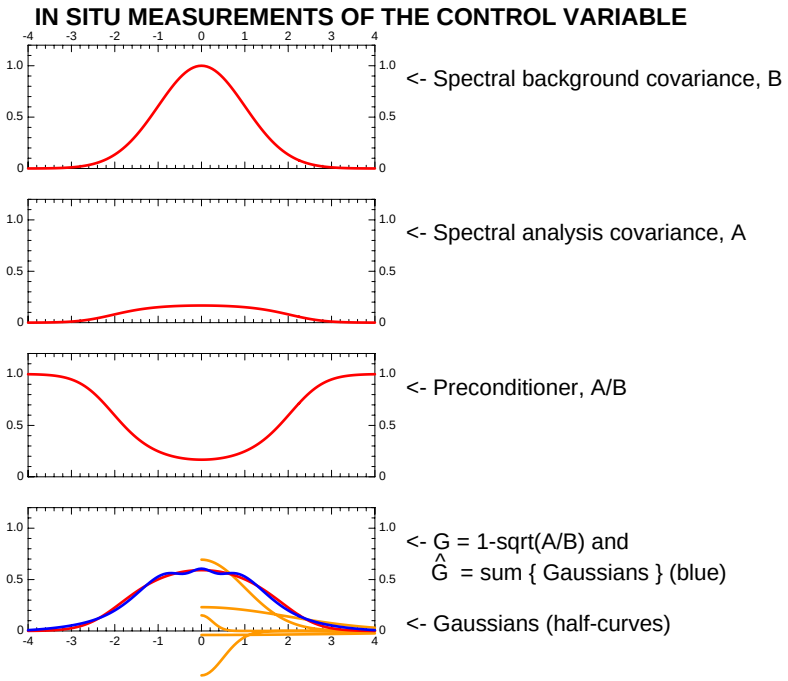

Figure 8: Idealized plots showing the shapes in the spectral domain of a Gaussian background field error covariance, B, and the analysis covariance A obtained after the assimilation of a spatially-homogeneous distribution of in situ measurements of the analysis variable. The filter having the response equal to the ratio of these, A/B, would serve as a preconditioner for this problem. The complement of the square-root of this preconditioning profile, shown as the red curve, G, in the bottom frame, can be approximated by a superposition, $\hat{G}$, of a small number of Gaussian filters, whose contributions are also seen as half-curves.

horizontally homogenized data of a given type should be approximately the spatial density of this data type, times the Fourier transform of the operator,

$$\hat{R}^{-1} = H^T R^{-1} H \tag{30}$$

that represents the projection of the measurement precision for a datum of this type into the analysis domain (we use here the standard notation suggested by Ide et al., 1997). The simplest case of in situ measurements of the actual analysis variable, illustrated schematically in Fig 8, suggests a way of generating a covariance roughly equivalent to the analysis covariance, by following the background error covariance filter with the filter associated with the profile, marked A/B, in the third frame. A way of generating an approximation to this operator is suggested by the bell-shaped profile of the complement of A/B or, more appropriately, the square-root of A/B, since the latter representation makes it easier to ensure that an approximating sequence of filters for A/B maintains positive definiteness. The fourth panel shows how well the complement, $G = 1 - \sqrt{(A/B)}$, can itself be approximated as a small superposition of appropriately weighted Gaussians which, of course, can be implemented efficiently. If our analysis variable is a streamfunction and our measurements are winds, then the operator $\hat{R}^{-1}$ has a parabolic spectral profile, which consequently changes the qualitative shape of the analysis, as shown in Fig. 9. Then again, if we interpret the abscissae as denoting vertical wavenumbers, we can also consider the profiles one might obtain from infrared or microwave passive sounder data (here modelled by the idealized profile suggested by Gautier and Revah, 1975), and where the smoothed out measurement of temperature represents the *vertical* derivative of a geopotential-like analysis variable. This case is illustrated in Fig. 10. Even in these difficult cases, the superposition of a handful of Gaussian filters seems capable of producing a reasonable fit to the ratio A/B. One use of such filters would then be to adjust the deviations of ensemble members from their mean in order to account for the proper shrinkage of an ensemble associated with the injection of information from measurements. Another possible role, as suggested by the side captions of the third frames of these figures, is to provide a preconditioner for the cost-function descent algorithm.

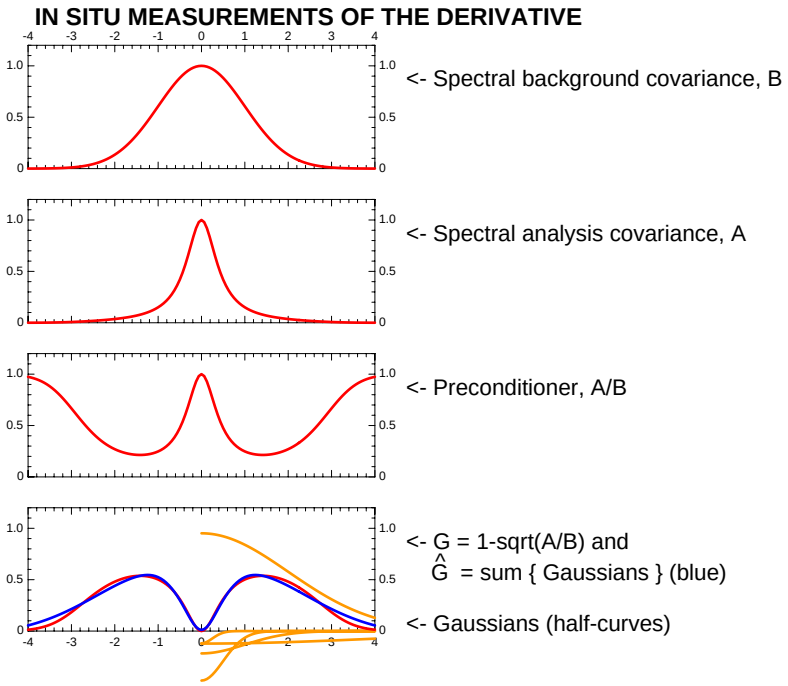

Figure 9: As in Fig. 8, except for the case of measurements of the derivative of the analysis variable.

8 Analysing variables with abrupt gradients

Earlier, we discussed the problem of adequately normalizing the amplitudes of filters when the aspect tensors vary gradually and smoothly with location, and mentioned that asymptotic methods show promise in performing this task efficiently. But, in cases where the covariance really does change abruptly, the parametrix expansion method can be of no value. However, such abrupt changes are often tied closely to topographic transitions, such as sharp gradients in terrain height, or land/water boundaries. The difficulties these sharp transitions cause are particularly troublesome when carrying out NCEP's 2D "Real-Time Mesoscale Analysis" (RTMA, de Pondeca et al., 2007) of surface analyzed parameters over the US, where such topographically-tied covariance changes do occur. One remedial approach we are giving some attention is based on a variant of Riishøjgaard's (1998) method.

The analysis could be carried out once for an elevation (say) that is somewhat higher than the local average obtained by heavily smoothing the true elevation, so that the covariances on this hypothetical surface can themselves now be taken to be smooth. Another, completely independent analysis is performed at a parallel smoothed elevation that is somewhat lower than the local average. The extra vertical interpolations are computed in accordance with the prescribed part of the covariance, which, as in Riishøjgaard's method, depends upon the extra variable (in this case, elevation). Thus, the observations are only permitted to influence the respective analyses in proportion to how close their actual elevations are relative to the nominal elevations of the two hypothetical surfaces. The final analysis is similarly the interpolation between the two surface's analyses to a target at the true elevation at each geographical position. The result is expected to give realistically rapid changes in surface analyzed variables at places where the topography changes rapidly, while the underlying covariances seen by the generating filters remain perfectly smooth, allowing a reliable and accurate normalization of them. It might be possible to extend these ideas to 3D if the duplicated analysis domains are confined to only the lowest levels.

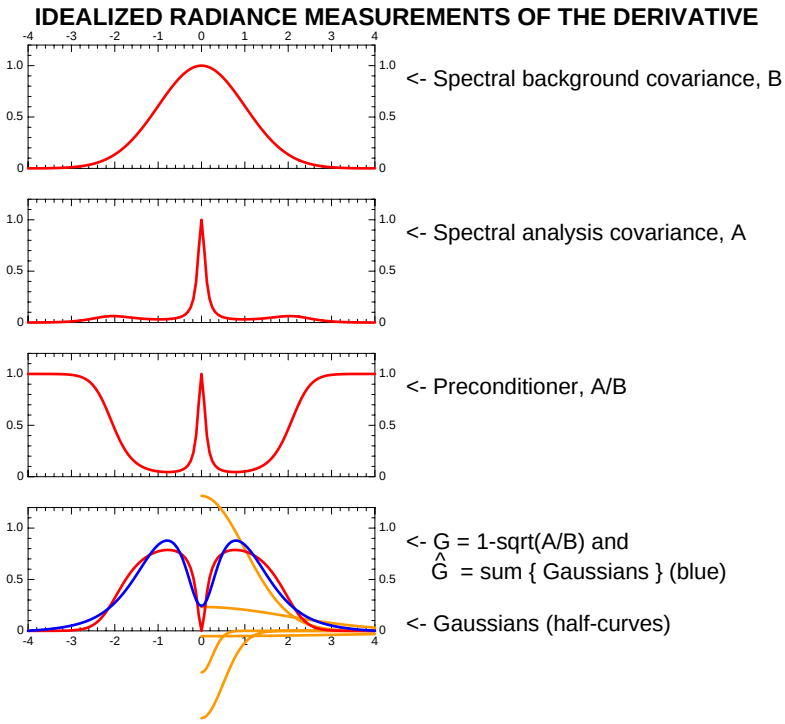

Figure 10: As in Fig. 8, but illustrating vertical spectral profiles associated with radiance measurements of the derivative (temperature) of the analysis variable (geopotential).

9 Conclusions

The numerical grid-smoothing technique of the Recursive Filter has shown itself to be a versatile tool for the efficient construction of approximate models of background error covariances. In combination with the Triad and Hexad methods, we can now generate smooth quasi-Gaussian covariance contributions with arbitrary degrees of anisotropy. Asymptotic methods based on differential geometric ideas are being developed and tested in the GSI at NCEP to make the normalization of these filters in inhomogeneous cases more accurate and efficient, while multigrid methods are being developed to facilitate the efficient construction of covariances having more general profile shapes than can be obtained with only a few Gaussians.

Finally, we have discussed some more speculative ideas concerning how future developments of these techniques might address the problems of analysis error characterization, preconditioning, and covariance modelling where abrupt changes occur.

Acknowledgements

The authors are grateful to Drs. John Derber, Geoff DiMego and Steve Lord for their support, encouragement and discussions.

References

Berger, M, P. Gauduchon, and E. Mazet, 1971: *Le Spectre d'une variété riemannienne,* Lecture Notes in Mathematics, vol. 194, Springer-Verlag, Berlin.

De Pondeca, M.S.F.V., and Coauthors, 2007: Analysis System at NCEP. Preprints, 22nd Conf. on WAF/18th Conf. on NWP, Park City, UT, Amer. Meteor. Soc., 4A.5.

Derber, J. C., and A. Rosati, 1989: A global ocean data assimilation system. *J. Phys. Ocean.,* **19,** 1333–1347.

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

Egbert, G. D., A. F. Bennett, and M. G. G. Foreman, 1994: TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res.,* **99,** 24821–24852.

Gautier, D., and I. Revah, 1975: Sounding of planetary atmospheres: A Fourier analysis of the radiative transfer equation. *J. Atmos. Sci.,* **32,** 881–892.

Gilkey, P. B., 1984: *Invariance Theory, the Heat Equation, and the Atiyah-Singer Index Theorem.* Publish or Perish, Wilmington, DE, 345 pp.

Grigor'yan, A., and M. Noguchi, 1998: The heat kernel on hyperbolic space. *Bull. London Math. Soc.,* **30,** 643–650.

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.

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.

Kreyszig, E., 1991: *Differential Geometry* Dover, New York. 352 pp.

Lafferty, J., and G. Lebanon, 2005: Diffusion kernels on statistical manifolds. *J. of Machine Learning Research,* **6,** 129–163.

Lorenc, A. C., 1992: Iterative analysis using covariance functions and filters. *Quart. J. Roy. Meteor. Soc.,* **118,** 569–591.

McKean, H. P., 1970: An upper bound to the spectrum of Δ on a manifold of negative curvature. *J. Differential Geometry,* **4,** 359–366.

Purser, R. J., 1983: Mesoscale analysis at U.K. Meteorological Office. In "Current problems in data assimilation". ECMWF Workshop Proceedings, 8–10 November 1982, 253–265.

Purser, R. J., 2005: A geometrical approach to the synthesis of smooth anisotropic covariance operators for data assimilation. NOAA/NCEP Office Note 447, 60 pp.

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. 17 pp.

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

Rabier, F. and P. Courtier, 1992: Four-dimensional data assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteor. Soc.,* **118,** 649–672.

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

Rosenberg, S., 1997: *The Laplacian on a Riemannian Manifold; London Mathematical Society Student Texts 31.* Cambridge, 174 pp.

Synge, J. L., and A. Schild, 1949: *Tensor Calculus.* Dover, New York, 324 pp.

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

Weaver, A., and S. Ricci, 2004: Constructing a background error correlation model using generalized diffusion operators. In "Recent developments in data assimilation for atmosphere and ocean". ECMWF Seminar Proceedings, 8–12 September 2003, 327–339.