



Covariance Modeling-- Principles

R. James Purser

SAIC at NOAA/NCEP/EMC

Camp Springs, Maryland.



Collaborators:

Dave Parrish

John Derber

Wan-Shu Wu

Manuel de Pondeca

Yoshiaki Sato

Dezso Devenyi

Which covariances?

In standard notation, and assuming linearity of all operators, two equivalent statements of the equation of statistical analysis are:

$$x^a = x^b + \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} (y - \mathbf{H} x^b)$$

and

$$x^a = x^b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (y - \mathbf{H} x^b)$$

When we talk about the “covariances” of data assimilation, it is normally the background error covariances gathered into the giant matrix, \mathbf{B} , that we are referring to.

\mathbf{B}^{-1} is generally not available to us, but \mathbf{C} is, where $\mathbf{C} \mathbf{C}^T = \mathbf{B}$.

Therefore, we might write:

$$x^a - x^b = \mathbf{C} v$$

Solve for v :

$$v = + (\mathbf{I} + \mathbf{C}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{H}^T \mathbf{R}^{-1} (y - \mathbf{H} x^b)$$

Then solve for x^a

The large linear system involved in the solution for v is generally done by some form of “Krylov” method, such as Conjugate Gradients.

Then the solution only requires application of the operators, \mathbf{I} , \mathbf{C} , \mathbf{H} , \mathbf{H}^T , \mathbf{R}^{-1} .

The nonlocal operator is \mathbf{C} . Multiplying by it could potential be very expensive.

It is always convenient to decompose the analysis variables into their different kinds of dynamic fields. For example, From variables, ϕ, u, v, h , we might derive:

- balanced (“geostrophic”) streamfunction
- velocity potential
- unbalanced streamfunction
- humidity

We simplify **B** by assuming the covariances among DIFFERENT members of the above set vanish.

In these talks I shall discuss methods that are devoted mostly to the efficient numerical execution of the operator, **C** (and hence **B**).

An explicit “matrix multiply” for **C** will not do --- for N grid points this would entail the vast number, $N*N$ operations. Any viable matrix operation for such a large N would have to be with a matrix that is narrowly banded, or the product of a small number of such matrices.

The methods I describe here do involve narrow-band matrices, but rather than directly multiplying with them, we shall find that the appropriate operations are “back-substitution”, which becomes possible when the matrices are both banded and triangular.

The operations we are describing are those associated with **RECURSIVE FILTERS**.

Let’s look at these filters for 1D univariate data.

Simplest recursive filters

(First order, constant coefficient)

Start with a gridded input of scalar field, V_0 . At grid point j let its value be $(V_0)_j$

Let the filter coefficient be α .

The output value of the FORWARD moving filter is $(V_1)_j$ and is given by,

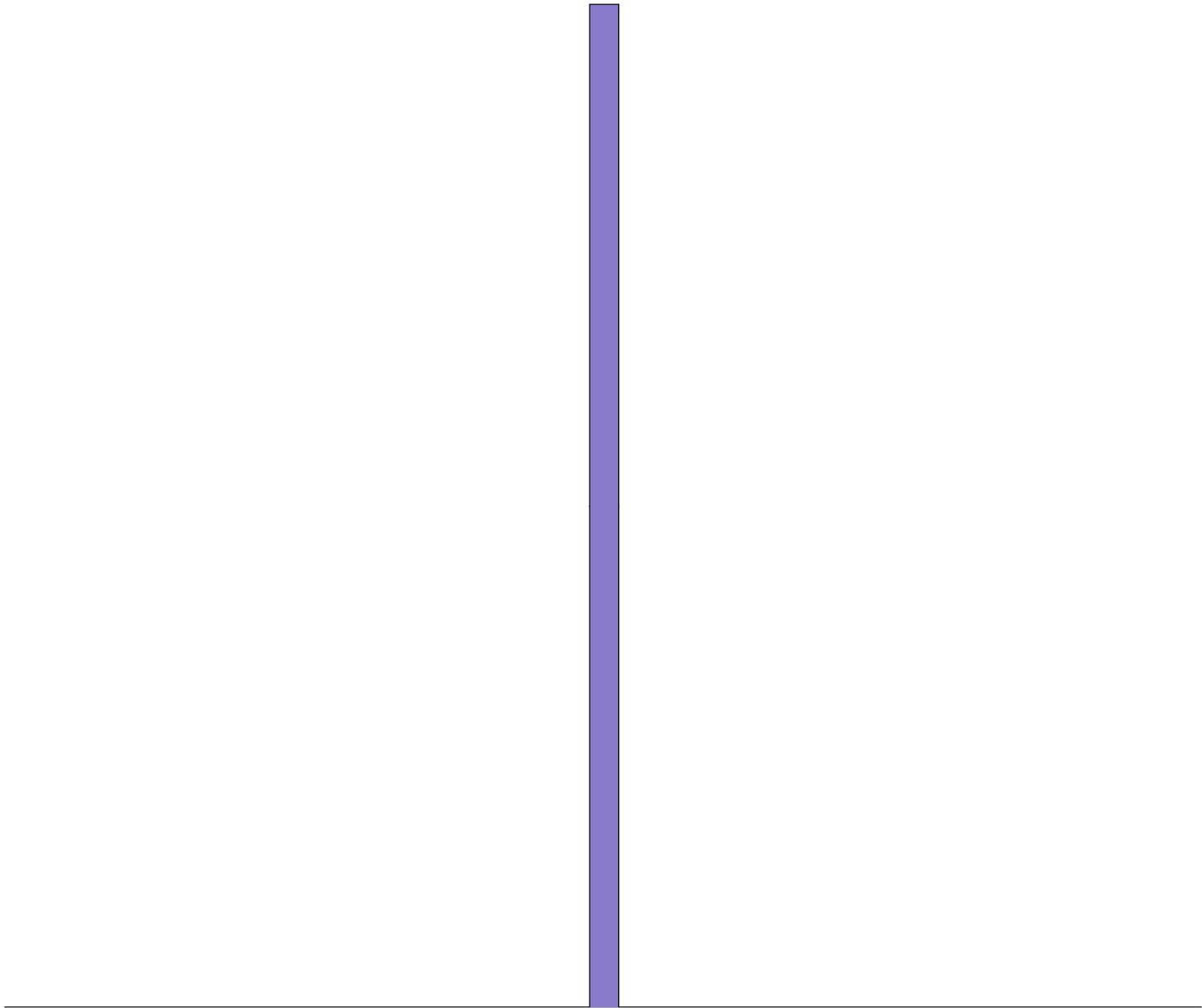
$$(V_1)_j = (1 - \alpha)(V_0)_j + \alpha(V_1)_{j-1}$$

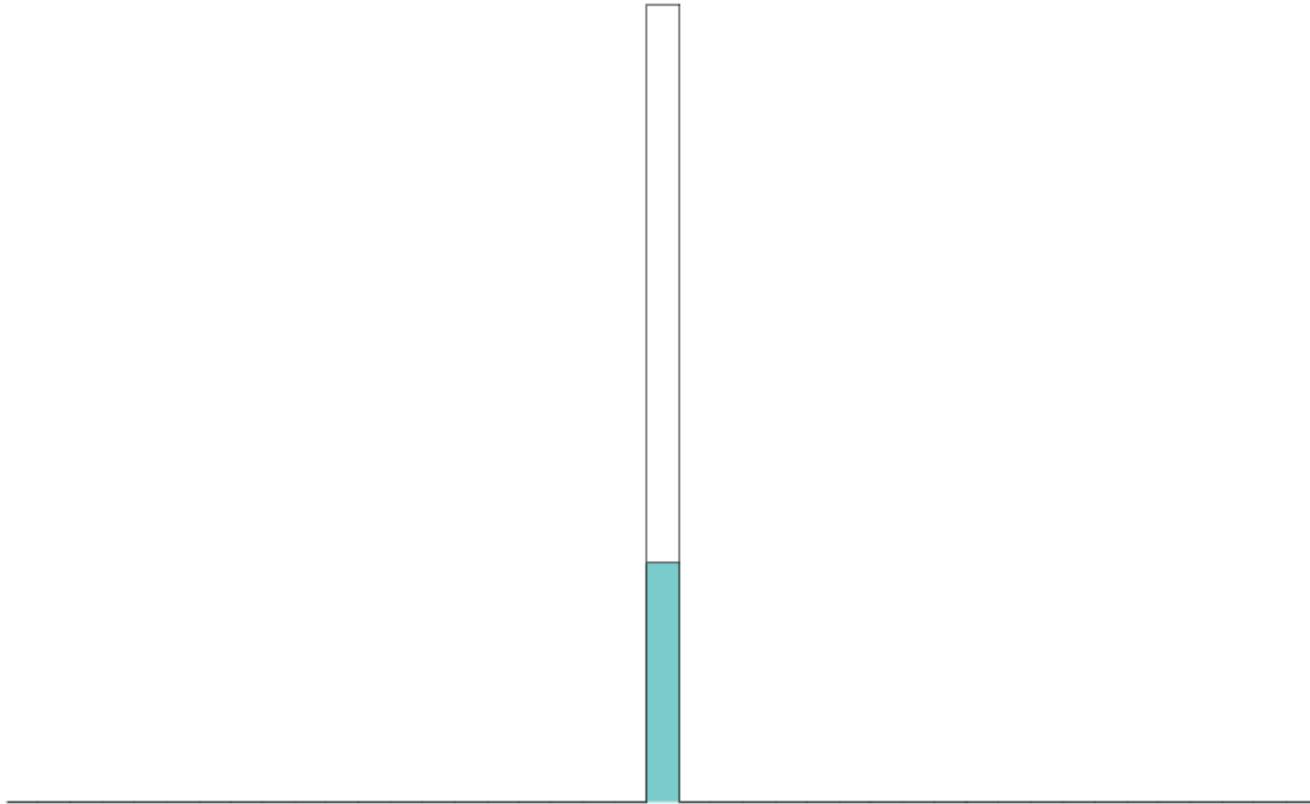
NOTE: This is OUTPUT

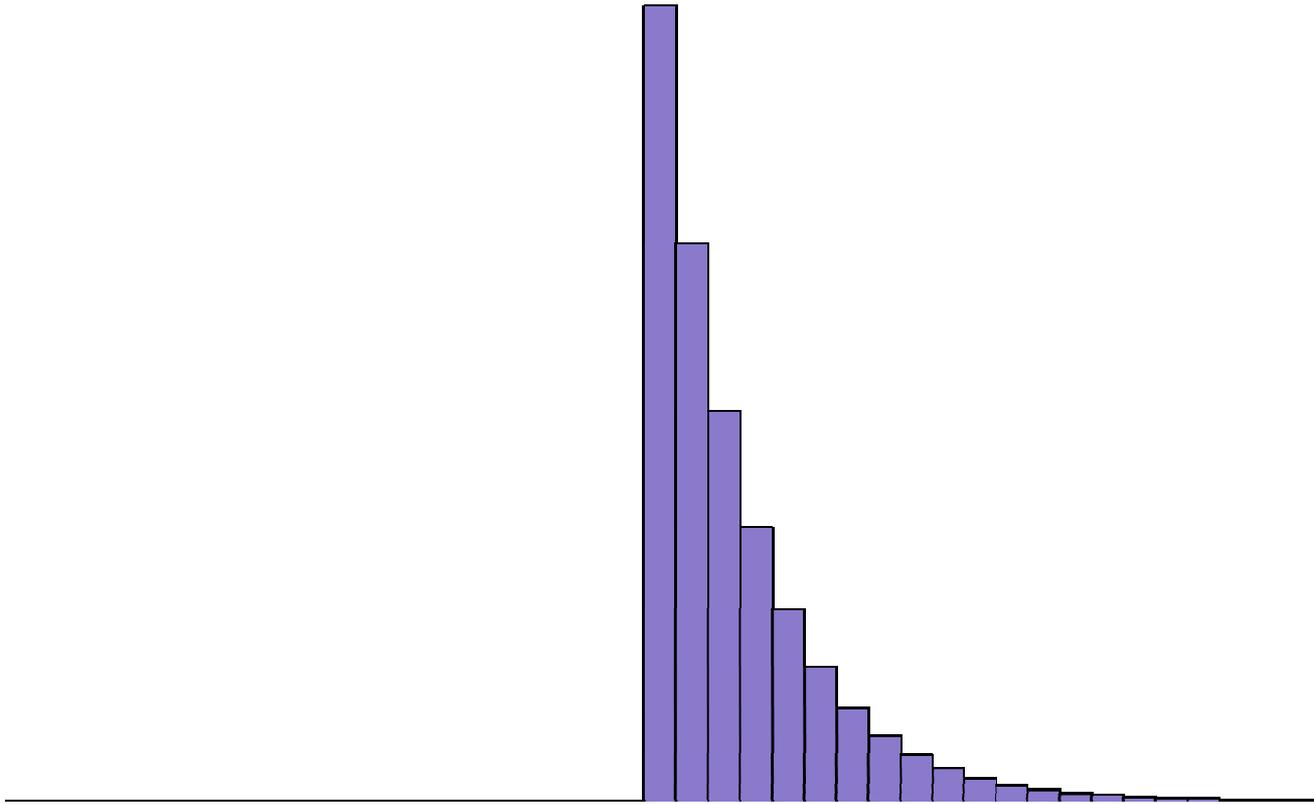
Note that the application, sequentially in index position, j , is indeed of a numerical cost proportional to the number, N , of grid points.

It is a highly efficient operator.

We will take a look at what it does.







The result is a one-side exponentially-weighted mean of the inputs.

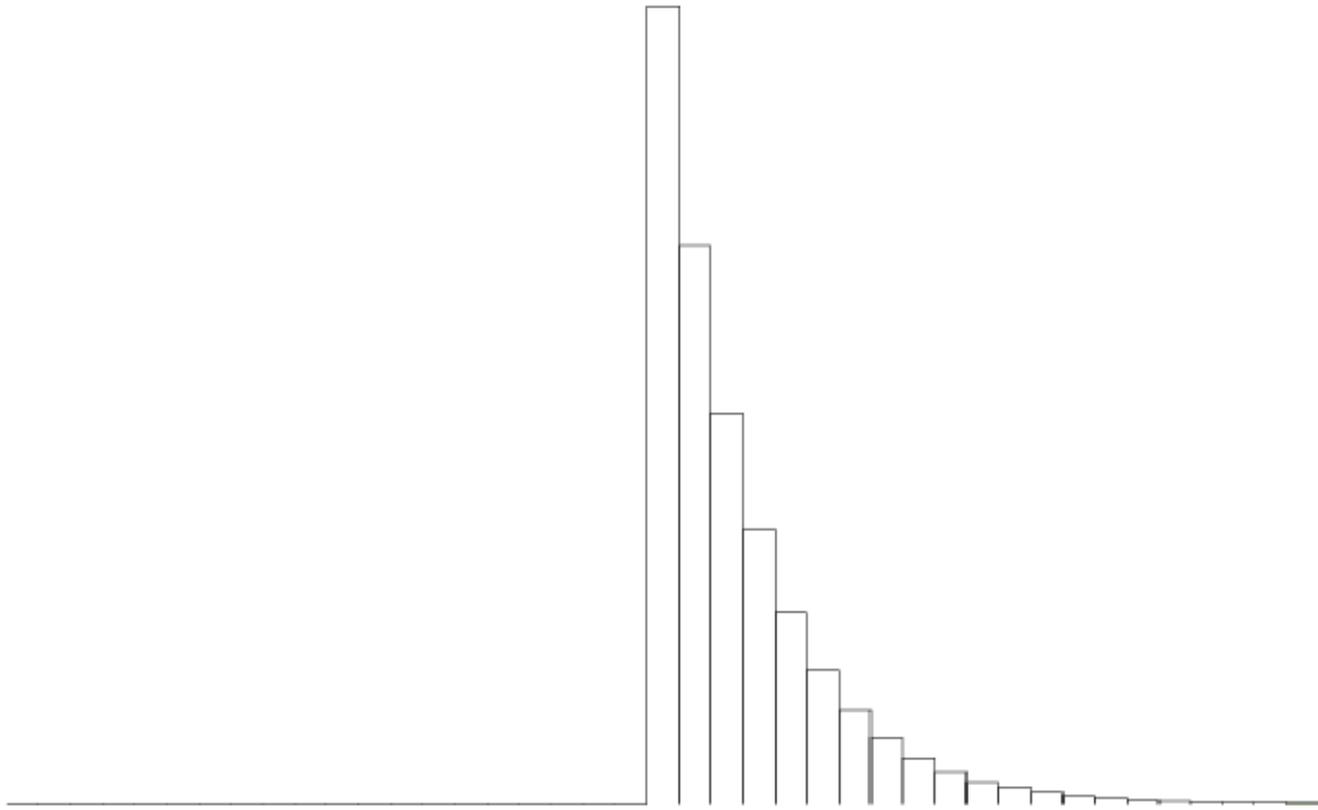
The spatial scale of the exponential depends upon the alpha coefficient. (A larger alpha means a larger scale).

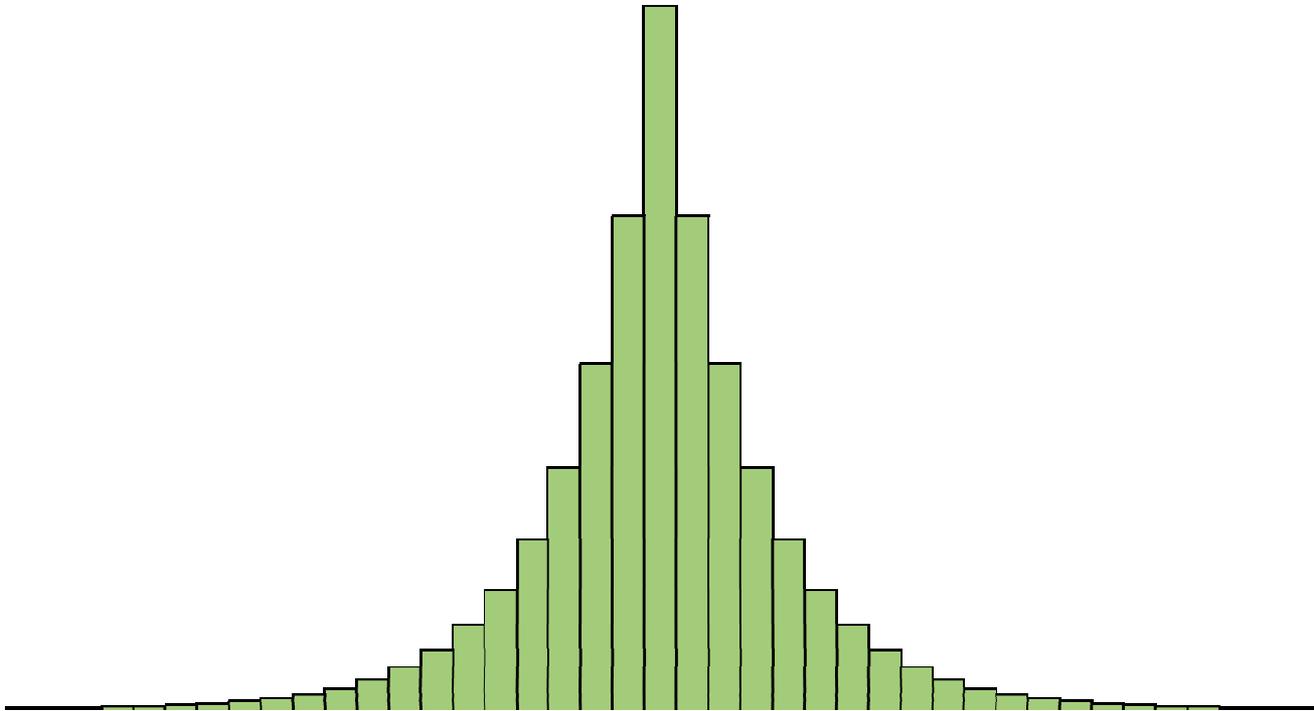
In order to get rid of the strong bias, we can use a second sequence of recursive operations, running backwards:

$$(V2)_j = (1 - \alpha)(V1)_j + \alpha(V2)_{j+1}$$



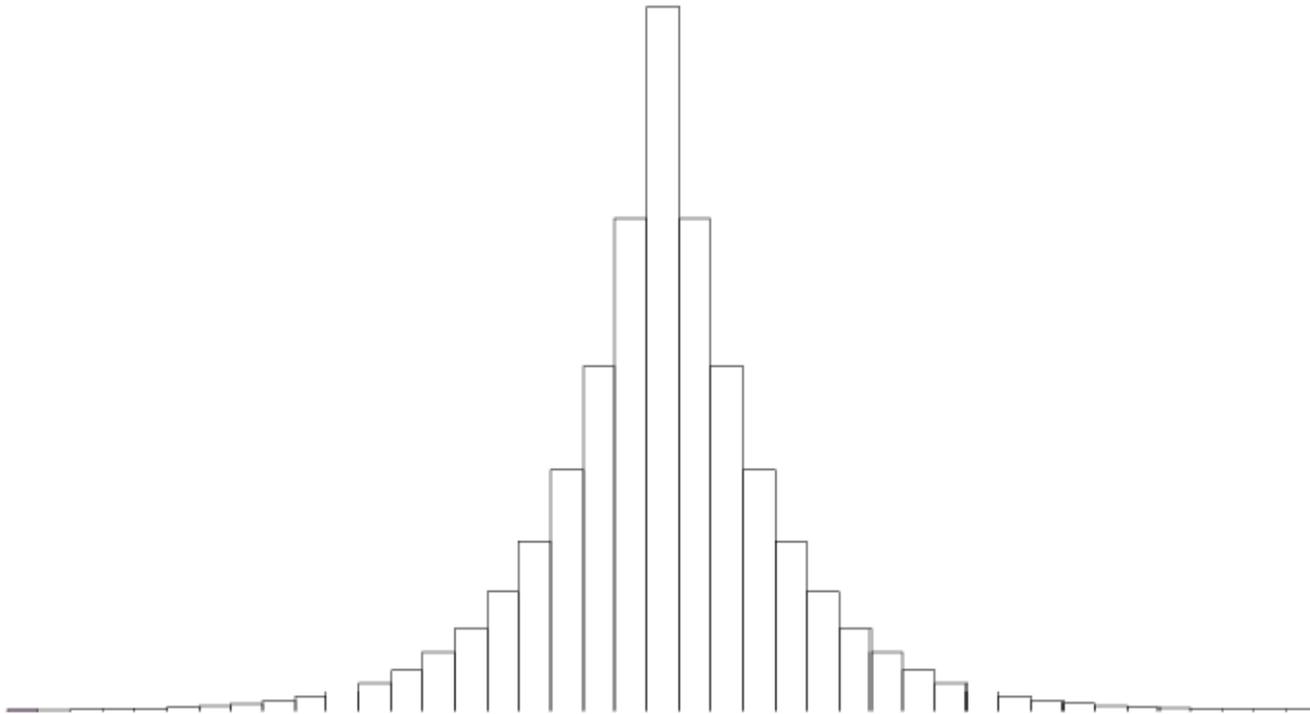
Output again, but now on the
“+” side.

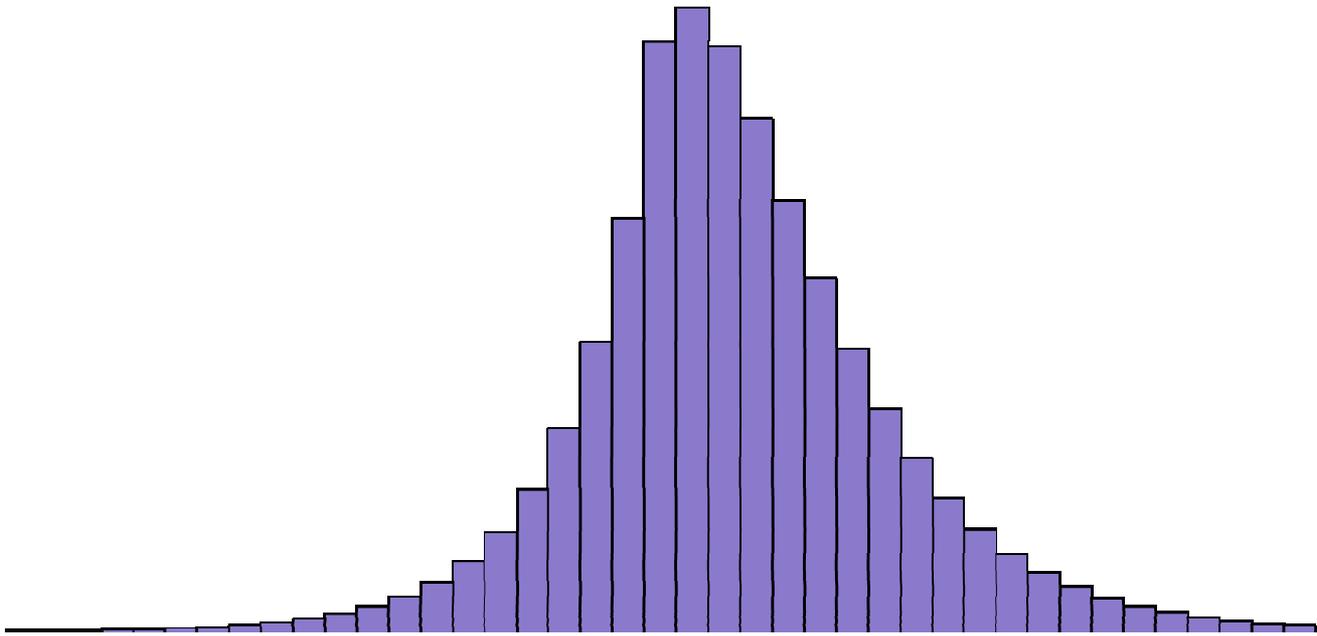


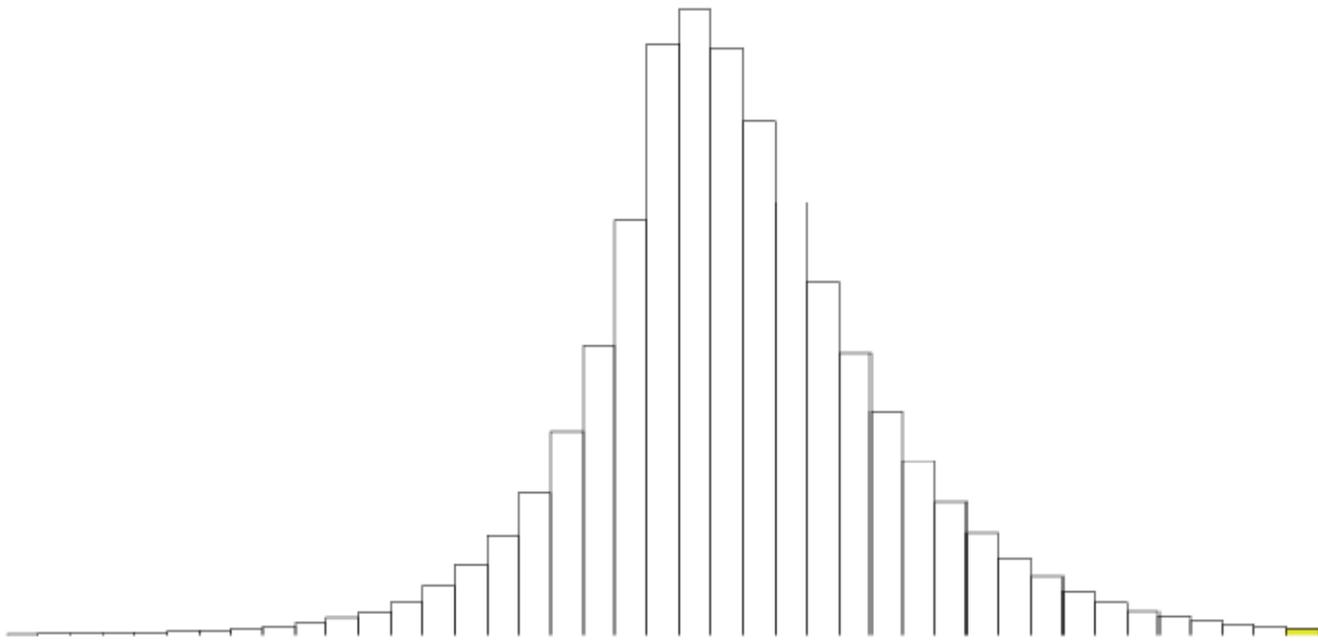


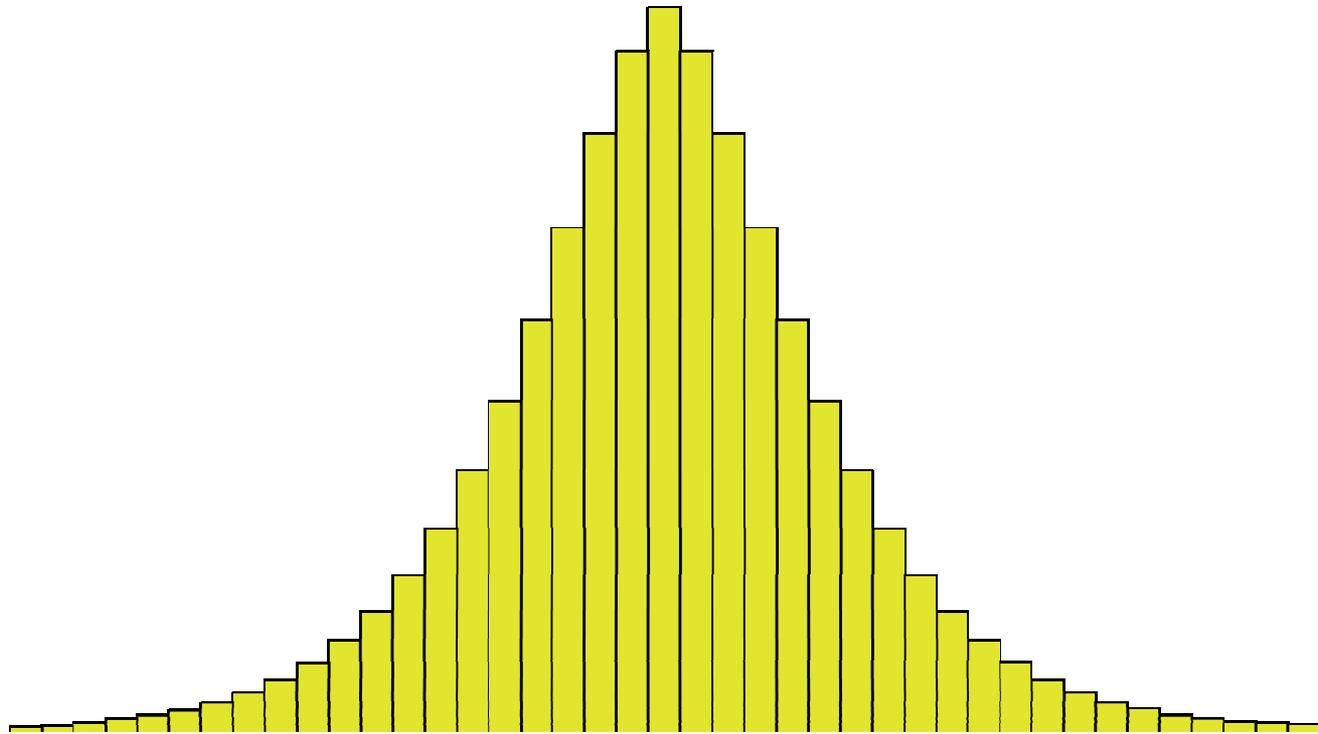
The bias has gone, but the response is too “spiky”.

The operation is quite cheap --- so let’s run the result through the forward-backward recursive filter sequence a second time.....









We have a nice smooth-looking bell-shaped profile for our “covariance”.

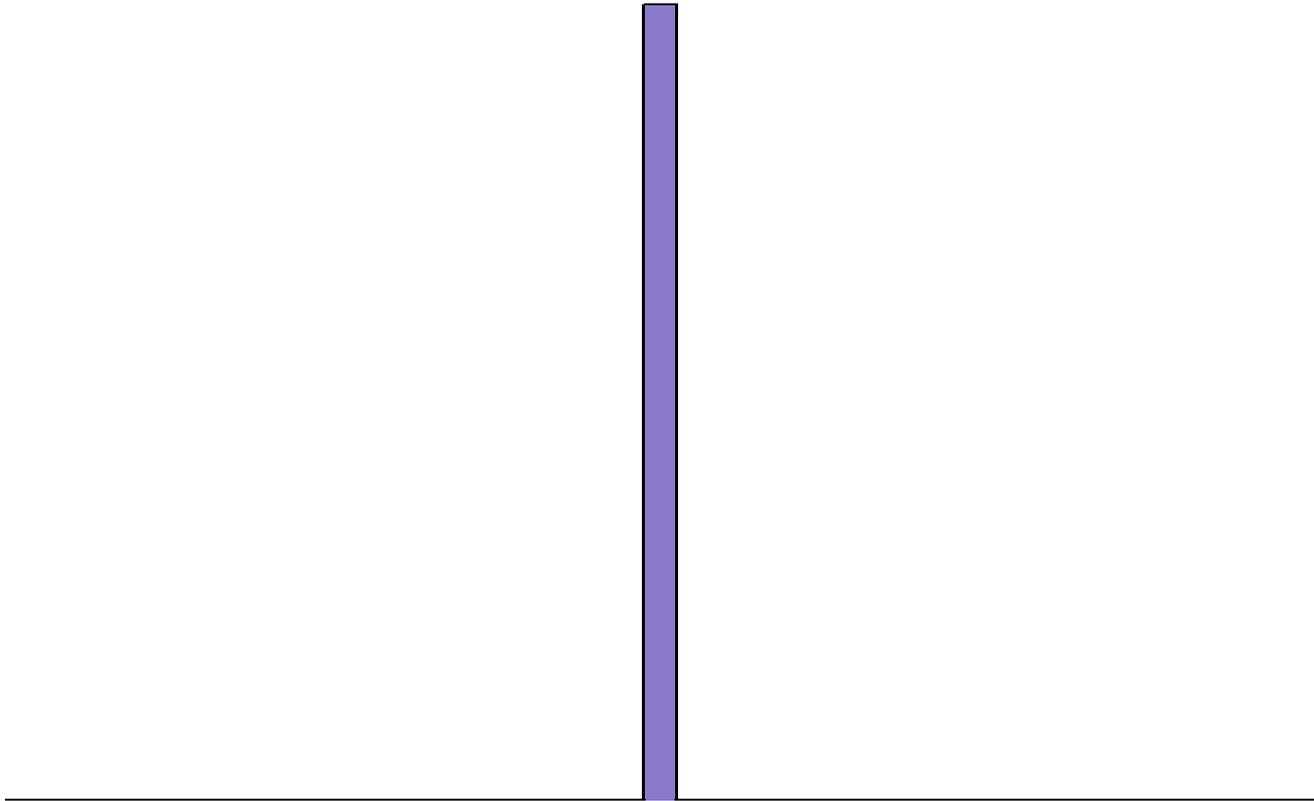
We used the sequence:

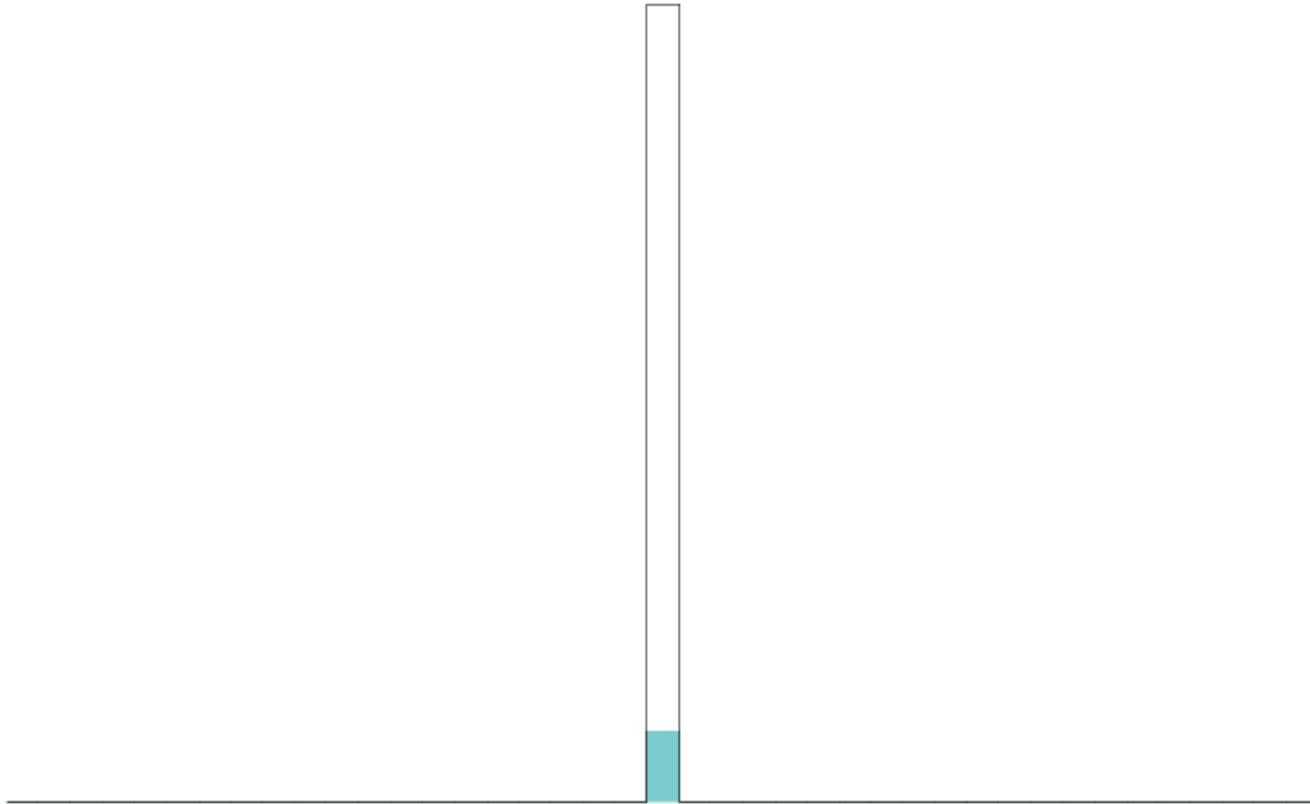
forward, backward, forward, backward.

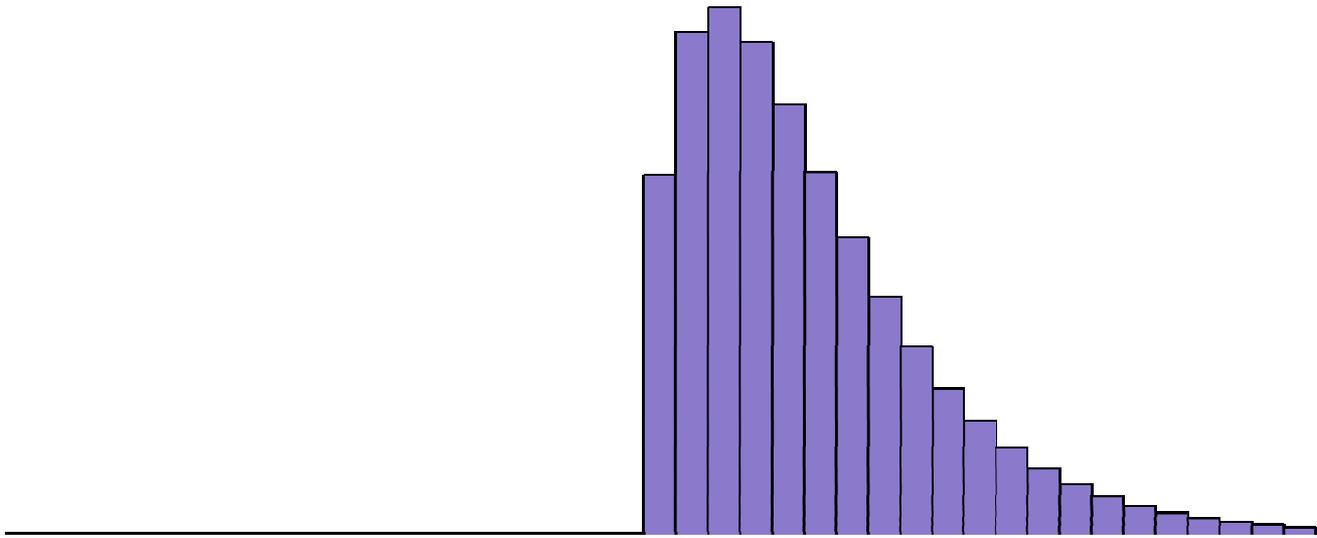
We could alternatively use the sequence:

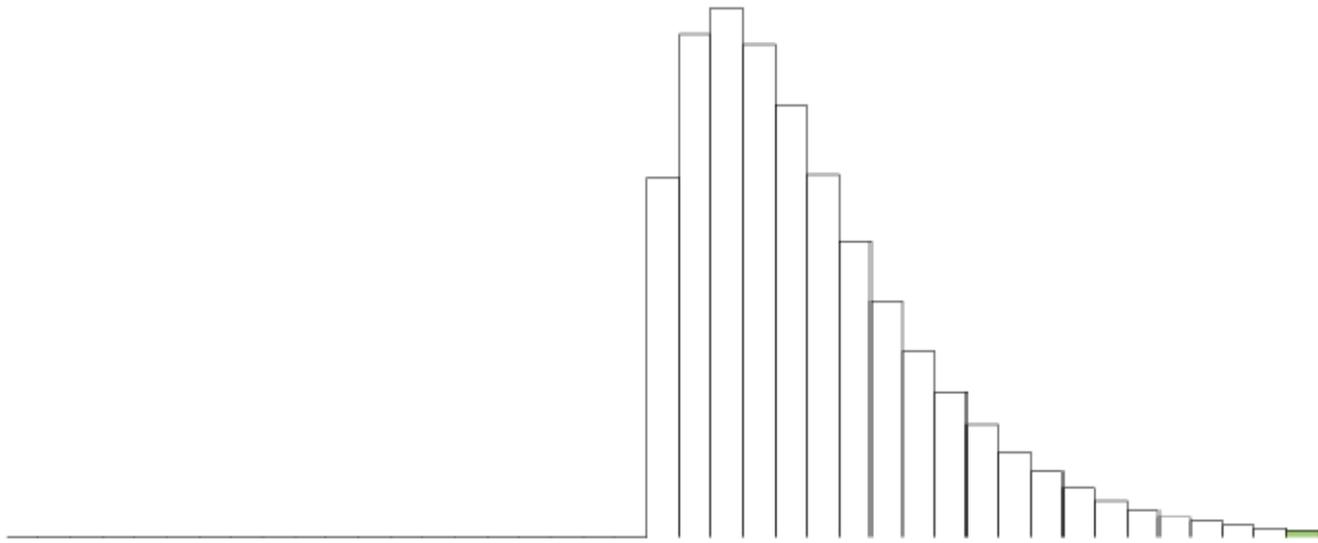
forward, forward, backward, backward.

In that case, the two forward sweeps could be combined into a single 2nd-order recursive filter running forwards, and the two backward sweeps could similarly be combined into a single 2nd-order recursive filter running backwards.





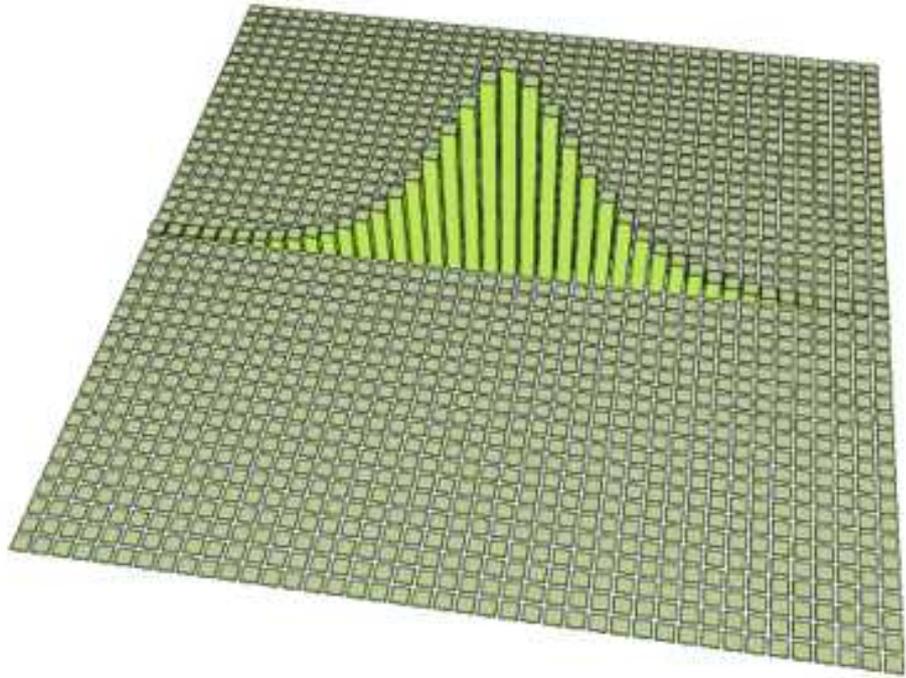


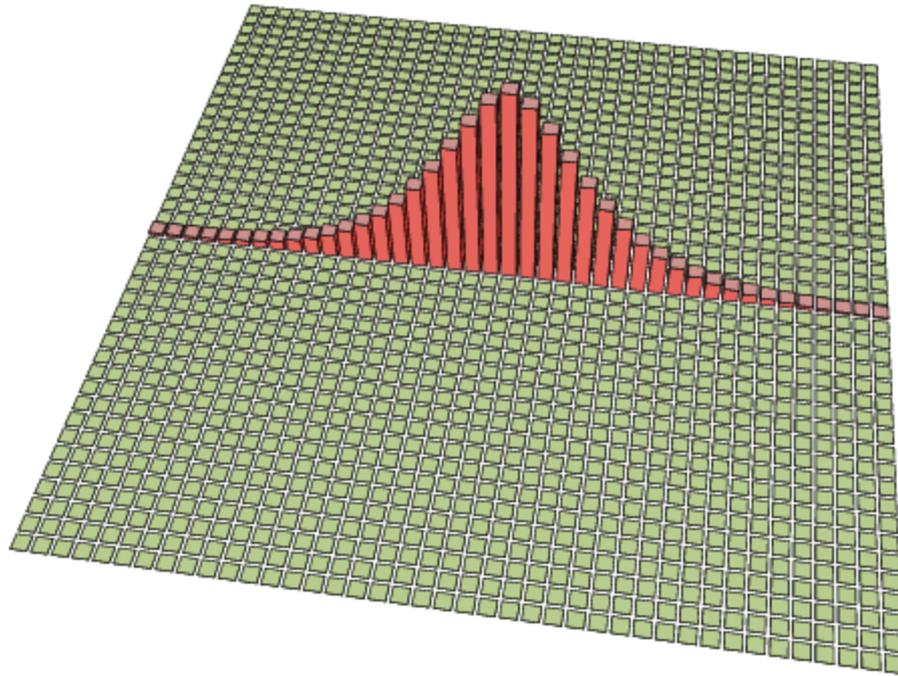


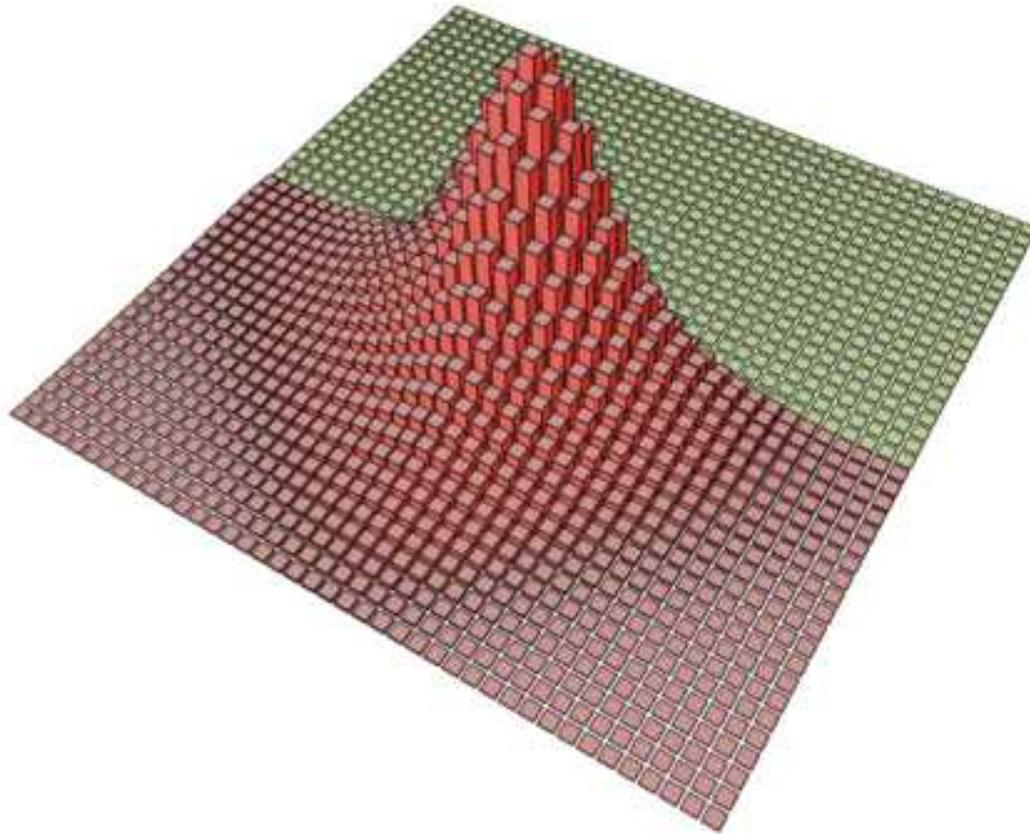
What about higher dimensions?

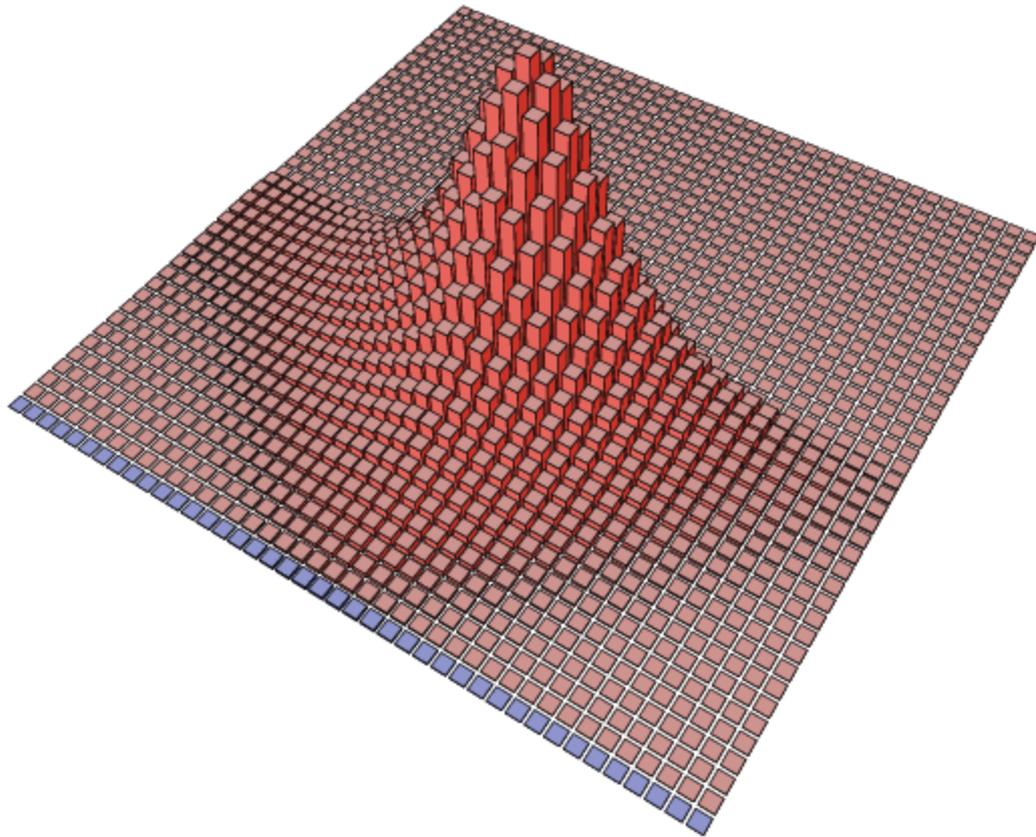
Suppose we follow a symmetrical filtering in the “x” direction (all such lines) with a similar symmetrical filtering in the “y” direction.

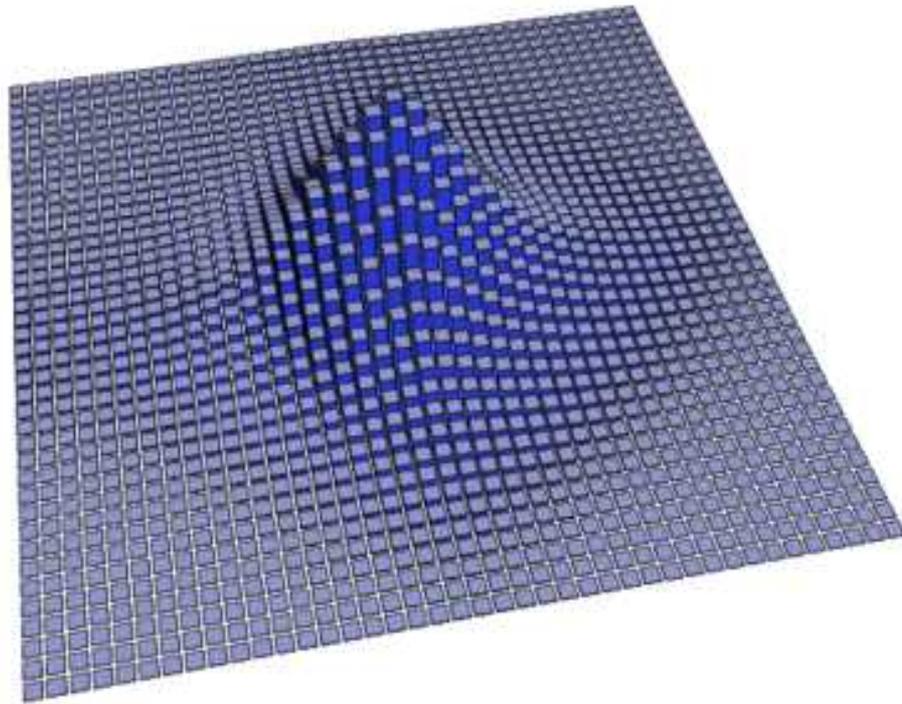
The filters associated with the parallel line directions can be done “in parallel” in every sense, since they do not mutually interfere.













The 2D distribution, after recursive filtering symmetrically in both x and y directions, certainly produces a bell-shaped response.

But is it sufficiently **rotationally** symmetric?



Not really!

In fact, even after multiple applications of the first-order recursive filters, the contours continue to show somewhat “diamond” shapes.

Instead of choosing the coefficients of the n th-degree recursive filters equivalent to n applications of the 1st-order filter, we can construct the best approximation, at this degree, to the shape $G(x)$ for which the contours of $G(x)G(y)$ are truly circular.

$$G[(x^2 + y^2)^{1/2}] = G(x) G(y)$$

Implies,

$$G(x) = \exp(a x^2)$$

For some a . Obviously, we need $a < 0$.

Hence, the “ideal” profile we seek for our filters is the

GAUSSIAN

NOTE: Our “Gaussian” is a function shape in physical space; NOT a probability density!

In the next lecture (practicalities) we shall discuss the ways in which

- (i) The degree of similarity with the Gaussian can be measured, and
- (ii) The ways in which the fit to the Gaussian can be improved.

The important point is that something closely resembling a Gaussian convolution CAN be efficiently implemented using spatial recursive filters that act along the lines of a smooth grid.

The isotropic Gaussian by itself is a very restricted form of covariance, but it is very simple and has often been adopted in both the old empirical “successive corrections” analysis schemes (Barnes) and in statistical, of “optimal” analyses (e.g., Derber and Rosati, in the ocean context).

But in the Derber and Rosati scheme, the “Gaussian” is generated by an explicit simulation of the process of diffusion.

There is no need for diffusion to be constant; can the filters also give a variable quasi-Gaussian, or quasi-diffusive, response?

Yes, they can!

Also, as Weaver and Courtier showed, the explicit diffusion method can be used to generate both inhomogeneous (spatially varying) and anisotropic (varying with angle) responses. Can the filters do this?

Yes they can!

Reinterpreting the filters

In fact, we can reinterpret the recursive filters that generate inhomogeneous and anisotropic quasi-Gaussian-shaped response profiles as highly accelerated numerical simulations of the diffusive process.

The “spread” of the response function is measured by its centered 2nd moment matrix, or “tensor”. This symmetric tensor contains information about the aspect ratio of characteristic scales in the principal orthogonal directions, -- but also encodes the information for those directions. It is convenient to give it a name – the “**aspect tensor**”. The aspect tensor equals the diffusivity when the time duration is fixed at $\frac{1}{2}$ nondimensional time units.

The different lines along which the filters act one-dimensionally can be thought of as the ways in which the numerical diffusion operator is “split”.

Given that our filters are equivalent (roughly) to generalized diffusion, we are faced with the decision about the manner in which the diffusion is envisaged to take place – there is an ambiguity, which we will now consider.



Ambiguity of diffusion equations corresponding to the same aspect tensor

The ambiguity concerns the implied choice of the “metric”.

For example, we could simulate a quasi-Gaussian component of a vertical covariance with a vertical scale of 4000m using physical height coordinates and a diffusivity of $(4000)^2$ applied for duration $t=1/2$, and get one result. Or we could switch to sigma coordinates and use a height-dependent diffusivity of about $(\text{sigma}/2)^2$, and express a simple diffusion equation in these sigma coordinates, and get a result corresponding to the same aspect tensor, but values that differ in detail.



For a diffusivity tensor, D^{ij} , the anisotropic diffusion in Cartesians may be written:

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x^i} D^{ij} \frac{\partial P}{\partial x^j}.$$

Transformed to another coordinate system, or generalized to fit a non-Euclidean domain, the diffusion equation becomes:

$$\frac{\partial P}{\partial t} = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \sqrt{g} D^{ij} \frac{\partial P}{\partial x^j},$$

But isotropic and homogeneous diffusion in such a domain, with diffusivity standardized to unity, becomes:

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



Resolving the ambiguity



The third diffusion equation above provides a natural way to resolve this ambiguity. I.e., for each quasi-Gaussian component with its given aspect tensor, we let the aspect tensor itself define the effective metric.



Remarks

By adopting the third form of the diffusion equation:

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

we are committed to filtering in a Riemannian, or non-Euclidean geometry.

What advantages are there to doing this?

Inhomogeneous diffusion and the amplitude problem

For homogeneous diffusion in a Euclidean geometry, the solution initiated by an impulse is always a GAUSSIAN.

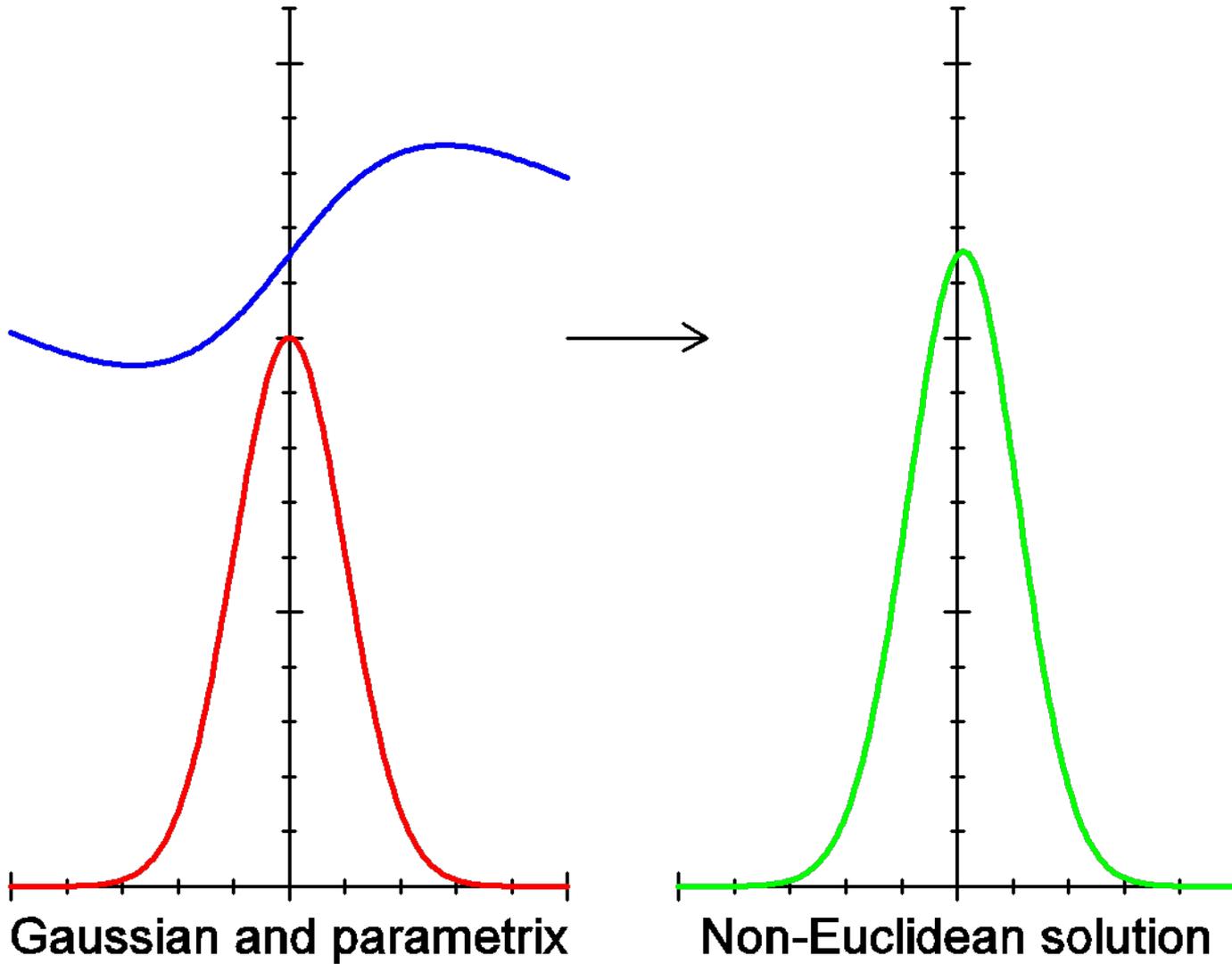
The amplitude is therefore trivially known.

But, even in Euclidean geometry, if the diffusivity is inhomogeneous, the outcome is no longer exactly Gaussian and the amplitude is uncertain.



For gently varying aspect tensor distributions, we find that
The amplitude correction factor (compared with the
standard Gaussian formula) depends only upon the
curvature diagnostic locally, and can be expressed as an
asymptotic power series expansion.

This is the so-called **PARAMETRIX EXPANSION**
method.



Parametrix Expansion Method

- Express the solution in “normal” coordinates as Gaussian*parametrix, T .
- Expand $T(\mathbf{x},t)$ as a power series in normal coordinates, \mathbf{x} , and “time”, t
- Evaluate the $t=1/2$ solution to obtain the amplitude correction factor, $T(\mathbf{0}, 1/2)$

The details of the parametrix method are highly technical, but the main idea is that the amplitude correction factor is made to depend only upon the CURVATURE and its higher derivatives. It does not depend upon the local value of the metric (=aspect tensor) itself and the curvature comes from the 2nd derivatives of this metric.

In the most general case, the “curvature” of a space is expressed as a 4th rank tensor, the Riemann-Christoffel tensor.

In 3D, this tensor can be condensed down to the 2nd rank Ricci symmetric tensor. In 2D, it further reduces to a scalar curvature, the Gaussian curvature (which is one half of the Ricci scalar).

Notation for curvature tensors

R^{ijkl} Riemann-Christoffel

$R^{ik} = R^{ijk}_{\ j}$ Ricci tensor

$R = R^k_{\ k}$ Ricci scalar

$\kappa = (1/2)R =$ Gaussian curvature (2D only)



The parametrix expansion for the amplitude quotient to second order:

In n dimensions:

$$A \approx 1 + \frac{R}{12} + \frac{1}{1440} \left(12\nabla^2 R + 5R^2 - 2R_{ij}R^{ij} + 2R_{ijkl}R^{ijkl} \right).$$

When $n = 3$:

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

When $n = 2$, and κ is the Gaussian curvature:

$$A \approx 1 + \frac{\kappa}{6} + \frac{1}{60}(\kappa^2 + \nabla^2 \kappa).$$

Further details can be found in:

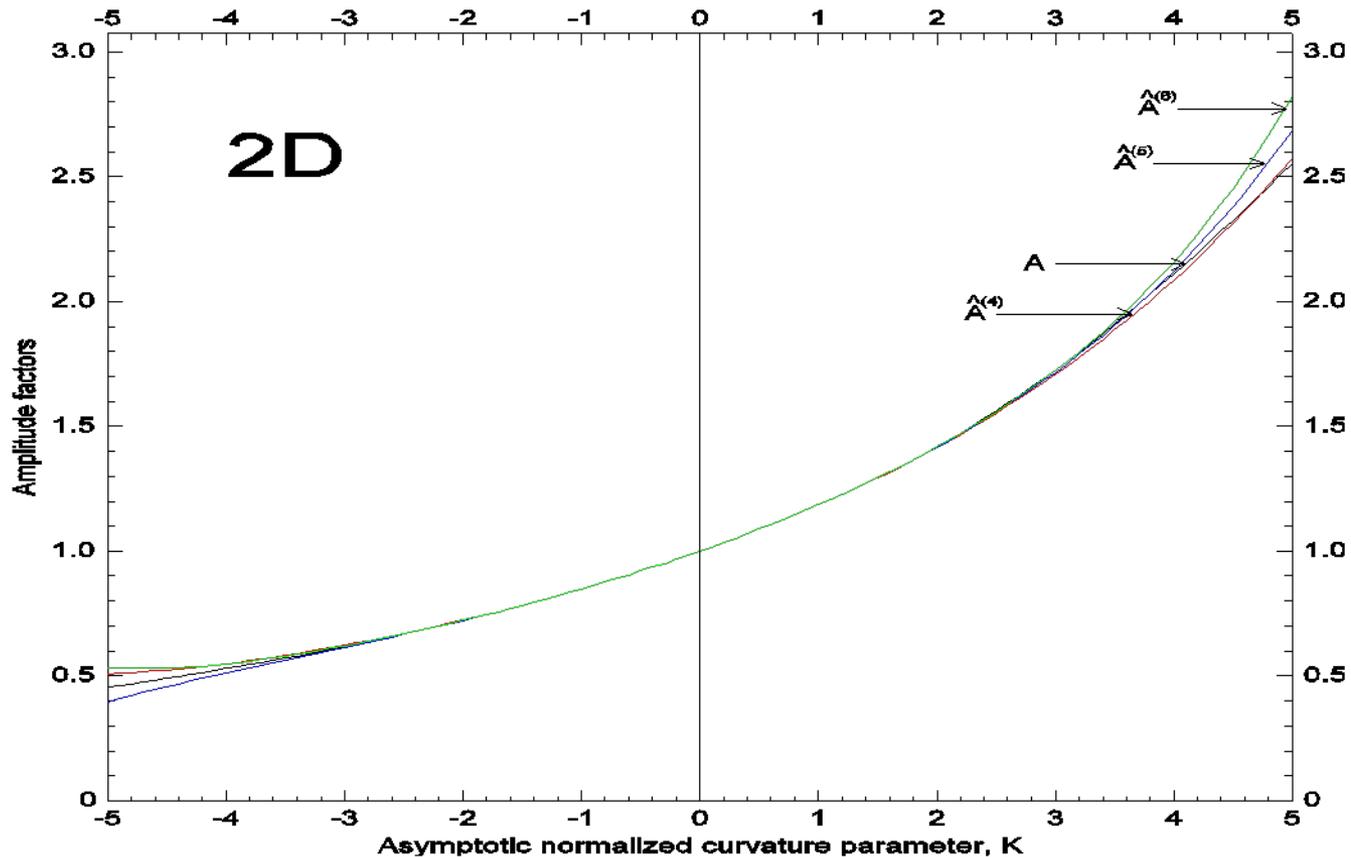
Purser, R. J., Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry; Part I: Analytic solutions for symmetrical configurations and the validation of practical Algorithms. NOAA/NCEP Office Note 456, 48 pp.; Part II: Riemannian geometry and the generic Parametrix expansion method. NOAA/NCEP Office Note 457, 55 pp.



Remarks

The solution involves only curvature, and covariant derivatives of the curvature. See, for greater detail,

Purser, R. J., 2008: Normalization of the diffusive filters that represent the inhomogeneous covariance operators of variational assimilation, using asymptotic expansions and techniques of non-Euclidean geometry; Part I: Analytic solution for symmetrical configurations and the validation of practical algorithms. NOAA/NCEP Office Note 456; Part II: Riemannian geometry and the generic parametrix expansion method. NOAA/NCEP Office Note 457.

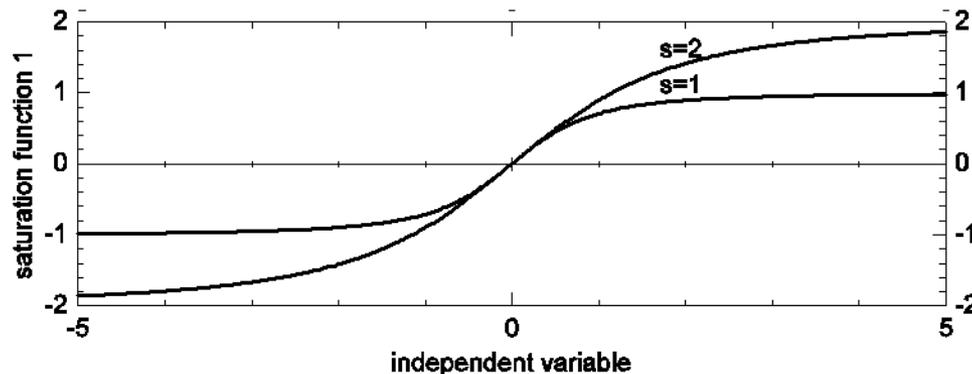


These graphs show the 2D results comparing the asymptotic expansion for the amplitude quotient with the true solution in the special case where the Gaussian curvature K is uniform. Even out to a curvature of ± 5 non-dimensional units, the asymptotic method with a few terms should give a very good approximation, as shown. However, the expansion is formally divergent. The true amplitude quotient is denoted “ A ”; other graphs show asymptotic expansions truncated to the degrees indicated by the superscript. 49

Remarks (continued)

The parametrix method is an asymptotic expansion and, as is typical of them, is of the divergent kind.

In practice, this means that the raw diagnostics “c” of curvature cannot be used. Instead, these diagnostics are “filtered” through the intermediary of a “saturation function”, $S(c)$, which limits their values at large c .



For example, having chosen the form of S , the practical estimation of the amplitude factor, A , in the 1st-order parametric expansion in 2D would be something like:

$$A = 1 + \kappa_0 S(\kappa/\kappa_0)/6$$

where κ_0 is an empirical characteristic scaling value.



The modified parametrix expansion for the amplitude correction factor will typically give an amplitude in 2D or 3D accurate to within about 5% when the variations of the aspect tensor are reasonably smooth.



Discussion

The adoption of Riemannian geometry standardizes the process by which covariance contributions are generated through quasi-diffusive processes.

Despite the apparent complication of dealing with an intrinsically curved geometry, the Riemannian formulation actually makes the estimation of amplitude easier and more accurate.

But real covariances aren't Gaussians are they?

That's true.

The diffusive structures are:

**Simple,
Cheap, and
Convenient.**

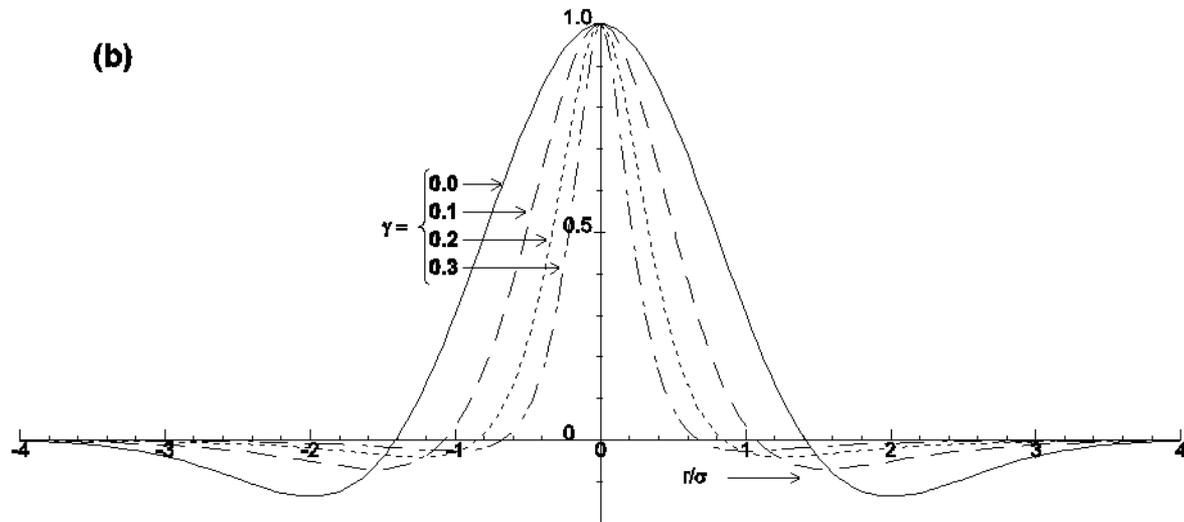
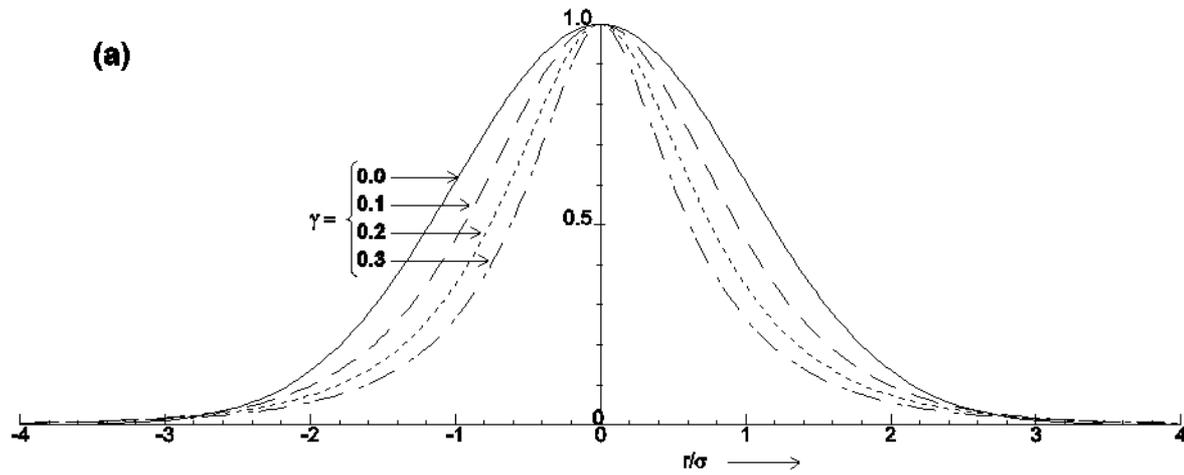
By appropriate superpositions, non-Gaussian filters can be designed.

This process is very similar to a form of (inverse) discrete Laplace transformation.

The main difference being that the weights vary.

In principle, a wide variety of profiles can be synthesized by the discrete superposition of quasi-Gaussian “building blocks”.

Because we are primarily synthesizing \mathbf{C} by this superposition, we guarantee that $\mathbf{B} = \mathbf{C}\mathbf{C}^T$ is positive semi-definite (good enough!) even when \mathbf{C} itself is not.



Some fat-tailed Gaussian mixtures, and their Laplacians

(See Purser et al., 2003: Mon. Wea. Rev., **131**, 1524—1548; 1536—1548.)

The ability to generate superpositions efficiently depends upon us being able to exploit the “Multigrid” method.

Coarse scales are primarily dealt with on a coarse grid, then later interpolated. (The adjoint interpolator is used to get the input data onto these coarse grids – the outcome of all the operations combined is then self-adjoint.)

The efficient generation of non-Gaussian filters opens up the possibility that analysis error can be characterized by filters. The best way to do this is to represent the information as the RATIO of background to analysis error.

In this form, the filters could also serve as very effective preconditioners in the cost-function reduction process.



Analysis error estimation and characterization. Preconditioning

These topics are somewhat speculative at present and it may eventually turn out that the kinds of filters we have discussed are not adequate to characterize the analysis error structure. The problem is that, while it seems reasonable to represent a background error covariance in a way that implicitly assumes that, locally, it could be characterized simply by its power spectrum, the degree of spatial inhomogeneity in a typical analysis error covariance, so strongly dominated by equally inhomogeneous observations, make it seem unreasonable that the same kind of characterization of the analysis error covariance would work.

However, even with its many imperfections, it would be at least informative to find out how well a filtering approach to representing analysis covariances might be made to work, once the necessary drastic simplifications in the measurement precision operators are made.



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

We ask: with realistic observation distributions, can we usefully impose this assumption anyway, and obtain anything of value from the resulting analysis covariance?

If so, filters might be useful for geographically smoothing a locally valid representation of each observation type's $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ operator to cause enough overlap among discretely distributed measurements of the same type to allow their measurement operators to **look** homogeneous locally and to make possible a local assessment of the density of such data.

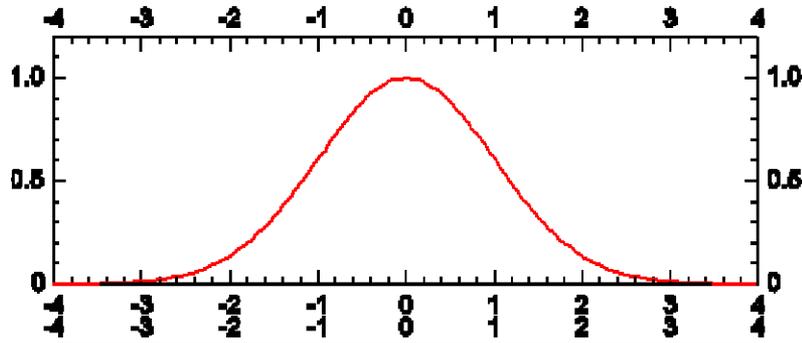


For example, a scatter of point-observations of the analysis variable would then give (locally) a white-noise distribution for the measurement precision. A horizontal scatter of satellite radiance measurements would, however, result in a very different appearance for the vertical part of the local precision spectrum, with a bell-shaped peak concentrating the precision at only the small vertical wavenumbers (the squared-absolute magnitude of the Fourier transform of the “typical” transmittance function). Each of these two sets of data would have, locally, a different spectral impact on the analysis that could be estimated, and hence represented by a combination of filters.

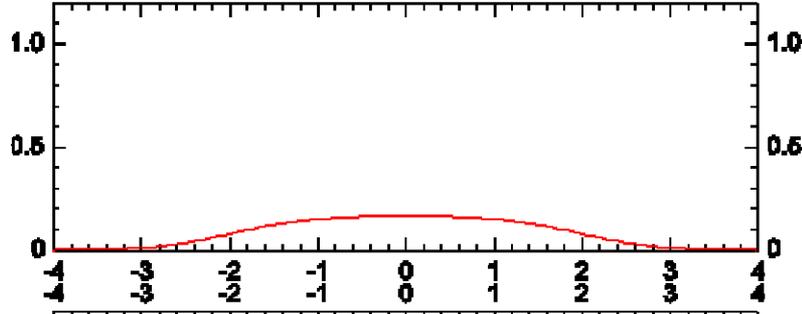
If it is possible to obtain a reasonable model of the analysis error covariance in terms of a filter, then that same filter might be a useful preconditioner for the analysis itself.

Also, it would provide a valuable tool for normalizing the spread of ensemble members in a way that reflects the influence of new data, but sidesteps the expense of adding an assimilation to each ensemble component.

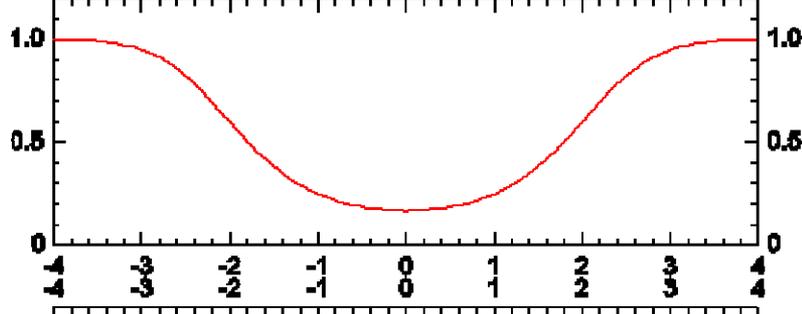
IN SITU MEASUREMENTS OF THE CONTROL VARIABLE



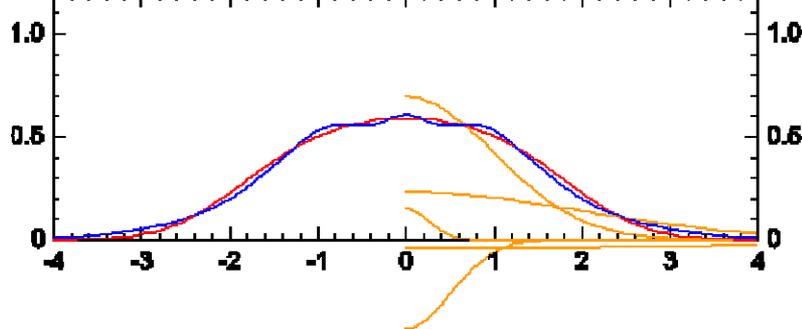
← Spectral background covariance, B



← Spectral analysis covariance, A



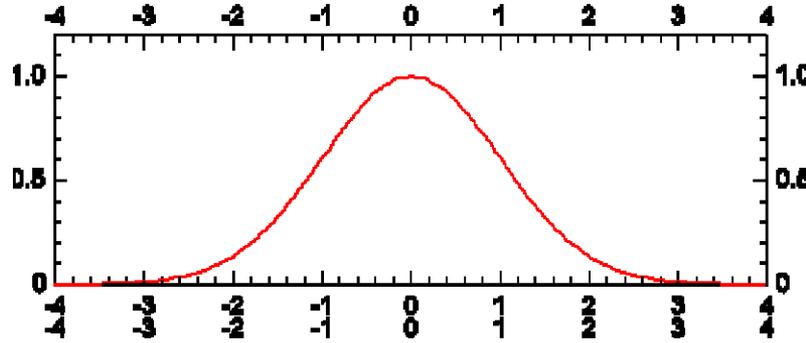
← Preconditioner, A/B



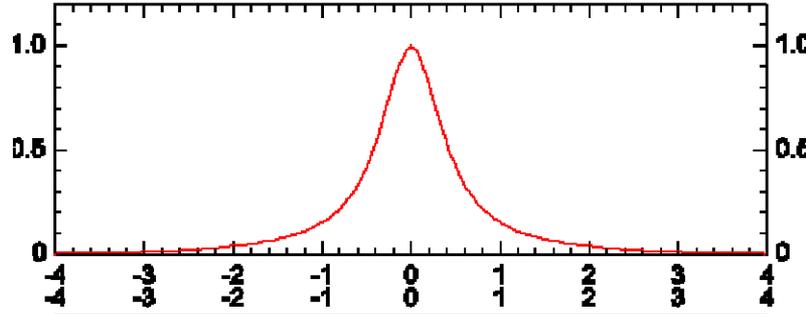
← $G = 1 - \sqrt{A/B}$ and
 $\hat{G} = \text{sum} \{ \text{Gaussians} \}$ (blue)

← Gaussians (half-curves)

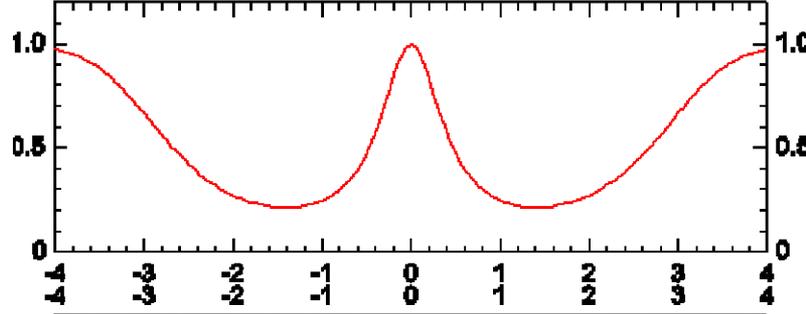
IN SITU MEASUREMENTS OF THE DERIVATIVE



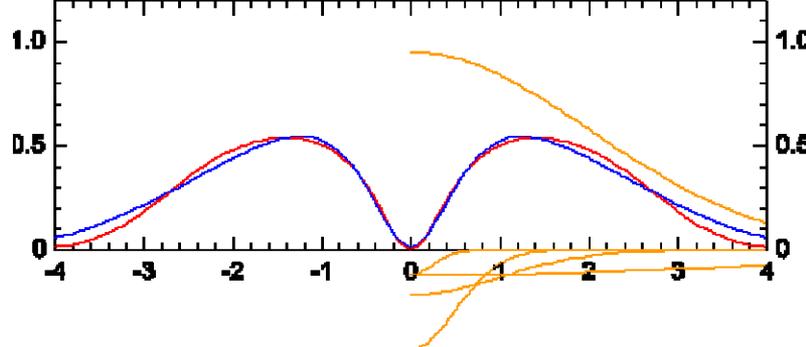
← Spectral background covariance, B



← Spectral analysis covariance, A



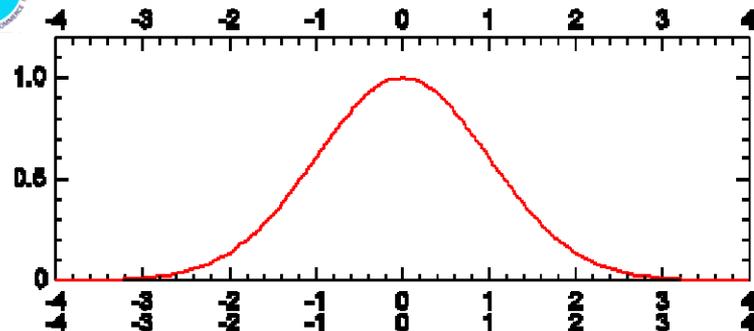
← Preconditioner, A/B



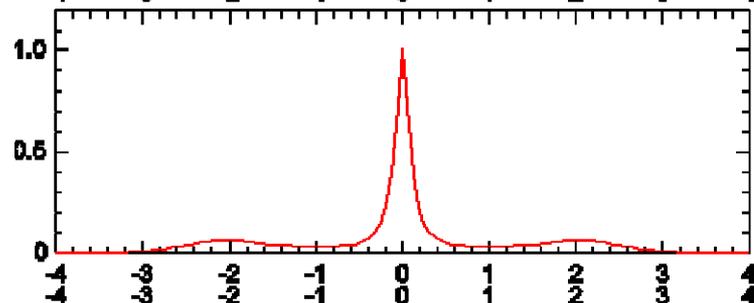
← $G = 1 - \sqrt{A/B}$ and
 $\hat{G} = \text{sum} \{ \text{Gaussians} \}$ (blue)

← Gaussians (half-curves)

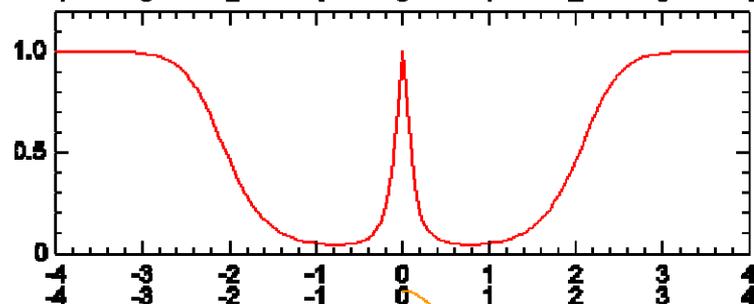
IDEALIZED RADIANCE MEASUREMENTS OF THE DERIVATIVE



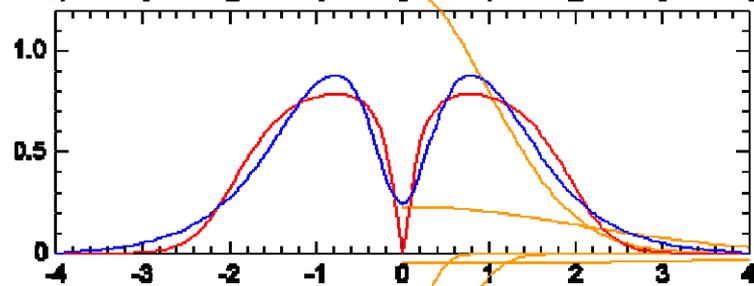
← Spectral background covariance, B



← Spectral analysis covariance, A



← Preconditioner, A/B



← $G = 1 - \sqrt{A/B}$ and
 $\hat{G} = \text{sum} \{ \text{Gaussians} \}$ (blue)

← Gaussians (half-curves)

I have described some of the “tricks” by which recursive filters can generate plausible covariance operators. (Some practical difficulties, and attempts at their solution, are described in the second talk.)

Given the ability to generate inhomogeneous and anisotropic covariance operators, how might we decide how to set the orientation and degree of anisotropy and scale --- i.e., the “aspect tensor”?

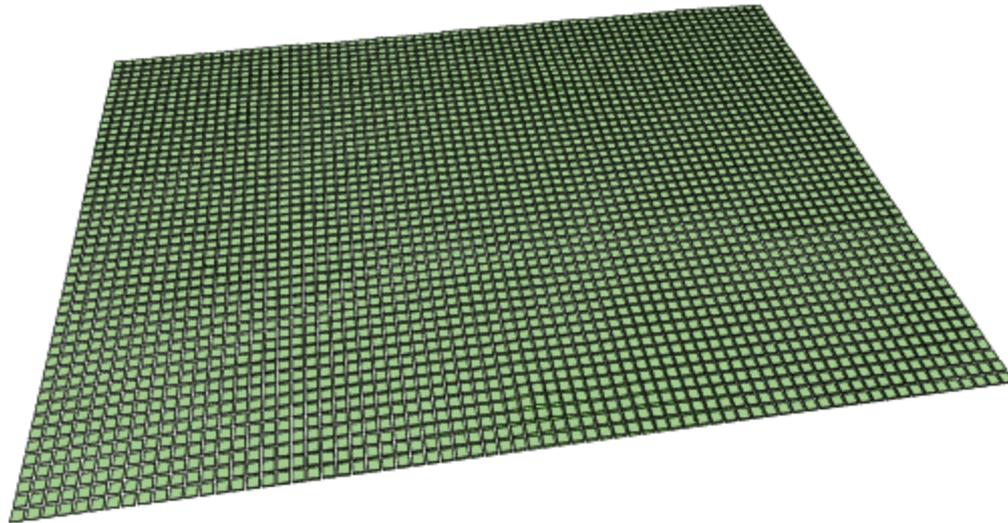
One way is to modify an idea by Riishojgaard – use variations in one (or several?) analysis variables to define additional effective dimensions of “distance”. Actually, a non-analysis variable will do.

To illustrate the idea, we take the analysis domain to be the two dimensions of surface analyzed data. (We actually do this at NCEP – it is called the RTMA.)

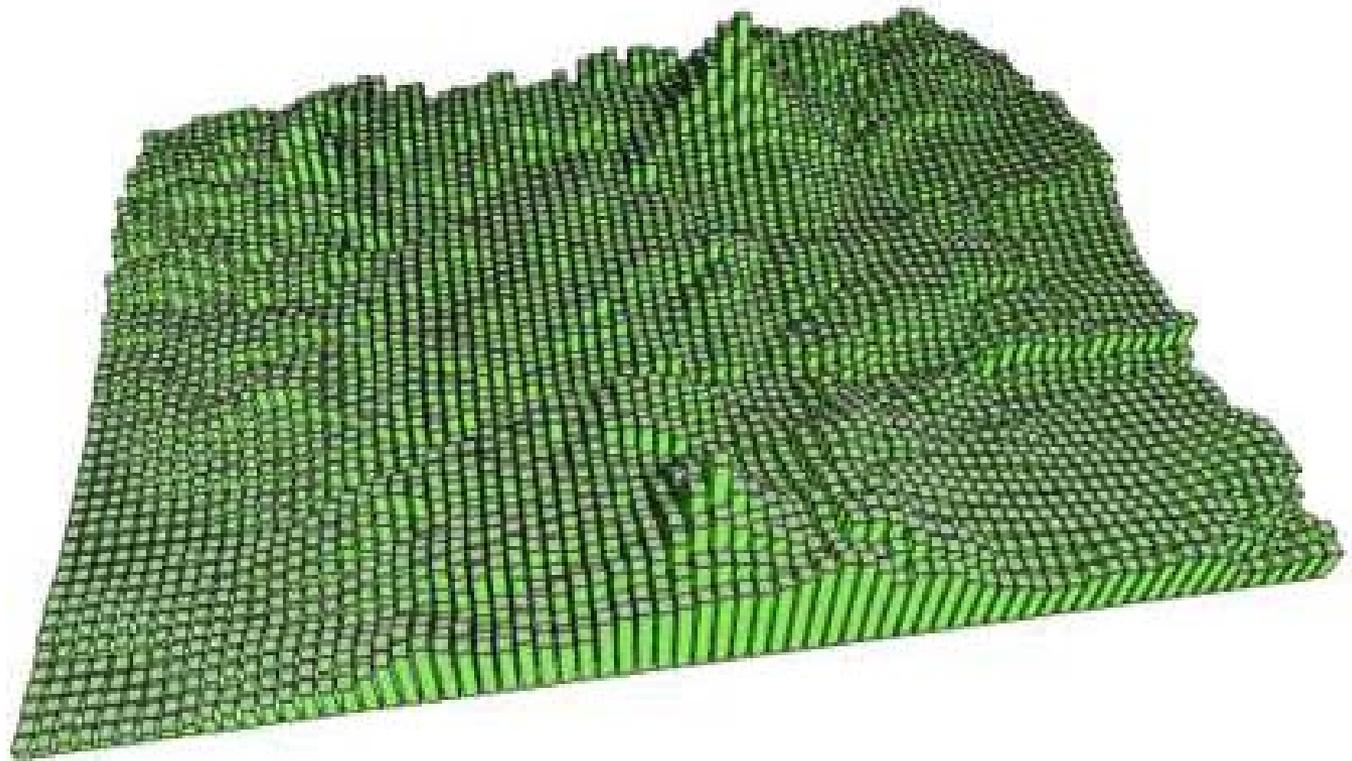
Take the “variable” that adds an extra “dimension” to be terrain height – but don’t be content with ordinary Euclidean composition of horizontal and vertical components to form effective distance – exaggerate the vertical!

Then make the covariance locally isotropic in total effective distance, and project it back into the 2D analysis domain again.

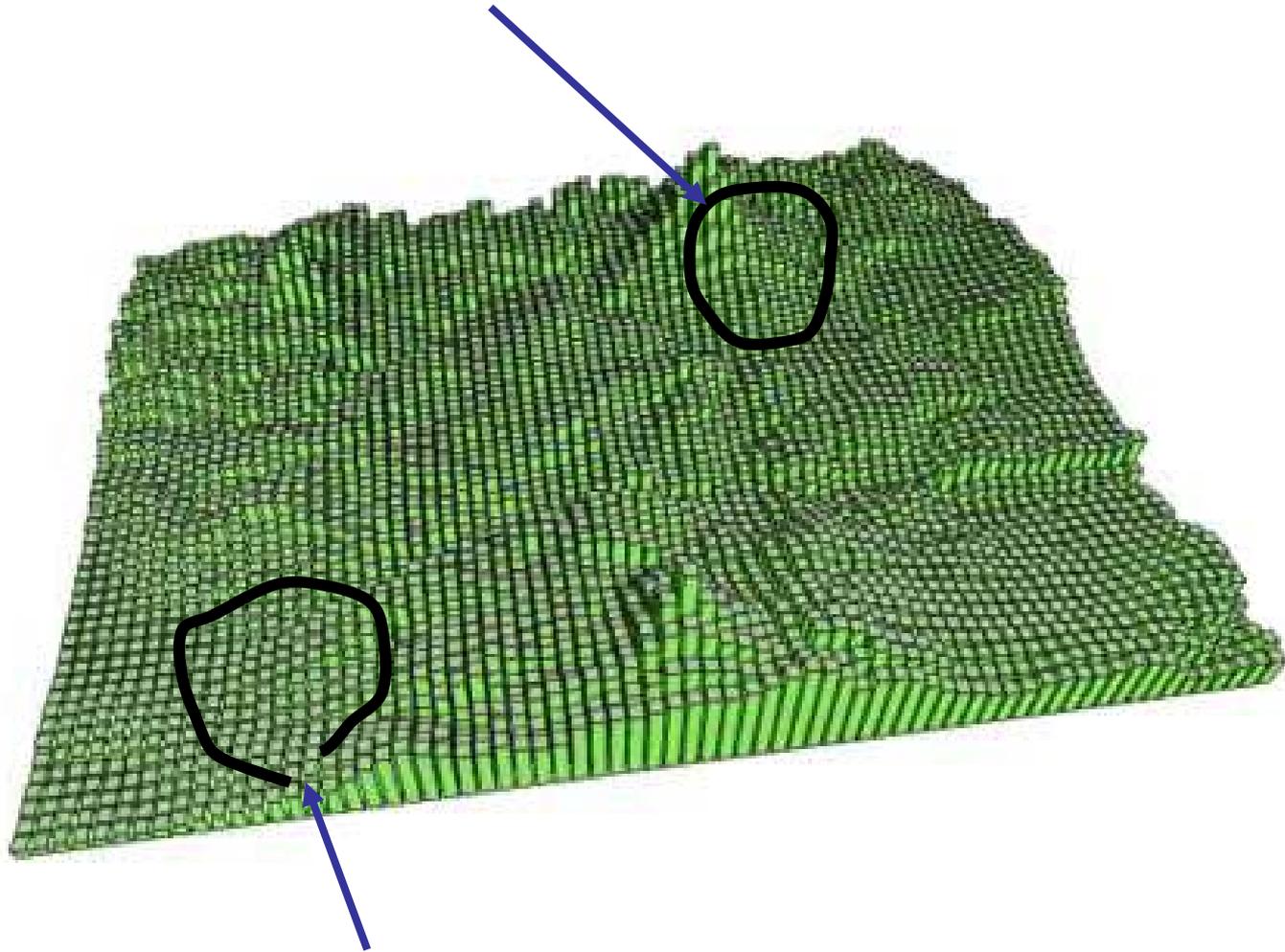
Grow the topography around the Columbia river valley:



(Data provided by Manuel de Pondeca)



Where terrain is sloping, a circular region projects back to an ellipse.

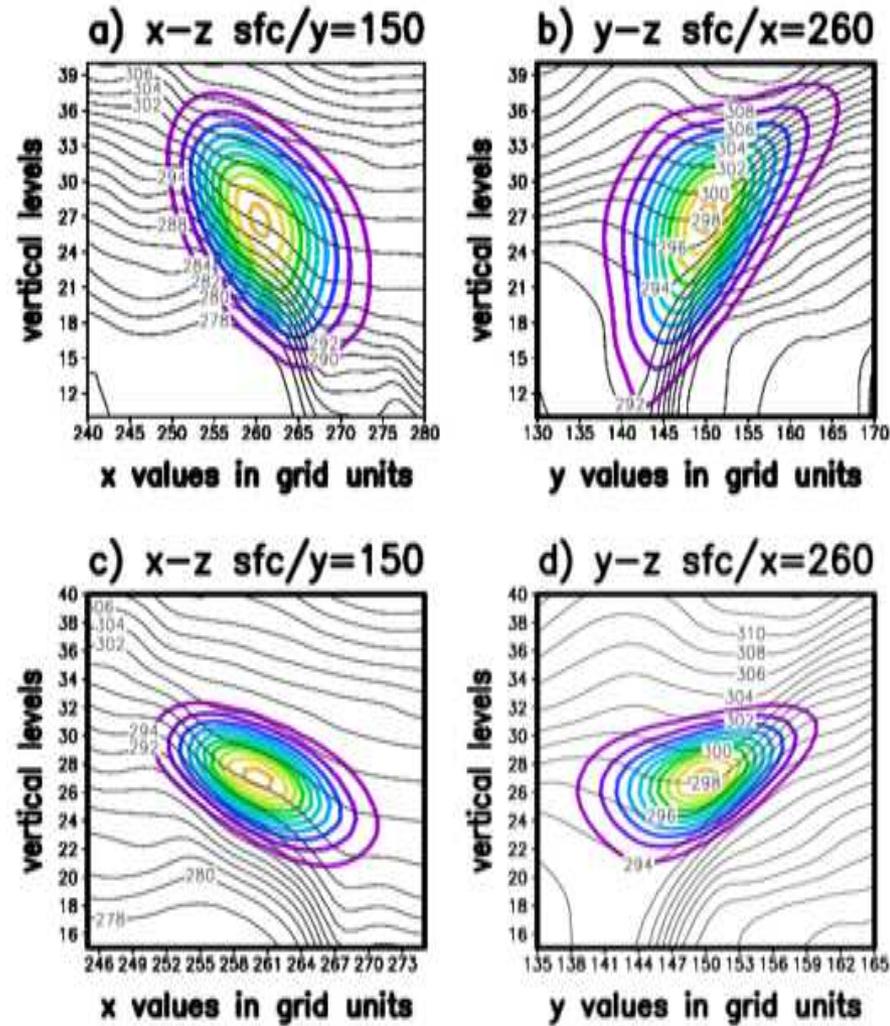


Where the terrain is flat, a circular region projects back to a circle

This kind of terrain-based adaptation is actually done in the RTMA. This inhibits analysis increments forced by an observation at one altitude from strongly influencing points that are horizontally nearby, but which have very different altitudes.

But it does not have to be “altitude” that exerts this control – it could be humidity, stream function, potential temperature --- or all of the above! And then the analysis need not be confined to two dimensions either.

Semigeos+Isent Model and Hybrid Model

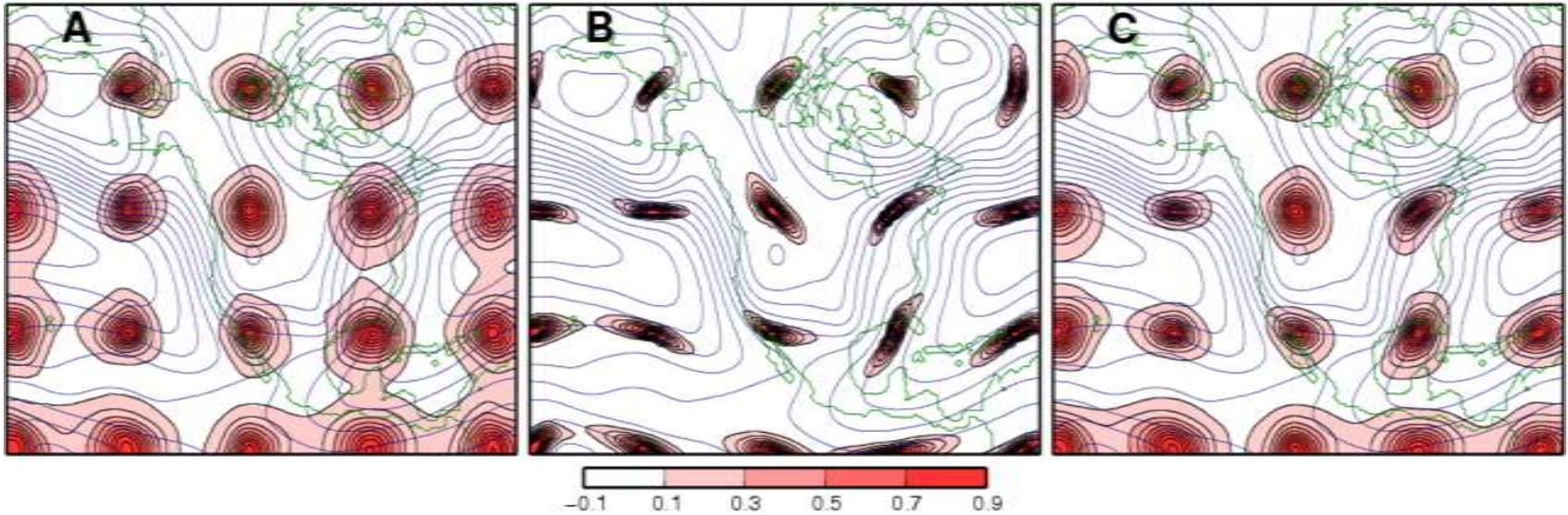


Black Contours are potential T in K
 Panels a) and b) are for the Semigeos+Isent Model
 Panels c) and d) are for the hybrid Model

Using existing variables in the background field, or even iterating to take them from the analysis, does not allow our “effective metric” (aspect tensor) to incorporate information about how well observed and how dynamically stable the atmosphere was that led to that background.

But that information should be implicitly contained in a forecast ensemble. For example, from some ensemble variable “P”:

$$g_{ij} = \text{sample mean}\{ (\text{grad } P)_i (\text{grad } P)_j / P^2 \}$$



Experiments carried out by Yoshiaki Sato et al. deriving adaptive covariances from ensembles using various methods (see NCEP Office Note 459).

Summary

Multivariate background covariances can be decomposed into quasi-independent “scalar” covariances.

Each scalar covariance, as an operator, can be simulated by recursive filters.

Such filters are numerically efficient and seem to be quite versatile – they allow both inhomogeneity and anisotropy.

By superposition, non Gaussian profiles are possible.

We still need to learn to use them adaptively to improve variational analyses and to exploit the ensemble data.



REFERENCES

- Derber, J. C., and A. Rosati, 1989: A global ocean data assimilation system. *J. Phys. Ocean.*, **19**, 1333–1347.
- Dickson, L. E., 1958: *Linear Groups (With an Exposition of the Galois Field theory)*. Dover, New York, 336 pp.
- Gilkey, P. B., 1984: *Invariance Theory, the Heat Equation, and the Atiyah-Singer Index Theorem*. Publish or Perish, Wilmington, DE, 345 pp.
- 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., 2008: Amplitude normalization of inhomogeneous diffusive filters using asymptotic expansions and techniques of non-Euclidean geometry; Part I: Analytic solutions for symmetric cases and the validation of practical algorithms; Part II: Riemannian geometry and the generic parametrix expansion method. NOAA/NCEP Office Notes (In preparation).
- 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.
- Rosenberg, S., 1997: *The Laplacian on a Riemannian Manifold; London Mathematical Society Student Texts 31*. Cambridge, 174 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.