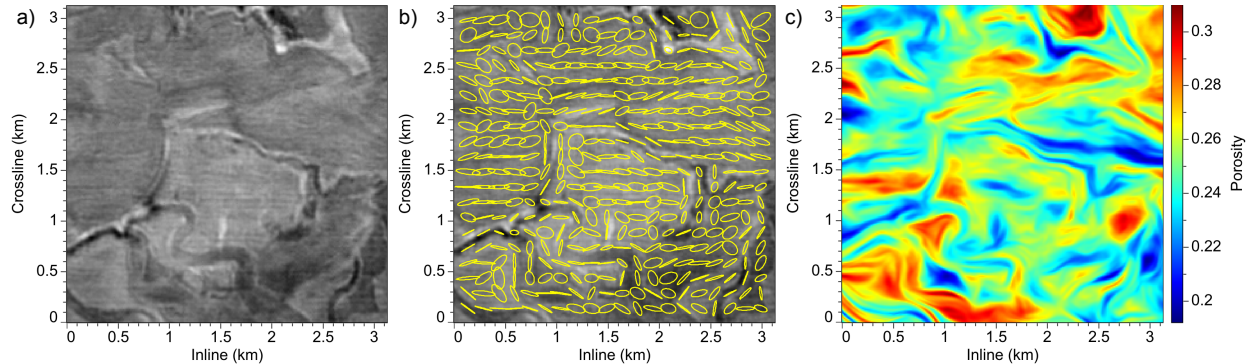# Implementing an anisotropic and spatially varying Matérn model covariance with smoothing filters

Dave Hale

Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401, USA

Figure 1. A horizon slice of a 3D seismic image (a) provides a model of spatial correlation (b) for an anisotropic and spatially varying Matérn model covariance used here in a geostatistical simulation of porosities (c). The model covariance is implemented with smoothing filters.

ABSTRACT

While known to be an important aspect of geostatistical simulations and inverse problems, an a priori model covariance can be difficult to specify and implement, especially where that model covariance is both anisotropic and spatially varying. The popular Matérn covariance function is extended to handle such complications, and is implemented as a cascade of numerical solutions to partial differential equations. In effect, each solution is equivalent to application of an anisotropic and spatially varying smoothing filter. Suitable filter coefficients can be obtained from auxiliary data, such as seismic images. An example with simulated porosities demonstrates the effective use of a Matérn model covariance implemented in this way.

Key words: inversion model covariance

1 INTRODUCTION

The solution of many inverse problems in geophysics is facilitated by a priori information about the desired solution. In least-squares inverse theory this information is provided in the form of an initial model estimate $\mathbf{m}_0$ and a covariance matrix $\mathbf{C}_M$, which can be used to compute a better (a posteriori) model estimate $\tilde{\mathbf{m}}$ as follows (Tarantola, 2005):

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_M\mathbf{G}^\top(\mathbf{G}\mathbf{C}_M\mathbf{G}^\top + \mathbf{C}_D)^{-1}(\mathbf{d} - \mathbf{G}\mathbf{m}_0). \quad (1)$$

Here, $\mathbf{d}$ denotes observed data, which are assumed to be approximately related to the true model $\mathbf{m}$ by $\mathbf{d} \approx \mathbf{Gm}$, for some linear operator $\mathbf{G}$. The data covariance matrix $\mathbf{C}_D$ quantifies uncertainties due to errors (e.g., measurement errors or ambient noise) in this approximation, while the model covariance $\mathbf{C}_M$ quantifies spatial correlation of the model $\mathbf{m}$.

The matrices in equation 1 can be viewed more generally as linear operators, and this view is especially useful for the model covariance matrix $\mathbf{C}_M$. For a model $\mathbf{m}$ with M parameters, the matrix $\mathbf{C}_M$ would contain M^2

elements, which for large M cannot be stored in computer memory. Moreover, analytical expressions for the elements of $\mathbf{C}_M$ may be unavailable, as when $\mathbf{m}$ is a sampled function of space with covariance that is both anisotropic and spatially varying. We are therefore motivated to implement multiplication by $\mathbf{C}_M$ in equation 1 as an algorithm that *applies* the linear operator $\mathbf{C}_M$ without explicitly constructing and storing a matrix.

Although we might avoid computing and storing the matrix $\mathbf{C}_M$, we must still specify parameters that describe this linear operator. This task can be especially difficult for anisotropic and spatially varying models of spatial correlation.

Figure 1 illustrates one way to parameterize the linear operator $\mathbf{C}_M$ using additional information. The spatial correlation of seismic amplitudes displayed in Figure 1a is clearly anisotropic and spatially varying. A quantitative measure of this spatial correlation is obtained from structure tensors (Weickert, 1999; Fehmers and Höcker, 2003) computed for every sample in the seismic image. In Figure 1b, a small subset of these structure tensors are represented by ellipses. Spatial correlation of seismic amplitudes is high at locations where ellipses are large and in directions in which they are elongated. In this paper I show how this tensor field can almost completely parameterize an anisotropic and spatially varying model covariance operator $\mathbf{C}_M$.

Moreover, if the model covariance operator can be factored such that $\mathbf{C}_M = \mathbf{F}\mathbf{F}^\top$, then we can easily simulate models $\mathbf{m}$ with covariance $\mathbf{C}_M$ by applying the operator $\mathbf{F}$ (or $\mathbf{F}^\top$) to an image of random numbers (Cressie, 1993). Figure 1c displays a simulated model $\mathbf{m}$ of porosities computed in this way. In this example, porosity is not directly correlated with seismic amplitude. In other words, we cannot accurately predict porosity at some location from the seismic amplitude at that location. However, the spatial correlation of porosities mimics that of seismic amplitudes, because the model covariance operator $\mathbf{C}_M$ used in this simulation was derived from the seismic image.

In this paper I describe a method for using tensor-guided smoothing filters to implement a linear operator $\mathbf{C}_M$ that approximates the Matérn covariance, which is widely used in geostatistics (Stein, 1999). My implementation of $\mathbf{C}_M$ is an approximation that extends the Matérn covariance to be both anisotropic and spatially varying. I illustrate the use of this $\mathbf{C}_M$ in a tensor-guided kriging method for gridding data sampled at scattered locations.

2   THE MATÉRN COVARIANCE

In its simplest form, the Matérn covariance function is defined by

$$c(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} r^\nu K_\nu(r), \qquad (2)$$
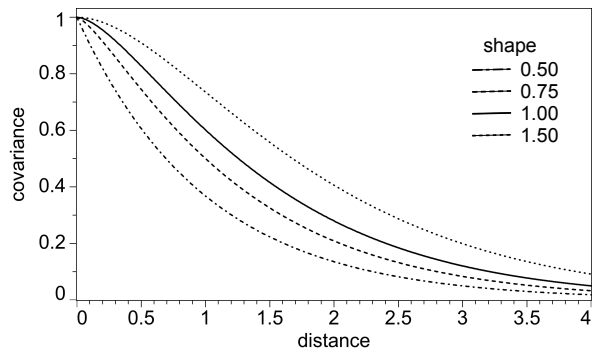
Figure 2. Matérn covariance functions $c(r)$ defined by equation 2, for four different values of the shape parameter ν.

where r is the Euclidean distance between two points in space, ν is a positive real number that controls the function's shape, and $K_\nu(r)$ denotes the modified Bessel function of the second kind with order ν. The function $c(r)$ is normalized to have unit variance $c(0) = 1$, but may easily be scaled to have any variance σ^2.

Figure 2 displays the Matérn covariance function $c(r)$ for four different choices of the shape parameter ν. For any value of ν, this function decays smoothly and monotonically with increasing distance r. For $\nu = 0.5$, the Matérn covariance is simply the exponential function $c(r) = e^{-r}$.

Despite its somewhat complex definition in terms of special functions, the Matérn covariance function is widely used in spatial statistics (Stein, 1999), partly because of the flexibility provided by the shape parameter ν. Indeed, equation 2 is sometimes described as defining the *Matérn family* of covariance functions, because any $\nu > 0$ yields a valid (positive definite) covariance function, for any number of spatial dimensions.

For d spatial dimensions, the Fourier transform of $c(r)$ is

$$C(k) = \frac{\Gamma(\frac{d}{2} + \nu)}{\Gamma(\nu)} \frac{(2\sqrt{\pi})^d}{(1 + k^2)^{\frac{d}{2} + \nu}}, \qquad (3)$$

where k is the magnitude of the wavenumber vector $\mathbf{k}$. Because the covariance function $c(r)$ is real and symmetric about the origin, its Fourier transform $C(k)$ is real and symmetric as well. Because $C(k)$ decays monotonically with increasing k, we may view the Matérn covariance function $c(r)$ as the impulse response of a smoothing filter that attenuates high spatial frequencies.

2.1   Range scaling

Figure 2 illustrates that the effective width or range of the Matérn covariance function increases as the shape parameter ν increases. In practice, we wish to specify both the shape and the range of this function, indepen-
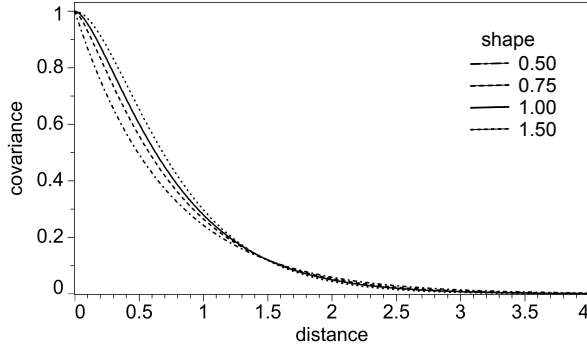
Figure 3. Matérn covariance functions $c(r)$ after range scaling, so that all functions shown here have an effective range $a = 1$ that is independent of the shape parameter ν.

dently. To do this, we make the following substitution suggested by Handcock and Wallis (1994):

$$c(r) \to c\left(\frac{2\sqrt{\nu}}{a}r\right), \qquad (4)$$

with a corresponding change to the Fourier transform

$$C(k) \to \left(\frac{a}{2\sqrt{\nu}}\right)^d C\left(\frac{a}{2\sqrt{\nu}}k\right). \qquad (5)$$

In both of these expressions the parameter a is the *effective range* of the covariance function $c(r)$; the additional factor $2\sqrt{\nu}$ compensates for the increase in range with increasing ν apparent in Figure 2.

Figure 3 displays unit range $(a = 1)$ Matérn covariance functions with this *range scaling* substitution. The effect on covariance shape caused by varying the parameter ν is more apparent in Figure 3 than in Figure 2. Note that the effect of varying ν is greatest for distances $r < 1$ or, more generally, $r < a$.

The scaling in equations 4 and 5 is isotropic; correlation varies only with distance, not with direction. Anisotropic covariance functions can be obtained by using a more general definition of distance r between two points $\mathbf{x}$ and $\mathbf{y}$, with the following substitution:

$$c(r) \to c\left(\sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{D}^{-1}(\mathbf{x} - \mathbf{y})}\right), \qquad (6)$$

with Fourier transform

$$C(k) \to |\mathbf{D}|^{\frac{1}{2}} C\left(\sqrt{\mathbf{k}^\top \mathbf{D} \mathbf{k}}\right). \qquad (7)$$

Here $\mathbf{D}$ denotes a metric tensor and $|\mathbf{D}|$ its determinant. In d spatial dimensions $\mathbf{D}$ is a symmetric positive definite $d \times d$ matrix. With these substitutions, correlation is highest in the direction of the eigenvector of $\mathbf{D}$ corresponding to its largest eigenvalue.

For simplicity in equations 6 and 7 and below, I let the range scaling factor $2\sqrt{\nu}/a$ in equations 4 and 5 be included in the tensor $\mathbf{D}$. In practice, where $\mathbf{D}$ varies spatially, such that $\mathbf{D} = \mathbf{D}(\mathbf{x})$, it is most convenient to keep these factors separate, so that we can adjust

the shape ν or effective range a without modifying the tensor field $\mathbf{D}(\mathbf{x})$.

Where the tensor $\mathbf{D}$ is spatially invariant, we can apply the Matérn model covariance operator $\mathbf{C}_{\mathrm{M}}$ by convolution with the function $c(r)$. Let $p(\mathbf{x})$ denote the input to the function that applies the operator $\mathbf{C}_{\mathrm{M}}$ to obtain an output $q(\mathbf{x})$. Then

$$q(\mathbf{x}) = \int p(\mathbf{y}) \, c\left(\sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{D}^{-1}(\mathbf{x} - \mathbf{y})}\right) \mathrm{d}\mathbf{y}. \qquad (8)$$

Equivalently, and perhaps more efficiently, we can

(i) Fourier transform $p(\mathbf{x})$ to obtain $P(\mathbf{k})$,
(ii) compute $Q(\mathbf{k}) = C\left(\sqrt{\mathbf{k}^\top \mathbf{D} \mathbf{k}}\right) P(\mathbf{k})$, and
(iii) inverse Fourier transform $Q(\mathbf{k})$ to obtain $q(\mathbf{x})$.

Note that, for either convolution with $c(r)$ or multiplication by $C(k)$, we need not construct and store a matrix representing $\mathbf{C}_{\mathrm{M}}$.

The problem addressed in this paper is that neither convolution in the space domain nor multiplication in the wavenumber domain is valid when the tensors $\mathbf{D}$ vary spatially. In this case, the output $q(\mathbf{x})$ should be computed as the solution to a partial differential equation with spatially varying coefficients.

2.2 Partial differential equations

To simplify the discussion below, let us consider only the 2D case for which $d = 2$, although the methods proposed in this paper can be extended to any number of spatial dimensions. In 2D, the (unscaled) Fourier transform of the Matérn covariance is simply

$$C(k) = \frac{4\pi\nu}{(1 + k^2)^{1+\nu}}. \qquad (9)$$

This simple form for $C(k)$ follows from equation 3 and the identity $\Gamma(1 + \nu) = \nu\Gamma(\nu)$.

Multiplication by the Fourier transform $C(k)$ of the 2D Matérn covariance function has been shown (Whittle, 1954; Guttorp and Gneiting, 2006) to be equivalent to solving the following partial differential equation (PDE):

$$(1 - \boldsymbol{\nabla} \bullet \boldsymbol{\nabla})^{1+\nu} \, q(\mathbf{x}) = 4\pi\nu \, p(\mathbf{x}). \qquad (10)$$

This equivalence results from the fact that multiplication by k^2 in the wavenumber domain is equivalent to applying the differential operator $-\boldsymbol{\nabla} \bullet \boldsymbol{\nabla}$ in the space domain.

Our reason for considering solution of partial differential equations like equation 10 is the need to apply the Matérn model covariance operator in contexts where the direction and extent of correlation are described by a spatially varying tensor field $\mathbf{D}(\mathbf{x})$. In such contexts equation 10 should be rewritten as

$$|\mathbf{D}|^{-\frac{1}{4}}(\mathbf{x}) \, (1 - \boldsymbol{\nabla} \bullet \mathbf{D}(\mathbf{x}) \bullet \boldsymbol{\nabla})^{1+\nu} \, |\mathbf{D}|^{-\frac{1}{4}}(\mathbf{x}) q(\mathbf{x})$$
$$= 4\pi\nu \, p(\mathbf{x}). \qquad (11)$$

This equation is analogous to equation 10, with the anisotropic range scaling of equation 7. Note that we must move the factor $|\mathbf{D}|^{\frac{1}{2}}$ in equation 7 to the left-hand side of equation 11 and split it into two parts, to ensure that the product of symmetric positive definite (SPD) operators on the left-hand side remains symmetric. That product is proportional to the inverse of the desired Matérn model covariance operator $\mathbf{C}_{\mathrm{M}}$, and so must be SPD.

For any tensor field $\mathbf{D}(\mathbf{x})$, solution of equation 11 is straightforward when the shape parameter ν is an integer. This is one reason that Whittle (1954) considered the integer shape $\nu = 1$ to be most natural for 2D problems. In a similar but more recent context, Fuglstad (2011) and Lindgren et al. (2011) have likewise assumed integer ν and thereby avoided the complexities of fractional PDEs.

In practical applications with spatially invariant 2D Matérn model covariances, commonly used values for the shape parameter ν lie in the interval $[0.5, 1.5]$, which includes only the one integer value $\nu = 1$. In practice, permitting only integer ν may reduce the Matérn family of covariance functions to just one function.

3   A SMOOTHING COVARIANCE

To facilitate more general (non-integer ν) shapes of covariance functions in the Matérn family, let us consider approximations to the fractional partial differential equation 11. For simplicity in developing these approximations, I temporarily omit the tensor field $\mathbf{D}(\mathbf{x})$ and use the Fourier transform $C(k)$ in equation 9 as a convenient shorthand for equations 10 and 11.

The approximation to $C(k)$ proposed here is of the form

$$\tilde{C}(k) = \frac{\gamma}{(1 + \alpha k^2)^l \, (1 + \beta k^2)}, \qquad (12)$$

where α, β, γ, and l are constants computed from the shape ν of the desired Matérn covariance function. Computation of the constant integer l is easy: $l = \lfloor 1 + \nu \rfloor$. For example, if $\nu = \frac{1}{2}$, then $l = 1$; if $\nu = 1$, then $l = 2$.

If ν is an integer, then $l = 1 + \nu$, $\alpha = 1$, $\beta = 0$, and $\gamma = 4\pi\nu$ yields an exact match to the Matérn covariance function. In this case $\tilde{C}(k)$ in equation 12 exactly equals $C(k)$ in equation 9, and no approximation is required.

Otherwise, after computing l, I compute the three non-negative constants α, β, and γ so that the approximate covariance function $\tilde{c}(r)$ corresponding to $\tilde{C}(k)$ matches exactly three values of the Matérn covariance function $c(r)$ given by equation 2. To obtain an approximation that is accurate for both large and small distances r, I choose to match the values 0.1, 0.9 and 1.0.

I first express the scale factor γ in terms of α and β so that $\tilde{c}(0) = 1$, which is one of the three values to be matched. Recall that $\tilde{c}(0)$ is just the 2D inverse Fourier transform of $\tilde{C}(k)$ evaluated at $r = 0$ which, in

Table 1. The scale factor γ in $\tilde{C}(k)$, as a function of α and β, chosen so that $\tilde{c}(0) = 1$. See equation 12.

l	γ
1	$\dfrac{4\pi(\alpha-\beta)}{\log(\alpha/\beta)}$
2	$\dfrac{4\pi(\alpha-\beta)^2}{\alpha-\beta-\beta\log(\alpha/\beta)}$
3	$\dfrac{8\pi(\alpha-\beta)^3}{(\alpha-3\beta)(\alpha-\beta)+2\beta^2\log(\alpha/\beta)}$

turn, equals the 2D integral of $\tilde{C}(k)$ divided by $4\pi^2$. By analytically performing this integration I obtained the expressions for γ listed in Table 1.

I then use bisection to find distances r_1 and r_9 such that $c(r_1) = 0.1$ and $c(r_9) = 0.9$. Finally, I use the iterative Newton-Raphson method to compute the constants α and β such that $\tilde{c}(r_1) = 0.1$ and $\tilde{c}(r_9) = 0.9$. The Newton-Raphson iterations require values and derivatives (with respect to α and β) of the approximate covariance function $\tilde{c}(r)$.

Using symbolic mathematical software, it is possible to obtain $\tilde{c}(r)$ as a function of α and β by analytical 2D inverse Fourier transform of $\tilde{C}(k)$ in equation 12. The value of $\tilde{c}(r)$ is given by one of the expressions in Table 2. Though straightforward to compute, these expressions are somewhat complicated, so I use finite differences to approximate derivatives of $\tilde{c}(r)$ with respect to α and β, as required by the Newton-Raphson method. I begin the Newton-Raphson iterations with initial values $\alpha = 1$ and $\beta = 0$.

The results of the fitting process are listed in Table 3. In practical applications we might use these tabulated values directly to obtain adequate approximations to $\tilde{c}(r)$ or $\tilde{C}(k)$. Note that for integer shapes ν, no approximation is required.

Figure 4 displays approximations to Matérn covariance functions computed in this way. All of these approximations match the Matérn covariance for at least three values of distance r, those with covariance values 0.1, 0.9 and 1.0. For shape $\nu = 1$, no approximation is necessary, since for this case we have simply $l = 2$, $\alpha = 1$, $\beta = 0$, and $\gamma = 4\pi$. For shape $\nu = 1.5$, the approximate covariance is almost indistinguishable from the Matérn covariance. The approximations are worse for $\nu < 1$, but even for $\nu = 0.5$ may be adequate.

Although I derived the covariance functions $\tilde{c}(r)$ displayed in Figure 4 as approximations, we may consider them to be a practical *alternative* to the Matérn family of covariance functions $c(r)$ displayed in Figure 2. We can control the shape of a function in this alternative family using a single parameter ν, just as we might do for the Matérn family. A key advantage of the alternative covariance functions $\tilde{c}(r)$ is that corresponding model covariance operators can be easily applied by solving partial differential equations.

Table 2. Covariance functions $\tilde{c}(r)$ for $\beta > 0$, which correspond to non-integer values of the shape parameter ν.

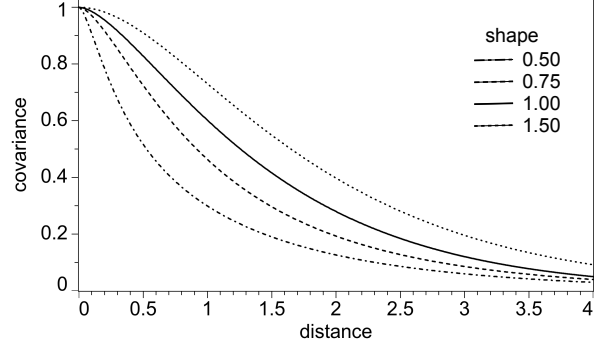| l	$\tilde{c}(r)$
1	$\frac{2}{\log(\alpha/\beta)}\left[K_0\left(\frac{r}{\sqrt{\alpha}}\right) - K_0\left(\frac{r}{\sqrt{\beta}}\right)\right]$
2	$\frac{1}{\alpha-\beta-\beta\log(\alpha/\beta)}\left[\frac{(\alpha-\beta)}{\sqrt{\alpha}}rK_1\left(\frac{r}{\sqrt{\alpha}}\right) - 2\beta K_0\left(\frac{r}{\sqrt{\alpha}}\right) + 2\beta K_0\left(\frac{r}{\sqrt{\beta}}\right)\right]$
3	$\frac{1}{2\alpha[(\alpha-3\beta)(\alpha-\beta)+2\beta^2\log(\alpha/\beta)]}\left[2\sqrt{\alpha}(\alpha-3\beta)(\alpha-\beta)rK_1\left(\frac{r}{\sqrt{\alpha}}\right) + \left(8\alpha\beta^2 + (\alpha-\beta)^2r^2\right)K_0\left(\frac{r}{\sqrt{\alpha}}\right) - 8\alpha\beta^2 K_0\left(\frac{r}{\sqrt{b}}\right)\right]$

Table 3. Parameters for covariance functions $\tilde{c}(r)$.

ν	l	α	β	γ
0.1	1	3.036078	0.000001	2.556095
0.2	1	3.632036	0.000019	3.749303
0.3	1	3.561636	0.000370	4.879283
0.4	1	3.318675	0.002999	5.944607
0.5	1	3.076319	0.012989	7.040834
0.6	1	2.851276	0.037789	8.177463
0.7	1	2.607699	0.087610	9.332528
0.8	1	2.312860	0.178474	10.469799
0.9	1	1.906318	0.351281	11.553648
1.0	2	1.000000	0.000000	12.566371
1.1	2	1.074940	0.005512	13.814264
1.2	2	1.139814	0.017076	15.071712
1.3	2	1.194873	0.035916	16.338362
1.4	2	1.239510	0.063831	17.610024
1.5	2	1.273500	0.102676	18.882510
1.6	2	1.292509	0.157953	20.155772
1.7	2	1.297433	0.232326	21.421235
1.8	2	1.281105	0.337481	22.676615
1.9	2	1.229161	0.500485	23.916472
2.0	3	1.000000	0.000000	25.132741
2.1	3	1.034468	0.016534	26.382951
2.2	3	1.065423	0.039589	27.644620
2.3	3	1.091981	0.070357	28.905683
2.4	3	1.115053	0.108546	30.168595
2.5	3	1.133170	0.157174	31.432672
2.6	3	1.145163	0.219478	32.696440
2.7	3	1.150721	0.297654	33.955588
2.8	3	1.145802	0.402307	35.210680
2.9	3	1.123103	0.554972	36.459091
3.0	4	1.000000	0.000000	37.699112

Figure 4. Alternative covariance functions $\tilde{c}(r)$. Compare with the Matérn covariance functions displayed in Figure 2.

For anisotropic and spatially varying tensor fields $\mathbf{D}(\mathbf{x})$, the corresponding PDE is

$$|\mathbf{D}|^{-\frac{1}{4}}(\mathbf{x})\left(1 - \alpha\boldsymbol{\nabla}\bullet\mathbf{D}(\mathbf{x})\bullet\boldsymbol{\nabla}\right)^l$$
$$\left(1 - \beta\boldsymbol{\nabla}\bullet\mathbf{D}(\mathbf{x})\bullet\boldsymbol{\nabla}\right)|\mathbf{D}|^{-\frac{1}{4}}(\mathbf{x})q(\mathbf{x}) \tag{14}$$
$$= \gamma\,p(\mathbf{x}).$$

I have again constructed the product of SPD operators on the left-hand side of equation 14 to be SPD. To see this, note that the differential operators $1 - \alpha\boldsymbol{\nabla}\bullet\mathbf{D}(\mathbf{x})\bullet\boldsymbol{\nabla}$ and $1 - \beta\boldsymbol{\nabla}\bullet\mathbf{D}(\mathbf{x})\bullet\boldsymbol{\nabla}$ share the same eigenvectors, which implies that they commute. Therefore, the composite left-hand-side operator is SPD.

To solve the partial differential equation 14 numerically, we could approximate the differential operators with finite differences, and then use the method of conjugate gradients to compute the output $q(\mathbf{x})$. However, the condition number for the complete left-hand-side operator grows exponentially with the number of differential operators in parentheses. As an example, for $\nu = 1$ ($l = 2, \alpha = 1, \beta = 0$), the condition number for the complete operator is the square of that for each of the two differential-operator factors, so that a large number of iterations may be required for convergence of the conjugate-gradient method.

A more efficient approach is to use the method of conjugate gradients multiple times, once for each of the differential operator factors on the left-hand side of equation 14. In this approach we solve the following

3.1 PDE implementations

Recall that the motive for an alternative family of covariance functions $\tilde{c}(r)$ is that factors in their Fourier transforms $\tilde{C}(k)$ have only integer exponents, which greatly simplifies PDE implementations of the corresponding model covariance operators $\mathbf{C}_\mathrm{M}$. The PDE corresponding to $\tilde{C}(k)$ in equation 12 is

$$\left(1 - \alpha\boldsymbol{\nabla}\bullet\boldsymbol{\nabla}\right)^l\left(1 - \beta\boldsymbol{\nabla}\bullet\boldsymbol{\nabla}\right)q(\mathbf{x}) = \gamma\,p(\mathbf{x}). \tag{13}$$

sequence of equations:

$$q_0(\mathbf{x}) = |\gamma^2 \mathbf{D}(\mathbf{x})|^{\frac{1}{4}} p(\mathbf{x})$$
$$(1 - \alpha \boldsymbol{\nabla} \cdot \mathbf{D}(\mathbf{x}) \cdot \boldsymbol{\nabla}) \, q_1(\mathbf{x}) = q_0(\mathbf{x})$$
$$(1 - \alpha \boldsymbol{\nabla} \cdot \mathbf{D}(\mathbf{x}) \cdot \boldsymbol{\nabla}) \, q_2(\mathbf{x}) = q_1(\mathbf{x})$$
$$\cdots$$
$$(1 - \alpha \boldsymbol{\nabla} \cdot \mathbf{D}(\mathbf{x}) \cdot \boldsymbol{\nabla}) \, q_l(\mathbf{x}) = q_{l-1}(\mathbf{x})$$
$$(1 - \beta \boldsymbol{\nabla} \cdot \mathbf{D}(\mathbf{x}) \cdot \boldsymbol{\nabla}) \, q_L(\mathbf{x}) = q_l(\mathbf{x})$$
$$q(\mathbf{x}) = |\gamma^2 \mathbf{D}(\mathbf{x})|^{\frac{1}{4}} q_L(\mathbf{x}). \quad (15)$$

In the special case where $\beta = 0$ (ν is an integer), solution of the last PDE for $q_L(\mathbf{x})$ is unnecessary. In any case, the total number of conjugate-gradient iterations grows only linearly, not exponentially, with the number of equations 15.

The solution of each PDE in the sequence of equations 15 is equivalent to applying a smoothing filter to a function $q_i(\mathbf{x})$ on the right-hand side. Tensor coefficients $\mathbf{D}(\mathbf{x})$ can be specified so that this smoothing is both anisotropic and spatially varying. At each location $\mathbf{x}$ the extent of smoothing is greatest in directions of eigenvectors corresponding to the largest eigenvalues of $\mathbf{D}(\mathbf{x})$. The cascade of tensor-guided smoothing filters in equations 15 approximates the application of a Matérn model covariance operator $\mathbf{C}_M$ in which covariance is both anisotropic and spatial varying.

Because this approximation is implemented with smoothing filters, it is henceforth referred to as a *smoothing covariance*.

4   TENSOR-GUIDED KRIGING

A simple and common use of a model covariance operator $\mathbf{C}_M$ is in the interpolation of measurements acquired at locations scattered in space. In this application, the linear operator $\mathbf{G}$ in equation 1 is simply a model-sampling operator $\mathbf{K}$ that extracts values of the model $\mathbf{m}$ at scattered locations to obtain data $\mathbf{d} \approx \mathbf{Km}$. Here, the approximation is due to measurement errors that may be non-zero. Substituting the model-sampling operator $\mathbf{K}$ for $\mathbf{G}$ in equation 1, we obtain

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_M \mathbf{K}^\top (\mathbf{K} \mathbf{C}_M \mathbf{K}^\top + \mathbf{C}_D)^{-1}(\mathbf{d} - \mathbf{Km}_0). \quad (16)$$

As noted by Hansen et al. (2006), the process of computing a model estimate $\tilde{\mathbf{m}}$ with equation 16 is equivalent to gridding with simple kriging, a process well known in geostatistics. In this process, we estimate the model $\mathbf{m}$ at locations on a uniform and dense sampling grid. The number of gridded-model samples in $\mathbf{m}$ is typically much larger than the number of scattered-data samples in $\mathbf{d}$. The operator $\mathbf{K}$ gathers values from a small subset of locations in the uniform grid, the scattered locations where data are available, and the operator $\mathbf{K}^\top$ scatters values into those same locations.

Tarantola (2005) shows that simple kriging can also

be performed in a different but equivalent way:

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + (\mathbf{C}_M^{-1} + \mathbf{K}^\top \mathbf{C}_D^{-1} \mathbf{K})^{-1} \mathbf{K}^\top \mathbf{C}_D^{-1}(\mathbf{d} - \mathbf{Km}_0). \quad (17)$$

This alternative is appealing because finite-difference approximations to $\mathbf{C}_M^{-1}$ are compact and can be applied more efficiently than those for $\mathbf{C}_M$, which requires solution of partial differential equations 15. However, equations 16 and 17 are equivalent only when the inverses of matrices in these equations exist. If measurement errors are negligible, so that $\mathbf{C}_D \approx \mathbf{0}$, then $\mathbf{C}_D$ is nearly singular and equation 17 is ill-conditioned. The fact that equation 17 is invalid without measurement error is just one reason to favor gridding with equation 16.

Another reason is that the size of the matrix $\mathbf{K} \mathbf{C}_M \mathbf{K}^\top + \mathbf{C}_D$ equals the number of scattered data samples, which is often small enough to enable efficient direct solution of the kriging equations 16. In contrast, the size of the matrix $\mathbf{C}_M^{-1} + \mathbf{K}^\top \mathbf{C}_D^{-1} \mathbf{K}$ equals the (typically) much larger number of gridded model samples, so that iterative solution of equation 17 is required. When $\mathbf{C}_D \approx \mathbf{0}$, and without a good preconditioner, convergence of iterative methods is slow.

For these reasons I choose to use equation 16 to implement gridding with an anisotropic and spatially varying model covariance. However, even with this choice, an iterative solution is required, because we lack analytic expressions for elements of the smoothing covariance matrix $\mathbf{C}_M$ and, hence, the composite matrix $\mathbf{A}_M \equiv \mathbf{K} \mathbf{C}_M \mathbf{K}^\top + \mathbf{C}_D$. The smoothing equations 15 provide only a method for applying (performing multiplication by) $\mathbf{C}_M$. Nevertheless, that method is sufficient for iterative conjugate-gradient solution of equation 16.

4.1   Paciorek's approximation

Convergence of conjugate-gradient iterations is greatly accelerated by a good preconditioning operator $\mathbf{P} \approx \mathbf{A}_M^{-1}$, one that can be computed and applied more quickly than $\mathbf{A}_M$ itself. Recalling that $\mathbf{A}_M \equiv \mathbf{K} \mathbf{C}_M \mathbf{K}^\top + \mathbf{C}_D$, one way to obtain such a preconditioner is to find a good approximation to the smoothing covariance operator $\mathbf{C}_M$.

Paciorek (2003) and Paciorek and Schervish (2006) propose an approximation $\mathbf{C}_P \approx \mathbf{C}_M$ whose elements can be computed quickly by the following modification of equation 6:

$$c(r) \rightarrow a(\mathbf{x}, \mathbf{y}) \, c\left(\sqrt{(\mathbf{x} - \mathbf{y})^\top \bar{\mathbf{D}}^{-1}(\mathbf{x}, \mathbf{y})(\mathbf{x} - \mathbf{y})}\right), \quad (18)$$

where

$$\bar{\mathbf{D}}(\mathbf{x}, \mathbf{y}) \equiv \frac{\mathbf{D}(\mathbf{x}) + \mathbf{D}(\mathbf{y})}{2}, \quad (19)$$

and

$$a(\mathbf{x}, \mathbf{y}) \equiv |\mathbf{D}(\mathbf{x})|^{\frac{1}{4}} |\mathbf{D}(\mathbf{y})|^{\frac{1}{4}} |\bar{\mathbf{D}}(\mathbf{x}, \mathbf{y})|^{-\frac{1}{2}}. \quad (20)$$

In effect, these expressions approximate the covariance of the model at location $\mathbf{x}$ and the model at location $\mathbf{y}$ using averages of tensors $\mathbf{D}(\mathbf{x})$ and $\mathbf{D}(\mathbf{y})$ at only those two locations. The approximation is best where $\mathbf{D}(\mathbf{x})$ varies slowly within the effective range of that function.

Assuming that the number of scattered measurements in $\mathbf{d}$ is sufficiently small, say, less than 1000, we can use Paciorek's approximation to quickly compute and store the elements of the approximate composite matrix $\mathbf{A}_P \equiv \mathbf{K}\mathbf{C}_P\mathbf{K}^\top + \mathbf{C}_D$. We can then use Cholesky decomposition to compute the preconditioner $\mathbf{P} = \mathbf{A}_P^{-1}$, for use in a conjugate-gradient solution of equation 16. The resulting process is a method for performing simple kriging with an anisotropic and spatially varying Matérn model covariance or, more simply, *tensor-guided kriging*.

4.2   Simulated model and data

To test the process, I synthesized data by sampling a known gridded model at 256 random locations. Figure 5a shows the known model porosities $\mathbf{m}$, while Figure 5b shows the sampled data porosities $\mathbf{d}$. The data were sampled without error; i.e., the data covariance $\mathbf{C}_D = \mathbf{0}$.

The model has a known covariance $\mathbf{C}_M$ that I derived from seismic amplitudes, also shown in Figure 5. (The images in Figures 1a and 1c are identical to those in Figures 5c and 5a, respectively.) These amplitudes were extracted from a 3D seismic image at depths corresponding to a seismic horizon. Stratigraphic features apparent in this image suggest an anisotropic and spatially varying model covariance $\mathbf{C}_M$.

I computed the known model $\mathbf{m} \equiv m(\mathbf{x})$ displayed in Figure 5a by smoothing an image $\mathbf{r} \equiv r(\mathbf{x})$ of pseudo-random values independently generated for a normal distribution $\mathcal{N}(0.25, 0.02)$. The smoothing was performed by solving a finite-difference approximation to the following partial differential equation:

$$(1 - \boldsymbol{\nabla} \bullet \mathbf{D}(\mathbf{x}) \bullet \boldsymbol{\nabla})\, m(\mathbf{x}) = |\gamma^2 \mathbf{D}(\mathbf{x})|^{\frac{1}{4}} r(\mathbf{x}), \qquad (21)$$

where $\mathbf{D}(\mathbf{x})$ are structure tensors computed from the seismic image displayed in Figure 5c.

Equation 21 is comparable to the "stochastic Laplace equation" described by Whittle (1954), but here with anisotropic and spatially varying coefficients $\mathbf{D}(\mathbf{x})$. Equation 21 is also equivalent to the first two equations in the sequence of equations 15, for the special case where the Matérn shape is $\nu = 1$ ($l = 2$, $\alpha = 1$, $\beta = 0$). In this case, if we define this first half of equations 15 as a linear operator $\mathbf{F}$, then the second half of those same equations is its transpose $\mathbf{F}^\top$. In summary, I used the factor $\mathbf{F}$ in $\mathbf{C}_M = \mathbf{F}\mathbf{F}^\top$ to generate a known porosity model $\mathbf{m}$ with model covariance $\mathbf{C}_M$ (Cressie, 1993).

It is important to emphasize that I did not assume any direct correlation between seismic amplitudes and
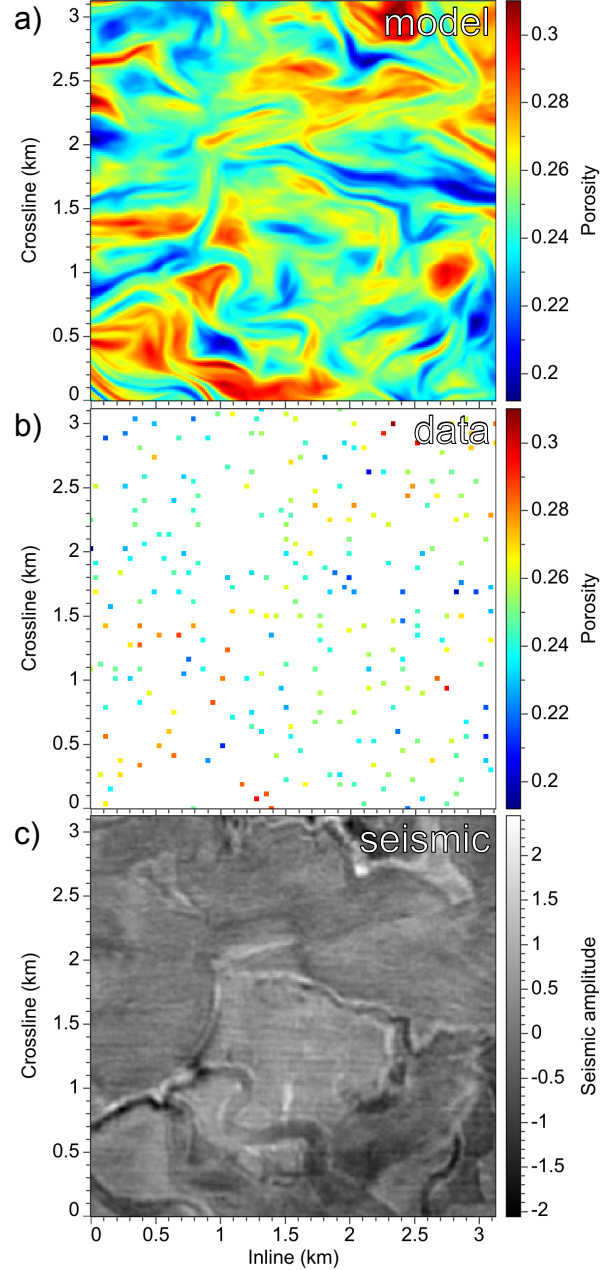
Figure 5. A simulated known gridded model $\mathbf{m}$ (a) and scattered data $\mathbf{d}$ (b) used to test tensor-guided kriging. The model covariance was derived from seismic amplitudes (c) in a horizon slice of a 3D seismic image. The size of the pixels in (b) has been exaggerated to make the locations of scattered data samples more visible.

porosities. Instead, I used the seismic image only to construct an anisotropic and spatially varying model covariance for porosity. Structure tensors $\mathbf{D}(\mathbf{x})$ computed from the seismic image enable the effective range of the model covariance $\mathbf{C}_M$ to be both anisotropic and

spatially varying. For this example I set the maximum range to be 1 km.

4.3 Gridding with kriging

Figure 6 displays model estimates $\tilde{\mathbf{m}}$ obtained by kriging via equation 16 the simulated data $\mathbf{d}$ for three different model covariances. In all three cases, I used the correct model mean $\mathbf{m}_0 \equiv m_0(\mathbf{x}) = 0.25$.

For Figure 6a, I used the correct smoothing covariance $\mathbf{C}_M$ (and $\mathbf{C}_D = \mathbf{0}$) in tensor-guided kriging with equation 16 to estimate the model $\mathbf{m}$. This model estimate $\tilde{\mathbf{m}}$ is most accurate because it is most consistent with the method (described above) used to synthesize the data $\mathbf{d}$. In the iterative conjugate-gradient solution of equation 16 I used a preconditioner based on Paciorek's approximation. The model estimate shown here was obtained after 16 iterations.

For Figure 6b, I replaced $\mathbf{C}_M$ in equation 16 with Paciorek's approximate model covariance $\mathbf{C}_P$, for which the matrix $\mathbf{A}_P = \mathbf{K}\mathbf{C}_P\mathbf{K}^\top + \mathbf{C}_D$ can be efficiently computed and factored directly. In this example, the resulting model estimate $\tilde{\mathbf{m}}$ differs significantly from that obtained using the correct model covariance $\mathbf{C}_M$ (Figure 6a), because the tensors $\mathbf{D}(\mathbf{x})$ vary significantly within the maximum range (1 km) of the covariance function. This result indicates that, while Paciorek's approximation may provide a useful preconditioner for tensor-guided kriging, it may not be an adequate substitute for the correct smoothing covariance $\mathbf{C}_M$.

To estimate the model displayed in Figure 6c, I replaced $\mathbf{C}_M$ with a simple isotropic and spatially invariant model covariance. The resulting model estimate $\tilde{\mathbf{m}}$ is least accurate of those displayed in Figure 6, because the correct model covariance $\mathbf{C}_M$ is both anisotropic and spatially varying.

The implementation of tensor-guided kriging demonstrated with this simple 2D example has a significant shortcoming not addressed in this paper. Recall that the number of scattered measurements in the data vector $\mathbf{d}$ in this example is 256. As that number increases to, say, 10,000 or more, the computational cost of the Paciorek preconditioner becomes prohibitive; this cost is mostly that of Cholelsky decomposition of the matrix $\mathbf{A}_P$, which grows with the cube of the number of data samples.

This high cost makes the implementation of tensor-guided kriging proposed in this paper infeasible for 3D gridding of scattered borehole data. For such large 3D subsurface gridding problems, we might exploit the fact that the range for vertical correlation of subsurface properties is typically much smaller than that for lateral correlation, and solve the large problem with a sequence of solutions to overlapping smaller problems.
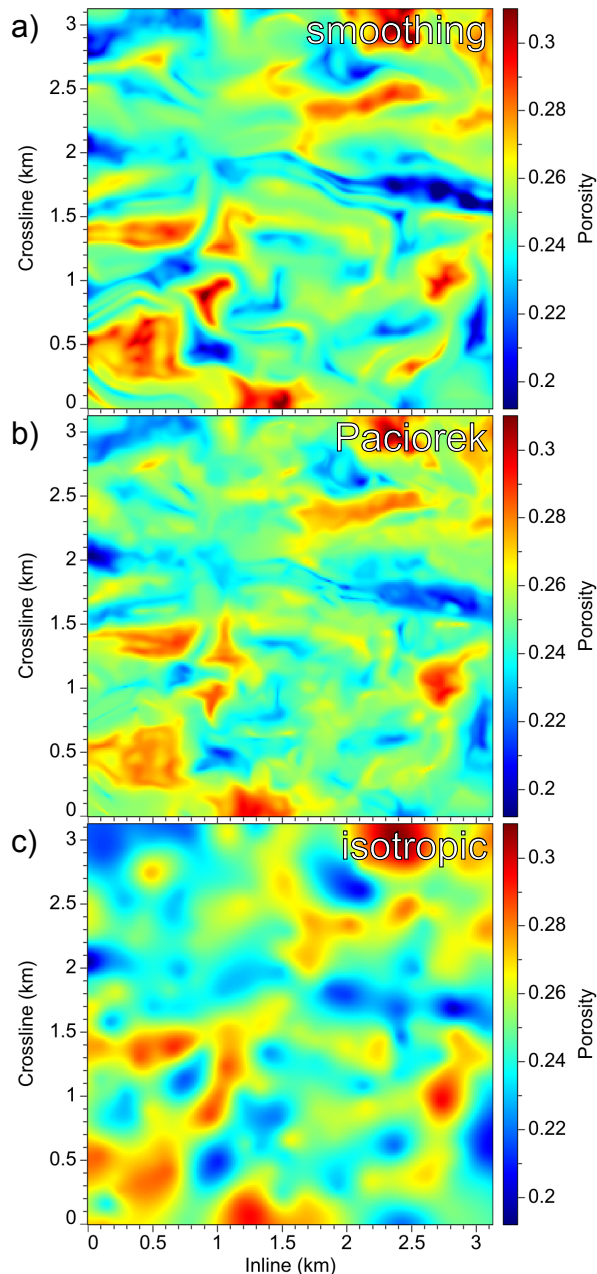
Figure 6. Porosity models estimated using (a) the correct anisotropic and spatially varying smoothing model covariance $\mathbf{C}_M$, (b) Paciorek's approximate model covariance $\mathbf{C}_P$, and (c) an isotropic and spatially invariant model covariance. The maximum range for all three model covariances is 1 km.

5 CONCLUSION

By solving a cascade of partial differential equations with tensor coefficients, we effectively implement an anisotropic and spatially varying Matérn model covariance $\mathbf{C}_M$. In addition to the tensor coefficients $\mathbf{D}(\mathbf{x})$, this model covariance requires only a few additional pa-

rameters: variance σ^2, maximum effective range a, and shape ν. The latter parameters are well-known to those familiar with the popular Matérn covariance function.

For some inverse problems, the necessary tensor field $\mathbf{D}(\mathbf{x})$ can be obtained directly from auxiliary data, such as seismic images. For other problems, a necessary first step might be for experts to specify this tensor field, perhaps using new interactive software tools developed for this purpose. In any case, it is difficult to imagine a geophysical inverse problem for which an isotropic or spatially invariant model covariance is appropriate.

ACKNOWLEDGMENT

REFERENCES

Cressie, N., 1993, Statistics for spatial data: Wiley.

Fehmers, G., and C. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, 1286–1293.

Fuglstad, G., 2011, Spatial modeling and inference with SPDE-based GMRFs: Master's thesis, Norwegian University of Science and Technology.

Guttorp, P., and T. Gneiting, 2006, Studies in the history of probability and statistics XLIX: on the Matérn correlation family: Biometrika, **93**, 989–995.

Handcock, M., and J. Wallis, 1994, An approach to statistical spatial-temporal modeling of meteorological fields: Journal of the American Statistical Association, **89**, 368–378.

Hansen, T., A. Journel, A. Tarantola, and K. Mosegaard, 2006, Linear inverse Gaussian theory and geostatistics: Geophysics, **71**, R101–R111.

Lindgren, F., H. Rue, and J. Lindström, 2011, An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach: Journal of the Royal Statistical Society: Series B (Statistical Methodology), 423–498.

Paciorek, C., 2003, Nonstationary Gaussian processes for regression and spatial modelling: PhD thesis, Carnegie Mellon University.

Paciorek, C., and M. Schervish, 2006, Spatial modelling using a new class of nonstationary covariance functions: Environmetrics, **17**, 483–506.

Stein, M., 1999, Statistical interpolation of spatial data: some theory for kriging: Springer, New York.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics.

Weickert, J., 1999, Coherence-enhancing diffusion filtering: International Journal of Computer Vision, **31**, 111–127.

Whittle, P., 1954, On stationary processes in the plane: Biometrika, **41**, 434–449.