The use of systems of stochastic PDEs as priors in seismic AVA inversion

Erlend Aune and Daniel P. Simpson

Abstract A challenge in seismic amplitude versus angle (AVA) inversion to elastic parameters is the inclusion of prior information from geology, geophysical relations, well logs and other sources. In a recent publication, Lindgren et al. (2011) developed the necessary background for using linear stochastic partial differential equations (SPDEs) as prior fields in latent Gaussian models and highlighted the link between such representations and Matérn covariance functions. This approach allows for flexible incorporation of nonstationarity and anisotropy in the prior model. Another advantage is that the prior field is Markovian and therefore the precision matrix is very sparse, introducing huge computational and memory benefits. The seismic AVA inversion problem is essentially a trivariate random field inversion problem and the extension of the univariate SPDE approach entails using a system of SPDEs as priors. This allows us to control stationarity, anisotropy and smoothness of the individual elastic parameters as well as for the link between them through the cross-covariance SPDEs and therefore allows us to make more realistic prior models. We explore the potential benefits this approach may have in seismic AVA inversion.

1 Introduction

The seismic AVA inversion problem is well studied in the geophysical litterature, and there are several incarnations of it with varying degrees of complexity. In this article, we concern ourselves with the inversion problem studied in (Buland and Omre, 2003; Buland et al., 2003; Rabben et al., 2008),

Erlend Aune
NTNU, Norway, e-mail: erlend.aune.1983@gmail.com

Daniel P. Simpson
University of Helsinki, Finland e-mail: dp.simpson@gmail.com

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

using the wavefield propagation approximations in Aki and Richards (1980), which results in linear systems of equations to solve. Variants and extensions of these equations are found in Stovas and Ursin (2003), including nonlinear approximations that may yield better inversion results in some situations.

2 Prior specification

The choice of prior in the inversion problem is of great importance when it comes to the performance of the inversion. It is vital to choose a "good" prior to emphasise the properties of $\boldsymbol{m}$ that we know it has. For us, $\boldsymbol{m}$ will denote the elastic parameters of interest, and it depends on position. How we do this is a matter of heuristics and geological knowledge. For a Gaussian prior model, the standard way of specifying the prior model is through the use of covariance functions for stationary fields (see, e.g. Buland and Omre (2003)). A covariance function in the stationary case is defined by a correlation function that defines how much a point is correlated with its neighbours and a marginal variance parameter, ϱ^2 through

$$\varrho^2 c(\|\boldsymbol{x}-\boldsymbol{y}\|) = \mathrm{Cov}(\boldsymbol{x},\boldsymbol{y}), \tag{1}$$

and in the Gaussian case, this defines a strictly stationary process. There is a list of widely used covariance functions in Cressie (1993). In the same reference, the theory of covariance functions is elaborated upon. We will throughout this text assume that the prior is from the Gaussian family. This family is defined by having density

$$p(\boldsymbol{x}|\mathbf{Q},\boldsymbol{\mu}_x) = (2\pi)^{n/2}\det(\mathbf{Q})^{1/2}\exp\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^T\mathbf{Q}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right), \tag{2}$$

where $\mathbf{Q}=\boldsymbol{\Sigma}^{-1}$ is the precision matrix – the inverse of the covariance matrix $\boldsymbol{\Sigma}$ – and $\boldsymbol{\mu}_x$ is the expectation, $\mathbb{E}(\boldsymbol{x}|\mathbf{Q},\boldsymbol{\mu}_x)$.

Moreover, the fields m_1, m_2, m_3 are assumed correlated with correlations specified by well data and/or other local knowledge. In the discretized domain, this allows for the following decomposition of the total covariance matrix

$$\boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}_{\text{space}} \otimes \boldsymbol{\Sigma}_0 \tag{3}$$

where $\boldsymbol{\Sigma}_{\text{space}}$ denotes the spatial covariance matrix, typically defined through a covariance function, and $\boldsymbol{\Sigma}_0$ the correlations between the elastic parameters. Since seismic observations typically are on a regular grid, either in 2-D or 3-D, it possible to let $\boldsymbol{\Sigma}_{\text{space}}$ be circulant by extending the grid by as many points as is needed to get the correlation below a threshold – typically 0.1 or 0.05. This allows us to use fast Fourier transforms for computing quantities of interest related to the covariance matrix. This, together with the Kronecker

structure of $\mathbf{\Sigma}_m$ allows for fast computations. See Buland et al. (2003); Rue and Held (2005); Gray (2006) for details. This approach also has very low memory requirements; since $\mathbf{\Sigma}_{\text{space}}$ is circulant it may be stored using only one vector. Hence storage is $\mathcal{O}(n)$ and computations (of any kind) are at most $\mathcal{O}(n \log n)$.

2.1 SPDE formulation

While this decomposition is sensible, it is also very inflexible and requires stationarity for low storage requirements. Another way of pursuing good prior models with fast computations and low memory requirements is through the use of elliptic (pseudo) differential operators (Ruzhansky and Turunen (2009), part 2 is an accessible source). The theory of pseudo differential operators is closely related to Weyl transforms and short-time Fourier transforms or Gabor transforms (Feichtinger et al. (2008)) and usual spectral considerations is seismology apply. In this approach, it is the sparsity of the resulting precision matrices that makes storage and computation manageable. Recently, Lindgren et al. (2011), studied how to apply such operators in a statistical setting. They studied a Laplace-like operator, $(-\triangle + \kappa)^{\alpha/2}$ and its relation to computation and Matérn covariance models (Matérn, 1960; Whittle, 1963). The main lessons are firstly, if

$$M_{\kappa,\alpha} x(s) = (\kappa^2 - \triangle)^{\alpha/2} x(s) = \mathcal{W}(s), \tag{4}$$

where $\mathcal{W}$ is spatial Gaussian white noise, then x has Matérn type covariance function, i.e.,

$$\rho(r) = \frac{\varrho^2}{\Gamma(\alpha - d/2) 2^{\alpha - d/2 - 1}} (\kappa r)^{\alpha - d/2} K_{\alpha - d/2}(\kappa r), \tag{5}$$

$$\varrho^2 = \frac{\Gamma(\alpha - d/2)}{\Gamma(\alpha)(4\pi)^{d/2} \kappa^{2(\alpha - d/2)}}, \tag{6}$$

where K_s is the modified Bessel function of the first kind. Secondly, fast computations through finite element methods or other discretisations of the differential operator in (4) are available through the induced Markov properties of the discretisation matrix, $\mathbf{Q}_{\text{space}}$. That essentially means that $\mathbf{Q}_m = \mathbf{Q}_{\text{space}} \otimes \mathbf{Q}_0$ is (very) sparse and with a structure ameanable to Cholesky factorisation. An alternative requirement is that we can construct the matrix vector product $\mathbf{Q}_m \boldsymbol{v}$ and $\det(\mathbf{Q}_m)$ relatively quickly through some iterative or direct procedure.

When addressing the "stationarity" of the field defined by (4), it is only stationary in the sense of (1) if it is defined on the whole of $\mathbb{R}^k$, where $k = 2, 3$ in our case – alternatively when the corresponding operator is defined on a

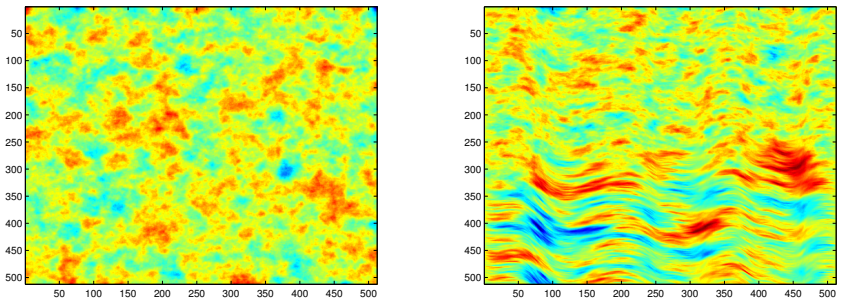
manifold without boundary. In our kase the domain on which (1) is defined is merely a subset, namely a square or box in $\mathbb{R}^2$ or $\mathbb{R}^3$. Hence boundary effects resulting from boundary conditions may destroy its direct interpretability in terms of this equation. It is, of course, possible to specify boundary conditions in such a way that you retain the property in (1), but usually there are more natural physical boundary conditions that in our opinion improves upon the specification through SPDEs compared to the model defined by covariance matrices through stationary covariance functions also in the stationary case.

There are two properties that are desirable to have in the prior model in AVA inversion. The first is being able to have different correlation length at different points in space. If a geologist have sound reasons to believe that a layer is very inhomogeneous, it may warrant putting a lower correlation length here than in a layer that is thought to be very homogeneous with very similar properties. Facilitating this is trivial - one merely lets $\kappa^2 = \kappa^2(s)$ vary with space. The other property that is very desirable to have is anisotropy. Letting the correlation length vary with direction is very natural given that the layers are typically not flat but are deformed in a specific way. The SPDE resulting is the following variant of (4):

$$M_{\kappa,\alpha,s}x(s) = (\kappa^2(s) - \nabla \cdot \mathbf{A}(s)\nabla)^{\alpha/2}x(s) = \mathcal{W}(s), \tag{7}$$

where $\mathbf{A}$ is a 3×3 symmetric positive definite matrix defining the anisotropy angle and principal correlation length in the three directions defined by the eigenvectors of the matrix. Realisations of the stationary model and the nonstationary model is given in Figure 1. Here we have illustrated the "layer" flexibility mentioned above, where the top layer is isotropic, and the bottom layer is anisotropic with deformation defined by the layer.

Fig. 1 Realisations from stationary model given by (4) (left) and nonstationary model given by (7) (right)



To see how this relates to the usual approach, consider $\mathbf{Q}_0 = \mathbf{\Sigma}_0^{-1}$ and say that m_1, m_2, m_3 have equal Matérn covariance models (this includes the

widely used exponential and Gaussian models), then the prior given as in (3) is given by the following system of stochastic differential equations:

$$(M_{\kappa,\alpha,s} \otimes \mathbf{Q}_0)\boldsymbol{m} = \boldsymbol{\mathcal{W}} \tag{8}$$

where $\boldsymbol{\mathcal{W}}$ is vector Gaussian spatial white noise. The experience in AVA-inversion is that at least one component of $\boldsymbol{m}$ worse resolved than the others, with m_1 being resolved the best (see Rabben et al. (2008) or any other article treating this problem). The obvious next question then is whether or not (8) specifies the best way of lending strength to the least resolved parameters. If not, can we find better operators on the diagonal in (8), and/or replace the off-diagonals with other operators that have better properties in the inversion problem? The answer to this question is not obvious, but we investigate some alternatives and see how they perform in our inversion problem; the criterion for a better prior in the synthetic case being that $\mathbb{E}((\boldsymbol{m_{true}} - \boldsymbol{m_{est}^{new}})^2) < \mathbb{E}((\boldsymbol{m_{true}} - \boldsymbol{m_{est}^{base}})^2)$, where $\boldsymbol{m_{est}^{base}}$ is given by the prior model (3).

It is possible to replace the operator $M_{\kappa,\alpha}$ in (8) by more general pseudo-differential operators. Representations of such operators in terms of its symbol are given by

$$(K_\sigma f)(x) = \int_{\mathbb{R}^d} \sigma(x,\xi)\hat{f}(\xi)e^{2\pi i x\cdot\xi}d\xi, \tag{9}$$

where $\hat{f}$ is the Fourier transform of f, and σ is the symbol of the operator. The symbol can be interpreted as defining the local spectrum of the operator. A deep theorem given in Rozanov (1977) states that a stationary random field is Markov (in the continuous sense) if and only if σ^{-1} is a polynomial. Hence Markov fields are represented by differential operators. Now, if the field in question is not Markov, it is possible to approximate σ by a rational approximation, $\sigma(x,\xi)^{-1} \approx \sigma_{\text{rat}}^{-1}(x,\xi) = \sum_{j=0}^{k} a_j(x)(2\pi i\xi)^j$. To find the a_js one can, for instance, use optimisation techniques. This is one way to do it, but we suspect that the time-frequency localisation of such an approach may be suboptimal, and discretization of the non-Markov operator may be better suited for time-frequency compressing approaches inducing approximate Markovity. We do not pursue these type of ideas here, but mention them as they are good candidates for future research.

3 Systems of SPDEs - generalising "$\mathbf{Q_0}$"

It is easy to write the form the generalised approach must have. First, for $i, j = 1, \ldots, 3$, let

$$K_{ij} = q_{ij}(\boldsymbol{s})(\kappa_{ij}(\boldsymbol{s}) - \nabla \cdot \mathbf{A}_{ij}(\boldsymbol{s})\nabla)^{\alpha_{ij}/2} \tag{10}$$

and define the following system of SPDEs

$$\mathbf{K}\boldsymbol{m}(\boldsymbol{s}) = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{12} & K_{22} & K_{23} \\ K_{13} & K_{23} & K_{33} \end{pmatrix} \boldsymbol{m}(\boldsymbol{s}) = \boldsymbol{\mathcal{W}}(\boldsymbol{s}) \tag{11}$$

For $q_{ij}(\boldsymbol{s}) = Q^0_{ij}$ and $K_{ij} = M_{\kappa,\alpha}$ we recover the structure in the previous section with stationarity. In Hu et al. (2012), they study the properties of this model in the stationary case, and give the link to the multivariate Matérn fields in Gneiting et al. (2010). Any choice of K_{ij} defines a valid Gaussian Markov random field, both in the continuous sense and when discretized. In order to reduce overparametrisation and increase interpretability, it is, however, natural to restrict the models somewhat.

One possible way to reduce the parametrisation demand is to do the modeling in the Cholesky domain. This is a simplification, but it is one we believe should increase interpretability and possibly estimation properties. To motivate this approach, consider the following: Suppose that the Cholesky factorisation of $\mathbf{Q}_0$ is given by $\mathbf{Q}_0 = \boldsymbol{L}_0\boldsymbol{L}_0^T$, and that $\mathbf{Q}_{\text{space}} = \mathbf{Q}_1^s\mathbf{Q}_2^s$, for some matrices $\mathbf{Q}_1^s, \mathbf{Q}_2^s$. Generating the matrices $\{\mathbf{Q}_i^s\}_{1,2}$ can for instance be done by using $\alpha_s = \alpha/2$ in (4) and discretizing this operator, but there exist many other factorisations that may behave in better way for the problem at hand. By a Kronecker product identity, $\mathbf{Q}_{\text{space}} \otimes \mathbf{Q}_0 = (\mathbf{Q}_1^s \otimes \boldsymbol{L}_0)(\mathbf{Q}_2^s \otimes \boldsymbol{L}_0^T)$. The intuition stemming from this identity carries over to the more general case in a natural way: Let $l_{ij}(\boldsymbol{s})$ be entry i, j of the Cholesky factor of the matrix $\{q_{ij}(\boldsymbol{s})\}_{ij}$ locally, and define locally operators that will correspond to some square root of its original form in (10). It is possible to define the operators in such a way that we get back (11), but this is of minor concern in practice as long as we get the interpretability we want. This is remniscient to the triangular approach mentioned in Hu et al. (2012).

3.1 Parametrising $q_{ij}(\boldsymbol{s})$

In general, it is both hard to interpret a local precision matrix, $\mathbf{Q}_0(\boldsymbol{s}) = \{q_{ij}(\boldsymbol{s})\}_{ij}$ defining how the individual parts of the multivariate fields is related to each other at position $\boldsymbol{s}$, and to ensure that this matrix is positive definite. It is much more natural to work with the inverse, namely the correlation matrix defining the local correlation of the fields, $\boldsymbol{\Sigma}_0(\boldsymbol{s}) = \mathbf{Q}_0^{-1}(\boldsymbol{s})$. Information about correlation in different layers may come from geologists or geophysicists who may know of phase changes when going from one layer to another in the different layers, or other, more complex phenomena. It may also come from well-logs that may contain information about such matters.

If $\mathbf{\Sigma}_0(\boldsymbol{s}) = \mathbf{\Sigma}_{0,1}$ for $\boldsymbol{s} \in S_1 \subset \mathbb{R}^d$ and $\mathbf{\Sigma}_{0,2}$ for $\boldsymbol{s} \in S_2 \subset \mathbb{R}^d$, there is obviously a transition between these two states. If the transition is discontinuous, this may be seen as a discontinuity in the realisation of the multivariate random field - rarely a desirable property - but it may make sense in some situations. There are obviously many ways of making a smooth transition between $\mathbf{\Sigma}_{0,1}$ and $\mathbf{\Sigma}_{0,2}$, but one key consideration is that $\mathbf{\Sigma}_0(\boldsymbol{s})$ must remain positive definite for all $\boldsymbol{s}$ in some transition domain S_T. One thing is certain - it is not necessarily enough to let the off-diagonals element in $\mathbf{\Sigma}_{0,1}$ change linearly in $\mathbb{R}^3$ to the corresponding off-diagonal elements in $\mathbf{\Sigma}_{0,2}$.

A very natural way of making such a transition between $\mathbf{\Sigma}_{0,1}$ and $\mathbf{\Sigma}_{0,2}$ is by considering geodesics on the manifold of symmetric positive definite matrices, denoted $\mathbb{P}_d$. The natural metric on this space has a reasonable statistical interpretation, closely related to information entropy and Kullback–Leibler divergence, and an accessible account for the theory is given in Bhatia (2007). Different treatments are given in (Ohara et al., 1996; Hiai and Petz, 2009). For completeness, we give a small account of the definition and properties we need related to this manifold. This exposition is based on Hiai and Petz (2009); Bhatia (2007).

The Boltzmann entropy of the Gaussian distribution (2), defining an information potential, is given by

$$B(p(\boldsymbol{x}|\mathbf{Q}, \boldsymbol{\mu}_x)) = B(\mathbf{Q}) = \frac{1}{2} \log \det \mathbf{\Sigma} + C, \tag{12}$$

where C is an arbitrary constant, $\mathbf{\Sigma} = \mathbf{Q}^{-1}$. The Riemannian metric based on this information potential is the Hessian

$$\mathrm{g}_{\mathbf{Q}}(\mathbf{H}, \mathbf{M}) = \frac{\partial^2}{\partial s \partial t} B(\mathbf{Q} + s\mathbf{H} + t\mathbf{M}) = \operatorname{tr} \mathbf{QHQK}, \tag{13}$$

where $\mathbf{H}, \mathbf{S} \in \mathbb{S}_d$, the tangent space of symmetric matrices, $\mathbb{S}_d = \{\mathbf{V} \in \mathbb{R}^{d \times d} | \mathbf{V} = \mathbf{V}^T\}$. This defines the line element

$$ds = \left(\operatorname{tr} \left[(\mathbf{Q}^{-1/2} d\mathbf{Q} \mathbf{Q}^{-1/2})^2 \right] \right)^{1/2}. \tag{14}$$

Hence, if we have a curve in $\mathbb{P}_d$, i.e. $\gamma : [a, b] \to \mathbb{P}_d$, its length can be calculated as

$$L(\gamma) = \int_a^b \left(\operatorname{tr} \left[(\gamma(t)^{1/2} \gamma'(t) \gamma(t)^{1/2})^2 \right] \right)^{1/2} dt \tag{15}$$

A nice property that follows from this is that lengths of curves are invariant under congruence transformations. That is, if $g(t) = \mathbf{X}^T \gamma(t) \mathbf{X}$, $L(\gamma) = L(g)$. The geodesic, the curve with minimal length, between two matrices, $\mathbf{A}$ and $\mathbf{A}$ can from this be deduced to be

$$\mathrm{g}_{\mathbf{A},\mathbf{B}}(t) = \mathbf{A}\#_t\mathbf{B} = \mathbf{A}^{1/2}\left(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\right)^t\mathbf{A}^{1/2}, \quad t\in[0,1]. \tag{16}$$

Obviously, $\mathrm{g}_{\mathbf{A},\mathbf{B}}(0) = \mathbf{A}$ and $\mathrm{g}_{\mathbf{A},\mathbf{B}}(1)$. It is this curve we use when we go from $\mathbf{A} = \mathbf{Q}_{0,1} = \boldsymbol{\Sigma}_{0,1}^{-1}$ to $\mathbf{B} = \mathbf{Q}_{0,2} = \boldsymbol{\Sigma}_{0,2}$ in different geological layers in our prior model, and this ensures that we are within the realm of positive definite matrices in a natural way. Noting that $(\mathbf{A}\#_t\mathbf{B})^{-1} = \mathbf{A}^{-1}\#_t\mathbf{B}^{-1}$, we see that it is unproblematic to work with precision matrices rather than covariance matrices. Integrating $\mathrm{g}_{\mathbf{A},\mathbf{B}}(t)$ yields the distance between the two matrices,

$$d_{\mathbb{P}_d}(\mathbf{A},\mathbf{B}) = \int_0^1 \mathrm{g}_{\mathbf{A},\mathbf{B}}(t) = \left(\mathrm{tr}\left[(\log\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2})^2\right]\right)^{1/2}. \tag{17}$$

A potential drawback of using this method is that if $\mathbf{Q}_{0,1}, \mathbf{Q}_{0,2}$ are correlation matrices, and what you want is a continuum of correlation matrices, $\mathrm{g}_{\mathbf{Q}_{0,1},\mathbf{Q}_{0,2}}(t)$ are not correlation matrices for $t\in(0,1)$. It is possible to correct for this by using geodesics on the submanifold of correlation matrices in $\mathbb{P}_d$. In practice, however, $\mathrm{g}_{\mathbf{Q}_{0,1},\mathbf{Q}_{0,2}}(t)$ are very close to being correlation matrices in most cases. We do not have any counterexamples.

We conclude this section by giving realisations of the four major prior models we have discussed. In Figure 2, no prior information about the geometry of the subsurface can be included. In Figure 3, geometric information has been incorporated, but no change in the correlation between the parameters in space can be included. In Figure 4, an example realisation from the full model is given. Pay attention to the rightmost field – here the correlation to the other two fields changes from being positive in the top layer to being negative in the bottom layer.

Fig. 2 Stationary model given by (3)

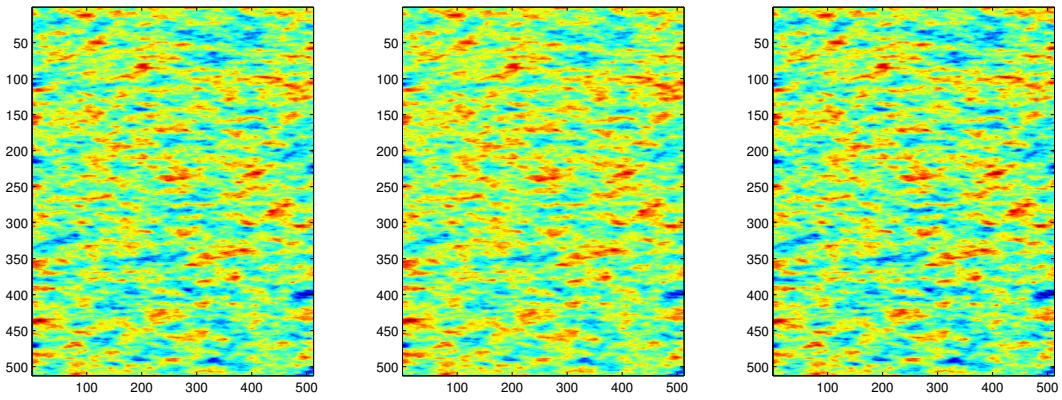


Fig. 3 Nonstationary model with fixed $\mathbf{Q}_0$

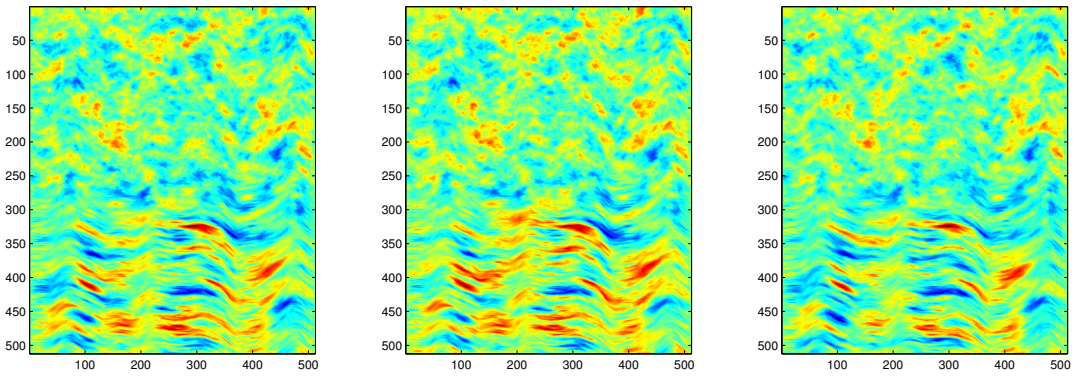
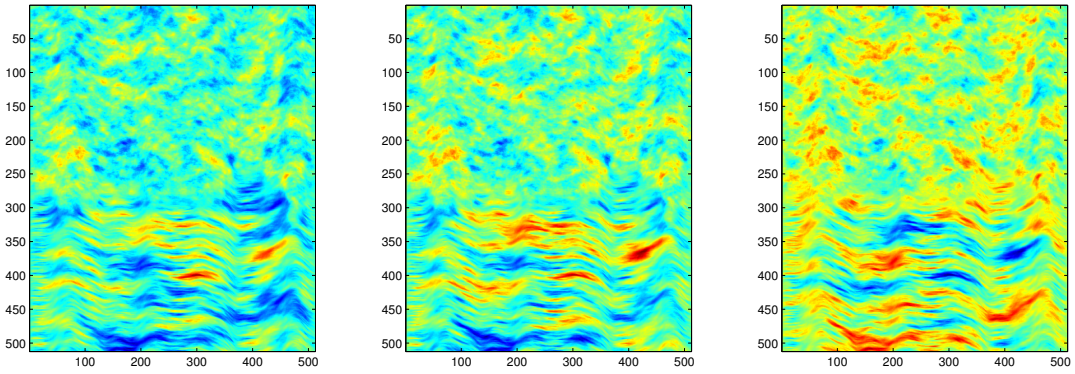


Fig. 4 Full nonstationary model with varying $q_{ij}(\boldsymbol{s})$ according to (16)



4 Conclusions and future work

In this text we have showed two things: First, how it is possible to incorporate information about the geometry of the subsurface. Secondly, how to facilitate changing covariance between elastic parameters depending on position. The first hinges on using SPDEs in order to specify local properties of the fields, and the second on how systems of SPDEs interrelate depending on position. This approach is not limited to the relatively simple models described here - rather, it may be used in any geological inversion problem with a natural geometry where soft constraints based on expert opinion may be used.

For future work, the major item is parameter estimation. This entails estimating the parameters $\boldsymbol{m}$ using both our approach and the stationary

and seeing how this affects the estimates. In a future work, this text will be expanded upon to include such estimates.

Appendix: Finite difference disretization - the gory details

This appendix is devoted to the finite difference scheme we used for discretizing the elliptic operator in (7). We employ a changed notation in this appendix for convenience, and we hope that it is transparent for readers. For a 2-dimensional field with $\alpha = 1$, we have

$$\begin{aligned}
&\nabla \cdot \begin{pmatrix} a_{11}(x,y)\ a_{12}(x,y) \\ a_{21}(x,y)\ a_{22}(x,y) \end{pmatrix} \begin{pmatrix} u_x(x,y) \\ u_y(x,y) \end{pmatrix} + \kappa(x,y)u(x,y) \\
=&\nabla \cdot \begin{pmatrix} a_{11}(x,y)u_x(x,y) + a_{1,2}(x,y)u_y(x,y) \\ a_{21}(x,y)u_x(x,y) + a_{22}(x,y)u_y(x,y) \end{pmatrix} + \kappa(x,y)u(x,y) \\
=&\partial_x(a_{11}u_x + a_{12}u_y) + \partial_y(a_{21}u_x + a_{22}u_y) + \kappa u \\
=&a_{11}^x u_x + a_{11}u_{xx} + a_{12}^x u_y + a_{12}u_{yx} + a_{21}^y u_x + a_{21}u_{xy} + a_{22}^y u_y + a_{22}u_{yy} \\
=&\mathrm{diag}(A)\nabla \cdot \nabla u + (a_{12} + a_{21})u_{xy} + a_{11}^x u_x + a_{12}^x u_y + a_{22}^y u_y + a_{21}^y u_x \qquad (18)
\end{aligned}$$

where $a_{ij}^v, v = x, y$ denotes differentation wrt. x or y of the i, j element of A, depending implicitly on the position. To discretize (18), we employ a finite difference scheme. We define the following finite difference operators

$$\begin{aligned}
\delta_x u &= \frac{1}{h}(u_{i+1}^j - u_i^j) \\
\delta_{\hat{x}} &= \frac{1}{h}(u_i^j - u_{i-1}^j),
\end{aligned}$$

where i, j are positions on the grid, with i denoting the x-direction and j denoting the y-direction. Now, we define the following operators

$$\begin{aligned}
\Lambda_{xx}u = \delta_x\left(\alpha_{11}\delta_{\hat{x}}u\right) &= \delta_x\left(\frac{1}{h}\alpha_{11}\left(u_i^j - u_{i-1}^j\right)\right) \\
&= \frac{1}{h^2}\left(\alpha_{11}^{i+1,j}\left(u_{i+1}^j - u_i^j\right) - \alpha_{11}^{i,j}\left(u_i^j - u_{i-1}^j\right)\right), \qquad (19)
\end{aligned}$$

where

$$\begin{aligned}
\alpha_{11}^{i,j} &= \frac{1}{2}\left(a_{11}^{i,j} + a_{11}^{i-1,j}\right) \\
\alpha_{22}^{i,j} &= \frac{1}{2}\left(a_{22}^{i,j} + a_{22}^{i,j-1}\right).
\end{aligned}$$

A equivalent expression holds for $\Lambda_{yy}u$. We define $\alpha_{kk}^{1,1} = a_{11}^{1,1}, k = 1, 2$. For the mixed operators we have

$$\Lambda_{xy}^{+}u = \frac{1}{2}\left(\delta_x\left(a_{12}\delta_y u\right) + \delta_{\hat{x}}\left(a_{12}\delta_{\hat{y}}u\right)\right) \tag{20}$$

and we have

$$\begin{aligned}
\delta_x\left(a_{12}\delta_y u\right) &= \frac{1}{h}\delta_x\left(a_{12}u_i^{j+1} - u_i^j\right) \\
&= \frac{1}{h^2}\left(a_{12}^{i+1,j}(u_{i+1}^{j+1} - u_{i+1}^j) - a_{12}^{i,j}(u_i^{j+1} - u_i^j)\right) \\
\delta_{\hat{x}}\left(a_{12}\delta_{\hat{y}}u\right) &= \frac{1}{h}\delta_{\hat{x}}\left(a_{12}(u_i^j - u_i^{j-1})\right) \\
&= \frac{1}{h^2}\left(a_{12}^{i,j}(u_i^j - u_i^{j-1}) - a_{12}^{i-1,j}(u_{i-1}^j - u_{i-1}^{j-1})\right).
\end{aligned}$$

Hence

$$\begin{aligned}
\Lambda_{xy}^{+}u = &\frac{1}{2h^2}\Big(\left(a_{12}^{i+1,j}(u_{i+1}^{j+1} - u_{i+1}^j) - a_{12}^{i,j}(u_i^{j+1} - u_i^j)\right) \\
&+ \left(a_{12}^{i,j}(u_i^j - u_i^{j-1}) - a_{12}^{i-1,j}(u_{i-1}^j - u_{i-1}^{j-1})\right)\Big)
\end{aligned} \tag{21}$$

For Λ_{yx}^{+} we reverse the order of the difference operators:

$$\begin{aligned}
\Lambda_{yx}^{+}u &= \frac{1}{2}(\delta_y(a_{12}\delta_x u) + \delta_{\hat{y}}(a_{12}\delta_{\hat{x}})) \\
&= \frac{1}{2h^2}\Big(\left(a_{12}^{i,j+1}(u_{i+1}^{j+1} - u_i^{j+1}) - a_{12}^{i,j}(u_{i+1}^j - u_i^j)\right) \\
&\quad + \left(a_{12}^{i,j}(u_i^j - u_{i-1}^j) - a_{12}^{i,j-1}(u_i^{j-1} - u_{i-1}^{j-1})\right)\Big)
\end{aligned}$$

And the complete discretisation is

$$(\Lambda_{xx} + \Lambda_{xy}^{+} + \Lambda_{yx}^{+} + \Lambda_{yy})u = f(u, W) \tag{22}$$

In Samarskii et al. (2002), it is proved that this scheme is convergent. If we assume that $\mathbf{A}$ does not vary in space, we can simplify the scheme;

$$
\begin{aligned}
\widehat{\Lambda_{xx}}u &= \frac{1}{h^2}(a_{11}(u_{i+1}^j - u_i^j) - a_{11}(u_i^j - u_{i-1}^j)) \\
&= \frac{1}{h^2}a_{11}(u_{i+1}^j - 2u_i^j + u_{i-1}^j) \\
\widehat{\Lambda_{yy}}u &= \frac{1}{h^2}a_{22}(u_i^{j+1} - 2u_i^j + u_i^{j-1}) \\
\widehat{\Lambda_{xy}^+}u &= \frac{1}{2h^2}\Big(\Big(a_{12}(u_{i+1}^{j+1} - u_{i+1}^j) - a_{12}(u_i^{j+1} - u_i^j)\Big) \\
&\quad + \Big(a_{12}(u_i^j - u_i^{j-1}) - a_{12}(u_{i-1}^j - u_{i-1}^{j-1})\Big)\Big) \\
&= \frac{a_{12}}{2h^2}\left(2u_i^j + u_{i+1}^{j+1} + u_{i-1}^{j-1} - u_{i+1}^j - u_i^{j+1} - u_i^{j-1} - u_{i-1}^j\right) \\
\widehat{\Lambda_{yx}^+}u &= \frac{1}{2h^2}\Big(\Big(a_{12}(u_{i+1}^{j+1} - u_i^{j+1}) - a_{12}(u_{i+1}^j - u_i^j)\Big) \\
&\quad + \Big(a_{12}(u_i^j - u_{i-1}^j) - a_{12}(u_i^{j-1} - u_{i-1}^{j-1})\Big)\Big) \\
&= \frac{a_{12}}{2h^2}\left(2u_i^j + u_{i+1}^{j+1} + u_{i-1}^{j-1} - u_{i+1}^j - u_i^{j+1} - u_i^{j-1} - u_{i-1}^j\right) \\
\left(\widehat{\Lambda_{xy}^+} + \widehat{\Lambda_{yx}^+}\right)u &= \frac{a_{12}}{h^2}\left(2u_i^j + u_{i+1}^{j+1} + u_{i-1}^{j-1} - u_{i+1}^j - u_i^{j+1} - u_i^{j-1} - u_{i-1}^j\right)
\end{aligned}
$$

This corresponds to the following stencil

$$
S = -\frac{1}{h^2}\begin{pmatrix} a_{12} & -a_{22} - a_{12} & 0 \\ -a_{11} - a_{12} & 2(a_{11} + a_{22} + a_{12}) & -a_{11} - a_{12} \\ 0 & -a_{22} - a_{12} & a_{12} \end{pmatrix} \tag{23}
$$

References

Aki, K. and Richards, P. G. (1980). *Quantitative Seismology*. WH Freeman and Co.

Bhatia, R. (2007). *Positive definite matrices*. Princeton Univ Press.

Buland, A., Kolbjørnsen, O., and Omre, H. (2003). Rapid spatially coupled AVO inversion in the Fourier domain. *Geophysics*, 68:824–836.

Buland, A. and Omre, H. (2003). Bayesian linearized AVO inversion. *Geophysics*, 68:185–198.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley.

Feichtinger, H., Helffer, B., Lamoureux, M., Lerner, N., Morel, J., Rodino, L., Takens, F., Teissier, B., Toft, J., and Wong, M. (2008). *Pseudo-Differential Operators: Quantization and Signals*. Springer-Verlag Berlin Heidelberg.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177.

Gray, R. (2006). *Toeplitz and circulant matrices: A review*. E-book.

Hiai, F. and Petz, D. (2009). Riemannian metrics on positive definite matrices related to means. *Linear Algebra and Its Applications*, 430(11-12):3105–3130.

Hu, X., Simpson, D. P., Lindgren, F., and Rue, H. (2012). Multivariate gaussian random fields using stochastic partial differential equations. *N/A*.

Lindgren, F., Lindstrøm, J., and Rue, H. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. *Journal of the Royal Statistical Society, Series B*, 73:423–498.

Matérn, B. (1960). Spatial variation. *Meddelanden fran Statens Skogsforskningsinstitut (Estocolmo)*, 49.

Ohara, A., Suda, N., and Amari, S. (1996). Dualistic differential geometry of positive definite matrices and its applications to related problems. *Linear Algebra and Its Applications*, 247:31–53.

Rabben, T. E., Ursin, B., and Tjelmeland, H. (2008). Non-linear Bayesian joint inversion of seismic reflection coefficients. *Geophysical journal international*, 173:265–280.

Rozanov, J. (1977). Markov random fields and stochastic partial differential equations. *Mathematics of the USSR-Sbornik*, 32:515.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*. Chapman & Hall.

Ruzhansky, M. and Turunen, V. (2009). *Pseudo-differential Operators and Symmetries: Background Analysis and Advanced Topics*. Birkhauser.

Samarskii, A. A., Matus, P. P., Mazhukin, V. I., and Mozolevski, I. E. (2002). Monotone Difference Schemes for Equations with Mixed Derivatives. *Computers and Mathematics*, 44:501 – 510.

Stovas, A. and Ursin, B. (2003). Reflection and transmission responses of layered transversely isotropic viscoelastic media. *Geophysical Prospecting*, 51:447–477.

Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Inst. Internat. Statist*, 40:974–994.