NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Directional Metropolis–Hastings updates for posteriors with nonlinear likelihoods

by

Håkon Tjelmeland and Jo Eidsvik

PREPRINT
STATISTICS NO. 5/2004

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL
http://www.math.ntnu.no/preprint/statistics/2004/S5-2004.ps
Håkon Tjelmeland has homepage: http://www.math.ntnu.no/∼haakont
E-mail: haakont@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

DIRECTIONAL METROPOLIS–HASTINGS UPDATES FOR POSTERIORS WITH NONLINEAR LIKELIHOODS

HÅKON TJELMELAND
Department of Mathematical Sciences, NTNU, 7491 Trondheim, Norway
JO EIDSVIK
Statoil Research Center, 7005 Trondheim, Norway

Abstract. In this paper we consider spatial problems modeled by a Gaussian random field prior density and a nonlinear likelihood function linking the hidden variables to the observed data. We define a directional block Metropolis–Hastings algorithm to explore the posterior density. The method is applied to seismic data from the North Sea. Based on our results we believe it is important to assess the actual posterior in order to understand possible shortcomings of linear approximations.

1 Introduction

Several applications in the earth sciences are preferably formulated by an underlying hidden variable which is indirectly observed via noisy measurements. Examples include seismic data, production data and well data in petroleum exploration: In seismic data the travel times and amplitude measurements are nonlinearly connected to the elastic parameters of the subsurface, see e.g. Sheriff and Geldart (1995). Production data contain the history of produced oil and gas, which is a very complex functional of the porosity and permeability properties in the reservoir, see e.g. Hegstad and Omre (2001). Well data of radioactivity counts need to be transformed into more useful information, such as clay content in the rocks, see e.g. Bassiouni (1994). The Bayesian framework is a natural approach to infer the hidden variable; this entails a prior model for the variables of interest and a likelihood function tying these variables to observations.

In this paper we consider Gaussian priors for the underlying spatial variable, and nonlinear likelihood models. When using a nonlinear likelihood function, the posterior is not analytically available. However, the posterior can be explored by Markov chain Monte Carlo sampling (see e.g. Robert and Casella (1999) and Liu (2001)), with the Metropolis–Hastings (MH) algorithm as a special case. We describe a directional MH algorithm in this paper, see Eidsvik and Tjelmeland (2003). We use this algorithm to update blocks of the spatial variable at each MH

iteration. MH block sampling is a trade off between single site sampling schemes and updating the entire field. We show results of our modeling procedures for seismic data from a North Sea petroleum reservoir.

2 Methods

In this section we present our assumptions about the Bayesian model and our methods for analyzing the posterior.

2.1 PRIOR AND LIKELIHOOD ASSUMPTIONS

The variable of interest is denoted $x = \{x_{ij} \in \mathcal{R}; i = 1, \ldots, n; j = 1, \ldots, m\}$, a spatial random field in two dimensions represented on a grid of size $n \times m$. We denote its probability density by $\pi(x) = \pi(x|\theta) = N(x; \mu(\theta), \Sigma(\theta))$, where $N(x; \mu, \Sigma)$ denotes a Gaussian density evaluated in x, with fixed mean μ and covariance matrix Σ. For generality we condition on hyperparameters θ, but in this study we treat θ as fixed parameters. The generalization to a vector variable, $x_{ij} \in \mathcal{R}^d$, is straightforward. For the application in Section 3 we have $x_{ij} \in \mathcal{R}^3$. A three dimensional grid, x_{ijk}, is of course also possible.

We assume the spatial variable x to be stationary and let the field be defined on a torus, see e.g. Cressie (1991). As explained below, this has important computational advantages, but the torus assumption also implies that one should not trust the results close to the boundary of the grid. Thus, one should let the grid cover a somewhat larger area than what is of interest.

The likelihood model for the data $z = \{z_{ij} \in \mathcal{R}; i = 1, \ldots, n; j = 1, \ldots, m\}$, given the underlying variable x, is represented by the conditional density $\pi(z|x) = N(z; g(x), S)$, where $g(x)$ is a nonlinear function. Hence, the conditional expectation of the data has a nonlinear conditioning to the underlying field. We assume that the likelihood noise is stationary with covariance matrix S. It is again straightforward to extend this model to vector variables at each location, $z_{ij} \in \mathcal{R}^d$, or three dimensional grids. For the application in Section 3 we have $z_{ij} \in \mathcal{R}^2$. We assume that a linearized version of the likelihood is available, and denote this by the conditional density $\pi^{lin}_{x_0}(z|x) = N(z; G_{x_0}x, S)$, where x_0 is the value of x used in the linearization.

The posterior of the hidden variable x conditional on the data is given by

$$\pi(x|z) \propto \pi(x)\pi(z|x), \tag{1}$$

an analytically intractable posterior. The linearized alternative;

$$\pi^{lin}_{x_0}(x|z) \propto \pi(x)\pi^{lin}_{x_0}(z|x), \tag{2}$$

for fixed x_0, can be written in a closed form, and is possible to evaluate and sample from directly. But note that in general this becomes computationally expensive in high dimensions. With our torus assumption discussed above, the

covariance matrices involved become circular and the linearized posterior can then be evaluated and sampled from effectively in the Fourier domain (Cressie (1991), Buland, Kolbjørnsen, and Omre (2003)). The actual nonlinear posterior can also be evaluated, up to a normalizing constant, in the Fourier domain, by treating the prior and likelihood terms in equation (1) separately.

2.2 METROPOLIS–HASTINGS BLOCK UPDATING

A MH algorithm is an iterative sampling method for simulating a Markov chain that converges to a desired posterior distribution, see e.g. Robert and Casella (1999) and Liu (2001). Each iteration of the standard MH algorithm consists of two steps: (i) Propose a new value for the underlying variable, (ii) Accept the new value with a certain probability, else keep the value from the previous iteration.

We describe a MH algorithm which updates blocks of the random field at each iteration. Let $X^i = x$ denote the variable after the i-th iteration of the MH algorithm. For the $(i+1)$-th iteration we draw a block of fixed size $k \times l$ at random, where $k < n$, $l < m$. Since the grid is on a torus, there are no edge problems when generating this block. We denote the block by A, a defined boundary zone of the block by B, and the set of nodes outside the block and boundary by C, see Figure 1. Further, we split the variable into these blocks; $x = (x_A, x_B, x_C)$ as the parts in

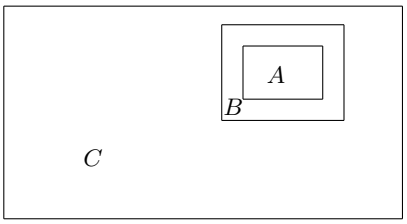

Figure 1. Block updating. The full size of the grid is $n \times m$. We illustrate the block A of gridsize $k \times l$, a boundary zone B, and the other parts of the field C.

the block, the boundary, and outside the block and boundary zone, respectively. Correspondingly, we denote the data vector by $z = (z_A, z_B, z_C)$. To propose a new value in the block we define a proposal density for the part in A, and denote the proposal on this block by y_A. The rest of x remains unchanged, and hence the proposed value is $y = (y_A, x_B, x_C)$. One proposal density in the block is the linearized posterior for y_A, conditional only on values in the boundary zone and data in A and B. We denote this by $\pi_x^{lin}(y_A|x_B, z_A, z_B) = N(y_A; m, T)$, where the mean m and covariance T can be calculated directly (Cressie, 1991). Note that the linearization is done at the current value $X^i = x$. The final step in the MH iteration is to accept or reject the proposed value y, and we thus obtain the next state X^{i+1}.

2.3 DIRECTIONAL METROPOLIS–HASTINGS

Directional MH algorithms are a special class of MH algorithms. Each iteration consists of the following steps; (i1) Generate an auxiliary random variable which defines a direction, (i2) Draw a proposed value on the line specified by the current state and the auxiliary direction, (ii) Accept the proposed value with a certain probability, else keep the variable from the previous iteration.

We present a directional updating scheme where the proposal step is done effectively in one dimension, see Eidsvik and Tjelmeland (2003). We outline our method following the block sampling setting in Section 2.2. Denote again the variable at the i-th iteration by $X^i = x = (x_A, x_B, x_C)$, and the data by $z = (z_A, z_B, z_C)$. We generate an auxiliary direction (step i1) as follows; First, draw w_A from $\pi_x^{lin}(\cdot|x_B, z_A, z_B)$. Next, define the auxiliary direction as $u = \pm\frac{w_A - x_A}{|w_A - x_A|}$, where we use $+$ or $-$ so that the first component of u is positive, see the discussion in Eidsvik and Tjelmeland (2003). Since this density for w_A is Gaussian, it is possible to calculate the density for the auxiliary unit direction vector u (Pukkila and Rao, 1988). We denote this density by $g(u|x_A, x_B, z_A, z_B)$.

At the last part of the proposal step (i2) we draw a one dimensional value t from some density $q(t|u, x, z)$, and set $y_A = x_A + tu$ as the proposed value for the block, and $y = (y_A, x_B, x_C)$ as the proposal for the entire field. This proposal is accepted, i.e. $X^{i+1} = y$, with probability

$$r(y|x) = \min\left\{1, \frac{\pi(y)}{\pi(x)} \cdot \frac{\pi(z|y)}{\pi(x|y)} \cdot \frac{g(u|y_A, x_B, z_A, z_B)}{g(u|x_A, x_B, z_A, z_B)} \cdot \frac{q(-t|u, y, z)}{q(t|u, x, z)}\right\}, \qquad (3)$$

else we have $X^{i+1} = x$.

In particular, if we choose the one dimensional density as

$$q^\star(t|u, x, z) \propto \pi(z|x_A + tu, x_B, x_C)\pi(x_A + tu, x_B, x_C)g(u|x_A + tu, x_B, z_A, z_B), \quad (4)$$

the acceptance probability in equation (3) is equal to unity, and hence the proposed variable is always accepted. Two elements are important when considering the unit acceptance rate proposal in equation (4); (i) An acceptance rate of one is not necessarily advantageous in MH algorithms. It is advantageous to obtain fast mixing, i.e. to have a small autocorrelation between successive variables. This is best achieved with large moves at each iteration of the MH algorithm. (ii) It is not possible to sample directly from the $q^\star$ density. To obtain a sample we have to fit an approximation to $q^\star$ in some way, either by a parametric curve fit, or by numerical smoothing of a coarse grid approximation.

In this paper we use an alternative density which does not give unit acceptance rate. The adjusted density is

$$\widetilde{q}(t|u, x, z) \propto (1 + |t|)^\lambda q^\star(t|u, x, z), \qquad \lambda > 0, \qquad (5)$$

where the $(1+|t|)^\lambda$ term encorages t values away from $t = 0$. This makes it possible to have larger moves in the MH algorithm since $t = 0$ corresponds to $y_A = x_A$. We fit an approximation by first calculating $\widetilde{q}$ on a coarse grid and then use linear

interpolation in the log scale between the evaluated grid points. The approximation that we obtain with this approach is our proposal denoted q.

In Figure 2 we show the proposal with acceptance one, $q^\star$, the adjusted proposal

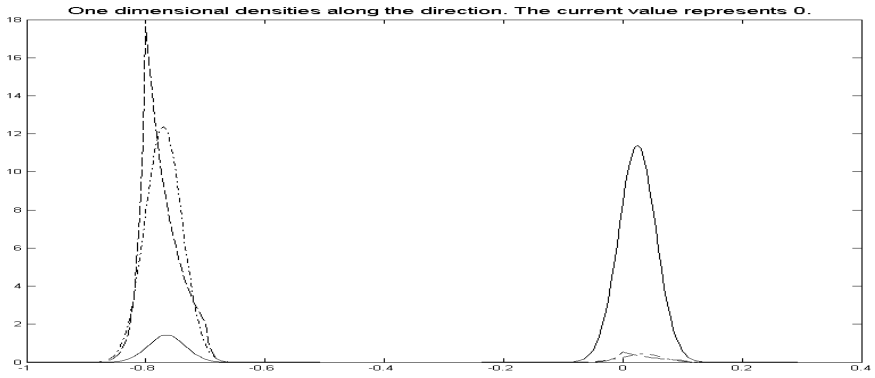

Figure 2. Sketch of the density $q^\star$(solid), the adjusted density $\widetilde{q}$ (dash-dots), and the fitted density q (dashed). This is a typical proposal density for t in our application in Section 3. The approximation q has an exponential form in each interval of length 0.1 on a grid (for example between -0.8 and -0.7).

$\widetilde{q}$ and its fitted proposal q. This particular plot is obtained in our application in Section 3. We tuned λ in equation (5) so that the acceptance rate was about 0.5 in our application. This results in a value of $\lambda = 10$. Note that the proposal density for t has a bimodal shape. One mode is usually close to $t = 0$, while the other mode is quite far from 0. The mixing of the MH algorithm improves if we reach the mode away from 0 more often. This can be established by $\widetilde{q}$ or q as shown in Figure 2. In the case of a linear likelihood the two modes illustrated in Figure 2 are always of equal size. A nonlinear likelihood causes them to have unequal mass, and most commonly the mode near $t = 0$ contains most of the probability mass.

3 Example

We apply our modeling procedures to seismic data from the North Sea. We first present the general motivation, then the statistical model formulated in the setting of Section 2, and finally the results.

3.1 SEISMIC AMPLITUDE VERSUS OFFSET DATA

Seismic amplitude versus offset (AVO) analysis is commonly used to assess the underlying lithologies (rocks and saturations) in a petroleum reservoir, see e.g. Sheriff and Geldart (1995), and Mavko, Mukerji, and Dvorkin (1998). The reflection amplitude of seismic data changes as a function of incidence angle and as a function of the elastic properties which are indicative of lithologies.

We analyze reflection data at two incidence angles in the Glitne field, North Sea. The Glitne field is an oil-producing turbidite reservoir with heterogeneous sand and shale facies, see Avseth et al (2001) and Eidsvik et al (2004). The domain of our interest is $2.5 \times 2.5\ km^2$, and it is split into a grid of size 100×100, with each grid cell covering $25 \times 25\ m^2$. The area covers what was interpreted as the lobe of the turbidite structure in Avseth et al (2001). Figure 3 shows the reflection amplitude along the grid at reflection angles zero (left) and thirty (right).

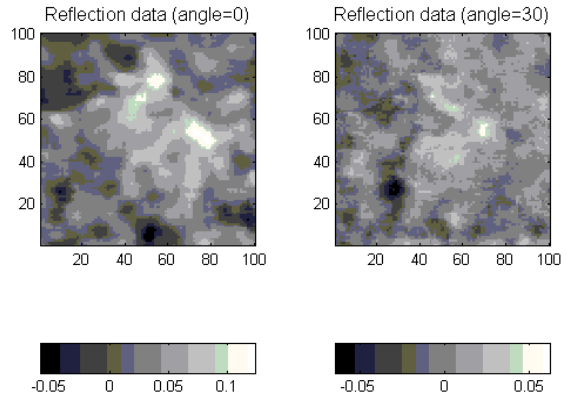

Figure 3. Reflection data at the 2D interface. Incidence angle 0 degrees (left) and 30 degrees (right).

The reflection amplitudes are hard to analyze directly because they are a result of the contrast in elastic properties in the cap rock covering the reservoir and the properties in the reservoir zone. We next use a statistical model for automatic analysis of the elastic reservoir properties.

3.2 STATISTICAL MODEL FOR SEISMIC DATA

The seismic data for the two reflection angles are denoted $z = (z^0, z^1)$, where z^0 refers to the zero offset reflection and z^1 to 30 degrees incidence angle (both are plotted in Figure 3). The statistical model that we use is closely connected to the one in Buland, Kolbjørnsen and Omre (2003).

The variables of interest are the pressure and shear wave velocities, and the density in the reservoir zone. We denote these by $x = (\alpha, \beta, \rho) = \{x_{ij} = (\alpha_{ij}, \beta_{ij}, \rho_{ij}); i = 1, \ldots, n; j = 1, \ldots, m\}$, where α is the logarithm of the pressure wave velocity, β is the logarithm of the shear wave velocity, and ρ is the logarithm of the density. The velocities and density are the elastic properties of the rocks which in some sense capture the rock mineral and saturation (Mavko, Mukerji, and Dvorkin, 1998). The units for the exponential equivalents of α, β and ρ are m/s for velocities and kg/m^3 for density. Let $(\alpha_0, \beta_0, \rho_0)$ be the logarithm of pressure and shear wave

velocities, and the logarithm of density for the cap rock. These cap rock properties are treated as fixed values in this study, and are equal for all locations in the grid.

We assign a Gaussian prior density to the reservoir variables of interest, i.e. $\pi(x) = N(x; \mu, \Sigma)$, where μ now becomes a $3mn$ vector with the mean of the three log elastic properties. These prior mean values are fixed, and set to $\mu_\alpha = E(\alpha_{ij}) = 7.86$, $\mu_\beta = E(\beta_{ij}) = 7.09$, $\mu_\rho = E(\rho_{ij}) = 7.67$, for all (i, j). These mean values are assessed from well logs in Glitne, see Avseth et al (2001). The prior covariance matrix, Σ, is a $3mn \times 3mn$ matrix defined by a Kronecker product, giving a 3×3 block covariance matrix at the diagonal, and 3×3 matrices with this covariance function and a spatial correlation function on the off-diagonal, see Buland, Kolbjørnsen, and Omre (2003). The diagonal covariance matrix describing the marginal variability at each location is defined by $Std(\alpha_{ij}) = 0.06$, $Std(\beta_{ij}) = 0.11$, $Std(\rho_{ij}) = 0.02$, and correlations $Corr(\alpha_{ij}, \beta_{ij}) = 0.6$, $Corr(\alpha_{ij}, \rho_{ij}) = 0.1$, $Corr(\beta_{ij}, \rho_{ij}) = -0.1$. These parameters capture the variability expected from previous studies, see Buland, Kolbjørnsen, and Omre (2003). The spatial correlation function is the same for all three reservoir variables and is an isotropic exponential correlation function with range $250m$ (10 grid nodes).

The likelihood function is nonlinear, and is defined by approximations to the Zoeppritz equations, see e.g. Sheriff and Geldart (1995). The density for the seismic AVO data, given the underlying reservoir properties, is $\pi(z|x) = N(z; g(x), S)$, where the nonlinear function goes only locationwise, i.e. at grid node (i, j) the expectation term in the likelihood is a function of the variables at this gridnode only. For each location (i, j) and angle $\gamma = 0$, 30 we have

$$g_{ij,\gamma}(x) = g_{ij,\gamma}(\alpha_{ij}, \beta_{ij}, \rho_{ij}) = a_0(\alpha_{ij} - \alpha_0) + a_{1,ij}(\beta_{ij} - \beta_0) + a_{2,ij}(\rho_{ij} - \rho_0), \quad (6)$$

where

$$\begin{aligned} a_0 &= \frac{1}{2}[1 + sin^2(\gamma)], \qquad a_{1,ij} = -4\xi_{ij} sin^2(\gamma), \\ a_{2,ij} &= \frac{1}{2}[1 - 4\xi_{ij} sin^2(\gamma)], \qquad \xi_{ij} = \frac{\exp(2\beta_{ij}) + \exp(2\beta_0)}{\exp(2\alpha_{ij}) + \exp(2\alpha_0)}. \end{aligned} \quad (7)$$

The noise covariance matrix of the likelihood, S, is a $2mn \times 2mn$ matrix defined from a Kronecker product. This covariance matrix has a block diagonal 2×2 matrix on the diagonal, and off-diagonal elements defined from an exponential correlation structure with range $250m$. The diagonal noise covariance matrix is defined by $Std(\gamma = 0) = 0.015$, $Std(\gamma = 30) = 0.012$, $Corr(\gamma = 0, \gamma = 30) = 0.7$. This likelihood noise model is specified using the parameters in Buland, Kolbjørnsen, and Omre (2003).

A linear likelihood model can be defined by fixing the ratio ξ_{ij} in equation (7). For a constant linearization point we have $\pi_\mu^{lin}(z|x) = N(z; G_\mu x, S)$. A similar linearization is used in Buland, Kolbjørnsen, and Omre (2003), and with this linearization they assess the analytically available posterior directly. For a linearized proposal density on the block A we have $\pi_x^{lin}(\cdot|x_B, z_A, z_B)$. For our block sampling algorithm we use a block size of 9×9, and a boundary zone of width one grid node.

3.3 RESULTS

The posterior is sampled using the block directional MH algorithm discussed above. We denote 144 updates as one iteration, i.e. on average each grid node is (proposed) updated about once in each iteration. Figure 4 shows trace plots for

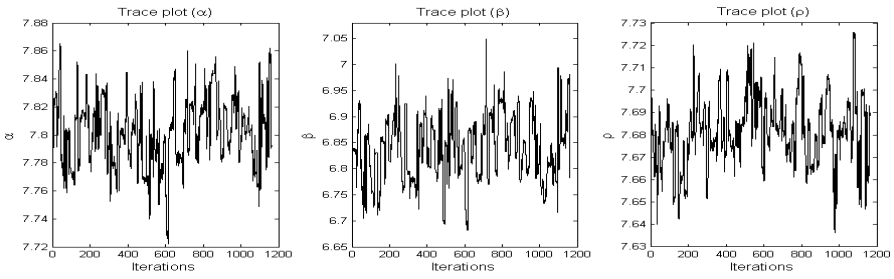

Figure 4. Trace plots of the three variables at one location in the grid. Log of pressure wave velocity α (left), log of shear wave velocity β (middle), and log of density (right).

the log elastic properties at one location in the grid. The traceplots are explained to some extent by the bimodal proposal density q (see Figure 2). In Figure 4 the variables move short distances at some iterations, while moves are large at other iterations, reflecting the bimodal density for the proposal.

In Figure 5 we show the estimates of the marginal mean and standard deviation for all three variables as images. Near grid coordinate (North,East) $= (60, 80)$ (see Figure 5, top left image, water sand arrow) both pressure and shear wave velocities are large. In the same area the reflection data (Figure 2) are large at both angles. Going south from gridnode $(60, 80)$ (see Figure 5, top left image, oil sand arrow) the pressure wave velocity decreases, and so does the shear wave velocity, but to a smaller degree. In Figure 2 the reflection data become smaller in this area. These two regions comprise the lobe of the turbidite structure, see Avseth et al (2001). In Eidsvik et al (2004) these two regions were estimated to be water and oil saturated sands, respectively. Pressure wave velocity is larger in water than oil saturated sands (Mavko, Mukerji, and Dvorkin, 1998) and our interpreted velocities are in accordance with the results in Eidsvik et al (2004). The western part of the domain were predicted to contain mostly shales (a low velocity rock) in Eidsvik et al (2004).

The prior standard deviations for the three variables are $(0.06, 0.11, 0.02)$. In Figure 5 (right) the mean standard deviations in the posterior are $(0.033, 0.065, 0.02)$. This indicates that there is information about α and β in the AVO data (standard deviation decreases by a factor two), but not much about ρ.

Note that the standard deviations in Figure 5 (right) varies quite a lot across the field (a factor of two). The standard deviation for β is smaller where the velocities large. For a linear model [Buland, Kolbjørnsen, and Omre (2003)] the

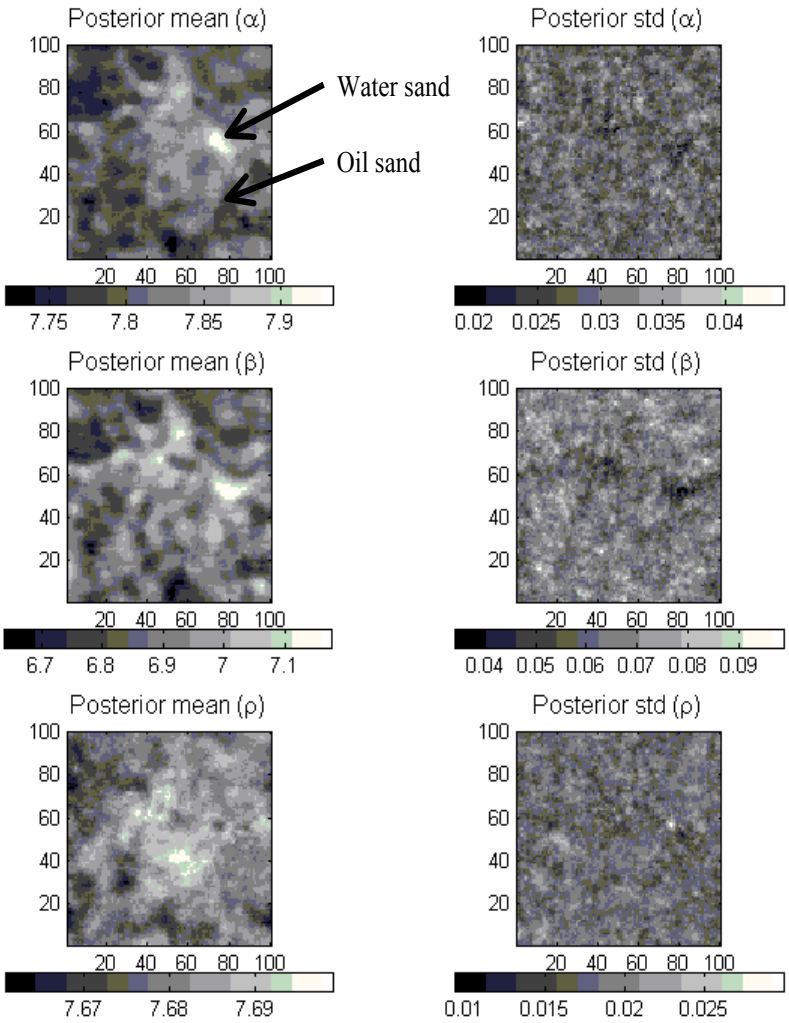

Figure 5. Mean and standard deviation of the three variables at each location in the grid. Log of pressure wave velocity α (top). Log of shear wave velocity β (middle), and log of density (bottom). The two regions indicated by arrows (upper left image) were interpreted to be oil and water saturated sands in a previous study.

standard deviations are constant across the field. The expected values also differ somewhat between a linear model and our nonlinear model; for example $E(\frac{\beta}{\alpha})$ is shifted significant between the two approaches. These differences suggest that

the linearized Gaussian posterior in Buland, Kolbjørnsen, and Omre (2003) might introduce a bias in the estimation of the elastic parameters.

4 Closing Remarks

In this paper we consider Bayesian models with a Gaussian random field prior and nonlinear likelihood functions. Such models are common in the earth sciences, but are usually simplified (linearized) to make the posterior analytically available. We propose a directional block Metropolis–Hastings sampler for exploring the original nonlinear posterior. When we apply our methods to a seismic dataset from the North Sea, we recognize some differences between our results and the ones obtained by a linearized model. These differences indicate that it is useful to check the validity of a simplified likelihood model by sampling the full nonlinear models.

One of the current challenges with the Glitne field is uncertainty in the thickness of the turbidite structure, associated with the noise in seismic data due to overburden effects. A natural extension is hence to study the full 3D seismic data. An extension of our statistical methods is to assign priors to the hyperparameters in the statistical model, and hence include the variability of these parameters into the final results.

References

Avseth, P., Mukerji, T., Jørstad, A., Mavko, G., and Veggeland, T., *Seismic reservoir mapping from 3-D AVO in a North Sea turbidite system*, Geophysics, vol. 66, no. 4, 2001, p. 1157-1173

Bassiouni, Z., *Theory, measurement, and interpretation of well logs*, Society of Petroleum Engineers, 1994

Buland, A., Kolbjørnsen, O., and Omre, H., *Rapid spatially coupled AVO inversion in the Fourier domain*, Geophysics, vol. 68, no. 3, 2003, p. 824-835

Cressie, N.A.C., *Statistics for spatial data*, Wiley, 1991

Eidsvik, J., and Tjelmeland, H., *On directional Metropolis–Hastings algorithms*, Submitted for publication, Available as a Technical Report, Department of Mathematical Sciences, Norwegian University of Science and technology, http://www.math.ntnu.no/preprint/statistics/2003/S6-2003.pdf

Eidsvik, J., Avseth, P., Omre, H., Mukerji, T., and Mavko, G., *Stochastic reservoir characterization using pre-stack seismic data*, Geophysics, v. 69, no. 4, 2004, p. 978-993

Hegstad, B. K., and Omre, H., *Uncertainty in Production Forecasts based on well observations, seismic data and production history*, Society of Petroleum Engineering Journal, December 2001, p. 409-424

Liu, J.S., *Monte Carlo Strategies in Scientific Computing*, Springer, 2001

Mavko, G., Mukerji, T., and Dvorkin, J., *The Rock Physics Handbook*, Cambridge, 1998

Pukkila, T.M., and Rao, C.R., *Pattern recognition based on scale invariant functions*, Information Sciences, v. 45, 1988, p. 379-389

Robert, C.P., and Casella, G., *Monte Carlo Statistical Methods*, Springer, 1999

Sheriff, R.E., and Geldart, L.P., *Exploration seismology*, Cambridge, 1995