Lithology/fluid prediction by simulating from the posterior of a Markov chain prior Bayesian model

Petroleum Geostatistics
Cascais Portugal

Hugo Hammer and Håkon Tjelmeland

10 – 14 September 2007

Abstract

We use a Markov chain prior Bayesian model to do lithology and fluid prediction from amplitude variations versus offset seismic data. We consider roughly the same model that is used in Larsen et al. (2006) which again is strongly influenced by the seismic inversion model found for example in Buland et al. (2003). We simulate from the resulting posterior distribution using a combination of the Metropolis–Hastings (MH) algorithm (Hastings, 1970) and the Gibbs sampler (Liu, 2001). We use the forward-backward algorithm (Liu, 2001) within the Metropolis–Hastings algorithm to update the elastic parameters and the lithology and fluid classes simultaneously. The result is an algorithm with very good convergence and mixing properties.

1 Introduction

Lithology and fluid prediction (LFP) from seismic data is a hard problem, actually an ill-posed inverse problem. In a Bayesian framework we combine available prior knowledge with measured data. This makes stable and reliable solutions possible. We also get the uncertainty of our solutions. Everything is summed up in the posterior distribution. The focus in this presentation is how to simulate from the posterior distribution.

We base our inversion on amplitude variations versus offset seismic data (AVO) similarly to Buland et al. (2003) and Larsen et al. (2006). We focus on predicting the Lithology and fluid classes (LF) along a vertical earth profile. The LF classes can for example be gas-, oil- or brine-saturated sandstone or shale. We consider roughly the same model that is used in Larsen et al. (2006) which again is strongly influenced by the seismic inversion model found for example in Buland et al. (2003).

2 Notation

Let $x = (x_1, x_2, \ldots, x_n)$ be LF classes for n positions along a vertical profile. We have in each position k possible LF classes, i.e. $x_i \in \{1, 2, \ldots, k\}$ for $i = 1, \ldots, n$. Further, $y \in \mathbb{R}^{n\times 3}$ are elastic parameters, where $y_i = (y_{i,1}, y_{i,2}, y_{i,3})$ represent the natural logarithm of P-wave velocity, S-wave velocity and density, respectively, at the location $i \in \{1, \ldots, n\}$. Let $r \in \mathbb{R}^{n\times n_\theta}$ and $z \in \mathbb{R}^{n\times n_\theta}$ be reflection coefficients and AVO seismic data for the n positions in n_θ observation angles, respectively. We further have that r_i and z_i denote the reflection coefficients and seismic data at position $i \in \{1, \ldots, n\}$ respectively. We also let r^j and z^j be the reflection coefficients and seismic data for angle $\theta_j,\ j \in \{1, \ldots, n_\theta\}$ for all the n positions.

3 The Bayesian model

We model the prior for x by a Markov chain, $\pi(x) = \pi(x_1)\pi(x_2|x_1)\cdots\pi(x_n|x_{n-1})$. Further we assume

$$y|x \sim \prod_{i=1}^{n} \pi(y_i|x_i),$$

where

$$\pi(y_i|x_i) = N_3(y_i^T; \mu(x_i), \Sigma(x_i)).$$

Here, $N_p(\chi; \mu, \Sigma)$ denotes the density of a p-dimensional multi Gaussian distribution with expectation μ and covariance matrix Σ. Thus, we assume that the elastic parameters for the different positions are conditionally independent given x. Further, we get the n_θ reflection coefficients

$$r_i^T = A\left(\frac{y_{i+1}^T - y_{i-1}^T}{2}\right)$$

for $i = \{1, 2, \ldots, n\}$. The matrix $A \in \mathbb{R}^{n_\theta\times 3}$ refer to the angle dependent linear weak contrast approximations (Buland et al., 2003).

To get the data z we suppose that there are two independent noise terms added to the reflection coefficients, the first part of the noise being smooth and of equal waveform as the signal, and the second part being white noise. i.e.

$$z^j = W_{\theta_j} r^j + \sigma_1 W_{\theta_j} \epsilon_1^j + \sigma_2 \epsilon_2^j = W_{\theta_j}(r^j + \sigma_1 \epsilon_1^j) + \sigma_2 \epsilon_2^j, \tag{1}$$

where $\epsilon_1^j, \epsilon_2^j \sim N_n(e; 0, I)$, $W_{\theta_j} \in R^{n \times n}$ is a wavelet matrix for the angle θ_j for $j = 1, \ldots, n_\theta$. Thus,

$$z^j \sim N_n(z^j; W_{\theta_j} r^j, \sigma_1^2 W_{\theta_j} W'_{\theta_j} + \sigma_2^2 I).$$

We assume that the variance σ_2^2 is small compared to σ_1^2, so most of the noise in the data are of the same wave-form as the wavelet. This is also used in Buland et al. (2003) and Larsen et al. (2006). We are now mainly interested in simulating x from the resulting posterior distribution.

4 Simulation algorithm

We are now interested in extracting information from the posterior distribution presented in the previous section. We choose to do this by generating realizations from this distribution. We use Markov chain Monte Carlo (MCMC) simulation and more specifically a combination of the Metropolis–Hastings (MH) algorithm (Hastings, 1970) and the Gibbs sampler (Liu, 2001).

A natural way to try to simulate from the model is to use the Gibbs sampler where we alternate between simulating y from $\pi(y|x, z)$ being multi normal and simulating x from $\pi(x|y, z)$. We are able to simulate effectively from $\pi(x|y, z) = \pi(x|y)$ because of the conditional independence structure in the distribution using the forward-backward algorithm (Liu, 2001). Unfortunately, such an algorithm performs poorly because x and y are strongly correlated. We therefore consider another layer, $u \in \mathbb{R}^{n \times n_\theta}$, in the model for simulation purposes. We use the same notation for u that we used for r and z. We define $u^j = r^j + \sigma_1 \epsilon_1^j$, so that we have from (1) that $z^j = W u^j + \sigma_2 \epsilon_2$. The posterior distribution of interest is thereby

$$\pi(x, y, u|z) \propto \pi(x, y, u, z) = \pi(x)\pi(y|x)\pi(u|y)\pi(z|u), \tag{2}$$

where calculations give that

$$\pi(u|y) = \prod_{i=1}^{n} \pi(u_i|y_{i-1}, y_{i+1}) = \prod_{i=1}^{n} N_{n_\theta}\left(u_i; A\left(\frac{y_{i+1}^T - y_{i-1}^T}{2}\right), \sigma_1^2 I\right)$$

and

$$\pi(z|u) = \prod_{j=1}^{n_\theta} N_n(z^j; W_{\theta_j} u^j, \sigma_2^2 I).$$

We simulate from (2) by using two update steps in each iteration of the algorithm. More specifically, in the first step we generate new values for (y, u) from the full conditional Gaussian distribution $\pi(y, u|x, z) \propto \pi(x, y, u|z)$, so this is a Gibbs step. The second step is a MH step where we propose new values for (x, y) and accept or reject this proposal according to the MH acceptance probability. The proposal distribution is close to the full conditional $\pi(x, y|u, z) = \pi(x, y|u) \propto \pi(x, y, u)$, see the next section for more details.

The idea for introducing u is that (posterior) correlation between u and x is much smaller than between y and x. Thereby the MH algorithm including u will have much better convergence and mixing properties.

4.1 Details the proposal for (x,y)

It is not straight forward to generate good proposals for the distribution $\pi(x, y|u)$. Note, however, that because of the independence structure in $\pi(x, y|u, z)$, we may write $\pi(x, y|u) \propto h_1(x_1, y_1, x_2, y_2, y_3, u) \cdot h_2(x_2, y_2, x_3, y_3, u) \cdot \ldots \cdot h_{n-1}(x_{n-1}, y_{n-1}, x_n, y_n, u) \cdot h_n(x_n, y_n, u)$ for some functions $h_i(\cdot, \ldots, \cdot), i = 1, \ldots, n-1$. This is in fact a chain structured distribution and a method similar to the forward-backward algorithm (Liu, 2001) can be constructed. We are then in principle able to calculate analytically all the conditional distributions in

$$\begin{aligned}\pi(x, y|u) = \\ \pi(y_1|x_1, x_2, y_2, y_3, u)\pi(x_1|x_2, y_2, y_3, u)\pi(y_2|x_2, x_3, y_3, y_4, u)\pi(x_2|x_3, y_3, y_4, u) \cdots \\ \cdots \pi(y_{n-1}|x_{n-1}, x_n, y_n, u)\pi(x_{n-1}|x_n, y_n, u)\pi(y_n|x_n, u)\pi(x_n|u).\end{aligned}$$

The $\pi(y_1|x_1, x_2, y_2, y_3, u)$ becomes a mixture of k normal densities, $\pi(y_2|x_2, x_3y_3, y_4, u)$ a mixture of k^2 normal densities, and so on. Ultimately, $\pi(y_n|x_n, u)$ becomes a mixture of k^n normal densities. Of course, we are not able to handle that many normal terms computationally. Our solution is to ignore the less important terms, i.e. terms where the constant in front of the normal density is small. Doing this trick result in a distribution that we are able to generate samples from fairly effectively for dimensions up to $n = 100 - 150$ on a modern computer. >From this distribution we propose new values for x and u, which in turn is accepted or rejected according to the corresponding MH acceptance probability,

5 Simulation example

We consider $n = 100$, $k = 4$ and $n_\theta = 5$. We use a wavelet of length approximately 60 positions and a signal-to-noise ratio of 1.2. For the remaining model parameters we adopt values similar to Larsen et al. (2006). We test the LF prediction method by first starting with some profile x_s generated from the Markov chain prior and generating synthetic data z from x_s using the model. Secondly, we generate samples from the posterior distribution using the simulation method presented above. Simulations show that the algorithm works very well in terms of convergence and mixing properties. In Figure 1 we have results from two simulations, one in each row. In the column to the left we have the simulated "true" x_s, the next three columns show three (independent) posterior samples of x, and in the plots to the right we show estimated marginal posterior probabilities for the different LF classes at each position. In the whole figure, the red, green, blue and black colours refer to gas-, oil-, brine-saturated sandstone and shale, respectively.

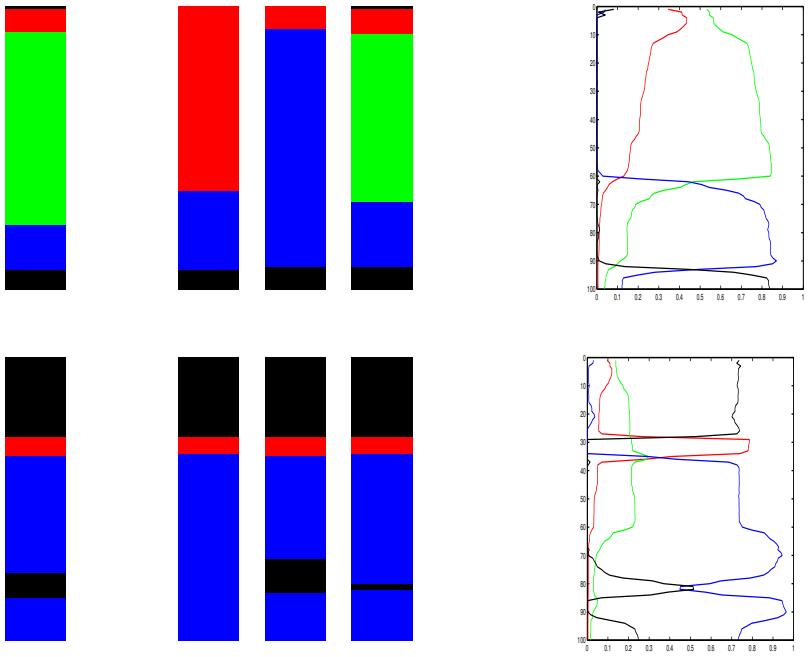

Figure 1: Results from two simulations, where the first row sum up the first simulation and the second row the second simulation. In the four columns to the left we have x_s and three samples from the posterior distribution and in the plots to the right we have estimated marginal posterior probabilities for the different LF classes at each position. In the whole figure, red, green, blue and black colour refer to gas-, oil-, brine-saturated sandstone and shale, respectively.

We see from the figure that the posterior distribution regain x_s in a good manner and also indicate the amount of uncertainty in the LF prediction.

6 Closing Remarks

We present an algorithm that simulates from the posterior distribution of a Markov chain prior Bayesian model. Simulations show that the algorithm converges fast and have good mixing properties. The realizations gives reliable and stable solutions to the problem. Future work is to extend the method to cope with longer profiles.

References

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

Hastings, W. (1970). Monte Carlo sampling using Markov chains and their applications, *Biometrica* **57**: 97–109.

Larsen, A. L., Ulvmoen, M., Omre, H. and Buland, A. (2006). Bayesian lithology/fluid prediction and simulation on the basis of a markov-chain prior model, *Geophysics* **71 issue 5**: R69–R78.

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