Empirical comparison of two Bayesian lithology–fluid prediction algorithms

Hugo Hammer and Marit Ulvmoen
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim, Norway

Abstract

We consider a Bayesian model for doing lithology–fluid prediction from prestack (amplitude versus offset) seismic data. Related to the Bayesian model, we look at two inversion algorithms. The first algorithm simulates from the posterior distribution with no approximations, but the algorithm is quite computer demanding. The second inversion algorithm introduces an approximation in the likelihood model and is in this way able to evaluate the resulting approximate posterior distribution very rapidly. The consequences of the approximation for the inversion result are not clear. The objective of this paper is to evaluate the consequences of the approximation in a synthetic but realistic empirical case study. The consequences are evaluated by comparing the inversion results from the two inversion algorithms. In the case study we observe that, dependent on the parameters in the model, typically the approximate likelihood model preserves between 55% and 80% of the information in the original likelihood model. The consequences of the approximation increase when the amount of noise in the model increases. The approximation works better when most of the variability is in the rock physics model and it is little seismic noise, compared to the opposite.

Keywords: amplitude versus offset seismic data, Bayesian model, empirical comparison, forward–backward algorithm, lithology–fluid prediction, seismic inversion

1 Introduction

In lithology–fluid (LF) prediction we want to determine the lithology and fluid properties of a reservoir using inversion techniques on seismic data. For an introduction to and presentation of several techniques for doing LF prediction we recommend Avseth et al. (2005) and references therein. We focus on LF prediction based on amplitude-versus-offset (AVO) seismic data in one vertical profile. Before the data are used for inversion, noise effects like moveout, multiples and the effect of geometrical spreading and absorption are removed. The data are also prestack migrated such that any dip related effects are removed. We focus on the Bayesian model presented in Larsen et al. (2006) and Hammer and Tjelmeland (2008). As a prior model for the LF properties a stationary Markov chain is used. Further, the model consists of a likelihood model that relates the LF properties to a set of elastic material properties and the AVO seismic data. The model is closely related to Buland and Omre (2003) and Buland et al. (2003).

To evaluate the model given in Larsen et al. (2006) and Hammer and Tjelmeland (2008) is a challenging task. Direct sampling from the posterior distribution is infeasible and single site Metropolis–Hastings algorithms give extremely slow convergence. Hammer and Tjelmeland (2008) simulate from the model using a more sophisticated block Metropolis–Hastings (MH) algorithm that proposes changes for all the variables in the model in each iteration. Larsen et al. (2006) introduce an approximation in the likelihood model and evaluate the resulting

approximate posterior exactly and rapidly. For example, the algorithm needs about a second to compute the approximate posterior distribution by recursion and generates therefore independent realizations every half millisecond. In comparison, the algorithm in Hammer and Tjelmeland (2008) needs hours to estimate properties like marginal distributions or maximum posterior value and generates independent realizations about every minute.

Central in both of the inversion algorithms is a Hidden Markov model (HMM). In a HMM we suppose that the observations are a noisy function of an underlying and unknown Markov process. In traditional HMM the underlying process is a Markov chain. Then the posterior is in fact a non-stationary Markov chain and can be evaluated analytically, exact and very efficiently using the Forward-backward (FB) algorithm (Scott, 2002). The FB algorithm consists of two parts. In the first part we iterate forward computing all the transition propabilities in the non-stationary Markov chain. In the second part we iterate backward evaluating the posterior distribution.

What consequences the approximation in Larsen et al. (2006) have on the inversion results are not clear. The objective of this paper is to evaluate these consequences in an empirical case study. Since the MH algorithm in Hammer and Tjelmeland (2008) simulates from the posterior distribution without approximations, we are able to analyse the consequence of the approximations by comparing the approximate inversion results with the exact inversion results.

The paper is organised as follows. In Section 2 we present the Bayesian model from Larsen et al. (2006) and Hammer and Tjelmeland (2008). In Section 3 we give a review of the inversion methods in the same two papers. Further, in Section 4 we discuss criteria to compare the inversion results and finally in Section 5 we analyse the consequence of the approximation in Larsen et al. (2006) in a synthetic but realistic case study.

2 Bayesian seismic model

The Bayesian seismic model relates the LF properties along a vertical profile with elastic material properties and observed AVO seismic data. The Bayesian model presented is identical to the one used in Hammer and Tjelmeland (2008) and only differs from Larsen et al. (2006) in the formulation of the rock physics model. We follow the notation in Larsen et al. (2006). We define a discretised Bayesian formulation. We use $i = 1, \ldots, n$ to denote n travel times along the vertical profile. Let π_i denote LF classes and let α_i, β_i and γ_i denote P-wave velocity, S-wave velocity and density, respectively, at position i. Further define $m_i = (m_{i1}, m_{i2}, m_{i3}) = (\ln(\alpha_i), \ln(\beta_i), \ln(\gamma_i))$. We consider s offset values $\theta_1, \ldots, \theta_s$ and let $c_i = (c_{i1}, \ldots, c_{is})^T \in \mathbb{R}^s$ and $d_i = (d_{i1}, \ldots, d_{is})^T \in \mathbb{R}^s$ denote reflection coefficients and seismic data, respectively, at position i for the s different offsets.

2.1 Prior model

As a prior model for the LF classes along the vertical profile, $\pi = (\pi_1, \ldots, \pi_n)^T$, we assume a stationary Markov chain

$$\mathrm{P} = [p(\pi_i|\pi_{i-1})]_{\pi_{i-1},\pi_i=1}^{L}. \tag{1}$$

Further, let $p(\pi_1)$ denote the marginal distribution for π_1.

2.2 Likelihood model

The distribution for $m = (m_1^T, \ldots, m_n^T)^T$ given π is related to a rock physics model (Mavko et al., 1998). We assume the relation to be Gaussian distributed

$$[m_i|\pi_i] \sim \mathrm{N}(m_i|\mu(\pi_i), \Sigma(\pi_i)), \tag{2}$$

where $\mathrm{N}(x|\mu, \Sigma)$ denotes a multivariate Gaussian distribution evaluated in x, with expectation μ and covariance matrix Σ. We assume $\mu(\pi_i)$ and $\Sigma(\pi_i)$ to be known and that $m_1, \ldots, m_n$ are conditionally independent given π.

Further, we assume that the reflection coefficients, $c = (c_1^T, \ldots, c_n^T)^T$, to be related to m through a weak contrast approximation of the Zoeppritz equations (Aki and Richards, 1980; Buland and Omre, 2003)

$$c_i = \Lambda^T \cdot \left(\frac{m_{i+1} - m_{i-1}}{2} \right) \text{ for } i = 2, \ldots, n-1, \tag{3}$$

where

$$\Lambda = \begin{bmatrix} a_\alpha(\theta_1) & a_\alpha(\theta_2) & \cdots & a_\alpha(\theta_s) \\ a_\beta(\theta_1) & a_\beta(\theta_2) & \cdots & a_\beta(\theta_s) \\ a_\rho(\theta_1) & a_\rho(\theta_2) & \cdots & a_\rho(\theta_s) \end{bmatrix} \tag{4}$$

and

$$a_\alpha(\theta) = \frac{1}{2}(1 + \tan^2(\theta)) \tag{5}$$

$$a_\beta(\theta) = -4\overline{\beta/\alpha}^2 \sin^2(\theta) \tag{6}$$

$$a_\rho(\theta) = -4\left(\overline{\beta/\alpha}^2 \sin^2(\theta)\right). \tag{7}$$

The differences in (3) represent approximations to derivatives originating from the continous version of the model. To avoid boundary problems for $i = 1$ and $i = n$ in (3), we use forward and backward differences, respectively. The relation between m and c may be written on a linear matrix form. Let $A = I_n \otimes \Lambda^T$, where I_n is the $n \times n$ identity matrix and $\otimes$ represent the Kronecker product. Moreover, let D be a matrix representing the approximations to the derivatives in (3). We then may write (3) on a linear form as $c = ADm$.

The seismic data $d = (d_1^T, \ldots, d_n^T)^T$ are related to the reflection coefficients through a wavelet convolution

$$d_{ij} = \sum_{u=-k}^{k} w(u, \theta_j) \cdot c_{i-u,j} + \varepsilon_{ij}, \tag{8}$$

where $w(\cdot, \cdot)$ is a wavelet function. We assume ε_{ij} to be coloured Gaussian noise, given as

$$\varepsilon_{ij} = \sum_{u=-k}^{k} w(u, \theta_j)\varepsilon_{i-u,j}^1 + \varepsilon_{i,j}^2 \tag{9}$$

where $\varepsilon_{i,j}^1$ and $\varepsilon_{i,j}^2$ are independent Gaussian white noise with $\mathrm{Var}(\varepsilon_{i,j}^1) = \sigma_1^2$ and $\mathrm{Var}(\varepsilon_{i,j}^1) = \sigma_2^2$. We see that the first part of the noise has the same waveform as the wavelet while the second part is white nose. The relation in (8) may also be written on linear matrix form

$$d = Wc + W\varepsilon^1 + \varepsilon^2 = Wc + \varepsilon. \tag{10}$$

Finally, by defining $G = WAD$, we get

$$d = WADm + W\varepsilon^1 + \varepsilon^2 = Gm + \varepsilon. \tag{11}$$

Remark 1. As mentioned above, the Bayesian model in Larsen et al. (2006) only differs from the model presented above in the rock physics formulation. Larsen et al. (2006) assume $p(m_i|\pi_i)$ to be an empirical distribution in stead of the Gaussian assumption in (2) and are in this way able to model more general rock physics models. In this paper we only consider the rock physics model in (2).

2.3 Posterior distribution

The posterior distribution resulting from the prior and likelihood models presented above can be written as

$$p(\pi|d) \propto \int \dots \int p(d|m) \prod_{i=1}^{n} p(m_i|\pi_i) \mathrm{d}m \prod_{i=1}^{n} p(\pi_i|\pi_{i-1}). \tag{12}$$

Here $p(d|m)$ and $p(m_i|\pi_i)$, refer to the relations in (11) and (2), respectively. In $p(\pi|d)$ and below, we let $p(\pi_1) = p(\pi_1|\pi_0)$ for notational convenience. We integrate out the elastic parameters, m, from the posterior distribution to emphasise that our focus is on the LF classes, π, in the posterior distribution. Larsen et al. (2006) introduce an approximation in the likelihood which makes it possible to evaluate the approximate posterior analytically and fast. The resulting approximate posterior distribution can be written on the form

$$\hat{p}(\pi|d) \propto \prod_{i=1}^{n} l(d|\pi_i) p(\pi_i|\pi_{i-1}). \tag{13}$$

In the next section we give the details in the approximation which results in $\hat{p}(\pi|d)$ and discuss how to simulate from $p(\pi|d)$ and $\hat{p}(\pi|d)$.

3 Inversion algorithms

In Sections 3.1 and 3.2 we give a review of the inversion methods in Larsen et al. (2006) and Hammer and Tjelmeland (2008), respectively.

3.1 Fast and approximate inversion

In this section we present the fast and approximate inversion method of Larsen et al. (2006). From (12), we may write

$$p(\pi|d) \propto \int \dots \int \frac{p(m|d)}{p(m)} \prod_{i=1}^{n} p(m_i|\pi_i) \mathrm{d}m \prod_{i=1}^{n} p(\pi_i|\pi_{i-1}). \tag{14}$$

The distribution $p(m)$ is a mixture of L^n Gaussian components. We are not able to evaluate this distribution or $p(m|d)$ analytically. Instead we substitute $p(m)$ and $p(m|d)$ with distributions $p^\star(m)$ and $p^\star(m|d)$, respectively, but in such a way that the distribution $p(\pi|d)$ is not changed. We acheve this by letting $p^\star(m)$ and $p^\star(m|d)$ satisfy the relation

$$p^\star(m|d) \propto p(d|m) p^\star(m). \tag{15}$$

Here $p(d|m)$ denotes (11) in the likelihood model. Based on (15) we can then write (14) as

$$p(\pi|d) \propto \int \dots \int \frac{p^\star(m|d)}{p^\star(m)} \prod_{i=1}^{n} p(m_i|\pi_i) \mathrm{d}m \prod_{i=1}^{n} p(\pi_i|\pi_{i-1}). \tag{16}$$

We observe that if let $p^\star(m)$ be Gaussian, then also $p^\star(d|m)$ becomes Gaussian based on (15). We therefore let $p^\star(m)$ be Gaussian with the possibility of spatial dependencies

$$p^\star(m) = \mathrm{N}(m; \mu_m^\star, \Sigma_m^\star), \tag{17}$$

where $\Sigma_m^\star = c(\delta) \otimes \Sigma_{m0}^\star$. Here $c(\delta)$ is a correlation function to model spatial dependencies along the vertical profile and $\Sigma_{m0}^\star$ is a 3×3 matrix representing intervariable covariance structure for the elastic parameters.

As a basis for a fast inversion we introduce an approximation by removing the spatial dependence in $p^\star(m)$ and $p^\star(m|d)$ in (16). This means that the covariance matrices in $p^\star(m)$ and $p^\star(m|d)$ are 3×3 block diagonal, only containing the intervariable dependencies. Note that we use $p^\star(m)$ with spatial dependencies when we calculate $p^\star(m|d)$ in (15) before we remove the spatial dependencies. An approximate posterior can then be written as

$$\hat{p}(\pi|d) \propto \prod_{i=1}^{n} \left[p(\pi_i|\pi_{i-1}) \int \frac{p^\star(m_i|d)}{p^\star(m_i)} p(m_i|\pi_i) \mathrm{d}m_i \right]. \tag{18}$$

Since we assume that the rock-physics model $p(m_i|\pi_i)$ is Gaussian distributed, the integrals in (18) are analytically available. The approximate posterior $\hat{p}(\pi|d)$ can then be written as given in (13). The posterior $\hat{p}(\pi|d)$ is in fact a first order hidden Markov model. The approximate posterior can therefore be exactly assessed very rapidly using the Forward-backward (FB) algorithm (Scott, 2002) shortly described in the introduction. Realizations can be generated extremely fast from $\hat{p}(\pi|d)$ of course.

3.2 MCMC simulation: Inversion with no approximations

In this section we summarise the Markov chain Monte Carlo (MCMC) simulation algorithm presented in Hammer and Tjelmeland (2008). The algorithm simulates from $p(\pi|d)$ with no approximations. To define an effective MCMC algorithm, we introduce $z = (z_1, \dots, z_n)^T$, where $z_i = (z_{i1}, \dots, z_{is})^T \in \mathbb{R}^s$ for $i = 1, \dots, n$,

$$z_i = c_i + \varepsilon_i^1 \tag{19}$$

and $\varepsilon_i^1 = (\varepsilon_{i1}^1, \dots, \varepsilon_{is}^1)^T$ as presented in the model in Section 2. Combining (8), (9) and (19) we get the relation between z and d,

$$d_{ij}|z \sim \mathrm{N}\left(d_{ij}; \sum_{u=-k}^{k} w(u, \theta_j) z_{i-u,j},\ \sigma_2^2 \right). \tag{20}$$

The posterior of interest is then

$$\begin{aligned} p(\pi, m, z|d) \propto p(\pi)p(m|\pi)p(z|m)p(d|z) = \\ \prod_{i=1}^{n} p(\pi_i|\pi_{i-1}) \cdot \prod_{i=1}^{n} \mathrm{N}\left(m_i|\mu(\pi_i), \Sigma(\pi_i)\right) \cdot \\ \prod_{i=1}^{n} \mathrm{N}\left(z_i|\, \Lambda^T \left(\frac{m_{i+1} - m_{i-1}}{2} \right), \sigma_1^2 I_s \right) \cdot \mathrm{N}(d|Wz, \sigma_2^2 I_{ns}), \end{aligned} \tag{21}$$

where W is the wavelet matrix introduced in Section 2. To simulate effectively from (21) we construct an MCMC algorithm consisting of two steps in each iteration. The first step is a Gibbs step (Liu, 2001) where we jointly update m and z conditioned on π and d. We are able to do this efficiently since the resulting distribution is Gaussian. The second step is a Metropolis–Hastings step where we propose new values for π and m conditioned on z from a proposal distribution, $q(\pi, m|z)$. Then the states of π from each iteration of the MCMC algorithm are (dependent) realizations from the posterior distribution in $p(\pi|d)$ after convergence of the Markov chain.

The key part in the MCMC algorithm is the proposal distribution $q(\pi, m|z)$. If we where able to generate proposals from $p(\pi, m|z, d) \propto p(\pi, m|z)$ it would be a Gibbs step and we would always get acceptance for our proposals. Our goal is therefore to generate proposals from a distribution close to $p(\pi, m|z)$. We start by giving an overview of an algorithm that generate realizations from $p(\pi, m|z)$, but this algorithm is not feasible to implement on a computer. Secondly we describe how we introduce approximations which result in the proposal distribution $q(\pi, m|d)$. For a detailed description of the proposal distribution, we refer to the original paper Hammer and Tjelmeland (2008). Here we only give a short review.

So ideally we would like to generate samples from $p(\pi, m|z)$ which is, using (1), (2) and (20),

$$p(\pi, m|z) \propto p(\pi)p(m|\pi)p(z|m) = \prod_{i=1}^{n} p(\pi_i|\pi_{i-1}) \cdot \prod_{i=1}^{n} \mathrm{N}\left(m_i|\mu(\pi_i), \Sigma(\pi_i)\right) \cdot \prod_{i=1}^{n} \mathrm{N}\left(z_i|\, \Lambda^T \left(\frac{m_{i+1} - m_{i-1}}{2}\right), \sigma_1^2 I\right). \tag{22}$$

This is a Hidden Markov model of order two with two hidden layers, π and m, and one observed layer, z. The FB procedure described for the traditional HMMs in Section 1 can also be used for this model, but the computational efficiency will be much reduced.

The forward part of a forward-backward algorithm for $p(\pi, m|z)$ sequentially integrates out m_i and π_i for $i = 1, \ldots, n$. Because of the Markovian structure of $p(\pi, m|z)$ and that the π_i's are discrete and the m_i's are Gaussian these integrals are analytically available. However, the result after having integrated over $(\pi_i, m_i), i = 1, \ldots, k$ is a mixture of L^k Gaussian densities. Thus, the number of mixture terms grows exponentially and the exact algorithm is computationally feasible only for very small values of n. A corresponding approximate forward integration algorithm is defined by ignoring the less important Gaussian terms, keeping a number of mixture terms that makes the algorithm feasible. Thereafter to do the backward simulation is computationally straight forward. We use the probability distribution defined by this approximate forward-backward algorithm as proposal distribution in a Metropolis–Hastings scheme.

Remark 2. In MCMC simulation one always needs to be careful when it comes to the analysis of the convergence and mixing properties of the Markov chain. The MCMC algorithm described above converges fast and mixes very well, which makes the results reliable. For an analysis of the convergence and mixing properties of the MCMC algorithm, see Hammer and Tjelmeland (2008).

Remark 3. Our main focus will, in the rest of the paper, be on the marginal distribution for the LF classes. The fast and approximate algorithm in Section 3.1 calculates the marginal probabilities by the FB algorithm, while the MCMC algorithm in Section 3.2 estimates these quantities through simulation. This makes the approximate algorithm a much faster alternative.

4 Comparison criteria

We seek for ways to quantify the differences between the true posterior $p(\pi|d)$ and the approximate posterior $\hat{p}(\pi|d)$. As mentioned above, we evaluate $\hat{p}(\pi|d)$ and $p(\pi|d)$ using the algorithms in Sections 3.1 and 3.2, respectively. Our focus will be on the posterior marginal distributions for the LF classes along the vertical profile for the two methods. This is typically the quantities from the inversions (posterior distributions) of main interest. We may observe differences by e.g. inspecting the marginal probabilities visually, but to really quantify the differences, we use univariate measures.

A natural quantity to measure, is the ability the posterior distributions have in regaining the true π. This does not cover all the aspects of the posterior marginal distributions but at least a very important one. A technique to archive this is the use of confusion matrices. We construct $L \times L$ confusion matrices, $C(p) = [c_{i,j}]$ such that row i give the posterior probabilities for the different LF classes when LF class i is the truth. More specifically we define

$$c_{i,j} = \frac{\sum_{k=1}^{n} \mathrm{I}(\pi_k^0 = i) P(\pi_k = j|d),}{\sum_{k=1}^{n} \mathrm{I}(\pi_k^0 = i)} \tag{23}$$

where $\mathrm{I}(\cdot)$ represent the indicator function and π_k^0 the true LF class at node k. The posterior marginal probabilities $P(\pi_k = j|d)$ can either be calculated from $p(\pi|d)$ or $\hat{p}(\pi|d)$. Clearly, we want the diagonals in the confusion matrices to be close to one.

Now we want to measure to what extent a posterior distribution do wrong classifications related to the true π, and define

$$\Delta(p) = \frac{1}{n} \sum_{k=1}^{n} \sum_{i \neq \pi_k^0} P(\pi_k = i|d) = \frac{1}{n} \sum_{k=1}^{n} \left[1 - P(\pi_k = \pi_k^0|d)\right]. \tag{24}$$

We see that for each node in (24) we have the posterior probability for all the LF classes except the true LF class π_k^0. Naturally we want $\Delta(p)$ to be small. We may calculate (24) based on either $p(\pi|d)$ or $\hat{p}(\pi|d)$ and denote this $\Delta(p)$ and $\Delta(\hat{p})$, respectively. We may also want to calculate (24) based on the prior marginal probabilities, $p(\pi_k)$, and denote the measure $\Delta(pr)$.

It is natural to assume that in a practical situation some wrong classifications have more consequences then others. For example confusing oil with gas may not be as dramatic as confusing oil with some other rock type, for example shale. We incorporate this by defining an $L \times L$ loss matrix $\Gamma = [\gamma_{i,j}]$, where $\gamma_{i,j}$ quantify the consequence of classifying to the class j if i is the true LF class. A natural generalisation of (24) is then

$$\Delta_\Gamma(p) = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{L} \gamma_{\pi_k^0, i} \; P(\pi_k = i|d). \tag{25}$$

Similar to (24) we also want (25) to be small. Similar to above we may calculate (25) based on prior probabilities, $p(\pi|d)$ or $\hat{p}(\pi|d)$ and denote the resulting measures $\Delta_\Gamma(pr)$, $\Delta_\Gamma(p)$ and $\Delta_\Gamma(\hat{p})$, respectively.

We also introduce a metric to quantify the difference between discrete distributions. The metric is inspired by Endres and Schindelin (2003), but we multiply by the factor $1/(2\ln(2))$ so that the maximal distance between the two distributions will be 1. Let $p = (p_1, \ldots, p_L)$

and $q = (q_1, \ldots, q_L)$ represent two discrete distributions. We quantify the difference between p and q with

$$d(p,q) = \sqrt{\frac{1}{2\ln(2)} \sum_{i=1}^{L} \left(p_i \ln\left(\frac{2p_i}{p_i + q_i}\right) + q_i \ln\left(\frac{2q_i}{p_i + q_i}\right)\right)}. \tag{26}$$

We then have that $d(p,q)^2$ satisfies the requirement of being a metric. At an arbitrary node k, we are typically interested in measuring differences between the marginal prior probabilities $p(\pi_k)$, posterior marginal probabilities $p(\pi_k|d)$ and approximate posterior marginal probabilities $\hat{p}(\pi_k|d)$. We also want to compare these marginal probabilities against the true π by defining a "probability distribution" for the true π at a node being 1 for the true LF class and 0 for the others. More specifically, at an arbitrary node k we denote this "probability distribution" $p_k^0 = (p_{k1}^0, \ldots, p_{kL}^0)$ with $p_{ki}^0 = \mathrm{I}(\pi_k^0 = i)$ for $i = 1, \ldots, L$. It is natural to evaluate the differences between the distribution above for the whole profile. We compute the difference by taking the average of the differences for each node. For example for the difference between the true π and the prior distribution we compute and denote this as follows

$$\overline{d}(p^0, pr) = \frac{1}{n} \sum_{k=1}^{n} d(p_k^0, p(\pi_k)). \tag{27}$$

We denote the difference from the true π to $p(\pi|d)$ and $\hat{p}(\pi|d)$ with $\overline{d}(p^0, p)$ and $\overline{d}(p^0, \hat{p})$, respectively. We denote other combinations of differences similarly.

The posterior distribution is a result of both the prior distribution and the likelihood model. The approximation introduced to get $\hat{p}(\pi|d)$ is done in the likelihood. A natural quantification of the consequences of the approximations is then to see how much the likelihood models correct on the prior information. We then define based on (24), (25) and (26)

$$\rho_\Delta = \frac{\Delta(pr) - \Delta(\hat{p})}{\Delta(pr) - \Delta(p)} \tag{28}$$

$$\rho_{\Delta_\Gamma} = \frac{\Delta_\Gamma(pr) - \Delta_\Gamma(\hat{p})}{\Delta(pr) - \Delta_\Gamma(p)} \tag{29}$$

$$\rho_d = \frac{\overline{d}(p^0, pr) - \overline{d}(p^0, \hat{p})}{\overline{d}(p^0, pr) - \overline{d}(p^0, p)}. \tag{30}$$

If the approximations introduced in the likelihood work well, we will expect the numerator to be almost as large as the denominator in ρ_Δ, ρ_{Δ_Γ} and ρ_d. If the approximations works poorly, the approximate likelihood adds little information to the posterior distribution and the numerator will be small compared to the denominator. This means that ρ_Δ, ρ_{Δ_Γ} and ρ_d can be interpreted as quantifying the portion of the information in the true likelihood preserved by the approximate likelihood.

5 Simulation example

We are now interested in evaluating the approximate posterior $\hat{p}(\pi|d)$ against the correct posterior $p(\pi|d)$ in a synthetic but realistic seismic inversion simulation example. We evaluate $\hat{p}(\pi|d)$ using the fast FB algorithm in Section 3.1 and simulate from $p(\pi|d)$ using the MCMC

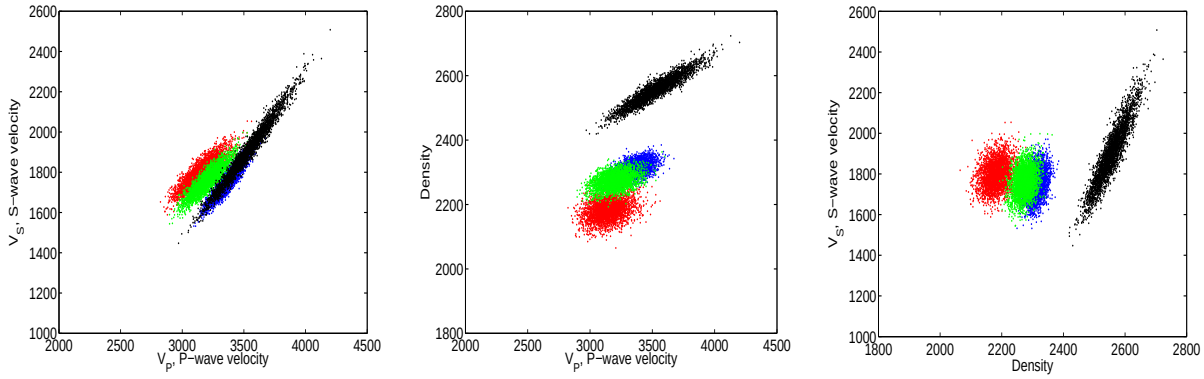

Figure 1: 3000 realizations from the rock physics model considered in the cases BC, NN, NL, MN and NZ. Red, green, blue and black represent gas, oil and brine saturated sandstone, and shale, respectively.

algorithm in Section 3.2. The general test design is to generate data from the forward model in Section 2 and secondly perform inversions using the two algorithms in Section 3. We evaluate the inversion results based on the criteria described in Section 4.

5.1 Model parameter values

We will evaluate the posteriors $p(\pi|d)$ and $\hat{p}(\pi|d)$ for seven different choices for the parameters in the model. The parameters we choose are closely related to the choices in Larsen et al. (2006) and Hammer and Tjelmeland (2008). We start by presenting one case, which we characterise as the base case and denote with BC. Secondly we present how the other six cases are different from BC. We use a length of the profile equal to $n = 100$. We consider $L = 4$ LF classes, representing gas, oil and brine saturated sandstone, and shale. We use the following prior transition matrix

$$P = \begin{bmatrix} 0.9441 & 0 & 0 & 0.0559 \\ 0.0431 & 0.9146 & 0 & 0.0424 \\ 0.0063 & 0.0230 & 0.9422 & 0.0284 \\ 0.0201 & 0.0202 & 0.1006 & 0.8591 \end{bmatrix} \tag{31}$$

defined upward along the vertical profile. The resulting marginal probabilities are $p(\pi_i) = [0.2419, 0.1552, 0.3830, 0.2199]$. The transition matrix is the same used in Hammer and Tjelmeland (2008) and closely related to the one in Larsen et al. (2006). The first row are the probabilities going from gas, the second row going from oil, the third brine saturated sandstone and the forth going from shale. We observe that we have zero probabilities for transitions from gas saturated sandstone to oil saturated sandstone, from gas saturated sandstone to brine saturated sandstone and from oil saturated sandstone to brine saturated sandstone, which seems reasonable due to gravity.

To illustrate the rock physics model, we generate 3000 realizations from $p(m_i|\pi_i)$ in (2) for each of the four LF classes and secondly calculate the resulting P-wave velocity, S-wave velocity and density, see Figure 1. In this figure and in the figures below, we let red, green, blue and black represent gas, oil and brine saturated sandstone, and shale, respectively. We

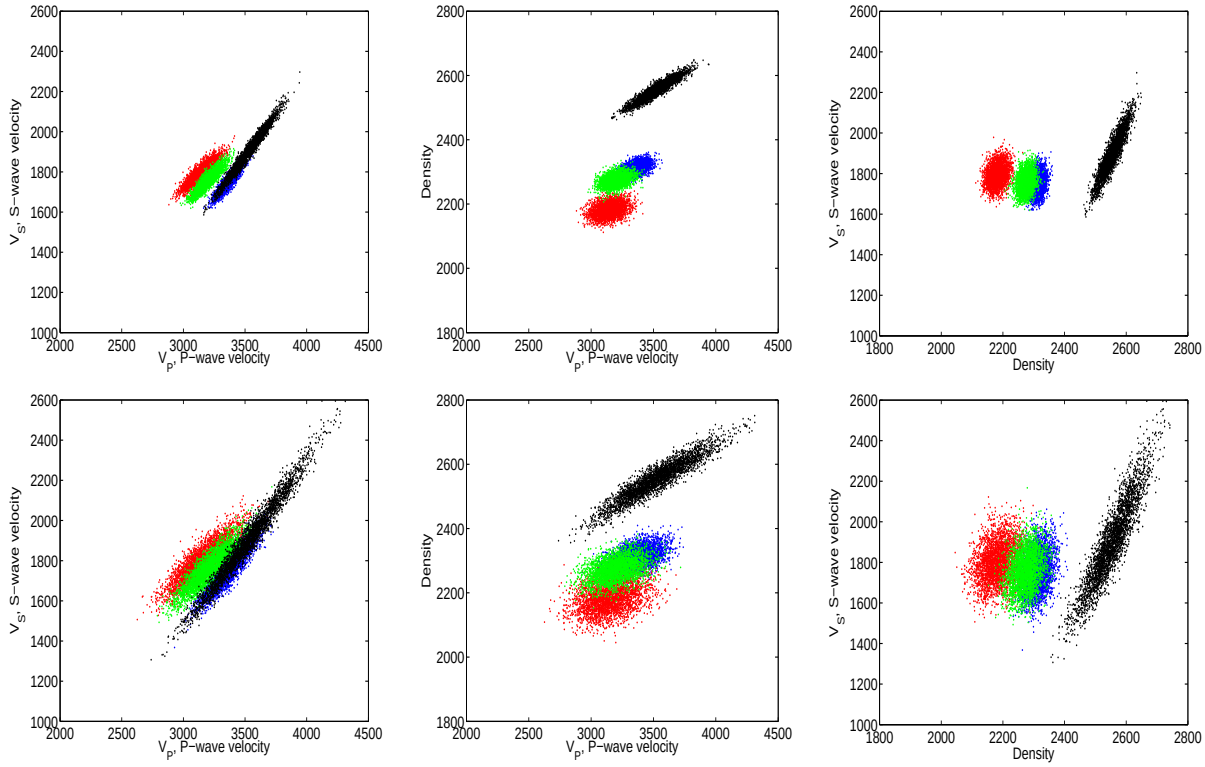

Figure 2: In row one and two we have 3000 realizations from the rock physics models considered in the cases RL and RM, respectively. Red, green, blue and black represent gas, oil, brine saturated sandstone and shale, respectively.

consider $s = 5$ angles $[0°, 10°, 20°, 30°, 40°]$ and use an angle independent Ricker wavelet

$$w(u) = (1 - 2(\pi\,\phi\,u)^2)\exp\left(-(\pi\,\phi\,u)^2\right),\ u = \{-k, \ldots, k\} \tag{32}$$

in (20), where we let $\phi = 0.11$ and $k = 10$.

In the base case BC, we set $\sigma_1 = 1.5 \cdot 10^{-2}$ and $\sigma_2 = 0.01\sigma_1$. This results in a reasonable Signal-to-noise ratio in the model. We return to this below. We consider six other cases. In the first case different from the BC, we change σ_1 to $5.0 \cdot 10^{-4}$ which can be interpreted as a case where we have essentially no noise added and denote the case NN. For the next two cases we change σ_1 to $0.85 \cdot 10^{-2}$ and $2.6 \cdot 10^{-2}$, which can be interpreted as cases with little and much noise and we denote the cases LN and MN, respectively. Similar to BC, we let in all the cases $\sigma_2 = 0.01\sigma_1$. For the next two cases we change the variability in the rock physics model. In the fifth case we divide the covariance matrices in (2) for the BC with two and in the sixth case we multiply the covariance matrices with two. The rock physics models are illustrated in Figure 2. We denote the cases RL and RM respectively, meaning rock physics models with less and more variability. For RL and RM we choose $\sigma_1 = 1.7 \cdot 10^{-2}$, $\sigma_1 = 0.95 \cdot 10^{-2}$, respectively and $\sigma_2 = 0.01\sigma_1$. We return to a motivation for these choices below. In the last case we change the prior transition matrix. We want to see if using a prior distribution with no zero elements, in contrast to (31), effects the difference between the posteriors $p(\pi|d)$ and $\hat{p}(\pi|d)$. In stead of (31) we use a transition matrix with 0.91 on the diagonal and 0.03 in all the other

Case	Description
BC	Base case
NN	No noise
LN	Little noise
MN	Much noise
RL	Rock physics model with less variability
RM	Rock physics model with more variability
NZ	Transition matrix with no zeros

Table 1: Description of the cases considered.

Case	Trans. matrix	Rock phys. model	σ_1	σ_2	SN	SN*
BC	(31)	Figure 1	$1.5 \cdot 10^{-2}$	$0.01\,\sigma_1$	2.91	1.27
NN	(31)	Figure 1	$5.0 \cdot 10^{-4}$	$0.01\,\sigma_1$	2350	3.49
LN	(31)	Figure 1	$0.85 \cdot 10^{-2}$	$0.01\,\sigma_1$	8.44	2.17
MN	(31)	Figure 1	$2.6 \cdot 10^{-2}$	$0.01\,\sigma_1$	0.95	0.64
RL	(31)	Row one Figure 2	$1.7 \cdot 10^{-2}$	$0.01\,\sigma_1$	1.90	1.28
RM	(31)	Row two Figure 2	$0.95 \cdot 10^{-2}$	$0.01\,\sigma_1$	8.44	1.31
NZ	(33)	Figure 1	$1.5 \cdot 10^{-2}$	$0.01\,\sigma_1$	3.02	1.32

Table 2: Parameter values and resulting Signal-to-noise ratios for the different cases considered.

elements

$$P_{NZ} = \begin{bmatrix} 0.91 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.91 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.91 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.91 \end{bmatrix}. \tag{33}$$

We denote this case NZ. For this case we use the same noise levels σ_1 and σ_2 as in BC, which result in a Signal-to-noise ratio equal to that in BC. All the cases are summarised in Tables 1 and 2.

The parameter sets in Table 2 are chosen to get reasonable Signal-to-noise ratios in the model. Similar to Hammer and Tjelmeland (2008) we find it natural to introduce two different interpretations of the Signal-to-noise ratio denoted with SN and SN*. For a more detailed description of the Signal-to-noise ratios, see Hammer and Tjelmeland (2008). For SN we consider m as the origin for the seismic signal and ε in (10) as the noise part. This is the way the Signal-to-noise ratio normally is defined in the seismic inversion literature, probably because the objective there is an inversion back to m and not all the way back to π. Typical values for SN in the literature are between 1.0 and 5.0. For SN* we consider π as the origin for the seismic signal and both the variability in the rock physics model in (2) and ε as noise parts. The Signal-to-noise values for the seven situations are given in Table 2. We see that SN* is approximately the same for the cases BC, RL, RM and NZ. It is easier to observe the effect of other changes in the model by holding SN* constant. The effect of changes in SN and SN* are studied by comparing the cases BC, NN, LN and MN.

In addition to the parameter choices in the model, we need to find a suitable correlation function $c(\delta)$ for the inversion method in Section 3.1. By generating samples of m from the stochastic model and inspecting the spatial correlations in the samples, we find a correlation

function of the form

$$c(\delta) = \exp\left(-\delta^{1/2}/3\right) \tag{34}$$

suitable.

5.2 Results with discussion

In this section we analyse the difference between the posteriors $\hat{p}(\pi|d)$ and $p(\pi|d)$ through several simulations using the algorithms in Sections 3.1 and 3.2. For each of the seven cases in Table 1 we generate ten independent realizations of π from the prior (1) and resulting seismic data d using the forward model in Section 2. Secondly we run both of the inversion algorithms for all the resulting 70 sets of data d.

We start by visually inspecting realizations from the ten inversions in each of the cases given in Table 1. Figures 3 and 4 shows a few samples from one of the ten inversions for each of the cases given in Table 1. In the figures, the first column shows the seismic data d, the second column the truth π, the next four columns independent samples from the fast and approximate algorithm in Section 3.1 and in the final four columns four independent samples from the MCMC algorithm in Section 3.2. We see that for both inversion algorithms the amount of variability in the realisations increases when the noise level increases. We see that in all the cases considered the realizations have similarities with the truth π. For the cases NN and LN there are little difference between the realizations from the same inversion algorithm, while for MN there are large differences, as one would expect. It also seems like the realizations from $\hat{p}(\pi|d)$ have some systematic differences compared to the true π while we do not see the same effect for $p(\pi|d)$. Except from these observations, it is hard to say anything more precise about the differences between the posteriors $p(\pi|d)$ and $\hat{p}(\pi|d)$.

A step further is to look at marginal probabilities for one set of seismic data for each case. This is given in Figures 5 and 6. In the two figures, in the first column we have the truth π. In the second and the fourth columns we have the marginal probabilities for the posteriors $\hat{p}(\pi|d)$ and $p(\pi|d)$, respectively. The first can be computed by the FB recursive algorithm while the latter must be assessed by sampling. In the third and the fifth columns we have the distance the marginal probabilities in each node are from the truth using the metric in (26). In the last column we have the difference between the marginal probabilities in columns two and four. It seems like the true posterior $p(\pi|d)$ over all are closer to the true π compared to the approximate posterior $\hat{p}(\pi|d)$ as one would expect.

To be able to draw more definite conclusions we need to summarise the results for all the ten inversions for each of the seven cases. In Table 3 we present confusion matrices for all the seven cases given in Table 1. We have calculated the elements in the confusion matrix according to (23). We have in addition, for each case taken the average of the calculated confusion matrix for each of the ten inversions. We see that the true posterior $p(\pi|d)$ regains the truth better then $\hat{p}(\pi|d)$, as one would expect. However the probabilities along the diagonal is also mostly large for the approximate posterior $\hat{p}(\pi|d)$, which means that also this posterior contains valuable information about the true π.

Now we want to quantify the difference in $p(\pi|d)$ and $\hat{p}(\pi|d)$ using (24) and (25). As mentioned in Section 4 it is natural to expect that some wrong classifications have more

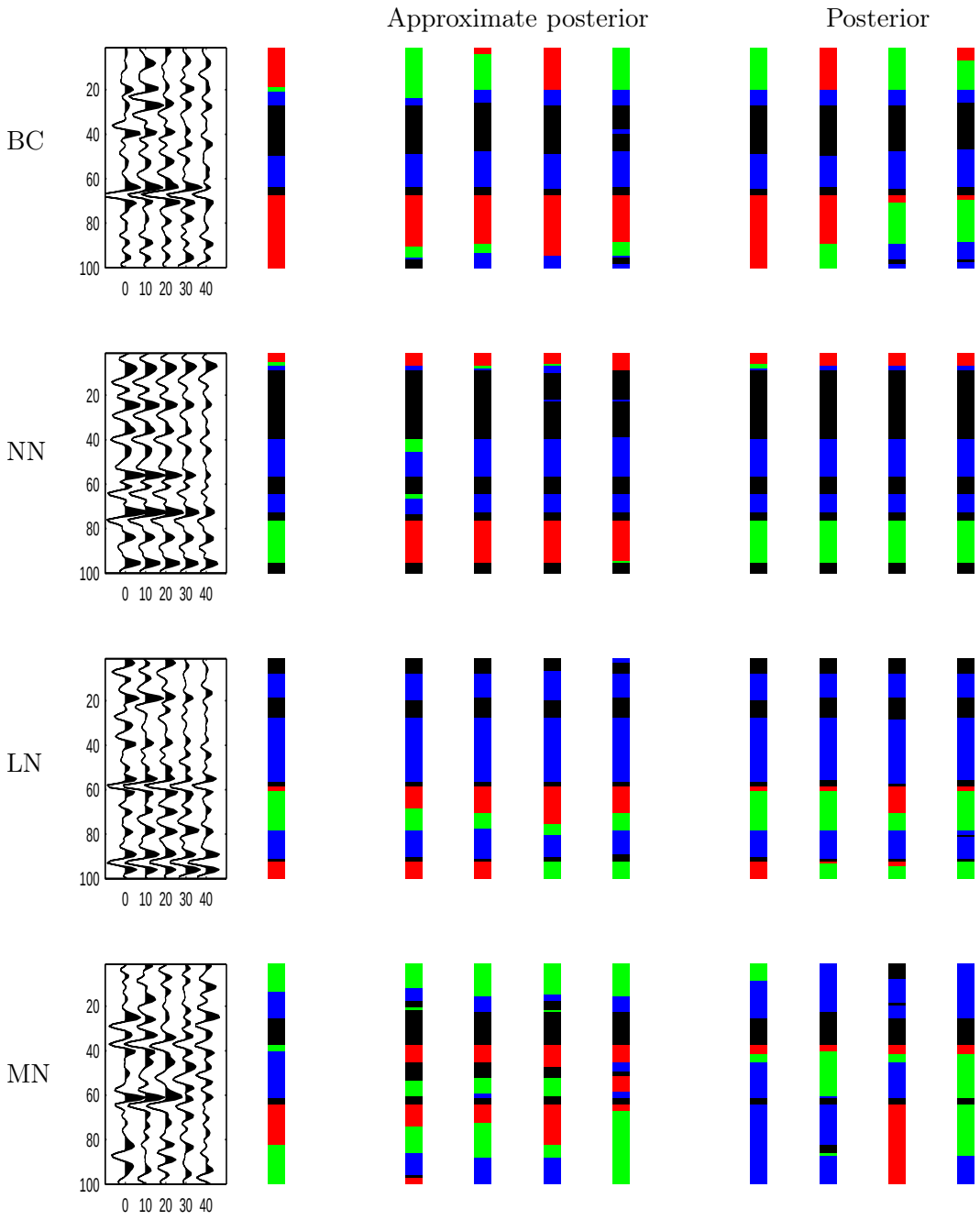

Figure 3: Row 1 to 4 give results for the cases BC, NN, LN and MN, respectively. In the first column we have the seismic data d, the second column the truth π, the next four columns independent samples from the fast and approximate algorithm in Section 3.1 and in the final four columns four independent samples from the MCMC algorithm in Section 3.2.

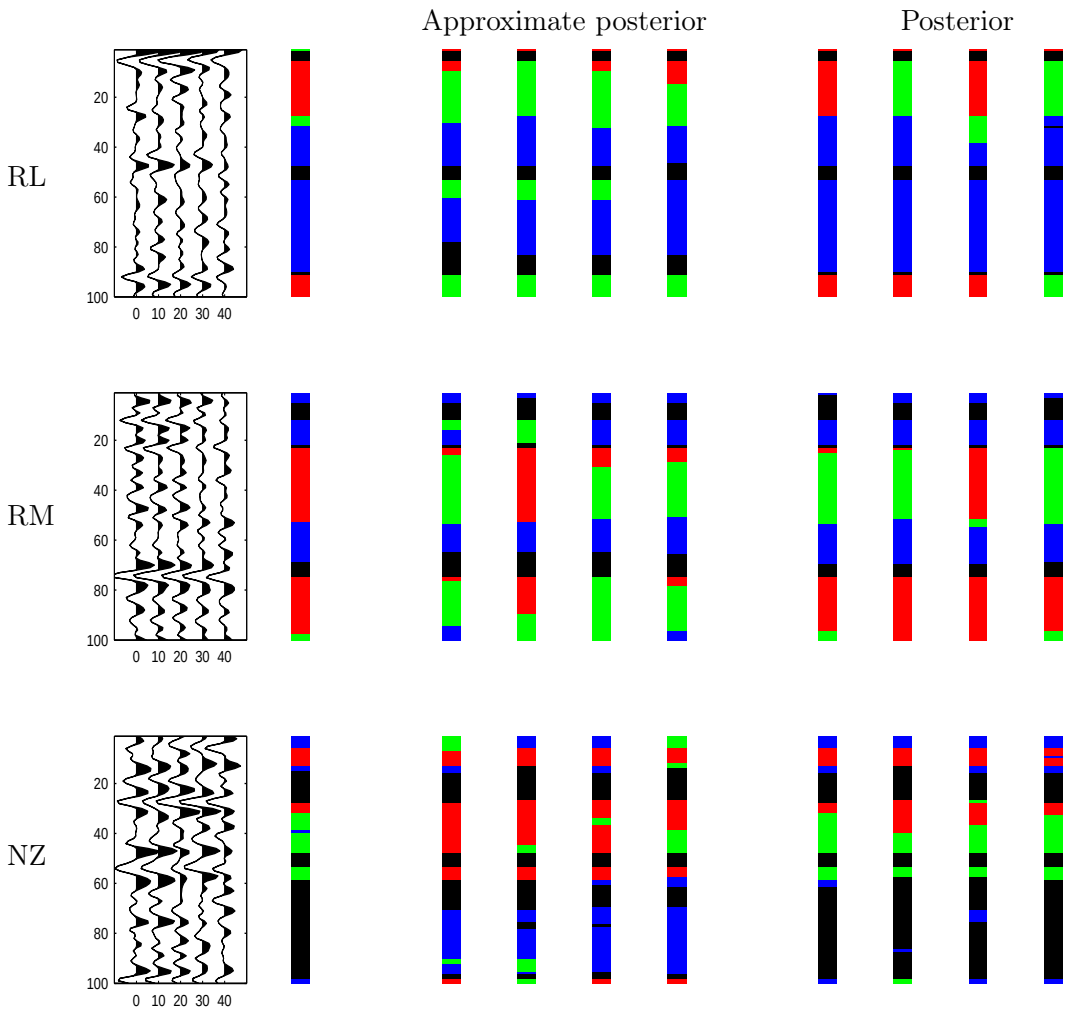

Figure 4: Row 1 to 3 give results for the cases RL, RM and NZ, respectively. In the first column we have the seismic data d, the second column the truth π, the next four columns independent samples from the fast and approximate algorithm in Section 3.1 and in the final four columns four independent samples from the MCMC algorithm in Section 3.2.

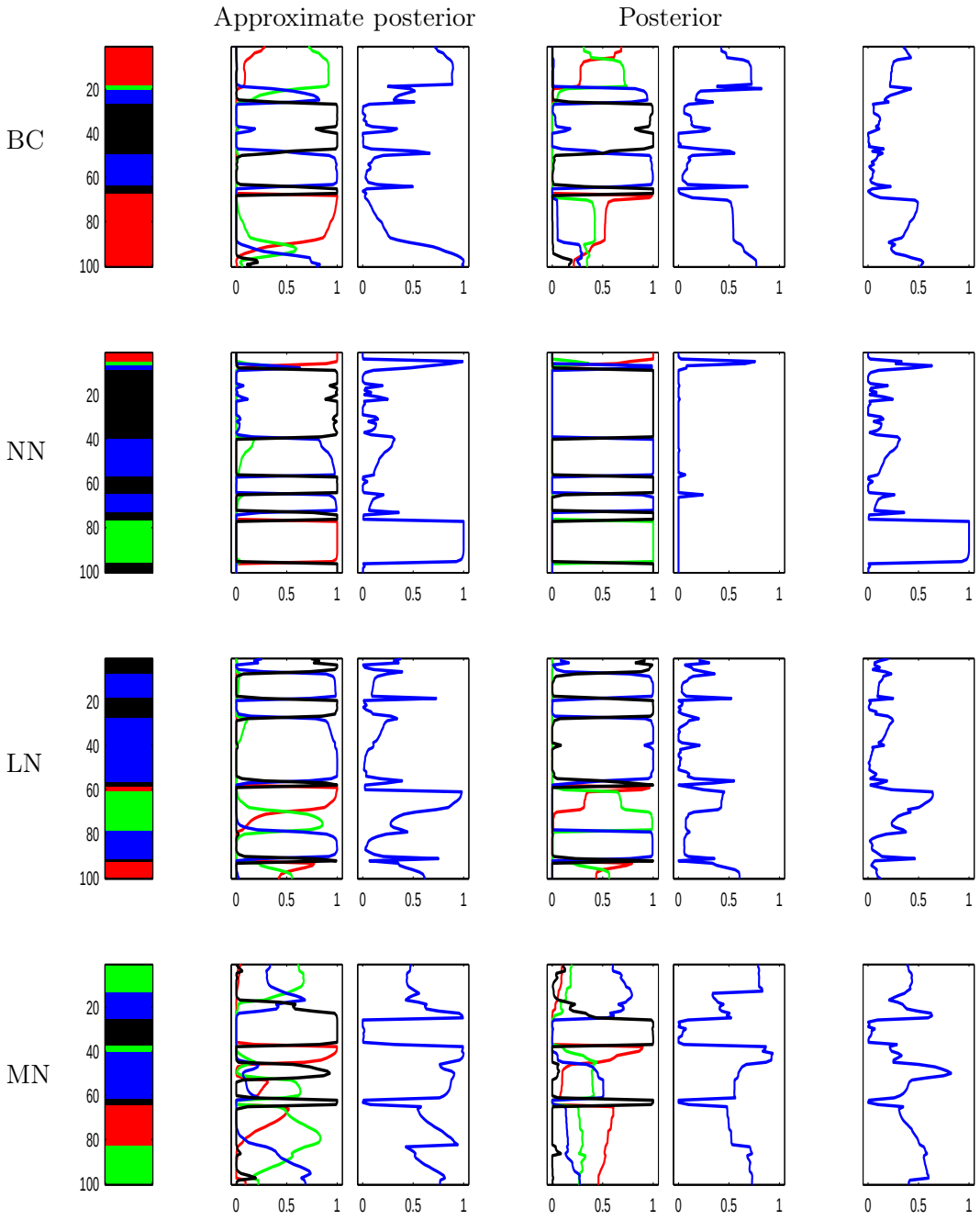

Figure 5: Row 1 to 4 give results for the cases BC, NN, LN and MN, respectively. In the first column we have the truth π. In the second and the fourth columns we have the marginal probabilities for the posteriors $\hat{p}(\pi|d)$ and $p(\pi|d)$, respectively. In the third and the fifth columns we have the distance the marginal probabilities in each node are from the truth using the metric in (26). In the last column we have the difference between the marginal probabilities in column two and four.

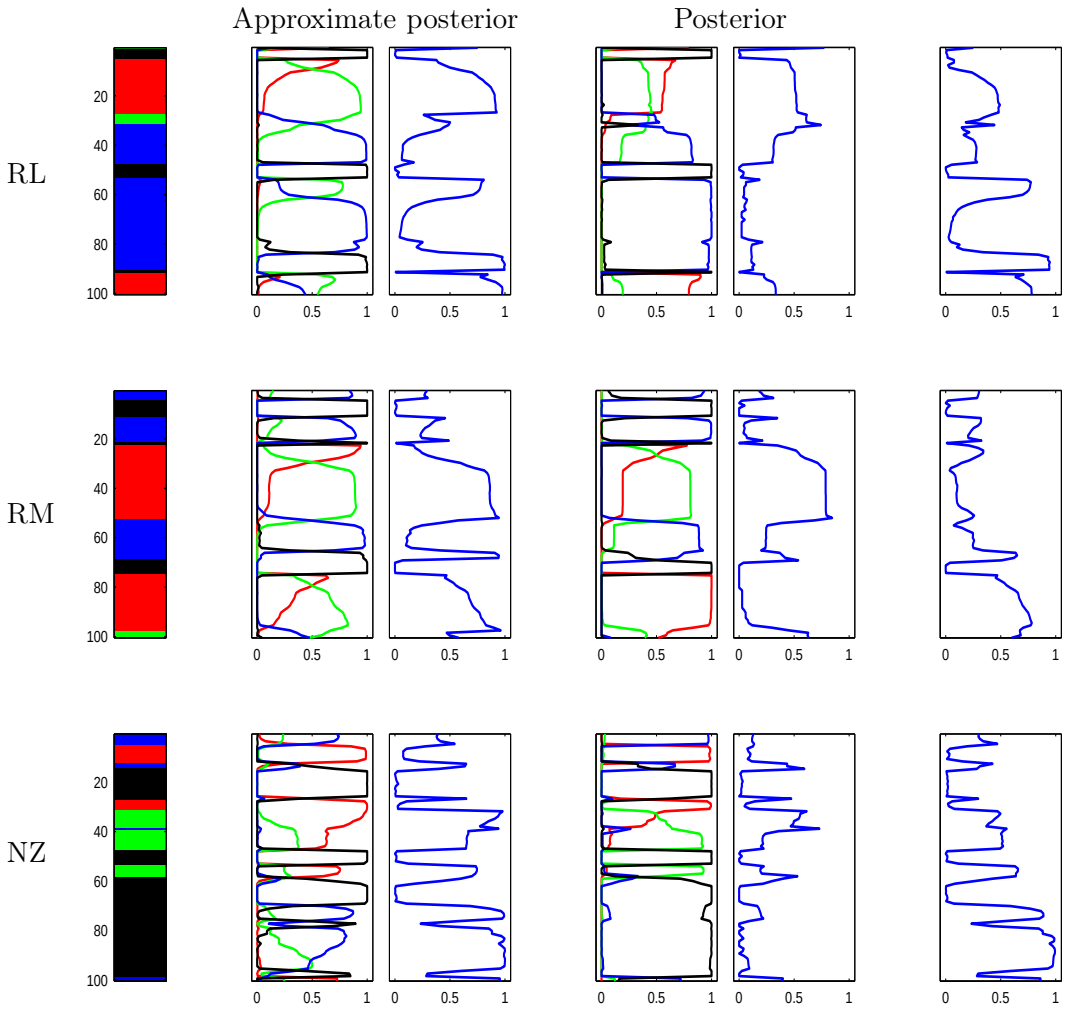

Figure 6: Row 1 to 3 give results for the cases RL, RM and NZ, respectively. In the first column we have the truth π. In the second and the fourth columns we have the marginal probabilities for the posteriors $\hat{p}(\pi|d)$ and $p(\pi|d)$, respectively. In the third and the fifth columns we have the distance the marginal probabilities in each node are from the truth using the metric in (26). In the last column we have the difference between the marginal probabilities in column two and four.

		Approximate posterior					Posterior			
		gas	oil	brine	shale		gas	oil	brine	shale
BC	gas	0.459	0.372	0.159	0.001	gas	0.794	0.168	0.035	0.003
	oil	0.488	0.451	0.061	0.000	oil	0.410	0.495	0.092	0.003
	brine	0.035	0.196	0.639	0.130	brine	0.003	0.053	0.874	0.069
	shale	0.013	0.002	0.047	0.939	shale	0.001	0.002	0.078	0.920
		gas	oil	brine	shale		gas	oil	brine	shale
NN	gas	0.881	0.118	0.000	0.000	gas	0.996	0.004	0	0
	oil	0.542	0.447	0.010	0.000	oil	0.018	0.976	0.007	0.000
	brine	0.002	0.044	0.936	0.018	brine	0	0.000	0.997	0.003
	shale	0.001	0.001	0.038	0.961	shale	0	0.000	0.006	0.994
		gas	oil	brine	shale		gas	oil	brine	shale
LN	gas	0.959	0.041	0.000	0.000	gas	0.983	0.017	0	0.000
	oil	0.224	0.595	0.171	0.010	oil	0.278	0.614	0.108	0.001
	brine	0.014	0.250	0.600	0.136	brine	0.001	0.014	0.953	0.032
	shale	0.006	0.032	0.214	0.748	shale	0.000	0.000	0.081	0.919
		gas	oil	brine	shale		gas	oil	brine	shale
MN	gas	0.253	0.458	0.232	0.057	gas	0.623	0.200	0.141	0.037
	oil	0.429	0.280	0.213	0.078	oil	0.413	0.313	0.253	0.021
	brine	0.115	0.218	0.474	0.194	brine	0.047	0.137	0.741	0.072
	shale	0.057	0.048	0.141	0.754	shale	0.013	0.029	0.222	0.735
		gas	oil	brine	shale		gas	oil	brine	shale
RL	gas	0.759	0.237	0.002	0.002	gas	0.888	0.111	0.000	0.000
	oil	0.190	0.333	0.471	0.006	oil	0.113	0.585	0.298	0.003
	brine	0.049	0.311	0.542	0.098	brine	0.003	0.079	0.900	0.018
	shale	0.003	0.012	0.134	0.850	shale	0.001	0.004	0.092	0.903
		gas	oil	brine	shale		gas	oil	brine	shale
RM	gas	0.716	0.214	0.061	0.010	gas	0.751	0.234	0.009	0.005
	oil	0.635	0.286	0.071	0.008	oil	0.416	0.542	0.035	0.007
	brine	0.014	0.186	0.689	0.112	brine	0.002	0.043	0.892	0.063
	shale	0.019	0.031	0.062	0.888	shale	0.001	0.007	0.073	0.919
		gas	oil	brine	shale		gas	oil	brine	shale
NZ	gas	0.539	0.319	0.140	0.002	gas	0.764	0.206	0.028	0.002
	oil	0.308	0.332	0.336	0.024	oil	0.258	0.632	0.096	0.014
	brine	0.090	0.294	0.407	0.209	brine	0.039	0.287	0.644	0.031
	shale	0.016	0.054	0.137	0.794	shale	0.003	0.025	0.101	0.872

Table 3: Confusion matrices. Row 1 to 7 give results for the cases given in Table 1.

	$\Delta(pr)$	$\Delta(\hat{p})$	$\Delta(p)$	$\Delta_\Gamma(pr)$	$\Delta_\Gamma(\hat{p})$	$\Delta_\Gamma(p)$	ρ_Δ	ρ_{Δ_Γ}
BC	0.722	0.391	0.177	0.503	0.178	0.056	0.605	0.733
NN	0.722	0.162	0.008	0.503	0.031	0.002	0.789	0.942
LN	0.722	0.320	0.120	0.503	0.168	0.034	0.671	0.703
MN	0.722	0.565	0.405	0.503	0.303	0.226	0.500	0.727
RL	0.722	0.388	0.140	0.503	0.261	0.086	0.557	0.562
RM	0.722	0.325	0.198	0.503	0.130	0.044	0.763	0.817
NZ	0.750	0.532	0.286	0.525	0.315	0.152	0.469	0.563

Table 4: Row 1 to 7 give results for the cases given in Table 1. In the columns one to three we have calculated (24) based on the prior distribution, $\hat{p}(\pi|d)$ and $p(\pi|d)$ respectively. In columns four to six we have the same based on (25). In columns seven and eight we have calculated ρ_Δ and ρ_{Δ_Γ} in (28) and (29), respectively.

	$\overline{d}(p^0, pr)$	$\overline{d}(p^0, \hat{p})$	$\overline{d}(p^0, p)$	$\overline{d}(\hat{p}, p)$	ρ_d
BC	0.719	0.430	0.249	0.373	0.615
NN	0.719	0.195	0.013	0.191	0.747
LN	0.719	0.368	0.163	0.353	0.637
MN	0.719	0.600	0.463	0.393	0.480
RL	0.719	0.427	0.214	0.340	0.567
RM	0.719	0.366	0.246	0.240	0.752
NZ	0.741	0.568	0.357	0.428	0.450

Table 5: Row 1 to 7 give results for the cases given in Table 1. For columns one to four we have calculated $\overline{d}(p^0, pr)$, $\overline{d}(p^0, \hat{p})$, $\overline{d}(p^0, p)$ and $\overline{d}(\hat{p}, p)$, respectively. In column five we have calculated ρ_d in (30). For each value, the result are an average over all the ten inversions.

serious consequences than others. We therefore evaluate (25), where we use

$$\Gamma = \begin{bmatrix} 0 & 0.1 & 1 & 1 \\ 0.1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0.1 \\ 1 & 1 & 0.1 & 0 \end{bmatrix}, \tag{35}$$

where the first row quantifies the negative consequences of classifying to gas, oil, brine saturated sandstone and shale if gas is the truth, the second to the forth row is the same for oil and brine santurated sandstone and shale respectively. We quantify that the consequence of confusing the type of hydrocarbon or confusing the non-hydrocarbon classes is not as dramatic as confusing hydrocarbon with non-hydrocarbon. The results are summarised in Table 4. Row 1 to 7 give results for the cases given in Table 1. In columns one to three we have calculated (24) based on the prior distribution, $\hat{p}(\pi|d)$ and $p(\pi|d)$, respectively. In columns four to six we have the same based on (25). In columns seven and eight we have calculated ρ_Δ and ρ_{Δ_Γ} in (28) and (29), respectively. We now quantify the difference between the posterior distributions considering the metric in (26). The results are summarised in Table 5. Row 1 to 7 give results for the cases given in Table 1. For columns one to four we have calculated $\overline{d}(p^0, pr)$, $\overline{d}(p^0, \hat{p})$, $\overline{d}(p^0, p)$ and $\overline{d}(\hat{p}, p)$, respectively. In column five we have calculated ρ_d in (30). Each value is an average of the result from each of the ten inversions.

From row one to four in Tables 4 and 5 we see that the consequences of the approximations

increase when the amount of noise in the model increases. From row five and six we see that the consequence of the approximation are smallest for RM meaning that the approximations seem to work better when the variability in the model are mostly in the rock physics model and less in the noise part ε. We get the largest consequences for the case NZ where we have the prior transition matrix without zeros. Using a transition matrix without zeros may be seen as a weaker prior distribution and it then becomes more important to have a likelihood without approximations. From columns seven and eight in Table 4 and column five in Table 5 we see that the approximate likelihood typically preserves between 55% and 80% of the information in the true likelihood function.

6 Closing remarks

We have considered a Bayesian model for doing LF prediction from AVO seismic data. The model contains a Markov chain prior, wavelet convolution and a colored noise term. Related to the Bayesian model, we have considered two inversion algorithms. The first algorithm, from Larsen et al. (2006), introduces an approximation in the likelihood model and is in this way able to evaluate exactly the resulting approximate posterior very efficiently. The other algorithm, presented in Hammer and Tjelmeland (2008), is a more computer demanding alternative, but simulate from the posterior model without approximations. The objective of this paper have been to evaluate the approximations introduced in the efficient algorithm, by comparing the inversion results from the two algorithms.

We have presented a synthetic but realistic study where several different parameter sets in the Bayesian model have been considered. The conclusions is that the approximate likelihood model typically preserves between 55 and 80% of the information in the true likelihood model. The consequences of the approximations increase when the amount of seismic noise in the model increases. The approximations work better in a situation where most of the variability is in the rock physics model and little is seismic noise, compared to the opposite case.

For an inversion problem of the size considered in this paper the algorithm without approximations seems to be the best alternative. In a real situation, we are normally interested in inverting a large amount of traces and then the algorithm without approximations can end up with problems, because it is quite computer demanding. For this situation the approximate alternative in Larsen et al. (2006) is the best alternative. It is also worth to note that the Bayesian model considered in this paper contains several assumptions and approximations. It is an open and relevant question how much the approximations introduced in the approximate algorithm will have on inversion results from real seismic data compared to the impact of the assumptions and approximations in the Bayesian model.

References

Aki, K. and Richards, P. G. (1980). *Quantitative seismology: Theory and methods*, W. H. Freeman and Company.

Avseth, P., Mukerji, T. and Mavko, G. (2005). *Quantitative Seismic interpretation : Applying rock physics tools to reduce interpretation risk*, Cambridge University 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.

Endres, D. M. and Schindelin, J. E. (2003). A New Metric for Probability Distributions, *IEEE Trans. Inform. Theory* **49**: 1858–1860.

Hammer, H. and Tjelmeland, H. (2008). Approximative forward-backward algorithm for a three layer hidden Markov model - with applications to seismic inversion, *Technical Report S4-2008*, Department of Mathematical Sciences, Norwegian University of Science and Technology.

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.

Mavko, G., Mukerji, T. and Dvorkin, J. (1998). *The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media*, Cambridge University Press.

Scott, A. L. (2002). Bayesian Methods for Hidden Markov Models: Recursive Compution in the 21st Century, *Journal of the American Statistical Association* **97**: 337–351.