

A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties

Martin Jullum¹ and Odd Kolbjørnsen²

ABSTRACT

Working in a Bayesian framework, we have derived a procedure for inverting rock properties based on geophysical data. The purpose was to arrive at a widely applicable and general procedure in which few and weak assumptions are required for application to various inverse problems within the geophysical industry. Our Bayesian statistical approach combines sampling-based techniques and Gaussian approximations to assess local approximations to quantities related to the posterior distribution of rock properties. These approximated quantities define the Bayesian inversion. A conceptual advantage of our approach is that there are few restrictions on the initial model, allowing realistic statistical models to be approximated directly. The methodology is easily parallelized and offers a range of procedures, which gives a trade-off between inversion speed and accuracy. We have tested the approach in a monitoring setting using seismic amplitudes by evaluating a synthetic case and real data from the Sleipner CO₂ injection project. For the synthetic case, the inversion results correspond well with the rock properties used to generate the data and the posterior distribution derived using an MCMC approach. We also found improved accuracy compared with a frequently used Gaussian inversion approach. In the real data case, we clearly identified high-saturation layers present in previous qualitative interpretations.

INTRODUCTION

In the geophysical industry, there is a great need for solutions to various types of inverse problems. Most of these problems are however ill posed and seldom have a unique or well-defined solution. The final objective of many of the inverse problems is to predict

rock properties such as porosity, lithology, saturation, permeability etc. from geophysical data such as seismic amplitudes (Mukerji et al., 2001; Doyen, 2007; Gunning and Glinsky, 2007; Avseth et al., 2010). The Bayesian approach (Tarantola and Valette, 1982) is a popular framework for solving inverse problems (Bosch et al., 2010). The main advantage of the Bayesian approach is the possibility to incorporate additional knowledge of the problem and assess the uncertainty after accounting for the data. In such a setting, computation or approximation of a posterior distribution, here corresponding to the probability distribution of the rock properties conditioned on the observed geophysical data, determines the inversion.

Unfortunately, full analytical evaluation of the posterior distribution is only possible for highly restricted classes of distributions. For instance, Buland and Omre (2003) perform inversion from seismic amplitude versus offset (AVO) data to elastic parameters (but not further to actual rock properties) by assuming that the two model components are jointly Gaussian, resulting in an analytic closed-form Gaussian posterior distribution for the elastic parameters. Because realistic models seldom fit such a formulation, the methodology may merely be viewed as an approximation, possibly far from the actual posterior distribution (Rimstad and Omre, 2014a). Rimstad and Omre (2014a, 2014b) relax the Gaussian model assumptions, but to evaluate the posterior distribution, they need to use Markov Chain Monte Carlo (MCMC; Robert and Casella, 2005) sampling procedures, which may be very time consuming in high dimensions. The Gaussian mixture approach of Grana and Della Rossa (2010) also relaxes the Gaussian assumption, and even goes all the way to rock properties, but it requires modeling approximations elsewhere and thereby also operates approximately.

Several authors have restricted their attention to discrete facies as the rock property of interest. Larsen et al. (2006) introduce Markov property dependencies to describe a vertical profile based on seismic AVO data. Buland et al. (2008), Ulvmoen and Omre (2010), and others follow along similar lines. These approaches include an initial step in which a distribution claimed to be multimodal is approximated by a unimodal Gaussian distribution based on Buland

Manuscript received by the Editor 13 June 2015; revised manuscript received 14 October 2015; published online 04 April 2016.

¹University of Oslo, Department of Mathematics, Oslo, Norway. E-mail: martinju@math.uio.no.

²Formerly Norwegian Computing Center, Oslo, Norway; presently Lundin Norway, Oslo, Norway. E-mail: odd.kolbjornsen@lundin-norway.no.

© 2016 Society of Exploration Geophysicists. All rights reserved.

and Omre (2003). Even if the multimodality is subsequently corrected for, and the results appear reasonable, the conflicting mode assumptions have unclear implications. Further, even if the used discrete Markov property is suitable in some situations, it generally restricts the dependence structure. Finally, because these methodologies rely on the discrete nature of the facies, the approaches cannot be directly transferred to situations with continuously distributed rock properties.

Although there is a wide range of techniques in the statistical literature for approximating Bayesian posterior distributions (for a review, see, e.g., Green et al., 2015), there are few examples of such techniques being directly applied to geophysical types of inverse problems. This is possibly caused by a gap between the computational efficiency of such techniques when being applied to large geophysical types of inverse problems, and what is acceptable for the industry (Mosegaard and Tarantola, 2002). Even though they are time consuming, attempts have been made to solve such inverse problems based on the MCMC approach (Mosegaard and Tarantola, 1995; Malinverno, 2002; Hammer et al., 2012).

It is evident that there is a need for a computationally feasible large-scale inversion methodology that can deal with more general model formulations than the existing ones. Short of computational advances, attempts to make statistical methods computationally tractable involve some kind of simplification, approximation, or intelligent elimination of redundancy. For these types of inverse problems, it often suffices to obtain inversion results for each individual cell in a grid of the region of interest. In a Bayesian framework, such local inversion corresponds to computing or approximating the marginal posterior distribution in each cell and using predictors and uncertainty measures based on those distributions as the inversion result.

Our methodology uses a local inversion approach that simplifies the problem. Inversion of a larger region then reduces to a large number of local inversions, which we handle individually. The efficiency of the local inversion approach lies partly in including only the variables and data most relevant for the current local inversion. This reduces the dimensionality of the problem to a magnitude that we can handle efficiently, while still taking the most relevant spatial dependence into account. The local inversion is carried out by relying on a certain Gaussian likelihood approximation and a weighted Monte Carlo sampling routine. Handling the local inversions individually allows us to parallelize the full-inversion problem, leading to heavy algorithmic speed-up compared with MCMC-type procedures. There are also few modeling limitations underlying our approach. In particular, our approach can handle any type of rock-physics model and any type of prior distribution for the rock properties. It is not limited by specific restrictions on the spatial structure typically present in other approaches (such as, e.g., certain Markov-type dependence structures), and it can in principle be used with any rock property: continuous, discrete, or a combination of the two.

METHODOLOGY

The full inverse problem is usually decomposed into two inversions: geophysical inversion (Buland and Omre, 2003) and

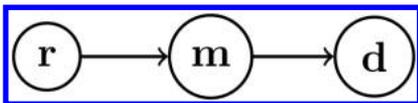


Figure 1. Forward model hierarchy.

rock-physics inversion (Avseth et al., 2010). The joint global problem may be described by the simple hierarchical formulation shown in Figure 1. Here, \mathbf{r} denotes a rock property, such as lithology, porosity, or saturation that we are interested in. The geophysical properties of \mathbf{m} typically consist of density and two elastic parameters (either P- and S-wave velocities, or acoustic and shear impedance) or any triplets of these. Finally, \mathbf{d} denotes the geophysical data often consisting of seismic AVO data at a few different offsets. The geophysical data may however also represent data sources such as root-mean-square (rms) velocity (Buland et al., 2011) or gravimetrics data (Hauge and Kolbjørnsen, 2015). Hence, the left arrow of Figure 1 represents the rock-physical relation, whereas the right arrow represents the geophysical relation. That is, all impact from the rock properties to the geophysical data goes through the geophysical properties.

We shall need a fair amount of notation when building our inversion framework. We shall assume that the global quantities in Figure 1 all operate on the same grid of the region of interest. (They may in principle work on different grids of the region, but we exclude that case for presentational simplicity.) Uppercase letters in calligraphic font ($\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$) are used to denote subsets of cells of the gridded region. A variable written in boldface roman font with a calligraphic subscript (e.g., $\mathbf{d}_{\mathcal{D}}$) refers to the subvector corresponding to the subset of that subscript. A boldface roman variable with no calligraphic subscript (e.g., \mathbf{d}) contains the individual variable(s) of the complete gridded region under consideration. For other quantities, we will use fairly standard statistical notation: We use a superscript roman T for the matrix transpose, $p(\cdot)$ as a generic notation for probability distributions, \sim for “distributed as,” $\boldsymbol{\mu}$ for the mean (vector), and $\boldsymbol{\Sigma}$ for the covariance matrix. The Gaussian distribution of a variable \mathbf{x} (with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$) is denoted by $N_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Noncalligraphic subscripts will be used to distinguish variables of similar types. A superscript asterisk will be specifically used to denote approximate quantities used in the framework, like, e.g., $p^*(\mathbf{x})$ and $\boldsymbol{\mu}_x^*$. The most important quantities are also given in Table 1.

The general framework

The overall goal of the inversion is to predict the rock properties \mathbf{r} from the obtained geophysical data \mathbf{d} over a gridded interest region \mathbb{A} . We will handle this by focusing on one grid cell at a time, and hence carry out predictions and uncertainty measures on grid cell level. Because the methodology will be the same for each cell in \mathbb{A} , we will present the methodology by considering a single cell \mathcal{A} — the extension to \mathbb{A} amounts to repeating the procedure for each $\mathcal{A} \in \mathbb{A}$. Working in the Bayesian framework, carrying out the inversion to rock properties in cell \mathcal{A} amounts to evaluating approximated quantities or measures related to the marginal posterior distribution $p(\mathbf{r}_{\mathcal{A}}|\mathbf{d})$ of the target variable $\mathbf{r}_{\mathcal{A}}$. Any preferred measure of central tendency of this posterior distribution may be used as a predictor for the true rock property in cell \mathcal{A} . Common selections are the mean, mode, and median. The uncertainty may be quantified by a measure of spread, such as the standard deviation, or through one or more suitably chosen credibility intervals. Local probability statements, such as the probability that the porosity in \mathcal{A} is more than 0.15, may also be computed. Our method may also produce an approximation to the complete marginal posterior distribution, should that be of interest.

By using Bayes' formula, the posterior distribution of the target variable is given by

$$p(\mathbf{r}_A|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{r}_A)p(\mathbf{r}_A). \quad (1)$$

Here, $p(\mathbf{r}_A)$ is the prior probability distribution for the target variable, whereas $p(\mathbf{d}|\mathbf{r}_A)$ is the likelihood of the geophysical data, conditioned only on the target variable. Hence, in $p(\mathbf{d}|\mathbf{r}_A)$, the geophysical properties \mathbf{m} are marginalized out along with the rock properties in other cells than \mathcal{A} . In terms of Figure 1, this model setup corresponds to a direct arrow from \mathbf{r}_A to \mathbf{d} , reducing the global two stage problem to a local one stage problem. This is beneficial because two-step approaches do not fully account for the dependence in Bayesian models (Bosch et al., 2010). It is not generally possible to give a closed-form expression for $p(\mathbf{d}|\mathbf{r}_A)$ based solely on the global geophysical and rock-physical likelihoods $p(\mathbf{d}|\mathbf{m})$ and $p(\mathbf{m}|\mathbf{r})$. Essentially, one would have to go through the following rewrite of equation 1:

$$\begin{aligned} p(\mathbf{r}_A|\mathbf{d}) &\propto \iint p(\mathbf{d}|\mathbf{m})p(\mathbf{m}|\mathbf{r})p(\mathbf{r}_{\mathbb{A}\setminus\mathcal{A}}|\mathbf{r}_A) \mathbf{d}\mathbf{m} \mathbf{d}\mathbf{r}_{\mathbb{A}\setminus\mathcal{A}} p(\mathbf{r}_A) \\ &= \iint p(\mathbf{d}|\mathbf{m})p(\mathbf{m}|\mathbf{r})p(\mathbf{r}) \mathbf{d}\mathbf{m} \mathbf{d}\mathbf{r}_{\mathbb{A}\setminus\mathcal{A}}, \end{aligned} \quad (2)$$

where $\mathbb{A}\setminus\mathcal{A}$ denotes all cells in \mathbb{A} except \mathcal{A} . The dimensions of these integrals depend on the number of cells in \mathbb{A} and whether the geophysical and rock-physical models possess independence between certain cells. For realistic problems, these are usually at least 100-dimensional, and in the densest cases, they might be of dimension 10^6 or more. Thus, we cannot tackle this problem directly via equations 1 and 2.

Instead of attempting to work with the global geophysical and rock-physical likelihoods, our approach aims at modeling only the part that is most relevant for the target variable \mathbf{r}_A . Let us thus introduce the local subsets \mathcal{B} , \mathcal{C} , and \mathcal{D} , which are sets of region cells reflecting the modeled part of respectively the rock properties \mathbf{r} , the geophysical properties \mathbf{m} , and the geophysical data \mathbf{d} in the local inversion for cell \mathcal{A} . Their corresponding variable sets \mathbf{r}_B , \mathbf{m}_C , and \mathbf{d}_D are named, respectively, the neighborhood variable, influence variable, and local data. To set the idea of the local subsets straight, the bulleted list below and illustration in Figure 2 provide basic guidelines for how these may be specified in an AVO data setting with vertical dependence:

- All local subset variables should be centered in cell \mathcal{A} .
- \mathcal{D} should include the cells for which the data \mathbf{d} are influenced by \mathbf{m}_A , that is, half a wavelet length above and below \mathcal{A} .
- \mathcal{C} should include the cells for which the geophysical properties \mathbf{m} influence the local data \mathbf{d}_D , i.e., one wavelet length above and below \mathcal{A} .
- \mathcal{B} should have size at least in order of the tuning thickness. This ensures that \mathbf{r}_B is the main source of variability for the data interfering with the contribution from \mathcal{A} .

Note, however, that our approach is not restricted to vertically defined local subsets; that is, lateral dependence may in principle also be modeled by our approach. The task of selecting the local subsets will be discussed in more depth later on.

Referring to Bayes' formula as in equation 1, our approximation approach is based on the relation

$$p(\mathbf{r}_A|\mathbf{d}_D) = \int p(\mathbf{r}_B|\mathbf{d}_D) \mathbf{d}\mathbf{r}_{\mathcal{B}\setminus\mathcal{A}} \propto \int p(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B) \mathbf{d}\mathbf{r}_{\mathcal{B}\setminus\mathcal{A}}, \quad (3)$$

where $\mathcal{B}\setminus\mathcal{A}$ denotes all cells in \mathcal{B} except \mathcal{A} . Replacing $p(\mathbf{d}_D|\mathbf{r}_B)$ (henceforth referred to as the local likelihood) by a Gaussian approximation yields the following integral form approximation:

$$p(\mathbf{r}_A|\mathbf{d}) \approx p^*(\mathbf{r}_A|\mathbf{d}_D) \propto \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B) \mathbf{d}\mathbf{r}_{\mathcal{B}\setminus\mathcal{A}}, \quad (4)$$

where the Gaussian approximation is given by

$$p^*(\mathbf{d}_D|\mathbf{r}_B) = N_{\mathbf{d}_D} \left(\boldsymbol{\mu}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B), \boldsymbol{\Sigma}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B) \right). \quad (5)$$

Table 1. Important quantities.

Symbol	Description
\mathbf{d}	Vector of geophysical data for all cells in the gridded interest region
\mathbf{m}	Vector of geophysical properties for all cells in the gridded interest region
\mathbf{r}	Vector of the rock property of interest for all cells in the gridded interest region
\mathbb{A}	Gridded interest region
\mathcal{A}	Cell in \mathbb{A} under consideration
\mathcal{B}	Subset of region cells for which the rock properties are modeled in the local inversion for \mathbf{r}_A
\mathcal{C}	Subset of region cells for which the geophysical properties are modeled in the local inversion for \mathbf{r}_A
\mathcal{D}	Subset of region cells for which the geophysical data are included in the local inversion for \mathbf{r}_A
\mathbf{r}_A	Target variable
\mathbf{r}_B	Neighborhood variable
\mathbf{m}_C	Influence variable
\mathbf{d}_D	Local data
\mathbf{G}	Matrix representing the impact \mathbf{m} has on the mean of \mathbf{d}
$\boldsymbol{\Sigma}_e$	Covariance matrix of the geophysical likelihood

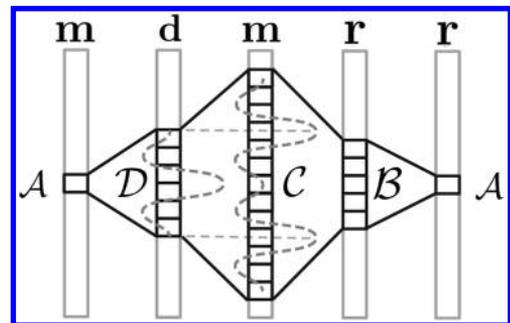


Figure 2. Illustration of sensible selection of the local subsets \mathcal{B} , \mathcal{C} , and \mathcal{D} for seismic AVO data with vertical dependence.

As will become clear in the following subsections, the Gaussian distribution in equation 5 will be established by merging a Gaussian approximation to the local rock-physical likelihood $p(\mathbf{m}_C|\mathbf{r}_B)$ with a Gaussian approximation to the local geophysical likelihood $p(\mathbf{d}_D|\mathbf{m}_C)$. Based on the integral form in equation 4, we define a weighted Monte Carlo routine enabling us to approximate in principle any quantities of interest related to $p(\mathbf{r}_A|\mathbf{d})$ (such as the mean, variance, probability of a certain event, or the complete density) by properly aggregating weighted samples from the prior of the target variable. The next three subsections present the details of the approximations for the local rock-physical and geophysical likelihoods, in addition to the proposed weighted Monte Carlo routine.

Local rock-physical likelihood

There is in general no simple description of the spatial distribution of the rock-physical likelihood $p(\mathbf{m}|\mathbf{r})$. To obtain a Gaussian approximation to the required local rock-physical likelihood $p(\mathbf{m}_C|\mathbf{r}_B)$, we therefore take a flexible sampling-based approach. This approach only requires that we are able to sample “pairs” $(\mathbf{m}_C, \mathbf{r}_B)$ from their joint distribution. The objective is to use these samples to fit the best possible mean and covariance matrix functions in a Gaussian approximation of the form

$$p^*(\mathbf{m}_C|\mathbf{r}_B) = N_{\mathbf{m}_C}(\boldsymbol{\mu}_{\mathbf{m}_C|\mathbf{r}_B}^*(\mathbf{r}_B), \boldsymbol{\Sigma}_{\mathbf{m}_C|\mathbf{r}_B}^*(\mathbf{r}_B)). \quad (6)$$

Note that a Gaussian approximation to $p(\mathbf{m}_C|\mathbf{r}_B)$ is less restrictive than the Gaussian approximation to the unconditional distribution of the geophysical properties $p(\mathbf{m})$ used in Buland and Omre (2003), Larsen et al. (2006), and related work. In principle, the mean function $\boldsymbol{\mu}_{\mathbf{m}_C|\mathbf{r}_B}^*(\mathbf{r}_B)$ and covariance function $\boldsymbol{\Sigma}_{\mathbf{m}_C|\mathbf{r}_B}^*(\mathbf{r}_B)$ could behave in completely unrestricted ways. In order not to make the fitting procedure too complicated, we do, however, suggest to divide the sampled pairs into K different nonoverlapping classes according to some specified criterion on \mathbf{r}_B . Within each such class k , separate mean functions $\boldsymbol{\mu}_{\mathbf{m}_C|\mathbf{r}_B,k}^*(\mathbf{r}_B)$ are fitted by a regression procedure, and the resulting residuals $\boldsymbol{\epsilon}' = \mathbf{m}_C - \boldsymbol{\mu}_{\mathbf{m}_C|\mathbf{r}_B,k}^*(\mathbf{r}_B)$ are used to estimate a fixed covariance matrix $\boldsymbol{\Sigma}_{\mathbf{m}_C|\mathbf{r}_B,k}^*$ for that class. Using this procedure, the mean function has an unrestricted dependence on the neighborhood variable \mathbf{r}_B , whereas the covariance matrix depends categorically on the class k of \mathbf{r}_B . Hence, the criterion that divides the samples into different classes should be chosen such that the dependence structure within the influence variable \mathbf{m}_C is fairly stable.

Although a simple linear regression (least squares) method may be used to fit $\boldsymbol{\mu}_{\mathbf{m}_C|\mathbf{r}_B,k}^*(\mathbf{r}_B)$ in the above procedure, we suggest using a technique that allows for increased fidelity in more complex situations. Examples are multivariate adaptive regression splines (MARS; Friedman, 1991), projection pursuit regression (Friedman and Stuetzle, 1981), neural networks (Cheng and Titterton, 1994), and generalized additive models (Hastie and Tibshirani, 1986). Flexibility is essential here because it allows the approximated dependence on \mathbf{r}_B to match that of the true model more closely. Also, the larger the sample size the more stable the approximations become.

One strength of the sampling-based approach is that the distribution fit may be checked by standard multivariate normality tests (Henze, 2002). If tests deem the Gaussian model acceptable, the approximations are guaranteed to be good. If not, one may attempt to correct for the non-Gaussianity by using a more flexible regression

procedure, increase the number of classes, redefine the local subsets, or reduce the influence of the outliers. The last option may for instance be done by using a range spanning covariance estimation routine to estimate $\boldsymbol{\Sigma}_{\mathbf{m}_C|\mathbf{r}_B,k}^*$ as opposed to using the standard sample covariance. Such a routine stretches the tails of the Gaussian model by using an estimate of the covariance that spans broadly enough for no sampled pairs $(\mathbf{m}_C, \mathbf{r}_B)$ to be very unlikely under the fitted model. This method weakens the impact of deviations from the Gaussian model's dependence structure, and it reduces to the standard sample covariance if the fit is already good. The suggested routine is outlined in Appendix A.

Local geophysical likelihood

The global geophysical likelihood model typically takes the form $p(\mathbf{d}|\mathbf{m}) = N_{\mathbf{d}}(\mathbf{G}\mathbf{m}, \boldsymbol{\Sigma}_{\mathbf{e}})$, where \mathbf{G} is a matrix of the appropriate dimension representing the linear dependence of the geophysical data \mathbf{d} on the geophysical properties \mathbf{m} . That is

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \boldsymbol{\epsilon}, \quad (7)$$

with $\boldsymbol{\epsilon}$ some error term with distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{e}})$. We seek an approximation for the local geophysical likelihood $p(\mathbf{d}_D|\mathbf{m}_C)$, where

$$\mathbf{d}_D \approx \tilde{\mathbf{G}}\mathbf{m}_C + \tilde{\boldsymbol{\epsilon}}, \quad (8)$$

with $\tilde{\mathbf{G}}$ corresponding to \mathbf{G} above, and $\tilde{\boldsymbol{\epsilon}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\epsilon}}})$. However, extracting the local part of equation 7 gives $\mathbf{d}_D = \mathbf{G}_D\mathbf{m} + \boldsymbol{\epsilon}_D$, where \mathbf{G}_D is the submatrix of \mathbf{G} containing only the rows corresponding to D . As illustrated upon introduction of the local subsets, \mathcal{C} is chosen as the region whose geophysical properties influence the local data (the most). Hence, it is reasonable to approximate $\mathbf{G}_D\mathbf{m}$ by $\mathbf{G}_{D,\mathcal{C}}\mathbf{m}_C$, where $\mathbf{G}_{D,\mathcal{C}}$ contains the columns of \mathbf{G}_D corresponding to the cells in \mathcal{C} . The sought-after approximation is consequently obtained by letting $\tilde{\mathbf{G}} = \mathbf{G}_{D,\mathcal{C}}$ and $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\epsilon}}} = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_D}$ in relation 8.

By a result in Appendix B, the established local rock-physical and geophysical likelihood approximations give a fully specified local likelihood approximation $p^*(\mathbf{d}_D|\mathbf{r}_B)$ as in equation 5 with

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B) &= \mathbf{G}_{D,\mathcal{C}}\boldsymbol{\mu}_{\mathbf{m}_C|\mathbf{r}_B,k}^*(\mathbf{r}_B), \\ \boldsymbol{\Sigma}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B) &= \mathbf{G}_{D,\mathcal{C}}\boldsymbol{\Sigma}_{\mathbf{m}_C|\mathbf{r}_B,k}^*\mathbf{G}_{D,\mathcal{C}}^T + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_D}, \end{aligned} \quad (9)$$

for each \mathbf{r}_B in class k .

Weighted Monte Carlo routine

The final part of the framework concerns the weighted Monte Carlo routine, which approximates inversion quantities of interest. The routine relies on the integral in relation 4 with $p^*(\mathbf{d}_D|\mathbf{r}_B)$ as specified by equations 5 and 9. The routine goes as follows:

- 1) Sample a large number L of \mathbf{r}_B -variables from its prior $p(\mathbf{r}_B)$.
- 2) For each sample $\mathbf{r}_B^{(l)}$, compute the mean $\boldsymbol{\mu}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B^{(l)})$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B^{(l)})$ of the local likelihood approximation from the formulae in equation 9.
- 3) For each sample $\mathbf{r}_B^{(l)}$, use the computed $\boldsymbol{\mu}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B^{(l)})$ and $\boldsymbol{\Sigma}_{\mathbf{d}_D|\mathbf{r}_B}^*(\mathbf{r}_B^{(l)})$ to evaluate the approximate local likelihood $p^*(\mathbf{d}_D|\mathbf{r}_B = \mathbf{r}_B^{(l)})$.

- 4) For each sample $\mathbf{r}_B^{(l)}$, extract $\mathbf{r}_A^{(l)}$ and define unnormalized and normalized weights by respectively

$$v^{(l)} = p^*(\mathbf{d}_D | \mathbf{r}_B = \mathbf{r}_B^{(l)}) \quad \text{and} \quad w^{(l)} = \frac{v^{(l)}}{\sum_{j=1}^L v^{(j)}}. \quad (10)$$

Each pair $(\mathbf{r}_A^{(l)}, w^{(l)})$, $l = 1, \dots, L$ may then be used to approximate in principle any quantity related to $p(\mathbf{r}_A | \mathbf{d})$. Some examples are:

- 1) $\boldsymbol{\mu}^*(\mathbf{r}_A | \mathbf{d}) = \sum_{l=1}^L w^{(l)} \mathbf{r}_A^{(l)}$.
- 2) $p^*(\mathbf{r}_A \in S | \mathbf{d}) = \sum_{l=1}^L w^{(l)} \mathbf{1}_{\{\mathbf{r}_A^{(l)} \in S\}}$ for some set $S \in (-\infty, \infty)$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.
- 3) For \mathbf{r}_A with discrete prior distribution $p(\mathbf{r}_A)$: $p^*(\mathbf{r}_A | \mathbf{d}) = \sum_{l=1}^L w^{(l)} \mathbf{1}_{\{\mathbf{r}_A^{(l)} = \mathbf{r}_A\}}$.
- 4) For \mathbf{r}_A with continuous prior distribution $p(\mathbf{r}_A)$: $p^*(\mathbf{r}_A | \mathbf{d}) = \sum_{l=1}^L w^{(l)} K_h(\mathbf{r}_A - \mathbf{r}_A^{(l)})$, where K_h is a scaled kernel density function with bandwidth h (see, e.g., Silverman, 1986).

In some cases, it is fruitful to sample from the prior conditioned on some criterion, rather than directly. This is especially the case when the interesting parts of the sample space are a priori unlikely or naturally separated in, e.g., a discrete and continuous part. Appendix C gives more details on this subject.

Full-region inversion

Two requirements must be met to use our approximation framework. First, relation 8 must hold with a Gaussian error term $\tilde{\mathbf{e}}$, at least approximately. Second, it must be possible to sample from $p(\mathbf{m}_C, \mathbf{r}_B)$. Sampling from $p(\mathbf{r}_B)$ is ensured by the latter requirement.

A major advantage of this framework first becomes apparent when considering inversion of a larger region. As mentioned earlier, inversion over a gridded region \mathbb{A} is carried out by applying the presented technique to each grid cell $\mathcal{A} \in \mathbb{A}$. This allows for parallelization that under the following additional stationarity assumptions results in a computationally efficient inversion procedure:

- $\mathcal{B} = \mathcal{B}(\mathcal{A})$, $\mathcal{C} = \mathcal{C}(\mathcal{A})$, $\mathcal{D} = \mathcal{D}(\mathcal{A})$ are all specified with relation to \mathcal{A} only.
- $p(\mathbf{d}_{\mathcal{D}(\mathcal{A})}, \mathbf{m}_{\mathcal{C}(\mathcal{A})}, \mathbf{r}_{\mathcal{B}(\mathcal{A})})$ is stationary with respect to \mathcal{A} .

These assumptions ensure that the local subsets follow \mathcal{A} when it shifts from one cell to another and that the joint distribution of the local subsets is independent of this shift. The mean and covariance matrix functions of the local likelihood approximation in equation 9 thus hold for all $\mathcal{A} \in \mathbb{A}$ and need to be computed only once. The only part of the procedure that changes from one position to another is the local geophysical data \mathbf{d}_D . Hence, global inversion of \mathbb{A} may be carried out by simply repeating steps 3 and 4 in the above Monte Carlo routine for each cell $\mathcal{A} \in \mathbb{A}$. The stationarity assumptions are not strictly required for our method to work out but are introduced for computational speed-up. There is speed-up also if the assumptions hold only for certain parts of \mathbb{A} .

Selecting local subsets

Selecting appropriate local subsets \mathcal{B} , \mathcal{C} , and \mathcal{D} is essential in this approximation framework. The dimension of these is an issue of approximation accuracy versus computation speed, but it is not

necessarily so that choosing them to be larger independently of each other will lead to a better approximation. A sensible and efficient approximation framework thus requires careful selection of these subsets.

By narrowing \mathbf{d} to the local data \mathbf{d}_D , we limit the information used for the local likelihood evaluation. Using too little data gives a considerable loss of information, but using too much data leads to infeasible computation. Also, including data with minor relation to the target variable \mathbf{r}_A does not bring anything new. Selecting \mathcal{D} to be the region directly influenced by the geophysical properties in \mathcal{A} gives a reasonable trade off. For other geophysical models, this region is not as easily determined as for seismic amplitude data. That situation is discussed in Appendix D.

As mentioned earlier, \mathcal{C} should be the region for which the geophysical properties influence the local data \mathbf{d}_D . Expanding \mathcal{C} beyond this would not improve the fit, but only make the task of fitting $p^*(\mathbf{m}_C | \mathbf{r}_B)$ more complex.

A common assumption within the geosciences is that all that could be learned about local geophysical properties from rock properties are found in the local variables; that is, $p(\mathbf{m}_X | \mathbf{r}) = p(\mathbf{m}_X | \mathbf{r}_X)$ for any set of region cells \mathcal{X} . Based on this relation, it is clear that it would be optimal to set \mathcal{B} equal to whatever \mathcal{C} is set to. Because the dimension of \mathcal{B} is exactly the dimension of the integral that the weighted Monte Carlo routine relies on, practicalities usually make such a choice impossible. The reason is that to maintain the accuracy of the Monte Carlo routine when the dimension of the integrand increases, a larger number of prior samples, and hence likelihood evaluations, is required. However, assuming that the Monte Carlo accuracy is maintained (at a higher computational cost), it is likely that expanding \mathcal{B} would lead to a better approximation of the true posterior and, consequently, more accurate predictions of the rock properties and their uncertainties. If \mathcal{B} is too small, important features of the local data \mathbf{d}_D , which are “transferred” to the influence variable \mathbf{m}_C , may be caused by characteristics of rock properties outside \mathcal{B} and thereby not be associated with appropriate values of the target variable \mathbf{r}_A .

The local subsets may also be selected by comparing the properties of samples from the local likelihood approximation $p^*(\mathbf{d}_D | \mathbf{r}_B)$ obtained for different choices of \mathcal{B} , \mathcal{C} and \mathcal{D} to those of the true local likelihood $p(\mathbf{d}_D | \mathbf{r}_B)$. Alternatively, confidence in certain local subsets may be built based on application to a synthetic case in which the resulting approximated posterior distribution can be compared with the true rock properties. We rely on this latter approach for selection of \mathcal{B} in the upcoming data illustrations.

SYNTHETIC DATA TEST

In the next section, we will consider a 4D survey from the Sleipner CO₂ injection project as a real data example. We shall use that case as a motivation in this synthetic example. The synthetic data shall reflect a region similar to the Utsira Formation in the Sleipner field offshore Norway, where CO₂ has been injected for storage and seismic base and monitor surveys have been conducted. The CO₂ is typically trapped underneath thin layers of shale within the formation or under the formation top. The main objective is to “map” the CO₂ based on the saturation in the region. Hence, we define the rock property of interest \mathbf{r} as the CO₂ saturation at the time point of one of the monitor surveys (herein for simplicity referred to as the saturation), assuming no injection prior to the base survey. The geophysical properties \mathbf{m} are defined as the change in the logarithm of

the P- and S-wave velocity and density from base to monitor time. Finally, \mathbf{d} denotes the change in properly aligned seismic angle gathers from base to monitor time.

In this synthetic example, we shall be content working with models and data specified on a gridded 2D region \mathbb{A} of size 3500 m \times 280 ms (length \times two-way traveltime [TWT]), where each of the 19,600 grid cells have size 25 m \times 2 ms. We shall perform inversion of synthetic geophysical data to saturation for the complete region \mathbb{A} . The synthetic case is constructed by specifying global stochastic rock-physics and geophysical models in this region. Note in particular that the stochastic rock-physics model involves also other rock properties than the saturation \mathbf{r} .

Let us first consider the rock-physics model. There are few reliable measurements of velocity and density from the injection well (Rabben and Ursin, 2011). The problems with the logging are probably due to issues with the very loose sand in Utsira. The P-wave velocity in brine-filled sand is slightly more than 2000 m/s, and

Table 2. Rock-physics model parameters. The mineral parameters have correlation 0.99 between properties. Details on the two and four parameter beta distributions are given in Appendix B.

Property	Distribution
Mineral bulk modulus (GPa)	N(35.4, 3.2)
Mineral shear modulus (GPa)	N(27.3, 7.4)
Mineral density (g/ccm)	N(2.647, 0.008)
Brine bulk modulus (GPa)	Fixed = 2.538
Brine density (g/ccm)	Fixed = 1.027
CO ₂ bulk modulus (GPa)	Fixed = 0.065
CO ₂ density (g/ccm)	Fixed = 0.686
Coordination number	Fixed = 7.3
Friction factor	Beta(5.0, 0.8)
Porosity	Beta ₄ (2.0, 2.0, 0.27, 0.42)
Pore pressure (MPa)	Fixed = 10
Effective pressure (MPa)	Fixed = 10
Temperature (°C)	Fixed = 36

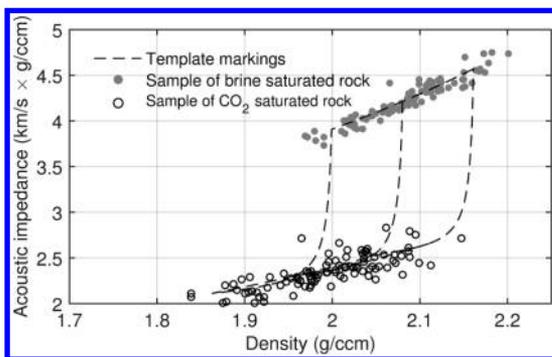


Figure 3. Rock-physics template: The boundaries of the template correspond to porosity of 40% and 30% and CO₂ saturation of 0 and 1. The center lines of the template correspond to porosity and CO₂ saturation of, respectively, 35% and 0.5. For saturation, that line is partly hidden close to the lower boundary of the template.

the rock matrix in this region corresponds to loose sand. The rock-physics model we use for the Utsira sand is consistent with these data. In our model, we use a Reuss mix of the mineral point and a high-porosity member constructed using Walton's model with 45% porosity (see, e.g., Mavko et al., 2009). The parameters and corresponding uncertainty model for this rock-physics model are given in Table 2, which also provides the temperature and pressures used to derive the fluid properties. The rock-physics model is similar to the model used by Arts et al. (2004), but a notable difference is that we use the velocity and density for brine, which are compatible with the most recent pressure and temperature measurements in the formation (Batzle and Wang, 1992; Alnes et al., 2011). When matching the observed velocity in Utsira, this gives softer sand than is used in Arts et al. (2004). The effect of the saturation on geophysical properties is computed using fluid substitution, that is, Gassmann's equations (Mavko et al., 2009) where a homogeneous fluid mix (Reuss) is assumed. Due to the nature of CO₂ and the soft rock at the Sleipner injection site, saturation is the main cause of variability for the change in geophysical properties from base to monitor time. This is illustrated in Figure 3, which displays samples of brine- and CO₂-saturated rocks from the rock-physics model, overlaid on a rock-physics template.

The geophysical likelihood model is of the form described by equation 7, i.e., Gaussian with linear mean function. The linear multiplicand here is $\mathbf{G} = \mathbf{WAD}$; \mathbf{W} is a block diagonal matrix representing the smoothing with a 25 Hz Ricker wavelet; \mathbf{A} is the matrix of weak-contrast coefficients for each of the offset angles 5° (near), 20° (mid), and 35° (far) as defined by Aki and Richards (1980); and \mathbf{D} is a differential matrix producing contrasts of the geophysical properties. The covariance matrix Σ_e for the error term is block diagonal with independence between the different offsets, and the standard deviations are 0.04 for near, 0.05 for mid, and 0.06 for far offset.

The synthetic saturation for our region of interest \mathbb{A} is constructed to mimic the high-saturation layers in the Utsira Formation. It contains multiple layers with a flat top and variable thickness. In particular, 5% of the region cells possess saturation greater than 0.6. The synthetic saturation and noisy geophysical data sampled from the rock model are displayed in Figure 4. Note in particular that the peak amplitude of the seismic data does not follow a flat top.

Let us now turn to the inversion, in which we will limit our approximations to include vertical dependence. We shall define a stochastic prior model for the saturation, and otherwise we use vertical analogs of the rock-physics and geophysical models described above as the basis for our approximations. The prior model for saturation is defined as a transformation of a Gaussian copula (Joe, 1997), which models the effect that adjacent cells are more correlated than distant ones. The marginal distribution in each cell has a point mass at zero and a continuous distribution from 0 to 1. In each cell, the prior probability for the saturation being zero is 99%, whereas the saturation is distributed as Beta(6, 1.5), as shown in Figure 5, when it is strictly positive (because the distribution of saturation has a discrete and a positive part, this example illustrates the applicability of our approach to both these types of rock properties). Consult Appendix B for details on the beta distribution.

The Gaussian copula has a stationary exponential covariance function with range parameter $R = 50$ ms in the parametrization $C(h) = \exp(-3|h|/R)$, where h is the vertical "distance" (in ms) between two locations.

Because we limit our approximations to only include vertical dependence, the local subsets \mathcal{B} , \mathcal{C} , and \mathcal{D} only include variables within the same vertical profile as \mathcal{A} . They are also all centered in \mathcal{A} . Specifically, we let \mathcal{D} include 20 ms above and below \mathcal{A} , having dimension 21, whereas \mathcal{C} includes 44 ms above and below \mathcal{A} , having dimension 45. As we shall see shortly, \mathcal{B} is selected on a trial and error basis through comparison with the true synthetic saturation.

In our approximations, we estimate $p^*(\mathbf{m}_C|\mathbf{r}_B)$ by splitting the sampled pairs $(\mathbf{m}_C, \mathbf{r}_B)$ into $K = 4$ classes depending on whether the saturation in the two boundaries of \mathcal{B} (i.e., the shallowest and deepest cells in \mathcal{B}) are zero or strictly positive. The mean functions $\mu_{\mathbf{m}_C|\mathbf{r}_B, k}^*$ for each class k are estimated by a MARS procedure, where generalized cross validation is used to specify the tuning parameters in the regression. The covariance matrices $\Sigma_{\mathbf{m}_C|\mathbf{r}_B, k}^*$, $k = 1, \dots, 4$ are estimated by the range spanning covariance estimation routine in Appendix A.

Between 45,000 and 100,000 samples are used for each of the $K = 4$ classes to fit the approximate local rock-physical likelihood $p^*(\mathbf{m}_C|\mathbf{r}_B)$. Because nonzero saturation occurs in only 1 out of 100 cells by direct sampling, it is beneficial to oversample positive \mathbf{r}_B . This will give more robust approximations for the whole sample space of the target variable \mathbf{r}_A . Hence, two sets of \mathbf{r}_B samples (one conditioned on $\mathbf{r}_A = 0$ and one conditioned on $\mathbf{r}_A > 0$) are used in the weighted Monte Carlo routine (see Appendix C for further details). This sampling procedure, with each of the two sets being of size 10^5 , will be used throughout the paper.

To fully specify the inversion, we must choose the size of \mathcal{B} for this synthetic case. (The chosen size will also be passed forward to the upcoming real case.) We do this by pointwise comparing predictors derived from the local inversions with the true synthetic saturation. For that comparison, and for the remainder of this paper, we will use the approximated marginal posterior mean saturation in each cell as a predictor for the true unknown saturation. Increasing the dimension of \mathcal{B} should in theory (on average) result in a closer match between the truth and the approximation, represented by the bias. At the same time, it increases the variability between the performances of different sets of prior samples, represented by the (Monte Carlo) variance. We define the optimal \mathcal{B} as the one minimizing the mean squared error (MSE), which decomposes nicely into squared bias plus variance. Empirical versions of squared bias, variance, and MSE averaged over the cells in the full 2D region are shown in Figure 6 for varying size of \mathcal{B} and constant sample size — when using the posterior mean as predictor. Note that the MSE is computed on a cell-by-cell basis; hence, it penalizes any minor misalignment in the predicted saturation severely. In terms of the bias versus variance trade-off (i.e., MSE minimization), we deem \mathcal{B} of dimension 17 the optimal for the current sampling regime. That is, the optimal \mathcal{B} includes 16 ms above and below \mathcal{A} . As seen from the figure, the squared bias is fairly flat to the right of $\dim(\mathcal{B}) = 17$. This indicates that increasing the sample size further would not improve the fit considerably. The slight increase in bias for the largest dimensions is an indirect effect of the relatively small sample size.

Figure 7 shows the marginal posterior means for the chosen \mathcal{B} and explained sampling regime, along with its difference from the synthetic truth. As seen from the figure, the posterior mean matches the true saturation well over the complete 2D region, at thicker and thinner layers of high saturation. Note the good positioning of the top and base of high-saturation layers also when the layer thickness

is below the tuning thickness. As expected, the positioning of thin layers far below the tuning thickness results in some misalignment of the high-saturation layers. They are, however, still detected, and vertical profile averages are generally well preserved. Compared with using the prior directly without support of the data (MSE = 0.038), our framework reduces the empirical MSE almost an order of magnitude (MSE = 0.0050). Also, the prior average saturation in

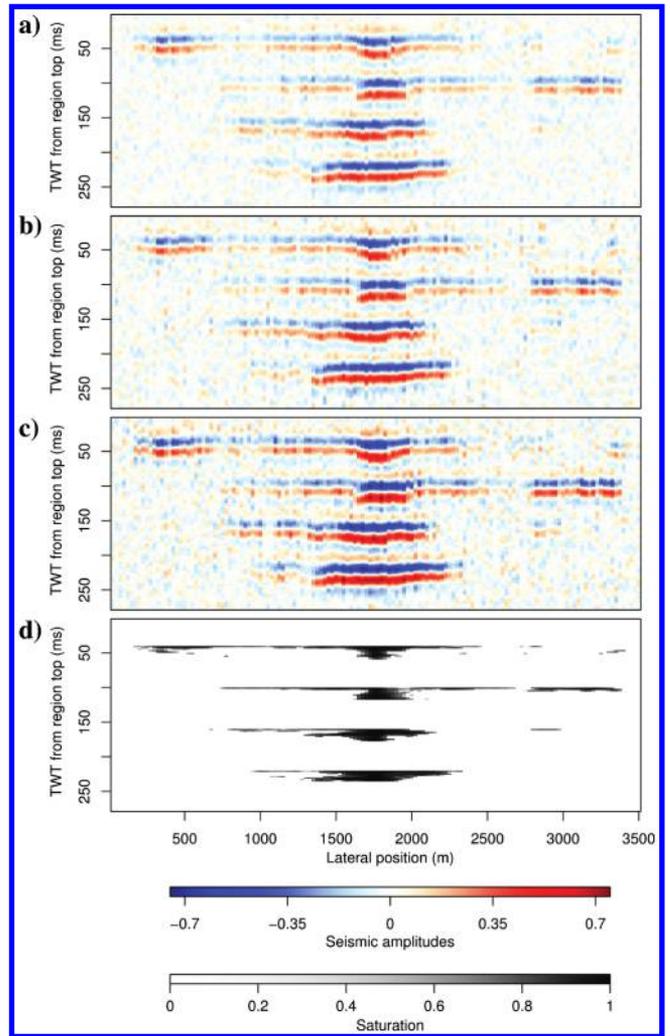


Figure 4. (a-c) Synthetic seismic difference data (near, mid, and far) relevant for the 2D region of interest. (d) Generated synthetic saturation at monitor time \mathbf{r} for the complete synthetic 2D region.

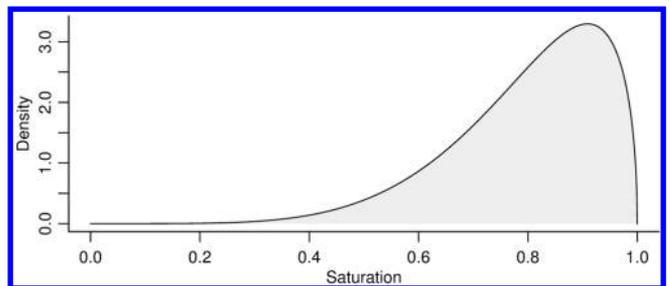


Figure 5. Marginal prior distribution of positive saturation: A beta distribution with shape parameters $a = 6$ and $b = 1.5$, having mean 0.8 and variance approximately 0.02.

the region is 0.0080, whereas the posterior regionwise average is 0.0455. These should be compared with the regionwise average of the true synthetic saturation, which is 0.0449. Thus, even if the approach operates locally, it still gives information about bulk properties. These results are not only an outcome of our framework and procedure, but they also depend on the statistical model and grid being used. Because the same grid is used to generate the synthetic case and perform inversion, potential bias caused by grid cells being misaligned with high-saturation layers are not present in this synthetic case. Also, the selection bias is not accounted for when the same data are used to tune a parameter and present performance results. However, the flat behavior of the MSE curve around the minimum point in Figure 6 indicates that such selection bias is insignificant here.

To properly evaluate the accuracy of our approximation method, the approximated marginal posterior distributions should be compared with the true posterior distribution. Because “exact” methods for

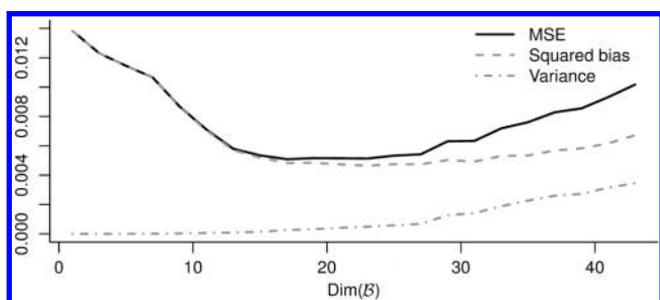


Figure 6. Empirical estimates of the MSE, squared bias, and Monte Carlo variance are plotted for different dimensions of \mathcal{B} in the synthetic 2D case. These are computed based on five different prior sampling seeds (with totally 2×10^5 samples) for each \mathcal{B} centered in \mathcal{A} .

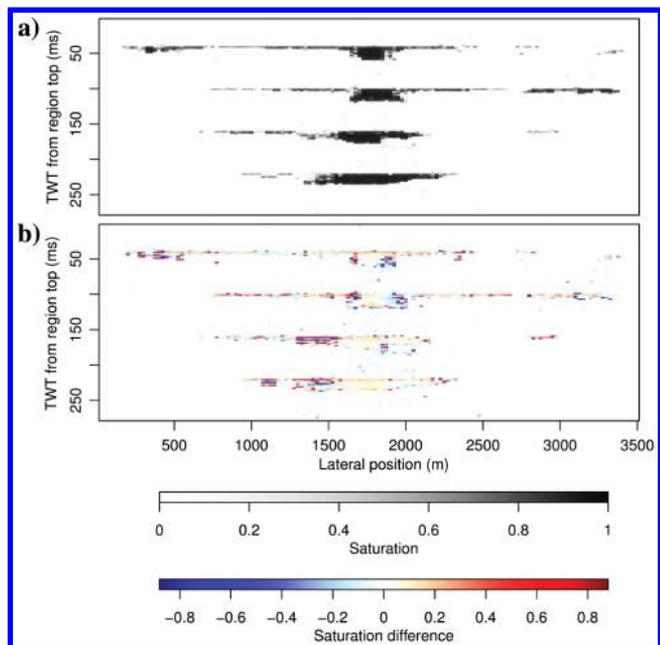


Figure 7. (a) Inversion results for the synthetic 2D region shown through approximated marginal posterior mean saturations in each cell. (b) The difference between the true synthetic saturation and the prediction in panel (a).

computing the posterior distribution are computationally extremely costly, this is not feasible for the full 2D region. To accompany the above evaluation and comparison with the actual synthetic saturation, we do however evaluate the true posterior distribution in a single vertical profile (positioned at 1625 m) for comparison with our procedure. The true posterior for the vertical trace is obtained by running a blockwise Metropolis Hastings MCMC scheme (Bolstad [2009], chapter 6.3) with an independence sampler corresponding to the conditional prior distribution. This “brute force” MCMC procedure required several days of CPU running time to provide reliable results — whereas our approach produced approximate results within seconds. For this vertical profile, Figure 8 shows pointwise 80% credibility intervals (CI) (the range between P10 and P90), posterior means and medians for the true posterior and our local approximation procedure with a few different sizes of \mathcal{B} . Also plotted are the seismic difference data and true synthetic saturation. The figure illustrates the typical behavior for our method when varying the size of \mathcal{B} , while the sample size is kept constant. When \mathcal{B} is too small, the approximated posteriors are simplified too much. Increasing the size of \mathcal{B} increases the level of detail, but a too-large \mathcal{B} generates unwanted noise with thin wrongly predicted high-saturation layers and gives unstable results due to the relatively small sample size. For the MSE-optimized procedure, where $\dim(\mathcal{B}) = 17$, the posterior mean, median, and the pointwise 80% credibility intervals match those of the true posterior very well. This indicates that the approximation method works as intended also on smaller scales.

Finally, we compare predictions from our procedure with the corresponding ones based on the frequently used Gaussian inversion approach of Buland and Omre (2003). Because the approach of Buland and Omre (2003) actually is a geophysical inversion approach, we temporarily change our focus to prediction of the change in the logarithm of the density ρ , instead of saturation. Figure 9 shows the posterior means for the 2D region discussed above for both approximation methods along with the synthetic true change in log density. As seen from the figure, the predictions based on our method are more distinct and clear compared with the vaguer predictions provided by the Gaussian inversion. The latter perform poorly on thin layers of reduced density, and also predict areas with a positive change in log density not present in the synthetic data. Hence, our method improves substantially upon Buland and Omre (2003) in terms of prediction accuracy.

REAL DATA CASE

The Sleipner CO₂ injection project aims at storing CO₂ captured from the gas production in the Sleipner field offshore Norway by leading compressed CO₂ down to the Utsira Formation through an injection well. We consider geophysical data from a seismic 4D survey of this formation and aim at monitoring or mapping the CO₂ based on the saturation. We concentrate on the changes from a base survey in 1994 (before injection) and until a monitor survey in 2006. Because the saturation is effectively zero everywhere prior to injection, the changes in saturation correspond to the amount at monitor time. The geophysical data consist of changes in seismic AVO data from base to monitor time for three different offsets. The data are aligned using rms and pushdown data prior to difference computation. We concentrate on a west–east-directed 2D region intersecting the injection well. The region spans more than $2900 \text{ m} \times 334 \text{ ms}$, has a seismic sampling resolution of $25 \text{ m} \times 2 \text{ ms}$, and is positioned approximately 800 m below sea level.

The setup of the model to be approximated and its various parameters are essentially the same as for the synthetic data test, with the exception of some parameter specifications for the geophysical likelihood $p(\mathbf{d}|\mathbf{m})$. In the present case, \mathbf{W} consists of 35, 30, and 25 Hz Ricker wavelets for the near (12.5°), mid (25°), and far offsets (40°), respectively. The frequency content of the wavelets was set to match those in the seismic data. For near and far stacks, this was done based on the common part of the base and monitor surveys in a region directly above the reservoir. This part has a slightly lower frequency content than the individual parts, indicating a lower signal-to-noise ratio at higher frequencies. Processing issues made such detailed analysis impossible for the far stack. Hence, the far wavelet was set only from the frequency content in the base survey, accounting also for a high end frequency loss. The near and mid wavelets are scaled by a factor two compared with the far offset. This was determined by analysis of the base data, which gave twice as strong a signal for the near and mid stack than for the far — most likely caused by survey and processing effects. The standard deviation of the error term in the geophysical likelihood model is 0.5 for near and mid offsets and 0.2 for far offset. This is larger than observed directly above the reservoir, reflecting that there are larger alignment errors and stronger amplitude effects in the target region than directly above.

All other model parameters, local subsets, and other parts of the inversion setup are the same as for the synthetic case. The optimal size of the neighborhood variable established for the synthetic case is used in the sampling. The seismic AVO data relevant for the 2D region and the resulting marginal posterior means are shown in Figure 10. The results indicate several wide sections of increased saturation, which seem to generally match well with the reflections from the seismic data.

Our approach provides more than the best estimate. Figure 11 shows more detailed inversion results for a vertical profile near the injection well, together with the seismic amplitudes of that profile. The figure indicates that there are five main high-saturation layers with tops approximately at vertical positions 60, 90, 110, 145–160, and 190 ms. This matches well with a straightforward visual interpretation of the reflections of the seismic amplitudes. The

approximate posterior uncertainty is small at the high-saturation areas corresponding to the three shallowest and the deepest of these positions, indicating that the presence of these layers is fairly certain. On the other hand, the contiguously wide pointwise 80% credibility intervals reaching all the way down to zero for TWTs of 145–180 ms,

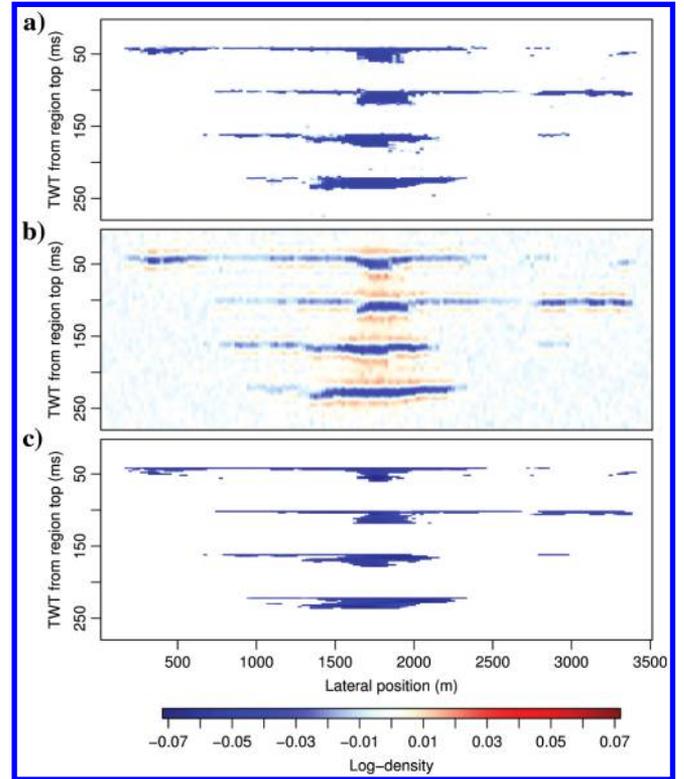


Figure 9. Comparison of predictions of changes in log density for the synthetic 2D region. (a) Predictions (means) based on our method with local subsets as above, (b) predictions (means) based on the Gaussian inversion approach (Buland and Omre, 2003), and (c) true synthetic change in log density.

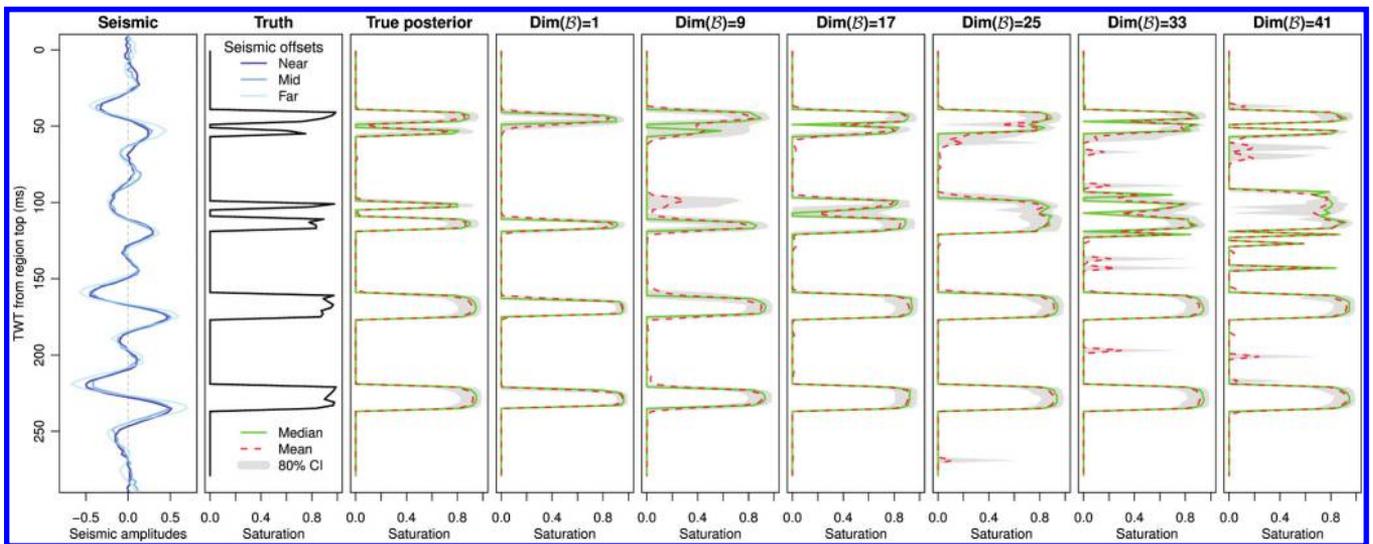


Figure 8. Seismic amplitudes, true saturation, and details of the true posterior and our approximation using different sizes of \mathcal{B} in a vertical profile of the synthetic data positioned at 1625 m. The 80% CI shows the pointwise range between the P10 and P90.

DISCUSSION

indicate larger uncertainty for the levels of saturation. Together with the disagreement of the mean and median predictors, this suggests that the presence and depth of one or two high-saturation layers are highly uncertain within this range.

Boait et al. (2012) study the CO₂ migration in the Utsira Formation after injection using seismic time-lapse data with several monitor surveys, and they present a qualitative interpretation of the horizons in a 2D region close to the injection point. The qualitative interpretations match our profile inversion results very well by indicating the same five layers. On a larger scale, the qualitative interpretations also match our inversion results well. Some deviations are, however, seen in an area slightly east of the injection well, and in the deepest areas where the signals in our data are weak.

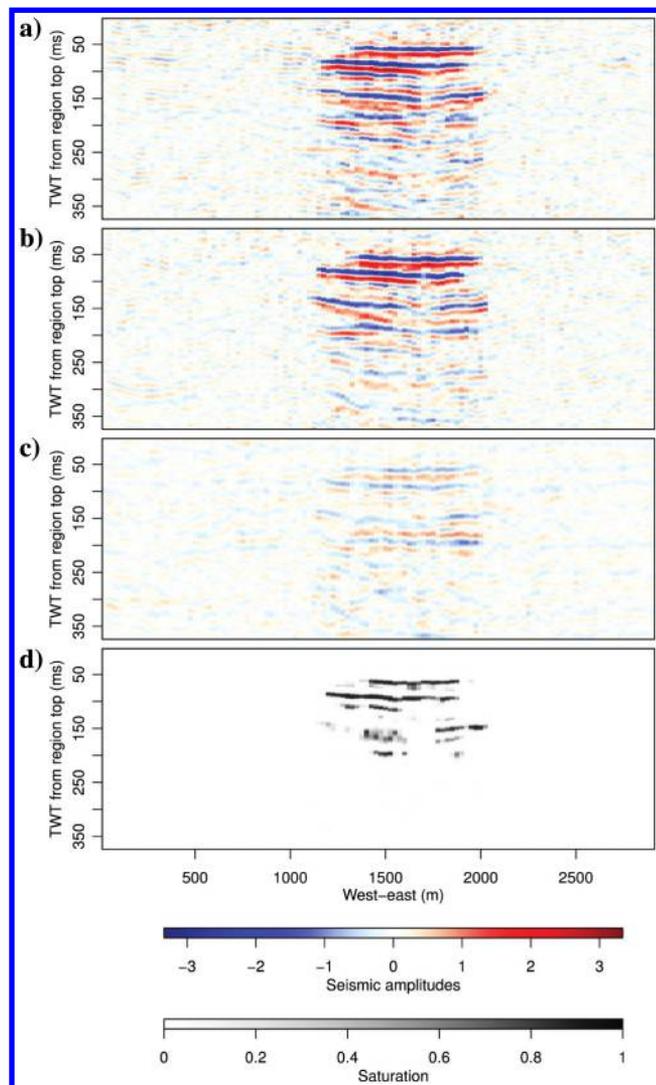


Figure 10. (a-c) The three different offsets (near, mid, and far) of the seismic AVO differences between base and monitor time relevant for the real case 2D region. The near and mid offsets are scaled by a factor two compared with the far offset. (d) Inversion results shown through marginal posterior mean saturations for each cell in the gridded 2D region of the real data case.

Our approach has connections to other general statistics-based and geophysical-motivated approximation techniques. Similar to our approach, the integrated nested Laplace approximation (INLA) method of Rue et al. (2009) approximates parts of the model by Gaussian distributions, uses their convenient properties, and solves a lower dimensional final integral by a numerical routine. The method does not involve local subset parameters, however. Those Gaussian approximations are also applied to different parts of the model than ours, and therefore require other types of assumptions — typically being unrealistic within geoscience applications. Of the existing geophysics types of techniques, the connection to Buland et al. (2008) is perhaps the closest. Like us, Buland et al. (2008) perform global inversion by focusing on one cell at a time and rely on Gaussian approximations. However, they only consider the direct behavior of the rock and geophysical properties in the cell under current consideration, and not the spatial neighbors, which is key in our framework. They do not require a Gaussian rock-physical likelihood (neither globally nor locally), but this also restricts the applicability of their technique to discrete rock properties, such as facies or lithology classes.

The overall procedure used in our framework can also be applied when $p(\mathbf{d}_D|\mathbf{r}_B)$ is approximated by a non-Gaussian distribution. There are, however, several benefits of relying on Gaussianity. Among them are the simple way that conditional distributions are handled, the computational savings available when preevaluating larger parts of the Gaussian distribution (under stationarity assumptions), and usage of the common assumption that the geophysical likelihood $p(\mathbf{d}|\mathbf{m})$ is Gaussian with a linear mean function. The

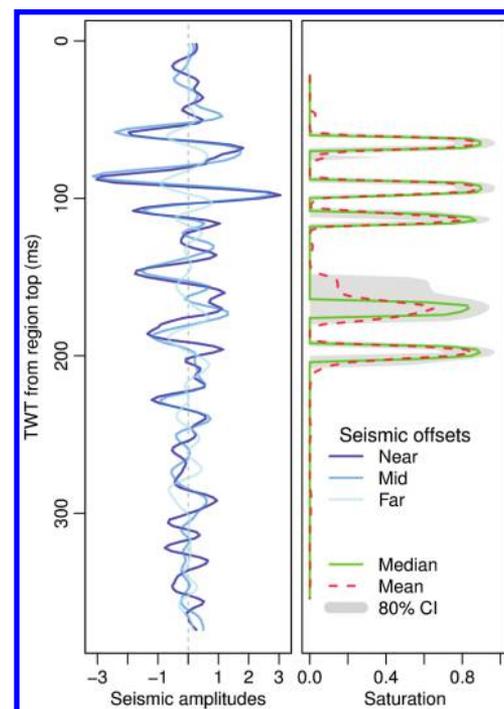


Figure 11. Seismic AVO data compared with the approximate marginal posterior distributions of saturation in a vertical profile close to the injection well (west-east position 1475 m in the real case 2D region).

Gaussian restriction can, however, be relaxed without losing all the beneficial properties. This is achieved by allowing the local geophysical likelihood $p(\mathbf{d}_D|\mathbf{m}_C)$ and/or the local rock-physical likelihood $p(\mathbf{m}_C|\mathbf{r}_B)$ to rather be approximated by Gaussian mixture distributions. This increases the flexibility because it allows for skewness and/or multimodal conditional distributions, which may improve the accuracy of the procedure if Gaussianity is inappropriate. This extension has the consequence that the local likelihood approximation $p^*(\mathbf{d}_D|\mathbf{r}_B)$ also becomes a Gaussian mixture. The total number of mixture components is the product of the number of mixture components of the two likelihoods. Because evaluation of a Gaussian mixture likelihood with q mixture components is computationally q times more expensive than evaluating a regular Gaussian likelihood, there should, however, be substantial reasons for including many extra levels of complexity. Another approach would be to approximate, e.g., $p(\mathbf{m}_C|\mathbf{r}_B)$ by the selection Gaussian distribution of Rimstad and Omre (2014b), with the consequence that $p^*(\mathbf{d}_D|\mathbf{r}_B)$ also becomes selection Gaussian distributed. However, further investigations are required for such an extension.

Although the methodology of the framework is motivated by inversion of single cells, the framework is not strictly restricted to this situation. All theory and methodology work out even if \mathcal{A} refers to several cells or some weighted sum of these, and also if \mathbf{r} corresponds to two or more rock properties. The latter is handled simply by sampling, say, the porosity and saturation jointly with the geophysical properties when approximating the local rock-physics likelihood in equation 6, and when sampling rock properties in the Monte Carlo routine. Further, the local subsets \mathcal{B} , \mathcal{C} , and \mathcal{D} are not restricted to be contiguous and centered around cell \mathcal{A} as illustrated in Figure 2, even though that is the most natural choice. For some type of applications (e.g., in seismic tomography with complex ray-paths), it may be more appropriate to include scattered cells in \mathcal{B} . This would, however, call for a more comprehensive local subset selection process.

Our approach does not require closed form expressions for the rock properties' prior distribution and the rock-physical likelihood, neither globally nor locally. We only require that samples from the local distributions are obtainable. This is advantageous in the fairly common setting where it is hard to specify closed form expressed models possessing the desired randomness features, but where sampling-based model specifications are easier to put up.

As mentioned, our framework is well suited for parallelization on multicore computers or graphics processing unit (GPU) accelerators. For our data application, the methodology was straightforwardly implemented in the statistical programming language R (R Core Team, 2014). The complete inversion procedure for the synthetic data test was run in parallel on a Windows-based laptop with an Intel i7 2.6Ghz processor and four cores in less than half an hour. This included the initial fitting procedure and the weighted Monte Carlo routine for all the 19,600 individual cells. Because no attempt was made on boosting this performance, there is still room for a significant reduction in the computing time. The main contributions would be a devoted implementation in a precompiled language and massive parallelization on high-performance cores.

The mixing of the CO₂ and brine is a frequently discussed topic. As an alternative to the homogeneous mix that we use in our applications, patchy mixtures of fluids have been suggested (Ghaderi and Landrø, 2009). For the real case, we also tested a simple model for a patchy fluid mix. The patchy model was obtained by substi-

tuting the Reuss average with a Voigt average when deriving the properties of the effective pore fluid. For the vertical profile of Figure 11, this reduced the uncertainty in the high-saturation layers and introduced the possibility of a low-saturated layer below the top CO₂ layer. Apart from this, the inversion results were very similar.

CONCLUSION

We have presented a general framework for approximate Bayesian inversion of geophysical data into rock properties. The methodology approximates quantities directly related to the marginal posterior distributions of the rock properties defining the Bayesian inversion. The framework is well suited for parallelization, producing potentially very fast inversion results. The generality, parallelization properties, and mild assumptions of the approach should make it attractive for a broad range of geophysical challenges in reservoir characterization, monitoring, and exploration.

The issue of the dimensionality in Bayesian inversion was reduced by considering local behavior for all model components (rock properties, geophysical properties, and geophysical data). This was done by merging a Gaussian approximation to a local rock-physical likelihood with a Gaussian approximation to a local geophysical likelihood and then applying a weighted Monte Carlo routine to compute the approximate quantities relevant for the inversion.

The synthetic data test showed that our local approach delivers acceptable errors compared to the much slower MCMC solution; and at least in this case, it is substantially more precise than a frequently used Gaussian inversion approach. For the real data case, the inversion results matched well with a previously published qualitative interpretation of the region.

ACKNOWLEDGMENTS

Thanks go to A. K. Furre and E. H. Jensen for providing the data and details of the rock-physics model for the Utsira Formation. We also thank the Sleipner license partners Statoil, ExxonMobil, and Total for permission to use the Sleipner data.

APPENDIX A

RANGE SPANNING COVARIANCE MATRIX

Consider estimation of a range spanning covariance matrix based on n residuals $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ of dimension q , with both n and q fairly large. Let $\delta_i = \mathbf{e}_i^T \hat{\Sigma}^{-1} \mathbf{e}_i$ (with $\hat{\Sigma} = \{1/(n-1)\} \sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T$ being the standard sample covariance matrix) be a measure of how likely \mathbf{e}_i is. When the residuals are Gaussian, it follows from arguments involving the central limit theorem that the distribution of δ_i is approximately Gaussian with mean q and variance $2q$. Clever use of Jensen's inequality and the moment generating function of the Gaussian distribution then shows that the maximum of n such Gaussian variables is bounded by $q + 2\sqrt{q \log(n)}$. The range spanning estimation method consists of checking whether any of the δ_i exceed this bound. If some do, we reestimate a covariance matrix $\hat{\Sigma}_{\max}$ representative for the most extreme residuals, and we cleverly merge that with $\hat{\Sigma}$. Here, $\hat{\Sigma}_{\max}$ is computed by taking the sample covariance of the most extreme residuals only, while still making sure sufficiently many are included to secure stability. The merged covariance matrix is denoted by $\hat{\Sigma}_{\text{span}}$ and is required to span both $\hat{\Sigma}_{\max}$ and $\hat{\Sigma}$ in the sense that both $\hat{\Sigma}^{-1} - \hat{\Sigma}_{\text{span}}^{-1}$ and

$\hat{\Sigma}_{\max}^{-1} - \hat{\Sigma}_{\text{span}}^{-1}$ are positive semidefinite, i.e., that $\mathbf{x}^T \hat{\Sigma}_{\text{span}}^{-1} \mathbf{x} \leq \max\{\mathbf{x}^T \hat{\Sigma}^{-1} \mathbf{x}, \mathbf{x}^T \hat{\Sigma}_{\max}^{-1} \mathbf{x}\}$ for all \mathbf{x} . This is obtained by ensuring that the inequality holds for all generalized eigenvectors of $\hat{\Sigma}_{\max}$ with respect to $\hat{\Sigma}$. After reestimation and merging, $\hat{\Sigma}$ is set equal to $\hat{\Sigma}_{\text{span}}$ and δ_j is recomputed. The procedure is repeated until no residuals are very large compared with the bound.

APPENDIX B DISTRIBUTIONS

Result: Merging two Gaussian distributions

Let $\mathbf{x}|\mathbf{y} \sim N_{\mathbf{x}}(\mathbf{H}\mathbf{y}, \Sigma_0)$ for some nonrandom matrix \mathbf{H} and $\mathbf{y} \sim N_{\mathbf{y}}(\boldsymbol{\mu}, \Sigma)$. Then, $\mathbf{x} \sim N_{\mathbf{x}}(\mathbf{H}\boldsymbol{\mu}, \mathbf{H}\Sigma\mathbf{H}^T + \Sigma_0)$.

Proof: The two relations are equivalent to $\mathbf{x} = \mathbf{H}\mathbf{y} + \boldsymbol{\varepsilon}_0$, $\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}_0 \sim N_{\boldsymbol{\varepsilon}_0}(0, \Sigma_0)$ independently of $\boldsymbol{\varepsilon} \sim N_{\boldsymbol{\varepsilon}}(0, \Sigma)$. Hence, $\mathbf{x} = \mathbf{H}(\boldsymbol{\mu} + \boldsymbol{\varepsilon}) + \boldsymbol{\varepsilon}_0 = \mathbf{H}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'$, where $\boldsymbol{\varepsilon}' = \mathbf{H}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}_0 \sim N_{\boldsymbol{\varepsilon}'}(0, \mathbf{H}\Sigma\mathbf{H}^T + \Sigma_0)$, which again is equivalent to $\mathbf{x} \sim N_{\mathbf{x}}(\mathbf{H}\boldsymbol{\mu}, \mathbf{H}\Sigma\mathbf{H}^T + \Sigma_0)$.

Two- and four-parameter beta distributions

The (two-parameter) beta distribution $\text{Beta}(a, b)$ has shape parameters a, b and continuous probability distribution function $f_2(y; a, b) \propto y^{a-1}(1-y)^{b-1}$, $y \in [0, 1]$. It has mean $a/(a+b)$ and variance $ab/[(a+b)^2(a+b+1)]$.

The four-parameter beta distribution $\text{Beta}_4(a, b, c, d)$ is a $\text{Beta}(a, b)$ -distribution scaled and shifted to match the support $[c, d]$. Its probability distribution function is scaled by $d-c$ and shifted by c : $f_4(y; a, b, c, d) \propto f_2((y-c)/(d-c); a, b)$, $y \in [c, d]$; its mean is $[a(d-c)/(a+b)+c]/[a(b+d-c)^2]$ and its variance is $[ab(d-c)^2]/[(a+b)^2(a+b+1)]$.

APPENDIX C

WEIGHTED MONTE CARLO WITH CONDITIONAL SAMPLING

In some cases, one may wish to apply the weighted Monte Carlo routine when conditioning on the event that $\mathbf{r}_{\mathcal{A}}$ is contained in some part or block of the sample space. That may in particular be the case when the prior distribution for $\mathbf{r}_{\mathcal{A}}$ has discrete and continuous parts because such parts are best treated separately. The technique may also be used to oversample a priori unlikely parts of the sample space of $p(\mathbf{r}_{\mathcal{A}})$ to increase the overall accuracy of the weighted Monte Carlo routine.

Choose the conditions, say, \mathcal{E}_j , $j = 1, \dots, J$, such that their corresponding blocks on the sample space form a partition, i.e., that the blocks are nonempty disjoint sets whose union is the sample space itself. Assume further that the enumerated routine defined in the weighted Monte Carlo section is carried out J times, once with prior samples conditioned on each of the \mathcal{E}_j . Denote, respectively, the extracted $\mathbf{r}_{\mathcal{A}}$ value and corresponding unnormalized and normalized importance weights for the l th prior sample in the j th run by $\mathbf{r}_{\mathcal{A}}^{(l,j)}$, $v^{(l,j)}$ and $w^{(l,j)}$, $l = 1, \dots, L_j$, $j = 1, \dots, J$. Quantities conditioned on \mathcal{E}_j may be computed analogous to the unconditional ones with weights and samples replaced by those for the j th condition. For instance, $p^*(\mathbf{r}_{\mathcal{A}}|\mathbf{d}, \mathcal{E}_j) = \sum_{l=1}^{L_j} w^{(l,j)} \mathbf{1}_{\{\mathbf{r}_{\mathcal{A}}^{(l,j)} = \mathbf{r}_{\mathcal{A}}\}}$ for discrete distributions, and $p^*(\mathbf{r}_{\mathcal{A}}|\mathbf{d}, \mathcal{E}_j) = \sum_{l=1}^{L_j} w^{(l,j)} K_h(\mathbf{r}_{\mathcal{A}} - \mathbf{r}_{\mathcal{A}}^{(l,j)})$ for continuous

distributions. Further, the posterior probability of each condition \mathcal{E}_j may be computed by

$$p^*(\mathcal{E}_j|\mathbf{d}) = \frac{p(\mathcal{E}_j) \frac{1}{L_j} \sum_{l=1}^{L_j} v^{(l,j)}}{\sum_{i=1}^J p(\mathcal{E}_i) \frac{1}{L_i} \sum_{l=1}^{L_i} v^{(l,i)}} \quad (\text{C-1})$$

Finally, the unconditioned approximated posterior distribution may be computed by $p^*(\mathbf{r}_{\mathcal{A}}|\mathbf{d}) = \sum_{j=1}^J p^*(\mathbf{r}_{\mathcal{A}}|\mathbf{d}, \mathcal{E}_j) p^*(\mathcal{E}_j|\mathbf{d})$. Other unconditioned quantities may be computed similarly using the derived quantities and posterior probabilities for each condition $p^*(\mathcal{E}_j|\mathbf{d})$, $j = 1, \dots, J$, or directly via $p^*(\mathbf{r}_{\mathcal{A}}|\mathbf{d})$.

APPENDIX D

RESOLUTION THEORY FOR SELECTING \mathcal{D} IN GENERAL GEOPHYSICAL PROBLEMS

The formulation we have used in the paper considers the geophysical relation of equation 7, that is, $\mathbf{d} = \mathbf{G}\mathbf{m} + \boldsymbol{\varepsilon}$, with the matrix \mathbf{G} given by the geophysical relations. This is indeed intended because these types of problems are what we aim to solve. However, this general form may be limiting when we consider selection of the local subset \mathcal{D} defining the local data $\mathbf{d}_{\mathcal{D}}$. For AVO data, the influence region may be bounded fairly easily, but when working with, e.g., the rms velocities of Buland et al. (2011), the influence region typically span all data below the position considered. For these cases, it is useful to consider resolution theory. Left multiplying both sides of equation 7 by $\mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}$ gives $\mathbf{d}' = \mathbf{G}'\mathbf{m} + \boldsymbol{\varepsilon}'$, where $\mathbf{G}' = \mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G}$ is the standard resolution kernel, and \mathbf{d}' and $\boldsymbol{\varepsilon}'$ are, respectively, rescaled data and noise. The resolution kernel will in general be much more locally focused. If the matrix $(\mathbf{G}\mathbf{G}^T)^{-1}$ is not invertible, one could include a ridge term before inverting or use some other type of pseudo inversion method to ensure stability of the inverse. It is also possible to perform a preinversion to focus the energy in the problem. This is done by left multiplying both sides of equation 7 by $\hat{\Sigma}_{\mathbf{m}}\mathbf{G}^T(\mathbf{G}\hat{\Sigma}_{\mathbf{m}}\mathbf{G}^T + \Sigma_{\boldsymbol{\varepsilon}})^{-1}$, where $\hat{\Sigma}_{\mathbf{m}}$ is an estimate of the covariance matrix of the geophysical properties and $\Sigma_{\boldsymbol{\varepsilon}}$ is the covariance matrix of the errors.

REFERENCES

- Aki, K., and P. G. Richards, 1980, Quantitative seismology: W. H. Freeman & Co.
- Alnes, H., O. Eiken, S. Nooner, G. Sasagawa, T. Stenvold, and M. Zumberge, 2011, Results from Sleipner gravity monitoring: Updated density and temperature distribution of the CO₂ plume: Energy Procedia, 4, 5504–5511, doi: 10.1016/j.egypro.2011.02.536.
- Arts, R., O. Eiken, A. Chadwick, P. Zweigel, B. van der Meer, and G. Kirby, 2004, Seismic monitoring at the Sleipner underground CO₂ storage site (North Sea): Geological Society, London, Special Publications, 233, 181–191.
- Avseth, P., T. Mukerji, and G. Mavko, 2010, Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk: Cambridge University Press.
- Batzle, M. L., and Z. Wang, 1992, Seismic properties of pore fluids: Geophysics, 57, 1396–1408, doi: 10.1190/1.1443207.
- Boait, F. C., N. J. White, M. J. Bickle, R. A. Chadwick, J. A. Neufeld, and H. E. Huppert, 2012, Spatial and temporal evolution of injected CO₂ at the Sleipner field, North Sea: Journal of Geophysical Research: Solid Earth, 117, 1–21, doi: 10.1029/2011JB008603.
- Bolstad, W. M., 2009, Understanding computational Bayesian statistics: Wiley.
- Bosch, M., T. Mukerji, and E. F. Gonzalez, 2010, Seismic inversion for reservoir properties combining statistical rock physics and geostatistics:

- A review: *Geophysics*, **75**, no. 5, 75A165–75A176, doi: [10.1190/1.3478209](https://doi.org/10.1190/1.3478209).
- Buland, A., O. Kolbjørnsen, and A. Carter, 2011, Bayesian Dix inversion: *Geophysics*, **76**, no. 2, R15–R22, doi: [10.1190/1.3552596](https://doi.org/10.1190/1.3552596).
- Buland, A., O. Kolbjørnsen, R. Hauge, Ø. Skjæveland, and K. Duffaut, 2008, Bayesian lithology and fluid prediction from seismic prestack data: *Geophysics*, **73**, no. 3, C13–C21, doi: [10.1190/1.2842150](https://doi.org/10.1190/1.2842150).
- Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: *Geophysics*, **68**, 185–198, doi: [10.1190/1.1543206](https://doi.org/10.1190/1.1543206).
- Cheng, B., and D. M. Titterton, 1994, Neural networks: A review from a statistical perspective: *Statistical Science*, **9**, 2–30, doi: [10.1214/ss/1177010638](https://doi.org/10.1214/ss/1177010638).
- Doyen, P., 2007, Seismic reservoir characterization: An earth modeling perspective: EAGE.
- Friedman, J. H., 1991, Multivariate adaptive regression splines: *The Annals of Statistics*, **19**, 1–67, doi: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Friedman, J. H., and W. Stuetzle, 1981, Projection pursuit regression: *Journal of the American Statistical Association*, **76**, 817–823, doi: [10.1080/01621459.1981.10477729](https://doi.org/10.1080/01621459.1981.10477729).
- Ghaderi, A., and M. Landrø, 2009, Estimation of thickness and velocity changes of injected carbon dioxide layers from prestack time-lapse seismic data: *Geophysics*, **74**, no. 2, O17–O28, doi: [10.1190/1.3054659](https://doi.org/10.1190/1.3054659).
- Grana, D., and E. Della Rossa, 2010, Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion: *Geophysics*, **75**, no. 3, O21–O37, doi: [10.1190/1.3386676](https://doi.org/10.1190/1.3386676).
- Green, P., K. Atuszyski, M. Pereyra, and C. Robert, 2015, Bayesian computation: A summary of the current state, and samples backwards and forwards: *Statistics and Computing*, **25**, 835–862, doi: [10.1007/s11222-015-9574-5](https://doi.org/10.1007/s11222-015-9574-5).
- Gunning, J., and M. E. Glinsky, 2007, Detection of reservoir quality using Bayesian seismic inversion: *Geophysics*, **72**, no. 3, R37–R49, doi: [10.1190/1.2713043](https://doi.org/10.1190/1.2713043).
- Hammer, H., O. Kolbjørnsen, H. Tjelmeland, and A. Buland, 2012, Lithology and fluid prediction from prestack seismic data using a Bayesian model with Markov process prior: *Geophysical Prospecting*, **60**, 500–515, doi: [10.1111/j.1365-2478.2011.01012.x](https://doi.org/10.1111/j.1365-2478.2011.01012.x).
- Hastie, T., and R. Tibshirani, 1986, Generalized additive models: *Statistical Science*, **1**, 297–310, doi: [10.1214/ss/1177013604](https://doi.org/10.1214/ss/1177013604).
- Hauge, V. L., and O. Kolbjørnsen, 2015, Bayesian inversion of gravimetric data and assessment of CO₂ dissolution in the Utsira Formation: *Interpretation*, **3**, no. 2, SP1–SP10, doi: [10.1190/INT-2014-0193.1](https://doi.org/10.1190/INT-2014-0193.1).
- Henze, N., 2002, Invariant tests for multivariate normality: A critical review: *Statistical Papers*, **43**, 467–506, doi: [10.1007/s00362-002-0119-6](https://doi.org/10.1007/s00362-002-0119-6).
- Joe, H., 1997, *Multivariate models and dependence concepts*: Chapman & Hall.
- Larsen, A. L., M. Ulvmoen, H. Omre, and A. Buland, 2006, Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model: *Geophysics*, **71**, no. 5, R69–R78, doi: [10.1190/1.2245469](https://doi.org/10.1190/1.2245469).
- Malinverno, A., 2002, Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem: *Geophysical Journal International*, **151**, 675–688, doi: [10.1046/j.1365-246X.2002.01847.x](https://doi.org/10.1046/j.1365-246X.2002.01847.x).
- Mavko, G., T. Mukerji, and J. Dvorkin, 2009, *The rock physics handbook: Tools for seismic analysis of porous media*: Cambridge University Press.
- Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: *Journal of Geophysical Research: Solid Earth*, **100**, 12431–12447.
- Mosegaard, K., and A. Tarantola, 2002, Probabilistic approach to inverse problems, in W. H. K. Lee, P. Jennings, C. Kisslinger, and H. Kanamori, eds., *International handbook of earthquake & engineering seismology, Part 1*: Academic Press, 237–265.
- Mukerji, T., A. Jørstad, P. Avseth, G. Mavko, and J. R. Granli, 2001, Mapping lithofacies and pore-fluid probabilities in a North Sea reservoir: Seismic inversions and statistical rock physics: *Geophysics*, **66**, 988–1001, doi: [10.1190/1.1487078](https://doi.org/10.1190/1.1487078).
- Rabben, T. E., and B. Ursin, 2011, AVA inversion of the top Utsira Sand reflection at the Sleipner field: *Geophysics*, **76**, no. 3, C53–C63, doi: [10.1190/1.3567951](https://doi.org/10.1190/1.3567951).
- R Core Team, 2014, *R: A language and environment for statistical computing*: R Foundation for Statistical Computing.
- Rimstad, K., and H. Omre, 2014a, Skew-Gaussian random fields: *Spatial Statistics*, **10**, 43–62, doi: [10.1016/j.spasta.2014.08.001](https://doi.org/10.1016/j.spasta.2014.08.001).
- Rimstad, K., and H. Omre, 2014b, Generalized Gaussian random fields using hidden selections: *ArXiv e-prints*.
- Robert, C., and G. Casella, 2005, *Monte Carlo statistical methods*: Springer.
- Rue, H., S. Martino, and N. Chopin, 2009, Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion): *Journal of the Royal Statistical Society, Series B*, **71**, 319–392, doi: [10.1111/j.1467-9868.2008.00700.x](https://doi.org/10.1111/j.1467-9868.2008.00700.x).
- Silverman, B., 1986, *Density estimation for statistics and data analysis*: Taylor & Francis.
- Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: *Reviews of Geophysics and Space Physics*, **20**, 219–232, doi: [10.1029/RG020i002p00219](https://doi.org/10.1029/RG020i002p00219).
- Ulvmoen, M., and H. Omre, 2010, Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations. Part 1: Methodology: *Geophysics*, **75**, no. 2, R21–R35, doi: [10.1190/1.3294570](https://doi.org/10.1190/1.3294570).

This article has been cited by:

1. Anne-Kari Furre, Ola Eiken, Håvard Alnes, Jonas Nesland Vevatne, Anders Fredrik Kiær. 2017. 20 Years of Monitoring CO₂-injection at Sleipner. *Energy Procedia* **114**, 3916-3926. [[Crossref](#)]
2. Mats Lundh Gulbrandsen, Knud Skou Cordua, Thomas Mejer Hansen, Klaus Mosegaard. Revealing Multiple Geological Scenarios Through Unsupervised Clustering of Posterior Realizations from Reflection Seismic Inversion 541-555. [[Crossref](#)]
3. Zhao-Yun Zong, Xing-Yao Yin, Kun Li. 2016. Joint AVO inversion in the time and frequency domain with Bayesian interference. *Applied Geophysics* **13**:4, 631-640. [[Crossref](#)]
4. Odd Kolbjørnsen, Arild Buland, Ragnar Hauge, Per Røe, Martin Jullum, Richard William Metcalfe, Øyvind Skjæveland. 2016. Bayesian AVO inversion to rock properties using a local neighborhood in a spatial prior model. *The Leading Edge* **35**:5, 431-436. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF w/Links](#)]