Bayesian rock physics inversion using a localized ensemble-based approach - with an application to the Alvheim field

Mina Spremić[*] **and Jo Eidsvik**[*] **and Per Avseth**[†]

Right Running Head: **Bayesian inversion - Alvheim field**

[*]*Norwegian University of Science and Technology, Department of Mathematical*

Sciences, NO-7491, Trondheim, Norway.

E-mail: mina.spremic@ntnu.no; jo.eidsvik@ntnu.no

[†] *Dig Science, Norway. E-mail: per.avseth@digscience.no*

(November 10, 2023)

ABSTRACT

We describe, implement and show results of a localized ensemble-based approach for seismic AVO inversion with uncertainty quantification. Ensembles are simulated from prior probability distributions for fluid saturations and clay content. Starting with continuous saturations and clay content variables, we use depth-varying models for cementation and grain contact theory, Gassmann fluid substitution with mixed saturations, and approximations to the Zoeppritz equations for the AVO attributes at the top-reservoir. The local conditioning to seismic AVO observations relies on i) the misfit between ensemble simulated seismic AVO data and the field observations in a local partition of the grid/local patch, of inlines/crosslines around the locations where we aim to predict, ii) correlations between the simulated reservoir properties and the data in local patches, and iii) local assessment to avoid unrealistic updates based on spurious correlations in the ensembles. Data from the Alvheim field in the North Sea are used to demonstrate the approach. The influence of the prior information from the well logs in combination with the seismic reflection data indicates presence of higher oil and gas saturation in the lobe structures of the field and increased clay content at their edges.

INTRODUCTION

Stochastic reservoir characterization from seismic data is commonly done today, see e.g. Bosch et al. (2010), Azevedo and Soares (2017), Grana et al. (2021) or Grana et al. (2022) for an overview of this field. Following the Bayesian approach, one specifies i) a prior model for the reservoir variables of interest based on geological knowledge and well log data in the area of interest, ii) a likelihood model based on rock physics relations between the reservoir properties and the seismic data as well as a specification of the seismic observation noise. The goal is then to assess the posterior distribution of the reservoir variables, conditional on the seismic data. Bayesian inversion of seismic amplitude data to predict density, P- and S-velocity is one example of this situation. Probabilistic lithology and fluid prediction from seismic amplitude data is another example.

Our focus in this paper is the prediction of reservoir properties from seismic amplitude versus offset (AVO) data. Unlike discrete facies and saturation models, see e.g. Larsen et al. (2006), Buland et al. (2008) and Grana et al. (2017), we represent the fluid saturations and clay content properties in the reservoir as continuous variables (Bachrach, 2006; Shahraeeni and Curtis, 2011). In doing so, we enable shaly sands or sand-clay mixtures and partial saturations in the sample space. A motivation for using continuous reservoir properties, is insight gained from the Alvheim field in the North Sea, which is used as a case study throughout this paper. Parts of the Alvheim field is dominantly oil saturated while other parts at similar depths are gas saturated. Yet, the seismic response at top-reservoir for different fluid scenarios can be similar.

This could be an effect of geological burial history, with laterally varying cementation volume within the reservoir sandstones associated with Miocene tectonic tilting as the East Shetland Platform was uplifted and a subsequent tectonic tilting process, see e.g. Rimstad et al. (2012) and Avseth and Lehocki (2016). The varying elastic stiffness could also be driven by shaly sands because the sands at Alvheim show a varying degree of clay content (Avseth et al., 2008), or there could be partial saturations which are seen in well log data from the Alvheim field. In this paper we gain further insight related to ambiguities and competing effects at Alvheim, showing results from our suggested model. Moreover, we argue that a continuous reservoir model representation avoids unrealistically strong classification tendencies and accuracies which can occur in discretized facies solutions.

We use data from four wells at Alvheim to build the prior model for oil and gas saturations and clay content. We also use these logs along with established rock physics models to train the likelihood model for the seismic AVO data, see e.g. Avseth et al. (2010). The seismic AVO inversion is focused on the top-reservoir horizon, extracted from a 3D seismic data volume that has been pre-processed for AVO analysis, see Rimstad et al. (2012). Eidsvik et al. (2004) similarly study the top-reservoir properties at the Glitne field, using a discrete Markov random field representation for the lithology and fluid saturation.

The Bayesian inversion is done by an ensemble-based approach. In particular, we use a local ensemble transform Kalman filter that predicts the reservoir properties at a location from a (local) patch of seismic AVO data surrounding it. Monte Carlo

4

realizations (ensembles) of reservoir variables and data are used to find the correlations going into the Kalman filter equations. In each patch, posterior samples are obtained as a linear predictor from the data in and around the patch.

Others have tried related ensemble-based approaches to seismic inversion: Liu and Grana (2018) apply an ensemble-based methods to seismic AVO inversion, but they do not use the local approach outlined here. Thurin et al. (2019) and Gineste et al. (2020) use the ensemble subspace version that we rely on here for ensemble-based iterative inversion of seismic waveform data in a layer-based depth section. The problem of full waveform inversion is a much more non-linear problem than that of AVO inversion. Since the problem is close to linear, a natural question is whether the forward model associated with the seismic likelihood model could instead be linearized analytically, see e.g. Buland and Omre (2003); Eidsvik et al. (2008); Grana (2016); Lang and Grana (2018). Nevertheless, the ensemble-based version avoids massive-size matrix calculations, which are instead performed at the size of the ensemble in our approach. Further, analytical linearization could be cumbersome only with a slight model change requiring tedious efforts to get derivatives, unless an automatic differentiation approach can be used in the calculations.

The paper is structured as follows: We start by presenting the Alvheim case as our motivation for implementing the models and methods of this paper. The approach used in the paper is exemplified in the context of this particular case. We next outline the Bayesian framework, including the prior modeling with multivariate geostatistical assumptions and the rock physics relations linking reservoir properties

to the seismic AVO data. We then describe the ensemble-based method that is used for approximate Bayesian inversion. We show results on the Alvheim field, and discuss the performance of this approach via skill scores, sensitivity analysis and comparison with other methods. In the end we conclude and point to further research.

CASE: ALVHEIM FIELD

Background description

The Alvheim field is a turbiditic oil and gas field of Paleocene age in the North Sea, offshore Norway (Figure 1a). The sand distribution and hydrocarbon trapping in this field is complex, with multiple submarine fan lobes and large variation in depositional facies, net-to-gross and sand texture, ranging from massive, thick-bedded sands to more heterogeneous, inter-bedded sand-shale intervals. The depositional pattern has been controlled by structural topography associated with deeper faults and salt tectonics.

The reservoir sands represent the Heimdal Formation, and are capped by Lista Formation shales. The rock physics properties of the Heimdal Formation sands are also affected by the compaction history, as the Alvheim field is buried approximately 2 km below the seafloor, corresponding to a temperature of around 70°C. This is around the temperature at which we expect transition from mechanical to chemical compaction, as smectite-rich shales start to transform to illite, and quartzose sands begin to cement with quartz overgrowths on grain surfaces. These diagenetic processes

result in significant elastic stiffening of the rock frame. It has been documented that both unconsolidated sands and cemented sandstones occur in the Alvheim field (Avseth et al., 2008). The presence of just a few percent of quartz cement at grain contacts will have a large effect on fluid sensitivity of the seismic signal, and it is challenging to predict the correct pore fluid from seismic amplitudes without taking into account these diagenetic changes.

The distribution of hydrocarbons is also quite complex in the Alvheim field, with both gas and oil present. The hydrocarbon migration and trapping history in the area is poorly understood, but is likely to have happened during different episodes, and strongly affected by the regional tectonics, including structural tilting and uplift events in Miocene and Quaternary. Hence, some of the lobes are filled predominantly with gas, whereas others are filled with only oil. The complex distribution of lithology and fluids makes it quite challenging to perform seismic reservoir characterization in the Alvheim field.

Seismic AVO data

Seismic AVO data are available for this study, including angle stacks from near (12°), mid (22°) and far (31°) angle ranges. The data have been processed and preconditioned to obtain seismic data quality suitable for quantitative analysis. The processing sequence is summarized by Rimstad et al. (2012). AVO attributes, including zero-offset reflectivity (R_0) and AVO gradient (G) have been extracted and calibrated using scalars estimated from the upscaled well log data (Avseth and Lehocki, 2016).

At each inline and crossline in addition to having two data points (R_0 and G), we have the associated traveltimes, shown in Figure 1b, which are later converted to the depth. The focus of this case study is on the top-reservoir horizon, as shown in Figure 1b, where we, in addition to having spatial dependence in inline-crossline domain, have depth of the 2D slice at each available inline-crossline location. Because we focus on this top-reservoir horizon and include depth information, we refer to this as a 2.5D inverse problem.

The AVO data from the Alvheim field are shown in Figure 2, where R_0 and G, are available for each inline and crossline of the top-reservoir. The available AVO data covers, in total, a region of approximately $12.35\,\text{km} \times 8.85\,\text{km}$. We use measurements from every 4 inlines and crosslines compared to the original acquisition. This corresponds to $50\,\text{m}$ in crossline direction and $50\,\text{m}$ in the inline direction. The spatial grid we study is represented through inlines and crosslines, meaning the unit conversion to distance metrics is not performed and all the modeling is done on a grid representation of the inline and crossline domain, as shown in Figure 3.

Well log data

Four wells are used in this study, their locations and outcomes at the top-reservoir are shown in Figure 3, and when relevant will be shown in figures throughout the paper, but the legends will be dropped. The wells include 24/6–2, 24/6–4, 25/4–7 and 25/4–8, which are all within the domain spanned by the seismic inline-crossline

---

*Map from `https://factmaps.npd.no`, accessed 20.12.22.

ranges. These wells are drilled on structural highs, in the shallower sections of the top-reservoir. The well log data include sonic velocities (V_p and V_s), densities, and petrophysical logs, see e.g. Rimstad et al. (2012). The latter comprise fluid saturations (oil, gas and brine) and clay content for each of the wells and are shown in Figure 4.

In well 24/6–2 there is commercial gas in the Heimdal Formation with a gas column of 52 m down to the gas-oil contact at 2151 m. There is also a thin oil column of 17 m down to the oil-water contact at 2168 m. Well 25/4–7 encountered commercial oil saturation, with an oil column of 48 m down to the oil–water contact at 2133 m. Hydrocarbon saturations are mostly within commercial ranges (0.6–0.9), but lower values are seen in heterolithic zones as presence of clay affects the transport properties of the reservoir. Thin clay laminations can also cause patchy saturations at the log scale. Wells 24/6–4 and 25/4–8 are interpreted as constant for different zones, due to missing saturation logs estimated from the resistivity logs.

MODEL

Bayesian model

To perform seismic AVO inversion, we define the problem in a Bayesian framework. This requires a prior model or prior probability density function (pdf) denoted by $p(\boldsymbol{x})$, a probabilistic representation of the continuous reservoir variables of interest $\boldsymbol{x}$. We use an informed type of prior, building on the expert knowledge of the domain

regarding geological trends, spatial smoothness and fluid sorting along with well data.

The Bayesian framework further requires a likelihood model denoted by pdf $p(\boldsymbol{y}|\boldsymbol{x})$ for the seismic AVO data $\boldsymbol{y}$, given the reservoir variables $\boldsymbol{x}$. This pdf includes a forward model, that often just goes into the expectation term of $p(\boldsymbol{y}|\boldsymbol{x})$. In our setting, the forward model builds on rock physics relations between the rock and fluid properties of interest and the seismic AVO reflection data.

The goal of Bayesian inversion is then to obtain the posterior pdf of the reservoir variables given the data,

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}). \tag{1}$$

Most often, in the case of inversion problems, we are facing weakly or highly non-linear forward models which entails complicated likelihood models, leading to posterior pdfs which are not analytically tractable. We assess the posterior in equation 1 by multiple realizations obtained by Monte Carlo sampling. Here, we choose to apply an ensemble based method to conduct approximate posterior sampling, which is further explained in the next section.

Before elaborating on the models mentioned above, we introduce the necessary notation. Reservoir variables are represented at locations $\boldsymbol{u} \in \mathcal{D}$, where $\mathcal{D}$ denotes the inline and crossline grid used to represent the seismic data at the top-reservoir. The grid size is 248×178 which gives about $N = 44144$ grid cells with inline-crossline locations $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$. Seismic AVO data at locations $\boldsymbol{u} \in \mathcal{D}$ are $\boldsymbol{y}(\boldsymbol{u}) = [R_0(\boldsymbol{u}), G(\boldsymbol{u})]$. The main interest lies in the oil and gas saturation, and clay content at each of the

inline-crossline locations, which we denote by $\boldsymbol{r}(\boldsymbol{u}) = [\boldsymbol{S}_\mathrm{g}(\boldsymbol{u}), \boldsymbol{S}_\mathrm{o}(\boldsymbol{u}), \boldsymbol{V}_\mathrm{clay}(\boldsymbol{u})]$.

Results will be displayed for the reservoir variables of interest $\boldsymbol{r}(\boldsymbol{u})$. However, in the suggested methodology and algorithm, for the purpose of efficient computations, we work with transformed saturations and clay content. Variables are transformed back and forth by applying logistic functions at all locations $\boldsymbol{u} \in \mathcal{D}$,

$$
\begin{aligned}
S_l(\boldsymbol{u}) &= \frac{\exp\left[x_l(\boldsymbol{u})\right]}{1 + \exp\left[x_\mathrm{g}(\boldsymbol{u})\right] + \exp\left[x_\mathrm{o}(\boldsymbol{u})\right]}, \quad l = \mathrm{o,g(oil,gas)} \\
S_\mathrm{b}(\boldsymbol{u}) &= \frac{1}{1 + \exp\left[x_\mathrm{g}(\boldsymbol{u})\right] + \exp\left[x_\mathrm{o}(\boldsymbol{u})\right]},
\end{aligned}
\tag{2}
$$

where $x_\mathrm{o}(\boldsymbol{u}) \in (-\infty, \infty)$ and $x_\mathrm{g}(\boldsymbol{u}) \in (-\infty, \infty)$. From equation 2, we then have that $S_\mathrm{g}(\boldsymbol{u}) + S_\mathrm{o}(\boldsymbol{u}) + S_\mathrm{b}(\boldsymbol{u}) = 1$ for every location $\boldsymbol{u}$. In doing so, we keep brine as a reference or background saturation. Similar for clay content $V_\mathrm{clay}(\boldsymbol{u})$,

$$
V_\mathrm{clay}(\boldsymbol{u}) = \frac{\exp\left[x_\mathrm{clay}(\boldsymbol{u})\right]}{1 + \exp\left[x_\mathrm{clay}(\boldsymbol{u})\right]},
\tag{3}
$$

where $x_\mathrm{clay}(\boldsymbol{u}) \in (-\infty, \infty)$. Altogether, we then get the reservoir variables of interest from a logistic function, compactly denoted by

$$
\boldsymbol{r}(\boldsymbol{u}) = \boldsymbol{f}_0\left[\boldsymbol{x}(\boldsymbol{u})\right], \quad \boldsymbol{u} \in \mathcal{D}.
\tag{4}
$$

In Figure 5 we show a graphical illustration of the reservoir variables and their connection to the seismic AVO data. The overview of the remaining notation used in the paper is provided in Table 2 in Appendix A.

Next, we describe the building blocks of this model. We start with the geostatistical prior model elements relevant for the Alvheim case, and provide the likelihood model elements which are using geophysical theory. Finally, we present the posterior

model and the chosen way of approximating it for this case. Note that our Bayesian model is using fixed hyperparameters in the prior and likelihood. An alternative fully Bayesian formulation would put prior distributions on these parameters, see e.g. Malinverno and Briggs (2004). In our setting we instead conduct sensitivity analysis to the most relevant input parameters.

Prior model

The prior model is based on spatial random field models for the reservoir variables $r(u)$, defined in equation 4. The prior pdfs for the transformed saturation variables $x_{\mathrm{g}} = [x_{\mathrm{g}}(u_1), \ldots, x_{\mathrm{g}}(u_N)]$, $x_{\mathrm{o}} = [x_{\mathrm{o}}(u_1), \ldots, x_{\mathrm{o}}(u_N)]$ and $x_{\mathrm{clay}} = [x_{\mathrm{clay}}(u_1), \ldots, x_{\mathrm{clay}}(u_N)]$ are here represented by independent Gaussian random fields. With the spatial discretization of locations, we get a Gaussian multivariate model so that,

$$x_{\mathrm{g}} \sim N(\boldsymbol{\mu}_{\mathrm{g}}, \boldsymbol{\Sigma}_{\mathrm{g}}), \quad x_{\mathrm{o}} \sim N(\boldsymbol{\mu}_{\mathrm{o}}, \boldsymbol{\Sigma}_{\mathrm{o}}), \quad x_{\mathrm{clay}} \sim N(\boldsymbol{\mu}_{\mathrm{clay}}, \boldsymbol{\Sigma}_{\mathrm{clay}}). \tag{5}$$

Here, the length N prior mean vectors $\boldsymbol{\mu}_{\mathrm{g}}$, $\boldsymbol{\mu}_{\mathrm{o}}$, and $\boldsymbol{\mu}_{\mathrm{clay}}$, and the $N \times N$ covariance matrices $\boldsymbol{\Sigma}_{\mathrm{g}}$, $\boldsymbol{\Sigma}_{\mathrm{o}}$ and $\boldsymbol{\Sigma}_{\mathrm{clay}}$ are set from well log information at various depths.

Figure 6 illustrates the well log data for transformed clay content, gas and oil saturations, along with the fitted mean (background models) and the associated uncertainty levels (90%).

The mean values induce a smooth background model with depth. High gas saturation is most likely at the very shallow inline-crossline locations, high oil saturation is most likely at quite shallow locations, while brine is very likely to be high at the

deeper locations. For each variable, the standard deviation is assumed not to be varying with depth.

As for the covariance matrix, we further include spatial correlation in the form of the Gaussian correlation function with effective correlation length of 15 grid cells. This is used for x_{o}, x_{g} and x_{clay}, and specified from the prior assumptions about the geological depositional environment with geographically separated lobe structures.

The suggested methodology requires Monte Carlo samples from the prior distribution. For this task, we use the Fast Fourier transform (FFT) routine as described by e.g. Davies and Bryant (2013), presented in Appendix B. Such approaches were previously applied to rapid 3D elastic inversion in Buland et al. (2003). The FFT routine, which includes embedding the covariance matrix onto a torus, allows for circumventing the construction of the full prior covariance matrix for the whole domain. The realizations are generated in the Fourier domain, making it an efficient method. Each of the realizations is conditioned on the well outcomes at the top-reservoir, and the realizations of the field are denoted by $\boldsymbol{x}^i, i = 1, \ldots, n_e$.

Figures 7 and 8 show prior ensemble means for the reservoir parameters $\boldsymbol{r}(\boldsymbol{u})$ for all locations $\boldsymbol{u}$ on the inline-crossline grid and the associated uncertainty. In Figures 7a–7c, one recognizes the regions of high oil or gas saturation, resulting from the well conditioning. In Figures 7d–7f we can see that the uncertainty is barely noticeable in the area around the wells, while it increases dramatically as we go away from these well locations. The ensemble members cover a wider range and have higher uncertainty in the remaining regions, where the prior ensemble mean has values in

13

the range between 0.3–0.5.

The well conditioning is done using information from the well logs (Figure 4) at top reservoir and at their respective inline-crossline locations. Because of the modeled spatial dependence, data tends to be informative in the vicinity of the well locations. Well conditioning has a substantial influence on the fluid saturation variables which are rather accurately observed at the top reservoir. The well log data for the clay content varies too much near the top reservoir to provide useful information for the clay content in well conditioning. Hence the mean and uncertainty for clay content in Figure 8 show less extreme values, and the uncertainty is overall quite high.

Figure 9 shows a ternary plot of realizations of oil, gas and brine saturation at two selected locations on the inline-crossline grid. The shallow point in Figure 9a contains a mixture of mostly gas and brine with a larger spread of realizations, while the deeper point in Figure9b contains predominantly brine and variation in realizations is rather small.

Likelihood model

Likelihood model is defined as Gaussian, with mean value coming from the nonlinear geophysical forward model and a spatially independent variance. Relations in the geophysical forward model are assumed valid for each inline and crossline, but depend on the known depth of the top-reservoir at each grid location. In this likelihood model we assume conditional independence, meaning that data at different locations

are independent, given the rock properties. This is reasonable given that the top reservoir reflection processing induces no coupling between the zero-offset reflectivity or AVO gradient information at different locations. At a location $\boldsymbol{u}$, we let $\boldsymbol{y}(\boldsymbol{u})$ denote the zero-offset coefficient R_0 and gradient G at that location. The model can be written in an additive manner,

$$\boldsymbol{y}(\boldsymbol{u}) = \boldsymbol{h}_0\left[\boldsymbol{x}(\boldsymbol{u})\right] + \boldsymbol{\epsilon}(\boldsymbol{u}), \quad \boldsymbol{\epsilon}(\boldsymbol{u}) \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_0), \tag{6}$$

where $\boldsymbol{h}_0\left[\boldsymbol{x}(\boldsymbol{u})\right] = \boldsymbol{g}_0\left\{\boldsymbol{f}_0\left[\boldsymbol{x}(\boldsymbol{u})\right]\right\}$, and $\boldsymbol{g}_0[\boldsymbol{r}(\boldsymbol{u})]$ is the rock physics forward model function working on the reservoir variables $\boldsymbol{r}(\boldsymbol{u})$ of main interest. The 2×2 matrix $\boldsymbol{\Omega}_0$ captures the variances in R_0 and G and the correlation between the two. Based on seismic reflection data near the four wells and associated well log data in the reservoir zones, we specify parameters $\mathrm{Var}(R_0) = 0.003$, $\mathrm{Var}(G) = 0.03$ and $\mathrm{Corr}(R_0, G) = -0.6$. These are quite similar to those used in Eidsvik et al. (2004).

By concatenating data at all inline-crossline locations into a length $2N$ vector $\boldsymbol{y} = [\boldsymbol{y}^t(\boldsymbol{u}_1), \dots, \boldsymbol{y}^t(\boldsymbol{u}_N)]^t$, we have probability density function $p(\boldsymbol{y}|\boldsymbol{x})$ equal to that of a Gaussian distribution $N(\boldsymbol{h}(\boldsymbol{x}), \boldsymbol{\Omega})$. With the assumption of conditional independence between the AVO data at different inline-crossline grid locations, the size $N \times N$ covariance matrix $\boldsymbol{\Omega}$ is block diagonal with the 2×2 matrix $\boldsymbol{\Omega}_0$ repeating along the block diagonal elements. The function $\boldsymbol{h}(\boldsymbol{x}) = (\boldsymbol{h}_0^t[\boldsymbol{x}(\boldsymbol{u}_1)], \dots, \boldsymbol{h}_0^t[\boldsymbol{x}(\boldsymbol{u}_N)])^t$ is built on the same rock physics relations working on different spatially varying input reservoir variables.

Rock physics and geophysical models, along with geological properties are forming a realistic forward model $\boldsymbol{h}_0(\cdot)$. As illustrated in Figure 5, the forward model

takes clay content and fluid saturation represented as random fields, as input. To account for the cementation effect which is assumed to occur at a certain depth, we impose rock physics models building on mechanical and chemical compaction. We use either an unconsolidated sand model or contact cement model, see for example Avseth et al. (2010), depending on the depth at which the input reservoir variables are situated. The effect of the change in the rock physics model is illustrated in Figure 10. Furthermore, the Gassmann fluid substitution is applied, followed by the Shuey approximation to obtain the AVO attributes. Details on the forward model are given in Appendix C.

LOCAL ENSEMBLE TRANSFORM KALMAN FILTER

Ensemble-based Kalman methods, first proposed by Evensen (1994), strive to approximate the posterior distribution with a limited set of realizations or ensembles. There are many variants of ensemble Kalman filters and smoothers. The one used in this paper is a square root approach called the local Ensemble transform Kalman filter (LETKF), see e.g. Li (2007) and Hunt et al. (2007). In the current setting with an inverse problem, one starts with an ensemble of realizations from the prior model $p(\boldsymbol{x})$, which we denote by $\boldsymbol{E} = (\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{n_e})$ for the n_e ensemble members. In our examples we use $n_e = 100$, which is a common number in these type of ensemble-based approximations (Van Leeuwen et al., 2015; Asch et al., 2016). The goal is to update the prior ensemble through conditioning, combining the information from the data and the physical forward model.

16

When updating the parameters, discretized on the grid of inlines and crosslines, we split the parameter domain into J patches of fixed size. We denote the patches by j, where $j = 1, \ldots, J$. The total number of patches varies depending on the chosen size of the patch. Each parameter patch has its corresponding and larger observation patch, which includes relevant information for the update. Parameter patches are centered in the middle of the observation patch, with some adjustment at the edges of the domain. The division into patches is further illustrated in Figure 11. The updating is hence local, honoring the data while avoiding spurious correlations which would occur with very large patches. Within each local patch, the parameters are updated using the LETKF method in Algorithm 1.

Note that the LETKF update is performed in the ensemble subspace of size $n_e <<$ n. In the conditioning we first update the mean of the ensemble with the Kalman gain $\boldsymbol{K}_j$ which is specified from ensemble perturbations and synthetic data perturbations. The Kalman gain works on the deviation between the real data $\boldsymbol{y}_j^{\mathrm{obs}}$ and the mean of the simulated data $\bar{\boldsymbol{y}}_j$,

$$\bar{\boldsymbol{x}}_j^a = \bar{\boldsymbol{x}}_j + \boldsymbol{K}_j(\boldsymbol{y}_j^{\mathrm{obs}} - \bar{\boldsymbol{y}}_j), \quad j = 1, \ldots, J. \tag{7}$$

Next, each of the ensemble members are set around the mean from equation 7, using the scaled ensemble perturbations $\boldsymbol{X}_j$ for patch j, see Algorithm 1.

Extensions of the outlined approach add an iterative loop over the Kalman equations. One then updates the perturbations, covariance and Kalman gain iteratively to improve the posterior approximation of the ensemble. This also involves adjustments of the perturbation matrix and the square root of the ensemble error covariance at ev-

ery iteration. This is referred to as the iterative ensemble Kalman smoother (IEnKS)

approach, see e.g. Asch et al. (2016).

Algorithm 1 Localized Ensemble Transform Kalman Filter. Subscripts j indicate a

subset of matrices associated with patch $j = 1, \ldots, J$.

$\boldsymbol{x}^i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad i = 1, \ldots, n_e$ $\qquad\qquad\qquad$ ▷ Prior ensemble

$\bar{\boldsymbol{x}} = \frac{1}{n_e} \sum_{i=1}^{n_e} \boldsymbol{x}^i$

$\boldsymbol{X} = (\boldsymbol{x}^1 - \bar{\boldsymbol{x}}, \ldots, \boldsymbol{x}^{n_e} - \bar{\boldsymbol{x}})$ $\qquad\qquad$ ▷ Prior ensemble perturbations

$\bar{\boldsymbol{y}} = \frac{1}{n_e} \sum_{i=1}^{n_e} \boldsymbol{y}^i, \quad \boldsymbol{y}^i = h(\boldsymbol{x}^i), \quad i = 1, \ldots, n_e$ $\qquad$ ▷ Synthetic data ensemble

$\boldsymbol{Y} = (\boldsymbol{y}^1 - \bar{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{n_e} - \bar{\boldsymbol{y}})$ $\qquad\qquad$ ▷ Synthetic data perturbations

repeat for each patch $j = 1, \ldots, J$:

$\qquad \boldsymbol{H}_j = [\boldsymbol{Y}_j^T \boldsymbol{\Omega}_j^{-1} \boldsymbol{Y}_j + (n_e - 1)\boldsymbol{I}]^{-1}$

$\qquad \boldsymbol{K}_j = \boldsymbol{X}_j \boldsymbol{H}_j \boldsymbol{Y}_j^T \boldsymbol{\Omega}_j^{-1}$ $\qquad\qquad\qquad$ ▷ Kalman gain

$\qquad \bar{\boldsymbol{x}}_j^a = \bar{\boldsymbol{x}}_j + \boldsymbol{K}_j(\boldsymbol{y}_j^{\text{obs}} - \bar{\boldsymbol{y}}_j)$ $\qquad\qquad\qquad$ ▷ Updated mean

$\qquad \boldsymbol{X}_j^a = \bar{\boldsymbol{x}}_j^a + \boldsymbol{X}_j[(n_e - 1)\boldsymbol{H}_j]^{1/2}$ $\qquad\qquad$ ▷ Updated ensemble

$\qquad \boldsymbol{x}_j^{a,i}$ is column i of $\boldsymbol{X}_j^a$, $i = 1, \ldots, n_e$

until

$\boldsymbol{E} = \{(\boldsymbol{x}_1^{a,i}, \ldots, \boldsymbol{x}_J^{a,i}), i = 1, \ldots, n_e\}$

PREDICTION, UNCERTAINTY AND VALIDATION

We now present ways of conducting prediction, uncertainty quantification and vali-

dation of the outlined LETKF methods.

Prediction of reservoir variables is done using the sample mean $\hat{\boldsymbol{x}}^a = \frac{1}{n_e} \sum_{i=1}^{n_e} \boldsymbol{x}^{a,i}$,

and similarly for oil and gas saturation and clay content which are the variables of primary interest. For such variables which are limited between 0 and 1, the sample variance is not very informative. Instead, we quantify the uncertainty in predictions by sorting variables. Say, for oil saturation $x_o^{a,i}(\boldsymbol{u})$, $i = 1, \ldots, n_e$ at a location $\boldsymbol{u}$, we sort these from smallest to largest such that $x_o^{a,(1)} < x_o^{a,(2)} < \ldots < x_o^{a,(n_e)}$. From this we compute the width between the 10th and the 90th percentile in the ensemble. This is done separately for oil, gas and brine saturation and for the clay content.

We use leave-one-out cross validation to study the predictive properties of the model. Doing so, one observation is excluded at the time, and its predictive distribution is constructed from all other data. Spatial cross-validation leaving out larger parts of the data was not attempted because it is difficult to combine with localised methods. Also, as illustrated next, we use a computationally efficient way of adjusting the ensemble that is easily implemented for the leave-one-out situation.

We construct the predictive leave-one-out distribution effectively by re-weighting the posterior ensemble to account for exclusion of the particular observation. At site $\boldsymbol{u}$, the predictive pdf of the data, leaving-out the particular observations at $\boldsymbol{u}$ is given by

$$p[\boldsymbol{y}(\boldsymbol{u})|\boldsymbol{y}(-\boldsymbol{u})] = \sum_{i=1}^{n_e} w^i(\boldsymbol{u})p[\boldsymbol{y}(\boldsymbol{u})|\boldsymbol{x}^{a,i}(\boldsymbol{u})]. \tag{8}$$

Here, $\boldsymbol{y}(-\boldsymbol{u})$ denotes all data except that at location $\boldsymbol{u}$. The pdf in equation 8 is a mixture of Gaussian pdfs $p[\boldsymbol{y}(\boldsymbol{u})|\boldsymbol{x}^{a,i}(\boldsymbol{u})]$, $i = 1, \ldots, n_e$ where the weights are

$$w^i(\boldsymbol{u}) = \frac{1/p[\boldsymbol{y}^{\mathrm{obs}}(\boldsymbol{u})|\boldsymbol{x}^{a,i}(\boldsymbol{u})]}{\sum_{i=1}^{n_e} 1/p[\boldsymbol{y}^{\mathrm{obs}}(\boldsymbol{u})|\boldsymbol{x}^{a,i}(\boldsymbol{u})]}, \quad i = 1, \ldots, n_e, \tag{9}$$

which are evaluated at the particular observation $\boldsymbol{y}^{\mathrm{obs}}(\boldsymbol{u})$. These weights result from

re-adjusting the equal weights in the posterior approximation of $p(\boldsymbol{x}|\boldsymbol{y})$. The adjustment is based on the conditional independence of data $\boldsymbol{y}$ given $\boldsymbol{x}$, which in this Bayesian formulation entails

$$p[\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{y}(\boldsymbol{u})|\boldsymbol{y}(-\boldsymbol{u})] = p[\boldsymbol{x}(\boldsymbol{u})|\boldsymbol{y}(-\boldsymbol{u})]p[\boldsymbol{y}(\boldsymbol{u})|\boldsymbol{x}(\boldsymbol{u})],$$

$$p[\boldsymbol{x}(\boldsymbol{u})|\boldsymbol{y}(-\boldsymbol{u})] = \frac{p[\boldsymbol{x}(\boldsymbol{u})|\boldsymbol{y}]p(\boldsymbol{y})}{p[\boldsymbol{y}(\boldsymbol{u})|\boldsymbol{x}(\boldsymbol{u})]p[\boldsymbol{y}(-\boldsymbol{u})]} \propto \frac{p[\boldsymbol{x}(\boldsymbol{u})|\boldsymbol{y}]}{p[\boldsymbol{y}(\boldsymbol{u})|\boldsymbol{x}(\boldsymbol{u})]}. \tag{10}$$

The associated cumulative distribution function (cdf) of equation 8 for the observed R_0 or G at location $\boldsymbol{u}$ is denoted by $\hat{F}(y) = F[y(\boldsymbol{u})|\boldsymbol{y}(-\boldsymbol{u})]$, for $R_0(\boldsymbol{u})$ or $G(\boldsymbol{u})$. We compute the percentiles of the observed values $y^{\text{obs}}(\boldsymbol{u})$ in this distribution from the mixture of Gaussian cdfs defined via equation 8. If the model is reasonable, we expect the percentiles of the observed R_0 or G in this cdf to be uniformly distributed. Hence, when this leave-one-out procedure is conducted over many different inline-crossline grid locations, the summary results should be approximately uniformly distributed. This approach is similar to that of rank histograms, where one simply tracks the rank of the observations $y^{\text{obs}}(\boldsymbol{u})$ in the sorted n_e ensemble members of the simulated observations at that location. For effective local calculations not involving the entire grid, we prefer the weighted form of the posterior ensemble in equation 8.

In a similar way, the continuous ranked probability score (CRPS) measures the integrated squared distance between the ensemble-based predictive cdf and the step indicator function at the observation. With y^{obs} either observed R_0 or G, we have

$$\text{CRPS}(\hat{F}, y^{\text{obs}}) = \int_{\mathcal{D}} [\hat{F}(y) - \mathbf{1}(y \geq y^{\text{obs}})]^2 dy. \tag{11}$$

20

In our setting, the integral in equation 11 is computed as a sum over the discretization of the cdf formed by the ensemble members. By doing this for many observations at many inline-crossline grid locations, we can compute the average CRPS. When comparing different models, smaller CRPS means better predictive performance.

RESULTS

The reference case we focus on is covering most of area of the grid representing the Alvheim field top-reservoir. The configuration of the reference case is as follows; The observation patch of size 16×16 grid nodes, corresponding to 60 inlines and crosslines, is used to update a smaller parameter patch of size 6×6 grid nodes. The cementation depth chosen for the reference case is set to be at $d^c = 2110\,\text{m}$, similar to Rimstad et al. (2012). Since we use larger patches of observations to update a smaller patch of observed data, a thin frame around the domain is not updated in the inversion and not visualized. However, we still show the location of one of the wells at the edge of the domain, as it has some influence on the update.

Using the LETKF we get $n_e = 100$ realizations of $\boldsymbol{x}(\boldsymbol{u})$ from the posterior on the spatial grid. Before the results are visualized, the realizations are transformed back to $\boldsymbol{r}(\boldsymbol{u})$, $\boldsymbol{u} \in \mathcal{D}$ as described in the prior model section in equations 2 and 3.

Figure 12 shows the posterior

ensemble mean for gas, oil and brine saturation, in addition to the uncertainty range shown as difference between 10th and 90th percentile, as described in the pre-

vious section. Similarly, the posterior mean and range of clay content are shown in Figure 13.

From Figures 12a–12c we see that oil and gas saturation are around zero for deeper parts, as in the prior model (Figure 7), and the conditioning has provided no significant update in these regions. In the lobes around the wells, there are clearly effects of the seismic AVO data conditioning. The posterior has much more detail about the oil and gas predictions than the prior (Figure 7). For the oil saturation, there are small sections of higher saturation scattered in the field, and increased saturation near the south-east well.

In the uncertainty plots in Figures 12d–12f, the uncertainty is reduced from the prior model (Figure 7d–7f). There is still much uncertainty for the gas and oil saturation at the edges of the lobes structures. For oil saturation, it is quite uncertain whether the small zones contain higher saturation of oil or not. These uncertainties near the gas and oil concentrations are mirrored in that for brine in Figure 7f.

The posterior for clay content in Figure 13 is changed from the prior (Figure 8) which showed rather low values of clay content across the field. In particular, clay content is higher at the edge of the shallower lobes. The increase in clay content occurs near parts where we predict higher saturation of oil, especially east of the oil discovery in the intra-lobe area. The prediction of the clay content mirrors the deposition patterns, and these results seem geologically plausible and in line with earlier works on Alvheim field, see for example Avseth et al. (2021).

In Figure 14 we show posterior ternary plots of oil, gas and brine saturations.

This is done for the same locations as in the prior (Figure 9). For the shallow location, seismic AVO data clearly indicate hydrocarbon presence. Also, the spread of the ensemble members is much narrower than for the prior, illustrating the conditional update performed through the LETKF. For the deeper location, we notice a more concentrated ensemble of high brine saturation.

Figure 15 shows the spread of the ensemble members for prior and posterior when we ignore the spatial correlation. For the shallow location, prior and posterior in Figures 15a and 15c show that there is a change after the update. However, one data point used in the update of the location is not giving so much information and the ensemble members indicate that there is a bit of each of the fluids in the mixture however more in the direction of gas, unlike from what we see in Figure 14a. For the deeper location in Figures 15d and 15f there is not much spread in the ensemble because there is strong prior information of brine and this is confirmed by the limited data.

Figure 16 shows the spread of the ensembles of both the prior and posterior along a chosen line in map view. This line is chosen because it crosses two of the wells, where the well outcomes are oil and gas (Figure 16a). In addition it crosses a region where we have a larger update due to the well with gas outcome around inline 700 and crossline 4900. Figures 16b–16d clearly show both a mean update with the AVO data and an uncertainty reduction from prior to posterior. For the gas saturation, in Figure 16b, the update is most visible in the part at inline 600–750, which corresponds to a region near a gas well. This indicates that information from both well data and

seismic data aid the prediction. Additionally, there is substantial update around inline 1000–1100, where the posterior mean of gas decreases and the update indicates increased oil saturation - Figure 16c. The clay content, shown in Figure 16f changes significantly after the update, as it has rather high uncertainty in the prior ensembles. The clay content increases around the oil well, and some at the edge of the lobes, close to a gas well.

DISCUSSION

We perform validation and sensitivity analysis for multiple cases including comparison with the IEnKS method. We test the sensitivity to prior correlation length, configurations of patch sizes, number of iterations and different cementation depths. Patch sizes are varied only for the ETKF type of method, while for the IEnKS, the patch size is set and same for both observations and parameters, and only number of iterations is varied.

Figure 17 shows that the integrated attributes are sensitive to the cementation depths. There is a bit less oil and a lot less integrated gas for a slightly larger value of the cementation depth. With larger cementation depth, there tends to be less of both, the decrease being significantly more dramatic for integrated gas. The decrease in the integrated gas with deeper cementation point indicates that we predict more gas when the rock is stiffer. However, overlap between the fluids is also possible due to higher stiffness of the rock and the oil being squeezed in the middle as illustrated in the curves of Figure 10.

Figure 17 shows predicted values of integrated oil and gas, with the standard deviation in parentheses. This is a measure of the total top-reservoir content of each element. Predictions of integrated attributes are calculated from the posterior ensemble, and we simply sum each reservoir variables over all grid nodes (not accounting for cell volume or expected depth of the reservoir unit).

The sensitivity to lower correlation range in Figure 17 is rather small and the uncertainty decreases. There appears to be an increase in oil and gas content, compared to the benchmark case, with larger correlation range in the spatial field.

The sensitivities to LETKF patch configurations are small for the integrated gas, where the uncertainty decreases some with the bigger observation and parameter patches. For the integrated oil on the other hand, the uncertainty increases with smaller observation patch, whereas increased patches seem to provide little change compared to the benchmark case.

The sensitivity to the number of iterations for IEnKS is also tested. With 2 (or 3) iterations, there is a tendency of getting more integrated oil and gas, than the benchmark case, and the uncertainty grows.

Using leave-one-out cross validation, as in equation 8, we study performance metrics of the statistical models. We compute the predictive percentiles of hold-out data and compare them to the theoretical ones. Based on about 4000 hold-out locations, the fraction of data below the 0.25 percentile in the predictive distribution should be near 25 %, and similarly for other percentiles to get ideal coverage.

In our setting, we have specified the prior mean and rock physics parameters from well logs. These wells are drilled at the structural high locations of the reservoir unit, and could give rise to a possible bias. In order to check this, we divide the depth sections into four bins and calculate the empirical percentiles for the hold-out data in different bins. We do this exercise for both R_0 and G. Values for all four bins for the base case are presented in Table 1. The percentiles for shallow parts (bin1 and bin2) are much better than the deeper ones (bin3 and bin4). This could come from sampling bias in the wells, challenges with the cementation model for deep locations, or the simplified caprock assumptions ignoring a potentially nonconstant caprock properties accounting for transition between smectite and illite above some parts of the reservoir.

Results of percentiles averaging and CRPS values of the shallower bin1 and bin2, are shown in Figure 18. On the x-axes of both plots, the theoretical percentiles are shown, while the y-axes indicates the empirical percentiles. In the ideal case, they should be the same, giving a centered straight line in these plots. We plot the values corresponding to different percentiles for all the cases. Focusing on the average of bin1 and bin2, we notice that values are a bit off the nominal levels, especially for the gradient G. The CRPS values are shown in the same color as the cases. We note that the model with smaller correlation and also the one with smaller observation patches get smaller CRPS values than the benchmark case, but they are not the ones showing the best results with respect to the percentile coverage. The case performing seemingly better for percentile ranges is the one with the largest correlation range,

but it has higher CRPS value than some others. One of the cases which stands out with poorer performance is that of the cementation depth being deeper than for the base case.

Iterations in the Kalman updates are also tested, and the empirical values corresponding to the 0.75th percentile are closer to the theoretical one, however the performance for the percentiles for the gradient are much poorer compared to the other cases. CRPS is worse when it comes to gradient G and better for zero-offset reflectivity R_0.

Overall, the validation and the sensitivity analysis indicate that changes in the forward model (cementation) are larger than those resulting from perturbations of statistical model parameters or LETKF tuning. However, the sensitivity and validation metrics show that it is not straightforward to determine which configuration of the model parameters is the best, particularly if we consider how differently the cases perform for the two AVO attributes.

CLOSING REMARKS

We present a Bayesian model for seismic inversion of AVO data to predict partial oil and gas saturation and clay content. The model builds on random fields and rock physics relations that connect the mixed fluid and clay-sand properties to the seismic AVO data. The posterior model is explored by an ensemble-based method, where we use Monte Carlo samples from prior and likelihood model to train the correlations between the reservoir variables of interest and the seismic AVO data. In our setting

with partial continuous sample space variables, this approach involves little manual specification of the algorithmic parameters and one can easily obtain posterior samples in local patches of the seismic inline-crossline grid.

We implement our approach for the top-reservoir domain of the Alvheim field in the North Sea. Well log information is used to specify model parameters and condition the saturations at well locations. We notice a considerable change in the posterior samples compared to the prior samples, indicating that there is much information in the seismic AVO data. Insight gained from our inversion results include potential variation in the sand quality at locations in the field. This has an influence on the rock physics forward model, and is included in our formulation. Validation results are sensitive to the cementation depth used in the seismic inversion, and we notice potential bias in the predictions, most likely due to preferential sampling of the boreholes at structural highs. The Alvheim field has a complex geological history with both gas and oil contained in close proximity. There is still large uncertainty for oil and gas saturation as well as for clay content.

Our sensitivity analyses show that there is more sensitivity to the rock physics and cementation model, and future work is needed to constrain some of these parameters and build rock physics relations that cover different configurations. Related to the statistical approach, it would be interesting to expand such approaches to time-lapse situations where uncertainty quantification is highly relevant for both predictive assessments and monitoring questions.

We assume a Gaussian distribution for the logistically transformed oil and gas

saturations and clay content. The depth varying means and the variance terms were estimated from well log data. One could potentially fit non-Gaussian prior distributions for these reservoir variables using the empirical marginal distributions seen in the well logs, but this would require much data and realistic assumptions about the depth trends. Such an approach would further need a more nuanced ensemble-based algorithm, possible using the normal score ensemble Kalman filter (Zhou et al., 2011).

ACKNOWLEDGMENTS

APPENDIX A

NOTATION

APPENDIX B

MONTE CARLO SAMPLING OF HIGH-DIMENSIONAL GAUSSIAN RANDOM FIELDS

Consider one field only, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ for the discretized variable on the inline-crossline grid. The spatial covariance matrix $\boldsymbol{\Sigma}$ is embedded on a torus structure to get at circulant matrix $\boldsymbol{C}$. The random variable of interest (transformed oil or gas saturation or clay content in our case), denoted $\boldsymbol{b} = (b_1, \ldots, b_N)'$, $N = n_1 n_2$, can be represented on the regular inline-crossline grid as $\boldsymbol{b}^m = (b_{0,0}^m, b_{0,1}^m, \ldots, b_{n_1-1,n_2-1}^m)$. Similarly, we arrange the covariance between $b_{0,0}^m$ and any other variable, as they are positioned on the gridded torus, and this forms the covariance entries of $\boldsymbol{C}$ in an $n_1 \times n_2$ matrix which we denote by $\boldsymbol{C}^m$. Next, dft2 denotes the two dimensional discrete FFT, i.e.

$$\text{dft2}(C^m)_{j_1', j_2'} = \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} C_{j_1', j_2'}^m \exp\left[-2\pi\iota \left(\frac{j_1 j_1'}{n_1} + \frac{j_2 j_2'}{n_2}\right)\right], \qquad \text{(B-1)}$$

for $j_1' = 1, \ldots, n_1$, $j_2' = 1, \ldots, n_2$, with $\iota = \sqrt{-1}$. Further, idft2$(\boldsymbol{d}^m)$ denotes the two dimensional inverse discrete FFT of $n_1 \times n_2$ matrix $\boldsymbol{d}^m$. Having prior mean $\boldsymbol{\mu}$, with size $n_1 \times n_2$ matrix associate $\boldsymbol{\mu}^m$, an unconditional sample is generated as follows:

$$\boldsymbol{b}^m = \boldsymbol{\mu}^m + \text{Re}\{\text{dft2}[\sqrt{\text{dft2}(\boldsymbol{C}^m)} \odot \text{idft2}(\boldsymbol{w}^m)]\}, \qquad \text{(B-2)}$$

where the square root works elementwise on the matrix and $\odot$ is elementwise multi-plication. Further, $\boldsymbol{w}^m$ are independent standard Gaussian variables of size $n_1 \times n_2$.

Well log data are available in the four wells, and we denote the logistically trans-formed variable (oil and gas saturation) in the wells by $\boldsymbol{y}_w$. To generate a conditional sample, given well information $\boldsymbol{y}_w$, we assume that the data are measured with in-dependent noise having small standard deviation $\tau^2 = 0.01$. A conditional sample is then obtained from the well log data and the unconditional sample using FFT routines:

$$
\begin{aligned}
\boldsymbol{x}^m &= \boldsymbol{b}^m + \mathrm{Re}\{\mathrm{dft2}[\mathrm{dft2}(\boldsymbol{C}^m) \odot \mathrm{idft2}(\boldsymbol{z}^m)]\}, \\
\boldsymbol{z} &= \boldsymbol{A}^T[\boldsymbol{A\Sigma A}^T + \tau^2 \boldsymbol{I}]^{-1}(\boldsymbol{y}_w - \boldsymbol{Ab}).
\end{aligned}
\tag{B-3}
$$

Here, $\boldsymbol{A}$ is a $4 \times N$ selection matrix picking the well locations and $\boldsymbol{x}^m$ and $\boldsymbol{z}^m$ are the matrix associate of vectors $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively. Note that this vector $\boldsymbol{z}$ only has four non-zero entries and the matrix $[\boldsymbol{A\Sigma A}^T + \tau^2 \boldsymbol{I}]$ is only 4×4, so this part is easily computed.

The procedures in equations B-2 and B-3 are repeated n_e times to get independent initial ensembles $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_e}$.

APPENDIX C

ROCK PHYSICS FORWARD MODEL

Rock physics forward model consists of multiple interconnected layers connecting physical properties. Model parameters, fluid saturation and clay content are taken in as input at different stages of the modelling. Due to the known existence of cementation effects in the area, we use models from contact theory as this first layer of our forward model.Quartz cementation is assumed to start at a certain depth d^c, which is unknown but can be estimated from a given temperature gradient. Above this cementation depth we use the unconsolidated sand model, while below d^c, we account for the cementation effects on the grains, and Dvorkin-Nur contact cement model is used, for overview see Lehocki and Avseth (2021).

As the assumed type of compaction changes at the given cementation depth, meaning the transition from mechanical compaction due to the effective stress, to the chemical compaction controlled by temperature and time, it is necessary to match the boundaries, by correcting for the pressure effect at the start of the cementation in the contact cement model, ensuring a smooth transition between the domains (Avseth and Lehocki, 2016; Torset et al., 2021).

In the unconsolidated sand model, assumption of the nonuniform contacts between the grains is made and accounted for through introduction of the volume fraction of no-slip contacts, f_t, see Bachrach and Avseth (2008). This quantity is determined through known geological properties of the area. For the contact cement model, the

type of cement deposition is specified and assumed to be evenly deposited on the grain surface, which is referred to as scheme 2 in Mavko et al. (2020). For more around these rock physical mixing properties on the Alvheim field, see Golikov et al. (2013). Both models are porosity dependent and we use empirical depth dependent models for porosity (Rimstad et al., 2012), with specified shale and sand porosities:

$$
\phi_{sh}(\cdot, d) = \phi_{sh}^0 \exp\left[-\alpha_{sh}\left(d - d^0\right)\right]
$$

$$
\phi_{ss}(\cdot, d) = \begin{cases} \phi_{ss}^0 \exp\left[-\alpha_{ss}\left(d - d^0\right)\right] & \text{if } d \leq d^c \\ \phi_{ss}(d^c) - \kappa_{ss}(d - d^c) & \text{if } d > d^c. \end{cases} \tag{C-1}
$$

Final porosity is computed as a weighted mean of sand and shale porosities

$$
\phi = V_{clay}\phi_{sh} + (1 - V_{clay})\phi_{ss}. \tag{C-2}
$$

Using the dry bulk and shear modulus provided from the contact theory models, Gassmann fluid substitution is performed to obtain the elastic parameters. The substitution is performed going from the dry rock to the one saturated with mixed saturation. Once the elastic parameters are obtained, these are provided as input into the last layer of the model. As the final layer of the forward model and inn order to obtain the data, zero-offset coefficient and the gradient, we use the Zoeppritz approximation by Shuey (1985), given as,

$$
R_0 = \frac{1}{2}\left(\frac{v_p - v_p^c}{\frac{1}{2}(v_p + v_p^c)}\right),
$$

$$
G = \frac{1}{2}\frac{v_p - v_p^c}{\frac{1}{2}(v_p + v_p^c)} - 2\frac{(v_s - v_s^c)^2}{\frac{1}{2}(v_s + v_s^c)^2}\left(\frac{\rho - \rho^c}{\frac{1}{2}(\rho + \rho^c)} + 2\frac{v_s - v_s^c}{\frac{1}{2}(v_s + v_s^c)}\right).
$$

The caprock properties used in the approximation are estimated from the well logs

34

and assumed constant over the whole grid, where $V_{\mathrm{p}}^{\mathrm{cap}} = 2650\,\mathrm{m/s}$, $V_{\mathrm{s}}^{\mathrm{cap}} = 1150\,\mathrm{m/s}$

and $\rho^{\mathrm{cap}} = 2350\,\mathrm{kg/m^3}$.

REFERENCES

Asch, M., M. Bocquet, and M. Nodet, 2016, Data assimilation: methods, algorithms, and applications: SIAM.

Avseth, P., A. Dræge, A.-J. van Wijngaarden, T. A. Johansen, and A. Jørstad, 2008, Shale rock physics and implications for avo analysis: A north sea demonstration: The Leading Edge, **27**, 788–797, doi: 10.1190/1.2944164.

Avseth, P., and I. Lehocki, 2016, Combining burial history and rock-physics modeling to constrain avo analysis during exploration: The Leading Edge, **35**, 528–534, doi: 10.1190/tle35060528.1.

Avseth, P., I. Lehocki, Ø. Kjøsnes, and O. Sandstad, 2021, Data-driven rock physics analysis of north sea tertiary reservoir sands: Geophysical Prospecting, **69**, 608–621, doi: 10.1111/1365-2478.12986.

Avseth, P., T. Mukerji, G. Mavko, and J. Dvorkin, 2010, Rock-physics diagnostics of depositional texture, diagenetic alterations, and. reservoir heterogeneity in high-porosity siliciclastic sediments and rocks - a review of selected models and suggested work flows: Geophysics, **75**, no. 5, A31–A47, doi: 10.1190/1.3483770.

Azevedo, L., and A. Soares, 2017, Geostatistical methods for reservoir geophysics: Springer.

Bachrach, R., 2006, Joint estimation of porosity and saturation using stochastic rock-physics modeling: Geophysics, **71**, no. 5, O53–O63, doi: 10.1190/1.2235991.

Bachrach, R., and P. Avseth, 2008, Rock physics modelling of unconsolidated sands: Accounting for nonuniform contacts and heterogeneous stress fields in the effective

media approximation with applications to hydrocarbon exploration: Geophysics, **73**, no. 6, E197–E209, doi: 10.1190/1.2985821.

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, A165–A176, doi: 10.1190/1.3478209.

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.

Buland, A., O. Kolbjørnsen, and H. Omre, 2003, Rapid spatially coupled avo inversion in the fourier domain: Geophysics, **68**, 824–836, doi: 10.1190/1.1581035.

Buland, A., and H. Omre, 2003, Bayesian linearized avo inversion: Geophysics, **68**, 185–198, doi: 10.1190/1.1543206.

Davies, T. M., and D. Bryant, 2013, On circulant embedding for gaussian random fields in r: Journal of Statistical Software, **55**, 1–21, doi: 10.18637/jss.v055.i09.

Eidsvik, J., P. Avseth, H. Omre, T. Mukerji, and G. Mavko, 2004, Stochastic reservoir characterization using prestack seismic data: Geophysics, **69**, 978–993, doi: 10.1190/1.1778241.

Eidsvik, J., D. Bhattacharjya, and T. Mukerji, 2008, Value of information of seismic amplitude and csem resistivity: Geophysics, **73**, no. 4, R59–R69, doi: 10.1190/1.2938084.

Evensen, G., 1994, Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics: Journal of Geophysical Research: Oceans, **99**, 10143–10162, doi: 10.1029/94JC00572.

Gineste, M., J. Eidsvik, and Y. Zheng, 2020, Ensemble-based seismic inversion for a stratified medium: Geophysics, **85**, no. 1, R29–R39, doi: 10.1190/geo2019-0017.1.

Golikov, P., P. Avseth, A. Stovas, and R. Bachrach, 2013, Rock physics interpretation of heterogeneous and anisotropic turbidite reservoirs: Geophysical Prospecting, **61**, 448–457, doi: 10.1111/j.1365-2478.2012.01073.x.

Grana, D., 2016, Bayesian linearized rock-physics inversion: Geophysics, **81**, no. 6, D625–D641, doi: 10.1190/geo2016-0161.1.

Grana, D., L. Azevedo, L. De Figueiredo, P. Connolly, and T. Mukerji, 2022, Probabilistic inversion of seismic data for reservoir petrophysical characterization: Review and examples: Geophysics, **87**, no. 5, M199–M216, doi: 10.1190/geo2021-0776.1.

Grana, D., T. Fjeldstad, and H. Omre, 2017, Bayesian gaussian mixture linear inversion for geophysical inverse problems: Mathematical Geosciences, **49**, 493–515, doi: 10.1007/s11004-016-9671-9.

Grana, D., T. Mukerji, and P. Doyen, 2021, Seismic reservoir modeling: Theory, examples, and algorithms: John Wiley & Sons.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007, Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter: Physica D: Nonlinear Phenomena, **230**, 112–126, doi: 10.1016/j.physd.2006.11.008.

Lang, X., and D. Grana, 2018, Bayesian linearized petrophysical avo inversion: Geophysics, **83**, no. 3, M1–M13, doi: 10.1190/geo2017-0364.1.

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.

Lehocki, I., and P. Avseth, 2021, From cradle to grave: how burial history controls the rock-physics properties of quartzose sandstones: Geophysical Prospecting, **69**, 629–649, doi: 10.1111/1365-2478.13039.

Li, H., 2007, Local ensemble transform kalman filter with realistic observations: University of Maryland, College Park.

Liu, M., and D. Grana, 2018, Stochastic nonlinear inversion of seismic data for the estimation of petroelastic properties using the ensemble smoother and data reparameterization: Geophysics, **83**, no. 3, M25–M39, doi: 10.1190/geo2017-0713.1.

Malinverno, A., and V. A. Briggs, 2004, Expanded uncertainty quantification in inverse problems: Hierarchical bayes and empirical bayes: Geophysics, **69**, 1005–1016, doi: 10.1190/1.1778243.

Mavko, G., T. Mukerji, and J. Dvorkin, 2020, The rock physics handbook: Cambridge University Press.

Rimstad, K., P. Avseth, and H. Omre, 2012, Hierarchical bayesian lithology/fluid prediction: A north sea case study: Geophysics, **77**, no. 2, B69–B85, doi: 10.1190/GEO2011-0202.1.

Shahraeeni, M. S., and A. Curtis, 2011, Fast probabilistic nonlinear petrophysical inversion: Geophysics, **76**, no. 2, E45–E58, doi: 10.1190/1.3540628.

Shuey, R., 1985, A simplification of the zoeppritz equations: Geophysics, **50**, 609–614, doi: 10.1190/1.1441936.

Thurin, J., R. Brossier, and L. Métivier, 2019, Ensemble-based uncertainty estimation in full waveform inversion: Geophysical Journal International, **219**, 1613–1635, doi: 10.1093/gji/ggz384.

Torset, S., R. M. Holt, and K. Duffaut, 2021, Integrating rock physics laboratory data and modelling to improve uplift characterization methodology: Geophysical Prospecting, **69**, 552–567, doi: 10.1111/1365-2478.12961.

Van Leeuwen, P. J., Y. Cheng, and S. Reich, 2015, Nonlinear data assimilation: Springer.

Zhou, H., J. J. Gomez-Hernandez, H.-J. H. Franssen, and L. Li, 2011, An approach to handling non-gaussianity of parameters and state variables in ensemble kalman filtering: Advances in Water Resources, **34**, 844–864, doi: 10.1016/j.advwatres.2011.04.014.

LIST OF FIGURES

percentile, PW(10,90).

8   Prior ensemble mean of the clay content and the associated uncertainty.(a) The prior ensemble mean of clay content. (b) The uncertainty in the clay content presented as the width between 10th and 90th percentile, PW(10,90).

9   Ternary plots show prior ensemble for the saturation at a (a) shallow location (800,4800) at 2112 m, and a (b) deep location (436,4532) at 2190 m.

10   Conceptual illustration of the transition between the mechanical and chemical compaction at the cementation depth.

11   Grid representation of the field, illustrating the idea of dividing of the grid into observation patches – white, which are centered around the smaller parameter patches – black.

12   Posterior ensemble mean of the fluid saturations and their associated uncertainties. (a)–(c) The posterior ensemble mean for the gas, oil and brine saturation after the conditioning on the seismic data. The well locations and their outcomes are shown. (d)–(f) The uncertainty in the three parameters is presented as the width between 10th and 90th percentile, PW(10,90).

13   Posterior ensemble mean of clay content with associated uncertainty. (a) The posterior ensemble mean for the clay content after conditioning on the seismic data. (b) The uncertainty is presented as the width between 10th and 90th percentile, PW(10,90).

14   Ternary plots show the posterior ensemble of the fluid saturation at two locations: (a) shallow location (800,4800) at 2112 m and (b) deep location (436,4532) at 2190 m.

15   Ternary plots show the prior and posterior ensemble of the fluid saturation at two locations, ignoring the spatial correlation. (a) Prior at a shallow location (800,4800) at 2112 m, (b) prior at a deep location (436,4532) at 2190 m, (c) posterior at a shallow location (800,4800) at 2112 m, and (d) posterior at a deep location (436,4532) at 2190 m.

16   (a) A line section of the grid nodes from the field, represented by the red line, including areas around the well log locations. The field in the background is posterior mean of brine saturation.(b)–(d) Prior and posterior ensemble spread along the red line showing the changes after the update with the seismic data. Prior and posterior means are also included. This is done for (b) gas saturation, (c) oil saturation and (d) clay content.

17   Tornado chart for values of integrated gas and oil for different cases. The black lines in the blue bars indicate the mean value, while the widths of the bars indicate 2 standard deviations on each side of the mean.

18   Deviation of empirical versus theoretical percentiles for the AVO attributes (a)R_0 and (b) G. Different colors signify various sensitivity tests, along with the associated CRPS for each case. Smaller CRPS means better predictive performance. The percentiles plotted are 0.25, 0.50 and 0.75.

LIST OF TABLES
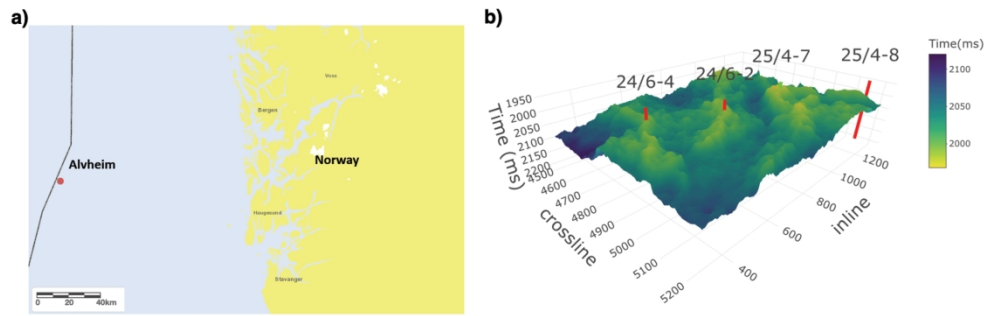
Figure_1_v Location of the Alvheim field off the Norwegian coast*. (b) Traveltimes to top-reservoir of the Alvheim field shown in an inline-crossline plane. Three of the gas wells are marked with red lines while the oil well is marked with a green line.

297x209mm (300 x 300 DPI)
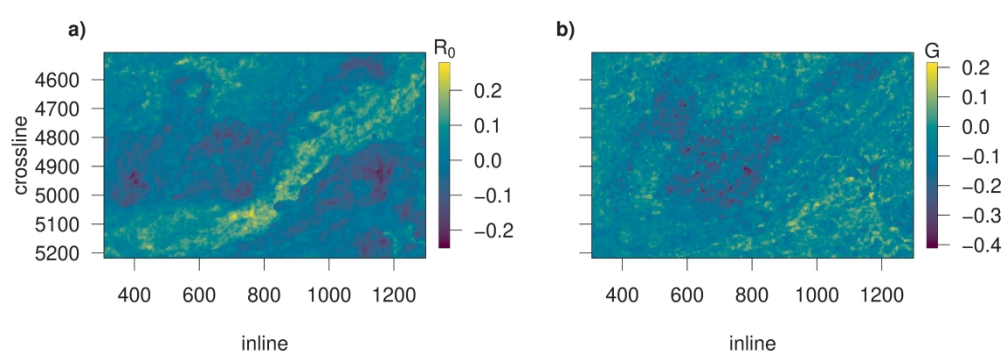
Figure_2_v AVO data presented in the inline-crossline coordinate system for the Alvheim field. (a) Unitless zero-offset intercept, R_0 . (b) AVO gradient, G.

419x152mm (300 x 300 DPI)
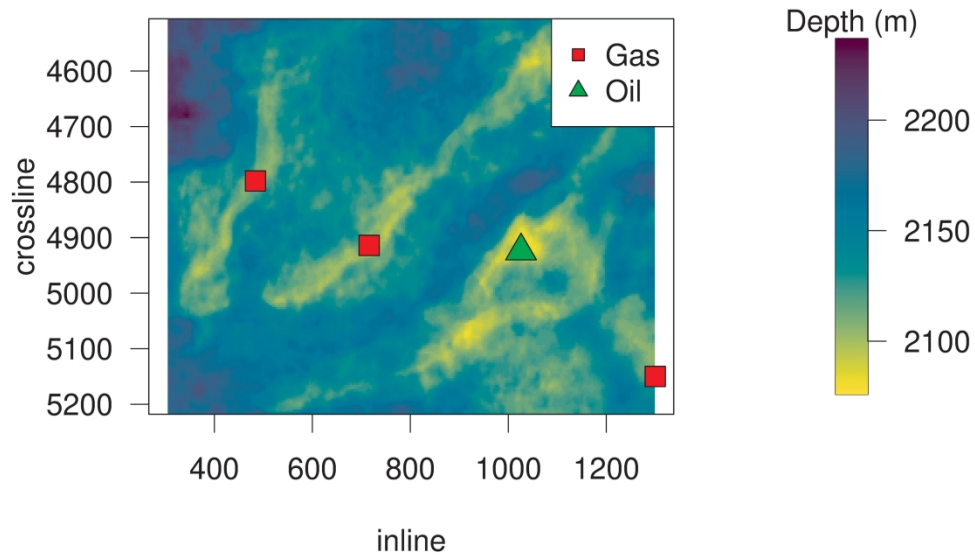
Figure_3_v Depth field of the top-reservoir of the Alvheim field on a grid with marked well locations and their fluid outcomes at the top-reservoir.

282x203mm (300 x 300 DPI)
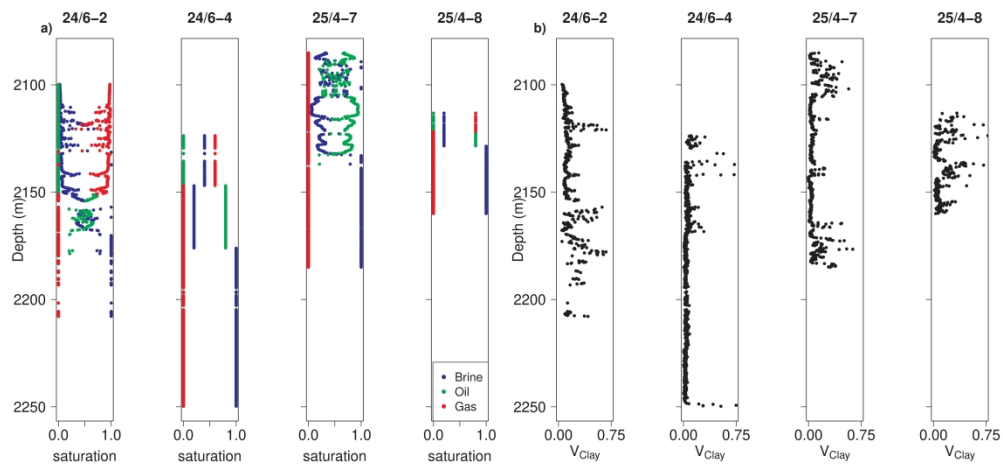
Figure_4_v Well logs of four different wells as a function of depth (MD-RKB).(a) Three different fluid saturations are shown: brine, gas, and oil. The values are scaled and shown in the range between 0 and 1 and they sum to 1 at each depth.(b) The fraction of clay content is shown with values ranging between 0 and 1.

533x254mm (300 x 300 DPI)
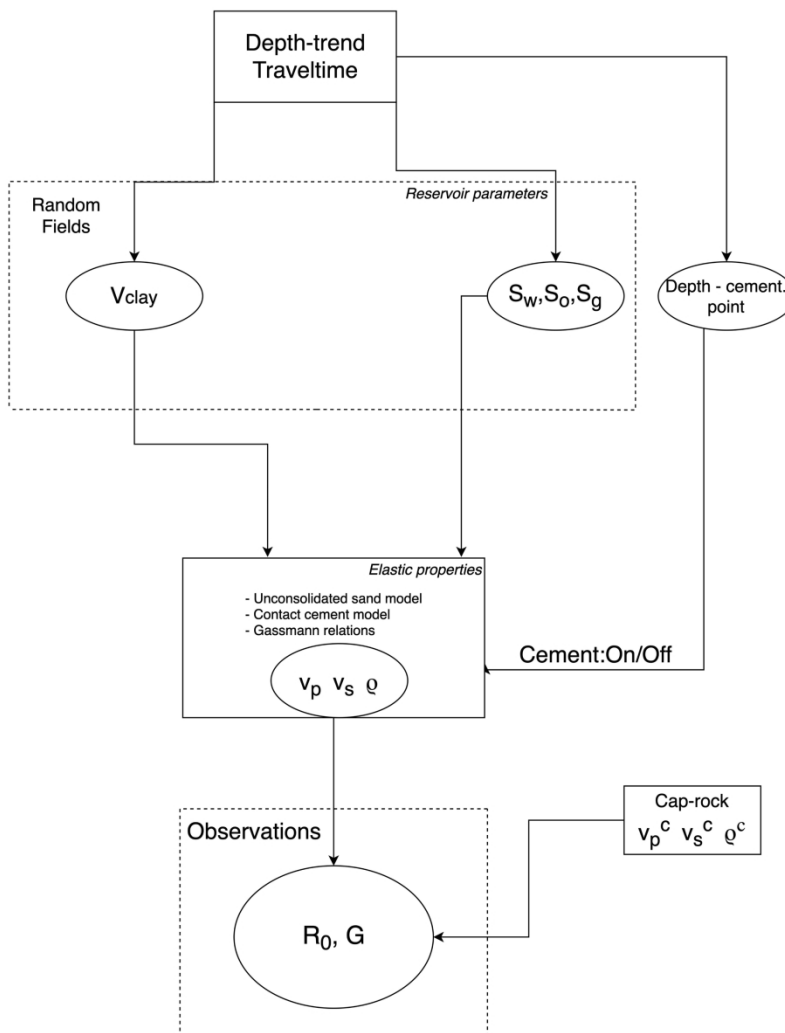
Figure_5_v Visualization of the model components and their relation.

209x297mm (300 x 300 DPI)
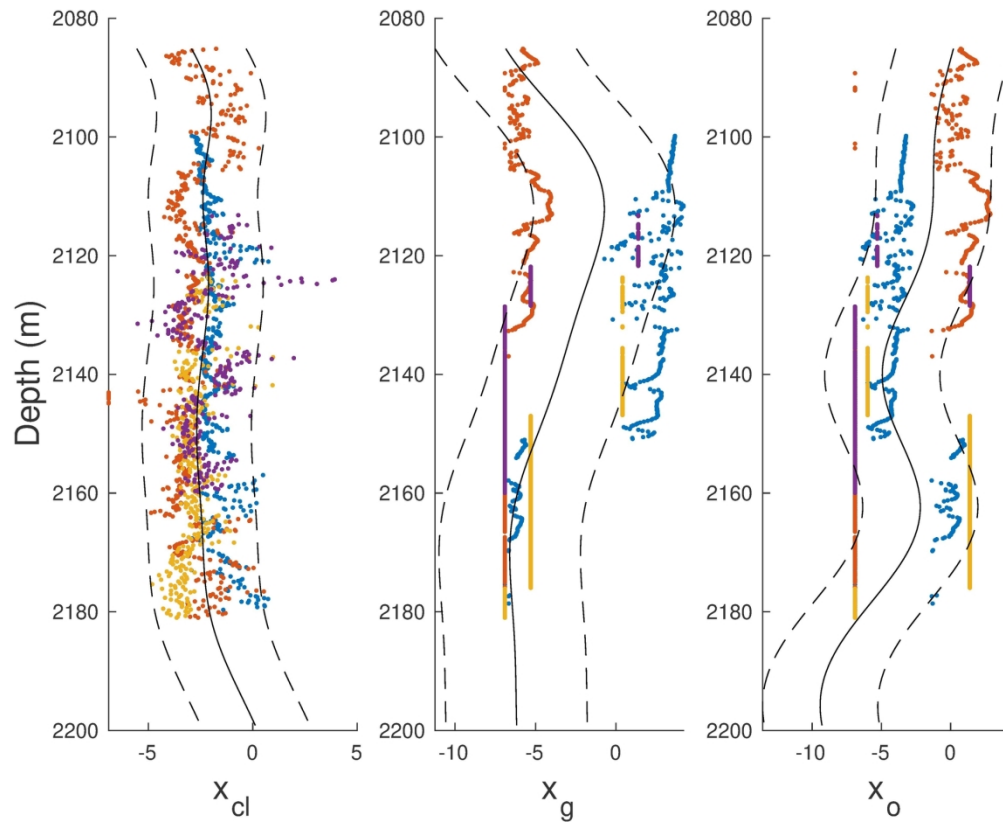
Figure_6_v Well data of transformed clay content and gas and oil saturation in the four wells (different colors). The fitted smooth background model (black lines) and its uncertainty (black dashed lines 90%) are constructed by a Gaussian process regression model to the well data as a function of the depth.

169x137mm (300 x 300 DPI)
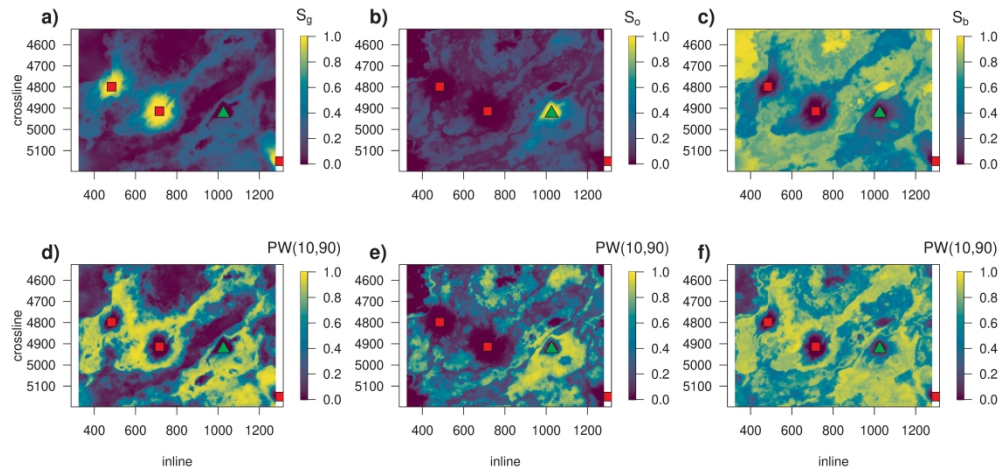
Figure_7_v Prior ensemble mean of fluid saturation and associated uncertainty. (a)–(c) Prior ensemble mean for gas, oil and brine saturation conditioned on the well-log outcomes at the well locations. The wells and their outcomes are marked with red squares representing gas and green triangle representing oil. (d)–(f) The uncertainty for gas, oil, and brine saturation is presented as the width between the 10th and 90th percentile, PW(10,90).

431x203mm (300 x 300 DPI)
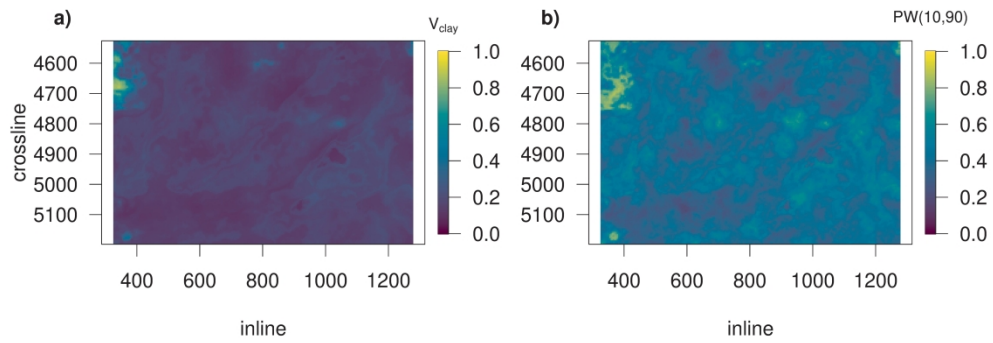
Figure_8_v Prior ensemble mean of the clay content and the associated uncertainty.(a) The prior ensemble mean of clay content. (b) The uncertainty in the clay content presented as the width between 10th and 90th percentile, PW(10,90).
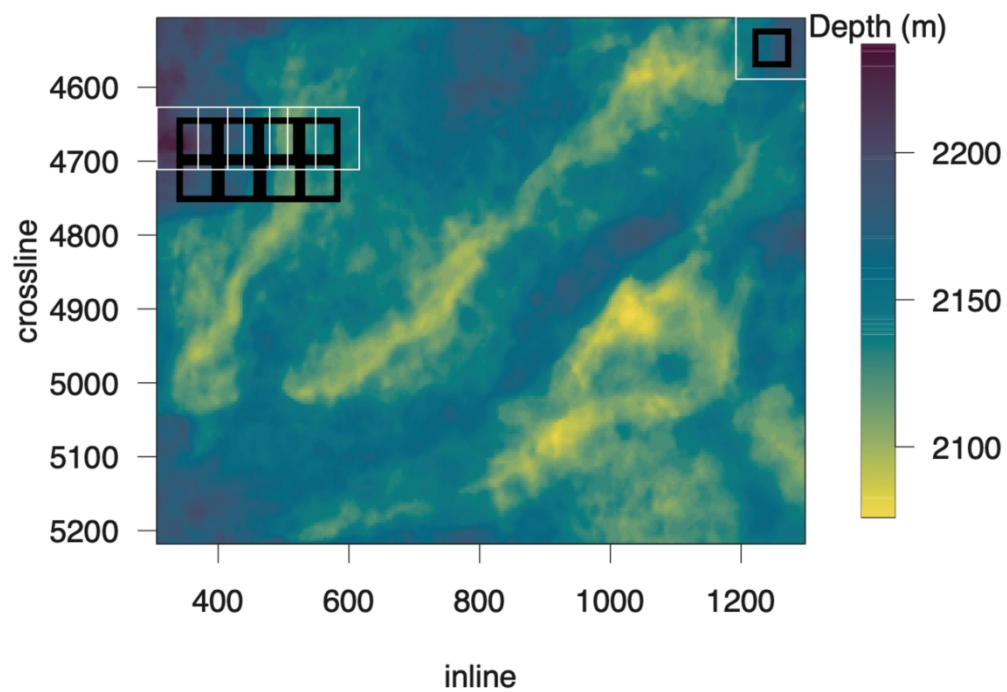
424x152mm (300 x 300 DPI)

a)

b)
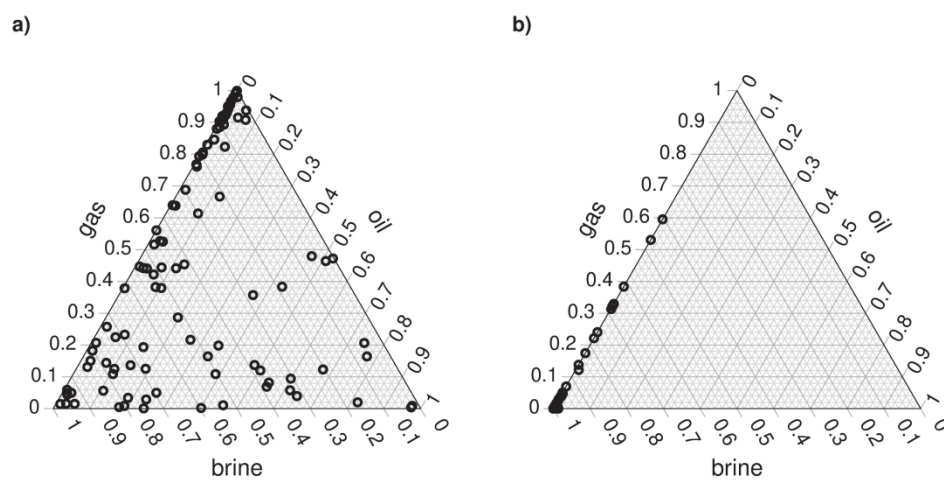
Figure_9_v Ternary plots show prior ensemble for the saturation at a (a) shallow location (800,4800) at 2112 m, and a (b) deep location (436,4532) at 2190 m.

304x175mm (300 x 300 DPI)

P−wave
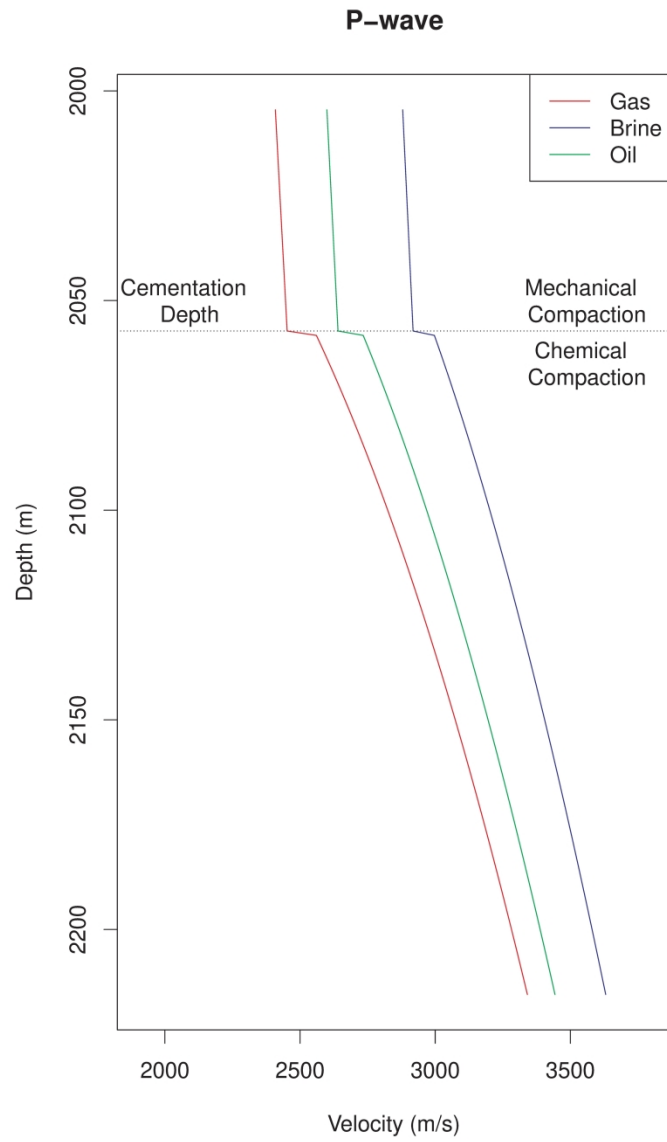
Figure_10_v Conceptual illustration of the transition between the mechanical and chemical compaction at the cementation depth.

228x381mm (300 x 300 DPI)

Figure_11_v Grid representation of the field, illustrating the idea of dividing of the grid into observation patches – white, which are centered around the smaller parameter patches – black.

209x162mm (300 x 300 DPI)
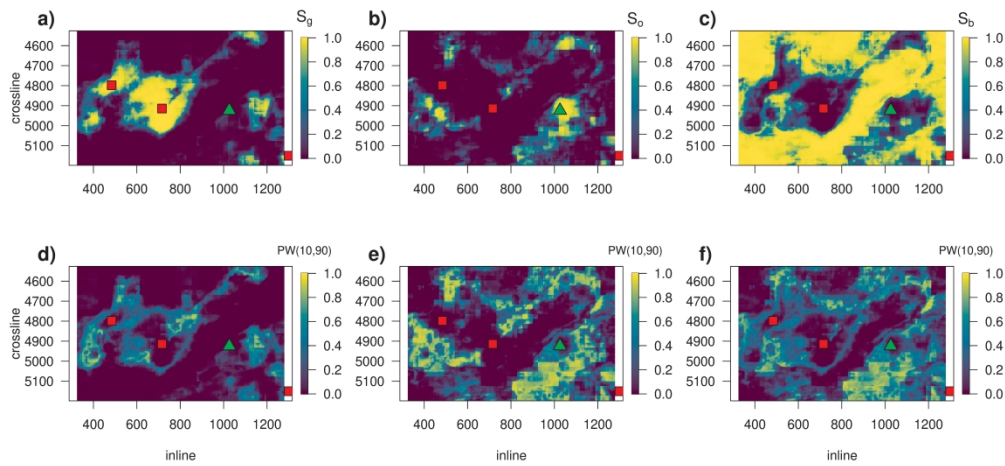
Figure_12_v Posterior ensemble mean of the fluid saturations and their associated uncertainties. (a)–(c) The posterior ensemble mean for the gas, oil and brine saturation after the conditioning on the seismic data. The well locations and their outcomes are shown. (d)–(f) The uncertainty in the three parameters is presented as the width between 10th and 90th percentile, PW(10,90).
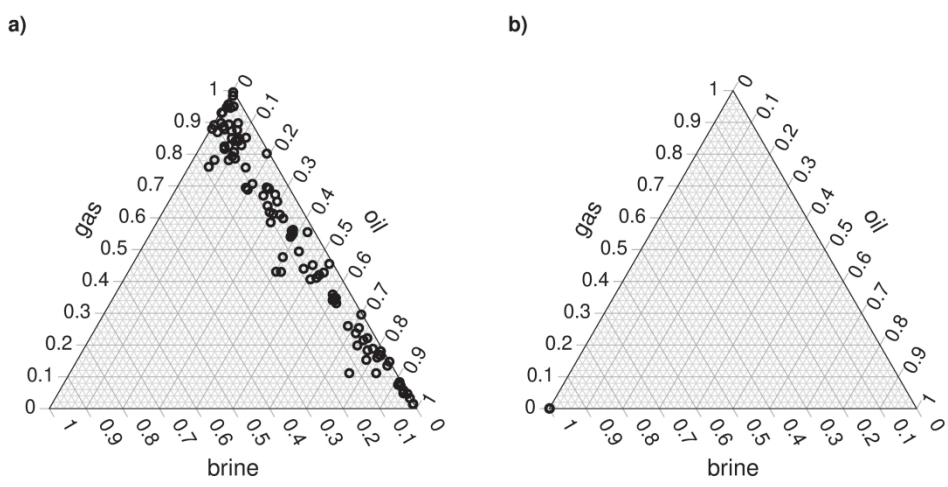
431x203mm (300 x 300 DPI)
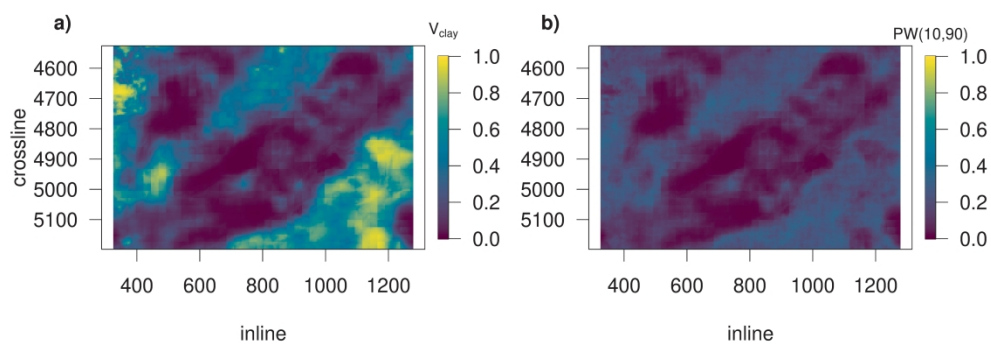
Figure_13_v Posterior ensemble mean of clay content with associated uncertainty. (a) The posterior ensemble mean for the clay content after conditioning on the seismic data. (b) The uncertainty is presented as the width between 10th and 90th percentile, PW(10,90).
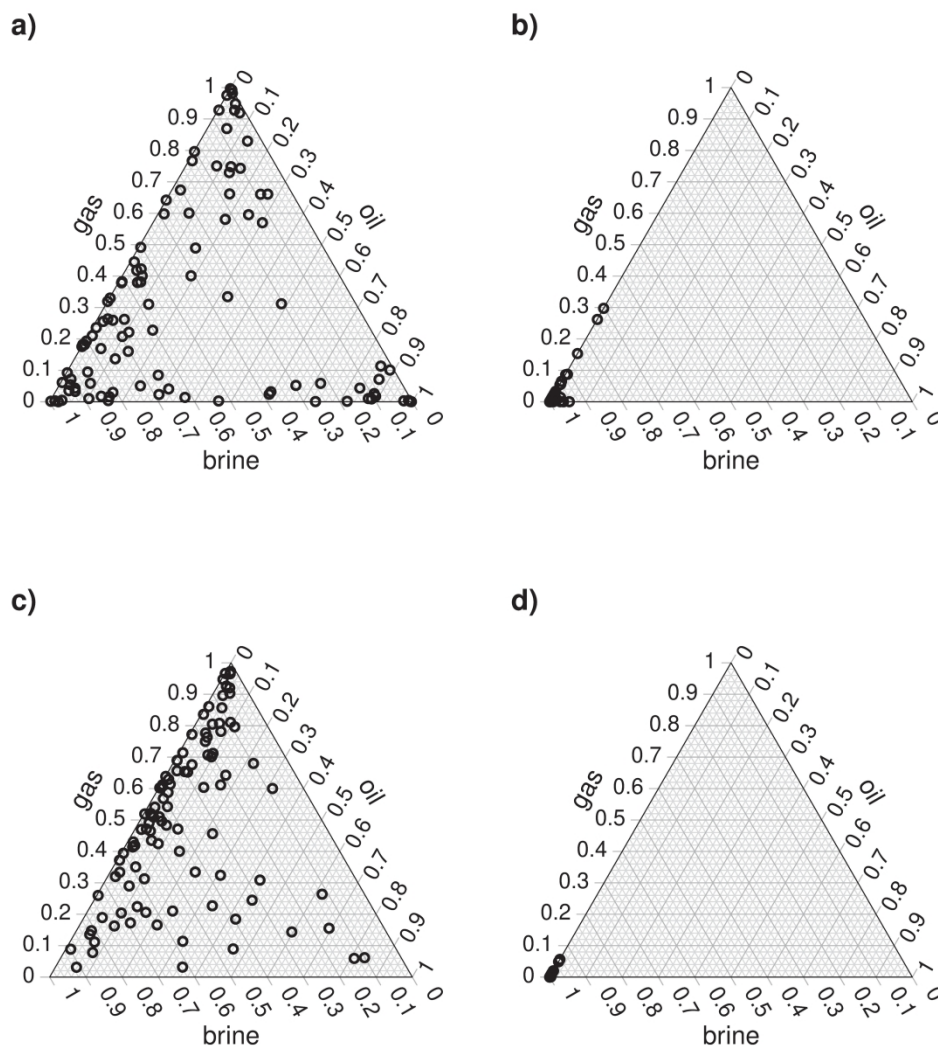
424x152mm (300 x 300 DPI)

Figure_14_v Ternary plots show the posterior ensemble of the fluid saturation at two locations: (a) shallow location (800,4800) at 2112 m and (b) deep location (436,4532) at 2190 m.

304x175mm (300 x 300 DPI)

a)

b)

c)

d)

Figure_15_v Ternary plots show the prior and posterior ensemble of the fluid saturation at two locations, ignoring the spatial correlation. (a) Prior at a shallow location (800,4800) at 2112 m, (b) prior at a deep location (436,4532) at 2190 m, (c) posterior at a shallow location (800,4800) at 2112 m, and (d) posterior at a deep location (436,4532) at 2190 m.

304x350mm (300 x 300 DPI)
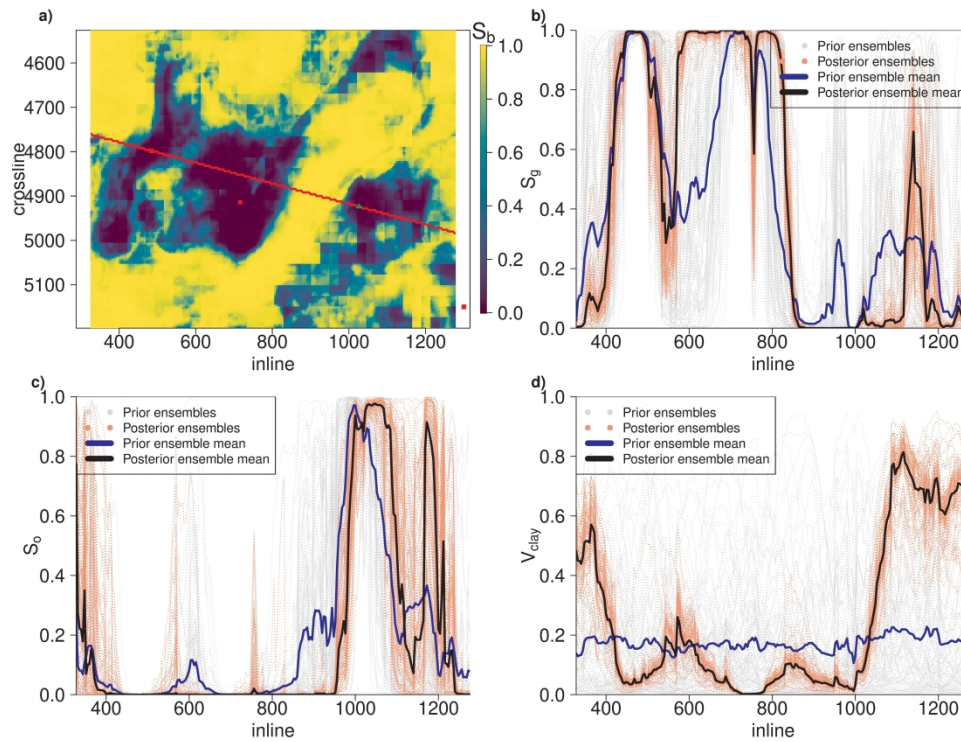
Figure_16_v (a) A line section of the grid nodes from the field, represented by the red line, including areas around the well log locations. The field in the background is posterior mean of brine saturation.(b)–(d) Prior and posterior ensemble spread along the red line showing the changes after the update with the seismic data. Prior and posterior means are also included. This is done for (b) gas saturation, (c) oil saturation and (d) clay content.

381x279mm (300 x 300 DPI)
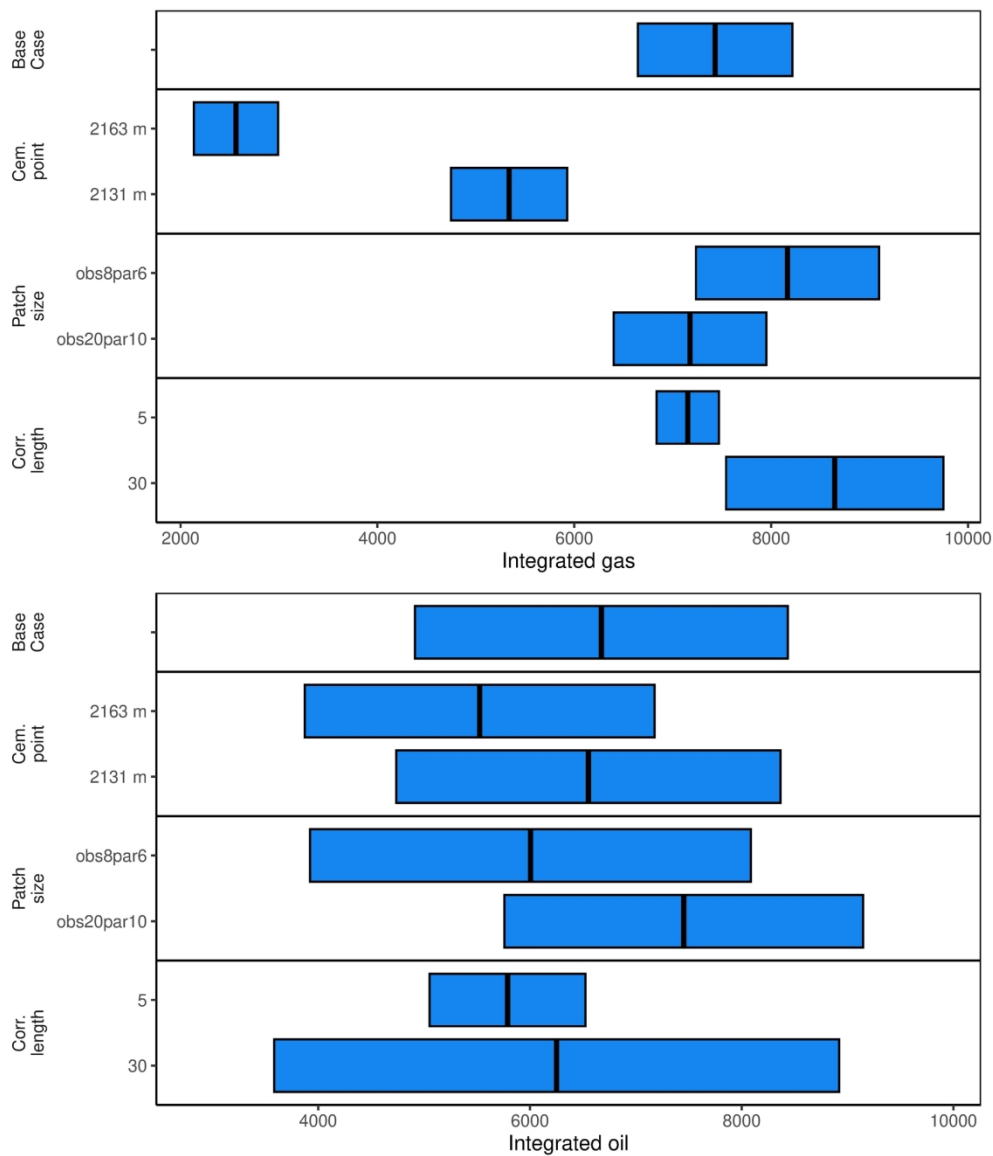
Figure_17_v Tornado chart for values of integrated gas and oil for different cases. The black lines in the blue bars indicate the mean value, while the widths of the bars indicate 2 standard deviations on each side of the mean.

185x215mm (300 x 300 DPI)
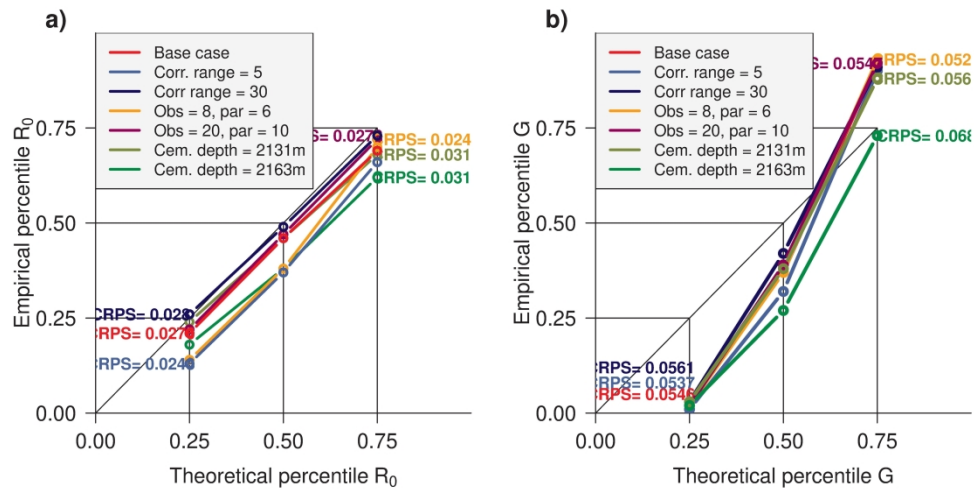
Figure_18_v Deviation of empirical versus theoretical percentiles for the AVO attributes (a)R_0 and (b) G. Different colors signify various sensitivity tests, along with the associated CRPS for each case. Smaller CRPS means better predictive performance. The percentiles plotted are 0.25, 0.50 and 0.75.

304x152mm (300 x 300 DPI)

	Percentile					
	R_0			G		
	0.75	0.50	0.25	0.75	0.50	0.25
Bin1	0.74	0.61	0.34	0.96	0.46	0.02
Bin2	0.63	0.30	0.07	0.87	0.28	0.01
Bin3	0.40	0.12	0.03	0.67	0.06	0.001
Bin4	0.27	0.02	0.0	0.47	0.0	0.0

Table 1:

N	number of inline and crossline grid cells
$1, .., j, .., J$	number of patches
n_e	number of ensembles
$\boldsymbol{u}$	inline-crossline location
$\boldsymbol{x} = (\boldsymbol{x}_{\mathrm{g}}, \boldsymbol{x}_{\mathrm{o}}, \boldsymbol{x}_{\mathrm{clay}})$	vector of transformed spatial
	gas and oil saturation and clay content
d	depth
d^{c}	cementation depth
R_0, G	zero-offset coefficient and gradient
$\boldsymbol{y}$	seismic AVO data vector
$\boldsymbol{r} = (\boldsymbol{S}_{\mathrm{g}}, \boldsymbol{S}_{\mathrm{o}}, \boldsymbol{V}_{\mathrm{clay}})$	vector of spatial reservoir variables
$\boldsymbol{S}_{\mathrm{b}}$	vector of spatial brine saturation
$\boldsymbol{\mu}_.$	mean vector of a given parameter/variable
$\boldsymbol{\Sigma}_.$	covariance matrix for a given parameter/variable
$\boldsymbol{y}^i, \quad i = 1, ..., n_e$	vector of simulated observations
$\boldsymbol{E} = (\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{n_e})$	ensemble of realizations from the prior model
$\boldsymbol{\Omega}$	observation error matrix

Table 2:

1