Assessing the value of seismic monitoring of CO_2 storage using simulations and statistical analysis

Susan Anyosa [a],*, Scott Bunting [a,b,1], Jo Eidsvik [a], Anouar Romdhane [c], Per Bergmo [c]

[a] Norwegian University of Science and Technology, Trondheim, Norway
[b] Visma, Oslo, Norway
[c] Sintef Industry, Trondheim, Norway

A B S T R A C T

Successful storage of CO_2 in underground aquifers requires robust monitoring schemes for detecting potential leakage. To aid in this challenge we propose to use statistical approaches to gauge the value of seismic monitoring schemes in decision support systems. The new framework is based on geostatistical uncertainty modeling, reservoir simulations of the CO_2 plume in the aquifer, and the associated synthetic seismic response for both leak and seal scenarios. From a large set of simulations we assess the leak and seal conditional probabilities given seismic data over time, and build on this to compute the value of information of the seismic monitoring schemes. The Smeaheia aquifer west of Norway is used to exemplify the approach for early leakage detection and decision support regarding CO_2 storage projects. For this case study, we find that the optimal monitoring time is about 10 years after injection starts.

1. Introduction

Geological carbon capture and storage (CCS) of CO_2 is considered as one of the most practical solutions to help meet the Paris agreement target on greenhouse gas emissions. The Sleipner project was the first industrial offshore project for CO_2 storage on the Norwegian continental shelf (Furre et al., 2017; Dupuy et al., 2017). Now, prospective operators are looking for new potential storage sites and one candidate is Smeaheia, Norway (see e.g. Ringrose et al., 2017; Dupuy et al., 2018). We will use Smeaheia as a case study in the current paper.

Our focus is on CO_2 leakage from an aquifer fault, which is one of the potential risks of CCS (Eiken et al., 2011; Jenkins et al., 2015). If CO_2 leaks from the aquifer and potentially to the surface, there are large additional costs associated with the storage operation and in some cases it might have to stop entirely. It is hence critical to look at monitoring strategies for early detection of leakage. For consistent evaluation of such monitoring schemes, we rely on simulations from realistic physical models of the subsurface, including uncertainty quantification. The approach presented here includes geostatistical methods for providing petrophysical properties, reservoir simulation, and rock physics models relating the CO_2 in the subsurface to the seismic response variables acquired at different monitoring times.

Building on a large number of simulation results, we use statistical learning methods to develop approaches for assessing the probabilities of early detection of leakage through a key fault at Smeaheia. This is done by computing the accuracy of characterizing the reservoir as leaking or sealing, given the simulated seismic data. We further conduct value of information (VOI) analysis of monitoring strategies (Eidsvik et al., 2015), considering an underlying decision situation connected to continued injection of CO_2, or termination of this process. In this setting, the geophysical monitoring has value if it leads to improved decisions for the injection program. A key question that we address in the paper is: when should one conduct seismic monitoring to optimally assist managers in this difficult injection decision under uncertainty?

Related research includes Sato (2011) which introduced VOI analysis in the monitoring of CO_2 storage. Trainor-Guitton et al. (2013) proposed a more nuanced VOI analysis methodology which consists of using numerical modeling of electrical resistivity data to detect CO_2/brine leakage. Other studies addressing this topic are Harbert et al. (2016) and Yang et al. (2018). Similar methods for VOI assessment have also been applied in reservoir monitoring for petroleum resources (Barros et al., 2016; Dutta et al., 2019a,b) and to evaluate distributed acoustic sensing (DAS) systems (Jreij et al., 2020). Our approach in the current paper is different from previous work in its focus on leakage monitoring and in

* Corresponding author.
  E-mail address: susan.anyosa@ntnu.no (S. Anyosa).
[1] Currently: Visma, Norway. Formerly: Norwegian University of Science and Technology.

how it combines reservoir simulation and synthetic seismic modeling with statistical learning methods and decision analytic concepts. The actual VOI calculation is in our case computed from the predicted probabilities on the hold-out test set, and differs from that of Jreij et al. (2020) using classification count results in the VOI expression.

This paper is organized as follows. In Section 2 we describe CO_2 storage and monitoring, focusing on the situation at the Smeaheia case. In Section 3 we outline our approach for simulation of reservoir properties and synthetic seismic amplitude data. In Section 4 the statistical learning approach for the estimation of classification probabilities is
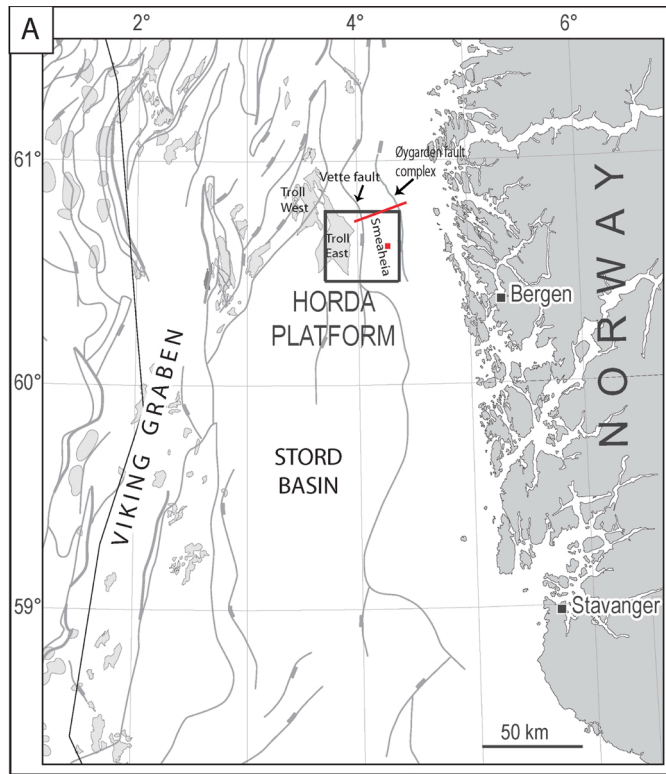
(a) Location of Smeaheia area. The red line depicts the seismic line and the red square depicts the injection point.
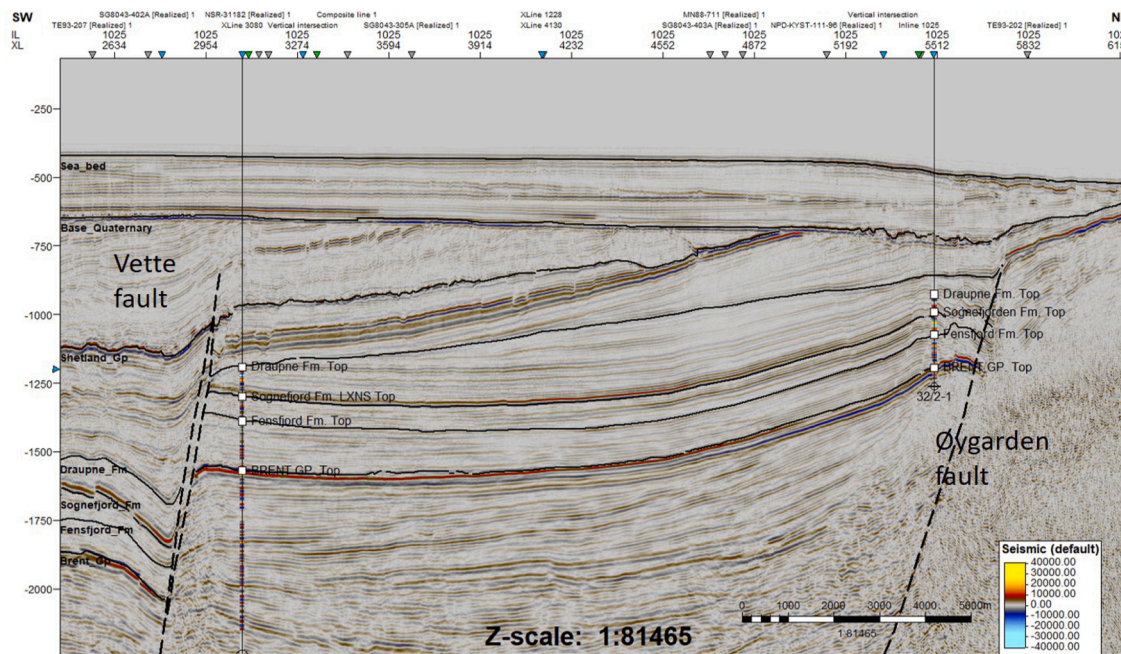
(b) Example of a 2D extracted seismic section from the Smeaheia area. The main faults, interpreted horizons, and well locations are indicated.

Fig. 1. Illustration of the Smeaheia case.

presented. In Section 5 the evaluation of the seismic monitoring strategy is formalized using VOI analysis. In Section 6 we discuss various assumptions used in the modeling and analyze the sensitivity of our results to seismic data accuracy. In Section 7 we summarize and provide perspectives for future work.

2. Background

2.1. CO₂ storage

For reliable storage of CO_2 in the subsurface, it is important to choose wisely among the possible storage resources available in different regions of the world. Possible storage locations include high-porosity and high-permeability saline aquifers with sealing cap rocks (Nordbotten and Celia, 2011).

The case study used in this paper is from Smeaheia offshore Norway (Fig. 1a). The Alpha and Beta structures are located East of the Vette fault and West of the Øygarden fault, respectively. The structures have been identified as possible locations for CO_2 storage. Fig. 1b is an extracted 2-dimensional seismic section showing the main faults, the key interpreted horizons, and the locations of the available wells in that area.

2.2. Monitoring

CCS operators are required to monitor storage safety with a long-term perspective. For large scale offshore CCS projects with significant upstream investments, this implies that tailored and cost effective measurement, monitoring and verification plans are of critical importance. Besides, an important challenge with subsurface storage of CO_2 is the possible leakage, which could ruin the purpose of the storage project. Hence, there has lately been large focus on how operators can conduct effective monitoring of CO_2 storage sites. Examples of approaches to develop monitoring strategies have been recently discussed in Sato (2011) and Dean and Tucker (2017). With this in mind it is of course useful to look at the experience that has been gained from recent projects (Ringrose et al., 2013; Furre et al., 2017), and in a recent overview of the qualitative monitoring components provided by Furre et al. (2020). But it is also highly relevant to use modern quantitative methodologies, data analysis tools and methods to study whether it will be possible to detect early-stage leakage in large scale industrial storage projects. We will use physical and mathematical modeling, coupled with statistical methods, to quantify the abilities of selected seismic monitoring strategies to detect leakage, and in doing so, evaluate whether seismic monitoring can be useful as an early warning decision support system.

Seismic reflection data are gathered by the emission of sound waves and recording the resulting echoes on a receiver grid. Depending on the application, the seismic data are processed in specific ways to describe (through interpretation) the geometry of the subsurface and/or characterize it in terms of spatial distributions of selected properties, such as saturation distributions in our case.

If the aquifer faults at Smeaheia are leaking, limitations in seismic vertical resolution suggest that it is difficult to detect leakage of CO_2 moving up along the fault. What can be possible to detect from seismic amplitudes is the relatively small changes in lateral extent of the CO_2 plume due to a change in the fault properties, e.g. from sealing to leaking cases. In this context, it might therefore be useful to focus the modeling, analysis, and monitoring efforts toward the top-aquifer properties (Dupuy et al., 2018).

As is common in geophysical data analysis for oil and gas exploration (Avseth et al., 2010), the seismic reflection data in CO_2 monitoring studies are processed to generate amplitude versus offset (AVO) data for each subsurface location. The resulting data are zero-offset reflectivity R_0 and AVO gradient G. When combined with appropriate rock physics models, these attributes constitute a very useful input to learn about fluid saturations (Dupuy et al., 2018). In our study, R_0 and G are extracted along the horizontal interface between the cap-rock formation and the top-aquifer zone.

2.3. Workflow

In the following sections, we discuss a workflow to estimate the value of monitoring schemes. The goal is to monitor possible leakage from one fault in the aquifer, and injection happen relatively near this fault. In doing so we use multiple realizations of the subsurface properties, reservoir simulation and seismic forward modeling, see Fig. 2. The fault probabilities of seal/leak are here set equal. This means that the fault has a large probability of leak, but the CO_2 volumes do not necessarily go to the surface, but migrate elsewhere in this aquifer or the structures above. A planned monitoring plan is hence considered for the aquifer, and a clear fault leak indication (as in the decision problem) would lead the company to stop injection until a thorough study is done as a secondary event triggered monitoring program, to further study possible CO_2 leakage from the entire structure (Furre et al., 2020).

In a setting with different probabilities, the workflow would be the same, but the probabilities must be weighted coherently according to their prior class probability in both the VOI sampling step and the conditional classification probabilities training.

Based on output simulations (Section 3) from both leaking and sealing fault scenarios, we use a statistical approach to see if seismic reflection data from the top-aquifer are useful for classifying the correct scenario, leakage or sealing (Section 4). We then couple this approach with the decision situation related to injection using VOI analysis on the monitoring time (Section 5).

3. Simulation of CO₂ saturation and seismic features

3.1. Physical modeling

Spatio-temporal reservoir simulation of the injected CO_2 is done by numerically solving the differential equations governing flow in porous media. Here, CO_2 is assumed to be injected from a well at the eastern part of the domain, at a depth of approximately 1250 m below the sea level, and the flux is dependent on amongst others the porosity and permeability in the aquifer. Many resources are available for conducting this kind of simulation. In this paper, we rely on the open-source Matlab Reservoir Simulation Toolbox (MRST) (Lie et al., 2012). The aquifer (reservoir) domain is discretized to a 3D rectangular grid and we use a
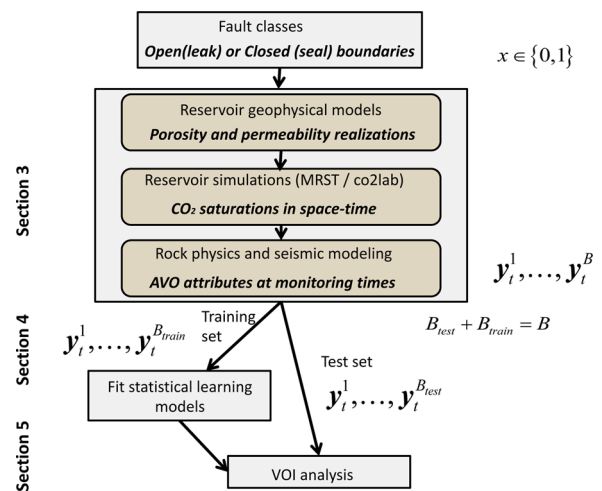

Fig. 2. Workflow describing the simulation of variables and data, and connections to the statistical analysis and VOI approaches where x denotes the classes (0: seal, 1: leak) and $\mathbf{y}$ denotes the seismic AVO attributes.

vertical equilibrium model in the simulations. Using the add-on module `co2lab` from MRST, we define a CO_2 injection rate of 8 million metric tons per year during 25 years (from the years 2021 to 2046). The injection well is located at cell (16, 10) in the size 18×18 lateral grid for the top-aquifer (see Fig. 3), with $2 \times 2 \, km^2$ grid cell size. In addition, we establish the reservoir boundaries to open or closed, corresponding to leaking or sealing faults. The outputs from the module are spatio-temporal CO_2 fluid saturation realizations in the aquifer. We simulate a total of 1000 realizations during the complete period of 25 years for both scenarios: open and closed faults. This set is balanced so that in the training there are 500 leak and 500 seal realizations.

The within-scenario variability is captured in the modeling by generating individual realizations of the uncertain porosity and permeability reservoir variables for each cell for each simulation run. This is done by specifying a constant mean in the porosity from well log data, and then simulating a truncated Gaussian process with a realistic level of spatial correlation and variance in a range set by log data. The mean value of the porosities is 0.28 and it was specified that they have to take values in the range [0.26, 0.30]. The permeability is linked to porosity using a calibrated empirical relationship found by regression from well log and core data (Dupuy et al., 2018). From the simulated porosity ϕ in a cell, we generated the permeability (in millidarcy), κ: $\kappa = 0.000029 \exp^{60.346\phi}$. The interval of porosities gives permeabilities in the interval [188.00 mD, 2112.4 mD]. Via the MRST simulations, these variable inputs will produce different CO_2 saturation (and pore pressure) maps in space-time.

Fig. 3 shows the average of the saturation realizations from CO_2 injection up to years 2029, 2035 and 2044 for the closed (first row) and open (second row) boundary cases. The plume is seen to grow away from the injection location, with some difference in the open and closed situations. The lateral extent of the plume is slightly larger for the sealing situation (top), especially at the latest time. CO_2 moves toward the boundary when the fault is open, so for the leaking situation (d–f), there is more CO_2 near the eastern boundary, and especially so at the latest time. However, it is not obvious to spot the minor differences.

In Fig. 4 boxplots of the saturation realizations for locations A and B (from Fig. 3) are plotted.

The saturations increase over the injection years, but the type of increase differs between the leak/seal situations. For the plume location (A), the sealing realizations have higher saturations, but with a rather large variability because of the reservoir uncertainty in porosity and permeability. For the location near the fault (B), the leaking realizations tend to have a higher saturation, again with rather large variability.

3.2. Seismic modeling

The seismic AVO data are modeled using rock physics relations. This involves conditioning on the CO_2 saturation levels and assigning a conditional distribution for the seismic zero offset reflectivity R_0 and AVO gradient G. In this model for the seismic response, conditional independence is assumed among all the grid responses, when the CO_2 is known. The modeling description is next hence focused on a single grid cell.

Fig. 5 shows a typical seismic AVO crossplot with R_0 on the first axis and G on the second axis. In this display, the joint bivariate predictive distribution of the seismic AVO response is plotted for three different CO_2 saturation levels.

Here, the mean values are calculated by the rock physics relations used in Dupuy et al. (2017). For no CO_2 saturation (black), the seismic AVO reflection is expected to have relatively low amplitude and small
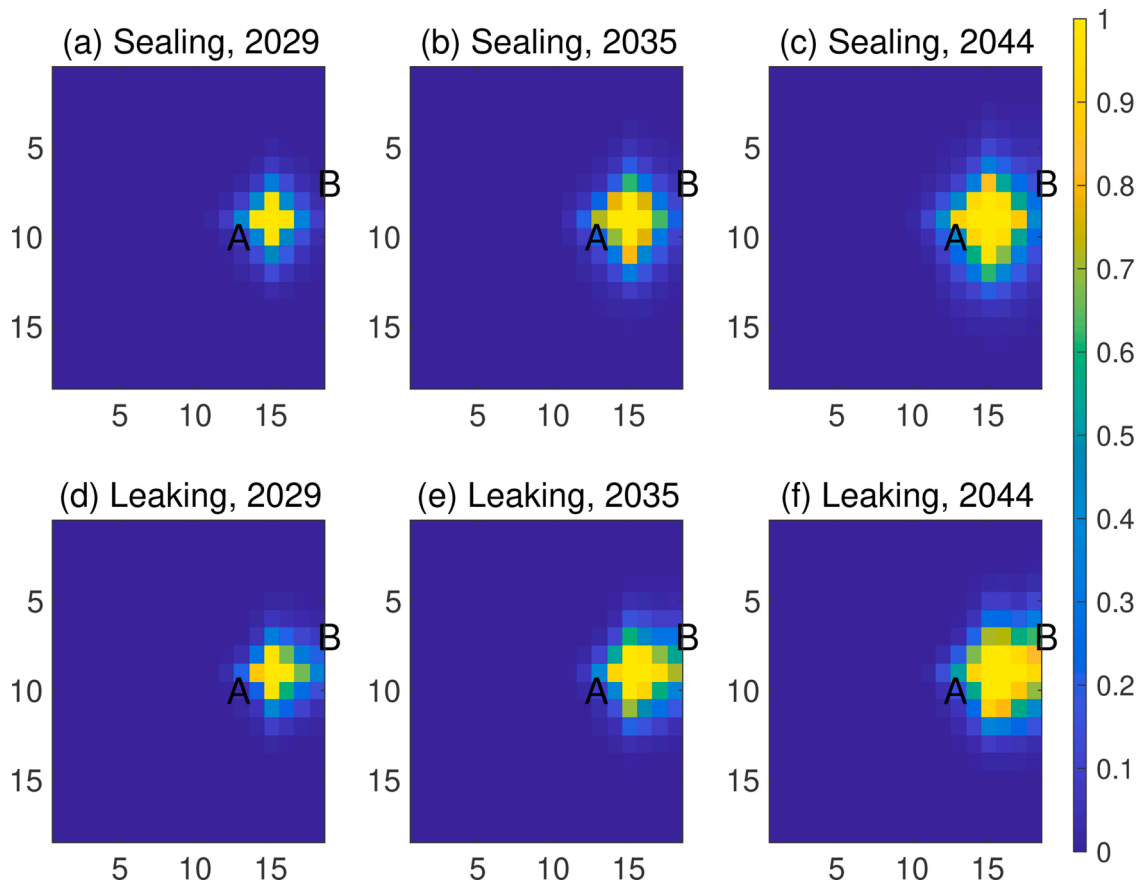

Fig. 3. Top-aquifer average of CO_2 saturation values over all realizations. This is plotted for various years (left to right) and for sealing fault (top) and leaking fault (bottom). A and B indicate two cells that we use to explain the plume development for the open and closed scenarios.
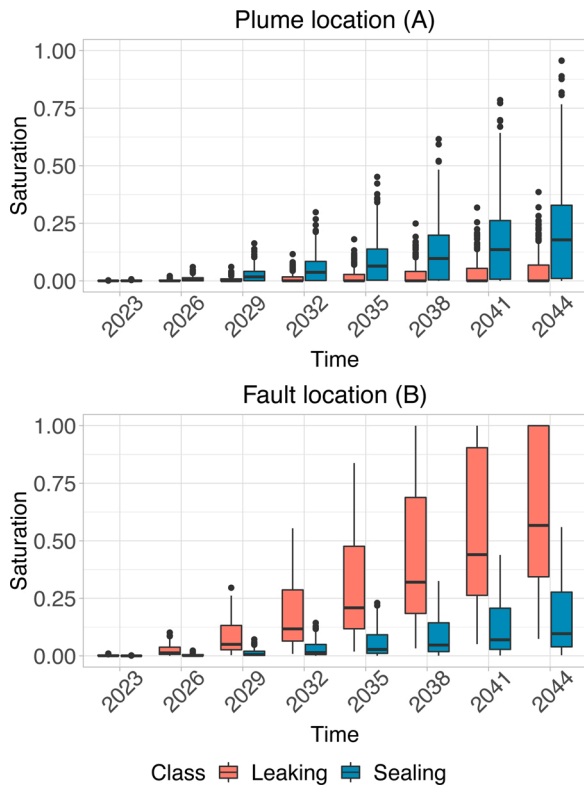
Fig. 4. Boxplots of saturation values over time in the plume and fault locations.
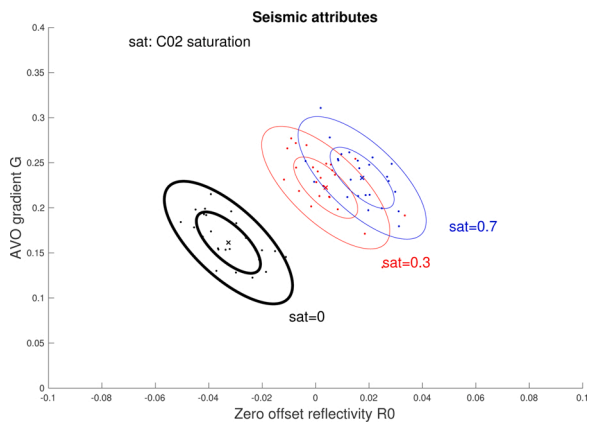

Fig. 5. Expected seismic AVO response (crosses) for different levels of CO_2 saturation along with 50% and 80% uncertainty contours in the seismic AVO observation model, and sample realizations (dots).

increase with angle. When the CO_2 saturation increases, the seismic response has a tendency of larger R_0 and G. The sensitivity to saturation in the seismic response occurs for rather small saturation levels (from 0 to sat = 0.3 in red), and then the additional shift is not so substantial (from 0.3 to sat = 0.7 in blue). These challenges of discriminating between high CO_2 saturations and the link to fluid mixing assumptions are discussed in Dupuy et al. (2017).

In order to generate AVO attributes we assume the processing to be noisy and to be normally distributed as $(R_0, G)^T|CO_2 \sim N(m(CO_2), T)$. The mean m is in this way calculated following the rock-physics equations of approximate relationships between AVO attributes and fluid properties (Avseth et al., 2010). For the covariance matrix T, the variances are here set to 0.006^2 for R_0 and 0.017^2 for G. The correlation in R_0 and G is set to -0.007. These uncertainty levels in the observation model are not straightforward to set from log and/or seismic data. They

would in general depend on the aquifer depth, overburden geometry, seismic acquisition design, and the heterogeneity in the elastic cap rock properties (which are assumed constant here). In the reference case of Fig. 5, the variances are 10 times smaller compared to what has been reported in petroleum exploration studies (e.g. Bachrach, 2006; Bhattacharjya et al., 2010). In our case, the top-aquifer is rather shallow (Dupuy et al., 2018), which rather suggests a relatively small level of noise in the associated seismic data. We therefore believe that the specified uncertainty levels are not unrealistic. We conduct a sensitivity analysis to the variability in the discussion section of the paper.

4. Classification probabilities

As previously stated, the injection duration in this application is considered to be 25 years, with the injection starting in 2021. During these years we consider to conduct one planned monitoring survey of the project to assess its state. We consider eight specific years for the planned monitoring survey: 2023, 2026, 2029, 2032, 2035, 2038, 2041 and 2044. We generate the AVO attributes for each saturation realization at these times. Since we have 1000 saturation realizations, we end up with eight arrays of 1000 R_0 values and eight arrays of 1000 G values. The generation of multiple realizations of AVO features covers the uncertainty in the domain of possible results derived from a seismic survey. We approach these features as Monte Carlo simulations and based on them, we present a variety of statistical learning methods for the classification of the seismic AVO features into seal or leak class. In doing so, we have to introduce the necessary mathematical notations summarized in Table 1. Some definitions include synonym terms in parenthesis. For example, the estimated probability of seal or leak scenario, given the seismic AVO features, is denoted $\widehat{P}(X = x|y_t)$, $x \in \{0, 1\}$.

4.1. Models

We now describe the statistical learning models that are used in the classification task. We set a common ground for the definition of the models: The partition of the whole dataset into a training and a testing set. With the training partition we build a statistical model and with the testing partition we evaluate the prediction performance of the model with previously unseen examples. The total size is $B = 1000$, which has partitions $B_{train} \approx 0.8B$ and $B_{test} \approx 0.2B$. Then we have $B_{train} + B_{test} = B$. The modeling is done in each time-step of monitoring independently from the other time-steps.

The chosen models are Gaussian process (GP), k-nearest-neighbors (KNN), random forest (RF), multi layer perceptron neural networks (MLP) and convolutional neural networks (CNN). We choose GP because one can intuitively consider the effect of the spatial correlations in the top-aquifer with this model. KNN, RF and MLP are considered because they are well-known benchmark methods, while CNN is chosen because of its common use in image classification analysis. For the tuning of the hyperparameters we test some variability around default software settings, aiming for good accuracy. For KNN, RF and MLP, we first consider default values given by scikit-learn (Pedregosa et al., 2011). Some of these values are then changed manually to improve the accuracy of the methods. For GP, different values of the spatial correlation parameter are considered. For CNN, an increase of the number of epochs was

Table 1
Glossary of notations.

Notation	Definition
t	Time-step of monitoring
s	Spatial location in the top-aquifer
b	Realization index
x	Scenario class (outcome)
y_t	Seismic response at time t (input)
$< y_t^b, x^b >, b = 1,...,B$	Dataset, instances

required to cope with the relatively small size of the training dataset. Note that none of these methods are exhaustively tuned, the purpose rather being to study variability around default models for classification and VOI results, in the setting of our workflow.

More details about the signatures of the models can be found in the code hosted in our repository.[2]

4.1.1. Gaussian process

A GP (Rasmussen and Williams, 2006; Banerjee et al., 2014) here means that at locations $s_1, \ldots, s_N$, the variable $y_t = \left(y_{t,1}, \ldots, y_{t,N}\right)$, $y_{t,n} = (R_{0,t,s_n}, G_{t,s_n})$, is distributed according to the multivariate Gaussian distribution. Knowing the scenario x, we use a GP to describe the variability and correlations in the seismic data response. This entails that the conditional probability density function (pdf) of the seismic data is $p_x(y_t) = \text{Normal}(y_t; \mu_{x,t}, \Sigma_{x,t})$, with mean vector $\mu_{x,t}$ and positive definite covariance matrix $\Sigma_{x,t}$. Here, the term 'Normal' denotes a multivariate Gaussian pdf.

For this GP model we specify the mean vector of each class $\widehat{\mu}_{x,t}$, $x \in \{0, 1\}$ from the empirical average of B_{train} training samples at each location at time t. Likewise, the diagonal terms of the covariance matrix $\widehat{\Sigma}_{x,t}$ are specified by the location-wise empirical variances. For the off-diagonal covariance entries we combine the empirical standard deviations with a specified Matern correlation function (Banerjee et al., 2014). The normalized classification probabilities are $\widehat{P}(X = x | y_t^b) = p_x(y_t^b)/(p_0(y_t^b) + p_1(y_t^b))$, where y_t^b is a test dataset at time (year) t.

4.1.2. k-nearest-neighbors

The kNN classifier (Hastie et al., 2009) uses a distance measure to identify the k nearest training points $y_t^b \in B_{train}$ to a query for prediction from the test partition. The classification to $x = 0$ or $x = 1$ is done by majority vote among the classes of these k neighbors. The classification probability $\widehat{P}(X = x | y_t^b)$ is equal to the relative number of votes for each class.

For this implementation we used scikit-learn, setting the number of neighbors to be equal to ten and using the Euclidean distance as measure. After the training and testing sets are split, a query for each one of the instances in the testing partition is generated. Then, the distance from the query to each one of the training points is computed. The value of k for the nearest points is tuned manually, and $k = 10$ is used to vote for the classification of the query.

4.1.3. Random forest

In RF (Breiman, 2001) many decision trees are grown. In the algorithm we draw M bootstrap samples from the original dataset $< y_t^b, x^b >$, $b \in B_{train}$. For each bootstrap sample, we grow a tree in which a subset of the predictors to be used are randomly sampled and the best split is chosen among them. The classification probability $\widehat{P}(X = x | y_t^b)$ is obtained from the mean of the classification probabilities of the trees in the forest.

To predict an outcome class ($x = 0$ or $x = 1$) the generated trees vote for the most popular class. RF is a powerful prediction tool that contributes to bias reduction compared with traditional parametric models.

For our implementation with scikit-learn, 100 decision trees are grown and the maximum depth of the decision trees is set to two. These hyperparameters are set manually, ensuring that high accuracy is obtained. Each of these trees predicts the class of the instance in the testing partition, therefore we have 100 predictions for each instance. The most popular class among the 100 predictions is selected as the classification for the instance.

4.1.4. Multi layer perceptron

Neural networks (Hinton, 1990) are nonlinear models represented by a network diagram which can solve both classification and regression tasks. They are composed by an input layer, hidden layer(s) and an output layer. In our case, the units in the input layer are the AVO features $y_{t,n}$, $n = 1, \ldots, N$. These inputs are linked to a hidden layer by an activation function f: $z = f(y_t, w)$, where w are weights from the connections between the units from the input layer and the hidden layer. Once the values of the hidden layer are obtained, z, they are used to calculate the classification probabilities $\widehat{P}(X = 1 | y_t^b)$ for test set data.

In our implementation we consider a MLP network with one hidden layer and use the ReLU (rectified linear unit) activation function and back-propagation in the training. Our parameterization uses a hidden layer of 20 nodes, a L_2 penalty equal to 0.0001 and the L-BFGS optimization algorithm.

4.1.5. Convolutional neural networks

CNN are popular for image classification. The main difference in architecture from a regular neural network consists in the introduction of a new kind of hidden layers called *convolutional layers*. The difference between a fully connected layer and a convolutional layer, is that fully connected layers learn global patterns in their input feature space, while convolutional layers learn local patterns in windows of the inputs. Jreij et al. (2020) used CNNs to evaluate various geophysical sensing techniques.

Our implementation of a CNN is done using Keras, where the realizations of seismic features are used as inputs. The convolutional layers use ReLU and max pooling, and dense layers are also included. With the limited data, we train the CNN for as much as 60 epochs. The implementation uses a batch size of 16.

4.2. Results

First, as part of an exploratory analysis, we compare specific cells from both scenarios. Notably, the VOI analysis below is done using all cells, but to gain insight in the problem, it helps to study what happens at individual cells as well. We choose to analyze one spatial location in the *plume* and one location near the *fault*. The selected grid cells are (10, 12) and (7, 18) in Fig. 3, denoted by letters A and B. For these two cells we compute the empirical distribution of R_0 over the B realizations for both seal and leak scenarios.

Fig. 6 depicts the empirical densities of R_0 for both scenarios (in pink and blue), for the two locations (plume left and fault right) and for two different years (2029 and 2044). In these displays, a classification boundary between the two classes is also shown (black vertical lines), where the GP classification boundary simply splits the variable domain (first axis) based on the largest probability density value (second axis), for leak an seal. In year 2029, the densities for the leak and seal cases are not very different, particularly for the grid cell located in the plume, suggesting that the classification task will not be straightforward. Although the seismic monitoring data are available at many cells, the responses are quite correlated and might therefore not carry as much information as one might expect. In year 2044, we notice more separation between the densities for leak and seal, and this helps in the classification task.

Next, we proceed with the analysis of the classification. In the testing partition we estimate $\widehat{P}(X = x | y_t^b)$, $x = 0, 1$ for each $y_t^b \in B_{test}$. The classes are assigned depending on which probability value is the highest. In a classification problem we can evaluate the performance of the prediction in the test set by counting the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) instances. Based on these counts we can compute the Accuracy score (ACC) which considers the total number of correctly classified instances and which is given by:
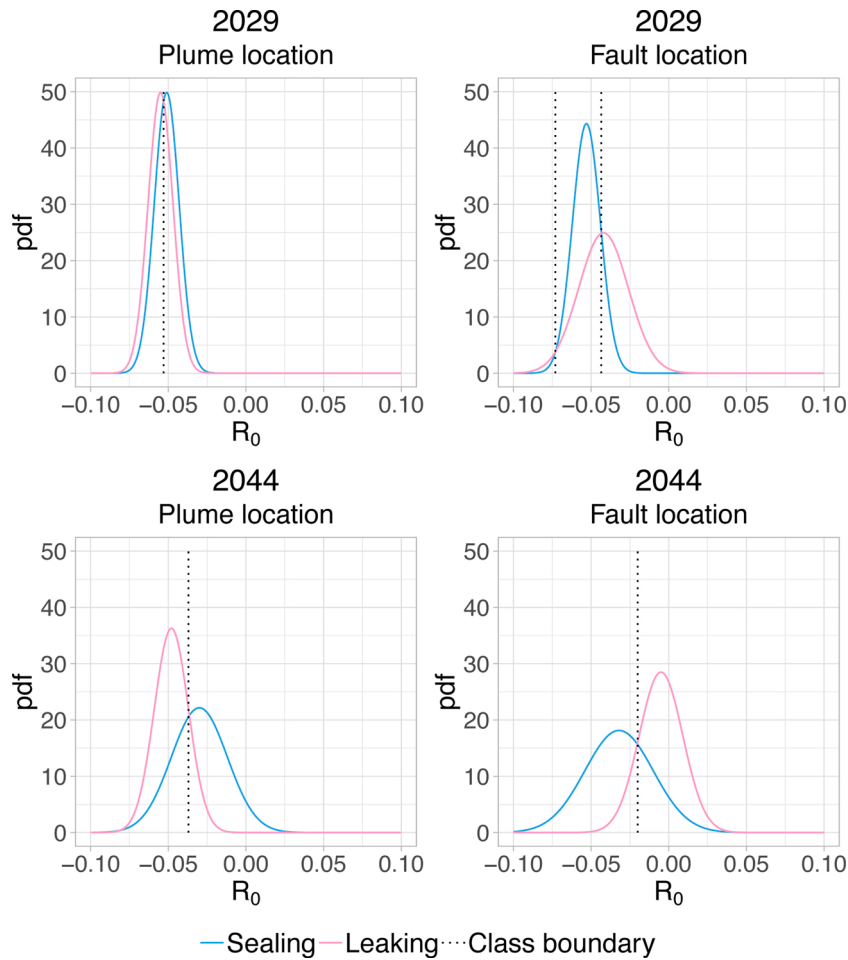

Fig. 6. Empirical distributions of R_0 for both leak and seal scenario, and GP classification boundaries (vertical lines). In plume location (left) and near fault location (right).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}.$$

A high value of accuracy means higher values of TP and TN and lower values of FP and FN. For our case study, a false positive indicates that a true sealing instance is predicted as leaking, whereas a false negative indicates that a true leaking instance is predicted as sealing. Since we want to avoid the high economic and environmental costs associated with not successfully identify leaking, we consider the false

negatives to be critical for a CO_2 storage project, but we frame this more clearly in the VOI analysis below.

Fig. 7 shows the accuracy values plotted over the years of monitoring, computed over ten Monte Carlo cross-validation samples. The curves tend to increase toward 1 as time goes by, with some random variation induced by the Monte Carlo sampling. Results indicate that it is easier to classify the leak and seal scenarios at later times, as we would expect. The curves generally increase faster toward 1 when both R_0 and G data are available (right display). The MLP approach appears to return
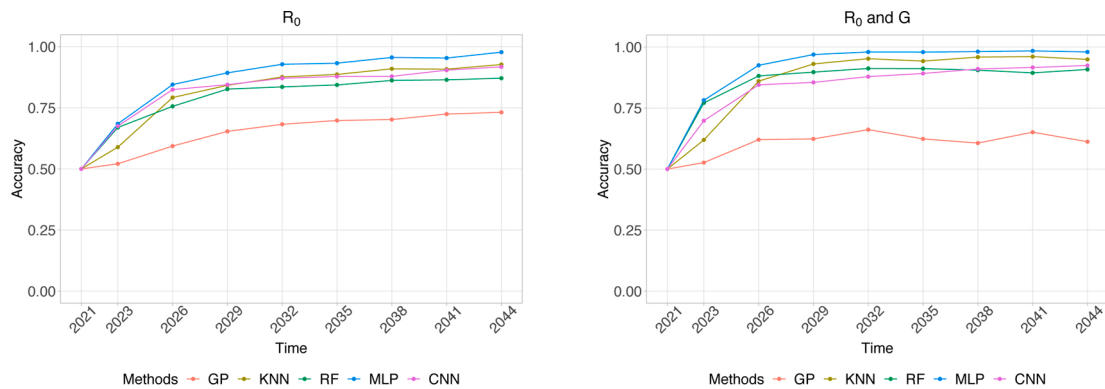
(a) Seismic zero-offset reflectivity data alone.　(b) Seismic zero-offset and AVO gradient data.

Fig. 7. Accuracy results of the different classification methods over monitoring year.

the largest accuracy values, possibly because the complexity of the problem requires rather complex interrelations as that provided by a non-linear network structure. The GP approach has the lowest accuracy values and seems more prone to Monte Carlo variability. This GP model performance would likely benefit from more refined parameter specification, but it does not seem to be the ideal model choice for characterizing the plume variability in our case.

To better assess the accuracy of these results, we also provide normalized histograms of the predicted probabilities of leakage in Fig. 8. Here, the color codes indicate the true class grouping in the test set: red is true leakage and blue is true sealing. The probabilities are close to 0.5 initially, but then move closer to the boundaries. The MLP results are better in the sense that the leakage probabilities move faster toward 0 or 1. Already for year 2029, a substantial separation is seen, but this is not that clear for GP, KNN, RF and CNN which perform about the same for this case.

Instead of looking at such accuracy and probability results one commonly studies other performance metrics that can be derived from the TP, FP, TN and FN (Powers, 2011). Our goal in this paper is to evaluate different monitoring times related to the CO_2 injection decision, so we next turn to VOI analysis of seismic AVO data in this context.

5. VOI analysis

In this section we go one step further in addressing the influence of monitoring. We build on the approach for computing classification probabilities for the test set to conduct VOI analysis of the seismic data.

5.1. Decision problem

Let us consider a company that wants to inject 8 million metric tons of CO_2 per year during 25 years in the aquifer at Smeaheia. Injection starts in 2021 and the decision situation is as follow: at time $t \in (0, 25)$, in the time from 2021 to 2046, the company has two alternatives $a \in \{0, 1\}$:

- stop injection at time t ($a = 0$),
- continue injection ($a = 1$).

At monitoring time *t*, we are imagining a one-time decision. This means that if alternative $a = 1$ is chosen, this will continue until the end of the 25 years period.

The choice between the stop or continue alternatives is difficult because there is uncertainty related to leakage. If the aquifer is leaking ($x = 1$), the company must pay a fine on top of the fixed cost of the injection operation. The costs in this problem are specified as follows, where 1 unit of CO_2 corresponds to the yearly injected 8 million metric tons:

- Fixed cost if injection is done: 5
- Cost of injecting per unit CO_2: 0.2
- Additional fixed cost if leakage: 2
- Fine if leakage per unit of injected CO_2: 1.2
- Cost of not injecting per unit CO_2: 0.8.

When injection has occurred through time *t*, the company would have injected *t* units of CO_2 at that time. The values $v_t(x, a)$ for the different alternatives and outcomes are then:

- $v_t(x = 0, a = 0) = -5 - 0.2t - 0.8(25 - t) = -25 + 0.6t$
- $v_t(x = 1, a = 0) = -5 - 0.2t - 0.8(25 - t) - 2 - 1.2t = -27 - 0.6t$
- $v_t(x = 0, a = 1) = -10$
- $v_t(x = 1, a = 1) = -42,$

where the value function for alternative $a = 1$ are the end values of the value function because of the continued injection.

These value functions are plotted in Fig. 9.

Here, for continued injection alternative, $a = 1$, values -10 and -42 represent the costs associated with continued injection for 25 years. If there is no leakage, this is just the initial costs of injection (5) plus the operating costs per unit $0.2 \cdot 25 = 5$. Even though the numbers used should not be taken in earnest, we foresee that CO_2 storage operators will have to deal with such difficult decision problems for the viability of future large scale CCS projects.

We assume that one must conduct further mandatory surveying as a secondary level after 25 years of injection, possibly with additional triggered monitoring components from the planned part we focus on here (Furre et al., 2020). Based on this second phase, we assume one will know for certain whether CO_2 has leaked or not. But the decision maker can benefit from doing the planned earlier monitoring at lower accuracy, and we analyze the optimal time of such monitoring. Hence, at the beginning of each time step *t* (year), the company has the option of conducting a seismic survey.

In such decision situations, analysts often compare the expected value with monitoring data with the expected value without any data. The prior value (PV_t) is obtained by finding the alternative with the lowest expected value (risk) at each time *t*. Without any survey data, the decision maker is indifferent to continued injection for all years or stop: The expected values of -26 are the same. Data can help the decision maker to stop the injection if the data indicate clearly that the fault is leaking, or to continue injecting with more certainty if the survey data indicate that there is no leakage. We consider the assumptions that the survey can only be performed at one time step *t*, and that the injection cannot be resumed once it has been stopped. Due to these constraints it will be useful to know when it is most valuable to perform this monitoring survey.

Our task is then to calculate the VOI of a seismic monitoring survey done at different times. The VOI is defined by the difference between the PV and the posterior value (PoV) with survey data $\mathbf{y}_t$. Mathematically, we have

$$\text{VOI}_t = \text{PoV}_t - \text{PV}_t \tag{1}$$
$$= \int max_a\{E[v_t(x, a)|\mathbf{y}_t]\}p(\mathbf{y}_t)d\mathbf{y}_t - max_a\{E(v_t(x, a))\},$$

where the PoV term involves informed decision making using the conditional expected value given the seismic data, while the PV is computed to assess the initial knowledge alone. But note that the PoV calculation is done in the prospective sense, so we must integrate over all possible seismic monitoring surveys that can be revealed, using the statistical model $p(\mathbf{y}_t)$ for these data. In practice these expressions are not straightforward to compute, and one must turn to approximations. In our setting, the approximate calculation is built from the simulations of CO_2 injection and synthetic seismic AVO data.

5.2. Approximating the VOI

The approximate conditional probabilities $\widehat{P}(X = x|\mathbf{y}_t)$ in Section 4 are used to approximate the VOI. For the posterior we have:

$$\text{PoV}_t = \int_{\mathbf{y}_t} max_{a \in A}\{\text{E}[v_t(x, a)|\mathbf{y}_t]\}p(\mathbf{y}_t)d\mathbf{y}_t \tag{2}$$
$$\approx \frac{1}{B_{\text{test}}} \sum_{b=1}^{B_{\text{test}}} max_{a \in A}\{\text{E}[v_t(x, a)|\mathbf{y}_t^b]\},$$

$$\text{E}[v_t(x, a)|\mathbf{y}_t^b] = \sum_{x=0}^{1} v_t(x, a)P(X = x|\mathbf{y}_t^b) \approx \sum_{x=0}^{1} v_t(x, a)\widehat{P}(X = x|\mathbf{y}_t^b).$$

For consistent approximation, the PV part is also assessed using the Monte Carlo realizations. By double expectation; $E[v_t(x, a)] = E_y\{E_{x|y}[v_t(x, a)|y)]\}$, and this is used to approximate the PV. By relying on the same Monte Carlo samples as for the PoV, we form the expectation
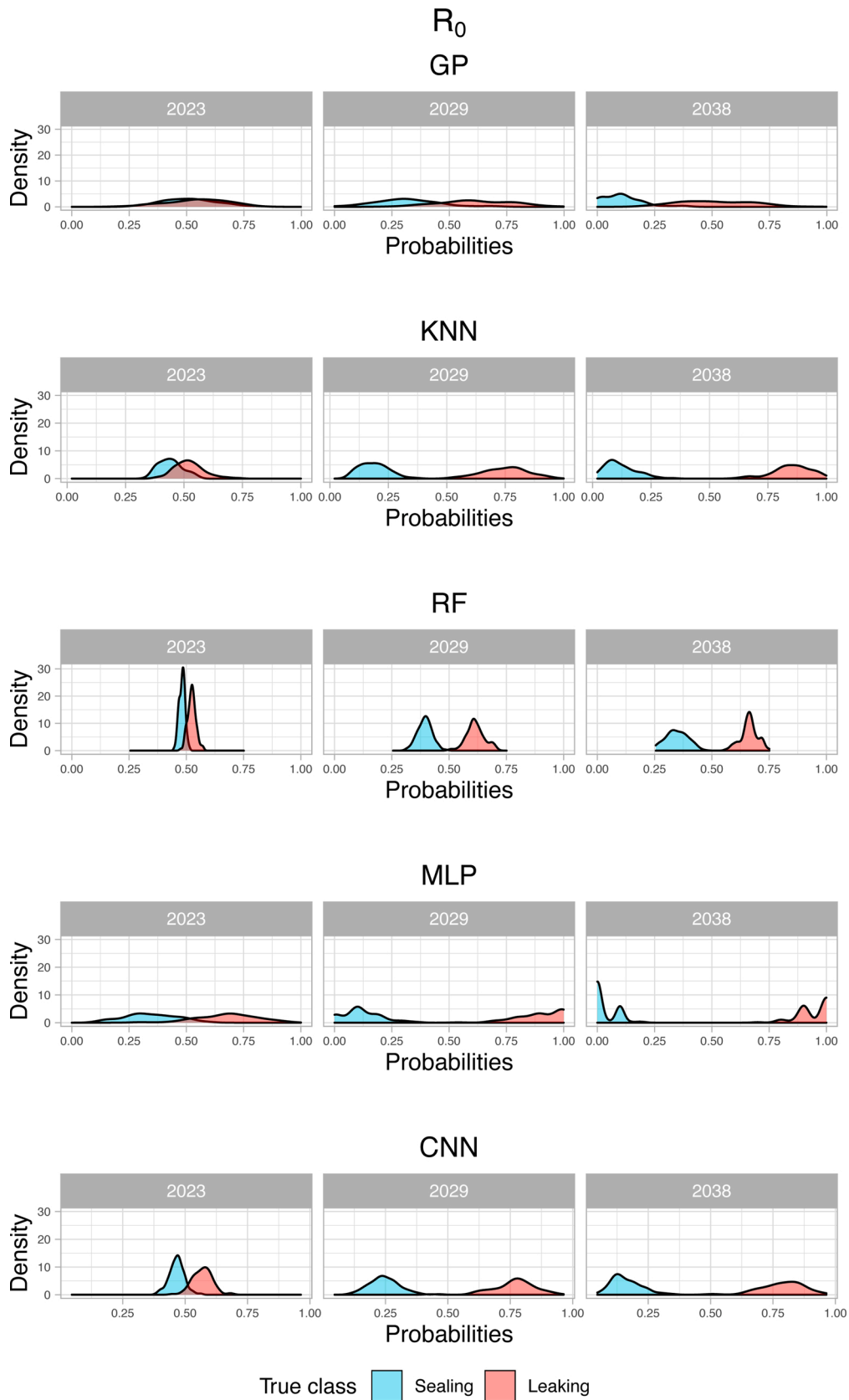
Fig. 8. Kernel densities (normalized histogram) plots of the leakage probabilities for the test set with R_0 data alone. This test set is grouped in the realizations for seal (blue) and leak (red).
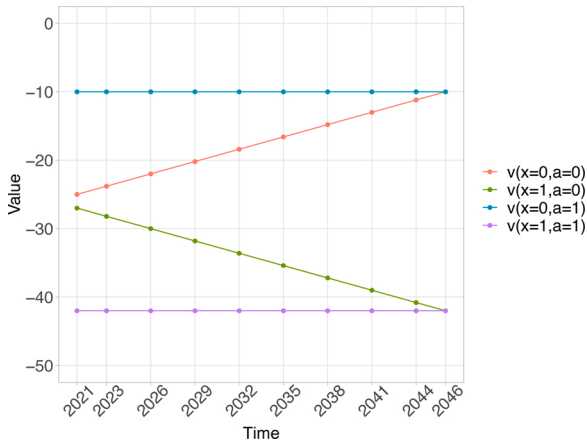
Fig. 9. Value function $v_t(x,a)$ for each of the alternatives considering leak and seal scenarios.

value calculations on the same basis and ensure non-negative VOI. This means that:

$$PV_t \approx max_{a \in A}\left\{\frac{1}{B_{\text{test}}} \sum_{b=1}^{B_{\text{test}}} E[v_t(x,a)|\mathbf{y}_t^b]\right\}, \qquad (3)$$
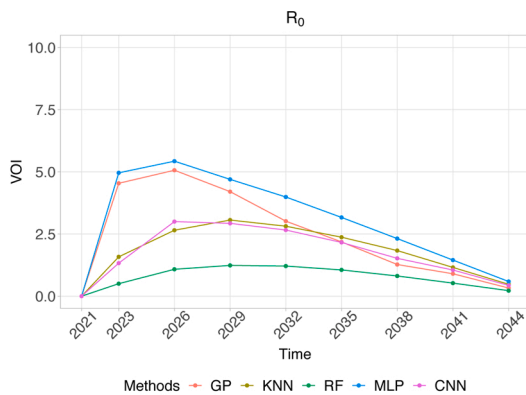
and the VOI approximation becomes:

$$
\begin{aligned}
VOI_t \approx \;& \frac{1}{B_{\text{test}}} \sum_{b=1}^{B_{\text{test}}} max_{a \in A}\left\{\sum_{x=0}^{1} v_t(x,a)\widehat{P}(X=x|\mathbf{y}_t^b)\right\} \\
& - max_{a \in A}\left\{\frac{1}{B_{\text{test}}} \sum_{b=1}^{B_{\text{test}}} \sum_{x=0}^{1} v_t(x,a)\widehat{P}(X=x|\mathbf{y}_t^b)\right\}
\end{aligned}
\qquad (4)
$$

This approximation tends to give robust approximation of the VOI (Eidsvik et al., 2017).

5.3. Results

Fig. 10 shows the results of the VOI analysis. This is shown for the various statistical methods presented in Section 4 for approximating the conditional probabilities. It is also shown for the cases with only the zero-offset seismic AVO attribute (left display) and with both zero-offset and gradient attributes (right display).

For all plots the VOI first increases from year 2021 to reach a peak value and then declines at the later years considered. The optimal monitoring time with respect to this VOI calculation appears to be

around 2026–2029 for most models. When both seismic AVO attributes are used the VOI is a little bit larger, with the exception of the GP results which potentially stem from having too much variability in the empirical parameter specification.

Just like with the accuracy values of the various models, we notice some difference in the VOI results between the statistical models. The MLP model gives the largest VOI and with a peak value appearing early in the time period. With the limited training data there might be some overfitting of the complex MLP model, leading to slightly larger VOI. The RF model gets the lowest VOI, possibly because it combines a lot of tree models and gets more of an averaging effect with reasonable classification ability but probabilities relatively near 0.5. The GP model gets the most variable results, which is seen by the plot fluctuating more than the others. With the limited training data, there is possibly not enough data to avoid Monte Carlo variability in the mean and covariance estimates of the GP model. The KNN and CNN approaches give very similar and smooth results.

We interpret the VOI results more closely by relating the expected values of each alternative to the probabilities of leakage plotted in Fig. 8. These expected values of the decision problem can be written as $v_t(0,a) + [v_t(1,a) - v_t(0,a)]\widehat{P}(X=1|\mathbf{y}_t^b)$. In Fig. 11, they are plotted as a function of the probability of leakage. For the alternative of continued injection the end points are -10 and -42. This is clearly the best alternative for leakage probabilities below 0.5. For the stop alternatives, which is the best alternative for higher leakage probabilities, the expected values are
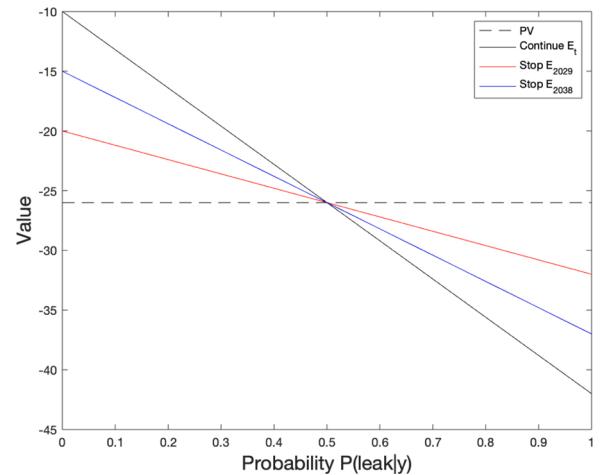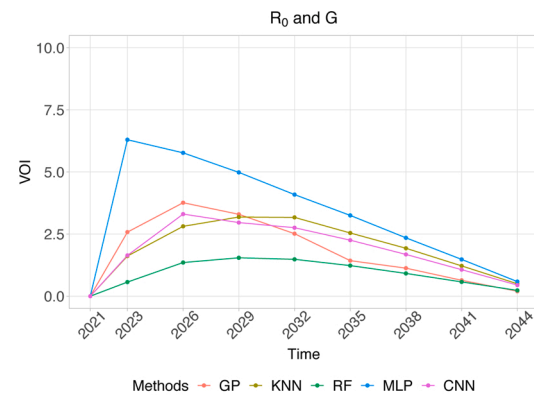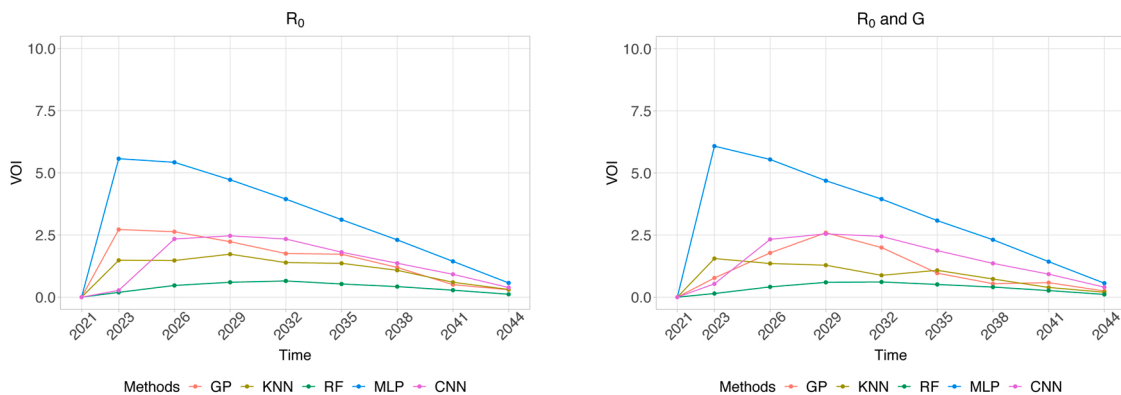

Fig. 11. Plot of the posterior expected values as a function of the leakage probabilities.

(a) VOI of zero-offset seismic AVO attribute.

(b) VOI of zero-offset and AVO gradient.

Fig. 10. VOI results for several statistical learning methods.

(a) VOI of zero-offset seismic AVO attribute.

(b) VOI of zero-offset and AVO gradient.

Fig. 12. VOI results for larger measurement noise (smaller signal to noise) in the seismic AVO data.

closer to -26, especially so for earlier years.

The largest VOI is achieved if the leakage probabilities are pulled close to the boundaries. If this happens at a later time, the gain over the continued injection is smaller (difference blue and black curve). So the VOI is a balance between the probability of leakage detection from the data and the difference in the value functions of the alternatives.

In our example the values are not directly related to real costs, so the results cannot be directly translated to an assessment of the price of the seismic testing at Smeaheia. Nevertheless, we believe that this could be done in a similar manner with real-world monetary inputs, and the VOI plot as a function of time can help decision makers select the optimal monitoring time.

6. Discussion

There are several assumptions in the model specification that could be chosen in a more nuanced manner. For instance, the leak or seal inputs are simply set to be fully leaking/sealing. A different approach would be to specify some kind of partial leakage near the fault, but this requires more knowledge on how this actually occurs. Another assumption is related to the size of the grid block for reservoir simulations. The grid sizes are rather large in our model, and smaller blocks (finer resolution) would likely lead to a higher level of detail in the solver for the differential equations and possibly better separation between the open/closed realizations and potentially a better classification process. The porosity and permeability variables are assigned

reasonable uncertainties and correlations. But one could of course imagine sensitivity to compartmentalization or three dimensional connections of volumes that makes this more complex than what we studied here. This would again require more detailed choices in the model formulation, which must be supported by available data or geological inputs, otherwise it could lead to bias effects in the modeling.

The precision of the seismic AVO data is critical for this monitoring strategy. We now study sensitivities to the variance levels of R_0 and G. In Fig. 12 we show results of increased level of variance in the seismic response. Here, we set the variance of the noise ten times larger than what used in Section 5. Fig. 13 shows the effects of decreased level of variance in the seismic data on the estimated VOI values. In this case, we set the variance of the noise ten times smaller than that considered in Section 5.

The VOI plots show tendencies similar to the reference case, but with less noise the peaks of the curves seem to appear a bit earlier in time. This indicates that with a lower signal to noise it is better to postpone the monitoring efforts, while for very accurate seismic AVO data the monitoring should take place earlier. With much noise, the VOI results are clearly a bit smaller, and they seem more variable. But the reduction in VOI is maybe smaller than we would expect with ten times increase/decrease, possibly because most of the uncertainty stems from the variability in the reservoir variables in this case. The MLP results appear to be the least sensitive to the variance level, possibly indicating overfitting to the modest training data.

Even though the case study aids in the testing of the proposed
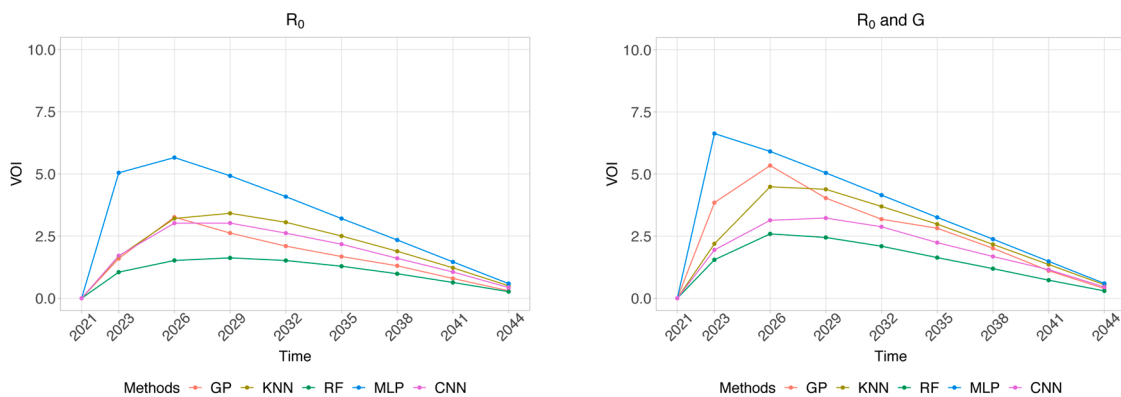
(a) VOI of zero-offset seismic AVO attribute.

(b) VOI of zero-offset and AVO gradient

Fig. 13. VOI results for smaller measurement noise (greater signal to noise) in the seismic AVO data.

workflow, the authors acknowledge that the inputs are case-specific and set by us in this study. Besides the discussed specification of reservoir model and data parameters, we set the injection rate of 8 million metric tons as an arbitrary and fixed amount which cannot be increased even if the monitoring indicates no leaking. Increasing (or decreasing) the amount of the injection would be part of a more complex decision problem that can be present in real-life applications.

7. Closing remarks

We have proposed a workflow for monitoring, framed to the decision situation of whether CO_2 injection should be stopped. The workflow includes the simulation of CO_2 saturations and AVO attributes, and the statistical modeling that enables computation of the conditional probabilities that are used to approximate the VOI. The suggested framework can provide valuable input to the storage operator. For the decision problem considered in this paper, the outcome of the VOI analysis can help choose the best monitoring time.

We only studied the influence of seismic AVO data on the CO_2 prediction. In a similar vein, using other geophysical forward models, one could analyze the value of electromagnetic data (Eidsvik et al., 2015; Furre et al., 2020), fiber-optic DAS systems (Jreij et al., 2020), or combinations of geophysical monitoring techniques. The workflow presented here can be done in a similar manner on a refined model when more information is available about aquifer variables. For all these possible data types, it would further be interesting to assess the value of time-lapse information (Dutta et al., 2019b), rather than just the setting with an optimal one-time planned monitoring.

As future work we further plan to consider a more flexible formulation of the boundary conditions. In this paper the boundary class is set to the reservoir as a whole, in a more complex scenario it can be presented as open or closed across different regions of the reservoir. Furthermore, we plan to consider a more realistic reservoir model with a more detailed top-aquifer topography which can play an important role since the current model is very coarse. The use of more realistic reservoir models needs to be analyzed against the computational burden to generate enough realizations.

Author contributions

Susan Anyosa: Methodology; software; validation; formal analysis, investigation; data curation; writing – original draft; visualization. Scott Bunting: Methodology; software; investigation, data curation; writing – original draft; visualization. Jo Eidsvik: Conceptualization; methodology; investigation; writing – review and editing; visualization; supervision; project administration; funding acquisition. Anouar Romdhane: Resources; methodology; writing – review and editing; visualization. Per Bergmo: Resources; methodology; writing – review and editing; visualization.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

References

Avseth, P., Mukerji, T., Mavko, G., 2010. Quantitative Seismic Interpretation: Applying Rock Physics Tools to Reduce Interpretation Risk. Cambridge University Press.

Bachrach, R., 2006. Joint estimation of porosity and saturation using stochastic rock-physics modeling. Geophysics 71 (5), O53–O63.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2014. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC.

Barros, E., Van den Hof, P., Jansen, J., 2016. Value of information in closed-loop reservoir management. Comput. Geosci. 20 (3), 737–749.

Bhattacharjya, D., Eidsvik, J., Mukerji, T., 2010. The value of information in spatial decision making. Math. Geosci. 42 (2), 141–163.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Dean, M., Tucker, O., 2017. A risk-based framework for Measurement, Monitoring and Verification (MMV) of the Goldeneye storage complex for the Peterhead CCS project, UK. Int. J. Greenh. Gas Control 61, 1–15.

Dupuy, B., Torres, V., Ghaderi, A., Querendez, E., Mezyk, M., 2017. Constrained AVO for CO_2 storage monitoring at Sleipner. Energy Procedia 114, 3927–3936.

Dupuy, B., Querendez, E., Ghaderi, A., Romdhane, A., Eliasson, P., 2018. Norwegian large-scale CO_2 storage project (Smeaheia): baseline geophysical models. In: 14th Greenhouse Gas Control Technologies Conference. Melbourne, pp. 21–26.

Dutta, G., Mukerji, T., Eidsvik, J., 2019a. Value of information analysis for subsurface energy resources applications. Appl. Energy 252, 113436.

Dutta, G., Mukerji, T., Eidsvik, J., 2019b. Value of information of time-lapse seismic data by simulation-regression: comparison with double-loop Monte Carlo. Comput. Geosci. 23 (5), 1049–1064.

Eidsvik, J., Mukerji, T., Bhattacharjya, D., 2015. Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis. Cambridge University Press.

Eidsvik, J., Dutta, G., Mukerji, T., Bhattacharjya, D., 2017. Simulation-regression approximations for value of information analysis of geophysical data. Math. Geosci. 49 (4), 467–491.

Eiken, O., Ringrose, P., Hermanrud, C., Nazarian, B., Torp, T.A., Høier, L., 2011. Lessons learned from 14 years of CCS operations: Sleipner, In Salah and Snøhvit. Energy Procedia 4, 5541–5548.

Furre, A.-K., Eiken, O., Alnes, H., Vevatne, J.N., Kiær, A.F., 2017. 20 years of monitoring CO_2-injection at Sleipner. Energy Procedia 114, 3916–3926.

Furre, A.-K., Meneguolo, R., Pinturier, L., Bakke, K., 2020. Planning deep subsurface CO_2 storage monitoring for the Norwegian full-scale CCS project. First Break 38 (10), 55–60.

Harbert, W., Daley, T.M., Bromhal, G., Sullivan, C., Huang, L., 2016. Progress in monitoring strategies for risk reduction in geologic CO_2 storage. Int. J. Greenh. Gas Control 51, 260–275.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.

Hinton, G.E., 1990. Connectionist learning procedures. Machine Learning. Elsevier, pp. 555–610.

Jenkins, C., Chadwick, A., Hovorka, S.D., 2015. The state of the art in monitoring and verification – ten years on. Int. J. Greenh. Gas Control 40, 312–349.

Jreij, S.F., Trainor-Guitton, W.J., Morphew, M., Chen Ning, I.L., 2020. The value of information from horizontal distributed acoustic sensing compared to multicomponent geophones via machine learning. J. Energy Resour. Technol. 143 (1).

Lie, K.-A., Krogstad, S., Ligaarden, I.S., Natvig, J.R., Nilsen, H.M., Skaflestad, B., 2012. Open-source MATLAB implementation of consistent discretisations on complex grids. Comput. Geosci. 16 (2), 297–322.

Nordbotten, J.M., Celia, M.A., 2011. Geological Storage of CO_2: Modeling Approaches for Large-Scale Simulation. John Wiley & Sons.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Int. J. Mach. Learn. Technol. 2, 37–63.

Rasmussen, C.E., Williams, C.K., 2006. Gaussian Processes for Machine Learning, vol. 38. The MIT Press, Cambridge, MA, USA, pp. 715–719.

Ringrose, P., Mathieson, A., Wright, I., Selama, F., Hansen, O., Bissell, R., Saoula, N., Midgley, J., 2013. The In Salah CO_2 storage project: lessons learned and knowledge transfer. Energy Procedia 37, 6226–6236.

Ringrose, P., Thorsen, R., Zweigel, P., Nazarian, B., Furre, A.-K., Paasch, B., Thompson, N., Karstad, P., 2017. Ranking and risking alternative CO_2 storage sites offshore Norway. Fourth Sustainable Earth Sciences Conference.

Sato, K., 2011. Value of information analysis for adequate monitoring of carbon dioxide storage in geological reservoirs under uncertainty. Int. J. Greenh. Gas Control 5 (5), 1294–1302.

Trainor-Guitton, W.J., Ramirez, A., Yang, X., Mansoor, K., Sun, Y., Carroll, S., 2013. Value of information methodology for assessing the ability of electrical resistivity to detect CO_2/brine leakage into a shallow aquifer. Int. J. Greenh. Gas Control 18, 101–113.

Yang, Y.-M., Dilmore, R.M., Bromhal, G.S., Small, M.J., 2018. Toward an adaptive monitoring design for leakage risk-closing the loop of monitoring and modeling. Int. J. Greenh. Gas Control 76, 125–141.