Value of information of time-lapse seismic data by simulation-regression: comparison with rigorous Monte Carlo

Geetartha Dutta[1], Tapan Mukerji[1] and Jo Eidsvik[2]
1) Department of Energy Resources Engineering, Stanford University, USA
2) Department of Mathematical Sciences, NTNU, Norway

Corresponding author: Geetartha Dutta (gdutta@stanford.edu, 1-6508043890),
Stanford University, California, USA

Abstract

Simulation-regression is a computationally efficient methodology to estimate the value of information (VOI), as it involves directly estimating the value outcomes corresponding to different data realizations by building a statistical relationship between the prospect values and the data, rather than estimating the model parameters from the data and then estimating the value outcomes given the model parameters. The simulation-regression workflow is applied to estimate the VOI of time-lapse seismic data in a 2D reservoir case using Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR), and the variance in the VOI result is estimated using bootstrapping for varying number of realizations. The VOI results from the two regression techniques are found to be consistent, and it is seen from the bootstrap results that the variance in the VOI decreases with increasing number of realizations and the VOI ranges obtained by higher number of realizations are captured by those obtained by fewer realizations. The VOI results from simulation-regression are then compared with those obtained by a rigorous Monte Carlo method, where the posterior model realizations are sampled using rejection sampling for each possible data realization, and then the prospect values are estimated for each model realization using flow simulation. Finally, the simulation-regression method is applied to estimate the VOI of time-lapse seismic data in a complex production optimization case involving sequential well placement and control decisions.

Keywords: value of information, reservoir development, time-lapse seismic data, simulation-regression

1 Introduction

Time-lapse seismic data is often acquired during reservoir development to monitor the fluid saturation changes as a result of production, with the goal of making better decisions towards the optimal development of the reservoir. Since time-lapse seismic data comes at a high cost, it is important to justify its cost by assessing the impact it could have on the decisions needed to be made. The value of information (VOI) is a decision-analytic metric that is suited for this purpose (Howard, 1966). The VOI is an estimate of the additional value that information brings to a decision situation, which can be compared with its cost to decide whether to collect the information or not.

From a decision-analytic perspective, information is valuable not only if it reduces uncertainty, but only if the reduction in uncertainty helps in making better decisions and maximizing the value outcome. Therefore, a VOI analysis can only be performed in the context of a particular decision situation. But since the VOI analysis is done a-priori, i.e. before actually collecting the information, the uncertainty in the information outcome also needs to be taken into account. Time-lapse seismic data can reduce the uncertainty in reservoir properties like porosity, permeability and saturation, and thus help in making better reservoir development decisions such as well placement and control decisions.

The conventional way of data assimilation in reservoir development is to build reservoir models conditioned to the data, which are then used to predict the future production. Extracting useful information from time-lapse seismic data using a model-driven approach is computationally expensive because the models are usually high-dimensional. This is especially true for VOI analysis as a set of models has to be built for every possible dataset, rendering the VOI analysis computationally intractable for complex problems. Therefore, there is a need for a data-driven approach (Satija et al., 2017) in VOI analysis for spatial problems, where the models are used to build a statistical relationship between the data and the forecast variables, and then the statistical relationship is used to predict the forecast. The main rationale behind this is that though the models are high dimensional and computationally expensive to build, they are highly uncertain because of the imperfect data, and even so, might not accurately assess the uncertainty of the forecast. On the contrary, the forecast variables that we are really interested in predicting, like the future production or the value outcomes, are usually low-dimensional variables and hence are easier to sample given the data. The VOI is even lower dimensional than the value outcomes, as it is obtained by taking the expectation of the value outcomes over the spatial uncertainty.

Most previous work on VOI analysis of geophysical data have employed a model-driven approach that involves explicit approximation of the posterior distribution of reservoir properties given the data. A characteristic of VOI analysis of geophysical data is the spatial nature of the problem – the decision alternatives, the uncertain reservoir properties and the data proposed to be collected are all spatial variables (Eidsvik et al., 2016). Houck (2007) used a 1D model for VOI analysis of time-lapse seismic data and hence did not address the spatial nature of the problem. Eidsvik et al. (2008) used 2D models for VOI analysis of seismic amplitude and CSEM resistivity data. However, the spatial dependence was modeled using simple Gaussian models which might not be very geologically realistic. Trainor-Guitton et al. (2013) proposed a VOI methodology for spatial problems incorporating complex geological spatial models simulated using multi-point geostatistics. To compute the VOI, they quantified the reliability of the information or the likelihood by forward modeling the geophysical attribute using the prior models. In their workflow, they used the data to inform about the geological scenario rather than the spatial heterogeneity of reservoir properties. Barros et al. (2015) proposed a VOI methodology for reservoir decisions which takes into account the stochastic variability in an ensemble of realizations. Their method relies on "twin-experiments", which refers to repeated application of reservoir life-cycle optimization and data assimilation. This method is computationally very expensive as it requires building a posterior ensemble of realizations for every possible dataset. Goda et al. (2018) used multilevel quasi-Monte Carlo methods to estimate the Expected Value of Partial Perfect Information (EVPPI), but their model too is parameterized using a few parameters, and thus might not capture the complex spatial heterogeneity of petroleum reservoirs.

In this paper, a data-driven approach known as simulation-regression (Eidsvik et al., 2017) is applied to evaluate the VOI of time-lapse seismic data in a 2D reservoir case, and the results compared with those obtained using a model-driven approach involving rigorous Monte Carlo simulation (Barros et al., 2015; Hong et al., 2018). The simulation-regression methodology has been used in the field of medicine by Strong et al. (2014) and Heath et al. (2016). However, these applications do not have the spatial characteristics that are typical of VOI problems in the earth sciences. The simulation-regression methodology for computing the VOI is similar to the direct forecasting methodology of Satija et al. (2017) but we attempt to predict the Net Present Value (NPV) rather than the production profiles, and our data is spatial rather than temporal. Various regression techniques are employed to perform the regression of the NPVs on the data.

This paper is organized as follows. A brief review of concepts and equations for VOI computation in spatial earth science problems is provided in Section 2. Section 3 provides a review of the methodologies that are used in this paper – simulation-regression and rigorous Monte Carlo – in the context of VOI estimation of time-lapse seismic data. Two regression techniques – Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) – that are appropriate for this purpose are also discussed. Section 4 presents the comparison of the results of VOI analysis of time-lapse seismic data in a 2D reservoir case using the two methodologies discussed in Section 3. In Section 5, simulation-regression is employed to perform VOI analysis in a complex production optimization case involving sequential well placement and control decisions. Finally, in Section 6, some concluding remarks and directions for future research are provided.

2 VOI in spatial decision situations

The VOI in any decision situation depends on three main factors: the prior uncertainty regarding the variables that affect the value outcomes of the decision, the decision situation comprising alternatives and prospect values, and the information reliability. The VOI is usually high if the prior uncertainty is high, because then there is more potential for the information to reduce the uncertainty. Also, the VOI is usually high if the number of decision alternatives is large, because then there is greater scope to tune our decision based on the information outcome. The VOI is also affected by how different the value outcomes corresponding to different alternatives are; if the value outcomes are very different, the information can potentially generate high value by enabling us to select the optimum alternative, and hence the VOI is high. Lastly, it is intuitive that the VOI is high if the information reliability is high.

In the case of VOI of time-lapse seismic data, the variables that affect the prospect values are the reservoir properties, and hence the prior uncertainty of the reservoir properties need to be taken into account. Reservoir properties such as facies, porosity, permeability, saturation, etc. affect the production and thereby the values. We denote these uncertain spatial reservoir properties by $\boldsymbol{x}$. Since $\boldsymbol{x}$ is very high dimensional, it is difficult to represent the distribution of $\boldsymbol{x}$ in analytical form, except for the special case of a multivariate Gaussian distribution (Eidsvik et al., 2016). Therefore, we usually approximate the distribution of $\boldsymbol{x}$ by Monte Carlo sampling, thus representing the prior distribution of $\boldsymbol{x}$ as an ensemble of realizations $\boldsymbol{x}^1$, $\boldsymbol{x}^2$,…, $\boldsymbol{x}^B$. The decision alternatives are denoted by $\boldsymbol{a} = \{a_i : i = 1,2, \dots, n\}$, where index i is associated with spatial location. The action or decision alternative $\boldsymbol{a}$ must be chosen from a set A of all possible alternatives, i.e., $\boldsymbol{a} \in A$. The values are functions of the particular realization and the decision alternative chosen, and are denoted by $v(\boldsymbol{x}, \boldsymbol{a})$.

Assuming a risk neutral decision maker, the prior value (PV) is defined as the maximum expected value over the uncertain reservoir properties, the maximization performed over the possible alternatives:

$$PV = max_{\boldsymbol{a} \in A} \left[\int v(\boldsymbol{x}, \boldsymbol{a}) p(\boldsymbol{x}) d\boldsymbol{x} \right] \approx max_{\boldsymbol{a} \in A} \left[\frac{1}{B} \sum_{b=1}^{B} v(\boldsymbol{x}^b, \boldsymbol{a}) \right]. \quad (1)$$

where $p(\boldsymbol{x})$ is the prior probability distribution over $\boldsymbol{x}$.

If we have perfect information about what value the variable $\boldsymbol{x}$ would take, we would choose the optimal action for that value of $\boldsymbol{x}$. However, since the VOI calculation is done before actually collecting the data, the posterior value (PoV) with perfect information is computed by taking the expectation over all possible values of $\boldsymbol{x}$. Therefore, we have:

$$PoV(\boldsymbol{x}) = \int max_{\boldsymbol{a} \in A} [v(\boldsymbol{x}, \boldsymbol{a})] p(\boldsymbol{x}) d\boldsymbol{x} \approx \frac{1}{B} \sum_{b=1}^{B} max_{\boldsymbol{a} \in A} [v(\boldsymbol{x}^b, \boldsymbol{a})]. \quad (2)$$

Now, let us assume that we collect time-lapse seismic data $\boldsymbol{y}$ which is tied indirectly to the reservoir properties by rock physics relations, such that $\boldsymbol{y} = f(\boldsymbol{x})$. In addition, the data $\boldsymbol{y}$ is at a lower resolution than the reservoir properties $\boldsymbol{x}$, and hence we cannot determine for certain the reservoir properties from the data, even if there is no noise in the data. For each realization $\boldsymbol{x}^b$, we can forward model the time-lapse seismic data $\boldsymbol{y}^b = f(\boldsymbol{x}^b)$. $\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^B$ represent the distribution of the data.

Then, the posterior value (PoV) with imperfect information is defined as

$$PoV(\boldsymbol{y}) = \int max_{\boldsymbol{a} \in A} [E(v(\boldsymbol{x}, \boldsymbol{a}) | \boldsymbol{y})] p(\boldsymbol{y}) d\boldsymbol{y} \approx \frac{1}{B} \sum_{b=1}^{B} max_{\boldsymbol{a} \in A} E[v(\boldsymbol{x}, \boldsymbol{a}) | \boldsymbol{y}^b]. \quad (3)$$

For a risk neutral decision maker, the VOI is given by the difference between the posterior value and the prior value (Bratvold et al., 2009):

$$VOI(\boldsymbol{y}) = PoV(\boldsymbol{y}) - PV. \quad (4)$$

The posterior value with imperfect information (Equation 3) can be estimated using rigorous Monte Carlo by first approximating the posterior distribution of $\boldsymbol{x}$ given each $\boldsymbol{y}^b$ and then computing the corresponding value outcomes. On the other hand, simulation-regression directly estimates the conditional expectation of values given the data. These methodologies are discussed in the next section.

3 VOI methodologies

3.1 Simulation-Regression

In the simulation-regression methodology, the VOI is computed by simulating the model parameters, the data and the prospect values, and then regressing the values on the data. The steps involved in this methodology are as follows:

a) Draw Monte Carlo samples of the model parameters ($\boldsymbol{x}^b$), like facies, porosity, permeability, etc., and generate the corresponding samples of data ($\boldsymbol{y}^b = f(\boldsymbol{x}^b)$) and prospect values ($v_{\boldsymbol{a}}^b = v(\boldsymbol{x}^b, \boldsymbol{a})$) for each alternative $\boldsymbol{a}$.

b) Regress the vector, $\boldsymbol{v_a}$, containing the samples of values for each alternative $\boldsymbol{a}$ on the data matrix, $\boldsymbol{Y}$, the rows of which correspond to different observations and the columns to different data dimensions, to obtain the regression model $F_{\boldsymbol{a}}(\boldsymbol{Y})$.

c) Fit values $\hat{v}_{\boldsymbol{a}}^b$ using the regression model:

$$\hat{v}_{\boldsymbol{a}}^b = F_{\boldsymbol{a}}(\boldsymbol{y}^b). \tag{5}$$

$\hat{v}_{\boldsymbol{a}}^b$ approximates the conditional expectation $E[v(\boldsymbol{x}, \boldsymbol{a})|\boldsymbol{y}^b]$.

d) The prior value is given by:

$$PV = max_{\boldsymbol{a} \in \boldsymbol{A}} \left[\frac{1}{B} \sum_{b=1}^{B} v_{\boldsymbol{a}}^b\right]. \tag{6}$$

e) The posterior value with imperfect information is given by:

$$PoV(\boldsymbol{Y}) = \frac{1}{B} \sum_{b=1}^{B} max_{\boldsymbol{a} \in \boldsymbol{A}}\, E[v(\boldsymbol{x}, \boldsymbol{a})|\boldsymbol{y}^b] \approx \frac{1}{B} \sum_{b=1}^{B} max_{\boldsymbol{a} \in \boldsymbol{A}}\, \hat{v}_{\boldsymbol{a}}^b. \tag{7}$$

f) Then, the VOI is given by:

$$VOI = PoV(\boldsymbol{Y}) - PV.$$

Various regression methods can be used to regress the values on the data. Simple regression methods like linear regression, k nearest neighbors and cubic smoothing splines work well for low dimensional problems. But these methods fail when the data is very high dimensional, especially when the number of dimensions in the data is much larger than the number of observations or realizations, as is usually the case for time-lapse seismic data. For these high dimensional problems, we need to use regression techniques like Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR). These techniques are reviewed below, and will be used in the VOI cases of Section 4 and Section 5.

3.1.1 Value regression by PLSR

Partial Least Squares (PLS) seeks to project two sets of variables to orthogonal latent variables by maximizing the covariance between them. PLS decomposes $\boldsymbol{Y}$ and $\boldsymbol{v_a}$ into the form (Rosipal & Kramer, 2006):

$$\begin{aligned} \boldsymbol{Y} &= TP^T + E \\ \boldsymbol{v_a} &= UQ^T + F \end{aligned} \tag{8}$$

where $\boldsymbol{Y}$ is the $B \times N$ data matrix with B observations of the N dimensional data variable, $\boldsymbol{v_a}$ is the $B \times 1$ vector of prospect values corresponding to alternative $\boldsymbol{a}$, T and U are $B \times M$ matrices containing the projections of $\boldsymbol{Y}$ (the $\boldsymbol{Y}$ scores) and the projections of $\boldsymbol{v_a}$ (the $\boldsymbol{v_a}$ scores) respectively, P and Q are matrices of loadings of size $N \times M$ and $1 \times M$ respectively, and E and F are matrices of residuals of size $B \times N$ and $B \times 1$ respectively. This decomposition is made such that the covariance between T and U is maximized. The method finds weight vectors w and c such that

$$[cov(t,u)]^2 = [cov(\boldsymbol{Y}w, \boldsymbol{v_a}c)]^2 = \max_{|r|=|s|=1}[cov(\boldsymbol{Y}r, \boldsymbol{v_a}s)]^2. \quad (9)$$

where $cov(t,u) = t^T u/n$ denotes the sample covariance between the score vectors t and u.

To use PLS as a regression method, two assumptions are made:
i) the score vectors $\{t_i\}_{i=1}^{M}$ are good predictors of $\boldsymbol{Y}$
ii) the score vectors t and u are linearly related as

$$U = TD + H. \quad (10)$$

where D is an $M \times M$ diagonal matrix and H is the matrix of residuals.

Substituting (10) in (8) we have

$$\boldsymbol{v_a} = TDQ^T + (HQ^T + F)$$

$$\text{Or, } \boldsymbol{v_a} = TC^T + F^*. \quad (11)$$

where $C^T = DQ^T$ is an $M \times 1$ matrix of regression coefficients and $F^* = HQ^T + F$ is the residual matrix.

Equation (11) represents the decomposition of $\boldsymbol{v_a}$ using ordinary least squares regression with orthogonal latent predictor variables T.

The quality of the PLS regression model can be evaluated by k-fold cross-validation, where the entire dataset is divided into k folds, a PLS regression model is trained on the entire dataset except each of the folds in turn and is used to fit the values in the fold left out. This process is repeated for all the folds. Thus, we obtain a vector of predicted values for each alternative, $\widetilde{\boldsymbol{v}}_{\boldsymbol{a}}$. The quality of the regression is then given by the Predicted Residual Sum of Squares (PRESS), which is computed as (Abdi, 2010):

$$PRESS = \|\boldsymbol{v_a} - \widetilde{\boldsymbol{v}}_{\boldsymbol{a}}\|^2. \quad (12)$$

The lower the value of PRESS, the more accurate the prediction is. Using the PRESS metric, the optimum number of latent vectors M can be determined. In most cases, the value of PRESS first decreases with increasing M, and then increases. This is because of the bias-variance tradeoff; at low values of M, the bias is high but the variance is low because of underfitting, while at high values of M, the bias is low but the variance is high due to overfitting. Thus, we usually obtain the minimum test error, which is proportional to the sum of variance and squared bias, for intermediate values of M. The value of M for which PRESS is minimum gives the optimum number of latent vectors. After the number of latent vectors to be used in the regression model has been determined, confidence intervals for the predicted values can be found by bootstrapping.

3.1.2 Value regression by PCR

In PCR, a Principal Component Analysis (PCA) is performed on the data matrix $\boldsymbol{Y}$, and then the PCA components are used as predictor variables in a linear regression model to predict the prospect values $\boldsymbol{v_a}$.

The PCA transformation of $\boldsymbol{Y}$ can be represented using the singular value decomposition (SVD) of $\boldsymbol{Y}$ as (Abdi, 2010):

$$\boldsymbol{Y} = R\Delta V^T. \tag{13}$$

with $R^T R = V^T V = I$, where R and V are the matrices of the left and right singular vectors respectively, and Δ is a diagonal matrix containing the singular values. The singular vectors are ordered according to their corresponding singular value which is the square root of the variance of $\boldsymbol{Y}$ explained by the singular vector. The columns of V are called the loadings, while the columns of $G = R\Delta$ are called the scores or principal components (PCs) of $\boldsymbol{Y}$. The PCs thus obtained are then used as regressors in a linear regression model to predict the values $\boldsymbol{v_a}$.

$$\boldsymbol{v_a} = GC^T + E. \tag{14}$$

where C^T = regression coefficient, and E = residual.

Typically, only a subset of the PCs which captures enough variance in the original predictor variables in $\boldsymbol{Y}$ are used in the regression model, resulting in a much reduced dimensional regression. To find the optimum number of PCs to use, we can employ the same cross-validation technique as discussed for PLSR. Also, as in the case of PLSR, the uncertainty in the predicted values obtained by PCR can be estimated using bootstrapping.

While both PLSR and PCR perform the value regression in a reduced dimensional space, there is some difference in the way the two methods project the data into components. PCR just takes into account components of the data that explain maximum variance in the predictor variables $\boldsymbol{Y}$, while PLSR seeks to find components along which there is maximum correlation between the predictor $\boldsymbol{Y}$ and the response $\boldsymbol{v_a}$, while also trying to capture maximum variance in $\boldsymbol{Y}$ and $\boldsymbol{v_a}$ (Hastie et al., 2009).

3.2 Rigorous Monte Carlo

In this paper, a rigorous Monte Carlo methodology proposed by Barros et al. (2015) with the modifications suggested by Hong et al. (2018) is employed to compute the VOI and compare it with the simulation-regression results. The steps involved in this rigorous Monte Carlo methodology are as follows:

a) Generate a prior ensemble of realizations conditioned to the prior information (in our case, the base seismic data, the hard well data and the production data).
b) Run flow simulation on the prior ensemble of realizations for all possible alternatives, and determine the alternative which results in maximum expected value over all the realizations. The maximum expected value is the prior value and the corresponding alternative is the optimal alternative given the prior information.
c) Generate synthetic time-lapse seismic data corresponding to each prior realization by forward modeling using a rock physics model.
d) Sample the posterior distribution conditioned to each time-lapse seismic dataset. In this paper, rejection sampling is used.
e) Run flow simulation on the posterior ensemble for the remaining time for all possible alternatives and determine the optimal alternative for the corresponding prior realization. Repeat this for every prior realization.
f) Run flow simulation on all the prior realizations using the optimal alternatives determined in step (e) and compute the prospect values given by the NPVs.

g) The posterior value with information is given by the mean of the NPVs computed in step (f) and the VOI for a risk-neutral decision maker is given by the difference between the posterior value and the prior value.

4 VOI case to compare methodologies

In this section, the VOI of time-lapse seismic data in a 2D reservoir case is computed using the two methodologies discussed in Section 3 – simulation-regression and rigorous Monte Carlo – and their results compared.

4.1 Problem formulation

Consider a reservoir, modeled using a grid of 120×220 cells, that has been under production for two years, with one producer at the center and four injectors at the corners, forming a typical five-spot pattern as shown in Figure 1. The injectors were drilled one after another at intervals of six months, and the order of drilling is indicated in the figure using numbers. The geological scenario of the reservoir has been interpreted as a channel scenario. In addition to the spatial uncertainty of the reservoir properties, there is uncertainty regarding the presence or absence of a flow barrier.

After two years of production, the decision of whether or not to drill another producer well at the location marked as "4" in Figure 1 has to be made. This infill well could help to drain the bypassed oil if there is a flow barrier. The objective here is to maximize the NPV after three more years of production. Before making this well placement decision, it might be worth collecting time-lapse seismic data because it informs about the saturation change in the reservoir and thus might help to identify the bypassed oil. The question of whether or not to collect time-lapse seismic data can be answered using the VOI metric; it is advisable to collect the data only if its cost is less than the VOI.

To compute the prior value using either methodology, the prior uncertainty in reservoir properties has to modeled. An ensemble of 4000 prior realizations of facies is first simulated using the multi-point geostatistical algorithm Single Normal Equation Simulation or SNESIM (Strebelle, 2002) using a training image with two distinct facies – channel and floodplain, with 65% floodplain facies and 35% channel facies. The porosity realizations are then simulated conditioned to the facies using the two-point simulation algorithm Sequential Gaussian Simulation (SGSIM). The permeability is computed using the Kozeny-Carman equation (Mavko et al., 2009) which relates porosity and permeability. Figure 2 shows some facies, porosity and permeability realizations that are part of the prior set of realizations.

To model the time-lapse seismic data, the acoustic impedance (AI) for each realization is modeled using the constant cement model (Avseth et. al., 2000) at both the initial time before production, and after two years of production. The Gassmann equation (Gassmann, 1951) is used to account for the saturation change. A moving mean filter is applied to the AI at the geostatistical scale to approximate what can be obtained by inverting seismic data with a resolution of 60 m. The difference between the AI at the seismic scale between the two instances of time represents the time-lapse seismic signature. Figure 3 shows the AI at both the geostatistical scale and at the seismic scale for the realizations shown in Figure 2.

Now, to compute the prior value, we need to evaluate the prospect values for each decision alternative corresponding to each realization. The values are given by the NPV which is a function of the oil production, the water production, the water injection and the cost of drilling wells:

$$V^b = \int_{t=0}^{T} \frac{q_o(t, \boldsymbol{x}^b) r_o - q_{wp}(t, \boldsymbol{x}^b) r_{wp} - q_{wi}(t, \boldsymbol{x}^b) r_{wi}}{(1+r)^t} dt - cost_{drill} * a. \qquad (15)$$

where V^b = NPV for the b^{th} realization, $\boldsymbol{x}^b = b^{th}$ realization, t = time, T = producing life, q_o = oil production rate, q_{wp} = water production rate, q_{wi} = water injection rate, r_o = price of oil produced = \$50/barrel, r_{wp} = cost of water produced = \$5/barrel, r_{wi} = cost of water injected = \$5/barrel, r = discount rate = 10% per year, $cost_{drill}$ = drilling cost = \$100 million per well, a = decision alternative ($a \epsilon \{0{:}\, do\ not\ drill\ well\ 4, 1{:}\, drill\ well\ 4\}$).

So, flow simulation is run on each realization for each decision alternative, and the NPVs (after three more years of production) calculated using the oil and water production rates. Figure 4 shows the oil and water production profiles for all 4000 realizations for each decision alternative.

The mean NPV for the alternative "Do not drill" is found to be \$381.6 million, while that for the alternative "Drill" is found to be \$380.8 million. Therefore, the optimal alternative without the time-lapse seismic data is not to drill the new producer well, and the prior value is the value of that alternative, i.e., \$381.6 million.

4.2 VOI by simulation-regression

To compute the VOI of the time-lapse seismic data by simulation-regression, we need to regress the NPVs on the data for each decision alternative. Since the number of predictor variables (dimensions of the time-lapse data) is much larger than the number of realizations, simple regression techniques like linear regression do not work in this case. To perform the regression in this case we need to reduce the dimensions of the data. We employed two different regression techniques to do this – PLSR and PCR.

To determine the optimum number of PLSR and PCR components to use, five-fold cross validation was performed to evaluate the PRESS for varying number of components. From Figure 5, we see that 9 PLSR components give the minimum value of PRESS for both decision alternatives. And for PCR, 141 components give the minimum value of PRESS for both decision alternatives. So, to regress the values on the data, PLSR models were built for number of components = 9, and PCR models were built for number of components = 141. Figure 6 shows the fitted values of NPV versus the observed values of NPV for both the PLSR and the PCR models. The posterior value with the time-lapse seismic data is then evaluated using the fitted values. The posterior value from PLSR is found to be \$420.1 million and that from PCR is found to be \$420.5 million. Since the prior value, as obtained is Section 4.1, is \$381.6 million and the VOI is given by the difference between the posterior value and the prior value, the VOI from PLSR is \$38.5 million and that from PCR is \$38.9 million.

Since the number of realizations is much lower than the number of dimensions in the data, the uncertainty in the regression models is quite high, even for techniques like PLSR and PCR. Hence, we employ bootstrapping to estimate the uncertainty in the VOI computed using simulation-regression. Figure 7 shows the bootstrap estimates of uncertainty in VOI by both PLSR and PCR for varying number of realizations. It is seen that, as the number of realizations increases, the variance in the VOI estimate decreases for both PLSR and PCR, which is intuitive because the uncertainty in the VOI estimate usually decreases with increasing sample size. It is also observed that the VOI ranges obtained using higher number of realizations are mostly captured by those obtained using fewer number of realizations. This uncertainty estimation also helps to identify the number of realizations required to obtain a reliable estimate of the VOI by observing the convergence of the VOI estimate. In this case, it is seen that the median VOI estimate varies significantly for sample sizes less than 2000, but is more or less stable for sample sizes greater than or equal to 2000. So, 2000 realizations might be enough to obtain a reliable estimate of VOI in this case.

4.3 VOI by Rigorous Monte Carlo

The rigorous Monte Carlo methodology to compute the VOI involves sampling the posterior ensemble of realizations for each possible dataset. In this work, a two-step procedure is used to sample the posterior realizations:

a) The presence or absence of the flow barrier is interpreted from each time-lapse seismic dataset.
b) A posterior ensemble of realizations is sampled using rejection sampling given the information about the flow barrier.

This two-step procedure is required because this case comprises two types of uncertainty – scenario uncertainty involving the presence of absence of the flow barrier, and spatial uncertainty involving stochastic variation of spatial reservoir properties within a scenario. To interpret the scenario, a Support Vector Machine (SVM, Steinwart & Christmann, 2008) classifier with a linear kernel is used to classify each time-lapse seismic dataset into two classes – "flow barrier" and "no flow barrier" – using the first 150 principal components of the data as features. To find the optimal value of the cost parameter in the SVM classifier, a 10-fold cross-validation is performed to minimize the classification error rate. It is found that the classification error is minimized for cost = 0.1, as shown in Figure 8. So, an SVM classifer with a linear kernel and cost = 0.1 is used to classify all the time-lapse seismic datasets corresponding to the 4000 prior realizations into the two classes. The confusion matrix in Table 1 shows the classification results. The overall classification accuracy is more than 98%.

	True "no flow barrier"	True "flow barrier"
Predicted "no flow barrier"	1686	44
Predicted "flow barrier"	25	2245

Table 1: Confusion matrix showing the results of classification of the time-lapse seismic data into the two classes.

Rejection sampling is then applied to sample the posterior ensemble of realizations for each of the 4000 prior realizations, conditioned to the class predicted by the SVM classifier. Then, flow simulation is run on each posterior ensemble for each decision alternative to find the optimal alternative for the corresponding prior realization. The posterior NPVs are then obtained by running flow simulation on the prior ensemble of realizations using the optimal alternatives given by the posterior ensembles. Finally, the value with information is given by the mean of the posterior NPVs, and it is found to be $419.9 million using all the 4000 prior realizations. Thus, the VOI of the time-lapse seismic data is $38.3 million.

To compare the VOI obtained by this rigorous Monte Carlo methodology with that obtained from PLSR and PCR, the VOI is evaluated using random samples drawn from the 4000 prior realizations available, with sample sizes varying from 150 to 4000 as was considered in Section 4.2. Figure 9 shows the comparison of the VOI obtained by the two methodologies for varying number of realizations. It is seen that the VOI obtained by rigorous Monte Carlo falls within the VOI ranges obtained by PLSR and PCR when sufficient number of realizations is considered. The VOI by rigorous Monte Carlo fluctuates a lot when the number of realizations is low, as was also seen in the case of VOI by PLSR and PCR. This is because the variance in the VOI estimate is high when the number of realizations is low. As the number of realizations increases, the VOI estimate by rigorous Monte Carlo stabilizes and falls within the VOI ranges given by PLSR and PCR. As was the case with PLSR and PCR, here too it is observed that about 2000 realizations is required to get a reliable estimate of the VOI.

	Simulation-regression	Rigorous Monte Carlo
Reservoir property simulations (facies, porosity, permeability)	20,000	1,020,000
Flow simulations	48,000	2,248,000
Time-lapse seismic forward modeling	4,000	204,000

Table 2: Comparison of the computational cost of VOI estimation by simulation-regression and rigorous Monte Carlo.

Table 2 shows a comparison of the computational cost of VOI estimation by simulation-regression and by rigorous Monte Carlo. It is seen that rigorous Monte Carlo is two orders of magnitude more computationally expensive than simulation-regression. In the simulation-regression methodology, 20,000 realizations of each reservoir property were initially simulated conditioned to the base seismic data and the well log data, 4,000 of which were retained after conditioning to the prior production data, forming the prior. In the case of the rigorous Monte Carlo methodology, in addition to the 20,000 realizations generated to obtain the prior, another 1,000,000 realizations were simulated to obtain the posterior for each prior realization by rejection sampling, thus amounting to a total of 1,020,000 realizations. In the case of simulation-regression, flow simulations for the first two years of production were run on all 20,000 realizations for each scenario (presence or absence of the flow barrier) resulting in 40,000 simulations, and then for the subsequent three years of production on all 4,000 prior realizations for both decision alternatives (drill or do not drill well 4) resulting in 8,000 simulations. Thus, a total of 48,000 flow simulations were run in the simulation-regression methodology. On the other hand, in the case of rigorous Monte Carlo, flow simulation for the initial two years of production was run on all the 1,000,000 realizations for each scenario to perform the rejection sampling, totalling 2,000,000 flow simulations. Additionally, flow simulation for the subsequent three years of production were run on 120,000 posterior realizations and 4,000 prior realizations for each alternative, totalling 248,000 flow simulations. Thus, the total number of flow simulations performed in the rigorous Monte Carlo methodology was 2,248,000. To perform simulation-regression, the time-lapse seismic data was forward modeled on all 4,000 prior realizations. In contrast, to evaluate the VOI by rigorous Monte Carlo, the time-lapse seismic data was forward modeled on all the 200,000 realizations used in the rejection sampling, in addition to modeling the data for the 4,000 prior realizations. So, the forward modeling was performed for a total of 204,000 realizations.

5 VOI in a production optimization case

This case illustrates how the VOI of time-lapse seismic data can be evaluated in the context of realistic field development decisions, comprising sequential well placement and well control optimization.

5.1 Problem Formulation

Consider a 2D reservoir, which has been producing for one year with two wells – one producer and one injector. Another six wells – one producer and five injectors – are to be drilled gradually over the next six years, one at the beginning of each year. The drilling order, injector placements and rates are to be optimized such that the oil recovery is maximized over the producing life of the reservoir, which is 10 years. There is an option to collect a time-lapse seismic survey at the present time, which would reduce the uncertainty in the optimization. The VOI of the time-lapse seismic data needs to be estimated to make the decision of whether or not to collect the data.

To represent the prior uncertainty, an ensemble of 1000 prior realizations of facies, porosity and permeability are simulated in the same way as in Section 4. Some facies, porosity and permeability realizations are shown in Figure 10. Then, the AI is modeled for each realization at both the initial time before production started and after one year of production, using the same rock physics model as in Section 4. The time-lapse seismic signature is then given by the difference in the AI between the two instances of time. Figure 11 shows the modeled time-lapse seismic data for the three realizations shown in Figure 10.

Since, for the VOI computation by simulation-regression, we need to evaluate the prospect values for each decision alternative corresponding to each realization, we can only consider a finite number of alternatives. In this case, we restrict the possible alternatives (or strategies) to the set of optimal alternatives for each prior realization, which is reasonable as the prior ensemble is assumed to include the truth. Thus, we have 1000 possible alternatives or strategies, which are found by optimizing the well placements and controls on each of the 1000 prior realizations.

For this sequential well placement and well control optimization problem, we use a modified version of the flow diagnostic measure, Lorenz coefficient, as the objective function. The optimization is implemented using the diagnostics module in the Matlab Reservoir Simulation Toolbox (MRST) (Lie et al., 2012; Matlab 2017). The Lorenz coefficient, which has been shown to have good correlation with oil recovery (Moyner et al., 2015), is defined as:

$$L_c = 2\int_0^1 [F - \varphi]\, d\varphi. \qquad (16)$$

where F is the flow capacity and φ is the storage capacity (Shook and Mitchell, 2009).

The storage capacity φ is defined as the cumulative pore volume as a function of the total traveltime, which is the sum of the forward time of flight and the backward time of flight (Moyner et al., 2015).

$$\varphi(\tau) = \int_0^{\tau} \phi[\vec{x}(\hat{\tau})] d\hat{\tau}. \tag{17}$$

where ϕ is the porosity, $\hat{\tau}$ is the total traveltime, and $\vec{x}(\hat{\tau})$ is the displacement in time $\hat{\tau}$.

$$\hat{\tau} = \tau_f + \tau_b. \tag{18}$$

where τ_f is the forward time of flight and τ_b is the backward time of flight.

The flow capacity F is defined as the cumulative flux as a function of the total traveltime.

$$F(\tau) = \int_0^{\tau} \phi[\vec{x}(\hat{\tau})]/\hat{\tau} d\hat{\tau}. \tag{19}$$

The time of flight can be computed from linear steady-state transport equations (Moyner et al., 2015),

$$\vec{v} \times \nabla\tau_f = \phi, \quad \tau_f|_{inflow} = 0.$$

$$-\vec{v} \times \nabla\tau_b = \phi, \quad \tau_b|_{outflow} = 0. \tag{20}$$

where $\vec{v}$ is the fluid velocity, which can be computed by solving an incompressible fluid pressure equation,

$$\vec{v} = -K\lambda\nabla p. \tag{21}$$

where K is the absolute permeability, λ is the fluid mobility and p is the pressure.

The Lorenz coefficient is a measure of heterogeneity which indicates how much the waterflood deviates from an ideal piston-like displacement. The optimization attempts to minimize the Lorenz coefficient by making the total traveltime more homogeneous throughout the reservoir, thereby improving the oil recovery if the reservoir is saturated with oil. In our case, however, the optimization will be performed on a reservoir which has already been waterflooded for one year. Hence, we use a modified form of the Lorenz coefficient that takes into account only the pore volume occupied by oil (Moyner et al., 2015):

$$L_{c,o} = 2\int_0^1 [F - \Phi]\, S_o\, d\Phi. \tag{22}$$

where S_o is the oil saturation.

The optimal strategy for each realization consists of the optimized drilling order, injector placements and injector rates. The initial drilling order and well placements for one of the realizations are shown in Figure 12 (a). Figure 12 (b) shows the optimized drilling order and well placements for the same realization. Also, for the same realization, the optimum injection rates, optimized every three months, for each injector are shown in Figure 13.

After finding the 1000 optimal strategies corresponding to the 1000 prior realizations, flow simulation is run on each prior realization using each strategy. Figure 14 shows the oil and water production profiles for all the realizations using the optimal strategy corresponding to the first realization. These production profiles are used to compute the NPV for each strategy-realization pair using the same NPV model as given by Equation 15, with r_o = \$70/barrel, r_{wp} = \$2/barrel, r_{wi} = \$2/barrel, r = 10% per year, $cost_{drill} * a$ = \$240 million for all a. The prior value without information is then given by the maximum expected value over the realizations, which is found to be \$2,136.3 million.

5.2 VOI by simulation-regression

As in the case presented in Section 4, here too we use PLSR and PCR to perform the regression of the NPVs on the data for each strategy. To find the optimum number of PLSR and PCR components to use, we again use cross-validation to minimize the PRESS. The cross-validation is run separately for each strategy, and thus the optimum number of components to use for the regression corresponding to each strategy is obtained. The variation of the PRESS as a function of the PLSR and the PCR components for all the strategies is shown in Figure 15.

To gauge the uncertainty in the VOI from PLSR and PCR, bootstrap estimation of the VOI was performed. Table 3 shows the 10th, 50th and 90th percentiles of the VOI obtained by bootstrapping. The probability density functions (PDFs) obtained using the bootstrap samples for both PLSR and PCR are shown in Figure 16. It is seen that the VOI for PLSR and PCR overlap. However, the VOI is still quite uncertain and to reliably estimate the VOI in this complex sequential decision case, it might be useful to increase the number of realizations or reduce the number of alternatives in creative ways. Still, the decision of whether or not to collect the time-lapse seismic data in this case is quite easy to make, as the VOI estimates obtained from both PLSR and PCR are much higher than the typical cost of a time-lapse seismic survey.

Regression technique	**VOI (\$ million)**		
	10th percentile	50th percentile	90th percentile
PLSR	92.2	108.4	118.6
PCR	77.6	91.8	106.3

Table 3: VOI percentiles for PLSR and PCR obtained by bootstrapping.

6 Conclusion

We have applied a computationally efficient and flexible simulation-regression methodology to perform VOI analysis of time-lapse seismic data in the context of reservoir development. Our first VOI case demonstrates the use of time-lapse seismic data in identifying bypassed oil and thereby making infill well drilling decisions. We have applied two different methodologies to compute the VOI in this case – simulation-regression and rigorous Monte Carlo – and found that they give comparable results, thus validating the simulation-regression methodology. We have also applied bootstrapping to estimate the uncertainty in the VOI for varying number of realizations, and observed that the variance in the VOI estimate decreases with increasing number of realizations, and that a certain minimum number of realizations is required to obtain a reliable estimate of the VOI.

A limitation of the simulation-regression methodology is that it only works for a finite number of alternatives because a regression model has to be built for each alternative. One way to extend this methodology to complex decision problems with a very high number of alternatives is to restrict the space of alternatives to include only the optimal alternatives for each of the prior realizations. Our second VOI case demonstrates such a decision situation involving sequential well placement and well control optimization. However, even simulation-regression can be very computationally expensive in such cases because the number of flow simulations required to be performed is of the order of the square of the number of prior realizations, rendering it difficult to reliably estimate the VOI. Clustering techniques in conjunction with proxy models can be used to lower the computational cost in such cases.

References

Abdi, H., 2010, Partial least squares regression and projection on latent structure regression (PLS Regression), *WIREs Comp Stat*, v. 2, p. 97-106, DOI 10.1002/wics.51.

Avseth, P., Dvorkin, J., Mavko, G., & Rykkje, J., 2000, Rock physics diagnostic of North Sea sands: Link between microstructure and seismic properties, *Geophysical Research Letters*, v. 27, no. 17, p. 2761-2764.

Barros, E. G. D., Van den Hof, P. M. J., & Jansen, J. D., 2015, Value of information in closed-loop reservoir management, *Computational Geosciences*, v. 20, no. 3, p. 737-749.

Bratvold, R. B., Bickel, J. E., & Lohne, H. P., 2009, Value of information in the oil and gas industry: past, present, and future, *Society of Petroleum Engineers*, SPE 110378.

Eidsvik, J., Bhattacharjya, D., & Mukerji, T., 2008, Value of information of seismic amplitude and CSEM resistivity, *Geophysics*, 73(4), R59-R69, doi: 10.1190/1.2938084.

Eidsvik, J., Mukerji, T., & Bhattacharjya, D., 2016, *Value of Information in the Earth Sciences*, Cambridge University Press, 396 p.

Eidsvik, J., Dutta, G., Mukerji, T., & Bhattacharjya, D., 2017, Simulation-Regression Approximations for Value of Information Analysis of Geophysical Data, *Mathematical Geosciences*, doi: 10.1007/s11004-017-9679-9.

Goda, T., Murakami, D., Tanaka, K., & Sato, K., 2018, Decision-theoretic sensitivity analysis for reservoir development under uncertainty using multilevel quasi-Monte Carlo methods, *Computational Geosciences*, https://doi.org/10.1007/s10596-018-9735-7.

Gassmann, F., 1951, Uber die Elastisitat Poroser Medien, Vierteljahrschr. Naturforsch. Ges. Zurich, 96, 1-23.

Hastie, T., Tibshirani, R., & Friedman, J., 2009, *The Elements of Statistical Learning*, Springer-Verlag, 763 p.

Heath, A., Manolopoulou, I., & Baio, G., 2016, Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation, *Stat in Med*, 35, 4264-4280.

Hong, A. J., Bratvold, R. B., Thomas, P., & Hanea, R. G., 2018, Value-of-information for model parameter updating through history matching, *Journal of Petroleum Science and Engineering*, 165, 253-268.

Houck, R. T., 2007, Time-lapse seismic repeatability – How much is enough?, *The Leading Edge*, 26(7), 828-834, doi: 10.1190/1.2756860.

Howard, R.A., 1966, Information value theory, *IEEE Transactions on Systems Science and Cybernetics*, v. 2, no. 1, p. 22-26.

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, *Computational Geosciences*, v. 16, no. 2, p. 297-322, doi 10.1007/s10596-011-9244-4.

Matlab, version 2017a, 2017, Natick, Massachusetts: The Mathworks, Inc.

Mavko, G., Mukerji, T., & Dvorkin, J., 2009, *The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media*, 2nd ed., Cambridge University Press.

Moyner, O., Krogstad, S., & Lie, K.-A., 2015, The application of flow diagnostics for reservoir management, *SPE Journal*, v. 20, no. 2, p. 306-323.

Rosipal, R., & Kramer, N., 2006, Overview and Recent Advances in Partial Least Squares, *Subspace, Latent Structure and Feature Selection*, p. 34-51, DOI 10.1007/11752790_2.

Satija, A., Scheidt, C., Li, L., & Caers, J., 2017, Direct forecasting of reservoir performance using production data without history matching, *Computational Geosciences*, DOI 10.1007/s10596-017-9614-7.

Shook, G., & Mitchell, K., 2009, A robust measure of heterogeneity for ranking earth models: The F Phi curve and dynamic Lorenz coefficient, *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, SPE-124625-MS.

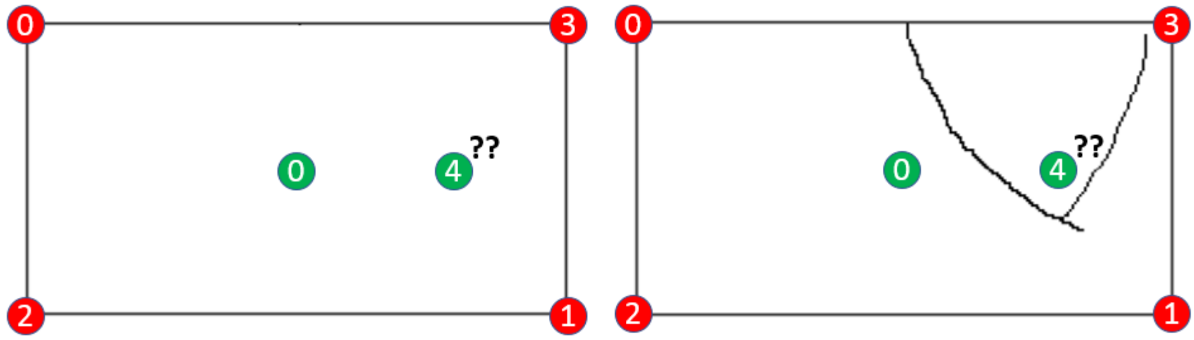
Steinwart, I., & Christmann, A., 2008, *Support Vector Machines*, Springer-Verlag.

Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol*, 34: 1-21.

Strong, M., Oakley, J., & Brennan, A., 2014, Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: A nonparametric regression approach, *Med Dec Making*, 34, 311-326.

Trainor-Guitton, W. J., Mukerji, T., & Knight, R., 2013, A methodology for quantifying the value of spatial information for dynamic Earth problems, *Stoch Environ Res Risk Assess*, 27: 969. DOI 10.1007/s00477-012-0619-4.

Figures



(a) (b)

Figure 1: Locations of injectors (red circles) and producers (green circles) in a reservoir with (a) no flow barrier, and (b) a flow barrier. Numbers within the circles indicate the order of drilling.

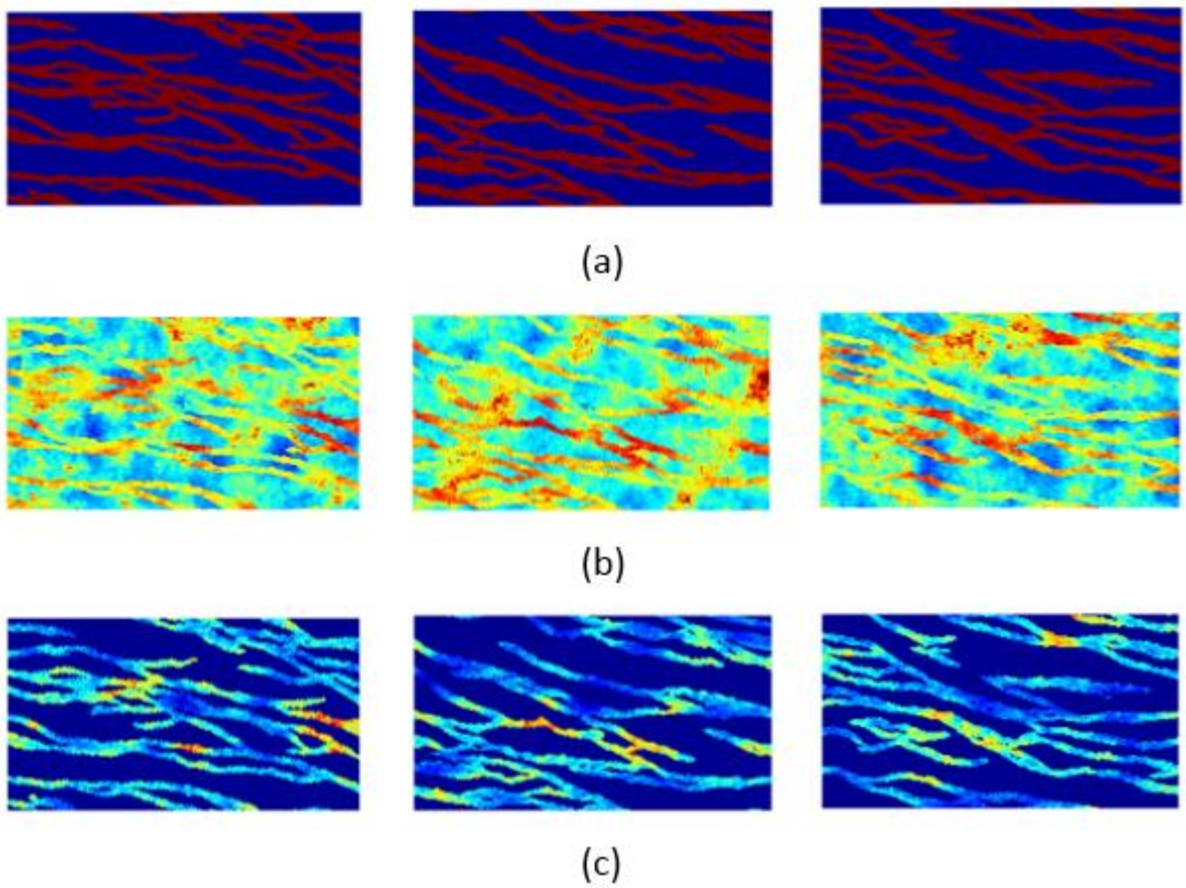


Figure 2: Three prior realizations each of (a) facies, (b) porosity, and (c) permeability.

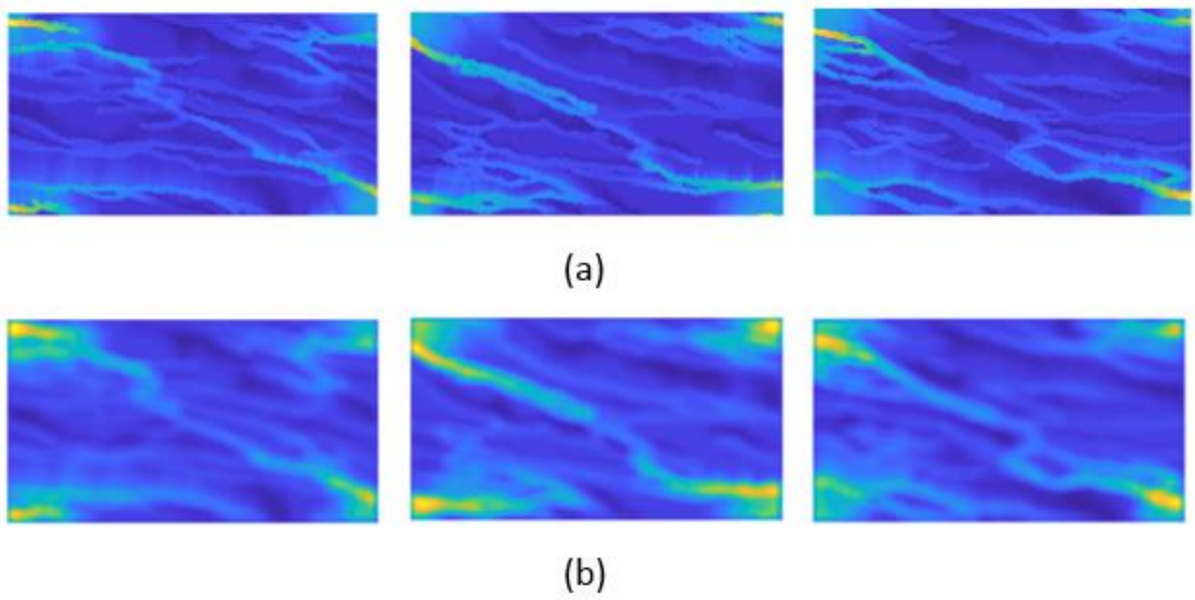


Figure 3: Change in AI at (a) the geostatistical scale, and (b) the seismic scale for the three realizations shown in Figure 2.

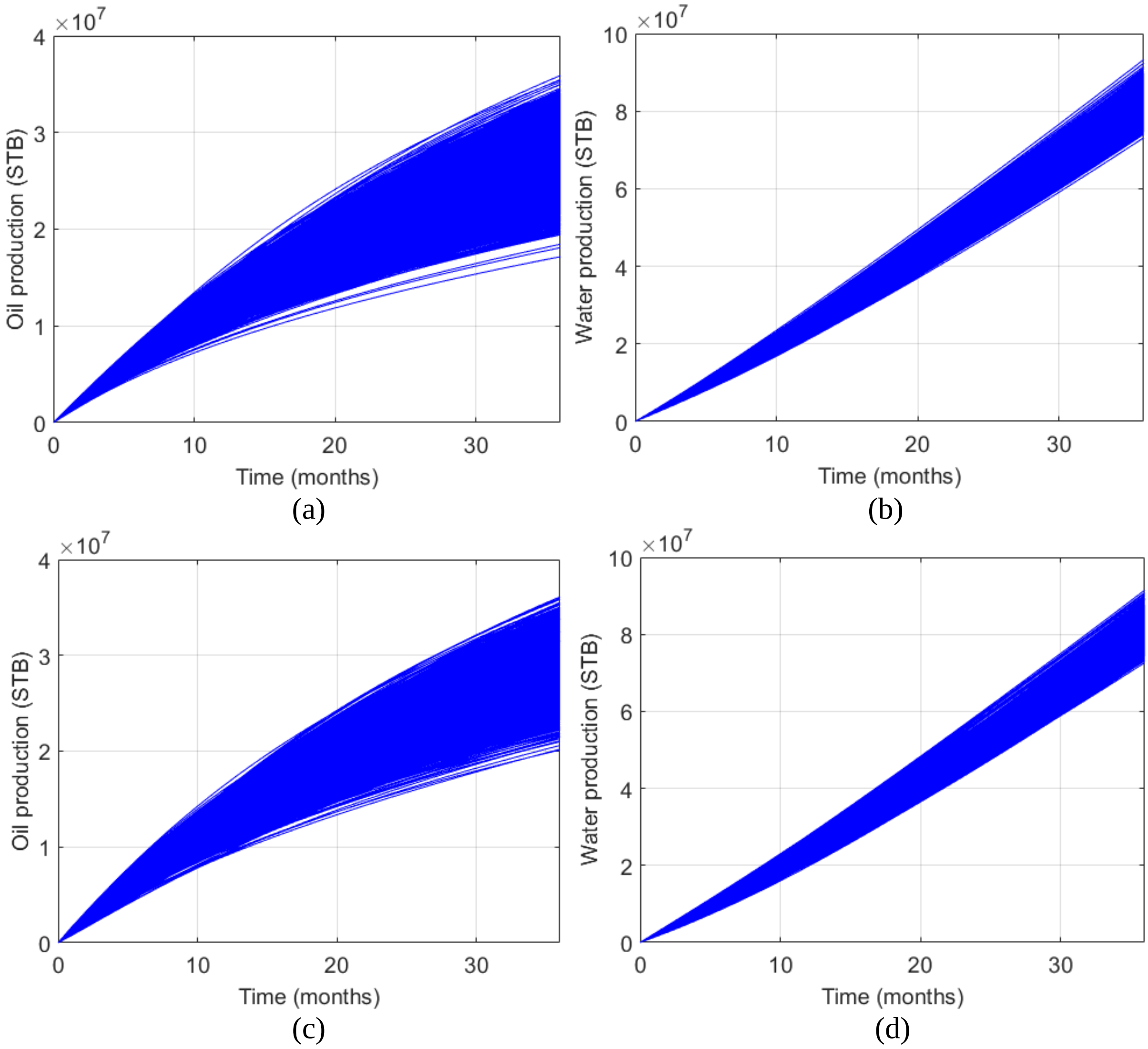


Figure 4: (a) The oil production, and (b) the water production profiles for all realizations for the alternative "Do not drill". (c) The oil production, and (d) the water production profiles for all realizations for the alternative "Drill".

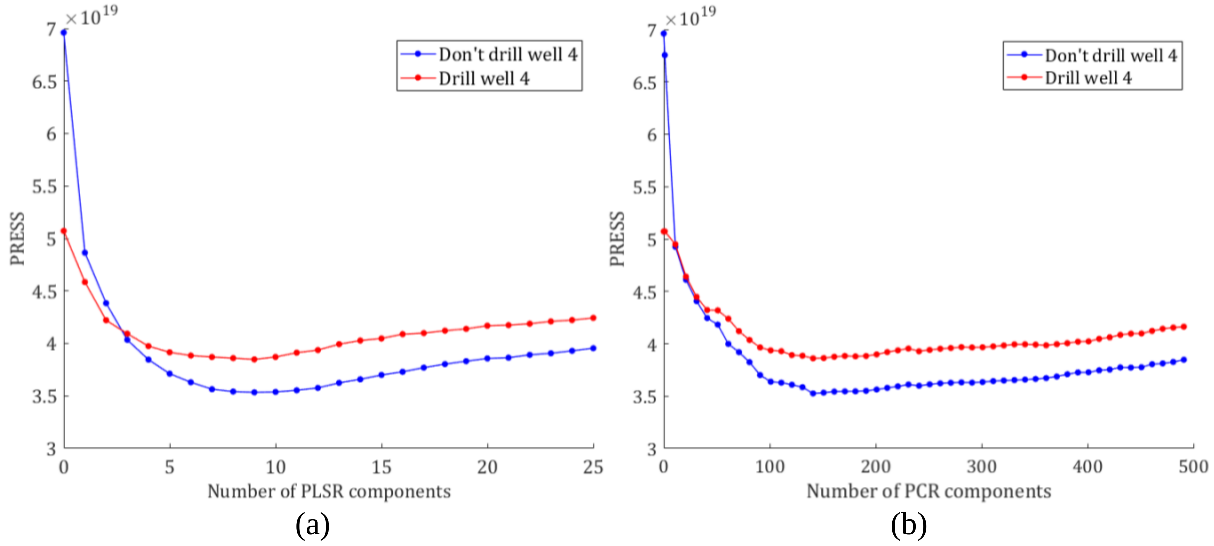


Figure 5: The PRESS as a function of the number of (a) PLSR components, and (b) PCR components.

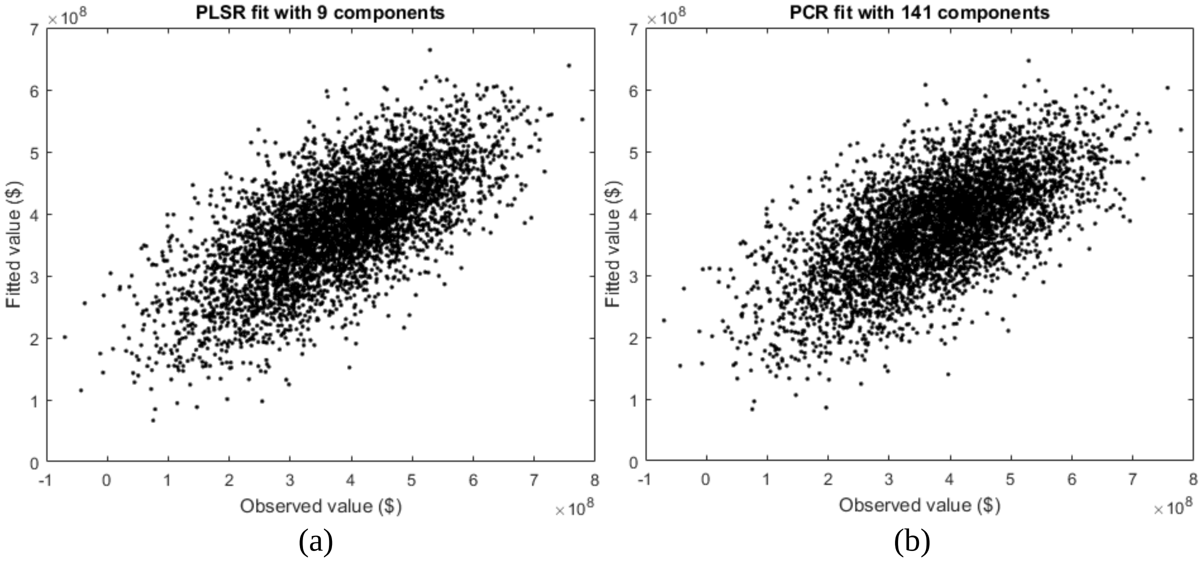
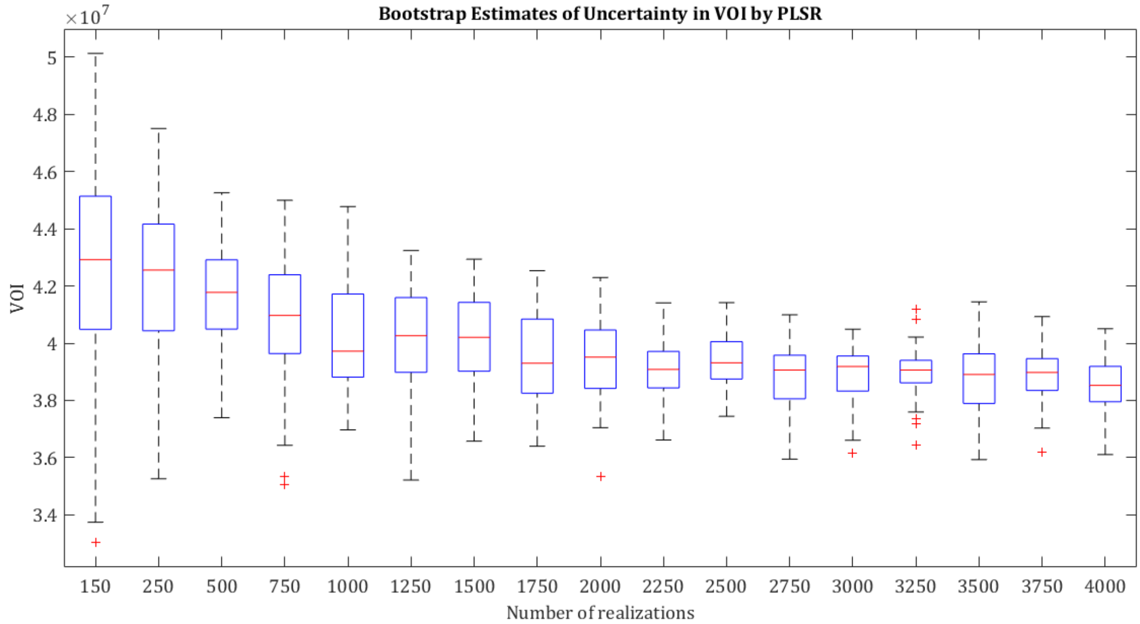
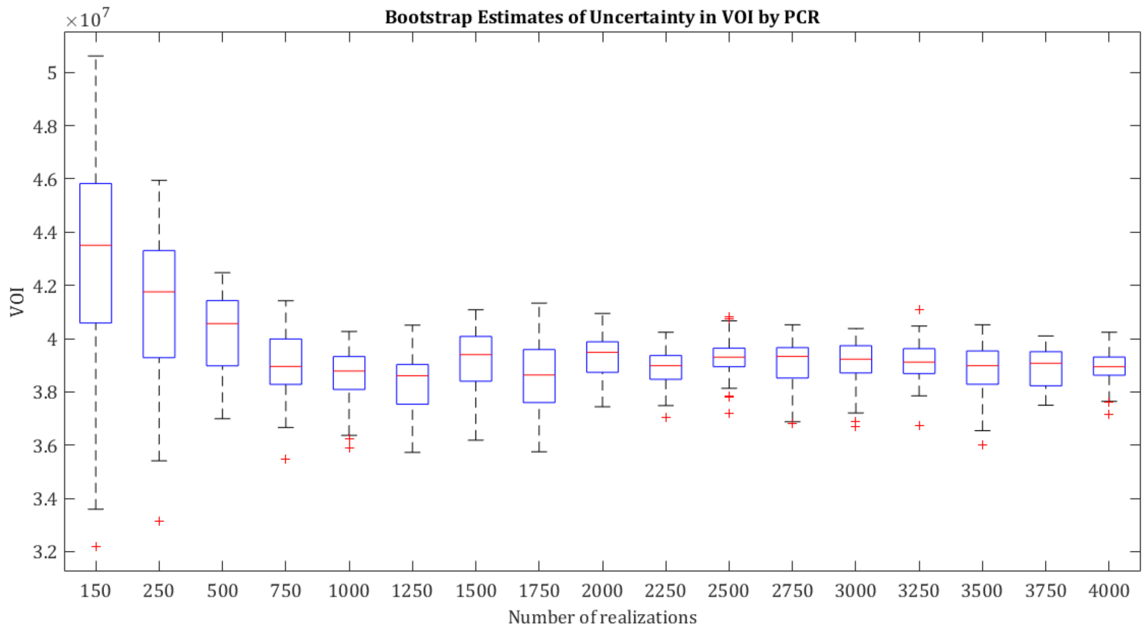


Figure 6: Plots of fitted values versus observed values for (a) a PLSR model with 9 components, and (b) a PCR model with 141 components.



(a)



(b)

Figure 7: Bootstrap estimates of uncertainty in the VOI computed using (a) PLSR, and (b) PCR.

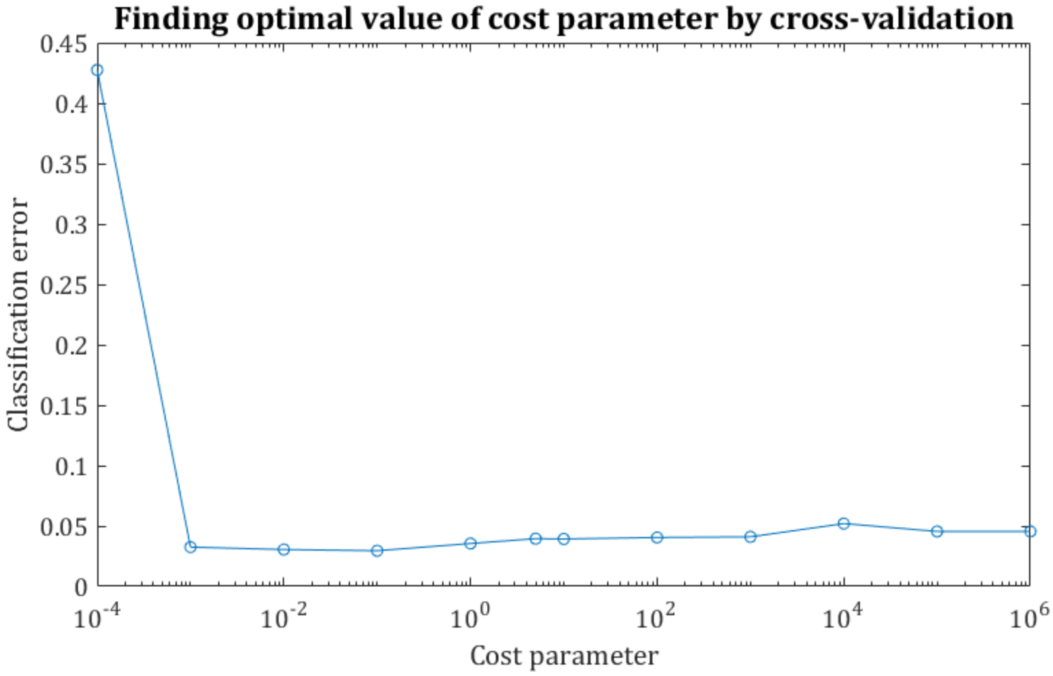


Figure 8: Cross-validation results showing the variation of the classification error with the value of the cost parameter.

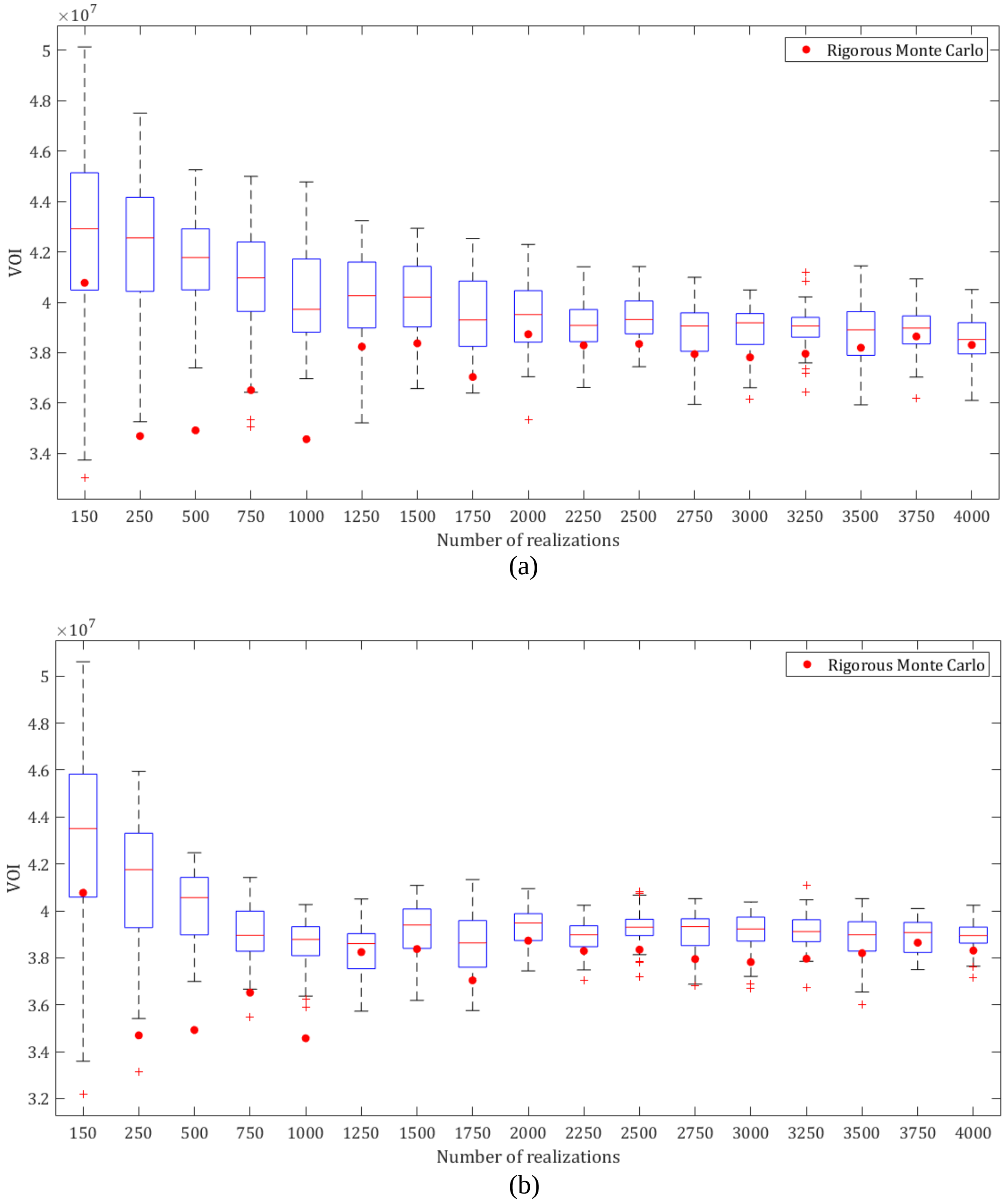


Figure 9: Comparison of the VOI results obtained by rigorous Monte Carlo with those obtained by (a) PLSR, and (b) PCR.

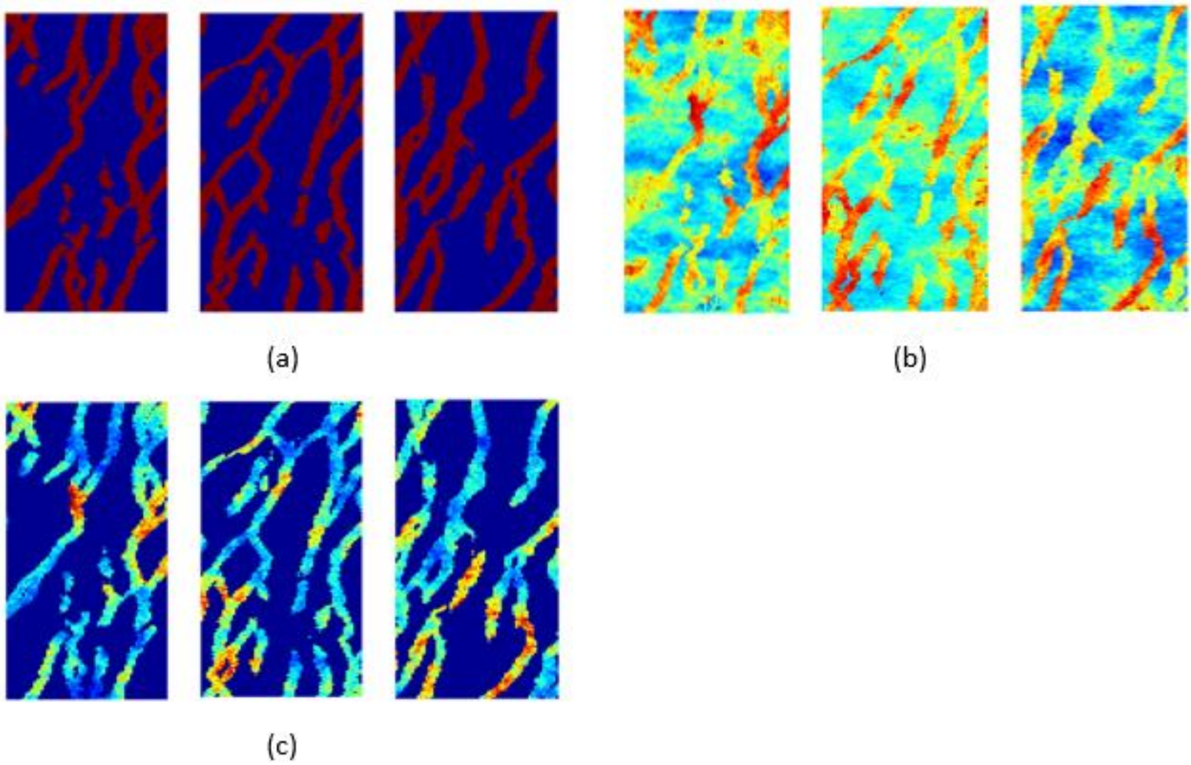


Figure 10: Three realizations each of (a) facies, (b) porosity, and (c) permeability.

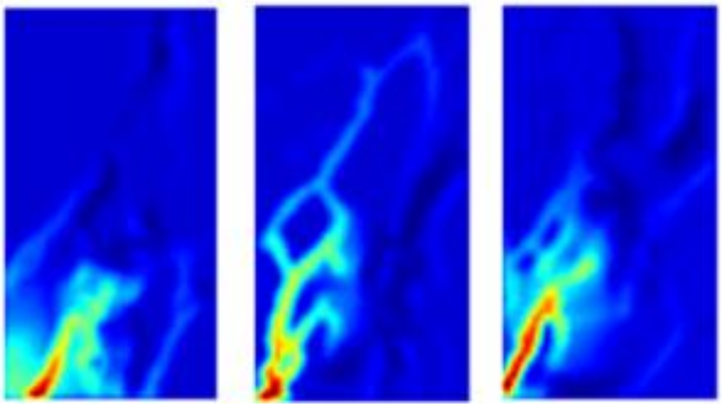
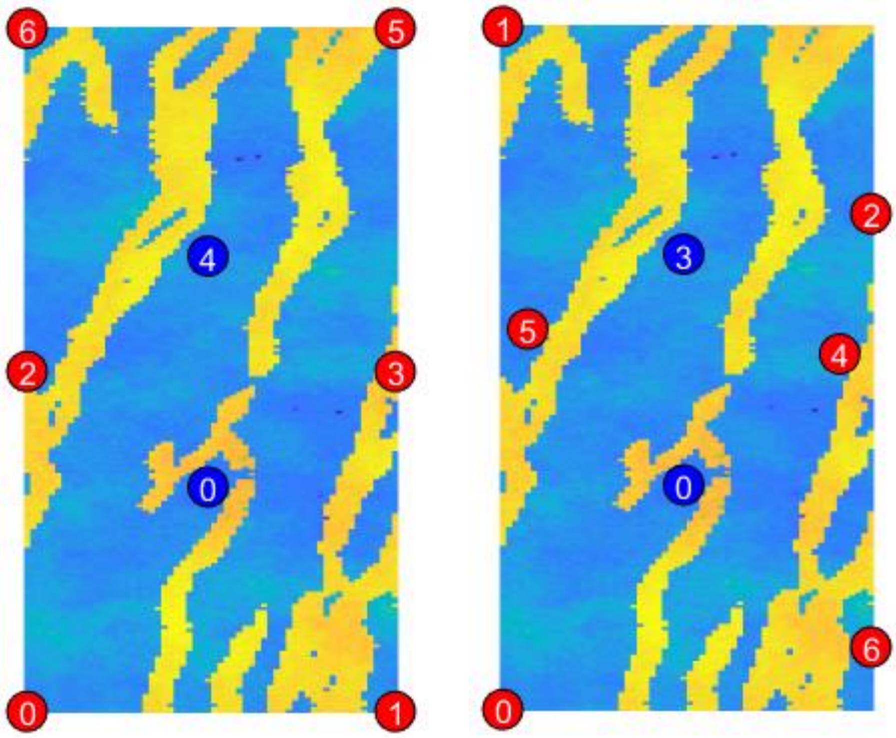


Figure 11: The modeled time-lapse seismic data for the three realizations shown in Figure 10.



(a) (b)

Figure 12: (a) Initial drilling order and well placements for one realization, and (b) optimized drilling order and well placements for the same realization.

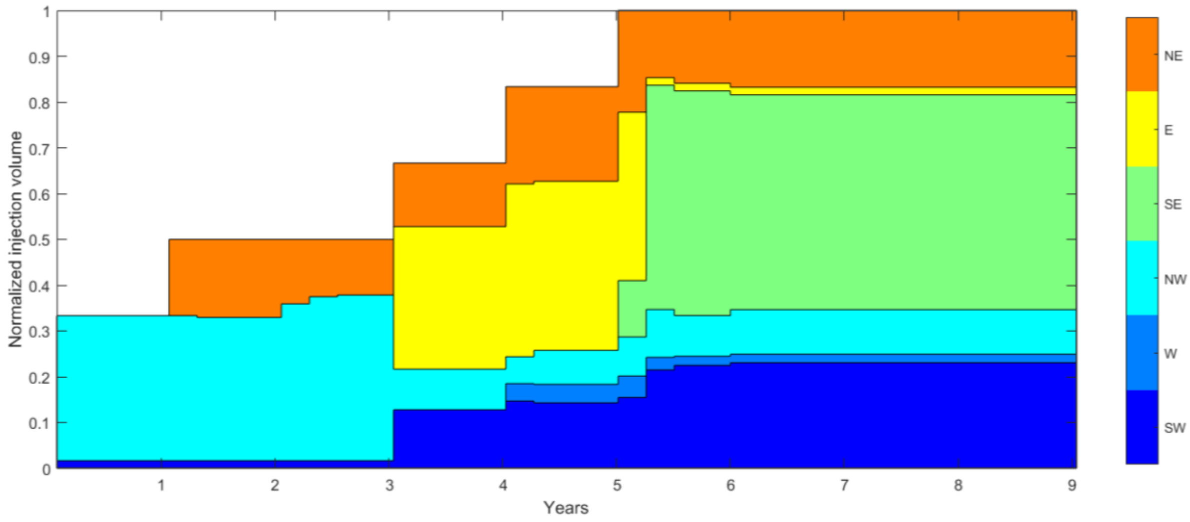
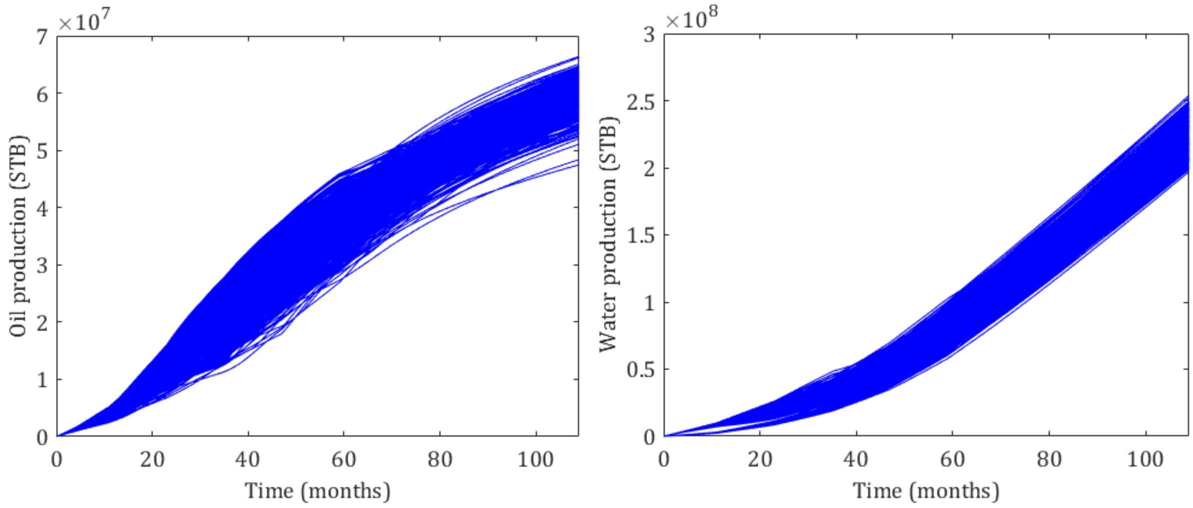


Figure 13: Optimized injection rates for each injector for the same realization as in Figure 12. The abbreviations on the colorbar indicate geographic locations of the injectors (e.g. 'NE' stands for North East, and so forth).



(a) (b)

Figure 14: (a) The oil production, and (b) the water production profiles for all the realizations using the optimal strategy corresponding to the first realization.

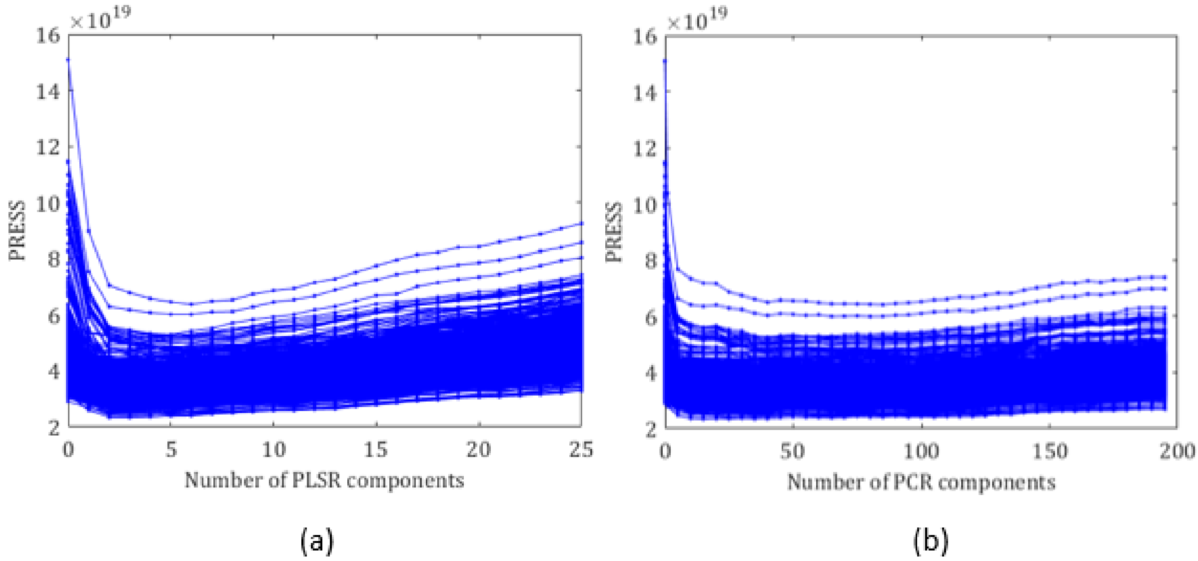


Figure 15: The variation of the PRESS with (a) the number of PLSR components, and (b) the number of PCR components for all the strategies.

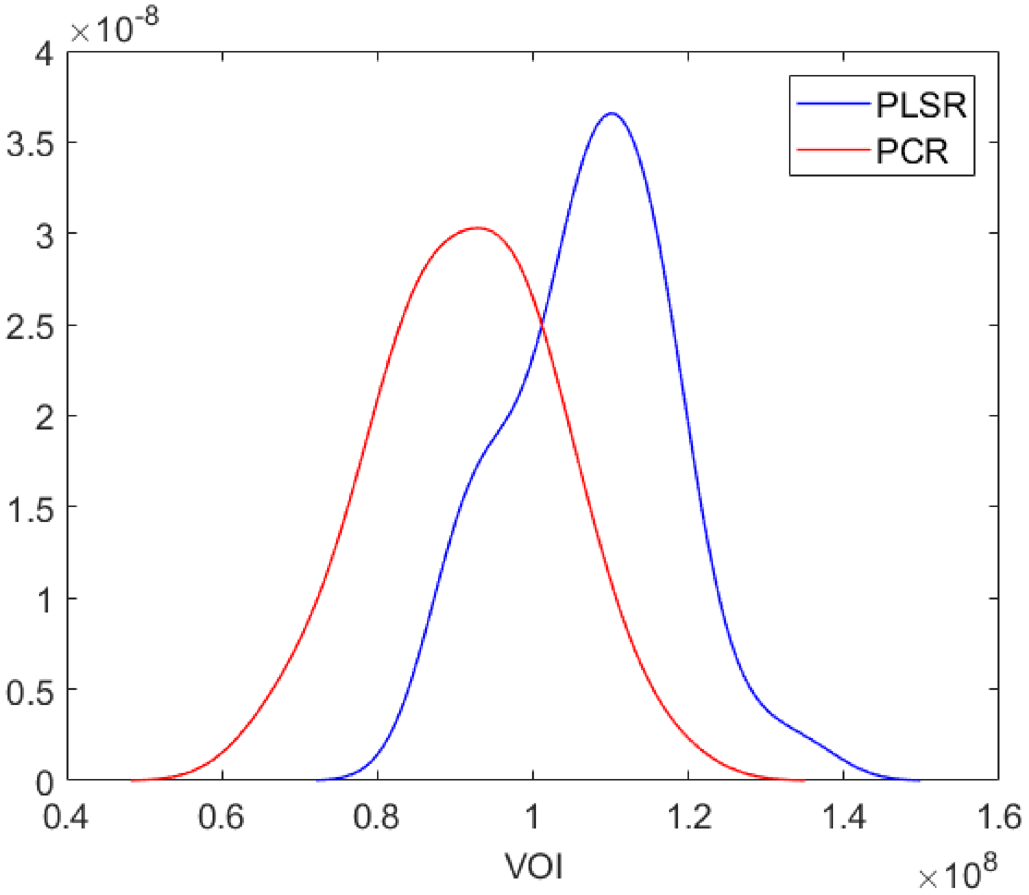


Figure 16: PDFs of VOI computed by PLSR and PCR obtained from bootstrap samples.