Geophysics-Based Fluid-Facies Predictions Using Ensemble Updating of Binary State Vectors

Margrethe Kvale Loe¹ · Dario Grana² ·
Håkon Tjelmeland¹

Received: 22 January 2020 / Accepted: 12 January 2021 / Published online: 8 February 2021
© The Author(s) 2021

Abstract Fluid flow simulations are commonly used to predict the fluid displacement in subsurface reservoirs; however, model validation is challenging due to the lack of direct measurements. Geophysical data can be used to monitor the displacement of the fluid front. The updating of the fluid front location in two-phase flow problems based on time-lapse geophysical data can be formulated as an inverse problem, specifically a data assimilation problem, where the state is a vector of binary variables representing the fluid-facies and the observations are measurements of continuous geophysical properties, such as electrical or elastic properties. In geosciences, many data assimilation problems are solved using ensemble-based methods relying on the Kalman filter approach. However, for discrete variables, such approaches cannot be applied due to the Gaussian-linear assumption. An innovative approach for mixed discrete-continuous problems based on ensemble updating of binary state vectors is presented for fluid-facies prediction problems with time-lapse geophysical properties. The proposed inversion method is demonstrated in a synthetic two-dimensional simulation example where water is injected into a reservoir and hydrocarbon is produced. Resistivity values obtained from controlled-source electromagnetic data are assumed to be available at different times. According to the results, the proposed inversion method is to a large extent able to reproduce the true underlying binary field of fluid-facies.

Keywords Data assimilation · Ensemble methods · Fluid-facies prediction · Markov chain models

✉ Margrethe Kvale Loe
margrethe.loe@ntnu.no

¹ Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

² Department of Geology and Geophysics, University of Wyoming, Laramie, WY, USA

1 Introduction

Monitoring the spatial distribution of fluids in the subsurface is a challenging modeling problem due to the uncertainty in rock and fluid properties. If the petrophysical properties, (e.g., porosity and permeability) and the subsurface conditions (e.g., temperature and pressure) are known, the fluid volumes and their spatial distribution can be predicted by solving partial differential equations for fluid flow in porous media (Bear 2013). However, in practical applications, the majority of the model parameters cannot be measured directly and can only be approximated from indirect geophysical measurements (Aki and Richards 1980; Doyen 2007). Hence, the accuracy in the model predictions depends on the quality of the data and the approximations in the physical models. For these reasons, it is necessary to update the fluid models every time new geophysical data are available.

One of the main challenges in monitoring fluid flow in the subsurface is obtaining accurate and precise predictions of the fluid front at different times. Geophysical surveys, acquired at the surface, provide measurements of physical properties whose changes in time reflect variations in rock and fluid properties. Hence, the location of the fluid front can be updated based on time-lapse geophysical data and fluid flow simulations. Examples of time-lapse geophysical surveys include seismic data that depend on changes in elastic properties and electromagnetic data that depend on changes in electrical properties.

Changes in fluid saturations modify the elastic and electrical response of reservoir rocks. Therefore, unknown fluid flow saturations can be predicted as the solution of an inverse problem where the data are geophysical observations and the governing equations are geophysical models (Doyen et al. 2000). However, the resolution of geophysical data measured at the surface is generally lower than the desired resolution of the fluid distribution model due to the bandlimited nature of the signal and the relatively low signal-to-noise ratio. Therefore, prediction of the fluid saturations is highly uncertain and saturation models can be inaccurate. In many applications, geophysical data can be used to interpret the fluid phase at a specific location; however, it is generally difficult to accurately predict fluid percentages. For this reason, rather than using continuous variables representing fluid volumes or fluid saturations, in this work, the inverse problem is formulated in terms of a discrete variable. For multiphase flow problems, the discrete variable represents fluid-facies (or fluid units), such as water-, air-, hydrocarbon-, or CO₂-saturated rocks, depending on the application. In particular, for two-phase flow problems, fluid-facies can be represented by a binary variable. The proposed methodology is defined in a probabilistic setting; therefore, the model property of interest is a discrete random variable.

In this work, the focus is on fluid properties that cause changes in the electrical response measured by electromagnetic data, such as electrical resistivity tomography (ERT) and controlled-source electromagnetic (CSEM). The goal of this work is to predict fluid-facies in subsurface reservoirs based on estimated resistivity values from time-lapse geophysical data. From a mathematical point of view, this is a data assimilation problem where the unknown state is a discrete random variable (the fluid-facies) and the observed data are geophysical measurements (the resistivity values). The resistivity of fluid-saturated porous rocks is obtained from electromagnetic measurements.

Here, it is assumed that resistivity values have been computed from CSEM data as a result of a preliminary inversion process. In near surface geophysics applications, ERT data are commonly acquired to investigate the seasonal water displacement (Flinchum et al. 2018; Kotikian et al. 2019; Claes et al. 2020). In hydrocarbon exploration and carbon sequestration, CSEM data are often measured to study the fluid spatial distribution and monitor the fluid front location (Weitemeyer et al. 2006; Lien and Mannseth 2008; Orange et al. 2009; Constable 2010; MacGregor 2012; Lien et al. 2014). The prediction of resistivity from electromagnetic data is itself an inverse problem (Gasperikova and Hoversten 2006; Buland and Kolbjørnsen 2012; MacGregor 2012; Bhuyian et al. 2012). Probabilistic and deterministic approaches have been proposed with different model parameterizations, in terms of resistivity or saturation. Time-lapse electrical resistivity inversions have been proposed in several applications (Berre et al. 2011; Shahin et al. 2012; Lien et al. 2014; Tveit et al. 2015; Commer et al. 2016; Bergmann et al. 2017).

Classification methods for litho-facies and fluid-facies based on geophysical data have also been proposed. Several clustering analysis methods described in Hastie et al. (2009) have been used for facies classification in geophysics inverse problems (Doyen 2007). Classification methods include supervised and unsupervised techniques (Hastie et al. 2009; Martinez and Martinez 2015). Clustering and pattern recognition methods have been used to classify geophysical measurements. However, the majority of these applications focuses on static characterization problems, with a spatial correlation component (i.e. facies are spatially correlated to mimic the geological continuity) but without a temporal component (i.e. facies are predicted at a given time step, typically before dynamic processes start).

The focus of this work is time-dependent fluid-facies characterization problems where the spatial distribution of fluid-facies changes in time and is monitored using time-lapse geophysical properties. Therefore, in this study a dynamic fluid-facies classification is presented, and it is applicable to geological dynamic problems where one fluid (e.g., water) replaces another fluid (e.g., hydrocarbon) in rock formations. Reservoir modeling with time-lapse data is a data assimilation problem where the model variables are predicted and updated when new measurements become available. Several stochastic optimization methods have been proposed for data assimilation problems, and during the last decade, ensemble-based methods have become the most popular stochastic data assimilation method in geoscience applications.

Data assimilation can refer to a range of different inference procedures, of which the two most common are filtering and smoothing. In the present article, the focus is exclusively on the filtering problem. There are two main classes of ensemble-based filtering methods: particle filters (Doucet et al. 2001) and ensemble Kalman filters (EnKFs) (Evensen 2009). Hybrid versions of these filters have also been proposed. Particle filters have the advantage of being exact in the sense that as the ensemble size goes to infinity, the ensemble representations of the series of filtering distributions converge to the corresponding correct series of distributions. Particle filters are also very general as they do not rely on any specific assumptions about the distributions of the model variables. Hence, particle filter methods are, in principle, applicable to both discrete and continuous variables. In practice, however, particle filters are known to collapse when the dimension of the state vector is large. The EnKF is a filtering

method which relies on a linear-Gaussian assumption about the underlying model. Despite the linear-Gaussian assumption, studies show that the EnKF provides good results even in non-linear, non-Gaussian problems, and unlike the particle filter it also scales well to problems with very high-dimensional state vectors. The EnKF has been applied to geophysical data assimilation and history matching problems using seismic and electromagnetic data (e.g. Tveit et al. 2015; Tveit et al. 2020). Recent publications focus on the integration of fluid flow simulation and geophysical data assimilation for the monitoring of the fluid front location (Trani et al. 2012; Leeuwenburgh and Arts 2014; Zhang and Leeuwenburgh 2017). However, the Gaussian approximations make the EnKF applicable only in situations with continuous variables. For problems with discrete variables, such as fluid-facies classification, the filter is not appropriate.

Ensemble filtering of discrete variables is a challenging problem which has received fairly little attention in the literature compared to filtering of continuous variables. Oliver et al. (2008) propose a strategy where the EnKF is used to update the discrete variables. Specifically, they propose a two-step strategy where, in the first step, the EnKF is used to update the discrete variables as if they were continuous, and in the second step, the updated continuous-valued variables are mapped back to the original discrete state space using the Viterbi algorithm (Viterbi 1967). Loe and Tjelmeland (2020) present an alternative updating method for binary vectors in one-dimensional space based on a generalized approach of the EnKF. Instead of using a linear-Gaussian model assumption in the ensemble update, as in the EnKF, they construct an update based on a hidden Markov model assumption. To capture as much information as possible from the forecast ensemble, including potential non-Markov properties, the expected number of components of the binary state vector that remain unchanged is maximized.

This paper presents an ensemble-based data assimilation method for a problem where the state vector at each time step is a vector of binary variables and the observations are continuous-valued estimated resistivity values. The binary variables of the state vector represent two different fluid-facies, for example water and hydrocarbon or CO₂, and each binary variable is connected to a continuous-valued variable representing water saturation. High water saturation values indicate the presence of the water facies, while low saturation values indicate the presence of the other fluid-facies. The proposed ensemble filtering method alternates between a forecast step performed in the continuous state space of the saturation variable and an update step performed in the discrete state space of the fluid-facies variable, and between each step an appropriate mapping from one state space to the other is performed. The update step is performed according to the updating procedure for binary state vectors proposed in Loe and Tjelmeland (2020). The proposed inversion method is demonstrated in a synthetic two-dimensional example representing a two-phase flow problem with resistivity values available at different times. According to the results, the proposed procedure is to a large extent able to reproduce the true underlying binary field of fluid-facies. Larger ensemble sizes provide more accurate results, but the results obtained with smaller ensemble sizes are also satisfactory.

The remains of this paper takes the following outline. First, Sect. 2 formulates the inverse problem more formally and presents the proposed ensemble-based inversion method. Next, Sect. 3 presents numerical results based on a two-dimensional synthetic

model for a two-phase fluid flow problem. Finally, a few closing remarks are given in Sect. 4.

2 Methodology

2.1 Inverse Problem Setting

The problem addressed in this work is the prediction of fluid-facies from time-lapse resistivity values. Consider a time series $\{k_i\}_{i=1}^{N_t}$ where $k_i = (k_i^1, \dots, k_i^{N_k})$ represents an N_k -dimensional vector of fluid-facies at a certain time t_i , $i = 1, \dots, N_t$ in a reservoir. Each component k_i^j of k_i can take a value in the set $\{0, 1, \dots, K - 1\}$, where K is the number of different fluid-facies. Given a corresponding series of resistivity data $\{d_i\}_{i \in \mathbb{T}}$, where $d_i = (d_i^1, \dots, d_i^{N_d}) \in \mathbb{R}^{N_d}$ is an N_d -dimensional vector of resistivity measurements recorded at time t_i , and $\mathbb{T} \subseteq \{1, \dots, N_t\}$, the goal is, for each time step $i = 1, \dots, N_t$, to assess the distribution of fluid-facies k_i in the reservoir. Notice from the set $\mathbb{T}$ that an observation d_i may be available at every time step $i = 1, \dots, N_t$, or just a subset of them.

In this work, each component k_i^j of k_i is assumed to be connected to a continuous variable $m_i^j \in [s_{wi}, 1]$ representing water saturation, where s_{wi} is an irreducible water saturation value; that is, the fraction of water that a porous rock can retain due to non-connected porosity, low permeability and/or capillary forces. Here, the irreducible water saturation value $s_{wi} = 0.2$ is assumed. Given $\{m_i\}_{i=1}^{N_t}$, the resistivity data $\{d_i\}_{i \in \mathbb{T}}$ are assumed to be conditionally independent, so that the vector d_i at time t_i depends only on m_i according to

$$d_i = f(m_i, e_i), \quad (1)$$

where f is a known, possibly non-linear function, and the variable $e_i \in \mathbb{R}^{N_e}$ is an N_e -dimensional vector of measurement random errors assumed to follow a known probability distribution. Similarly, the saturation m_{i+1} at time step $i + 1$, given all the saturation values $m_1, \dots, m_i$ up to time step i , depends only on m_i according to a known forward model,

$$m_{i+1} = g(m_i), \quad (2)$$

for $i = 1, \dots, N_t$, where g is the fluid flow simulation, generally given by a system of partial differential equations solved by finite difference methods (Aziz 1979).

The goal of this work is, for each time step $i = 1, \dots, N_t$, to assess the filtering distribution $p(k_i | d_{1:i})$, where $d_{1:i} = \{d_j; j \in \mathbb{T} \cap \{j \leq i\}\}$, that is, the distribution of fluid-facies k_i given all the resistivity data up to time t_i . Only $K = 2$ fluid-facies are assumed in this work: facies 1 represents water and facies 0 represents another fluid-facies. The relationship between the fluid-facies $k_i^j \in \{0, 1\}$ and the saturation

value $m_i^j \in [s_{wi}, 1]$ is assumed as

$$k_i^j = \begin{cases} 0, & \text{if } m_i^j \in [s_{wi}, r], \\ 1, & \text{if } m_i^j \in (r, 1], \end{cases} \quad (3)$$

where $r \in (s_{wi}, 1)$ is some appropriate threshold. The value of the parameter r might vary from one application to another. A reasonable choice is to set $r = 0.5$ such that each fluid facies is named after the predominant fluid component. However, from a reservoir management perspective, the focus is generally on areas with a high concentration of hydrocarbon. Therefore one could choose a lower value to identify the regions that are economically valuable.

2.2 Forward Model

The prediction of the time-dependent electrical response of a reservoir model requires a rock-physics model to link the petrophysical properties, such as porosity and fluid saturations, to the resistivity of the saturated porous rocks and a fluid flow simulation model to compute the saturation at a given time step, given the saturation at the previous time step. In the proposed approach, porosity and permeability are assumed to be estimated from pre-injection geophysical measurements (e.g., seismic data). Alternatively, multiple geostatistical simulations of porosity and permeability can be generated to repeatedly apply the methodology to an ensemble of realizations; however, the computational cost would linearly increase with the number of realizations.

A rock-physics model is a relationship to predict the geophysical response of saturated porous rocks. Assuming that the porosity ϕ of the porous rock is known, the resistivity R (the measured data d in the inverse problem) of the porous rock saturated with water saturation s_w can be predicted using Archie's law (Mavko et al. 2009),

$$R = \frac{R_w}{\phi^a s_w^b}, \quad (4)$$

where R_w is the resistivity of formation water, a is the cementation exponent, and b is the saturation exponent (Mavko et al. 2009). The parameters R_w , a and b in Eq. (4) are assumed to be constant in time. Archie's equation is valid for clean sandstone formations. For formations with a small to medium clay volume, Archie's equation can be modified to account for the conductivity of the clay mineral as in Simandoux and Poupon-Leveaux models (Mavko et al. 2009).

The dynamic model that governs two-phase fluid flow in porous media is based on the constitutive equations of mass and momentum balance. The model is numerically solved using the black-oil framework to predict the saturation and pressure at each time step, given the initial rock and fluid parameters (Aziz 1979). In this work, the MATLAB Reservoir Simulation Toolbox (Lie 2019) is adopted.

2.3 Inversion Method

To solve the inverse problem presented above, an ensemble-based strategy where the forecast step is performed in the continuous domain of m_i and the update step is performed in the discrete domain of k_i is adopted. At each time t_i , an ensemble of fluid-facies fields $\{k_i^{(1)}, \dots, k_i^{(M)}\}$ represents the distribution of k_i given the resistivity data up to time t_{i-1} , that is $d_{1:i-1}$. Likewise an ensemble of saturation fields $\{m_i^{(1)}, \dots, m_i^{(M)}\}$ represents the distribution of m_i given the same resistivity data. Correspondingly, the distributions of k_i and m_i given resistivity data up to time t_i , that is $d_{1:i}$, are also represented by ensembles, which are denoted by $\{\tilde{k}_i^{(1)}, \dots, \tilde{k}_i^{(M)}\}$ and $\{\tilde{m}_i^{(1)}, \dots, \tilde{m}_i^{(M)}\}$, respectively. The main steps of the inversion procedure are summarized in Algorithm 1, while each step is studied in closer detail in the following sections.

Initialize: $\{m_1^{(1)}, \dots, m_1^{(M)}\}$.

For $i = 1, \dots, N_t$ do

1. Update:

(a) Map $m_i^{(l)} \rightarrow k_i^{(l)}$, $l = 1, \dots, M$, using Eq. (3).

(b) Update $k_i^{(l)} \rightarrow \tilde{k}_i^{(l)}$, $l = 1, \dots, M$, as discussed in Sect. 2.3.1.

2. Forecast:

(a) Map $\tilde{k}_i^{(l)} \rightarrow \tilde{m}_i^{(l)}$, $l = 1, \dots, M$, as discussed in Sect. 2.3.2.

(b) Generate $m_{i+1}^{(l)} = g(\tilde{m}_i^{(l)})$, $l = 1, \dots, M$, using Eq. (2).

End

Algorithm 1: Inversion method

2.3.1 The Update Step

As summarized in Algorithm 1, the update step of the proposed approach involves two parts. First, the ensemble $\{m_i^{(1)}, \dots, m_i^{(M)}\}$ is mapped to a corresponding ensemble $\{k_i^{(1)}, \dots, k_i^{(M)}\}$ using the assumed relation between k_i and m_i in Eq. (3). Second, $\{k_i^{(1)}, \dots, k_i^{(M)}\}$ is updated to take the new observation d_i at time t_i into account. In the following, the two parts of the update step are discussed in more detail.

The ensemble of saturation fields $\{m_i^{(1)}, \dots, m_i^{(M)}\}$ is mapped to a corresponding ensemble of fluid-facies fields $\{k_i^{(1)}, \dots, k_i^{(M)}\}$ by simply applying Eq. (3) to each element in each of the ensemble members, that is, set

$$k_i^{(l),j} = \begin{cases} 0, & \text{if } m_i^{(l),j} \in [s_{wi}, r], \\ 1, & \text{if } m_i^{(l),j} \in (r, 1], \end{cases} \quad (5)$$

for each location $j = 1, \dots, N_k$ for each ensemble member $l = 1, \dots, M$.

To update the ensemble $\{k_i^{(1)}, \dots, k_i^{(M)}\}$ to take the new observation d_i into account, the procedure proposed in Loe and Tjelmeland (2020) is adapted to the situation considered in the present article. In the present article, it is assumed that the fluid-facies k_i at each time step $i = 1, \dots, N_t$ is defined on a two-dimensional lattice. However, the method in Loe and Tjelmeland (2020) is applicable only for vectors with a one-dimensional spatial arrangement. Therefore, in order to apply their procedure, the updating of each column in the lattice is done independently of the others. Of course, this is not an ideal approach since it means that some of the spatial correlation in the horizontal direction is lost; however, since the forecast step incorporates spatial correlation in both directions, one may still obtain satisfactory results. Let C denote the number of columns in the lattice and let $k_{i,c}^{(l)}$ and $\tilde{k}_{i,c}^{(l)}$ for $c = 1, \dots, C$ denote the values in column number c of $k_i^{(l)}$ and $\tilde{k}_i^{(l)}$, respectively. The procedure used for the updating of $k_{i,c}^{(l)}$, $l = 1, \dots, M$ is inspired by the updating procedure used in the ensemble Kalman filter (EnKF), but as the elements of $k_{i,c}^{(l)}$ are binary variables, the updating procedure is based on a first order Markov chain instead of a Gaussian distribution as in EnKF. Thus, the update of $k_{i,c}^{(l)}$, $l = 1, \dots, M$ starts by estimating a (non-stationary) Markov chain for column c . Using a Bayesian model for this estimation, the $k_{i,c}^{(l)}$, $l = 1, \dots, M$ are considered as independent realizations from the assumed Markov chain, and independent uniform priors on the unit interval are adopted for the initial distribution and for each transition probability. The maximum a posteriori estimators are then used to estimate the initial distribution and the transition probabilities. The estimated Markov chain is used as a prior distribution in a new Bayesian model and combined with an assumed likelihood model for the part of d_i related to column c . It is here assumed that d_i contains one component d_i^j for each element k_i^j in k_i , that the components of d_i are conditionally independent given k_i , and that d_i^j depends only on k_i through k_i^j . The likelihood for the part of d_i related to column c can then be expressed as

$$p(d_{i,c}|k_{i,c}) = \prod_{j \text{ in column } c} p(d_i^j|k_i^j). \quad (6)$$

The likelihood model $p(d_i^j|k_i^j)$ is specified by first defining d_i^j to be given by replacing the saturation value s_w in the rock physics model in Eq. (4) by an auxiliary random variable $u_i^j \in [s_{wi}, 1]$, that is,

$$d_i^j = \frac{R_w}{\phi^a (u_i^j)^b}. \quad (7)$$

The distribution of the latent u_i^j should depend on the fluid-facies value k_i^j , and it is assumed that

$$p_0(u_i^j) = p(u_i^j | k_i^j = 0) = \begin{cases} c_0 e^{-\lambda_0 r} & \text{when } u_i^j \in [s_{wi}, r], \\ c_0 e^{-\lambda_0 u_i^j} & \text{when } u_i^j \in (r, 1], \end{cases} \quad (8)$$

and

$$p_1(u_i^j) = p(u_i^j | k_i^j = 1) = \begin{cases} c_1 e^{\lambda_1 u_i^j} & \text{when } u_i^j \in [s_{wi}, r], \\ c_1 e^{\lambda_1 r} & \text{when } u_i^j \in (r, 1], \end{cases} \quad (9)$$

where c_0 and c_1 are normalizing constants, and λ_0 and λ_1 are parameters specifying the level of noise in the resistivity measurements. Small values of λ_0 and λ_1 reflect noisy resistivity data, while higher of λ_0 and λ_1 reflect less noisy resistivity data. Essentially, the auxiliary variable u_i^j can be interpreted as a noisy realisation of the saturation value m_i^j . The logic behind the choice of distributions in Eqs. (8) and (9) is that it should be more likely to generate u_i^j -values in the correct interval and less likely to generate u_i^j -values in the wrong interval; for example, given that $k_i^j = 0$, it should be more likely to generate u_i^j -values in $[s_{wi}, r]$ than in $(r, 1]$. Moreover, the distributions in Eqs. (8) and (9) ensure that it becomes more and more unlikely to generate u_i^j -values the further you step away from the correct interval. The parameters λ_0 and λ_1 determine how fast this decrease in probability occurs and thereby the spread in the u_i^j -values. The spread in the u_i^j -values, in turn, controls the spread, or the level of noise, in the corresponding d_i^j -values obtained from Eq. (6). If λ_0 and λ_1 are relatively large, most of the generated u_i^j -values will be located within the correct intervals, which in turn results in d_i^j -values with relatively little noise. Likewise, if λ_0 and λ_1 are relatively small, many of the generated u_i^j -values will be located outside the correct intervals, which results in a larger spread in the d_i^j -values and hence more noise. The left plot in Fig. 1 shows $p_0(u_i^j)$ and $p_1(u_i^j)$ when $\lambda_0 = 9.8$, $\lambda_1 = 5$ and $r = 0.3$. Combining that d_i^j is a transformation of u_i^j as given in Eq. (7) and that the distribution for u_i^j is as specified in Eqs. (8) and (9), the likelihood model $p(d_i^j | k_i^j)$ can be derived. When $a = b = 2$ it can be shown that

$$\begin{aligned} p(d_i^j | k_i^j = 0) &= \begin{cases} \frac{c_0 \sqrt{R_w}}{2\phi^j} (d_i^j)^{-3/2} \exp \left\{ -\frac{\lambda_0}{\phi^j} \sqrt{\frac{R_w}{d_i^j}} \right\} & \text{when } d_i^j \in \left(\frac{R_w}{(\phi^j)^2}, \frac{R_w}{r^2 (\phi^j)^2} \right], \\ \frac{c_0 \sqrt{R_w}}{2\phi^j} (d_i^j)^{-3/2} \exp \{-\lambda_0 r\} & \text{when } d_i^j \in \left(\frac{R_w}{r^2 (\phi^j)^2}, \frac{R_w}{s_w^2 (\phi^j)^2} \right] \end{cases} \end{aligned} \quad (10)$$

Fig. 1 Likelihood model: Plots of Eqs. (8) and (9) to the left, and Eqs. (10) and (11) to the right, when $\phi = 0.15$, $R_w = 0.1$, $\lambda_0 = 9.8$ and $\lambda_1 = 5$

and

$$\begin{aligned}
 & p(d_i^j | k_i^j = 1) \\
 &= \begin{cases} \frac{c_1 \sqrt{R_w}}{2\phi^j} (d_i^j)^{-3/2} \exp\{\lambda_1 r\} & \text{when } d_i^j \in \left(\frac{R_w}{(\phi^j)^2}, \frac{R_w}{r^2(\phi^j)^2} \right], \\ \frac{c_1 \sqrt{R_w}}{2\phi^j} (d_i^j)^{-3/2} \exp\left\{ \frac{\lambda_1}{\phi^j} \sqrt{\frac{R_w}{d_i^j}} \right\} & \text{when } d_i^j \in \left(\frac{R_w}{r^2(\phi^j)^2}, \frac{R_w}{s_w^2(\phi^j)^2} \right]. \end{cases}
 \end{aligned} \quad (11)$$

The right plot in Fig. 1 shows $p(d_i^j | k_i^j = 0)$ and $p(d_i^j | k_i^j = 1)$ when $\lambda_0 = 9.8$, $\lambda_1 = 5$, $a = b = 2$, $R_w = 0.1$, $\phi^j = 0.15$ and $r = 0.3$.

Combining the estimated prior Markov chain for column c with the likelihood model specified above, the corresponding posterior distribution also becomes a non-stationary Markov chain. The properties of this posterior Markov chain are computationally easy to compute, and in particular the bivariate distributions for every two neighbor nodes in column c can be found. To update the prior ensemble members of column c , a distribution $q(\tilde{k}_{i,c}^{(l)} | k_{i,c}^{(l)})$ which preserves these bivariate distributions is constructed. More specifically, under the assumption that the estimated prior Markov chain for column c is correct, $k_{i,c}^{(l)}$ is updated by simulating from a conditional distribution $q(\tilde{k}_{i,c}^{(l)} | k_{i,c}^{(l)})$ such that the bivariate distribution for every pair of neighbor nodes in $\tilde{k}_{i,c}^{(l)}$ is equal to the corresponding bivariate distribution of the posterior Markov chain for column c .

The chosen distribution $q(\tilde{k}_{i,c}^{(l)} | k_{i,c}^{(l)})$ for updating the prior ensemble members of column c can be expressed as

$$q(\tilde{k}_{i,c}^{(l)} | k_{i,c}^{(l)}) = q_1(\tilde{k}_{i,(1,c)}^{(l)} | k_{i,(1,c)}^{(l)}) \prod_{s=2}^S q_s(\tilde{k}_{i,(s,c)}^{(l)} | \tilde{k}_{i,(s-1,c)}^{(l)}, k_{i,(s,c)}^{(l)}), \quad (12)$$

where S is the number of rows in the lattice used to represent k_i , and $k_{i,(s,c)}^{(l)}$ and $\tilde{k}_{i,(s,c)}^{(l)}$ are the s 'th elements in column c of $k_i^{(l)}$ and $\tilde{k}_i^{(l)}$, respectively. Thus, given

the prior ensemble member $k_{i,c}^{(l)}$, the distribution of the posterior ensemble member $\tilde{k}_{i,c}^{(l)}$ is a Markov chain with initial distribution specified by $q_1(\cdot|\cdot)$ and transition probabilities specified by $q_s(\cdot|\cdot, \cdot)$, $s = 2, \dots, S$. To specify the updating procedure completely it now remains to specify $q_1(\cdot|\cdot)$ and $q_s(\cdot|\cdot, \cdot)$, $s = 2, \dots, S$. These are specified to accomplish two goals. First, considering $k_{i,c}^{(l)}$ as a sample from the estimated prior Markov chain, the marginal bivariate distributions for every pair $(\tilde{k}_{i,(s-1,c)}^{(l)}, \tilde{k}_{i,(s,c)}^{(l)})$, $s = 2, \dots, S$ should be identical to the corresponding bivariate distribution in the posterior Markov chain discussed above. This requirement ensures that the updated fluid-facies values reflect the new resistivity data d_i . However, with only this requirement many possible solutions exist for $q_1(\cdot|\cdot)$ and $q_s(\cdot|\cdot, \cdot)$, $s = 2, \dots, S$, so there is room for formulating another goal. Still considering $k_{i,c}^{(l)}$ as a sample from the estimated prior Markov chain, the second goal for the updating of $k_{i,c}^{(l)}$ is to maximise the expected number of elements in $k_{i,c}^{(l)}$ that remain unchanged; that is, the goal is to maximize

$$\mathbb{E} \left[\sum_{s=1}^S I \left(k_{i,(s,c)}^{(l)} = \tilde{k}_{i,(s,c)}^{(l)} \right) \right], \quad (13)$$

where $I(\mathcal{A})$ equals one if the event $\mathcal{A}$ is true, and zero otherwise, and the expectation is taken with respect to the joint distribution of $k_{i,c}^{(l)}$ and $\tilde{k}_{i,c}^{(l)}$. This requirement makes the updating robust with respect to the a priori Markov chain assumption made for $k_{i,c}^{(l)}$, $l = 1, \dots, M$. If the true distribution of $k_{i,c}^{(l)}$, $l = 1, \dots, M$ is not a Markov chain, many of its non-Markov properties will prevail into $\tilde{k}_{i,c}^{(l)}$, $l = 1, \dots, M$ since it is specified that as few changes as possible should be made to $k_{i,c}^{(i)}$ in the generation of $\tilde{k}_{i,c}^{(l)}$. Numerically, it turns out that in that the maximization of the expression in Eq. (13) under the constraints for the specified bivariate distributions for the pairs $(\tilde{k}_{i,(s-1,c)}^{(l)}, \tilde{k}_{i,(s,c)}^{(l)})$, $s = 2, \dots, S$ can be efficiently computed using a combination of dynamic programming and linear programming. The details of the optimization algorithm are discussed in Loe and Tjelmeland (2020).

2.3.2 The Forecast Step

The forecast step of the proposed approach also involves two parts. First, the ensemble $\{\tilde{k}_i^{(1)}, \dots, \tilde{k}_i^{(M)}\}$ is mapped to a corresponding ensemble $\{\tilde{m}_i^{(1)}, \dots, \tilde{m}_i^{(M)}\}$. Second, the forward model in Eq. (2) is used to generate $\{m_{i+1}^{(1)}, \dots, m_{i+1}^{(M)}\}$ from $\{\tilde{m}_i^{(1)}, \dots, \tilde{m}_i^{(M)}\}$. In the following, the two parts of the forecast step are discussed in more detail.

To generate the saturation field $\tilde{m}_i^{(l)}$ based on a given fluid-facies field $\tilde{k}_i^{(l)}$, the fluid-facies indicators in $\tilde{k}_i^{(l)}$ are first used to define a lattice of distances, δ , from each node j to a node with the opposite value of node j . More precisely, the values in δ are

Fig. 2 To the left: True porosity and log permeability models, with the locations of the production and injection wells marked P and I, respectively. To the right: Assumed porosity and log permeability models

defined sequentially as follows. First, $\delta^j = 0$ is set for all nodes j that has one or more neighbor node j' so that $\tilde{k}_i^{(l),j} \neq \tilde{k}_i^{(l),j'}$. Thereafter, $\delta^j = 1$ is set for all nodes j for which δ^j is still undefined and which has a neighbor node j' with $\delta^{j'} = 0$. Thereafter, $\delta^j = 2$ is set for all nodes j for which δ^j is still undefined and which has a neighbor node j' with $\delta^{j'} = 1$. This process is continued until δ^j is defined for all nodes j . The next step is to scale the δ^j values into the $[0, 1]$ interval. Letting Δ denote the scaled field, the value for node j is defined as

$$\Delta^j = \begin{cases} 0 & \text{if } \delta^j > \delta_{\max} \text{ and } \tilde{k}_i^{(l),j} = 0, \\ \frac{1}{2} - \frac{\delta^j + \frac{1}{2}}{2\delta_{\max} + 1} & \text{if } \delta^j \leq \delta_{\max} \text{ and } \tilde{k}_i^{(l),j} = 0, \\ \frac{1}{2} + \frac{\delta^j + \frac{1}{2}}{2\delta_{\max} + 1} & \text{if } \delta^j \leq \delta_{\max} \text{ and } \tilde{k}_i^{(l),j} = 1, \\ 1 & \text{if } \delta^j > \delta_{\max} \text{ and } \tilde{k}_i^{(l),j} = 1, \end{cases} \quad (14)$$

where $\delta_{\max} > 0$ is a parameter controlling the size of the transition zone from s_{wi} to 1. The larger the value of $\delta_{\max}$, the larger the size of the transition zone. One should choose a value for $\delta_{\max}$ based on what one believes is a realistic transition for the application in consideration. In the numerical examples in Sect. 3, $\delta_{\max} = 8$ is used. The Δ field defines a trend for the $\tilde{m}_i^{(l)}$ values. To add noise to this trend a slightly modified version of the so-called smoothesstep function is first used to transform the

Fig. 3 Reference model: **a** Fluid-facies k_i , **b** water saturation values m_i and **c** log resistivity values d_i at time steps (from left to right) $i = 6, 12, 18$ and 24

Δ^j values over to the real line,

$$\begin{aligned} v^j = & 2\Phi^{-1}(0.99999) \left[6(\Delta^j)^5 - 15(\Delta^j)^4 + 10(\Delta^j)^3 \right] \\ & - \Phi^{-1}(0.99999) + \Phi^{-1}(r), \end{aligned} \quad (15)$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution and $\Phi^{-1}(\cdot)$ is the inverse of this function. The effect of Eq. (15) is that the value v^j is in the left tail of a normal distribution with mean $\Phi^{-1}(r)$ and unit variance when $\Delta^j = 0$, or in the right tail of the same distribution when $\Delta^j = 1$, and with a smooth transition between these two extremes. Moreover, the last term in Eq. (15) ensures that $v^j = \Phi^{-1}(r)$ when $\Delta^j = 0.5$. A noisy version z of v is then defined by setting

$$z^j = \sqrt{1 - \alpha^2}(v^j - \Phi^{-1}(r)) + \alpha \varepsilon^j, \quad (16)$$

Table 1 Experimental settings for the three case studies

	# of measurements	Measurement noise	Likelihood model parameters
Case 1	24 (every 6 months)	Low	$\lambda_0 = 9.8, \lambda_1 = 5.0$
Case 2	4 (every 3 years)	Low	$\lambda_0 = 9.8, \lambda_1 = 5.0$
Case 3	4 (every 3 years)	High	$\lambda_0 = 7.8, \lambda_1 = 2.5$

Fig. 4 Log resistivity observations d_i at time steps (from left to right) $i = 6, 12, 18$ and 24 , in **a** for cases 1 and 2, and in **b** for case 3

where ε is a Gaussian field with zero mean, unit variance and an exponential correlation function, and $\alpha > 0$ is a parameter controlling the noise level. Finally, the saturation field is defined by transforming the z field back to the $(s_{wi}, 1)$ interval,

$$\tilde{m}_t^{(i),j} = s_{wi} + (1 - s_{wi})\Phi\left(z^j + \Phi^{-1}\left(\frac{r - s_{wi}}{1 - s_{wi}}\right)\right). \quad (17)$$

The second part in the forecast step, to generate the ensemble $\{m_{i+1}^{(1)}, \dots, m_{i+1}^{(M)}\}$ from the ensemble $\{\tilde{m}_i^{(1)}, \dots, \tilde{m}_i^{(M)}\}$, is simply done by using the forward model in Eq. (2) for each ensemble member separately; that is, by setting

$$m_{i+1}^{(l)} = g\left(\tilde{m}_i^{(l)}\right) \quad (18)$$

for $l = 1, \dots, M$.

Fig. 5 Results from case 1: Images of estimated marginal probabilities $p(k_i^j = 1 | d_{1:i})$ at time steps (from left to right) $i = 6$, $i = 12$, $i = 18$, and $i = 24$ using three different ensemble sizes

3 Application

The proposed inversion method is tested using a synthetic reservoir model. The model consists of a two-dimensional reservoir, $25 \text{ m} \times 25 \text{ m}$, with constant thickness and four main channels with high-porosity rocks surrounded by low-porosity rocks; see Fig. 2. The fluid system includes two fluid phases: oil and water. Therefore, two fluid-facies are defined: oil-saturated rocks (corresponding to the value 0) and water-saturated rocks (corresponding to the value 1). The discretized reservoir is defined on a 128×128 grid, and the well configuration includes four injectors and six producers as shown in Fig. 2. The oil production mechanism is based on water injection simulated using the MATLAB Reservoir Simulation Toolbox (Lie 2019) for a time period of 12 years. The 12 year time period is discretized into 24 equidistant time points $t_1, \dots, t_{24}$ such that each step of the simulation involves propagating the system six months forward in time. During the simulation, injection rates are kept constant at the injector locations, and bottom hole pressure is kept constant at the producer locations. Initially, the entire reservoir is filled with hydrocarbon with the irreducible water saturation value of

Fig. 6 Results from case 2: Images of estimated marginal probabilities $p(k_i^j = 1 | d_{1:i})$ at time steps (from left to right) $i = 6$, $i = 12$, $i = 18$, and $i = 24$ using three different ensemble sizes

$s_{wi} = 0.2$. Based on a preliminary analysis, the threshold parameter r in Eq. (3) is set to $r = 0.3$. Figure 3a, b show the fluid-facies k_i and saturation values m_i , respectively, in the reservoir at the time steps $i = 6$, $i = 12$, $i = 18$ and $i = 24$; that is, after $t_6 = 3$, $t_{12} = 6$, $t_{18} = 9$ and $t_{24} = 12$ years of the simulation. Figure 3c shows corresponding reference resistivity values (in log-scale); that is, the resistivity values one obtains by inserting the true water saturation values into Archie's law in Eq. (4). Pretending that the fluid-facies and saturation values used to generate the plots in Fig. 3 are unknown, the goal of the simulation experiment is to estimate the fluid-facies field at each time step based on noisy resistivity data. In this example, the resistivity data d_i at time t_i includes a two-dimensional map of resistivity measurements; specifically, the dimensionality N_d of d_i is equal to the dimensionality N_k of k_i (and m_i) so that an observation d_i^j is available for each variable k_i^j of k_i . Since a 128×128 grid is considered, with a fluid-facies variable k_i^j in every cell j , the dimensions N_k and N_d are $N_k = N_d = 128 \cdot 128 = 16384$.

The porosity and permeability models shown in Fig. 2 (left plots) are the true porosity and permeability models of the reservoir. These were the values used to generate the reference model shown in Fig. 3. Since porosity and permeability are

Fig. 7 Results from case 3: Images of estimated marginal probabilities $p(k_i^j = 1 | d_{1:i})$ at time steps (from left to right) $i = 6, i = 12, i = 18$, and $i = 24$ using three different ensemble sizes

generally not known, a reservoir model of assumed porosity and permeability models is built to mimic the resolution of a reservoir model estimated from pre-production seismic data. The assumed porosity and permeability models are shown in Fig. 2 (right plots).

Three case studies are presented, differing in the frequency with which the resistivity measurements are collected and the amount of noise in the measurements; see Table 1. The first case, referred to as case 1, represents an idealized situation where resistivity measurements are recorded frequently and the degree of noise in the data is small. Specifically, observations are assumed to be recorded every six months, or at every time step $i = 1, \dots, N_t$. Hence the set $\mathbb{T}$ introduced in Sect. 2.1 is $\mathbb{T} = \{1, 2, \dots, 24\}$. Figure 4a shows the simulated resistivity measurements d_i (in log-scale) at the four time steps $i = 6, 12, 18$ and 24 for case 1. The resistivity data were generated with the likelihood model specified in Sect. 2, using the true fluid-facies shown in Fig. 3a and the assumed porosity model shown to the right in Fig. 2, and with the parameters λ_0 and λ_1 set to $\lambda_0 = 9.8$ and $\lambda_1 = 5$. These values for λ_0 and λ_1 represent optimistic noise conditions. In the second case, referred to as case 2, the same data as in case 1 are considered, but observations are assumed to be acquired only every three years of the simulation period; that is, an observation is recorded after 3, 6, 9, and 12 years, or at

Fig. 8 Results from case 1: Empirical standard deviations of $\{\hat{p}_{A,1}^M(k_i^j = 1|d_{1:i})\}_{A=1}^{10}$ at time steps (from left to right) $i = 6, i = 12, i = 18$ and $i = 24$ using three different ensemble sizes

the time steps $i = 6, 12, 18$ and 24 . Hence the set $\mathbb{T}$ is in this case $\mathbb{T} = \{6, 12, 18, 24\}$, and the likelihood parameters λ_0 and λ_1 are the same as in case 1. In the third case, referred to as case 3, observations, as in case 2, are acquired only every 3 years, but a different set of data with a much higher level of noise is considered. Hence this case represents the most realistic of the three cases. Figure 4b shows the simulated resistivity measurements for case 3. Similarly to the resistivity data for cases 1 and 2, the resistivity measurements for case 3 were generated using the likelihood model specified in Sect. 2, but with the parameters λ_0 and λ_1 set to $\lambda_0 = 7.8$ and $\lambda_1 = 2.5$. These parameter values represent realistic noise conditions.

For all three case studies the proposed inversion method is tested using three different ensemble sizes: $M = 20$, $M = 100$ and $M = 500$. The parameters $\delta_{\max}$ and α in Eqs. (14) and (16) are set to $\delta_{\max} = 8$ and $\alpha = 0.2$. The ensembles were initialised by first introducing an initial field of fluid facies k_0 for which it is assumed that $k_0^j = 0$ for every cell j in the reservoir, and thereafter generate each $m_1^{(l)}$ from k_0 as discussed in Sect. 2.3.2. To evaluate the results, an estimate $\hat{p}(k_i^j = 1|d_{1:i})$ for each marginal probability $p(k_i^j = 1|d_{1:i})$, $j = 1, \dots, N_k$, is computed; specifically, each $p(k_i^j = 1|d_{1:i})$ is estimated as the fraction of updated k_i^j -samples equal to one.

Fig. 9 Results from case 2: Empirical standard deviations of $\{\hat{p}_{A,2}^M(k_i^j = 1|d_{1:i})\}_{A=1}^{10}$ at time steps (from left to right) $i = 6, i = 12, i = 18$ and $i = 24$ using three different ensemble sizes

Mathematically, that is

$$\hat{p}(k_i^j = 1|d_{1:i}) = \frac{1}{M} \sum_{l=1}^M \tilde{k}_i^{j,(l)}. \quad (19)$$

Figures 5, 6 and 7 present images of these estimated marginal probabilities for cases 1, 2 and 3, respectively. Comparison between the plots in Figs. 5 to 7 and the reference k_i -values in Fig. 3a shows that the proposed inversion method to a large extent has captured the true underlying binary field of fluid-facies in all three cases, even when using the small ensemble size $M = 20$. As expected, larger ensemble sizes provide more accurate results, but the results obtained with $M = 20$ are also satisfactory. However, a few short vertical lines have a tendency to appear in some of the figures, especially in the results from case 3 in Fig. 7. This is an inevitable spatial effect due to the columns of the grid being updated independently of each other in the conditioning step of the inversion method. It is reasonable that the effect is more apparent in the results from case 3 than in the results from cases 1 and 2, since the quality of the geophysical data in case 3 is lower (i.e., noisier).

Fig. 10 Results from case 3: Empirical standard deviations of $\{\hat{p}_{A,3}^M(k_i^j = 1 | d_{1:i})\}_{A=1}^{10}$ at time steps (from left to right) $i = 6, i = 12, i = 18$ and $i = 24$ using three different ensemble sizes

To evaluate how sensitive the proposed inversion method is on the ensemble size M , ten independent runs are performed for each of the ensemble sizes $M = 20$, $M = 100$ and $M = 500$ in each of the three cases. Thereafter the marginal probabilities $p(k_i^j = 1 | d_{1:i})$, $j = 1, \dots, N_k$ are estimated cf. Eq. (19). Thereby, in each of the three cases, ten different estimates $\hat{p}(k_i^j = 1 | d_{1:i})$ of $p(k_i^j = 1 | d_{1:i})$ are obtained for each ensemble size. In the following, let $\hat{p}_{A,B}^M(k_i^j | d_{1:i})$ denote the estimate of $p(k_i^j | d_{1:i})$ obtained in run number $A = 1, \dots, 10$ of case number $B = 1, 2, 3$ when using ensemble size $M = 20, 100, 500$. To evaluate the accuracy of the ten estimates $\hat{p}_{1,B}^M(k_i^j | d_{1:i}), \dots, \hat{p}_{10,B}^M(k_i^j | d_{1:i})$ of $p(k_i^j = 1 | d_{1:i})$ obtained in the ten runs of case B when using ensemble size M , the standard deviation of these ten estimates is computed. Results are shown in Figs. 8, 9 and 10 for cases 1, 2 and 3, respectively. Similarly to the other results presented above, the results obtained with the higher ensemble sizes $M = 100$ and $M = 500$ are overall smoother and less noisy than those obtained with the rather small ensemble size $M = 20$. A general trend, however, for all three ensemble sizes and all three cases, is that the standard deviations tend to be higher near the boundary of the fluid front, which is reasonable, since this is the most uncertain area where changes occur. Moreover, the results from case 3 in Fig. 10 are considerably

noisier than the results from cases 2 and 3 in Figs. 8 and 9. This means that the case 3 results tend to vary more from one run to another. Again, this is reasonable, since the resistivity measurements in case 3 are more uncertain.

4 Conclusions

A novel method for monitoring and updating the evolution of fluid-facies from time-lapse geophysical properties in a two-phase flow problem has been presented. The inversion method is based on an ensemble filtering method where the updating of the prior ensemble at each time step is performed using a particular updating method for binary vectors. The main novelty of the work is the extension of ensemble-based methods to mixed discrete-continuous problems to update the spatial distribution of fluid-facies. In the proposed application, the geophysical dataset includes time-lapse resistivity values that are assumed to have been estimated from CSEM data through a preliminary inversion process. The proposed method is tested in a synthetic example with a two-dimensional reservoir model. The results from this synthetic example are accurate and support the validation of the proposed methodology. In real data applications, the accuracy of the results depends on the quality of the data in terms of resolution and signal-to-noise ratio, and also on the accuracy of the fluid flow simulator. The main limitation of this work is that uncertainty in the estimation of porosity and permeability are not taken into account. Future research directions aim to extend the proposed method so that porosity and permeability are also treated as random variables and so that the geophysical dataset includes measured data such as electromagnetic amplitude and phase.

Funding Open Access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aki K, Richards PG (1980) Quantitative seismology: theory and methods. W.H. Freeman and Co, New York
- Aziz K (1979) Petroleum reservoir simulation, vol 476. Applied Science Publishers, London
- Bear J (2013) Dynamics of fluids in porous media. Courier Corporation, Chelmsford
- Bergmann P, Schmidt-Hattenberger C, Labitzke T, Wagner FM, Just A, Flechsig C, Rippe D (2017) Fluid injection monitoring using electrical resistivity tomography—five years of CO₂ injection at Ketzin, Germany. *Geophys Prospect* 65(3):859–875
- Berre I, Lien M, Mannseth T (2011) Identification of three-dimensional electric conductivity changes from time-lapse electromagnetic observations. *J Comput Phys* 230(10):3915–3928

- Bhuyian AH, Landrø M, Johansen SE (2012) 3D CSEM modeling and time-lapse sensitivity analysis for subsurface CO₂ storage. *Geophysics* 77(5):E343–E355
- Buland A, Kolbjørnsen O (2012) Bayesian inversion of CSEM and magnetotelluric data. *Geophysics* 77(1):E33–E42
- Claes N, Paige GB, Grana D, Parsekian AD (2020) Parameterization of a hydrologic model with geophysical data to simulate observed subsurface return flow paths. *Vadose Zone J* 19(1):e20024
- Commer M, Doetsch J, Dafflon B, Wu Y, Daley TM, Hubbard SS (2016) Time-lapse 3-D electrical resistance tomography inversion for crosswell monitoring of dissolved and supercritical CO₂ flow at two field sites: Escatawpa and Cranfield, Mississippi, USA. *Int J Greenhouse Gas Control* 49:297–311
- Constable S (2010) Ten years of marine CSEM for hydrocarbon exploration. *Geophysics* 75(5):75A67–75A81
- Doucet A, de Freitas N, Gordon N (2001) *Sequential Monte Carlo methods in practice*. Springer, New York
- Doyen P (2007) *Seismic reservoir characterization: an earth modelling perspective*, vol 2. EAGE Publications, Houten
- Doyen P, Psaila D, Astratti D, Kvamme L, Al Najjar N (2000) Saturation mapping from 4-D seismic data in the Statfjord field. Paper OTC 12100 presented at the Offshore Technology Conference, Houston, Texas, USA, 1–4 May
- Evensen G (2009) *Data assimilation: the ensemble Kalman filter*. Springer, Berlin
- Flinchum BA, Holbrook WS, Grana D, Parsekian AD, Carr BJ, Hayes JL, Jiao J (2018) Estimating the water holding capacity of the critical zone using near-surface geophysics. *Hydrol Process* 32(22):3308–3326
- Gasperikova E, Hoversten GM (2006) A feasibility study of nonseismic geophysical methods for monitoring geologic CO₂ sequestration. *Lead Edge* 25(10):1282–1288
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, Berlin
- Kotikian M, Parsekian AD, Paige G, Carey A (2019) Observing heterogeneous unsaturated flow at the hillslope scale using time-lapse electrical resistivity tomography. *Vadose Zone J* 18(1):1–16
- Leeuwenburgh O, Arts R (2014) Distance parameterization for efficient seismic history matching with the ensemble Kalman filter. *Comput Geosci* 18(3–4):535–548
- Lie KA (2019) *An introduction to reservoir simulation using MATLAB/GNU Octave*. Cambridge University Press, Cambridge
- Lien M, Mannseth T (2008) Sensitivity study of marine CSEM data for reservoir production monitoring. *Geophysics* 73(4):F151–F163
- Lien M, Mannseth T, Agersborg R (2014) Assimilation of time-lapse CSEM data for fluid flow monitoring. In: 76th EAGE conference and exhibition-workshops, 1, European Association of Geoscientists & Engineers, pp 1–3
- Loe MK, Tjelmeland H (2020) Ensemble updating of binary state vectors by maximising the expected number of unchanged components. *Scand J Stat*. <https://doi.org/10.1111/sjos.12483>
- MacGregor L (2012) Integrating seismic, CSEM and well log data for reservoir characterization. *Lead Edge* 31(3):258–265
- Martinez WL, Martinez AR (2015) *Computational statistics handbook with MATLAB*. Chapman and Hall/CRC, Cambridge
- Mavko G, Mukerji T, Dvorkin J (2009) *The rock physics handbook*, 2nd edn. Cambridge University Press, Cambridge
- Oliver DS, Reynolds AC, Liu N (2008) *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press, Cambridge
- Orange A, Key K, Constable S (2009) The feasibility of reservoir monitoring using time-lapse marine CSEM. *Geophysics* 74(2):F21–F29
- Shahin A, Key K, Stoffa P, Tatham R (2012) Petro-electric modeling for CSEM reservoir characterization and monitoring. *Geophysics* 77(1):E9–E20
- Trani M, Arts R, Leeuwenburgh O et al (2012) Seismic history matching of fluid fronts using the ensemble Kalman filter. *SPE J* 18(01):159–171
- Tveit S, Bakr SA, Lien M, Mannseth T (2015) Ensemble-based Bayesian inversion of CSEM data for subsurface structure identification. *Geophys J Int* 201(3):1849–1867
- Tveit S, Mannseth T, Park J, Sauvin G, Agersborg R (2020) Combining CSEM or gravity inversion with seismic AVO inversion, with application to monitoring of large-scale CO₂ injection. *Comput Geosci* 24:1201–1220

- Viterbi A (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans Inf Theory* 13(2):260–269
- Weitemeyer K, Constable S, Key K, Behrens J (2006) First results from a marine controlled-source electromagnetic survey to detect gas hydrates offshore Oregon. *Geophys Res Lett* 33(3):L03304
- Zhang Y, Leeuwenburgh O (2017) Image-oriented distance parameterization for ensemble-based seismic history matching. *Comput Geosci* 21(4):713–731