Seismic monitoring of CO_2 plume dynamics using ensemble Kalman filtering

Grant Bruer, Abhinav Prakash Gahlot, Edmond Chow, Felix Herrmann

***Abstract*—Monitoring carbon dioxide (CO_2) injected and stored in subsurface reservoirs is critical for avoiding failure scenarios and enables real-time optimization of CO_2 injection rates. Bayesian sequential data assimilation (DA) is a statistical method for combining information over time from multiple sources to estimate a hidden state, such as the spread of the subsurface CO_2 plume. Existing literature in the seismic-CO_2 monitoring domain uses small physical systems with unscalable DA algorithms, ignores the CO_2 flow dynamics, or simulates seismic data without the wave equation or with unrealistic survey designs. We improve upon existing DA literature in this domain by applying the scalable ensemble Kalman filter (EnKF) DA algorithm to a high-dimensional CO_2 reservoir using two-phase flow dynamics and time-lapse full waveform seismic data with a realistic surface-seismic survey design. We show this DA method is more accurate compared to using either the seismic data or the fluid physics alone. Furthermore, we show the stability of this method by testing a range of values for the EnKF hyperparameters and give guidance on their selection for seismic CO_2 reservoir monitoring.**

***Index Terms*—carbon capture and storage, Kalman filters, nonlinear dynamical systems, seismic measurements**

I. INTRODUCTION

CARBON capture and storage is a recently sought-after technology involving capturing carbon dioxide (CO_2) for long-term storage. Geologic storage sites such as saline aquifers and depleted oil fields can store large amounts of CO_2 due to their size and the naturally high pressure deep underground [1]–[3]. Monitoring the state of such CO_2 reservoirs is critical to avoid failure scenarios, such as leaks or man-made seismic activity (earthquakes) [4], [5]. Monitoring also enables real-time optimization of reservoir performance using computer simulations. However, due to the limited accessibility of underground reservoirs, traditional measurements of the reservoir state are confined to the surface and a relatively small number of boreholes. This limits the predictive power of simulations that are based on traditional measurements. Since injected CO_2 displaces brine in saline aquifers, it affects the gravitational properties, electromagnetic properties [6], and seismic wave propagation in the rock. Thus, indirect measurements reveal information at otherwise inaccessible points. Non-intrusive methods for determining the spread of the CO_2 in saline aquifers include electromagnetic, gravitational, and seismic measurements. Seismic methods have been shown to provide higher resolution estimates than electromagnetic or gravitational methods [7]–[11]. Seismic data have been used to estimate the spread of the injected CO_2, such as for the Sleipner field offshore Norway [12] and the Ketzin pilot site in Germany [13]. We work with active-source surface seismic, but passive seismic monitoring has also been investigated [14].

The contribution of this research is combining fluid-flow simulations with time-lapse full waveform seismic data to predict CO_2 saturations in underground reservoirs as CO_2 is injected. The literature on seismic-CO_2 monitoring lacks solutions that can be applied to large, complex systems. Existing literature in this domain makes one or more limiting choices: (1) uses small physical systems [15], [16] with unscalable data assimilation (DA) algorithms [17]; (2) does not use the CO_2 flow dynamics [17]–[20]; or (3) simulates seismic data without the wave equation [15], [20], [21] or (4) with unrealistic survey designs [16]. While most seismic-CO_2 monitoring approaches use time-lapse seismic data by itself, the combination of physical simulation and full waveform seismic data has recently begun to be studied [15], [18]. Challenges that hinder progress in this area include the cost of high-resolution physical simulations and the cost of storing the covariance matrix necessary for properly combining information from simulation and from system measurements. These problems have been tackled in other domains, in particular weather forecasting, using algorithms such as ensemble Kalman filters, particle filters, and variational optimization. These methods are all forms of Bayesian data assimilation.

Bayesian sequential data assimilation aims to estimate a state vector over time with quantified uncertainty given a sequence of noisy time-lapse measurements. For our purposes, the state is the CO_2 saturation throughout the reservoir, and the measurements come from seismic waveform data. The state follows known two-phase flow dynamics with uncertainty coming from the unknown subsurface permeability field, which describes the rate at which the CO_2 and brine flow through the porous rock. In general, there are many Bayesian sequential data assimilation algorithms for appropriately combining uncertain simulation predictions with noisy time-lapse measurements. We demonstrate the predictive efficacy of the ensemble Kalman filter technique on CO_2 reservoirs with multiple time-lapse surveyed seismic data.

The main issue with data assimilation for CO_2 reservoirs is that the sheer size of the problem makes the simulation and

---

Manuscript received 9 September 2024. This work was supported by the National Science Foundation under Grant 2203821.

(Corresponding author: Edmond Chow.)

Grant Bruer and Edmond Chow are with the School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA.

Abhinav Prakash Gahlot is with the School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332 USA.

Felix Herrmann is with the School of Computational Science and Engineering, the School of Earth and Atmospheric Sciences, and the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA.

Software will be made available online prior to publication.

the data assimilation computations expensive. CO_2 reservoirs are represented in simulation by discretizing the reservoir into a three-dimensional (3D) grid and associating at the very least one CO_2 saturation value and one pressure value at each point in the grid. The system is governed by nonlinear dynamics, and thus demands a relatively high-resolution grid in order to be acceptably accurate. For example, a "small" reservoir can have a volume of 10^{10} cubic meters (i.e., a few dozen cubic kilometers) and be divided into cells on the order of $10^2 \sim 10^3$ cubic meters. The state vector (the CO_2 saturation) has 10^7 degrees of freedom. Seismic measurements are collected by recording seismic waves with a resolution of about 10^{-3} seconds (i.e., a few milliseconds) for a duration on the order of a second. A "small" seismic survey records these waves for 10–100 sources for each of 100–1000 receivers, making the total length of the observation vector approximately 10^8. While this is not prohibitively large for a single vector (a few hundred megabytes), data assimilation requires storing data many times that size, which can quickly cause issues with computation. In our experiments, we model a two-dimensional (2D) reservoir with volume $\sim 10^8$ cubic meters, so the state vector is of size $\sim 10^5$, and the observation vector is of size $\sim 10^6$. For classical data assimilation techniques that require storing the square of the data size, this 2D system requires almost one hundred gigabytes of computer memory. The three-dimensional 10^{10} cubic meter system requires more than one terabyte of memory.

A classical data assimilation technique is the Kalman filter (KF), which computes the optimal state for a given set of measurements assuming linear transition and measurement functions with Gaussian noise. With N_x as the length of the system state vector and N_y as the length of the measurement vector, the standard KF computes the Gaussian distribution of the state vector $\mathbf{x}$ in $O(N_x^3 + N_y^3)$ time and $O(N_x^2 + N_y^2)$ storage, excluding the cost of the simulations [22]. Other data assimilation research on CO_2 reservoirs has handled the challenge of large covariance matrix sizes from seismic waveform observations. For example, one can use a hierarchical sparse matrix structure [18], [23] or use ensemble-based methods [17] with the neighborhood algorithm [24]. However, these past data assimilation approaches with full waveform measurements of CO_2 reservoirs have ignored the CO_2 dynamics.

CO_2 reservoirs, like many real-world systems, are governed by nonlinear dynamics, thereby invalidating the Gaussianity assumption of the KF. The KF can be applied to nonlinear systems (in which case it is called the extended KF) successfully if the time between measurements is so small that the transition is approximately linear. But the high cost of seismic surveys leads to a long time between reservoir measurements, which breaks the linear approximation. Furthermore, seismic observation is a nonlinear operator, further breaking the assumptions of the KF.

Weather forecasting is the well-researched domain that routinely handles nonlinearity, noise, and large data sizes for data assimilation. Major weather forecasting centers rely on various data assimilation algorithms, including variational methods such as four-dimensional variational assimilation and ensemble methods such as the ensemble Kalman filter (EnKF).

Variational methods, on the one hand, minimize the observation error regularized with the distance from the predicted state. These methods can have a high implementation cost because they require adjoint operators for computing gradients. Additionally, for large systems, storing the state covariance matrix becomes infeasible. Bannister [25], [26] describes how weather forecasting models reduce the matrix storage size by diagonalizing the covariance matrix using a physics-based transform. However, this transform becomes less accurate as the model resolutions are increased, and as a result, many weather systems have switched to an ensemble-based estimate of the covariance. Furthermore, variational methods do not inherently estimate the posterior uncertainty, so they must use expensive Hessian-based approximations or be combined with ensemble methods to obtain a covariance estimate.

Ensemble methods, on the other hand, are comparatively much easier to implement because they do not require adjoint operators. Given past samples of the system state representing the uncertainty at an earlier time, the basic EnKF predicts the current state samples using known transition dynamics and then uses the sample covariance to update the samples based on observations. Although this algorithm is optimal only for linear models, it does not rely on linearizing the transition or observation models and thus handles nonlinearities much better than the standard KF. Furthermore, using a low-rank form for the covariance simplifies the update cost to $O(N_x N_e^2 + N_y N_e^3)$ time and $O(N_x N_e + N_y N_e)$ storage for an ensemble of size N_e [27]. The EnKF has been shown to scale well with N_x both in accuracy and computational cost for many data assimilation problems [27]–[29]. Several variants of the EnKF exist, such as those described by Evensen [30], but we leave the application of other variants to future work.

In seismic measurement domains, several variations on the KF of varying complexity have been used to address nonlinearity, noise, and large data, but none yet have addressed the combination of large CO_2 reservoirs with CO_2 dynamics and full waveform seismic measurements. Eikrem *et al.* [19] address the issue of observation nonlinearity by using the iterated extended KF to estimate wave velocities from the full waveform data. This method linearizes the model at multiple points, thereby better approximating the nonlinearity, but they note both that it requires an adjoint observation model and that the matrix storage causes difficulty for scaling to large data. Li *et al.* [31] address the issue of scalability with a low-rank KF to estimate CO_2 reservoir state. They show its performance on seismic travel time measurements or 115 wells on a small 0.2 square kilometer domain, but this method has not been shown on high-dimensional seismic data. Alfonzo and Oliver [20] apply the ensemble transform KF to estimate porosity based on acoustic impedance inverted from seismic measurements, but this does not estimate the dynamic state of the reservoir. Grana *et al.* [15], Ma *et al.* [21], and Guzman *et al.* [32] apply the EnKF to estimate CO_2 reservoir state and geological parameters using CO_2 dynamics but not full waveform observations. A scalable application of the KF using full waveform data and CO_2 dynamics has yet to be achieved.

In the machine-learning (ML) domain, research that uses

CO_2 dynamics has focused on building surrogate flow models for use in Bayesian inversion algorithms. These algorithms estimate static properties, such as porosity and permeability or some other parameter for the surrogate model [33]–[40]. The surrogate models are much faster than solving the partial differential equations (PDEs) governing flow, thus allowing the use of inversion algorithms that are normally prohibitively expensive but more accurate than the KF. Such methods include rejection sampling [33]; randomized maximum likelihood [34], [35]; ensemble smoothing with multiple data assimilation [35]–[37]; and variational methods [38]–[40]. While one reference uses full-waveform seismic data [38], most do not, instead relying on less informative data such as well data [34], [35], [40] or surface deformation data [33], [37]. Those using seismic data approximate the seismic image directly [36], [39], ignoring the more complex seismic wave physics. We are also investigating ML-based DA research using full-waveform seismic data and CO_2 fluid dynamics; see [41]–[43] for preliminary results. While ML-based methods show promise, they have not reached the level of hyperparameter robustness, scalability, and interpretability that the EnKF has has shown in the field of weather forecasting. In particular, ML-based methods tend to rely on arbitrary hyperparameter tuning, making them difficult to apply in practice. The EnKF relies on the transition and observation physics as well as estimates of observation noise, all of which are typically not arbitrary.

The contribution of this research is showing that the EnKF with full waveform data and CO_2 dynamics estimates the CO_2 reservoir state with higher accuracy than two non-data-assimilation baselines without needing a prohibitively large ensemble. We improve upon existing literature by using a relatively large physical system with full waveform seismic data, CO_2 dynamics, and a scalable DA algorithm. The paper is structured as follows: Section II describes the mathematical background for DA filters, the CO_2 dynamics, and seismic measurements; Sections III-A and III-B explain how we apply the fluid dynamics and seismic operators in the EnKF; Section III-C gives a high-level overview of the workflow for this method; Section III-D defines the algorithms we compare with the EnKF; Section III-E describes the EnKF hyperparameters that we examine; Section IV introduces our high-dimensional synthetic experiment; Section V presents the results of our tests with the EnKF; and Section VI concludes.

II. BACKGROUND

A. Filters

Sequential data assimilation is the process of estimating a hidden system state using available observations as well as knowledge of the system's dynamics. The simulated dynamics are represented with the Markovian transition function f, which updates the state $\mathbf{x}^n$ at time step n to the next time step. System measurement is represented with the observation function h. Both functions may depend on noise $\boldsymbol{\eta}$ sampled at each time step independent of $\mathbf{x}$, which represents any stochasticity of the system or unknown system dynamics. The transition and observation operators may depend on time, but we will typically omit the time parameters as they are clear in context. We also leave off the noise parameter as necessary when indicating a non-noisy evaluation. Mathematically, the operators are written as

$$\mathbf{x}^{n+1} = f(\mathbf{x}^n, \boldsymbol{\eta}_f; t_n, t_{n+1}) \text{ and} \tag{1}$$

$$\mathbf{y}^{n+1} = h(\mathbf{x}^{n+1}, \boldsymbol{\eta}_h; t_{n+1}), \tag{2}$$

for time t and noise samples $\boldsymbol{\eta}_f$ and $\boldsymbol{\eta}_h$ for f and h.

Using a sequence of observation data vectors $\{\mathbf{y}^1, \mathbf{y}^2, \dots \mathbf{y}^n\}$, data assimilation algorithms estimate the hidden states $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^n\}$. Given an observation with a known noise model, the likelihood $p(\mathbf{y}|\mathbf{x})$ could be maximized to find the best estimate for $\mathbf{x}$. However, we are typically unable to observe the full hidden state, making this an ill-posed inverse problem requiring regularization. Regularization can be expressed as a prior distribution in Bayes' formula, which states

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{3}$$

The prior $p(\mathbf{x})$ can be chosen based on physics or smoothness knowledge. In sequential data assimilation, it is conditioned on previous data and the prediction of a simulation. Let $\mathbf{y}^{1:n}$ denote all observations up to time step n. Then the conditional distribution $p(\mathbf{x}^n|\mathbf{y}^{1:n-1})$ represents the probability of a certain value of $\mathbf{x}$ at time t_n given all past observations. Sequential data assimilation starts from a given prior distribution $p(\mathbf{x}^0)$, which represents knowledge of the state at the initial time. We denote this as $p(\mathbf{x}^0|\mathbf{y}^{1:0})$ to have the same form as other conditional distributions. The prior is typically chosen by trying to be uninformative, trying to be easy to work with computationally, or using prior knowledge from domain experts [44]. The distributions at later times can be expressed based on the transition and observation operators applied to past distributions. Given a past distribution $p(\mathbf{x}^{n-1}|\mathbf{y}^{1:n-1})$, the prior (or predictive or forecast) distribution $p(\mathbf{x}^n|\mathbf{y}^{1:n-1})$ conditioned on all data from previous time steps is given by

$$p(\mathbf{x}^n|\mathbf{y}^{1:n-1}) = \int p(\mathbf{x}^n|\mathbf{x}^{n-1})p(\mathbf{x}^{n-1}|\mathbf{y}^{1:n-1})\, d\mathbf{x}^{n-1}, \tag{4}$$

where $p(\mathbf{x}^n|\mathbf{x}^{n-1})$ is the probability density determined by the transition function from Equation (1). This integral is computationally difficult for high-dimensional probability distributions, which leads to simplifying assumptions by the Kalman filter described in the next section. With $p(\mathbf{x}^n|\mathbf{y}^{1:n-1})$ as the prior defined in Equation (4), Bayes' formula for sequential data assimilation becomes

$$p(\mathbf{x}^n|\mathbf{y}^{1:n}) = \frac{p(\mathbf{y}^n|\mathbf{x}^n)p(\mathbf{x}^n|\mathbf{y}^{1:n-1})}{p(\mathbf{y}^n|\mathbf{y}^{1:n-1})}, \tag{5}$$

with $\mathbf{y}^{1:0}$ being an empty set and $p(\mathbf{x}^0|\mathbf{y}^{1:0}) \equiv p(\mathbf{x}^0)$ fixed to represent the initial knowledge of the system before any measurements.

Since we always incorporate available measurements, we simplify the notation by defining the distributions at time step n to be implicitly conditioned on measurements from the previous time steps. We define $p(\mathbf{x}^n) \equiv p(\mathbf{x}^n|\mathbf{y}^{1:n-1})$ and $p(\mathbf{x}^n|\mathbf{y}^n) \equiv p(\mathbf{x}^n|\mathbf{y}^{1:n})$. That also allows us to drop the time

Algorithm 1: Generic data assimilation loop

```
Input: prior, data $\mathbf{y}^n$ at time $t_n$
posterior ← prior
$t_{n-1} \leftarrow 0$
for ($\mathbf{y}^n$, $t_n$) in observations do
    // Equation (6)
    prior ← Predict(posterior, $t_{n-1}$, $t_n$)
    // Equation (7)
    posterior ← Update(prior, $\mathbf{y}^n$)
    $t_{n-1} \leftarrow t_n$
```

step superscript when looking at data at a single time step. The form of data assimilation in Equations (4) and (5) can be written as two repeated phases, shown both in Algorithm 1 and here mathematically as

$$\text{predict:} \quad p(\mathbf{x}^n) = \int p(\mathbf{x}^n|\mathbf{x}^{n-1})p(\mathbf{x}^{n-1}|\mathbf{y}^{n-1})\, d\mathbf{x}^{n-1}, \tag{6}$$

$$\text{update:} \quad p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{7}$$

with implicit n superscripts in the update phase. The predict phase advances the distribution $p(\mathbf{x}^{n-1}|\mathbf{y}^{n-1})$ to $p(\mathbf{x}^n)$ using the transition function from Equation (1). The update phase updates the prior $p(\mathbf{x}^n)$ to the posterior $p(\mathbf{x}^n|\mathbf{y}^n)$ using the likelihood $p(\mathbf{y}^n|\mathbf{x}^n)$, which is based on the observation function from Equation (2). In Section II-B, we explain this update phase for the Kalman filter and the EnKF, with common EnKF difficulties in Section II-C. In Sections II-D and II-E, we introduce the transition and observation operators we use with the EnKF.

B. Kalman filters

The standard Kalman filter (KF) [45] is a sequential application of Bayes' rule that assumes each distribution is Gaussian in the state $\mathbf{x}$. This requires a linear transitions and a linear observation $\mathbf{y} = \mathbf{H}\mathbf{x}$. For linear operators with Gaussian noise, an input Gaussian distribution $p(\mathbf{x}^{n-1}|\mathbf{y}^{n-1})$ always transforms to a Gaussian prior $p(\mathbf{x}^n)$ and posterior $p(\mathbf{x}^n|\mathbf{y}^n)$. That reduces Equations (6) and (7) to linear equations for the updated Gaussian mean and covariance. The KF is a linear update to the mean and covariance based on the observation covariance and the state-observation cross-covariance. Let the forecast Gaussian $p(\mathbf{x})$ be a Gaussian distribution with mean $\boldsymbol{\mu}_f$ and covariance $\mathbf{B}_f$. Then the analysis (or posterior) distribution is $p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_a, \mathbf{B}_a)$, where the moments can be calculated using the Kalman update equations,

$$\boldsymbol{\mu}_a = \boldsymbol{\mu}_f + \mathbf{K}(\mathbf{y} - h(\boldsymbol{\mu}_f)) \text{ and} \tag{8}$$

$$B_a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}_f. \tag{9}$$

The Kalman gain matrix $\mathbf{K}$ can be expressed in terms of the covariance of the predicted $\mathbf{x}_f \sim p(\mathbf{x}_f)$ and $\mathbf{y}_f = h(\mathbf{x}_f)$. Recall the definition of the covariance $\operatorname{cov}(\cdot)$ and the cross-covariance $\operatorname{cov}(\cdot,\cdot)$ as the expectation of the outer product of the deviations from the mean,

$$\operatorname{cov}(\mathbf{x}_f, \mathbf{y}_f) = \mathop{\mathbb{E}}_{(\mathbf{x}_f,\mathbf{y}_f)\sim p(\mathbf{x}_f,\mathbf{y}_f)} (\mathbf{x}_f - \boldsymbol{\mu}_f)(\mathbf{y}_f - \boldsymbol{\mu}_y)^T, \tag{10}$$

where $\boldsymbol{\mu}_y$ is the expectation of $\mathbf{y}_f$ over $p(\mathbf{y}_f)$. Then the Kalman gain matrix is

$$\mathbf{K} = \operatorname{cov}(\mathbf{x}_f, \mathbf{y}_f) \operatorname{cov}(\mathbf{y}_f)^{-1}, \tag{11}$$

with $\operatorname{cov}(\mathbf{y}_f) = \operatorname{cov}(\mathbf{y}_f, \mathbf{y}_f)$.

For large state data, the state covariance is too large to be directly stored. Instead, it must be expressed in a low-rank or sparse form. Similarly, for large observational data, directly forming $\mathbf{K}$ is impossible, and a low-rank or sparse form of $\operatorname{cov}(\mathbf{y}_f)$ must be used in order to apply $\mathbf{K}$ in Equation (9). Furthermore, for nonlinear operators f and h, the distributions are not Gaussian.

The ensemble Kalman filter (EnKF) [46], [47] ameliorates these problems using a Monte-Carlo method. The prior is represented by an ensemble of samples. The predict phase advances each sample forward in time using the transition operator. The update phase simulates observations of each sample and applies the linear Kalman update from Equation (8) to each sample to better match the true observations. The update assumes the state and observation samples are jointly Gaussian, so this method is optimal only for the linear KF case. However, because the samples do not have to be Gaussian, this method has the capability of expressing non-Gaussian distributions. Specifically, if the transition operator is nonlinear, the prior ensemble for the update step is not Gaussian, and if h is nonlinear, the likelihood $p(\mathbf{y}|\mathbf{x})$ for the update step is not Gaussian. Since the update step is linear, it can only update the non-Gaussian prior to a non-Gaussian posterior.

Let the ensemble be forecasted samples $\mathbf{x}_{f,i}$ indexed by i from 1 to N_e. They can be described as samples of $\mathcal{N}(\boldsymbol{\mu}_f, \mathbf{B}_f)$ with forecasted sample mean $\boldsymbol{\mu}_f = \sum \mathbf{x}_{f,i}/N_e$ and sample covariance $\mathbf{B}_f = \mathbf{X}_f\mathbf{X}_f^T$, where $\mathbf{X}_f$ is a matrix with the i-th column being $(\mathbf{x}_{f,i} - \boldsymbol{\mu}_f)/\sqrt{N_e - 1}$. Equation (8) applies to each sample to update the samples with the observation $\mathbf{y}$, written as

$$\mathbf{x}_{a,i} = \mathbf{x}_{f,i} + \mathbf{K}(\mathbf{y} - \mathbf{y}_{f,i}) \text{ for } i \in \{1, 2, \dots, N_e\}, \tag{12}$$

where $\mathbf{y}_{f,i} = h(\mathbf{x}_{f,i}, \boldsymbol{\eta}_{h,i})$ is a simulated observation, $\boldsymbol{\eta}_{h,i}$ is a sample of the noise for the observation operator, and $\mathbf{x}_{a,i}$ is a sample of the posterior. Note that the observation operator here on each sample is simulated with noise, whereas in Equation (8), the mean is observed without noise, since the noise is assumed to have zero mean. Let $\boldsymbol{\mu}_y = \sum_j \mathbf{y}_{f,j}/N_e$ be the sample mean of the predicted observations. Let $\mathbf{B}_y = \mathbf{Y}_f\mathbf{Y}_f^T$ be the sample observation covariance, where $\mathbf{Y}_f$ is a matrix with the i-th column being $(\mathbf{y}_{f,i} - \boldsymbol{\mu}_y)/\sqrt{N_e - 1}$. Then the Kalman gain matrix is

$$\mathbf{K} = \mathbf{X}_f\mathbf{Y}_f^T(\mathbf{Y}_f\mathbf{Y}_f^T)^{-1}, \tag{13}$$

which requires solving a linear system for each ensemble member in Equation (12). Note that the update to each $\mathbf{x}$ resides in the range of $\mathbf{X}_f$, so it is important that $\mathbf{X}_f$ have full column rank, requiring the ensemble members to stay separated.

The ensemble size N_e is typically chosen to be as large as possible for a given computational budget. However, there there are diminishing returns for increasing ensemble size, since the error due to limited ensemble size decreases with $O(1/\sqrt{N_e})$ [27]. For large systems, the cost and diminishing returns limit the ensemble size to a few hundred. In seismic applications, the observations are typically seismic waves or seismic images with millions of degrees of freedom N_y. Since we have $N_e \ll N_y$, the sample observation covariance $\mathbf{Y}_f\mathbf{Y}_f^T$ is explicitly in a low-rank form and is thus singular.

If true observation noise covariance $\mathbf{R}^*$ is known, the sample observation covariance is $\hat{\mathbf{Y}}_f\hat{\mathbf{Y}}_f^T + \mathbf{R}^*$, where $\hat{\mathbf{Y}}_f\hat{\mathbf{Y}}_f^T$ is the non-noisy sample observation covariance obtained from simulating observations without noise. In practice, $\mathbf{R}^*$ is not known. Instead, a nonsingular approximation $\mathbf{R}$ can be constructed, e.g., as a diagonal matrix. Then the observation covariance can be approximated without noise samples as $\hat{\mathbf{Y}}_f\hat{\mathbf{Y}}_f^T + \mathbf{R}$ or with noise samples as $\mathbf{Y}_f\mathbf{Y}_f^T + \mathbf{R}$, where $\mathbf{Y}_f\mathbf{Y}_f^T$ is the noisy sample observation covariance. In both cases, the added matrix approximates the components of the noise covariance not captured in the sample covariance. $\mathbf{R}$ is typically scaled smaller when using the noisy $\mathbf{Y}_f\mathbf{Y}_f^T$ to avoid over-inflating the noise variance in directions captured in the sample covariance.

The standard way to handle seismic data for parameter estimation is by solving an optimization problem minimizing the data misfit. To facilitate comparison with the Kalman filters, we show the optimization formulations of these methods. The mean $\boldsymbol{\mu}_a$ in the Kalman filter update in Equation (8) minimizes the linearized optimization problem

$$\min_{\mathbf{x}} \|h(\boldsymbol{\mu}_f) + \mathbf{H}(\mathbf{x} - \boldsymbol{\mu}_f) - \mathbf{y}\|^2_{\mathbf{R}^{-1}} + \|\mathbf{x} - \boldsymbol{\mu}_f\|^2_{\mathbf{B}_f^{-1}}, \quad (14)$$

and then the covariance estimate $\mathbf{B}$ is the inverse Hessian of the optimization objective at $\mathbf{x} = \boldsymbol{\mu}_a$. Similarly, the EnKF can be formulated as updating each ensemble member by solving N_e linearized optimization problems below to yield $\mathbf{c}_i$,

$$\min_{\mathbf{c}_i \in \mathbb{R}^{N_e}} \|\mathbf{y}_{a,i} - \mathbf{y}\|^2_{\mathbf{R}^{-1}} + \|\mathbf{x}_{a,i} - \mathbf{x}_{f,i}\|^2_{\mathbf{B}_f^{-1}}, \quad (15)$$

with $\mathbf{x}_{a,i} = \boldsymbol{\mu}_f + \mathbf{X}_f\mathbf{c}_i$ and $\mathbf{y}_{a,i} = \boldsymbol{\mu}_y + \mathbf{Y}_f\mathbf{c}_i \approx h(\mathbf{a}_{a,i})$.

This linearization is similar to a full-waveform-inversion-type optimization, which minimizes the nonlinear misfit term typically with repeated linearized optimizations.

The Kalman filter solves the optimization problem for the mean using one linear solve, linearizing the observation operator h about the mean. To get the updated covariance, it requires solving N_x linear systems to apply the Kalman gain to a matrix.

The EnKF solves N_e linear optimization problems, linearizing the observation operator to best fit the ensemble. The updated covariance is implicitly represented by the updated ensemble.

C. *EnKF difficulties*

Although the EnKF efficiently represents the covariance, the EnKF encounters similar challenges as other variants of the KF, specifically in choosing noise statistics. When noise statistics are underestimated, the filter's estimated state covariance becomes increasingly small, such that new observations have negligible effect on the estimated state and the filter diverges from the truth. This is especially an issue for the EnKF with small sample size and small transition uncertainty.

To manage this filter divergence, a conventional approach is to increase the variance in the state. Covariance inflation, described by Anderson and Anderson [48], artificially inflates the covariance of the ensemble by an ad hoc factor. Hamill and Snyder [49], Whitaker and Hamill [50], and Whitaker *et al.* [51] showed that this simple, efficient approach improved the predictive accuracy of atmospheric data assimilation models. In the context of seismic measurements, Alfonzo and Oliver [20] increased the state covariance by estimating observation noise using observation error, thereby avoiding filter divergence.

Filter divergence is especially a problem when observations are assimilated often. For seismic surveys, we assume observations are expensive and therefore temporally sparse, and we do not assimilate enough observations to encounter filter divergence. Our EnKF simulates reservoir dynamics for a year with unknown permeability and then assimilates a seismic survey. The uncertainty in the parameters governing the reservoir dynamics over this large of a time period is enough to ensure the filter covariance does not become too small.

The novelty of this work is applying the EnKF to the seismic-CO_2 monitoring domain with reservoir fluid dynamics. The next two subsections describe the reservoir dynamics and the seismic observation operator we use for the EnKF.

D. *Two-phase flow*

Multi-phase porous flow is a standard approximation for CO_2 reservoir dynamics. For our purposes, the reservoir consists of briny water and injected supercritical CO_2 flowing through porous rock. Simulations may also account for aqueous CO_2, but for ease of presentation, we consider the two fluids to be immiscible.

For each fluid, a saturation field describes what proportion of the pore space is filled with that fluid. Let S_g and P_g be the CO_2 saturation and pressure fields, usually written with no subscript because they are the main quantities of interest. For two-phase flow, the water saturation S_w and pressure P_w can be directly computed as $S_w = 1 - S_g$ and $P_w = P_g + P_c$, where P_c is the capillary pressure. Capillary pressure can be a complex function [52], but we do not attempt to address that field of research. Instead, we focus on the effects of permeability and take capillary pressure to be 0.

The governing PDE is the mass balance equation with Darcy flow applied to each fluid i,

$$\frac{\partial(\phi\rho_i S_i)}{\partial t} - \nabla\cdot(\rho_i v_i) = \rho_i q_i, \quad (16)$$

$$v_i = -k_{ri}/\mu_i K \nabla(P_i - \rho_i g Z), \quad (17)$$

$$S_i, P_i \text{ given at } t = 0, \quad (18)$$

with g being gravitational acceleration and Z being depth. Fluid i has fluid density ρ_i, velocity v_i, volume injection rate q_i, relative permeability k_{ri}, and fluid viscosity μ_i. The permeability field K describes the relationship between fluid flow rate and an applied pressure gradient. The flow rate is slowed by viscosity and relative permeability. Permeability is usually expressed in millidarcies and is anisotropic, typically favoring horizontal directions due to sedimentary layering. We represent it with a diagonal tensor separated into vertical and horizontal components,

$$K = \begin{bmatrix} K_v & 0 \\ 0 & K_h \end{bmatrix}, \tag{19}$$

with K_h being a heterogeneous scalar field of the horizontal permeability and K_v/K_h fixed to a constant less than 1. The porosity ϕ is a scalar field with values in the range 0–1 representing how much of the rock is accessible by the fluids (typically about 20%). The relative permeability is a nonlinear function of the saturation. In line with existing literature, we use a modified Brooks-Corey model with residual saturation r for both CO_2 and brine, given by

$$\begin{aligned} k_{ri} &= \operatorname{clamp}((1-2r)^{-1}(S_i - r),\ 0,\ 1)^2 \\ &= \begin{cases} 1 & \text{if } S_i \geq 1 - r \\ 0 & \text{if } S_i \leq r \\ (1-2r)^{-2}(S_i - r)^2 & \text{else} \end{cases} \end{aligned} \tag{20}$$

Let $\mathcal{M}$ represent the nonlinear operator that numerically solves the PDE in Equations (16) to (18) and (20) from time t_n to t_{n+1} with (spatially discretized) initial conditions $\mathbf{S}^n$ and $\mathbf{P}^n$, permeability $\mathbf{K}$, and porosity $\boldsymbol{\phi}$, written as

$$\mathbf{S}^{n+1}, \mathbf{P}^{n+1} = \mathcal{M}(\mathbf{S}^n, \mathbf{P}^n, \mathbf{K}, \boldsymbol{\phi}; t_n, t_{n+1}). \tag{21}$$

The resulting saturation is then used to simulate seismic measurements, as described in the next subsection.

E. Seismic observation

Surface-seismic data allows measuring subsurface changes without drilling wells. In active-source surface seismic, a seismic source, such as an explosive or air gun, on the surface sends seismic waves into the reservoir. Spatial differences in subsurface wave velocity cause the waves to be reflected or transmitted back to receivers on the surface. The data recorded for this measurement consists of a time series for each possible source-receiver pair.

The wave dynamics are governed by the wave equation. To simulate seismic measurements, the wave equation must be solved for each source to determine the data collected at the receiver locations. In terms of the acoustic pressure field δ_p, the governing wave dynamics are

$$\frac{m}{\rho}\frac{\partial^2 \delta_p}{\partial t^2} - \nabla \cdot \left(\frac{1}{\rho}\nabla \delta_p\right) + w\frac{\partial \delta_p}{\partial t} = \frac{q}{\rho}, \tag{22}$$

where m is the P-wave squared slowness field, ρ is the density field, w is a spatially-varying dampening parameter that is nonzero only in the absorbing boundary layer, and q is an acoustic source. The boundary layer is used to simulate an infinite domain, ensuring any waves that would escape the system of interest are absorbed by the boundary instead of being reflected back into the system.

Seismic measurements can detect CO_2 because the density and squared slowness at each point is dependent on the material composition at that point, including the rock composition and the relative concentrations of CO_2 and brine. This relation is specified by a rock physics model, and a typical rock physics model for CO_2 reservoirs is the patchy-saturation model. Past research has investigated the effects of pressure in the rock physics model [53], [54], but we neglect that here to focus on the leading order effect. The patchy-saturation rock physics model, described in Chapter 1 of [55], expresses the density and squared slowness as a pointwise function of the CO_2 saturation and a baseline model representing the density and squared slowness before injection. This model is a nonlinear interpolation of the squared slowness with 0% CO_2 and 100% CO_2 in the pore space, based on the P-wave modulus of the rock and fluids. We use this model to compute the squared slowness field and density field as the saturation changes. Details are given in Appendix A.

Given the spatially discretized squared slowness field $\mathbf{m}$ and density field $\boldsymbol{\rho}$ from the rock physics model, let $\mathbf{d} = \mathcal{H}(\mathbf{m}, \boldsymbol{\rho}) + \boldsymbol{\eta}$ be the waveforms representing simulating the seismic waves (i.e., solving Equation (22)) and obtaining measurements at the receivers with noise $\boldsymbol{\eta}$. In terms of the CO_2 saturation $\mathbf{S}$,

$$\mathbf{d}(\mathbf{S}; \boldsymbol{\eta}) = \mathcal{H}(\mathbf{m}(\mathbf{S}), \boldsymbol{\rho}(\mathbf{S})) + \boldsymbol{\eta}, \tag{23}$$

where $\mathbf{d}$ being a vector with components $d_{ij}(t)$ as the acoustic pressure measurement at receiver i from source j at time t. These are the full-waveform measurements used for seismic monitoring of CO_2.

III. EnKF APPLIED TO CO_2 PLUME WITH SEISMIC OBSERVATIONS

We apply the scalable, efficient EnKF to the two-phase flow system with seismic observations and show that it gives good estimates of the CO_2 saturation in a high-dimensional (high-resolution) system with temporally sparse seismic measurements. As our seismic model does not depend on the CO_2 pressure, we focus solely on estimating the CO_2 saturation over time. The novelty of this manuscript is that we show viability of addressing all three issues of (1) relatively large problem size with (2) CO_2 dynamics and (3) seismic waveform data. Details on the transition and observation operators are below. In this section, we first describe the particular CO_2 dynamics (transition operator) and the seismic calculations (observation operator) we use, and then we present the ensemble-based workflow for seismic-CO_2 monitoring. We then present simple but comparably scalable algorithms to compare to the EnKF. Finally, we describe the hyperparameters of the EnKF that we vary to show how to pick reliable hyperparameters. Implementation details required for reproducibility are given in the appendices.

A. Transition operator

We use Equation (21) as the transition operator. We assume to know all the transition function parameters except for the saturation $\mathbf{S}$, pressure $\mathbf{P}$, and permeability $\mathbf{K}$. The transition function f updates the saturation $\mathbf{S}$ and pressure $\mathbf{P}$ given the permeability $\mathbf{K}$ and leaves the permeability unchanged. Using $\mathcal{M}$ described by Equation (21), the transition operator for ensemble member i can be written as $\mathbf{x}_i^{n+1} = f(\mathbf{x}_i^n)$, where $\mathbf{x}_i^n$ is a tuple $\mathbf{x}_i^n = (\mathbf{S}_i^n, \mathbf{P}_i^n, \mathbf{K}_i)$ and

$$\mathbf{S}_i^{n+1}, \mathbf{P}_i^{n+1} = \mathcal{M}(\mathbf{S}_i^n, \mathbf{P}_i^n, \mathbf{K}_i, \boldsymbol{\phi}; t_n, t_{n+1}). \tag{24}$$

Note that the stochastic term $\boldsymbol{\eta}_f$ from Equation (1) is not present. This term must be used in systems with large forecast uncertainty, such as inherently stochastic systems or chaotic systems. Specifically, the $\boldsymbol{\eta}_f$ term represents uncertainty in the system even when all parameters are known to high precision. We consider the flow physics to be deterministic, and therefore we express the uncertainty in the transition explicitly in the input parameters instead of adding stochasticity. Typically, when all parameters are known, the physics model is accurate, and we do not address model misspecification here. Although there are methods to account for model misspecification (e.g., [56]), better results are obtained by including misspecified parameters as explicit unknowns as we do here.

The permeability and porosity fields cannot be known exactly, which must be accounted for in our EnKF. As a first step, we focus on the uncertainty created by the unknown permeability field and assume the porosity field is known. We assume that knowledge of the permeability is represented as a distribution from which we can sample permeability fields. Accounting for the uncertainty in the permeability amounts to initializing each ensemble member with a different realization of the permeability field, with each ensemble member's permeability field fixed across time. In practice, the permeability can be characterized (but not computed) by seismically measuring the velocity field and correlating this measurement with permeability measurements along wells. Since permeability is not a one-to-one function of the velocity or density that can be measured seismically, this characterization inherently has uncertainty. In our simulated experiment, we represent this uncertainty by sampling permeability fields via an empirical noisy function of perturbations of the ground-truth velocity model based on the exponential relation described in [57], with details in Appendix B.

B. Observation operator

The wave dynamics for the seismic operator take place on a time scale up to a few seconds with a resolution on the order of milliseconds. Changes in the CO_2 saturation occur on a much longer time scale, so we consider the saturation field to be static when simulating seismic measurements.

Recall $\mathcal{H}$ as the operator mapping from squared slowness $\mathbf{m}$ and density $\boldsymbol{\rho}$ to receiver waveforms $\mathbf{d}$ in Equation (23). In line with typical seismic imaging techniques, we construct an approximation of the true pre-injection baseline parameters $\mathbf{m}_B$ and $\boldsymbol{\rho}_B$ to obtain a smooth baseline model $\mathbf{m}_0$ and $\boldsymbol{\rho}_0$. We compute the reverse-time migration (RTM) image by applying the adjoint of the background Jacobian $\mathbf{J}_0$ with a linear post-processor $\mathbf{P}$ that mutes the water layer and scales the image as a function of depth, as suggested by Herrmann *et al.* [58]. The baseline image is computed by the operator h_B, represented as

$$\mathbf{J}_0 = \frac{d\mathcal{H}}{d\mathbf{m}_0}(\mathbf{m}_0, \boldsymbol{\rho}_0), \tag{25}$$

$$h_B(\nu_B^* \boldsymbol{\eta}_B^*) = \mathbf{P}\mathbf{J}_0^T \Big(\mathcal{H}(\mathbf{m}_B, \rho_B) + \nu_B^* \boldsymbol{\eta}_B^* \Big), \tag{26}$$

where $\boldsymbol{\eta}_B^*$ is true noise with a fixed norm and ν_B^* controls the noise magnitude. The noisy monitor image is computed similarly by operator h on a state $\mathbf{x}$ with noise $\boldsymbol{\eta}$ and noise level ν, and a time-lapse image $\mathbf{A}$ is obtained by subtracting the baseline image from the monitor image,

$$h(\mathbf{x}, \nu\boldsymbol{\eta}) = \mathbf{P}\mathbf{J}_0^T \Big(\mathcal{H}\big(m(\mathbf{S}), \rho(\mathbf{S})\big) + \nu\boldsymbol{\eta} \Big), \tag{27}$$

$$A(\mathbf{x}, \nu\boldsymbol{\eta}, \nu_B^* \boldsymbol{\eta}_B^*) = h(\mathbf{x}, \nu\boldsymbol{\eta}) - h_B(\nu_B^* \boldsymbol{\eta}_B^*), \tag{28}$$

where the measurements depend only on the saturation of the state tuple $\mathbf{x} = (\mathbf{S}, \mathbf{P}, \mathbf{K})$. Note that the image $\mathbf{A}$ is a constant offset from the observation operator h, so it makes no difference in the filter whether we use h or $\mathbf{A}$ as the observation.

C. Workflow overview

Figure 1 shows the workflow for the EnKF-based monitoring workflow. The left half shows the steps taken in the digitally-simulated system, and the right half shows steps taken in the physical system. This setup is known as a digital shadow because the simulated system takes information from the real system without automatically affecting the real system [59].

In the real system, seismic surveys of an injection site are conducted. In the simulated system, transitions and observations of an ensemble of possible subsurface states are simulated, and the ensemble is updated based on field surveys using the EnKF update formula. Figure 2 shows the steps applied independently to each ensemble member to simulate the transition and the observation.

D. Comparison to other algorithms

For the EnKF, we initialize $N_e = 256$ ensemble members with different permeability models and zero saturation. These states represent possibilities for the state of the reservoir at the initial time. For each ensemble member, we simulate two-phase flow with CO_2 injection until the measurement time. The EnKF uses the ensemble covariance and the measurement covariance to update the saturations of each ensemble member towards a value consistent with the observations, according to Equation (12). Then the predict-update process is repeated. The fluid flow simulation advances each ensemble member to the next observation time, at which point a seismic measurement is assimilated that updates the ensemble states.

The EnKF assimilates data from simulations and observations. In order to be worth the extra computational effort, the EnKF must perform better than similar algorithms that use only data from simulations or only data from observations. We propose a comparison with two baseline ensemble methods of updating the forecast, referred to as no-observations method *NoObs* and a just-observations method *JustObs*.

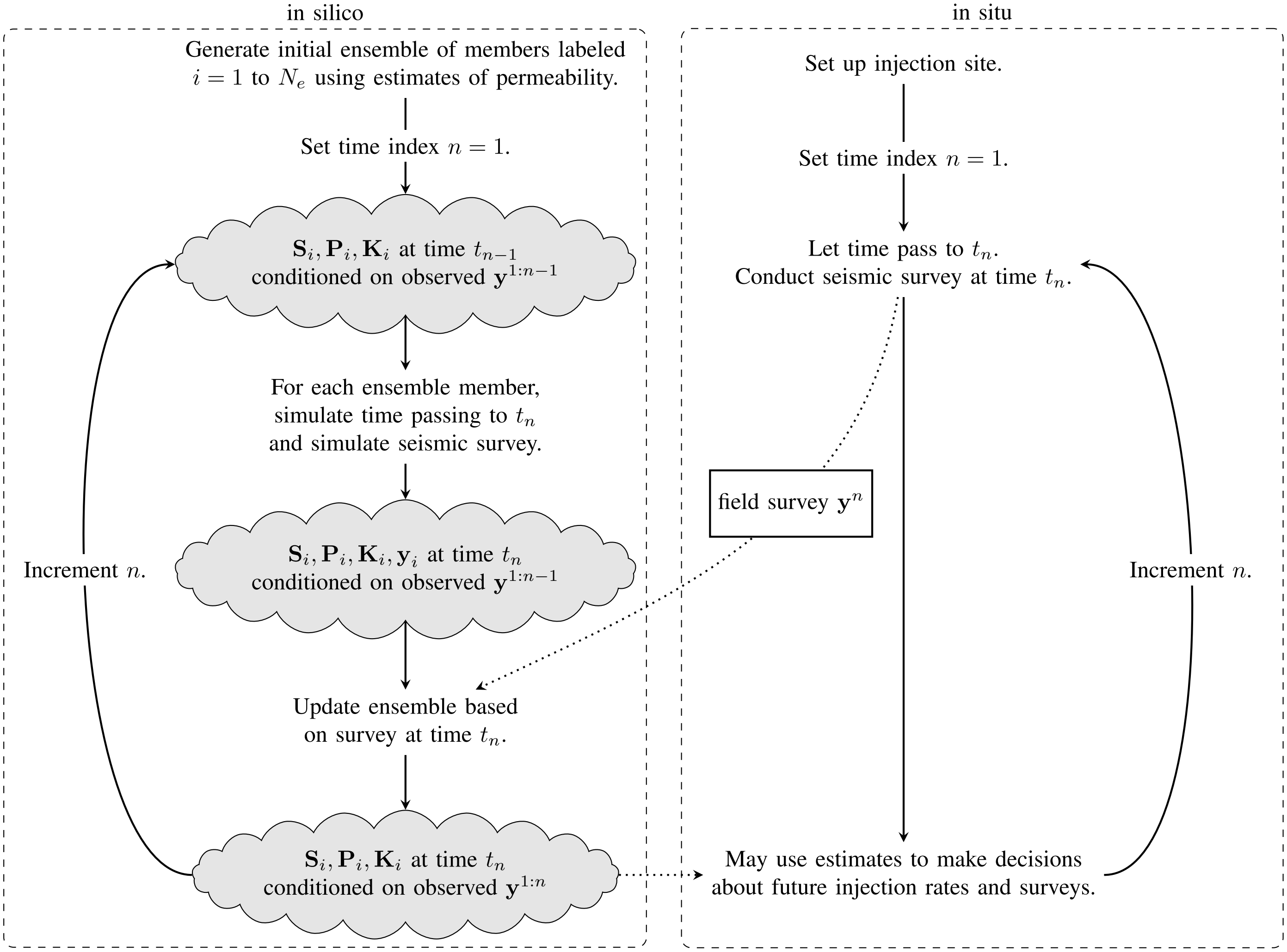


Fig. 1: Workflow diagram for sequential data assimilation for seismic monitoring. Distributions are shown as shaded clouds, and the measurement is shown in a rectangle. The left dashed box contains the steps run in computer simulation. The right dashed box shows actions to be taken in the field. The dotted lines show where information crosses from one side to the other. In simulation, an ensemble represents possible states of the reservoir. Known fluid dynamics are used to simulate the ensemble members forward in time. Known seismic physics are used to simulate the seismic surveys. Then the ensemble is updated based on the field data, and the predict-update cycle repeats.

The NoObs method does not incorporate observations. Instead, it represents the best prediction we can obtain based on prior physics knowledge. Specifically, it uses the two-phase flow model along with the probability distribution describing the permeability field.

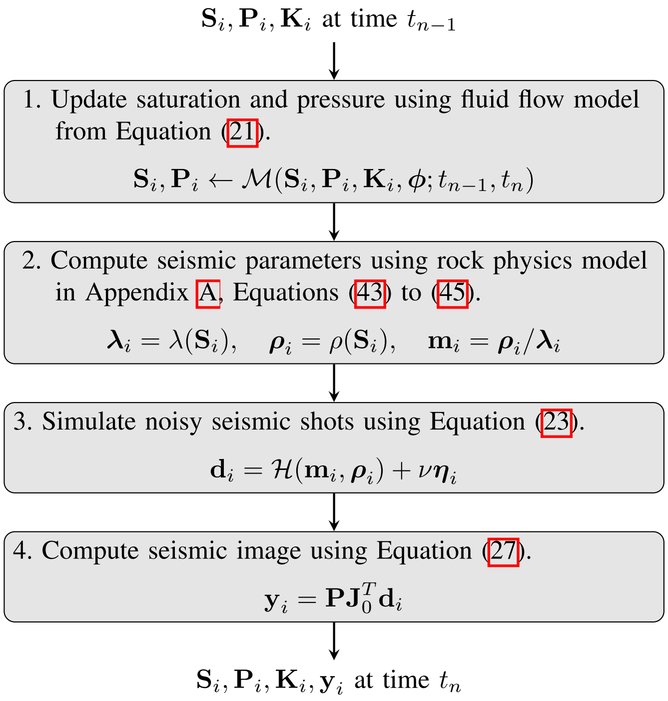
The JustObs method does not use the fluid-flow physics knowledge. Instead, it represents the best prediction we can obtain based solely on observations. Specifically, it uses waveform observations from Equations (38) and (39).

The JustObs optimization is ill-posed and can be sensitive to the initial guess for optimization, so we initialize it with the forecasted ensemble mean and add a regularization term $C(\mathbf{x})$, defined in Appendix C. A strict JustObs formulation that just minimizes the observation misfit performs much worse than the EnKF, but regularization and initialization can improve it. Unfortunately, we were unable to achieve feasible saturation estimates with standard FWI-based regularizations. A highly-tailored JustObs baseline may perform better than the EnKF, but designing it for this problem is complex and not the focus of this paper. We leave the details of our JustObs implementation in Appendix C.

The forecast for each method is computed by applying the transition operator in Equation (24) to each ensemble member i of the previous analysis state $\mathbf{x}_a$ at time step $n-1$, written as

$$\mathbf{x}_{f,i}^{n} = f(\mathbf{x}_{a,i}^{n-1}). \tag{29}$$

We can see the similarities between three methods by writing them each as an optimization, dropping the time superscript from $\mathbf{x}$ since they are all at time step n:



$\mathbf{S}_i, \mathbf{P}_i, \mathbf{K}_i$ at time t_{n-1}

1. Update saturation and pressure using fluid flow model from Equation (21).

$$\mathbf{S}_i, \mathbf{P}_i \leftarrow \mathcal{M}(\mathbf{S}_i, \mathbf{P}_i, \mathbf{K}_i, \boldsymbol{\phi}; t_{n-1}, t_n)$$

2. Compute seismic parameters using rock physics model in Appendix A, Equations (43) to (45).

$$\boldsymbol{\lambda}_i = \lambda(\mathbf{S}_i), \quad \boldsymbol{\rho}_i = \rho(\mathbf{S}_i), \quad \mathbf{m}_i = \boldsymbol{\rho}_i / \boldsymbol{\lambda}_i$$

3. Simulate noisy seismic shots using Equation (23).

$$\mathbf{d}_i = \mathcal{H}(\mathbf{m}_i, \boldsymbol{\rho}_i) + \nu \boldsymbol{\eta}_i$$

4. Compute seismic image using Equation (27).

$$\mathbf{y}_i = \mathbf{P}\mathbf{J}_0^T \mathbf{d}_i$$

$\mathbf{S}_i, \mathbf{P}_i, \mathbf{K}_i, \mathbf{y}_i$ at time t_n

Fig. 2: Operations applied to ensemble member i to advance time to t_n and generate observation $\mathbf{y}_i$. These steps can be applied simultaneously to each ensemble member.

NoObs maximizes the probability of $\mathbf{x}^n$ given the initial state:

$$\begin{aligned} \mathbf{x}_{a,i} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{x}^0_{f,i}) \\ &= \arg\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{x}_{f,i}\|^2_{\mathbf{B}_f^{-1}}. \end{aligned} \tag{30}$$

JustObs maximizes the current data likelihood with ad-hoc regularization C:

$$\begin{aligned} \mathbf{x}_{a,i} &\approx \arg\max_{\mathbf{x}} p(\mathbf{y}^n|\mathbf{x}) \exp(-C(\mathbf{x})/2) \\ &= \arg\min_{\mathbf{x}} \|h(\mathbf{x}) - \mathbf{y}^n\|^2_{\mathbf{R}^{-1}} + C(\mathbf{x}). \end{aligned} \tag{31}$$

EnKF maximizes the posterior likelihood by using the forecast as a prior:

$$\begin{aligned} \mathbf{x}_{a,i} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}^{1:n}) = \arg\max_{\mathbf{x}} p(\mathbf{y}^n|\mathbf{x}) p(\mathbf{x}|\mathbf{y}^{1:n-1}) \\ &= \arg\min_{\mathbf{x}} \|\hat{h}(\mathbf{x}) - \mathbf{y}^n\|^2_{\mathbf{R}^{-1}} + \|\mathbf{x} - \mathbf{x}_{f,i}\|^2_{\mathbf{B}_f^{-1}}. \end{aligned} \tag{32}$$

Recall from Equation (15) that the EnKF uses a linearized observation operator (written $\hat{h}(\mathbf{x})$ here) based on the ensemble members and constrains $\mathbf{x}$ to be expressed by the range of the ensemble mean deviations. The version here is simplified to more easily compare to NoObs and JustObs.

Because these are ensemble methods, each ensemble member has a separate realization for $\mathbf{x}_{f,i}$ and $\mathbf{x}_{a,i}$. Each of the above equations is solved for each ensemble member. Since each ensemble member has a different $\mathbf{x}_{f,i}$, the NoObs and EnKF methods include uncertainty in the resulting $\mathbf{x}_{a,i}$. The NoObs optimization has a simple solution $\mathbf{x}_{a,i} = \mathbf{x}_{f,i}$ because the observations are not included. In JustObs, the resulting $\mathbf{x}_{a,i}$ is identical for each ensemble member because $\mathbf{x}_{f,i}$ does not appear in the optimization expression. For all methods, due to each ensemble member's unique permeability field, the ensemble states tend to drift away from both each other and the ground-truth state when the transition function is applied.

E. Noise tests

We additionally test a range of values for the EnKF's noise parameters while keeping the transition and observation models fixed. We show the EnKF's resiliency to deviations from the ideal parameters, and we provide guidance on how to choose some parameters. Specifically, the EnKF algorithm has choices for how the noise is handled and estimated, and we are interested in how the accuracy of the EnKF's state estimate accuracy changes under incorrect assumptions and different ways to handle the noise.

Recall that the Kalman update can be written in terms of covariances, and the EnKF uses sample covariances from the ensemble, denoted here with $\widehat{\mathrm{cov}}$,

$$\mathbf{x}_{a,i} = \mathbf{x}_{f,i} + \widehat{\mathrm{cov}}(\mathbf{x}_f, \mathbf{y}_f)(\widehat{\mathrm{cov}}(\mathbf{y}_f) + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{y}_{f,i}). \tag{33}$$

$\mathbf{R}$ acts as a regularization in the inversion and is a non-singular approximation of the covariance of any observation noise not already represented in $\widehat{\mathrm{cov}}(\mathbf{y}_f)$.

Our seismic data is represented as $\mathbf{y}_{f,i} = h(\mathbf{x}_{f,i}, \nu\boldsymbol{\eta}_i)$, where the norm of $\boldsymbol{\eta}_i$ is fixed and ν determines the signal-to-noise ratio (SNR), expressed as $-20\log\nu$ dB. If the noise covariance is known precisely, $\widehat{\mathrm{cov}}(\mathbf{y}_f)$ should be computed without noise as $\widehat{\mathrm{cov}}(h(\mathbf{x}_f, 0))$, and R should fully represent the noise covariance.

However, we don't generally know the covariance of the noise; instead, we can simulate noise. If we compute $\widehat{\mathrm{cov}}(\mathbf{y}_f)$ with simulated noise as $\widehat{\mathrm{cov}}(h(\mathbf{x}_f, \nu\boldsymbol{\eta}_h))$, then $\widehat{\mathrm{cov}}(\mathbf{y}_f)$ implicitly contains an estimate of the noise covariance. Unfortunately, the number of ensemble members is much smaller than the observation size, which makes $\widehat{\mathrm{cov}}(\mathbf{y}_f)$ singular. This necessitates regularization in the inversion of the sample observation covariance, in which case, the matrix $\mathbf{R}$ is ideally close to the true noise covariance only in the directions that are not already sampled in $\widehat{\mathrm{cov}}(\mathbf{y}_f)$.

Typical handling of the regularization simply parametrizes $\mathbf{R}$ as a diagonal scaling of the identity, and we follow suit here. The noise variance is proportional to ν^2, so we parametrize $\mathbf{R}$ as $\mathbf{R}(\nu, \beta) = \nu^2\beta^2\mathbf{I}$, with scaling parameter β^2 being the estimate of the average variance of the observation vector's entries when the SNR is 0 dB. If we have the true observation noise covariance $\mathbf{R}^*$, a typical choice is $\nu^2\beta^2 = \mathrm{mean}(\mathrm{diag}(\mathbf{R}^*))$. However, we assume not to know the ground-truth noise covariance of the seismic images. We estimate noise covariance $\hat{\mathbf{R}}$ with simulated noisy and non-noisy observations at the initial time step, thus defining β as

$$\hat{\mathbf{R}} = \widehat{\mathrm{cov}}(h(\mathbf{x}_f, \nu\boldsymbol{\eta}_h) - h(\mathbf{x}_f, 0)), \tag{34}$$

$$\beta = \frac{1}{\nu}\sqrt{\mathrm{mean}(\mathrm{diag}(\hat{\mathbf{R}}))}. \tag{35}$$

For our noise tests, we consider changing the following four noise parameters: first, β, which determines the amount of regularization in the inversion of $\widehat{\operatorname{cov}}(\mathbf{y}_f)$; then $\alpha \in \{0, 1\}$, a binary value which determines whether to simulate the noise in the estimate of $\widehat{\operatorname{cov}}(\mathbf{y}_f)$; and finally, ν and ν^*, which determine the magnitude of the noise in the simulation and in the synthetic ground-truth observations. These scalar parameters are shown underlined in the update equation as

$$\mathbf{x}_{a,i} = \mathbf{x}_{f,i} + \widehat{\operatorname{cov}}\big(\mathbf{x}_f, h(\mathbf{x}_f, \underline{\nu}\boldsymbol{\eta})\big) \left(\widehat{\operatorname{cov}}(h(\mathbf{x}_f, \underline{\alpha}\underline{\nu}\boldsymbol{\eta})) + \underline{\nu}^2 \underline{\beta}^2 \mathbf{I}\right)^{-1} \left[h(\mathbf{x}^*, \underline{\nu^*}\boldsymbol{\eta}^*) - h(\mathbf{x}_{f,i}, \underline{\nu}\boldsymbol{\eta}_i)\right]. \quad (36)$$

We refer to changing β and α as regularization tests, changing ν as simulated noise tests, and changing ν^* as true noise tests, described in more detail below.

Regularization tests: The regularization tests measure performance with different values of β and α, keeping $\nu = \nu^*$ fixed. As β is increased, the update to $\mathbf{x}_i$ goes to 0, so we expect the EnKF estimates to approach the NoObs estimates. For $\alpha = 1$, the noise is included in the samples for the sample covariance, and β is simply a regularization parameter. For $\alpha = 0$, $\mathbf{R}$ should represent the covariance of the noise in the observation operator, and β is chosen to make the resulting variance match the noise covariance on the diagonal.

Simulated noise tests: For these tests, we measure performance when the noise estimate $\nu \neq \nu^*$ while keeping β, α, and ν^* fixed. If ν equals the true noise ν^*, then our ensemble's observation noise variance is an unbiased estimate of the true observation noise variance. When the noise estimate is too high, we expect to regress to the NoObs results, and when the noise estimate is too low, we expect to fit the noisy data too closely. In practice, the true noise variance is not known exactly but can be estimated based on observations and technical specifications of the sensors. Note that we consider strictly unbiased (zero-mean) noise. Biased errors typically occur from incorrect forward models, but we do not address those here.

True noise tests: Finally, we further measure the EnKF's performance with different values of the true noise magnitude ν^*. We are especially interested in how the performance decays for larger values of noise. For these tests, we keep $\nu = \nu^*$ and keep β and α fixed. Large noise should simply give results similar to the NoObs baseline.

IV. TEST SCENARIO

We show the ensemble Kalman filter performance on a synthetic CO_2 injection problem based on the Compass model, which is a synthetic benchmark for seismic full waveform inversion [60]. We simulate the CO_2 injection for 5 years, with new seismic observations every year. The simulation domain is shown in Figure 3, and the scalar simulation parameters are shown in Table I. We use the fluid and mineral parameters from Li *et al.* [61] for brine, supercritical CO_2, and quartz.

The Compass model is a large 3D model based on surveys in the North Sea. It is designed to capture the geological complexities of seismic behaviors in real systems. Running simulations with the full model is expensive, so we limit the domain to a 2D slice, shown in Figure 4.

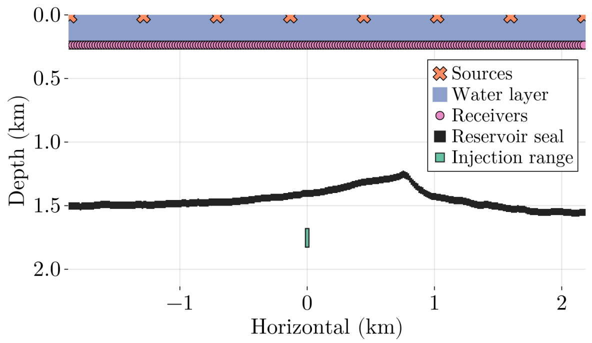


Fig. 3: The simulation domain consists of seismic sources and receivers along the sea surface and sea bottom. The reservoir is approximately 1.5 km deep with an impermeable seal.

Domain size	4.05 km × 2.125 km
Grid size	325 × 341
Cell size	12.5 m × 6.25 m

(a) Domain parameters.

Simulation length	5 years
Injection depth	1.76 km
Injection extent	37.5 m
Injection rate	0.8 Tg/year
Residual saturation	0.1
K_v/K_h	0.36

(b) CO_2 plume parameters.

Simulation length	1.8 s
Source dominant frequency	24 Hz
Source maximum amplitude	7.8 MPa
Signal-to-noise ratio	8 dB
Time step size	0.004 s
Number of receivers	200
Number of sources	8

(c) Seismic survey parameters.

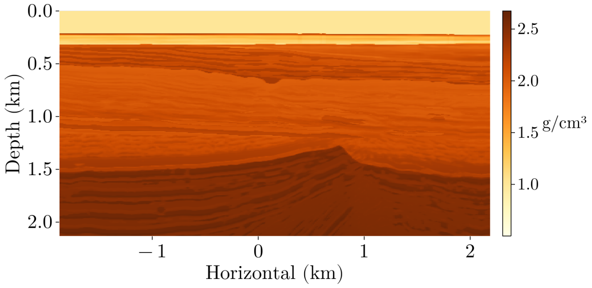
Brine viscosity	10^{-3} Pa·s
CO_2 viscosity	10^{-4} Pa·s
Brine density	1053.0 kg/m^3
CO_2 density	776.6 kg/m^3
Brine bulk modulus	2735 MPa
CO_2 bulk modulus	125 MPa
Density reference pressure	15 MPa

(d) Fluid and mineral parameters.

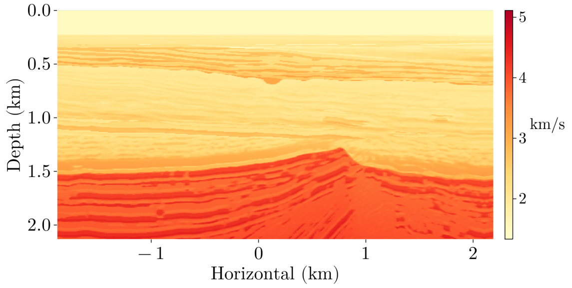
TABLE I: Simulation parameters for the CO_2 plume, seismic survey, and rock physics model.

Each ensemble member has a different permeability model, which is computed from the ground-truth velocity model as detailed in Appendix B. To avoid extremely high pressure, we choose the simulated injection depth for each permeability model to be the maximum permeability value in the 120 meter range labeled in Figure 3, and we initialize the saturation field with a random-valued octagonal patch of 57 grid cells centered on that location.

Figure 5 shows the ground truth permeability, an example ensemble member's permeability, and the mean and standard deviation of the 256 permeability samples used for the ensemble. Figure 6 shows the ground-truth saturation and observation at two times. We set the porosity to a constant 25%.



(a) Density field



(b) P-wave velocity field

Fig. 4: Seismic field parameters. These come from a 2D slice of the Compass model, which is based on the subsurface of the North Sea.

A. Notes on scaling larger

We simulate a relatively small number of sources and receivers on a 2D system with an ensemble of 256 members. For large systems, the ensemble size is typically a few hundred and is independent of the grid size, so we do not expect to change the ensemble size much when scaling to 3D. The EnKF scales linearly with the number of degrees of freedom in the state and observations, so it scales the same as the grid size for scaling from 2D to 3D. The computational cost of running the linear Kalman update used in the EnKF is negligible compared to the computational cost of simulating the transition and observation for each ensemble member. Fortunately, the transition and observation operations are trivially parallelizable across the ensemble.

Compared to 2D, the cost of each flow simulation in 3D is certainly much higher and may dominate the cost. The cost of each seismic simulation is proportional to the number of sources. Increasing the number of sources and receivers affects only the cost of simulating observations, which is proportional to the number of sources. With uncorrelated sensor noise, increasing the number of sources and receivers leads to lower noise images that better show the plume. This decreases the observation noise and allows the EnKF to better match the true plume.

B. Seismic noise

Each frequency component of the noise is drawn from a zero-centered normal distribution with standard deviation proportional to the corresponding frequency's contribution to the source Ricker wavelet. We express the noise norm in terms of the signal-to-noise ratio (SNR) $\gamma = \nu^{-2}$, typically expressed in decibels as $10 \log \gamma$. We consider the signal to be the difference in the observation from the smooth baseline model observation $\mathbf{d_0}$. We scale the noise $\boldsymbol{\eta}$ to have the same norm as the signal so that $\nu\boldsymbol{\eta}$ has the signal-to-noise ratio γ,

$$\gamma = \frac{\|\mathcal{H}(\mathbf{m}(\mathbf{S}), \boldsymbol{\rho}(\mathbf{S})) - \mathbf{d_0}\|^2}{\|\nu\boldsymbol{\eta}\|^2} = \frac{1}{\nu^2}, \tag{37}$$

and similarly for $\boldsymbol{\eta}_B^*$, where the norm is defined as $\|\boldsymbol{\eta}\|^2 = \sum_{i,j} \int \eta_{ij}(t)^2 \, dt$, summed over the sources and receivers and integrated over time.

C. Linearization

The Kalman filter is exact for linear transition and observation operators with additive Gaussian noise. To avoid introducing too much complexity at once to this research, we use the nonlinear rock physics model (see Appendix A) in conjunction with a linearized seismic operator $\mathcal{H}$. Nonlinear seismic data has been successfully assimilated using a machine-learning method by Gahlot *et al.* [43], but for the EnKF, we leave that to future work. Here, we take the initial step of showing that the EnKF accommodates the nonlinearity from the two-phase flow transition and the patchy-saturation model. We use the linear model to generate our synthetic ground-truth observations and to simulate observations for the ensemble.

The seismic operator can be re-parameterized in terms of squared slowness and impedance as $\mathcal{H}_{mz}(\mathbf{m}, \mathbf{z})$. We choose the acoustic impedance, computed pointwise as $\mathbf{z} = \boldsymbol{\rho}/\sqrt{\mathbf{m}}$, because the resulting seismic images (computed with inverse scattering imaging conditions) lack low-frequency updates, thereby revealing the shape of the plume better. We linearize this operator $\mathcal{H}_{mz}$ using the Jacobian $\hat{\mathbf{J}}_0$ with respect to the acoustic impedance about the smooth baseline $\mathbf{z_0} = \boldsymbol{\rho_0}/\sqrt{\mathbf{m_0}}$. Thus, for the linearized model, we replace $\mathcal{H}(\mathbf{m}, \boldsymbol{\rho})$ in Equations (25) to (28) and (37) with

$$\overline{\mathcal{H}}_{mz}(\mathbf{m}, \mathbf{z}) = \mathcal{H}_{mz}(\mathbf{m}_0, \mathbf{z}_0) + \hat{\mathbf{J}}_0(\mathbf{z} - \mathbf{z}_0), \tag{38}$$

where the impedance is computed using the patchy-saturation model (Equations (43) and (44)) as

$$z(\mathbf{S}) = \sqrt{\rho(\mathbf{S})\lambda(\mathbf{S})}, \tag{39}$$

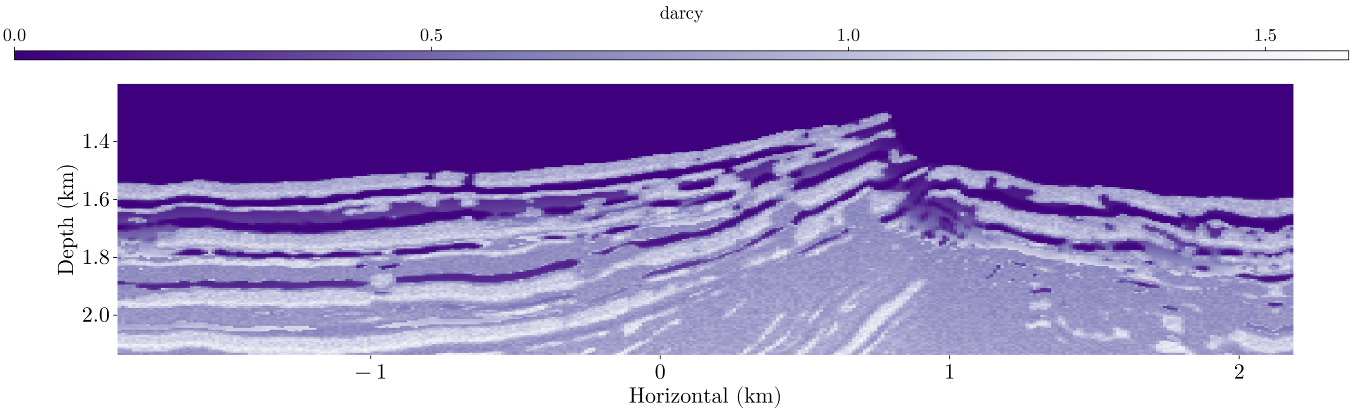
so that the linearized observation operator is

$$h(\mathbf{x}, \nu\boldsymbol{\eta}) = P\hat{\mathbf{J}}_0^T\Big(\hat{\mathbf{J}}_0\big(z(\mathbf{S}) - \mathbf{z}_0\big) + \nu\boldsymbol{\eta}\Big), \tag{40}$$

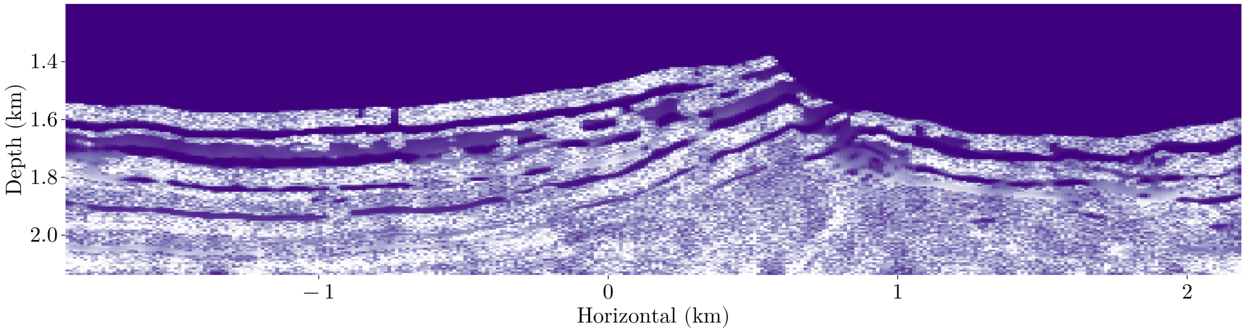
with ground-truth observation $\mathbf{y} = h(\mathbf{x}^*, \nu^*\boldsymbol{\eta}^*)$.

D. Baseline models

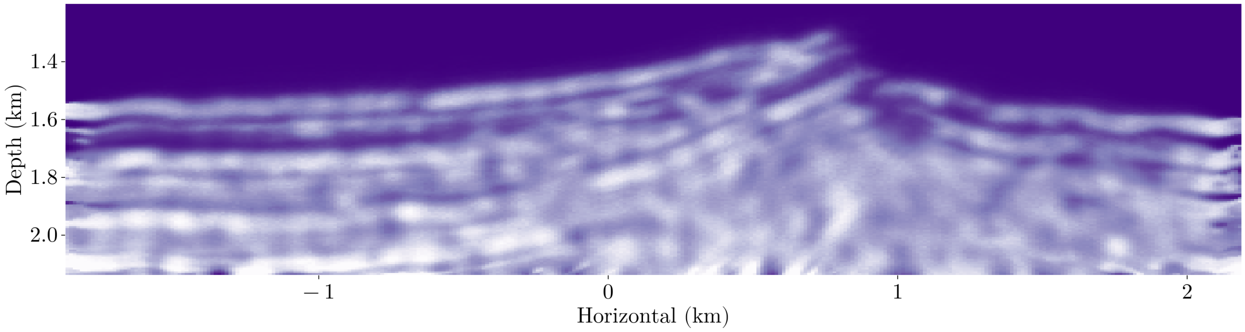
Inverting seismic data itself is difficult, so we do not confound that difficulty with extra uncertainty in the rock physics model. Instead, we assume the pre-injection parameters $\mathbf{m_B}$



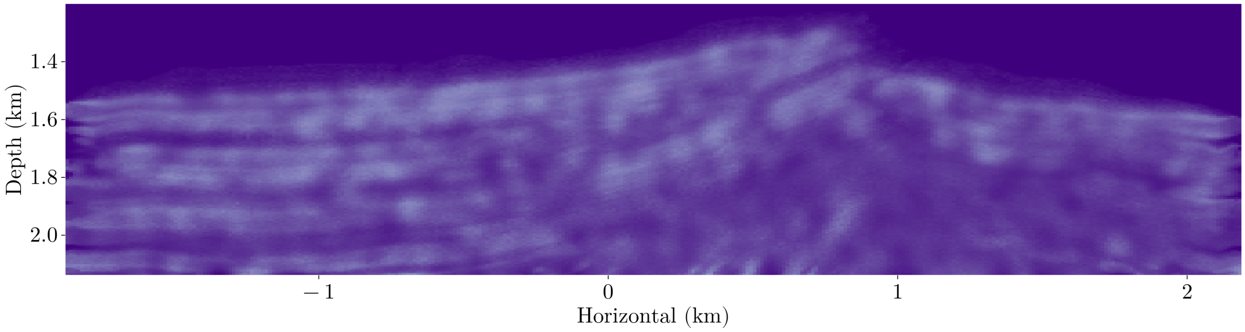
(a) Ground-truth permeability field



(b) Example permeability sample for ensemble

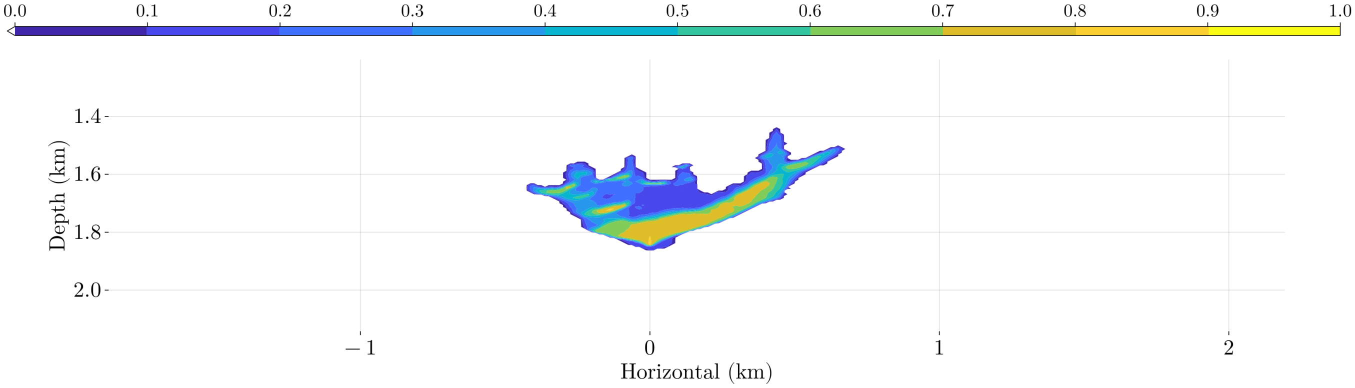


(c) Mean permeability field for ensemble

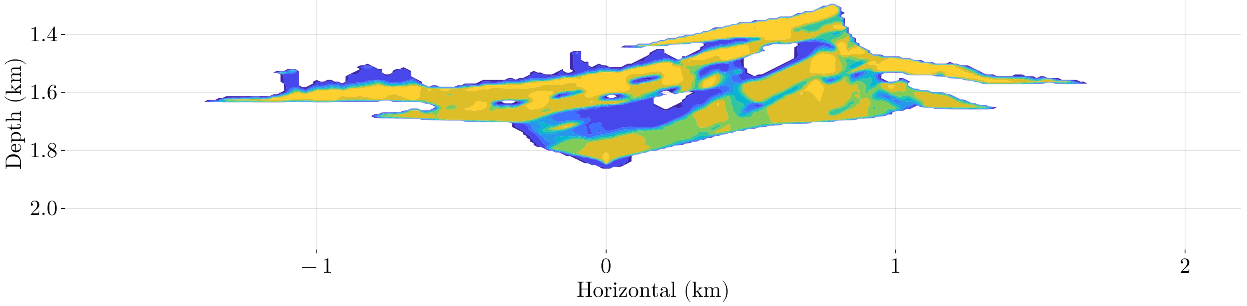


(d) Standard deviation of permeability fields

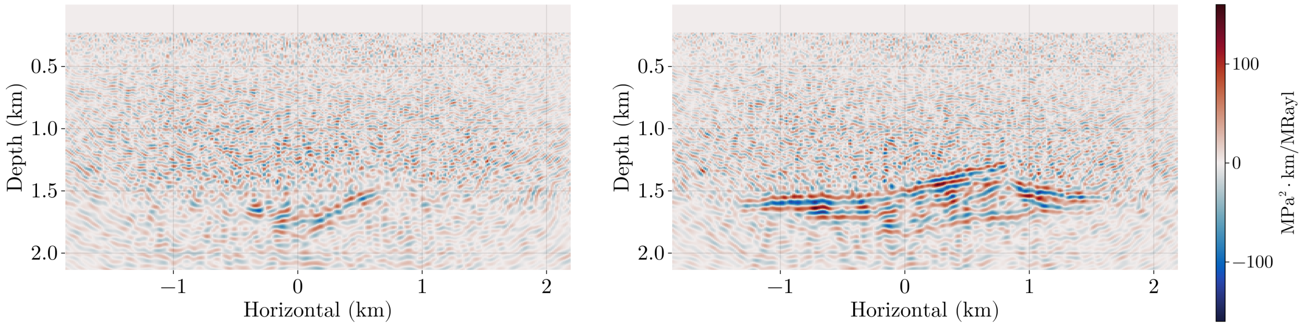
Fig. 5: The permeability fields are computed from perturbations of the velocity field as described in Appendix B. The ground truth permeability is chosen as a small perturbation from the ground truth velocity field, while the ensemble permeabilities are larger, noisier perturbations. The mean ensemble permeability is blurry due to averaging different spatial perturbations.



(a) CO_2 saturation at year 1. Locations with zero CO_2 are transparent.



(b) CO_2 saturation at year 5.



(c) RTM image (year 1)

(d) RTM image (year 5)

Fig. 6: CO_2 plume evolution over time with states in (a), (b) and corresponding noisy observations in (c), (d). Observation units for RTM are squared pressure per impedance times depth, with pressure in megapascals, impedance in megarayls ($g/cm^3 \cdot km/s$), and depth in kilometers.

and $\boldsymbol{\rho}_\mathbf{B}$ are known exactly for the patchy-saturation model of Equations (43) and (44). For the seismic model, we compute the smooth baseline model parameters $\mathbf{m_0}$ and $\boldsymbol{\rho}_\mathbf{0}$ from a Gaussian blur of $\mathbf{m_B}$ and $\boldsymbol{\rho}_\mathbf{B}$ with standard deviation 62.5 m depth-wise and 125 m horizontally.

E. Discretization and PDE solution

For the two-phase flow system, we use the JutulDarcy simulator [62], [63]. JutulDarcy is implemented in the Jutul framework [64]. This code discretizes spatial fields with finite volumes, discretizes time with an implicit Euler step, and chooses time step sizes automatically with an adaptive time stepper. Jutul solves the discretized system with Newton's method with the necessary Jacobian obtained via automatic differentiation.

For the seismic observation, we use the JUDI software [65], [66], which solves the wave equation using the Devito package [67], [68]. The method is an 8th order spatial finite differencing scheme with 2nd order in time with time step size chosen based on the CFL conditions described by Lines *et al.* [69].

Our code will be made open source and available at publication time.

F. Limitations

We make several simplifying assumptions in order to apply the EnKF to the large problem of CO_2 monitoring with seismic data. We explicitly define and discuss several limitations here. To understand how these assumptions limit the results we show, consider that the sources of error in the EnKF are sampling error, linearization error, and Gaussianity error. The sampling error can be decreased by increasing the ensemble size or by including known information into the covariance estimates via methods like covariance deflation and localization. The errors due to the linearity and Gaussianity assumptions of the EnKF are problem dependent, and as we increase the complexity of the problem, this source of error will increase. Decreasing these two sources of error require a nonlinear assimilation method, such as the conditional normalizing flow filter applied to this CO_2 system in [43].

First, in our two-phase fluid flow model, we assume immiscible fluids with a simple relative permeability relation, zero capillary pressure, fixed temperature, and a homogeneous porosity. The immiscibility assumption is valid on short time scales, but dissolution becomes an important trapping mechanism when storing CO_2 for hundreds of years. Heterogeneous in the porosity field affects the plume migration and seismic measurement. Capillary pressure can have a relatively large effect on CO_2 plume migration, although estimating it is very site specific. Introducing capillary pressure adds a nonlinear term to the governing PDE, which we expect to increase the error in the EnKF since it implicitly assumes linearity. Future work should include heterogeneous porosity and the effect of capillary pressure.

Second, in our seismic model, we use the Born approximation, which simplifies computation and interpretation of the results. While the full observation operator is still nonlinear due to the nonlinear rock physics, the seismic linearization makes the assimilation problem easier. In real-world systems, nonlinear effects, e.g., from multiple reflections, must be accounted for in order to accurately estimate subsurface parameters. Due to the EnKF's linearity assumption, we expect to perform worse on nonlinear seismic data at the noise levels we show here. This will be addressed in future work.

Third, we account for colored instrumental noise but not any environmental noise. Instrumental noise is independent for each receiver. Real-world systems are subject to environmental noise, which is non-Gaussian and is spatially correlated across receivers. The EnKF is fit well for correlated Gaussian noise, but non-Gaussianity introduces error in the analysis step. We leave testing more complex noise to future work.

V. NUMERICAL RESULTS

First, we define the metrics we use to measure performance. Since this is a synthetic experiment, we benefit by being able to compare each saturation estimate $\mathbf{S}$ directly to the ground truth saturation $\mathbf{S}^*$. The ensemble mean is the standard estimate to show and is the best estimate of the true saturation field in terms of the ℓ_2 norm assuming a Gaussian distribution. Since our transition and observation operators are nonlinear, our distribution is not Gaussian. Still, there is not a clear choice for a different function of the ensemble states to use, so we show the mean and compute error statistics using the mean. We use root-mean squared error (RMSE) defined as

$$\mathrm{RMSE}(S, S^*) = \sqrt{\frac{1}{|\Omega|}\int_\Omega (S(\mathbf{r}) - S^*(\mathbf{r}))^2 \, dA}, \tag{41}$$

where Ω is the simulation domain, $|\Omega|$ is the area of the domain, $\mathbf{r}$ is the integration variable and the spatial coordinate in the domain, dA is the infinitesimal area at coordinate $\mathbf{r}$, $S(\mathbf{r})$ is the ensemble mean saturation at coordinate $\mathbf{r}$, and $S^*(\mathbf{r})$ is the ground-truth saturation at coordinate $\mathbf{r}$. For our piecewise constant discretization with fixed cell size, this RMSE can be computed as the mean of the squared error in each grid cell. We also computed the structural similarity index measure (SSIM) metric defined in [70]. SSIM takes into account edges as well as values and is strongly correlated with how humans perceive similarity between two images. For our tests, we found the SSIM metric to not yield any more information than the RMSE, so we show only RMSE for most plots.

A. EnKF compared to baselines

Figure 7 shows a comparison of the final plumes for EnKF and NoObs, and Figure 8 shows the error in the final plumes. The plume estimates all have similar spatial extents because they are generated with permeability values in approximately the same range. However, since the locations of the high permeability channels are not known, the NoObs and EnKF predictions cannot capture the interior of the plume. Note that the ground-truth plume has sharp edges due to the sharp edges in the true permeability field. The ensemble estimates are smoother due to each ensemble member having different realizations of the permeability field with different locations for the sharp edges. In order from most smooth to least smooth, we have the NoObs plume, the EnKF prediction, the EnKF analysis, and the ground-truth. This shows that assimilating observations achieves sharper plume estimates, both in forecasting (the EnKF prediction) and filtering (the EnKF analysis).

The most notable differences occur in areas of low permeability in the ground-truth. Specifically, the ground-truth has empty pockets that the CO_2 does not reach due to the low permeability, and the ground-truth CO_2 saturation does not spread out as much. The NoObs estimate misses all except the single pocket at 1.5 km depth, 0.6 km horizontal. The EnKF recovers that pocket and shows signs of the 5 pockets along 1.6 km depth from -0.3 km to 0.3 km horizontal. Each time the transition operator is applied, CO_2 flows into the pockets for members that have high permeability there. Thus, it is very hard to recover these without updating the ensemble members' permeabilities.

Furthermore, the estimated states show larger plume boundaries due to the uncertain permeability allowing the CO_2 to spread faster for some permeability samples. The EnKF estimate shows signs of correcting for this, seen by the noisy scatterings of minuscule but nonzero saturations on the edge of the estimate plume. This is due to the EnKF updating the saturation based on seismic data. Saturation updates outside the boundary of the ground-truth plume tend to push the saturation to be small or even negative, which is then clamped to zero.

Figure 9 shows the error for each method over time, with discontinuities when observations are assimilated. The x-axis starts at year 1 when the first observation is assimilated, because, before observations, each method is identical. JustObs was unable to reduce the RMSE with noisy observations, so we show JustObs results solely for non-noisy data. The NoObs error growth over time shows how much the ensemble's distribution of permeability fields causes the mean estimate to diverge from the ground-truth plume. Comparing the JustObs error with the typically lower NoObs error, we conclude the permeability's uncertainty in this experiment is small enough that the knowledge of the CO_2 dynamics is more informative than seismic measurements. Real applications may have much higher uncertainty in the CO_2 dynamics, which may make JustObs perform better than NoObs.

The EnKF combines the knowledge of the CO_2 dynamics with the seismic measurements and consistently achieves lower error than NoObs and JustObs, thereby showing the benefit of data assimilation. The uncertainty in the permeability causes a sharp increase in error between observations, which shows a limited ability for making accurate predictions. Future research can use the seismic observations to reduce the uncertainty in the CO_2 dynamics, which will decrease the growth in error between observations.

B. EnKF noise parameter tests

By comparing to the baselines, we conclude that the EnKF with our initial choice of EnKF algorithm parameters performs well. Now we examine performance when modifying the EnKF parameters. Figures 10 to 12 show the results, discussed below, of testing values for the different noise parameters in Equation (36), duplicated here,

$$\mathbf{x}_{a,i} = \mathbf{x}_{f,i} + \widehat{\operatorname{cov}}\left(\mathbf{x}_f, h(\mathbf{x}_f, \underline{\nu}\boldsymbol{\eta})\right) \left(\widehat{\operatorname{cov}}(h(\mathbf{x}_f, \underline{\alpha\nu}\boldsymbol{\eta})) + \underline{\nu^2\beta^2}\mathbf{I}\right)^{-1} \left[h(\mathbf{x}^*, \underline{\nu^*}\boldsymbol{\eta}^*) - h(\mathbf{x}_{f,i}, \underline{\nu}\boldsymbol{\eta}_i)\right]. \quad \text{(36 revisited)}$$

Regularization tests: We first vary β with $\nu = \nu^*$ to ensure the EnKF performance is not sensitive to the magnitude of regularization. β scales the regularization covariance $\mathbf{R} = \nu^2\beta^2\mathbf{I}$ in the observation covariance matrix, and has units of $\mathrm{MPa}^2\cdot\mathrm{km}/\mathrm{MRayl}$. For $\alpha = 0$ and making the poor assumption that the seismic image has spatially uncorrelated noise, $\nu^2\beta^2$ should be the variance of the noise. To choose β, we use noise samples to form an estimate $\hat{\mathbf{R}}$ of the true noise covariance $\mathbf{R}^*$. We empirically determine the average noise standard deviation across the seismic observations to be $\nu\beta \approx 11$ for an SNR of 8 dB, which corresponds to $\nu = 10^{-8\ \mathrm{dB}/20} \approx 0.4$ and $\beta \approx 28$. Scaling by the average noise standard deviation is the typical method for regularizing the observation covariance inversion in the EnKF, although there are more advanced methods.

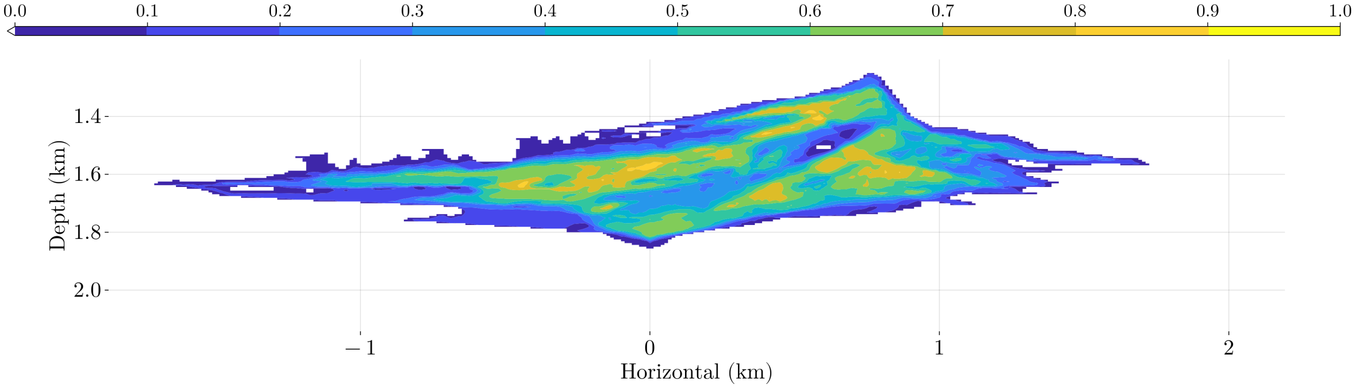
Additionally, we compute the largest eigenvalues of the sample estimate of the true noise covariance. The largest eigenvalue leads to a more unbiased estimate of the regularized observation noise covariance when approximating correlated noise with a multiple of the identity. This leads to an estimate of $\beta \approx 400$. This approximation is an extreme case of a low-rank PCA approximation discussed, for instance, by Woolrich *et al.* [71]. More accuracy could be obtained by using the eigenvectors as part of the regularization, but we attempt a fair comparison with the typical identity scaling by choosing the approximation rank to be zero and simply using the largest eigenvalue.

In Figure 10a, we plot the error as a function of β with simulated noise in the observation covariance ($\alpha = 1$). A wide range of β values from 10^{-4} to 10^3 achieve equivalent results. The error sharply increases for very small β corresponding to $\nu\beta$ approximately six orders of magnitude smaller than the standard deviation of the ground truth noise. As β becomes very large, the post-assimilation error begins to approach the pre-assimilation error as expected. We conclude that if the noise is included in the observation covariance ($\alpha = 1$), then the choice of regularization magnitude has very little effect on the results. The regularization should be chosen within a couple of orders of magnitude of the true noise standard deviation and err on the side of being smaller.

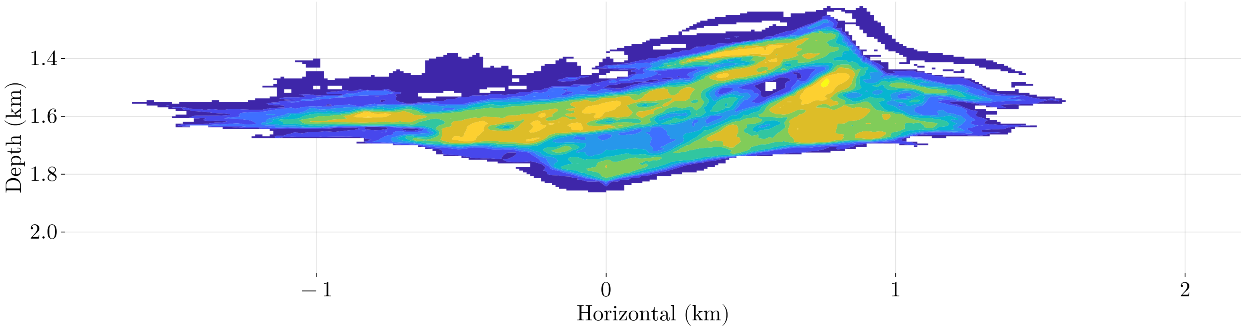
We then vary β without simulated noise in the observation covariance ($\alpha = 0$) to see if simulating the noise gives the EnKF better performance than fully approximating the noise covariance as a multiple of the identity matrix. In Figure 10b, we plot the error as a function of β. We find that for $\alpha = 0$, the RMSE is much more sensitive to the magnitude of the regularization. The minimum RMSE here is not significantly different than the RMSE for $\alpha = 1$ over $\beta \in [10^{-4}, 10^3]$. Furthermore, we highlight the range of β based on the diagonal of the observation noise sample covariance and the range of β based on the largest eigenvalues. The minimum error for $\alpha = 0$ is achieved when choosing the regularization magnitude based on the largest eigenvalues. This can be useful for systems where simulating noise is more costly. For example, the regularization estimate can be computed offline by generating one set of noise samples and computing the resulting noise covariance eigenvalues.

For full waveform observations, simulating noise ($\alpha = 1$) is inexpensive compared to the cost of the transition and observation operators. The benefit from having insensitivity with regards to the noise scale estimate β for $\alpha = 1$ likely offsets the low cost of simulating the noise without significantly affecting the RMSE. Furthermore, in practice, choosing β exactly to get the minimum RMSE may not be possible, especially when the noise is not known exactly. Therefore, we recommend choosing $\alpha = 1$ to avoid sensitivity to β.

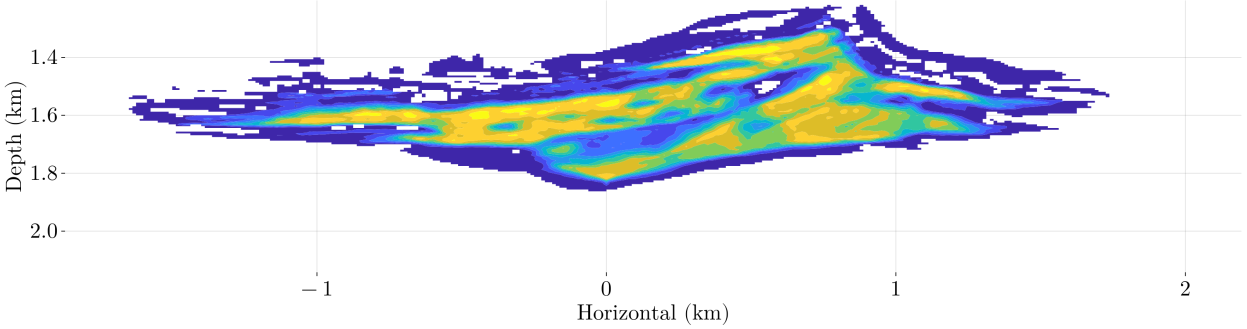
Simulated noise tests: We then vary ν with ν^* fixed and simulating the observation noise samples ($\alpha = 1$) to ensure the EnKF performance is not sensitive to the estimate of the observation noise magnitude. In Figure 11, we plot the error versus the simulated SNR in decibels, which is $-20\log_{10}\nu$. As expected, when the estimated noise magnitude is chosen very large (low SNR), the filter approaches the NoObs results, essentially ignoring observations. On the opposite end, when the simulated SNR is too high, the filter puts too much weight



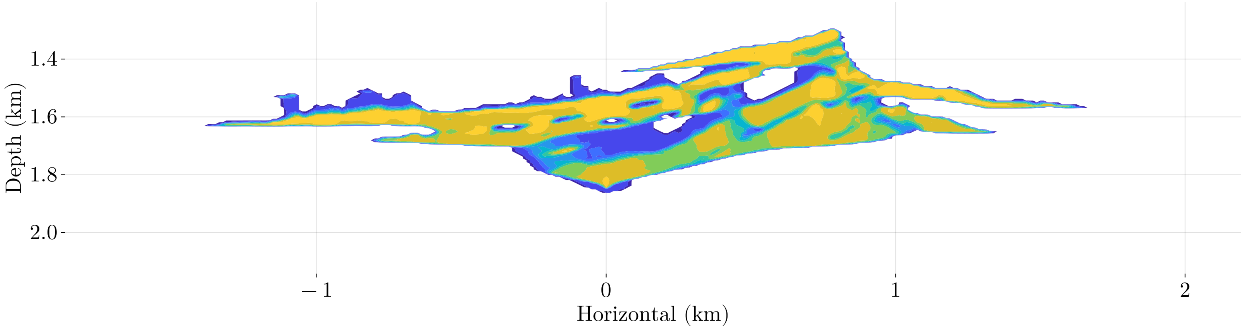
(a) NoObs prediction at the end of year 5 (zero yearly surveys)



(b) EnKF prediction at the end of year 5 (four yearly surveys)

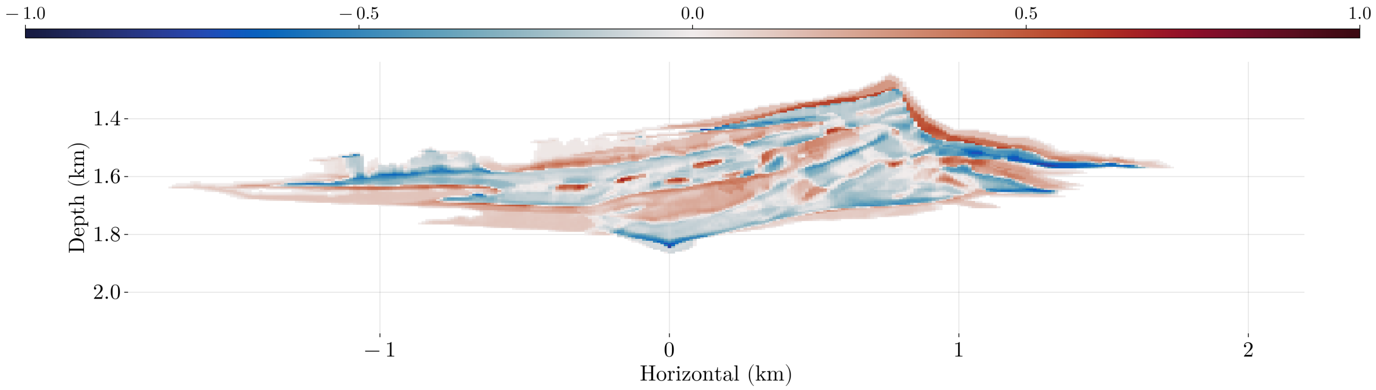


(c) EnKF analysis at the end of year 5 (five yearly surveys)

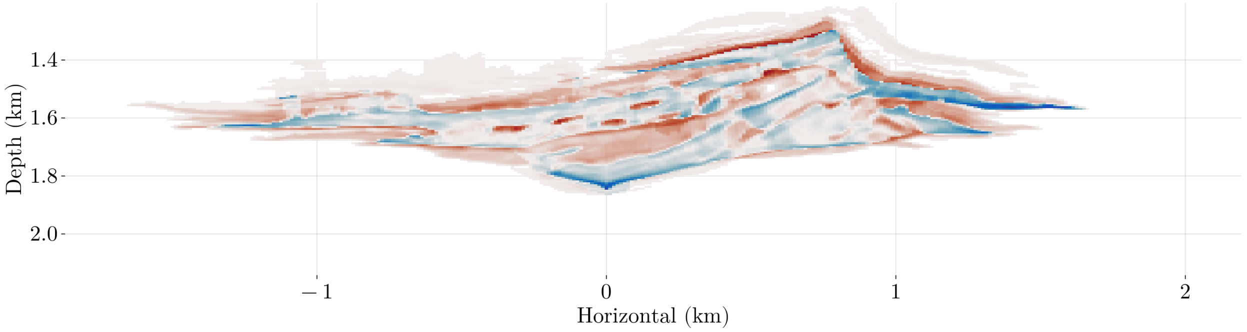


(d) Ground truth at the end of year 5

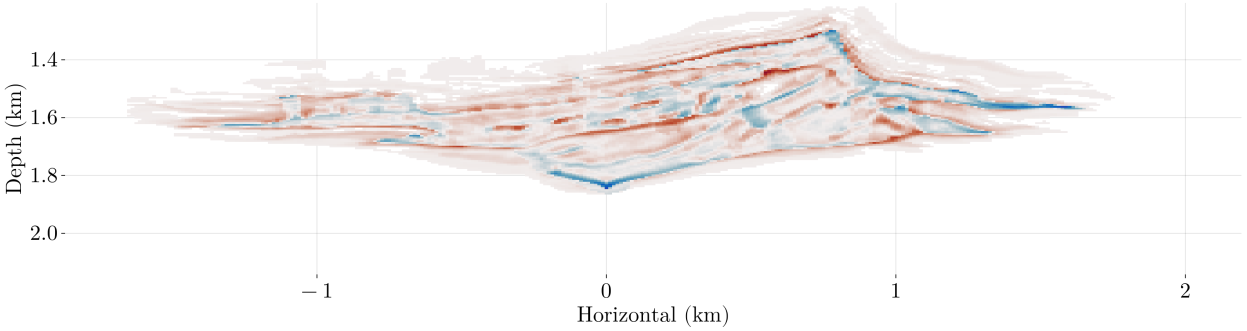
Fig. 7: CO_2 saturation after five years for NoObs, EnKF, and ground-truth. The NoObs (a) and EnKF (b) predictions have blurry plume interiors because they are computed from averaging predictions with different permeability models. The EnKF analysis (c) shows a sharper structure for the plume since it has a current seismic observation.



(a) NoObs prediction error at the end of year 5 (zero yearly surveys)



(b) EnKF prediction error at the end of year 5 (four yearly surveys)



(c) EnKF analysis error at the end of year 5 (five yearly surveys)

Fig. 8: CO_2 saturation error after 5 years for NoObs and EnKF. The similarity in the errors (a), (b) for the NoObs and the EnKF predictions show the limitations of forecasting with incorrect permeabilities. The EnKF analysis error (c) that a seismic survey greatly reduces the error of the estimate.

on the observations, thus overfitting them and raising the error.

Estimating the noise correctly with estimated SNR = γ^* should give the lowest error in the long term. However, Figure 11 shows the optimum estimate is higher than γ^*, with evidence that the optimum estimate is shifting towards γ^* over time. In Figure 11 at year 1, the minimum RMSE is achieved by overestimating the SNR by 10 dB, but by year 5, the minimum RMSE is achieved by overestimating SNR by 5 dB. This result can be explained based on the interplay between the prior and the observations. In particular, overestimating the SNR is equivalent to the covariance inflation technique used in Kalman filtering to ensure the state covariance is not underestimated. The decrease in error that we see here indicates that the ground-truth is too far from our initial prior.

When the observations are estimated to have high uncertainty compared to the prior distribution, the EnKF posterior estimate is closer to the prior estimate. Here, we achieve lower error by estimating lower uncertainty, indicating that the error in the prior estimate is underestimated by the prior uncertainty. While the rigorous solution for this is to increase variability in the prior, it may be beneficial to initially underestimate observation noise to decrease dependence on the prior. As

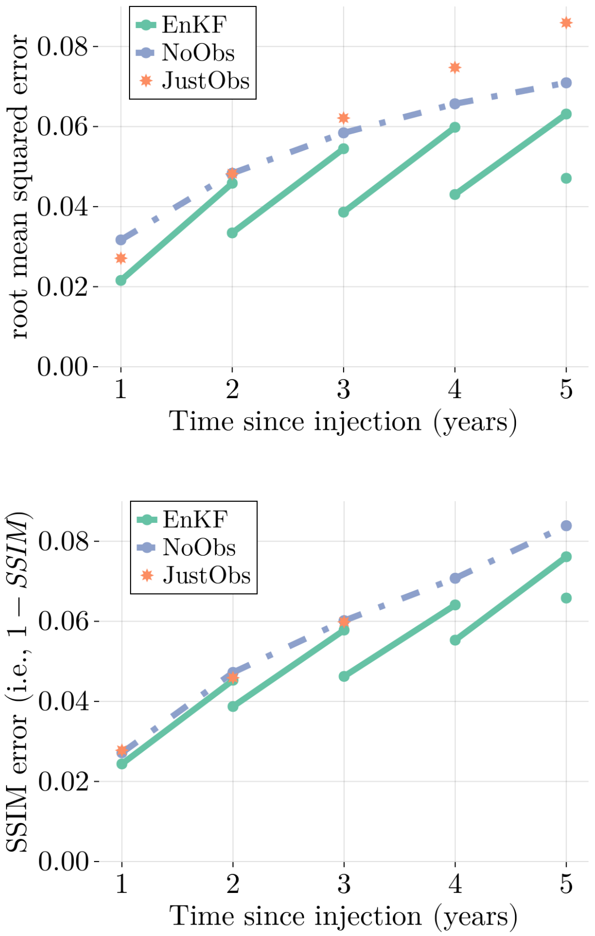
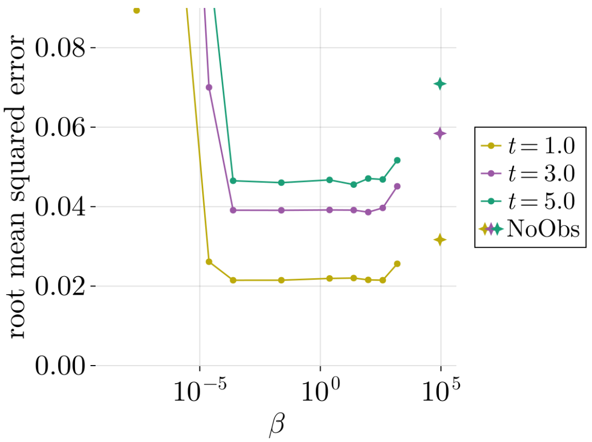


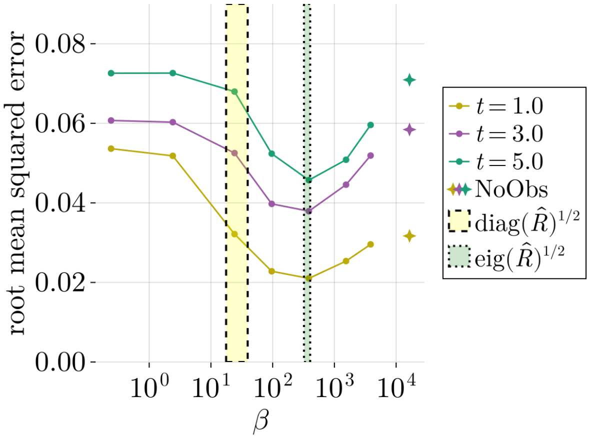
Fig. 9: Error in predicted CO_2 saturation over time. The error between observation steps are linearly interpolated. The errors increase over time when forecasting due to uncertainty in the permeability field. Yearly observations decrease the error in the EnKF estimate. These JustObs errors were generated with non-noisy observations and a hybrid ℓ_1/ℓ_2 regularization on the saturation gradient. The SSIM error for JustObs is too high to be shown on this plot at years 4 and 5.

more observations are collected, the effect of the initial prior decreases, and the minimum should shift to coincide with the true SNR γ^*.

True noise tests: Finally, we record the RMSE for varing ν^*. Here, we keep $\nu = \nu^*$ and simulate the observation noise samples ($\alpha = 1$) to ensure the EnKF can still perform well on varying levels of noisy data. In Figure 12, we plot the error versus the true SNR in decibels. As expected, we find that the EnKF error approaches the NoObs error for large noise magnitudes (low SNR). Except for the first two time steps, we find that the error increases as the noise becomes very small. This can be explained by the results with large SNR in the ν tests in Figure 11 that have large error due to overfitting the noise. As we decrease ν^* and ν together, we see better results, until we reach a point where the $\nu = \nu^*$ does not properly account for the noise. Specifically, for high SNR, we expect simulation error to dominate the noise, and therefore, the error increases. This is not a problem in real scenarios as the real-world seismic noise always dominates floating-point arithmetic errors.



(a) $\alpha = 1$



(b) $\alpha = 0$

Fig. 10: RMSE vs β ($\mathrm{MPa}^2 \cdot \mathrm{km/MRayl}$) with 8 dB SNR ($\nu = \nu^* \approx 0.4$) at three time steps after the EnKF update. In (a), the noise covariance includes simulated noise plus a diagonal identity scaled by β. A wide range of β achieves similar error, with a sharp increase in error for small regularization. In (b), the noise covariance is approximated solely as a diagonal matrix, causing more sensitivity to the choice of β. In either case, the error with large regularization should approach the NoObs case, which is equivalent to $\beta = \infty$.

VI. CONCLUSION

Since the EnKF is known to be capable of scaling to large, nonlinear geophysical systems, we expect it to be a valuable tool for seismic monitoring of CO_2 injection sites. The existing literature has shown various Kalman filters applied to monitoring CO_2 plume on a relatively small scale or without seismic measurements or without CO_2 dynamics. We apply the EnKF to a high-dimensional CO_2 plume monitored with

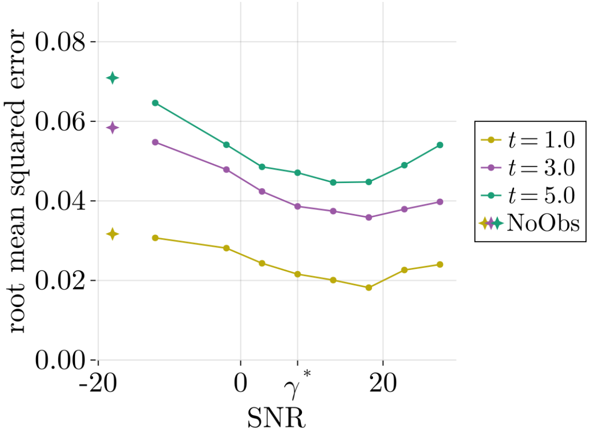


Fig. 11: RMSE vs simulated SNR $\gamma = -20\log_{10}\nu$ dB with fixed ground-truth SNR $\gamma^* = 8$ dB, $\beta = 96.5$ MPa^2 $\cdot$ km/MRayl, and $\alpha = 1$ at three time steps after the EnKF update. Underestimating the SNR should approach the NoObs case (SNR $= -\infty$) as observations are treated as uninformative. Overestimating the SNR by 10 dB gives lower error as it relies less on the initial prior distribution, but overestimating the SNR higher leads to an increase in error as the filter starts to fit the noise.

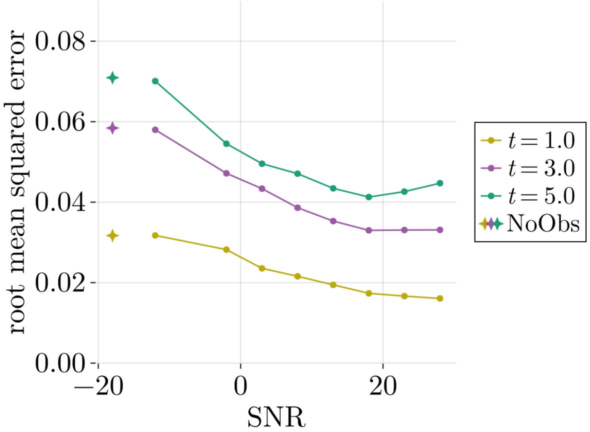


Fig. 12: RMSE vs ground-truth SNR $\gamma^* = -20\log_{10}\nu^*$ dB with $\nu = \nu^*$ and fixed $\beta = 96.5$ MPa^2 $\cdot$ km/MRayl and $\alpha = 1$ at three time steps after the EnKF update. Low SNR (large noise) should approach the NoObs case (SNR $= -\infty$) in the two-norm (in expectation). The post-assimilated error increases with SNR at later time steps. That may indicate ν^* does not capture all the sources of noise, meaning there may be noise even when $\nu^* = 0$ due to numerical simulation. To avoid this issue, it may be better to overestimate the noise instead of underestimating it.

seismic measurements. We show that it achieves lower error than the non-data-assimilation baselines, which ignore the CO_2 dynamics or ignore the observations. This difference in error is small but significant. While the cost of simulating an ensemble of hundreds of reservoirs is expensive compared to doing a single forward simulation, it is far outweighed by the cost of the seismic survey. Conducting a seismic survey is *very* expensive, while the computations required for simulating the seismic survey and flow physics are relatively very cheap. Field-scale applications may require extra modifications to the EnKF to work well, but the benefit in requiring fewer seismic surveys greatly outweighs the computational cost. By showing the EnKF's performance on a synthetic CO_2 monitoring example, we hope to encourage its application in field-scale monitoring.

In addition to comparing to the two baselines, we also examine the sensitivity of the EnKF to the choice of noise parameters. Lower error is achieved by simulating the observation noise and implicitly obtaining a sample observation noise covariance, compared to approximating the observation noise covariance as a diagonal matrix. Since simulating noise for seismic waveform measurements is inexpensive, we recommend doing this to increase the robustness with respect to the choice of regularization magnitude in the EnKF observation covariance. If the noise is not simulated, we see that the regularization parameter should be chosen based on the largest eigenvalues, not the average standard deviation. Regarding the magnitude of the simulated noise, we find slightly better performance with overestimating the true noise, but this result is specific to the case of relatively low uncertainty in the CO_2 dynamics and may not apply in the general case.

Future work: We considered all parameters to be known except for the permeability and the CO_2 saturation and pressure, and we updated only the CO_2 saturation when assimilating observations. Because we do not update the permeability, the error in the EnKF forecast states is very similar to the error from ignoring observations. To achieve accurate predictions of the CO_2 plume, the geological parameters controlling the flow must be updated as well. Therefore, we plan to apply the EnKF to a high-dimensional CO_2 plume with *nonlinear* seismic observations *and* update the permeability and porosity fields along with the CO_2 saturation, following the work of Li *et al.* [16] who updated the permeability and CO_2 saturation with an EnKF variant, Ma *et al.* [21] who updated the permeability and porosity based on the EnKF with approximated seismic images, and Li *et al.* [61] and Yin *et al.* [72] who demonstrated end-to-end inversion for permeability based on time-lapse FWI of CO_2 plumes.

APPENDIX A
PATCHY-SATURATION MODEL

The P-wave modulus λ is related to the density and squared slowness by $\lambda = \rho/m$, and the shear wave modulus G is related to the density and shear wave squared slowness m_G by $G = \rho/m_G$. We assume a typical relation of $m_G = 3m$, but in general, the relation can be more complex. The bulk modulus B is related to the P-wave and shear wave moduli by $B = \lambda - 4/3G$.

Algorithm 2: Pre-computation for patchy-saturation rock physics model in Equations (43) and (44).
Note that each computation here is done pointwise.

```
Input: $\boldsymbol{\rho}_{wr}$, $\mathbf{m}_{wr}$, $\boldsymbol{\phi}$, $B_w$, $B_g$, $B_r$
$\boldsymbol{\lambda}_{wr} \leftarrow \boldsymbol{\rho}_{wr} / \mathbf{m}_{wr}$
$\mathbf{m}_G \leftarrow 3\mathbf{m}_{wr}$
$\mathbf{G} \leftarrow \boldsymbol{\rho}_{wr} / \mathbf{m}_G$
$\mathbf{B}_{wr} \leftarrow \boldsymbol{\lambda}_{wr} - 4\mathbf{G}/3$
$\mathbf{B}_{gr} \leftarrow$ solve Equation (42) for $\mathbf{B}_{gr}$
$\boldsymbol{\lambda}_{gr} \leftarrow \mathbf{B}_{gr} + 4\mathbf{G}/3$
```

Gassman's equation for porous media describes the relation between composite bulk moduli and constituent bulk moduli based on porosity [73]. Here, the constituent bulk moduli are B_r for the mineral making up the porous rock, B_w for brine, and B_g for CO_2. Given the density and velocity before injection, the bulk modulus $B_{wr} = \frac{5}{9}\rho/m$ of the pre-injection brine-rock system is computed using the relations $B_{wr} = \lambda_{wr} - 4/3G$, $m_G = 3m$. Then Gassman's equation below can be solved for the bulk modulus B_{gr} of the rock fully saturated with CO_2,

$$\frac{\phi^{-1}B_w}{B_r - B_w} - \frac{B_{wr}}{B_r - B_{wr}} = \frac{\phi^{-1}B_g}{B_r - B_g} - \frac{B_{gr}}{B_r - B_{gr}}. \tag{42}$$

Replacing brine with CO_2 does not change the shear modulus, so the resulting P-wave modulus of rock saturated with CO_2 is $\lambda_{gr} = B_{gr} + 4/3G$. Algorithm 2 shows the steps to compute λ_{gr}.

The patchy-saturation model uses an arithmetic average for the density and a harmonic average for the P-wave modulus,

$$\rho(S; \rho_{wr}) = \rho_{wr} + S\phi(\rho_g - \rho_w), \tag{43}$$

$$\lambda(S; \lambda_{wr}) = \left[(1-S)\lambda_{wr}^{-1} + S\lambda_{gr}^{-1}\right]^{-1}, \tag{44}$$

where $\lambda_{wr} = \rho_{wr}/m$ with pre-injection density ρ_{wr} and squared slowness m_{wr}. The harmonic average is a result of assuming the pressure is equalized in the mixed fluid. Increasing the CO_2 saturation decreases the density and tends to decrease the P-wave modulus, depending on porosity. The squared slowness is a quadratic function of the saturation,

$$\begin{aligned} m(S; \rho_{wr}, m_{wr}) &= \rho(S)/\lambda(S) \\ &= \left[\rho_{wr} + S\phi(\rho_g - \rho_w)\right]\left[(1-S)\lambda_{wr}^{-1} + S\lambda_{gr}^{-1}\right]. \end{aligned} \tag{45}$$

APPENDIX B
PERMEABILITY AND VELOCITY RELATION

For each ensemble member, the permeability is a function of a deformed version of the true velocity field v. The applied distortion is an elastic deformation provided by the Augmentor.jl software [74]. Let L_i be the domain extent in either the vertical or horizontal direction. The deformation is expressed as a regular coarse grid of displacement vectors sampled uniformly with components between $-L_i$ and L_i and then normalized such that the 2-norm over the grid along each displacement component of the vectors is $0.2L_i$. For a displacement field $\mathbf{d}(\mathbf{r})$, the deformed velocity field is expressed as $v'(\mathbf{r}) = v(\mathbf{r} + \mathbf{d}(\mathbf{x}))$.

The dimensions of the coarse grid are chosen uniformly between 30 and 50, and the distortion vectors along the boundary are set to 0. Furthermore, a Gaussian kernel with length scale between $L_i/15$ and $L_i/25$ is applied to smooth the distortion field. For our grid, the root-mean-squared distortion is 20.3 meters in the horizontal direction and 10.7 meters in the vertical direction.

Following [72], we assume the permeability field follows a similar heterogeneity as the brine-saturated velocity field, and we use a pointwise relation to compute permeability from the deformed velocity field. This relation makes the logarithm of the permeability be a noisy linear function of v based on the exponential dependence shown in [57]. We modify the relation from [72] to increase the variability among samples. We generate the horizontal permeability $K_h(\mathbf{r})$ at spatial coordinate $\mathbf{r}$ as a pointwise random function of a given velocity field $v(\mathbf{r})$ using

$$K_h = \begin{cases} 10^{-2}e^{v-3.5} & \text{if } v < 3.6 \\ 10^{-4(3.85-v)}|c|^{2(v-3.35)} & \text{if } 3.6 \le v < 3.85 \\ |c|e^{v-3.7-w} & \text{if } v \ge 3.85 \end{cases} \tag{46}$$

where K_h is in millidarcies, v is in km/s, and c and w are random fields representing random coefficients for the relation defined in [72]. Specifically, c is a Gaussian random field, fixed across all ensemble samples, with mean 1200, standard deviation 3000, and a Gaussian covariance kernel with length scale 62.5 meters in the horizontal direction and 31.25 meters in the vertical direction. The pointwise random field w is generated from a discrete uniform distribution between 0 and 1.4 with step size of 0.1. This relation separates the scales of the permeability such that high velocity sections have about 10^5 times higher permeability than low velocity sections. Middling velocities have highly variable permeabilities. The ground-truth permeability is generated via the same methodology with the standard deviation of c replaced with 5.

APPENDIX C
JUSTOBS IMPLEMENTATION NOTES

For the JustObs optimization we use a projected quasi-Newton algorithm with spectral projected gradient algorithm described by Schmidt *et al.* [75] and implemented in Julia by Louboutin *et al.* [76]. For regularization $C(\mathbf{x})$, we use $C(\mathbf{x}) = \|\lambda_h^{-1}\mathbf{L}_h\mathbf{S}\| + \|\lambda_v^{-1}\mathbf{L}_y\mathbf{S}\|$, with length scales λ_h and λ_v and linear operators $\mathbf{L}_h$ and $\mathbf{L}_v$ that compute the horizontal and vertical gradients of the saturation field S with finite differences. We experimented with the following norms to regularize the gradient of the saturation: ℓ_1, known as total variation (TV) regularization [77]; ℓ_2, known as Tikhonov regularization [78]; and a hybrid ℓ_1/ℓ_2 norm [79] that benefits from the sparsity behavior of the ℓ_1 norm without requiring the ℓ_1-norm projection. These norms have successfully been used for inverting seismic velocity. A useful, albeit simplified, characterization of these regularizations is TV yields piecewise-constant solutions, Tikhonov yields smoothly-varying solutions, and the hybrid yields piecewise-smooths solutions. The CO_2 plume should be piecewise-smooth instead of piecewise-constant, so we expect the hybrid regularization to yield better results. Unfortunately, the optimization with each of these norms gave very poor estimates of CO_2 saturation, with nonzero CO_2 saturation placed across much of the domain even with non-noisy data, perfect velocity models, and tuned regularization scales. More work could be done to improve the JustObs algorithm, e.g., based on seismic inversion recommendations in [80]. This research may benefit from attribute analysis such as time-frequency analysis, e.g., in [81]. However, since the focus of this paper is the EnKF, comparing a simple JustObs algorithm to the simplest EnKF algorithm is reasonable.

REFERENCES

[1] S. Solomon, G. Bureau-Cauchois, N. Ahmed, J. Aarnes, and P. Holtedahl, "CO_2 storage capacity assessment of deep saline aquifers in the mozambique basin," *Energy Procedia*, vol. 63, pp. 5266–5283, 2014. DOI: 10.1016/j.egypro.2014.11.558.

[2] A. Raza *et al.*, "Integrity analysis of CO_2 storage sites concerning geochemical-geomechanical interactions in saline aquifers," *Journal of Natural Gas Science and Engineering*, vol. 36, pp. 224–240, Nov. 2016. DOI: 10.1016/j.jngse.2016.10.016.

[3] A. Raza, R. Gholami, R. Rezaee, V. Rasouli, and M. Rabiei, "Significant aspects of carbon capture and storage – a review," *Petroleum*, vol. 5, no. 4, pp. 335–340, Dec. 2019. DOI: 10.1016/j.petlm.2018.12.007.

[4] C. B. Raleigh, J. H. Healy, and J. D. Bredehoeft, "An experiment in earthquake control at rangely, colorado," *Science*, vol. 191, no. 4233, pp. 1230–1237, Mar. 26, 1976. DOI: 10.1126/science.191.4233.1230.

[5] M. He, S. Luis, S. Rita, G. Ana, V. Euripedes, and N. Zhang, "Risk assessment of CO_2 injection processes and storage in carboniferous formations: A review," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 3, no. 1, pp. 39–56, Mar. 2011. DOI: 10.3724/SP.J.1235.2011.00039.

[6] N. Yu *et al.*, "Advancing CO_2 storage monitoring via cross-borehole apparent resistivity imaging simulation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023. DOI: 10.1109/TGRS.2023.3331421.

[7] W. J. Stephenson, R. B. Smith, and J. R. Pelton, "A high-resolution seismic reflection and gravity survey of quaternary deformation across the wasatch fault, utah," *Journal of Geophysical Research: Solid Earth*, vol. 98, pp. 8211–8223, B5 May 10, 1993. DOI: 10.1029/92JB02873.

[8] G. Gao, A. Abubakar, and T. M. Habashy, "Joint petrophysical inversion of electromagnetic and full-waveform seismic data," *GEOPHYSICS*, vol. 77, no. 3, WA3–WA18, May 1, 2012. DOI: 10.1190/geo2011-0157.1.

[9] A. Abubakar, T. M. Habashy, M. Li, and J. Liu, "Inversion algorithms for large-scale geophysical electromagnetic measurements," *Inverse Problems*, vol. 25, no. 12, p. 123 012, Dec. 1, 2009. DOI: 10.1088/0266-5611/25/12/123012.

[10] E. S. Um, M. Commer, and G. A. Newman, "A strategy for coupled 3d imaging of large-scale seismic and electromagnetic data sets: Application to subsalt imaging," *GEOPHYSICS*, vol. 79, no. 3, pp. ID1–ID13, May 1, 2014. DOI: 10.1190/geo2013-0053.1.

[11] I. Filina *et al.*, "Integration of seismic and gravity data — a case study from the western gulf of mexico," *Interpretation*, vol. 3, no. 4, SAC99–SAC106, Nov. 1, 2015. DOI: 10.1190/INT-2015-0050.1.

[12] R. Arts, A. Chadwick, O. Eiken, S. Thibeau, and S. Nooner, "Ten years' experience of monitoring CO2 injection in the utsira sand at sleipner, offshore norway," *First Break*, vol. 26, no. 1, Jan. 1, 2008. DOI: 10.3997/1365-2397.26.1115.27807.

[13] F. Huang *et al.*, "The first post-injection seismic monitor survey at the ketzin pilot CO_2 storage site: Results from time-lapse analysis," *Geophysical Prospecting*, vol. 66, no. 1, pp. 62–84, Jan. 2018. DOI: 10.1111/1365-2478.12497.

[14] T. Li *et al.*, "Monitoring CO_2 injection at the CaMI field research station using microseismic noise sources," *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 12, e2022JB024719, Dec. 2022. DOI: 10.1029/2022JB024719.

[15] D. Grana, M. Liu, and M. Ayani, "Prediction of CO_2 saturation spatial distribution using geostatistical inversion of time-lapse geophysical data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 3846–3856, May 2021. DOI: 10.1109/TGRS.2020.3018910.

[16] J. Y. Li, A. Kokkinaki, E. F. Darve, and P. K. Kitanidis, "Smoothing-based compressed state ¡span style="font-variant:small-caps;"¿k¡/span¿ alman filter for joint state-parameter estimation: Applications in reservoir characterization and ¡span style="font-variant:small-caps;"¿CO¡/span¿ $_2$ storage monitoring," *Water Resources Research*, vol. 53, no. 8, pp. 7190–7207, Aug. 2017. DOI: 10.1002/2016WR020168.

[17] B. Dupuy, A. Romdhane, P.-L. Nordmann, P. Eliasson, and J. Park, "Bayesian rock-physics inversion: Application to CO_2 storage monitoring," *GEOPHYSICS*, vol. 86, no. 4, pp. M101–M122, Jul. 1, 2021. DOI: 10.1190/geo2020-0218.1.

[18] C. Huang and T. Zhu, "Towards real-time monitoring: Data assimilated time-lapse full waveform inversion for seismic velocity and uncertainty estimation," *Geophysical Journal International*, vol. 223, no. 2, pp. 811–824, Nov. 1, 2020. DOI: 10.1093/gji/ggaa337.

[19] K. S. Eikrem, G. Nævdal, and M. Jakobsen, "Iterated extended kalman filter method for time-lapse seismic full-waveform inversion," *Geophysical Prospecting*, vol. 67, no. 2, pp. 379–394, Jan. 29, 2019. DOI: 10.1111/1365-2478.12730.

[20] M. Alfonzo and D. S. Oliver, "Seismic data assimilation with an imperfect model," *Computational Geosciences*, vol. 24, no. 2, pp. 889–905, Apr. 2020. DOI: 10.1007/s10596-019-09849-0.

[21] W. Ma, B. Jafarpour, and J. Qin, "Dynamic characterization of geologic CO_2 storage aquifers from monitoring data with ensemble kalman filter," *International Journal of Greenhouse Gas Control*, vol. 81, pp. 199–215, Feb. 2019. DOI: 10.1016/j.ijggc.2018.10.009.

[22] M. Verhaegen and P. Van Dooren, "Numerical aspects of different kalman filter implementations," *IEEE Transactions on Automatic Control*, vol. 31, no. 10, pp. 907–917, Oct. 1986. DOI: 10.1109/TAC.1986.1104128.

[23] J. Y. Li, S. Ambikasaran, E. F. Darve, and P. K. Kitanidis, "A kalman filter powered by h2-matrices for quasi-continuous data assimilation problems," *Water Resources Research*, vol. 50, no. 5, pp. 3734–3749, May 2014. DOI: 10.1002/2013WR014607.

[24] M. Sambridge, "Geophysical inversion with a neighbourhood algorithm—i. searching a parameter space," *Geophysical Journal International*, vol. 138, no. 2, pp. 479–494, Aug. 1, 1999. DOI: 10.1046/j.1365-246X.1999.00876.x.

[25] R. N. Bannister, "A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 637, pp. 1951–1970, Oct. 2008. DOI: 10.1002/qj.339.

[26] R. N. Bannister, "A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 637, pp. 1971–1996, Oct. 2008. DOI: 10.1002/qj.340.

[27] G. Evensen, "The ensemble kalman filter: Theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, Nov. 1, 2003. DOI: 10.1007/s10236-003-0036-9.

[28] M. Butala, R. Frazin, Y. Chen, and F. Kamalabadi, "Tomographic imaging of dynamic objects with the ensemble kalman filter," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1573–1587, Jul. 2009. DOI: 10.1109/TIP.2009.2017996.

[29] N. Kumari, R. Kulkarni, M. R. Ahmed, and N. Kumar, "Use of kalman filter and its variants in state estimation: A review," in *Artificial Intelligence for a Sustainable Industry 4.0*, S. Awasthi, C. M. Travieso-González, G. Sanyal, and D. Kumar Singh, Eds., Springer International Publishing, 2021, pp. 213–230, ISBN: 978-3-030-77070-9. DOI: 10.1007/978-3-030-77070-9_13.

[30] G. Evensen, "Sampling strategies and square root analysis schemes for the EnKF," *Ocean Dynamics*, vol. 54, no. 6, pp. 539–560, Dec. 2004. DOI: 10.1007/s10236-004-0099-2.

[31] J. Y. Li, A. Kokkinaki, H. Ghorbanidehno, E. F. Darve, and P. K. Kitanidis, "The compressed state kalman filter for nonlinear state estimation: Application to large-scale reservoir monitoring," *Water Resources Research*, vol. 51, no. 12, pp. 9942–9963, Dec. 2015. DOI: 10.1002/2015WR017203.

[32] J. D. L. T. Guzman, M. Babaei, J.-Q. Shi, A. Korre, and S. Durucan, "Coupled flow-geomechanical performance assessment of CO_2 storage sites using the ensemble kalman filter," *Energy Procedia*, vol. 63, pp. 3475–3482, 2014. DOI: 10.1016/j.egypro.2014.11.376.

[33] M. Tang, X. Ju, and L. J. Durlofsky, “Deep-learning-based coupled flow-geomechanics surrogate model for CO_2 sequestration,” *International Journal of Greenhouse Gas Control*, vol. 118, p. 103 692, Jul. 2022. DOI: 10.1016/j.ijggc.2022.103692.

[34] M. Tang, Y. Liu, and L. J. Durlofsky, “A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems,” *Journal of Computational Physics*, vol. 413, p. 109 456, Jul. 2020. DOI: 10.1016/j.jcp.2020.109456.

[35] G. S. Seabra, N. T. Mücke, V. L. S. Silva, D. Voskov, and F. C. Vossepoel, “AI enhanced data assimilation and uncertainty quantification applied to geological carbon storage,” *International Journal of Greenhouse Gas Control*, vol. 136, p. 104 190, Jul. 2024. DOI: 10.1016/j.ijggc.2024.104190.

[36] H. Tang *et al.*, “A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage,” *International Journal of Greenhouse Gas Control*, vol. 112, p. 103 488, Dec. 2021. DOI: 10.1016/j.ijggc.2021.103488.

[37] H. Tang *et al.*, “Deep learning-accelerated 3d carbon storage reservoir pressure forecasting based on data assimilation using surface displacement from InSAR,” *International Journal of Greenhouse Gas Control*, vol. 120, p. 103 765, Oct. 2022. DOI: 10.1016/j.ijggc.2022.103765.

[38] Z. Yin, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Learned coupled inversion for carbon sequestration monitoring and forecasting with fourier neural operators,” in *Second International Meeting for Applied Geoscience & Energy*, ser. SEG Technical Program Expanded Abstracts, Society of Exploration Geophysicists and American Association of Petroleum Geologists, Aug. 8, 2022, pp. 467–472. DOI: 10.1190/image2022-3722848.1.

[39] M. Liu, D. Vashisth, D. Grana, and T. Mukerji, “Joint inversion of geophysical data for geologic carbon sequestration monitoring: A differentiable physics-informed neural network model,” *Journal of Geophysical Research: Solid Earth*, vol. 128, no. 3, e2022JB025372, Mar. 2023. DOI: 10.1029/2022JB025372.

[40] T. Sakai, M. Nagao, C. H. Chan, A. Datta-Gupta, and C. Yao, “CO_2 plume imaging with accelerated deep learning-based data assimilation using distributed pressure and temperature measurements at the illinois basin-decatur carbon sequestration project,” in *Proceedings of the 2024 Carbon Capture, Utilization, and Storage Conference*, American Association of Petroleum Geologists, 2024, ISBN: 978-1-73619-224-5. DOI: 10.15530/ccus-2024-4014874.

[41] A. P. Gahlot, H. T. Erdinc, R. Orozco, Z. Yin, and F. J. Herrmann, *Inference of CO_2 flow patterns – a feasibility study*, Version Number: 2, 2023. DOI: 10.48550/ARXIV.2311.00290.

[42] A. P. Gahlot, H. Li, Z. Yin, R. Orozco, and F. J. Herrmann, *A digital twin for geological carbon storage with controlled injectivity*, Version Number: 1, 2024. DOI: 10.48550/ARXIV.2403.19819.

[43] A. P. Gahlot, R. Orozco, Z. Yin, and F. J. Herrmann, *An uncertainty-aware digital shadow for underground multimodal CO2 storage monitoring*, Oct. 2, 2024. DOI: 10.48550/arXiv.2410.01218. arXiv: 2410.01218[physics].

[44] P. B. Stark, “Constraints versus priors,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 3, no. 1, pp. 586–598, Jan. 2015. DOI: 10.1137/130920721.

[45] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1, 1960. DOI: 10.1115/1.3662552.

[46] G. Evensen, “Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics,” *Journal of Geophysical Research: Oceans*, vol. 99, pp. 10 143–10 162, C5 May 15, 1994. DOI: 10.1029/94JC00572.

[47] G. Burgers, P. Jan Van Leeuwen, and G. Evensen, “Analysis scheme in the ensemble kalman filter,” *Monthly Weather Review*, vol. 126, no. 6, pp. 1719–1724, Jun. 1998. DOI: 10.1175/1520-0493(1998)126⟨1719:ASITEK⟩2.0.CO;2.

[48] J. L. Anderson and S. L. Anderson, “A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts,” *Monthly Weather Review*, vol. 127, no. 12, pp. 2741–2758, Dec. 1999. DOI: 10.1175/1520-0493(1999)127⟨2741:AMCIOT⟩2.0.CO;2.

[49] T. M. Hamill and C. Snyder, “A hybrid ensemble kalman filter–3d variational analysis scheme,” *Monthly Weather Review*, vol. 128, no. 8, pp. 2905–2919, Aug. 2000. DOI: 10.1175/1520-0493(2000)128⟨2905:AHEKFV⟩2.0.CO;2.

[50] J. S. Whitaker and T. M. Hamill, “Ensemble data assimilation without perturbed observations,” *Monthly Weather Review*, vol. 130, no. 7, pp. 1913–1924, Jul. 2002. DOI: 10.1175/1520-0493(2002)130⟨1913:EDAWPO⟩2.0.CO;2.

[51] J. S. Whitaker, G. P. Compo, X. Wei, and T. M. Hamill, “Reanalysis without radiosondes using ensemble data assimilation,” *Monthly Weather Review*, vol. 132, no. 5, pp. 1190–1200, May 2004. DOI: 10.1175/1520-0493(2004)132⟨1190:RWRUED⟩2.0.CO;2.

[52] A. Y. Beliaev and S. M. Hassanizadeh, “A theoretical model of hysteresis and dynamic effects in the capillary relation for two-phase flow in porous media,” *Transport in Porous Media*, vol. 43, no. 3, pp. 487–510, 2001. DOI: 10.1023/A:1010736108256.

[53] C. MacBeth, “A classification for the pressure-sensitivity properties of a sandstone rock frame,” *GEOPHYSICS*, vol. 69, no. 2, pp. 497–510, Mar. 2004. DOI: 10.1190/1.1707070.

[54] G. Côrte, H. Amini, and C. MacBeth, “Bayesian inversion of 4d seismic data to pressure and saturation changes: Application to a west of shetlands field,” *Geophysical Prospecting*, vol. 71, no. 2, pp. 292–321, Feb. 2023. DOI: 10.1111/1365-2478.13304.

[55] P. Avseth, T. Mukerji, and G. Mavko, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press, 2010.

[56] J. Baird and C. Dawson, “The representer method for two-phase flow in porous media,” *Computational Geosciences*, vol. 11, no. 3, pp. 235–248, Aug. 2, 2007. DOI: 10.1007/s10596-007-9048-8.

[57] M. Prasad, “Velocity-permeability relations within hydraulic units,” *GEOPHYSICS*, vol. 68, no. 1, pp. 108–117, Jan. 2003. DOI: 10.1190/1.1543198.

[58] F. J. Herrmann, C. R. Brown, Y. A. Erlangga, and P. P. Moghaddam, “Curvelet-based migration preconditioning and scaling,” *GEOPHYSICS*, vol. 74, no. 4, A41–A46, Jul. 2009. DOI: 10.1190/1.3124753.

[59] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “Digital twin in manufacturing: A categorical literature review and classification,” *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018. DOI: 10.1016/j.ifacol.2018.08.474.

[60] C. E. Jones, J. A. Edgar, J. I. Selvage, and H. Crook, “Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies,” presented at the 74th EAGE Conference and Exhibition incorporating EUROPEC 2012, Jun. 4, 2012. DOI: 10.3997/2214-4609.20148575.

[61] D. Li, K. Xu, J. M. Harris, and E. Darve, “Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation,” *Water Resources Research*, vol. 56, no. 8, Aug. 2020. DOI: 10.1029/2019WR027032.

[62] O. Moyner, “JutulDarcy.jl - a fully differentiable high-performance reservoir simulator based on automatic differentiation,” in *ECMOR 2024*, European Association of Geoscientists & Engineers, 2024, pp. 1–37. DOI: 10.3997/2214-4609.202437111.

[63] O. Møyner, B. M. Pacheco, A. Riedemann, G. Bruer, T. Holy, and Z. Yin, *Sintefmath/JutulDarcy.jl: V0.2.11*, version v0.2.11, Nov. 17, 2023. DOI: 10.5281/ZENODO.10150938.

[64] O. Møyner, M. Johnsrud, H. M. Nilsen, X. Raynaud, K. O. Lye, and Z. Yin, *Sintefmath/jutul.jl: V0.2.16*, version v0.2.16, Nov. 17, 2023. DOI: 10.5281/ZENODO.10150601.

[65] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, May 1, 2019. DOI: 10.1190/geo2018-0174.1.

[66] M. Louboutin *et al.*, *Slimgroup/JUDI.jl: V3.3.10*, version v3.3.10, Jan. 23, 2024. DOI: 10.5281/ZENODO.10556864.

[67] M. Louboutin *et al.*, “Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration,” *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, Mar. 27, 2019. DOI: 10.5194/gmd-12-1165-2019.

[68] F. Luporini *et al.*, “Architecture and performance of devito, a system for automated stencil computation,” *ACM Transactions on Mathematical Software*, vol. 46, no. 1, pp. 1–28, Mar. 31, 2020. DOI: 10.1145/3374916.

[69] L. R. Lines, R. Slawinski, and R. P. Bording, “A recipe for stability of finite-difference wave-equation computations,” *GEOPHYSICS*, vol. 64, no. 3, pp. 967–969, May 1999. DOI: 10.1190/1.1444605.

[70] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. DOI: 10.1109/TIP.2003.819861.

[71] M. Woolrich, L. Hunt, A. Groves, and G. Barnes, “MEG beamforming using bayesian PCA for adaptive data covariance matrix regularization,” *NeuroImage*, vol. 57, no. 4, pp. 1466–1479, Aug. 2011. DOI: 10.1016/j.neuroimage.2011.04.041.

[72] Z. Yin, M. Louboutin, O. Møyner, and F. J. Herrmann, “Time-lapse full-waveform permeability inversion: A feasibility study,” *The*

Leading Edge, vol. 43, no. 8, pp. 544–553, Aug. 2024. DOI: 10.1190/tle43080544.1.

[73] F. Gassmann, “Uber die elastizitat poroser medien,” *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich*, vol. 96, pp. 1–23, 1951.

[74] M. D. Bloice, C. Stocker, and A. Holzinger, *Augmentor: An image augmentation library for machine learning*, Aug. 11, 2017. arXiv: 1708.04680[cs,stat].

[75] M. Schmidt, E. Berg, M. Friedlander, and K. Murphy, “Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm,” in *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, D. van Dyk and M. Welling, Eds., ser. Proceedings of Machine Learning Research, vol. 5, PMLR, Apr. 16, 2009, pp. 456–463.

[76] M. Louboutin, Z. Yin, and F. Herrmann, *Slimgroup/SlimOptim.jl: V0.2.5*, version v0.2.5, Mar. 19, 2024. DOI: 10.5281/ZENODO.10839737.

[77] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, Nov. 1992. DOI: 10.1016/0167-2789(92)90242-F.

[78] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*. V. H. Winston & Sons, 1977, xiii+258.

[79] K. P. Bube and R. T. Langan, “Hybrid $\ell 1/\ell 2$ minimization with applications to tomography,” *GEOPHYSICS*, vol. 62, no. 4, pp. 1183–1195, Jul. 1997. DOI: 10.1190/1.1444219.

[80] R. Modrak and J. Tromp, “Seismic waveform inversion best practices: Regional, global and exploration test cases,” *Geophysical Journal International*, vol. 206, no. 3, pp. 1864–1889, Sep. 1, 2016. DOI: 10.1093/gji/ggw202.

[81] N. Liu, Y. Lei, R. Liu, Y. Yang, T. Wei, and J. Gao, “Sparse time–frequency analysis of seismic data: Sparse representation to unrolled optimization,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–10, 2023. DOI: 10.1109/TGRS.2023.3300578.