Detection of potential leakage pathways from geological carbon storage by fluid pressure data assimilation

Ana González-Nicolás^{1*}, Domenico Bau², and Ayman Alzraiee³

¹Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA

²University of Sheffield, Department of Civil and Structural Engineering, Sheffield, UK

³Sandia National Laboratories, Carlsbad, New Mexico, USA

*Corresponding author: T: +1-970-4037858

E-mail address: anagna@gmail.com (A. González-Nicolás)

Postal address: Bureau of Economic Geology, The University of Texas at Austin, University Station, Box X, Austin, TX 78713-8924, USA

Abstract

One of the main concerns of geological carbon storage (GCS) systems is the risk of leakage through “weak” permeable areas of the sealing formation or caprock. Since the fluid pressure pulse travels faster than the carbon dioxide (CO₂) plume across the storage reservoir, the fluid overpressure transmitted into overlying permeable formations through caprock discontinuities is potentially detectable sooner than actual CO₂ leakage occurs. In this work, an inverse modeling method based on fluid pressure measurements collected in strata above the target CO₂ storage formation is proposed, which aims at identifying the presence, the location, and the extent of possible leakage pathways through the caprock. We combine a three-dimensional subsurface multiphase flow model with ensemble-based data assimilation algorithms to recognize potential caprock discontinuities that could undermine the long-term

safety of GCS. The goal of this work is to examine and compare the capabilities of data assimilation algorithms such as the ensemble smoother (ES) and the restart ensemble Kalman filter (REnKF) to detect the presence of brine and/or CO₂ leakage pathways, potentially in real-time during GCS operations. For the purpose of this study, changes in fluid pressure in the brine aquifer overlying to CO₂ storage formation aquifer are hypothetically observed in monitoring boreholes, or provided by time-lapse seismic surveys. Caprock discontinuities are typically characterized locally by higher values of permeability, so that the permeability distribution tends to fit to a non-Gaussian bimodal process, which hardly complies with the requirements of the ES and REnKF algorithms. Here, issues related to the non-Gaussianity of the caprock permeability field are investigated by developing and applying a normal score transform procedure. Results suggest that the REnKF is more effective than the ES in characterizing caprock discontinuities.

Keywords: CO₂ storage, leakage, ensemble Kalman filter, ensemble smoother, normal score transform, ECLIPSE

1 Introduction

In the last decades, geological carbon storage (GCS) has been identified as a technology of great potential for reducing anthropogenic emissions of carbon dioxide (CO₂) in the atmosphere. However, while technically feasible, GCS must be carefully evaluated with respect to environmentally threatening side effects, such as the leakage of CO₂ through sealing formations or caprock. When CO₂ is injected underground it displaces the resident fluid in the target geological formation, which, in the case of deep aquifers, is constituted mainly by high-density saline water, or brine. If brine and/or CO₂ find a pathway through the caprock, they migrate into overlying formations, which may lead to deteriorating the quality of shallow fresh water resources [1]. In particular, CO₂ can produce pH changes of groundwater resources by increasing the concentration of carbonates, which can consequently influence dissolution and sorption of minerals and hazardous trace metals [2, 3].

The 2005 report of the Intergovernmental Panel on Climate Change [4] provided a list of potential CO₂ leakage pathways: (1) “weak” areas of the caprock (permeable areas), where CO₂ breaks into the caprock if capillary entry pressure is exceeded; (2) faults and fractures; and (3) poorly completed and/or abandoned wells. In order to monitor the migration of the CO₂ plume different techniques have been proposed. The US Department of Energy [5] classifies these techniques into three main groups: I) atmospheric monitoring techniques, II) near-surface monitoring techniques, and III) subsurface monitoring techniques. From this perspective, Seto and McRae [6] highlighted the importance of developing model-based framework for risk-based monitoring design prior to large scale deployment of GCS. These frameworks are crucial to ensure safe CO₂ storage by gaining insight in site-specific processes, thus reducing uncertainties and allowing for the formulation of appropriate plans to reduce risk.

Measuring fluids pressure is a key component of subsurface monitoring (Group III). Since the fluid pressure pulse travels faster than the CO₂ plume across the storage reservoir, leakage of brine through possible “weak” areas of the caprock into overlying permeable formations should be identifiable sooner than CO₂ leakage by measuring even small and localized fluid pressure variations in these formations. Fluid pressure changes can be detected, for example, by pressure-monitoring wells [5] or four-dimensional (4D) time-lapse seismic data [7-9].

To date, a number of studies related to monitoring pressure changes in observation wells have been published. Chabora and Benson [10] presented a method to assess the usefulness of pressure monitoring in overlying formations based on the correlation between calculated pressure changes and a proposed detection factor. Similarly, Zeidouni et al. [11] developed an analytical solution to detect leakage through pressure-monitoring wells screened in the geologic formations overlying the injection point. They showed that their model is capable of locating and quantifying the leakage of CO₂ and found a positive correlation between the accuracy of the estimation and the number of available monitoring wells. Also, Nogues et al. [12] developed an analytical solution to estimate the CO₂ and brine leakage from pressure variations observed at monitoring wells and investigated optimal location of the monitoring

wells to improve leakage detection. Park et al. [13] proposed a methodology to detect CO₂ leakage by measuring pressure changes at monitoring wells with a constrained distribution. Sun and Nicot [14] presented an inversion method based on a global optimization algorithm to identify CO₂ leakage through leaky wells from pressure anomalies observed in the layers overlying the injected aquifer.

Analytical solutions and optimization algorithms constitute appealing tools for leakage detection. However, typical analytical solutions are based on highly idealistic assumptions and do not account for system heterogeneities. Improved accuracy of estimation can be achieved by using numerical models. Optimization algorithms may present a high computational cost. A reduction of the computation cost may be achieved by resorting to Monte Carlo inverse methods and ensemble-based Kalman filter (KF) [15] data assimilation methods. With respect to these two methods, Hendricks Franssen and Kinzelbach [16] performed a comparison between an ensemble-based KF and a Monte Carlo sequential self-calibration inverse method to characterize a heterogeneous transmissivity field. They found that the ensemble-based KF method requires a less central processor unit (CPU) effort than the Monte Carlo based method.

For CO₂ leakage detection problems, an attractive feature of KF methods lies in their intrinsic ability to assimilate data into uncertain model results as they are collected, that is, in real time. Data assimilation methods have been used in various earth science disciplines to update model states [17-19] and system parameters [17, 20] based on field observations. The classical KF [15] provides an optimal solution in the case of linear Gaussian systems and unbiased measurements. Expanding the applicability of KF to nonlinear problems can be achieved by using an ensemble of realizations to approximate the prior uncertainty in states and parameters [21]. According to Evensen [22], ensemble-based KF methods can be subdivided into three main categories depending on the scheme adopted to assimilate measurements: (1) ensemble Kalman smoother (EnKS) algorithms, (2) ensemble Kalman filter (EnKF) algorithms, and (3) ensemble smoother (ES) algorithms. Some examples of successful applications of the EnKS can be found in the literature [23-25]. Examples of EnKF applications are reported by Chen and Zhang [17], Hendricks Franssen and Kinzelbach [26], and Li et al. [27]. Skjervheim and Evensen [28] effectively applied the ES to solve the history-matching problem in a petroleum reservoir and compared it

with the EnKF. Bailey and Bau [29] used the ES to obtain the hydraulic conductivity through assimilation of water table height and stream flow rate data. Herrera and Simuta-Champo [30] applied the ES to optimize the location of sampling wells and the depth of the measurements in an aquifer.

To provide an optimal solution, these ensemble-based KF methods require: (a) unbiased and uncorrelated observation errors, (b) parameters and state variables to fit to multi-Gaussian distributions, and (c) a linear relationship between predicted data and model data. Since in practical applications at least one of these assumptions is often not satisfied, there is a need to devise approaches to circumvent these limitations. For example, to avoid the problem of non-Gaussianity, Gu and Oliver [31] successfully applied the normal score transform (NST) [32] to saturation data in a one-dimensional multiphase reservoir. Other more recent studies have addressed issues of non-Gaussianity of state variables or parameters [e.g. 20, 33, 34-37].

In this work, we propose an inverse modeling method based on the assimilation of fluid pressure data collected in strata above the target carbon storage formation. The inverse modeling framework relies on the combination of a subsurface multiphase flow model with ensemble-based data assimilation algorithms. The objective of this study is to investigate and compare the capabilities of the ES and a modified “restart” version of the EnKF (REnKF) to identify the presence of brine and/or CO₂ leakage pathways through the caprock formation during GCS operations. To the best of our knowledge, this is the first time that the ES and the REnKF are applied to detect caprock discontinuities by collection of pressure data from the aquifer overlying the storage formation.

Issues related to the non-Gaussianity of caprock permeability field are also examined. For the purpose of this study, we assume that changes in pressure in the upper aquifer are either observed at monitoring wells or provided by 4D seismic time-lapse surveys. To pose the CO₂ detection problem as an inverse estimation of caprock permeability from known pressure measurements, we assume that leakage areas can be represented as regions with high permeability, resulting in a non-Gaussian bimodal distribution of the caprock permeability. To overcome issues of non-Gaussianity of the caprock permeability spatial distribution a NST procedure is developed and implemented. The multiphase flow

model used in this work is ECLIPSE [38], a reservoir simulator widely used in the petroleum industry. For the practical purposes of our numerical investigations, idealized assumptions of multiphase flow are made for the model setup.

This paper is organized as follows. Section 2 presents first the multiphase flow governing equations, followed by the ES and REnKF algorithms. Section 3 presents and compares the results of the application of these two methods for identifying potential leakage areas of the caprock during GCS operations and provides a thorough discussion on model assumptions. Finally, section 4 summarizes the major findings of the present work.

2 Methodology

In this section, a basic description of the multiphase flow model ECLIPSE [38] used in the numerical experiments is given. Next the ES and the REnKF algorithms are described, followed by the definition of parameters used to assess their performance.

2.1 Multiphase flow model

Movement of CO₂ injected in brine aquifers is simulated using the reservoir model ECLIPSE [38]. In particular, its compositional version E300 is adopted to perform two-phase “compositional” simulations of gas-brine systems, based on the mass balance equations of N_c generic fluid components. E300 relies on the formulation of Trangenstein and Bell [39], which considers the total molar density (moles per pore volume) of each component $\hat{\rho}_j$ ($j = 1, 2, \dots, N_c$) as partitioned among different phases. Assuming the presence of only two fluid phases, a CO₂-rich gas phase, denoted as g , and H₂O-rich liquid phase, denoted as w , $\hat{\rho}_j$ must equal the sum of the gas molar density $\hat{\rho}_{jg}$, and the liquid molar density $\hat{\rho}_{jw}$:

$$\hat{\rho}_j = \hat{\rho}_{jg} + \hat{\rho}_{jw} \quad (j = 1, 2, \dots, N_c) \quad (1)$$

In our case the components are CO₂, water and salt (NaCl), thus $N_c = 3$. For these components, the mass conservation laws in a three-dimensional (3D) setting may be expressed as:

$$\nabla \left(\hat{\rho}_{jg} \frac{\mathbf{v}_g}{s_g} + \hat{\rho}_{jw} \frac{\mathbf{v}_w}{s_w} \right) + q_j = \frac{\partial}{\partial t} (\varphi \hat{\rho}_j) \quad (j = 1, \dots, N_c) \quad (2)$$

152 subject to the condition:

$$s_g + s_w = 1 \quad (3)$$

153 where: φ is the porosity of the medium, s_α and $\mathbf{v}_\alpha$ are the saturation and the Darcy velocity of phase α
 154 (α denotes the phase type, either w or g), respectively, and q_j denotes the source term of component j .

155 Darcy's velocity of phase α is given by:

$$\mathbf{v}_\alpha = -\frac{k k_{r\alpha}}{\mu_\alpha} (\nabla p_\alpha - \rho_\alpha \mathbf{g}) \quad (4)$$

156 where: k is the intrinsic permeability of the porous medium; $\mathbf{g}$ is the gravitational acceleration vector
 157 $(0,0,g)$; and $k_{r\alpha}$, ρ_α , μ_α , and p_α are the relative permeability, the density, the dynamic viscosity, and the
 158 partial pressure of phase α , respectively. The capillary pressure p_c is defined as:

$$p_c = p_g - p_w \quad (5)$$

159 Relative permeabilities (k_{rg} and k_{rw}), phase saturations (s_g and s_w), and capillary pressure can be linked
 160 together by close-form relationships derived by fitting experimental data to capillary models, such as
 161 those developed by Van Genuchten [40] and Brooks and Corey [41]. ECLIPSE allows for specifying
 162 these relationships in tabular form. By substituting Equations (1) and (4) into PDEs (2), the mass
 163 conservation laws for all components can be rewritten as:

$$\begin{aligned} & -\nabla \left[\frac{\hat{\rho}_{jg}}{s_g} \frac{k k_{rg}}{\mu_g} (\nabla p_g - \rho_g \mathbf{g}) + \frac{\hat{\rho}_{jw}}{s_w} \frac{k k_{rw}}{\mu_w} (\nabla p_w - \rho_w \mathbf{g}) \right] + q_j \\ & = \frac{\partial}{\partial t} [\varphi (\hat{\rho}_{jg} + \hat{\rho}_{jw})] \quad (j = 1, \dots, N_c) \end{aligned} \quad (6)$$

164 To solve these PDEs the processes of dissolution, evaporation, and condensation must be taken into
 165 account. This is carried out using the CO2STORE option of ECLIPSE, which allows for the gas phase to
 166 contain supercritical CO₂ and evaporated water, and the liquid phase to contain water, dissolved CO₂ and
 167 dissolved NaCl. The partitions of CO₂ and H₂O between phases are computed by following fugacity
 168 equilibrium equations [42, 43]. The fugacity measures how the available energy of a real fluid differs

from the available energy of an ideal gas. To calculate CO₂ fugacity ECLIPSE uses a modified Redlich-Kwong equation of state [44], whereas water fugacity is computed using Henry's law. ECLIPSE also computes densities and viscosities of CO₂ and water phases. CO₂ gas density is calculated by a cubic equation of state according to Spycher and Pruess [42], whereas CO₂ gas viscosity is computed based on relationships developed by Vesovic et al. [45] and Fenghour et al. [46].

In ECLIPSE, Equations (6) are solved using a finite-difference discretization in space. The time variable is integrated with a finite-difference fully implicit Euler scheme (selected by the E300 option FULLIMP) to ensure stability of the numerical solution. Newton's method is used to linearize and solve a set non-linear equations derived from Equation (6). In E300, the primary solution variables are the phase pressures p_g and p_w , and the molar densities of each component $\hat{\rho}_j$. The linear equations produced by Newton's method are preconditioned and solved simultaneously by a nested-factorization procedure that uses the "Orthomin" iterative method [47].

ECLIPSE output includes, among others, the spatial distributions of fluid phase pressures and saturations in the model domain. For each cell, centered at the generic location (x, y, z) , the equivalent fluid pressure $\bar{p}$ at time t is calculated as the average of the phase pressures, p_w and p_g , weighted with respect to the phase saturations, s_w and s_g [48, 49]:

$$\bar{p} = p_w s_w + p_g s_g \quad (7)$$

Therefore the fluid pressure change at location (x, y, z) and time t is defined as:

$$\Delta p(x, y, z; t) = \bar{p}(x, y, z; t) - p_0(x, y, z) \quad (8)$$

where $p_0(x, y, z)$ is the initial fluid pressure (at time $t = 0$).

2.2 Ensemble based data assimilation

The problem of detecting potential CO₂ leakage pathways is approached herein as a data assimilation problem, in which observations of fluid pressure at a number of given locations and subsequent times are assimilated into the multiphase model results to characterize the spatial distribution of permeability of the caprock. Typically, the amount of available pressure data is less than the

unknowns, that is, the model output at the cells of the finite-difference grid, leading to an ill-posed problem. In this case, since the inverse problem has no unique solution, a Bayesian framework is adopted. Prior information on the system state, characterized by the vector $\boldsymbol{\phi}$ including the fluid pressure distribution within the model domain at a given time, can be updated based on pressure data collected at that time using Bayes' law:

$$pdf(\boldsymbol{\phi}|\mathbf{m}, \mathbf{I}) = \frac{pdf(\mathbf{m}|\boldsymbol{\phi}, \mathbf{I})pdf(\boldsymbol{\phi}, \mathbf{I})}{pdf(\mathbf{m}, \mathbf{I})} \quad (9)$$

where the vector $\mathbf{m}$ includes local pressure observations, $pdf(\boldsymbol{\phi}|\mathbf{m}, \mathbf{I})$ represents the posterior probability distribution function (PDF) of $\boldsymbol{\phi}$ given the observations $\mathbf{m}$ and a generic prior information $\mathbf{I}$ on the system, $pdf(\mathbf{m}|\boldsymbol{\phi}, \mathbf{I})$ is the so-called “likelihood” PDF, $pdf(\boldsymbol{\phi}, \mathbf{I})$ is the prior PDF of $\boldsymbol{\phi}$, and $pdf(\mathbf{m}, \mathbf{I})$ constitutes a normalization term. An exact solution to Equation (9) can be obtained when the measurements $\mathbf{m}$ are related to the state $\boldsymbol{\phi}$ through a linear relationship, and when all PDFs conform to Gaussian joint distributions. This solution is referred to as the Kalman filter (KF) [15]. In the KF, the assimilation of system observations into the system state is carried out with a two-step process, defined by a forecast stage and an update stage. In the forecast stage, a forward in time prediction of the state and its error covariance is made. This prediction is carried out at the time in which system observations are collected. In the update stage, the system state is corrected to resemble field measurements.

The application of the KF to 3D system states is computationally expensive and is limited to Gaussian linear systems. Evensen [21] devised the ensemble Kalman filter (EnKF) in order to extend the KF to nonlinear systems. Within the EnKF, the prior PDF of $\boldsymbol{\phi}$, $pdf(\boldsymbol{\phi}, \mathbf{I})$ (Equation (9)) is approximated by an ensemble of realizations that characterize the prior uncertainty in the system state.

The subpanels (a) and (b) in Figure 1 illustrate the procedural structures of the EnKF and the ES, respectively. The horizontal red arrows indicate forecast simulation stages. The vertical blue arrows represent collection of data into the forecast model results. Similar to the classic KF, the EnKF [21] (Figure 1a) calculates system updates sequentially as system measurements are made available; the forward model forecast simulation proceeds from the current to the next data assimilation time. In this

case, measurement collected at a given time are used to condition the system state at that time, which is then used to restart the forward model until the next data assimilation time. Different from the EnKF, the ES [50] (Figure 1b) uses a forecast state formed by the system states at all measurement times. All available data are assimilated simultaneously and thus are used to correct the system states at all previous and future data assimilation times.

Figure 1. Data assimilation processes for (a) the EnKF, (b) the ES (adapted from Evensen [22]), and (c) the REnKF. Horizontal red arrows indicate forecast simulation stages. Vertical blue arrows represent collection of data, whereas vertical black arrows indicate assimilation of data. The green arrows in (c) indicate restart from initial time after the update stage. Δp_i denotes the vector $\mathbf{m}$ of fluid overpressure data collected at a number N_t of generic times $\mathbf{t}_i$ ($i=1,2,\dots,N_t$).

The EnKF and the ES can be used to update not only the system states, but also the system parameters. This is achieved by “augmenting” the forecast state ensemble $\boldsymbol{\phi}$ with a corresponding ensemble of the system parameters $\mathbf{P}$, thus resulting in a state-parameter forecast ensemble $[\boldsymbol{\phi} \mathbf{P}]$.

A well-known inconsistency of these data assimilation algorithms, when applied to non-linear processes, is constituted by their tendency to provide corrected state values that are not coherent with the

physical laws behind the model equations. This inconsistency is particularly relevant in the case of the EnKF since in the update stage, the solution of the simulation model calculated in the forecast step is corrected statistically to resemble field observations, and thus may deteriorate the accuracy of the forecast simulation through the next data assimilation time. These issues are further accentuated when the EnKF is used to update not only the state but also the system parameters, as is the case in this study. To avoid the inconsistency between the updated system states and system parameters several methods have been introduced in the literature.

Wen and Chen [51] and Gu and Oliver [31] developed an iterative application of the EnKF. In their procedures, after the update stage, the model is rerun from the previous time step using the updated system parameters to obtain an update of the system states at the current data assimilation time. Other approaches that use the governing equations of the model to update the system parameters can be found in Hendricks Franssen and Kinzelbach [26], Nowak [52], and Crestani et al. [33]. Hendricks Franssen and Kinzelbach [26] presented a modified EnKF, termed the restart EnKF (REnKF). With the REnKF, the forecast ensemble is formed by the system state and the system parameters, and at any given assimilation time these are updated in a fashion similar to the EnKF. However, after each update step, the forward model is restarted from the initial time to the next assimilation time using the newly updated parameters. Restarting the simulation from the initial time is intended to eliminate the previously mentioned inconsistencies of the EnKF.

The REnKF procedure is shown in Figure 1c. Note that, at any data assimilation time, the updated model does not necessarily honor the state observations. However, this is a secondary concern in this case, since the primary goal is to update sequentially the system parameters rather than its states. In this study, to remove inconsistencies between system states and system parameters only the ES and the REnKF are employed. The REnKF presents the advantage of processing observations sequentially, that is, as they become available. Instead the ES processes all observations together at the final time. However, the REnKF involves a greater CPU time than the ES, since it requires running and restarting the simulation model from the initial time at each data assimilation time.

2.2.1 Forecast of system parameters

In this work, the uncertain parameters of interest are the spatial distribution of permeability in the caprock layer overlying the CO₂ storage formation. As explained above, we use data assimilation algorithms to estimate system parameters by considering an “augmented” state-parameter ensemble $[\phi \mathbf{P}]$. For the purposes of this study, the spatial distribution of the caprock permeability k is simulated geostatistically, using as a two-dimensional (2D), stationary, spatially-distributed random process [53]. According to this process, at any given location the log-transformed caprock permeability, $Y = \log k$, is assumed to fit a bimodal PDF, characterized by two average values of facies permeability, μ_{Y_1} and μ_{Y_2} , with probabilities of occurrence P_1 and $P_2 = 1 - P_1$, respectively.

$$Y = \log k = B(\mu_{Y_1}, \sigma'_{1}, P_1; \mu_{Y_2}, \sigma'_{2}, P_2) \quad (10)$$

The first value, μ_{Y_1} , denotes a sealing caprock, whereas the second value, μ_{Y_2} , denotes a permeable, potentially leaky, area of the caprock ($\mu_{Y_1} < \mu_{Y_2}$). In addition, local variability of the caprock Y is prescribed by adding a “nugget” Gaussian noise with standard deviation equal to σ'_i to the facies mean μ_{Y_i} ($i=1, 2$).

In this work, a sequential categorical indicator Kriging simulator (CIKSIM) algorithm (see Appendix A) is used to generate an ensemble of realizations of N_{MC} equally likely realizations of the caprock permeability field. In CIKSIM, the probability of occurrence of each facies at any given point is assumed to be conditional to the presence of that facies in surrounding points. To calculate this conditional probability, we use a Kriging interpolation based on a 2D isotropic, exponential covariance model, expressed as:

$$C_i(\mathbf{d}; \sigma_i, \lambda_i) = \sigma_i^2 \cdot \exp\left(-\sqrt{\sum_{j=1}^2 \frac{d_j^2}{\lambda_i^2}}\right) \quad (i = 1, 2) \quad (11)$$

where $\mathbf{d} \equiv (d_1, d_2)$ is the horizontal distance vector between any two points, λ_i is the facies correlation length, and σ_i^2 is the facies covariance, which is calculated as $P_i \cdot (1 - P_i)$. Note that in this case $\sigma_1^2 =$

278 $\sigma_2^2 = P_1 \cdot P_2$. Since this model is 2D, no variability of the caprock permeability is hypothesized in the
 279 vertical direction. Figure 2 shows a demonstrative example of a bimodal PDF for the caprock
 280 permeability, obtained using $P_1 = 0.8$, $P_2 = 0.2$, $\sigma^2 = 0.16$, $\mu_{Y_1} = -21$, $\mu_{Y_2} = -12$, and $\sigma'_1 = \sigma'_2 = 0.5$.

281 Figure 2. Example of bimodal PDF for the caprock permeability obtained using parameters $P_1 = 0.8$, $P_2 =$
 282 0.2 , $\mu_{Y_1} = -21$, $\mu_{Y_2} = -12$, and $\sigma'_1 = \sigma'_2 = 0.5$.

283 The generated ensemble of caprock permeability fields is denoted as $\mathbf{P} = [\mathbf{Y}_1, \dots, \mathbf{Y}_{N_{MC}}]$, where
 284 $\mathbf{Y}_j$ ($j = 1, \dots, N_{MC}$) is a $n \times 1$ column vector, and n is the number of cells in the finite-difference grid
 285 used to discretize the model domain. This ensemble constitutes a sample of the prior joint PDF of Y . It is
 286 worth pointing out that the set of geostatistical parameters $(P_1, P_2, \lambda_1, \lambda_2, \mu_{Y_1}, \mu_{Y_2}, \sigma'_1, \sigma'_2)$ necessary to
 287 generate the forecast parameter ensemble $\mathbf{P}$ are assumed to be known, and truly constitute the prior
 288 information $\mathbf{I}$ of this problem (Equation (9)). Wen and Chen [51] compared the use of different
 289 covariance models to generate ensembles of reservoir permeability and concluded that the covariance
 290 model is not important if the available permeability dataset is large enough to capture the distinctive and
 291 characteristic changes of the parameter in the model.

2.2.2 Forecast of system states

The forecast ensemble $\boldsymbol{\phi}$ of fluid pressure distributions in the model domain is obtained through a stochastic, or Monte Carlo, multiphase flow simulation performed with ECLIPSE [38]. In this step, the PDEs (6) are solved numerically for each realization $\mathbf{Y}_j$ ($j = 1, \dots, N_{MC}$) of the forecast permeability ensemble $\mathbf{P}$ in order to compute the corresponding forecast state ensemble $\boldsymbol{\phi} = [\mathbf{p}_1, \dots, \mathbf{p}_{N_{MC}}]$, where $\mathbf{p}_j$ ($j = 1, \dots, N_{MC}$) is the $n \times 1$ vector including the fluid pressure values (Equation (8)) at the cells of the finite-difference grid.

Note that the primary variables of the simulation model, that is, the phase pressures p_g and p_w and the molar densities of each component $\hat{\rho}_j$ (Equations (2)), are not included in the state ensemble, since neither the ES nor the REnKF requires feedback from the updated states. In the case of the ES, this is because all observations are assimilated together at the final time. In the case of the REnKF the parameter update is operated sequentially, at each assimilation time, but the updated primary variables are not used since the simulation is restarted from the initial condition after each update step. The fact that neither algorithm makes use of updated states is an attractive feature in that it avoids the introduction of spurious solutions that do not necessarily meet the model equations and thus may be physically inconsistent.

The augmented forecast state-parameter matrix is constructed as $\mathbf{X}_f = [\boldsymbol{\phi} \mathbf{P}]^T$. The size of this matrix is $n_f \times N_{MC}$, where n_f is the number of degrees of freedom used to characterize the state-parameter system. In the case of the ES, the state is formed by the system states at all N_t measurement times and by the system parameters, therefore $n_f = n(N_t + 1)$. In the case of the REnKF, the state is formed by the system state at the current assimilation time and by the system parameters, therefore $n_f = 2n$.

The forecast ensemble $\mathbf{X}_f$ is used to calculate the following $n_f \times n_f$ ensemble prior covariance matrix:

$$\mathbf{C}_f = \frac{(\mathbf{X}_{f,t} - \bar{\mathbf{X}}_f)(\mathbf{X}_{f,t} - \bar{\mathbf{X}}_f)^T}{N_{MC} - 1} \quad (12)$$

where $\bar{\mathbf{X}}_f$ is the prior ensemble mean obtained as $\bar{\mathbf{X}}_f = \mathbf{X}_f \bar{\mathbf{1}}_{n_f \times N_{MC}}$ and $\bar{\mathbf{1}}_{n_f \times N_{MC}}$ is a matrix of size $n_f \times N_{MC}$ with each element equal to $1/N_{MC}$.

2.2.3 Update of system states and parameters

Both the ES and the REnKF produce a Bayesian least-square estimate, in which the state measurements are used to “correct” the forecast matrix $\mathbf{X}_f$ and thus obtain updated ensembles of system states and parameters. The measurements are assembled into a $m \times 1$ vector $\mathbf{m}$. In the investigated problem, it is assumed that fluid pressure measurements are made at a number n_m of locations over N_t measurement times, therefore the size m is equal to $n_m N_t$ in the case of the ES, and to n_m in the case of the REnKF. Based on the KF theory [15], the updated state-parameter matrix $\mathbf{X}_u$ [$n_f \times N_{MC}$] and the updated covariance matrix $\mathbf{C}_u$ [$n_f \times n_f$] are calculated as follows:

$$\mathbf{X}_u = \mathbf{X}_f + \mathbf{K}_t(\mathbf{D} - \mathbf{H}\mathbf{X}_f) \quad (13)$$

$$\mathbf{C}_u = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{C}_f(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (14)$$

where: $\mathbf{D}$ is a $m \times N_{MC}$ matrix obtained by perturbing the measurement vector $\mathbf{m}$ with a Gaussian noise representing unbiased measurement errors, $\mathbf{H}$ is a $m \times n_f$ matrix that maps each measurement to its location in the finite difference grid and in time, so that the product $\mathbf{H}\mathbf{X}_f$ holds model results at measurement locations and times. $\mathbf{R}$ is the $m \times m$ measurement error covariance matrix that includes the variance of the measurement values:

$$\mathbf{R} = \frac{\mathbf{E}\mathbf{E}^T}{N_{MC} - 1} \quad (15)$$

where the matrix $\mathbf{E}$ [$m \times N_{MC}$] is assembled by storing in each of its rows an ensemble of measurement errors sampled from a Gaussian distribution with mean equal to zero and a prescribed standard deviation. This ensemble represents the noise associated with the each of the measurements. The matrix $\mathbf{K}$ in Equation (13) is referred to as “Kalman gain”, and is computed as:

$$\mathbf{K}_t = \mathbf{C}_f \mathbf{H}^T (\mathbf{H} \mathbf{C}_f \mathbf{H}^T + \mathbf{R})^{-1} \quad (16)$$

Note that, in the context of parameter estimation addressed here, we are interested in the update of the spatial distribution of permeability, rather than the fluid pressure distribution within the model domain. As a result, the updated ensemble of log-transformed caprock permeability is extracted from the update state as $\mathbf{Y}_u = \mathbf{X}_u(n_f - n + 1:n_f, 1:N_{MC})$. The mean of the posterior ensemble of the log-transformed caprock permeability is computed as $\bar{\mathbf{Y}}_u = \mathbf{Y}_u \bar{\mathbf{1}}_{N_{MC} \times 1}$, where $\bar{\mathbf{1}}_{N_{MC} \times 1}$ is a $N_{MC} \times 1$ vector with all components equal to $1/N_{MC}$.

2.2.4 Assessment of ES and REnKF performance

To evaluate the performance of the ES and the REnKF, “synthetic” fluid-pressure data are collected from a synthetic “true” reference state ϕ_{ref} . The reference state is obtained with the multiphase flow model (section 2.1) using a corresponding “true” permeability distribution $\mathbf{Y}_{ref}$. The purpose of the “true” reference system is twofold: i) simulate available fluid pressure data (from ϕ_{ref}); and ii) evaluate the performance of the DA algorithm, which is achieved by computing and comparing the deviations from the “true” permeability distribution of the forecasted and updated parameter ensembles. Such comparison allows for determining the degree to which the forecast is corrected by data assimilation. This correction is quantified by the two associated global parameters AE (absolute error) and AES (average ensemble spread) [26], which are defined as:

$$AE(\mathbf{X}) = \frac{1}{N_{MC}n} \sum_{j=1}^{N_{MC}} \sum_{i=1}^n |X_{i,j} - X_{i,ref}| \quad (17)$$

$$AES(\mathbf{X}) = \frac{1}{N_{MC}n} \sum_{j=1}^{N_{MC}} \sum_{i=1}^n |X_{i,j} - \bar{X}_i| \quad (18)$$

In Equation (17), $X_{i,ref}$ is the true state variable at location i , $X_{i,j}$ is the state variable at location i and realization j , and $\bar{X}_i$ is the average over all realizations of the state variable at location i . The absolute error AE measures the deviation between the reference value and the updated ensemble. On the other

hand, AES measures the deviation between the updated ensemble and its mean. Lower values of AE correspond to a model state approaching the reference state; whereas lower values of AES signify reduced uncertainty in model parameters.

For a visual assessment of the data assimilation performance, the maps of the updated ensemble median (50th percentile) and maps of the 10th percentile distributions may be plotted and compared to the reference state $\mathbf{X}_{ref}$ (i.e. $\boldsymbol{\phi}_{ref}$ or $\mathbf{Y}_{ref}$) map. In the case of $\mathbf{Y}$, 10th percentile maps show permeability values that have a 90% probability of being exceeded. Where these values are "high" (on the order typical of aquifer permeabilities) there is a high chance to find leaky areas of the caprock. Also, maps of the local ensemble spread $AES_i = \sum_{j=1}^{N_{MC}} |Y_{i,j} - \bar{Y}_i| / N_{MC}$ ($i=1,2,\dots,n$) may be useful to analyze the spatial variability of the updated parameter ensemble and assess the level of confidence achieved by data assimilation.

3 Numerical experiments

This section presents the results of numerical tests performed to analyze the effectiveness of the proposed methodologies. First, we introduce the model setup, then we present two scenarios in which the ES and the REnKF are applied and, last, we show and discuss the obtained results.

3.1 Model setup

The methodology is applied to a hypothetical geological system consisting of a 100-m thick saline aquifer, denoted as "Aquifer 1", which constitutes the target CO₂ storage formation, overlain by another 100-m thick aquifer, denoted as "Aquifer 2". A low permeable 30-m thick formation, representing the caprock, separates the two aquifers. The bottom of Aquifer 1 lies at a depth of 1,600 m, whereas the top of Aquifer 2 lies at a depth of 1,370 m. The extension of the 3D domain is 4,100 m × 4,100 m × 230 m. Figure 3 shows a plan view and a cross section of the model domain. Within the model grid, the two aquifers and the caprock are each divided vertically into 4 layers for a total of 12 layers, which are enumerated in ascending order starting from the bottom.

Horizontally, the model domain is uniformly discretized with $100 \text{ m} \times 100 \text{ m}$ grid blocks. Overall, 20,172 cells ($41 \times 41 \times 12$) make up the model grid. A single CO_2 injection vertical well is considered at the center of the model domain and screened within Aquifer 1. The resolution of the finite-difference grid in the area surrounding this well is increased in order to better represent steep horizontal pressure gradients and obtain an appropriate grid size for a well ($\sim 0.5 \text{ m}$). The suitability of the grid adopted in our numerical investigations is discussed in section 3.5.3.

The CO_2 injection well is assumed to operate at a constant volumetric rate of $1,395,575 \text{ sm}^3/\text{d}$, corresponding to $946,080 \text{ ton/yr}$, for an indefinite period of time. For demonstrative purposes, we focus on assimilating fluid pressure data collected during the first two years of operation.

Figure 3. (a) Plan view of Aquifer 2 with detail of the grid, and (b) Vertical cross section of the domain with grid sub-layers in each formation. Measurement locations of Aquifer 2 are indicated as black dots. A fully penetrating injection well is located at the center of the domain in Aquifer 1. Note that vertical cross section is vertically exaggerated. Gray cells denote the cells in which pore-volume is modified to account for boundary conditions.

Initially, all formations are assumed saturated with only brine (with salinity equal to 0.2 kg salt/kg solution) and under hydrostatic pressure conditions. To simulate a laterally infinite aquifer system, the pore-volume of boundary cells is multiplied by a factor of 10^6 . Van Genuchten's model [40] is used to calculate capillary pressure and relative permeability of CO₂ and brine, using data from Zhou et al. [54]. Figure 4 illustrates k_{rg} , k_{rw} , and p_c as functions of s_w based on such model. Neither capillary heterogeneity nor hysteresis is considered in this study. All simulations are performed using a homogeneous porosity distribution. This is set to 0.15 for Aquifers 1 and 2, and 0.10 for the caprock.

Figure 4. (a) Relative permeability curves of CO₂ (dashed line) and brine (solid line) and (b) capillary pressure as function of brine saturation.

Fluid-pressure data are assumed to be available at the $n_m = 49$ nodes of a uniform 7×7 square grid, whose locations are indicated in Figure 3. These data are collected in proximity of the bottom of Aquifer 2 (layer 9) every three months, for a total of $N_t = 8$ measurement times over two years. Therefore, in the ES a total of $m = 392$ measurements (49×8) are assimilated into the simulation model response simultaneously, whereas in the REnKF $m = 49$ measurements are assimilated at each of the $N_t = 8$ observation times. Fluid pressure measurements are characterized by a Gaussian noise with a standard deviation of 0.1 bar (1×10^4 Pa).

In this work, the uncertain parameters of interests consist of the spatial distribution of permeability in the caprock separating Aquifers 1 and 2. The geostatistical parameters $(P_1, P_2, \lambda_1, \lambda_2, \mu_{Y_1}, \mu_{Y_2}, \sigma'_1, \sigma'_2)$ identifying the bimodal geostatistical model used to simulate the spatial distribution of the caprock and permeability variability are given in Table 1. Table 1 includes parameters sets used to generate both the forecast ensemble and the reference system for the caprock permeability. Note that these sets are different since in practical instances the geostatistics of the true system are unknown.

Table 1. Parameters used to generate the facies distribution and permeability variability at the caprock for the reference (“true”) system and the forecast ensemble. Units of caprock permeability are in m^2 .

	Category	Facies	Covariance Type	μ_Y	σ'	P (/)	λ (m)
Reference	1	Caprock	Exponential	-20	0.2	0.99	1,000
	2	Weak areas	Exponential	-13	0.2	0.01	100
Forecast Ensemble	1	Caprock	Exponential	-21	0.3	0.98	1,500
	2	Weak areas	Exponential	-12	0.3	0.02	200

To evaluate the performance of the methodology to estimate the caprock permeability, two scenarios are considered for the reference system. The two scenarios differ for the permeability assumed in Aquifers 1 and 2. In the first scenario (scenario 1), permeability of both aquifers is assumed homogeneous and isotropic, with a deterministic value equal to $1 \times 10^{-13} \text{ m}^2$. In the second scenario (scenario 2), aquifers' permeabilities $k_{aq,i}$ ($i=1,2$) are assumed heterogeneous and fitting a 3D stationary Gaussian process [53] in a log-transformed space. The following anisotropic exponential covariance model characterizes this process:

$$\log k_{aq,i} = Y_{aq,i} = N(\mu_{Y_{aq,i}}; \sigma_{Y_{aq,i}}) \quad (i = 1,2) \quad (19)$$

$$C_{Y_{aq,i}Y_{aq,i}}(\mathbf{d}; \sigma_{Y_{aq,i}}^2, \lambda_1, \lambda_2, \lambda_3) = \sigma_{Y_{aq,i}}^2 \exp\left(-\sqrt{\sum_{j=1}^3 \frac{d_j^2}{\lambda_j^2}}\right) \quad (20)$$

where: $\mu_{Y_{aq,i}}$ and $\sigma_{Y_{aq,i}}^2$ are the mean and variance of the log-transformed aquifer permeability; d_1 , d_2 , and d_3 are the components of the distance vector $\mathbf{d}$ along the coordinate directions x , y , and z , respectively; and λ_1 , λ_2 and λ_3 are the corresponding spatial correlation scales. The parameters $(\mu_{Y_{aq,i}}, \sigma_{Y_{aq,i}}^2, \lambda_1, \lambda_2, \lambda_3)$ of this process are given in Table 2. Note that for both aquifers λ_3 is equal to 500 m, which is 20 times larger than the thickness of aquifers' sublayers (Figure 3b), thus low variability of the permeability in the vertical direction is expected. In scenario 2, aquifers' permeability fields are generated using a 3D sequential Gaussian simple Kriging algorithm [55].

In the forecast simulations, permeabilities of both aquifers are considered deterministic, homogeneous and equal to $1 \times 10^{-13} \text{ m}^2$. Therefore $Y_{aq,1}$ and $Y_{aq,2}$ are not included in the parameter ensemble $\mathbf{P}$. The aquifers' permeabilities in the coordinate direction z are assumed to be 1/10 of their respective permeability in the horizontal directions x and y .

Table 2. Lognormal PDF parameters used to generate the permeabilities of the target storage formation (Aquifer 1) and the overlying aquifer (Aquifer 2) of the reference ("true") system. Aquifer permeability units are m^2 .

Scenario	Covariance type	$\mu_{Y_{aq}}$	$\sigma_{Y_{aq}}^2$	$\lambda_x = \lambda_1$ (m)	$\lambda_y = \lambda_2$ (m)	$\lambda_z = \lambda_3$ (m)
1	-	-13	0	-	-	-
2	Exponential	-13	0.25	1,000	1,000	500

In both scenarios 1 and 2, the spatial distribution of the caprock Y for the "true" reference system (Table 1) is the same. This distribution is mapped in Figure 5. Since the caprock permeability does not vary vertically, Figure 5 is representative of any of the four layers used to discretize the caprock (Figure 3b). This map reveals the presence of three distinct potential leakage pathways, which are referred as A, B and C hereafter. Figure 6 shows the Y_{aq} maps for Aquifer 1 (layer 1) and Aquifer 2 (layer 9) for the "true" reference system in scenario 2. Note that in this, and the following maps representing aquifers or caprock

properties, boundary cells in which the pore volume is modified to represent infinite aquifer or caprock
are omitted.

Figure 5. “True” reference spatial distribution of Y in the caprock for the scenarios 1 and 2. Units of caprock permeability are m^2 . Measurement locations are indicated with white dots.

Figure 6. “True” reference spatial distributions of Y_{aq} of (a) Aquifer 1 at layer 4, and (b) Aquifer 2 at layer 9 for scenario 2. Units of aquifer permeability are m^2 .

The temporal evolutions of the fluid pressure distribution within the model domain for the “true” reference systems in scenarios 1 and 2 are obtained with multiphase flow ECLIPSE simulations making use of the corresponding “true” reference fields. Figure 7 and Figure 8 show the reference distributions of fluid overpressure and CO₂ saturation, respectively, at $t = 2$ years for scenarios 1 and 2.

Figure 7. Reference fluid pressure change (in bar) distribution in proximity of the bottom of Aquifer 2 (layer 9) at $t = 2$ years for (a) scenario 1 and (b) scenario 2. Measurement locations are indicated with white dots.

Figure 8. Reference CO₂ saturation distribution in proximity of the top of Aquifer 1 (layer 4) at $t = 2$ years for (a) scenario 1 and (b) scenario 2.

Figure 7a and Figure 7b show the “true” spatial distributions of fluid overpressure Δp (Equation (8)) at $t = 2$ years in proximity of the bottom of Aquifer 2 (layer 9) where measurements are taken, for scenarios 1 and 2, respectively. The Δp pulse is more contained in scenario 2 than in scenario 1 as a consequence of the aquifers’ heterogeneity, which prevents the spread of the pressure in areas away from the injection well. Values of Δp close to 1 bar can be observed in scenario 1 over the leaky area A, whereas in scenario 2 the maximum Δp values are around 0.7 bar over the leaky area B. Both scenarios present fluid pressure changes above the three leaky areas indicating that brine leakage is occurring. Pressure changes over the leaky area C are generally smaller in magnitude, which can be explained as its area is smaller than A and B’s (Figure 5).

Figure 8a and Figure 8b show the reference distribution of the CO₂ plume. These maps emphasize that by monitoring early pressure changes in Aquifer 2 one can infer the presence of potential leakage pathways before CO₂ actually reaches them. At $t = 2$ years CO₂ plume has not yet invaded any of the leaky areas in both scenarios 1 or 2.

3.2 Model forecast

In this stage, the forecast ensemble $\boldsymbol{\phi} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_{N_{MC}}]$ of fluid pressure distributions in the model domain is computed with a Monte Carlo multiphase flow ECLIPSE [38] simulation using the generated ensemble $\boldsymbol{P} = [\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{N_{MC}}]$ of realizations of the caprock Y field. The results of this stochastic simulation are used to assemble the forecast state-parameter matrix $\boldsymbol{X}_f = [\boldsymbol{\phi} \boldsymbol{P}]^T$.

As previously commented, ensemble-based KF algorithms (Equations (13)-(14)) provide an optimal solution to Equation (9) under the hypotheses of linear systems with state variables conforming to multi-Gaussian joint distributions, and unbiased, uncorrelated and normal-distributed measurement errors. Since in the considered case the multiphase flow model (section 2.1) is non-linear, the ES and REnKF solutions can be sub-optimal. This sub-optimality could be partially reduced if all requirements of the KF

other than model linearity are met. It is thus suitable to require that: (i) the state and parameter ensembles in the matrix $\mathbf{X}_f$ be Gaussian distributed; (ii) all fluid-pressure measurement errors be unbiased and normal distributed (see section 2.2.3). To meet these conditions, adequate transformations may be applied to the parameters and the state variables that are not Gaussian.

Here, since the caprock permeability is simulated as a bimodal random process, a NST algorithm is employed to transform permeability values to fit to a Gaussian PDF. A modified NST based on the works of Zhou et al. [34] and Crestani et al. [33] is here proposed to transform caprock Y values. The “local” NST used by Zhou et al. [34] builds a cumulative distribution function (CDF) for each model cell and each observation time for the system state, while the CDF for the system parameter remain the same for all assimilation times. Crestani et al. [33] show that a “global” NST, with a CDF composed by model results at all cells and observation times, may provide better results as it avoids the disruption of the parameter-state cross-correlation structure occurring when a different NST is applied for each cell.

In this study, a single global NST is applied in combination with the ES. In the case of the REnKF, the global NST is instead recalculated after each data assimilation time, to account for the newly updated Y field. Both NST approaches are based on a sample CDF defined by all simulated $Y_{i,j}$ values of each cell ($i=1,2,\dots,n$) and all realizations of the ensemble ($j=1,2,\dots,N_{MC}$). The size of the sample used to build this CDF is $n \cdot N_{MC}$. After ordering the $Y_{i,j}$ values in ascending order, $Y_1 < Y_2 < \dots < Y_{n \cdot N_{MC}}$, the corresponding CDF values are calculated as $CDF(Y_l) = (l - 0.5)/(n \cdot N_{MC})$ [56], where $l=1,2,\dots,n \cdot N_{MC}$. For both the ES and the REnKF, after the update of state and parameters, the inverse of the adopted NST needs to be applied to the parameter ensemble in order to retrieve the updated spatial distribution of k in the caprock.

In the considered problem, the state variables of interest, that is, the Δp (Equation (8)) distribution within the model domain, require no transformation since a fit analysis indicates that this variable is normal-distributed. This fit analysis is based on a linear regression of the dataset $\left[f_{\Phi}(\Delta p); \Phi^{-1}\left(\frac{l-0.5}{N_{MC}}\right) \right] (l = 1, 2, \dots, N_{MC})$, where Δp represents fluid-overpressure values at the

observation points in Aquifer 2 sampled by stochastic simulation, Φ^{-1} is the inverse of the theoretical CDF, numerically computed, and f_Φ is a function of Δp that depends on the type of CDF under consideration. The fit analysis is conducted by testing different types of CDF. The “goodness” of fit for the sampled distribution is assessed based on the accuracy with which the regression line approximates the sample, as quantified by the coefficient of determination R^2 and the Kolmogorov-Smirnov statistics KS . Values of R^2 close to 1 and low values of KS are indicators of a good fit.

Figure 9 shows the results of a fit test for Δp values sampled at a cell of Aquifer 2 at layer 9 and observation time $t = 0.25$ years when Φ is chosen as a Gaussian CDF. For this example, values of $R^2 = 0.975$ and $KS = 0.063$ are obtained, indicating a substantially Gaussian behavior. One must be aware, however, that, whenever possible, transformations of parameters and/or states can at most meet conditions of “marginal” Gaussianity for each of the state variables, but do not guarantee that states and parameters fit a multi-Gaussian joint distribution, as the KF hypotheses would require. Therefore applying transformations to states and parameters does not always improve the optimality of the ES and the REnKF solutions.

Figure 9. Inverse CDF of the fluid pressure change in one cell of Aquifer 2 at observation time $t = 0.25$ years, showing its Gaussian distributed behavior.

Table 3 summarizes the numerical tests that are considered in this work. The ES and the REnKF are applied with NST of the forecast parameter ensemble $\mathbf{P}$. These tests, for both scenarios 1 and 2, are termed I and II and correspond to the application of the ES-NST and the REnKF-NST, respectively. Preliminary data assimilation experiments are conducted to evaluate a sufficient ensemble size. Based on these tests, a sample size $N_{MC}=1,000$ is adopted for the ES, whereas a sample size $N_{MC}=200$ is used for the REnKF. These values of N_{MC} are large enough to provide a fairly accurate representation of the cross-covariance of model parameters and state variables at measurement locations.

The adoption of an ensemble size for the ES larger than that for the REnKF stems from the idea of having a similar ratio between the ensemble size N_{MC} and the number of assimilated measurements m . In both the REnKF and the ES, the matrix $\mathbf{H}\mathbf{C}_f\mathbf{H}^T + \mathbf{R}$ in Equation (16) is calculated using a finite ensemble of realizations and has rank m , which equals 49 for the REnKF, and 392 for the ES (section 2.2.3). Therefore, one needs to account for the fact that in order to sample $\mathbf{H}\mathbf{C}_f\mathbf{H}^T + \mathbf{R}$ with comparable levels of accuracy, the ensemble size for the ES has to be considerably larger than that used in the REnKF.

Table 3. Scenario definition.

Scenario	Algorithm	Reference aquifers 1 and 2	Ensemble aquifers 1 and 2	N_{MC}
1-I	ES-NST	Homogeneous	Homogeneous	1,000
1-II	REnKF-NST	Homogeneous	Homogeneous	200
2-I	ES-NST	Heterogeneous	Homogeneous	1,000
2-II	REnKF-NST	Heterogeneous	Homogeneous	200

3.3 Model update: scenario 1

In scenario 1, both Aquifers 1 and 2 are homogeneous and isotropic. In scenario 1-I, the ES-NST is applied, whereas in scenario 1-II the REnKF-NST is applied. Figure 10 and Figure 11 summarize

547 results of these two scenarios at time $t = 2$ years for caprock permeability. Additionally, the maps of the
 548 updated ensemble median (50th percentile) of the state are also included. In both figures, subpanel (a)
 549 shows the map of the updated median value of the caprock Y , subpanel (b) displays the map of the 10th
 550 percentile of the caprock Y , subpanel (c) includes the map of the ensemble spread (section 2.2.4), and
 551 subpanel (d) shows the map of the updated median value of the fluid overpressure at the bottom of
 552 Aquifer 2 where fluid pressure is measured.

553 Figure 10. Results of scenario 1-I at $t = 2$ years. Maps of the updated caprock Y : (a) median (or 50th
 554 percentile), (b) 10th percentile, and (c) ensemble spread. Subpanel (d) shows the map of the updated fluid
 555 overpressure median in proximity of the bottom of Aquifer 2. Measurement locations are indicated with
 556 white dots. Units of caprock permeability are m^2 .

Figure 11a presents the map of the median caprock Y derived with the REnKF-NST (scenario 1-II). In this case, all three leaky areas A, B, and C are substantially identified. However, the sizes of the leakage areas are smaller in magnitude in comparison with the reference system (Figure 5). The permeability values of these areas are identified with fairly good accuracy. The map of the 10th percentile of the caprock Y obtained (Figure 11b) shows two areas with a value of Y greater than -12 ($k > 10^{-12} \text{ m}^2$), which have a high probability (90%) of being exceeded. These two areas correspond to leaky areas B and C. Figure 11c shows a contained spatial distribution of the ensemble-spread distribution with the highest values where leaky areas A and B are located in the reference system (Figure 5). In addition, the REnKF-NST exhibits high values (around 2) of the ensemble spread close to the actual leaky area C. Figure 11d displays the map of the spatial distribution of the updated changes in pressure at the bottom of Aquifer 2. In this figure, three different areas with greater pressure changes ($\Delta p > 0.4 \text{ bar}$) are shown, which correspond to the three leaky areas A, B, and C, respectively.

Figure 11. Results of scenario 1-II at $t = 2$ years. Maps of the updated caprock Y : (a) median (or 50th percentile), (b) 10th percentile, and (c) ensemble spread. Subpanel (d) shows the map of the updated fluid overpressure median in proximity of the bottom of Aquifer 2. Measurement locations are indicated as white dots. Units of caprock permeability are m^2 .

3.4 Model update: scenario 2

In scenario 2, the performances of the ES-NST and the RENKF-NST are investigated in a more realistic setting, which assumes Aquifers 1 and 2 as heterogeneous (see Figure 6). Figure 12 report results obtained at time $t = 2$ years with the RENKF-NST (scenario 2-II). In this figure, subpanel (a) illustrates the map of the updated median value of the caprock Y , subpanel (b) displays the map of the 10th percentile of the caprock Y , subpanel (c) shows the map of the ensemble spread (section 2.2.4), and

subpanel (d) shows the map of the updated median value of the fluid overpressure at the bottom of Aquifer 2. Results of scenario 2-I are not shown, since they are very similar to those observed in scenario 1-I, with the ES-NST not being able to locate any leaky area. No high probability (90%) area of leakage is found with the 10th percentile map. The ensemble-spread map only presents uncertainties around the leaky area B, with spread values between 1 and 1.5. The update of the fluid overpressure highlights that the only noticeable pressure changes are situated above the leaky area B, which is the largest.

On the other hand, Figure 12a shows that the REnKF-NST can detect caprock discontinuities of leaky areas B and C and also predict correctly their permeability values. However, as seen in Figure 11a, the REnKF-NST cannot accurately predict the size of these caprock discontinuities, probably due to the reduced amplitude of fluid pressure perturbations at measurement locations at $t = 2$ years, which is of the same order magnitude of the measurement noise. The 10th percentile map (Figure 12b) shows that a caprock discontinuity, located where area B is, has a 90% of probability of having permeability larger than 10^{-12} m^2 . The ensemble spread presented in Figure 12c shows that uncertainty is well contained with the exception of the three leaky areas A, B, and C, which have spread values slightly larger than 2 for leaky areas A and C, and around 4 for leaky area B. Figure 12d displays the distribution of the updated fluid overpressure at layer 9. Significant pressure changes are predicted close in Aquifer 2 above the leaky areas A, B, and C.

Figure 12. Results of scenario 2-II at $t = 2$ years. Maps of the updated caprock Y : (a) median (or 50th percentile), (b) 10th percentile, and (c) ensemble spread. Subpanel (d) shows the map of the updated fluid overpressure median in proximity of the bottom of Aquifer 2. Measurement locations are indicated with white dots. Units of caprock permeability are m^2 .

As injection proceeds in time, the intensity of overpressure observations become stronger and the effectiveness of the algorithm improves. Further numerical tests reveal that if the final simulation time is increased beyond $t = 2$ years, variation of fluid pressure in Aquifer 2 due to leakage become more substantial and the REnKF-NST is able to detect the three leaky areas with increased levels of confidence. Figure 13 shows the spatial distribution of the median of the caprock Y ensemble updated with the REnKF-NST at time $t = 2.25$ years (Figure 13a) and $t = 2.75$ years (Figure 13b) in scenario 2-II. Map in Figure 13a shows that caprock discontinuity at the leaky area A is now identified; however, the identified

location is south of the actual one. Also this map shows that the updated leaky area B is larger than in Figure 12a. At $t = 2.75$ years (Figure 13b), the REnKF-NST is able to better identify the location of leaky area A, which confirms that continued data collection is instrumental towards improving the identification of leakage areas. The comparison of scenarios 2-I (results not shown) and 2-II (Figure 12) suggests that the REnKF-NST is in general more effective than the ES-NST in identifying potential caprock leakage pathways.

Figure 13. Spatial distribution of updated caprock Y for scenario 2-II at: (a) $t = 2.25$ years and (b) $t = 2.75$ years. Measurement locations are indicated with white dots. Units of caprock permeability are m^2 .

3.5 Discussion

3.5.1 Comparing algorithms' performances

The low-quality results achieved with the ES-NST can be attributed to the fact that the algorithm updates the parameter ensemble in a single comprehensive data assimilation step, thus not letting the parameter ensemble evolve gradually at each of the assimilation times. The features of sequential data assimilation are better exploited when using the REnKF-NST. The REnKF-NST incorporates information of the observation updates backward in time and is able to steer the ensemble towards the reference field. In addition, the sequential assimilation applied with the REnKF-NST is somehow able to moderate the effects of non-linearity by performing a progressive update. In this study, the non-linearity lies in the

complex relationship between fluid pressure and caprock permeability. Similar conclusions about the sequential update process in comparison to the free model evolution of the ES are drawn in [33] and [57].

Figure 14 reports updates of the median caprock Y of scenarios 1-II and 2-II at different times. The main caprock discontinuity, leaky area B, is detected by the combined REnKF-NST early on at $t = 0.5$ years for scenarios 1-II (Figure 14a) and 2-II (Figure 14c). The REnKF-NST detects sooner this discontinuity because leaky area B is the largest and has the highest permeability, and thus produces the largest overpressure observations. Leaky area C is not detected by the REnKF-NST until $t = 0.75$ years for scenario 1-II (Figure 14b) and until $t = 1$ year for scenario 2-II (Figure 14d). Leaky area A is identified at $t = 1.75$ years in scenario 1-II (figure not shown here) and is not detected at all in scenario 2-II, at least within $t = 2$ years. However, as reported in Figure 13b, increasing the assimilation times to $N_t = 11$ (i.e., $t = 2.75$ years) allows for identifying the leaky area A even in scenario 2-II. Similar results have been observed by Nævdal et al. [19, 58] in reservoir history-matching experiments.

637 Figure 14. Spatial distribution of updated caprock Y for scenario 1-II at: (a) $t = 0.5$ and (b) $t = 0.75$ years;
 638 and for scenario 2-II at: (c) $t = 0.5$ years and (d) $t = 1$ years. Measurement locations are indicated with
 639 white dots. Units of caprock permeability are m^2 .

640 Figure 15 summarizes the performance assessment (Equations (17)-(18)) for the investigated data
 641 assimilation algorithms in scenarios 1 and 2. Figure 15a shows the AE for the fluid overpressure, whereas
 642 Figure 15b shows the AES for the fluid overpressure.

Figure 15. Performance assessment of scenarios 1 and 2 for: (a) $AE(\Delta p)$, (b) $AES(\Delta p)$. See Table 3 for description of scenarios 1 and 2.

In Figure 15a, the $AE(\Delta p)$ profiles decrease after the first assimilation time ($t = 0.25$ years) in all scenarios. Afterwards, $AE(\Delta p)$ for scenarios 1-I and 2-I remains constant, thus no further improvement is obtained. On the other hand, the $AE(\Delta p)$ profiles in scenarios 1-II and 2-II continue decreasing at each assimilation time. Similarly to Figure 15a, Figure 15b shows a similar trend of the $AES(\Delta p)$ profiles. In scenarios 1-I and 2-I, where the ES-NST is applied, uncertainty decreases only at the first assimilation time ($t = 0.25$ years) and no further improvement is observed afterwards. $AES(\Delta p)$ profiles of the REnKF-NST (scenarios 1-II and 2-II) continue decreasing until $t = 2$ years.

To further highlight the benefits of our approach, we compared maps of fluid pressure change in the proximity of the bottom of Aquifer 2 at $t = 3$ years for scenario 2-II in Figure 16. Figure 16a shows the map of fluid overpressure median obtained with the forecast caprock Y ensemble and no data assimilated. Figure 16b shows the fluid overpressure median produced by using the updated ensemble of caprock Y at $t = 2.75$ years. Comparison of these two figures with Figure 16c, which shows the corresponding reference fluid pressure change, shows the clear improvement brought about by data assimilation.

Figure 16. Fluid pressure change maps in proximity of the bottom of Aquifer 2 at $t = 3$ years for scenario 2: (a) Median produced by the initial forecast caprock Y ensemble fields (no update is applied), (b) Median produced by caprock Y ensemble updated values at $t = 2.75$ years (scenario 2-II), (c) Reference fluid pressure change. Measurement locations are indicated as white dots.

Scenario 2 presents the challenge that Aquifer 1 and Aquifer 2 are heterogeneous. Jafarpour and McLaughlin [59] remarked the importance of the geology represented by each members of the ensemble, that in order to obtain a good characterization realistic reproductions of the geology are needed. In scenario 2, the fact that both aquifers are heterogeneous and unknown has a negative effect on the characterization of the caprock through the data assimilation updating process, as observed in section 3.4.

To further understand results obtained and demonstrate the importance of the NST process, CDFs of caprock permeability before NST and after NST for the forecast and update steps are included in Figure 17. The top row of this figure displays CDF plots of the caprock Y in scenario 1-I, whereas the bottom row shows CDF plots in scenario 1-II, at $t = 2$ years. Figure 17a and Figure 17c include the CDF plot for the caprock permeability of the forecast (before applying NST process) and update (after applying the back-NST process) steps for scenarios 1-I and 1-II, respectively. Figure 17c and Figure 17d include the CDF plots after the NST methodology is applied for the forecast and update steps of scenarios 1-I and 1-II, respectively. These plots indicate that the NST and back-NST methods accomplish their goal: transform a non-Gaussian distribution to Gaussian, and transform back the Gaussian distribution to its original distribution. In addition, the forecast and update CDFs before and after applying the NST are practically overlapped.

Details of the CDFs between 0.95 and 1 of the caprock Y for scenarios 1-I and 1-II are shown in Figure 18. This figure also includes the CDF of the caprock Y for scenario 1-II at $t = 0.25$ years. Although the forecast and update of the caprock Y CDFs are very similar, Figure 18 illustrates the differences among CDFs between scenarios 1-I and 1-II and also between the two different time steps of scenario 1-II. CDF of scenario 1-I shows a higher probability of finding a caprock discontinuity in the ensemble, even if the ensemble is generated with the same parameters (Table 1). In addition, a slight difference for scenario 1-II at $t = 0.25$ and $t = 2$ years is found. At $t = 2$ years, the CDF of the caprock Y shows a slightly greater probability of finding a caprock discontinuity, even though the difference is minor.

687 Figure 17. CDFs of caprock Y at $t = 2$ years. Scenario 1-I: (a) CDFs of caprock Y of the forecast and
 688 update steps, (b) CDFs of the transformed caprock Y of the forecast and update steps. Scenario 1-II: (c)
 689 CDFs of caprock Y of the forecast and update steps, (d) CDFs of the transformed caprock Y of the
 690 forecast and update steps.

Figure 18. Detail of caprock Y CDFs for scenarios 1-I at $t = 2$ years, 1-II at $t = 0$ years, 1-II at $t = 0.25$ years, and 1-II $t = 2$ years.

With regard to computational cost, the CPU running time for a single ECLIPSE simulation for the ES ($N_t = 8$) is under 2 minutes on a Unix server with 2xIntel Xeon Six core (3.47Ghz) with 40GB RAM. The running time for a single simulation for the REnKF is between 30 s and 2 minutes (corresponding to the first and last assimilation time, respectively). Thus, after the eight assimilation times, the REnKF total CPU time required is greater than the total CPU time of the ES. This is shown in Figure 19. The computational time of the data assimilation methodology is on the order of seconds.

Figure 19. CPU time of the ES and REnKF algorithms.

3.5.2 Model assumptions

The modeled geological setting described in section 3.1 and the hypothesized pressure monitoring network represent idealized conditions, which have been adopted only for the sake of comparing the performances of the ES and the REnKF. A monitoring network on a regular grid such as that shown in Figure 3 along with the hypothesized conditions of vertical homogeneity for formation permeability hypothesized are such that leakages could be simply detected by mapping overpressure data in Aquifer 2. In real-world scenarios however, vertical heterogeneities may be significant, and observation points (e.g. pre-existing boreholes) are often sparse and irregularly distributed, so that pressure monitoring alone may be not so informative unless the leakage spots are found in close proximity of monitoring locations. In these instances, the application of methodologies such as those presented here becomes a necessary step.

Another aspect worth of attention concerns the adopted grid resolution (section 3.1). In this work, $100 \text{ m} \times 100 \text{ m}$ horizontal grid cells have been adopted after preliminary numerical tests indicate that a sufficient model accuracy is achieved with respect to fluid overpressure in Aquifer 2, which is where hypothetical observations are made and where flow remains single phase until CO_2 escapes into it.

Note that this resolution may be rather coarse in the regions of Aquifer 1 where flow is two-phase and heterogeneity of the capillary pressure influences the CO₂ capillary trapping. Saadatpoor et al. [60] show that heterogeneity of the capillary pressure controls the CO₂ plume advancement and reduces CO₂ leaks. In most of our simulations, however, CO₂ leakage into Aquifer 2 is not observed within the simulated injection time ($t = 2$ years). Thus capillary pressure processes in Aquifer 1 seem to have a marginal effect on spatial distributions of fluid pressure in Aquifer 2.

In addition, fluid pressure forces in Aquifer 1 are more significant in comparison to capillary pressure forces during the CO₂ injection. Large fluid pressure forces are expected for the flow scenarios in comparison to the capillary pressure forces produced by sub-scale heterogeneity, which results a secondary concern for the purpose of this study. In essence, a grid finer than that shown in Figure 3 is likely to provide more accurate results on fluid pressure and capillary pressure especially in Aquifer 1 where two-phase flow occurs. However, the coarse grid model seems quite adequate to demonstrate the potential of the proposed methodology within an affordable computational cost.

In this study, hysteresis processes are not considered. The importance of hysteresis on CO₂ stored by capillary trapping is demonstrated in [61, 62] especially when modeling post-injection phase (or water injection) since re-imbibition phenomena are expected. Since this study only considers continuous CO₂ injection, hysteresis effects are deemed minor. Note that in instances where hysteresis effects are relevant higher grid resolutions are required, which may increase CPU times considerably and make our data assimilation approach computationally prohibitive.

Finally, it is worth mentioning that the positive results exhibited by the REnKF-NST are in part due to the assumption that the forecast ensemble of caprock permeability fits to a bimodal process similar to that characterizing the “true” reference system. The choice of a bimodal PDF relies on the common knowledge of caprock layers in sedimentary basins, which are typically low permeable layers characterized by more or less localized intrusions, discontinuities, and fractures that may constitute preferential pathways to fluid vertical flow. The nature of the NST is such that, when the updated caprock Y distributions in the NST space are transformed back to the original space, they get redistributed over to

the original forecast PDF. Potential intermediate permeability values are thus hardly found when the nugget variances σ_1' and σ_2' are much smaller than the absolute value of the difference between the averages μ_{Y_1} and μ_{Y_2} , as in our tests. Adopting a more complex forecast geostatistical model constitutes certainly a more rigorous approach, but this choice needs to be supported by available prior information. In practice, our approach relies on assuming a “simplistic” forecast bimodal PDF (Table 1) and let the data assimilation algorithm identify “equivalent” leakage regions, that is, weak spots having areal extent that, however inaccurate, make the product area $\times$ permeability match the overpressure observations. In the tests presented in sections 3.3 and 3.4 the areal extension of caprock discontinuities is found with relatively good accuracy also because the true reference caprock distribution is, at least in statistical terms, not very different from spatial distribution that conforms to the hypothesized forecast bimodal PDF. If this was not the case, one should expect to find the REnKF-NST still able to localize “equivalent” leakage spots in the sense explained above.

To summarize, our work is meant to present a first proof-of concept investigation to understand the strengths and the weaknesses of the ES and the REnKF in relation to fluid leakage detection in GCS systems. Future work needs to address more realistic model setups, accuracy issues related to grid resolution, effects of heterogeneity of other hydrogeological parameters such as the porosity, vertical heterogeneity of caprock permeability, and typical conditions in which little prior information is available to characterize the geostatistics of the storage formation and the confining caprock.

4 Summary and conclusions

In this work, we compared the capabilities of the ES and the REnKF algorithms to identify possible escape pathways of brine and CO₂ that could undermine the effectiveness of GCS systems. We developed and implemented a framework for the assimilation of fluid pressure data into multiphase models. We used a hypothetical 3D geological setup consisting of two deep saline aquifers separated by a sealing caprock formation. Supercritical CO₂ is injected into the deepest aquifer. The caprock includes areas of low permeability (sealing formation) and high permeability (leaky pathways) generated with a

sequential indicator Kriging simulator based on a bimodal PDF. Taking advantage of the fact that fluid pressure travels and propagates faster than CO₂ plume, measurements of fluid pressure at the overlying aquifer were used to identify and locate these discontinuities.

A subsurface multiphase flow model and two ensemble-based data assimilation algorithms were combined to characterize the caprock discontinuities. The assimilated data consisted of fluid overpressures measured at several locations of the aquifer above the caprock and taken with regular frequency over a prescribed period of time. A NST technique was applied to transform caprock Y bimodal PDFs into Gaussian. The ES and the REnKF methods were applied to two different virtual reference scenarios: in one in which permeability of both aquifers is considered homogeneous, and another in which these permeabilities are heterogeneous.

Results show that the ES performs poorly as this method relies on a single update of the system. On the other hand, the REnKF proves much more effective tool for inverting pressure data in order to detect potential leakage pathway in the geological storage of supercritical CO₂. In the case of heterogeneous aquifers, several assimilation times are necessary to improve the detection of caprock.

The use of the REnKF allows for a successful identification of caprock discontinuities through fluid pressure data assimilation, which is instrumental to simulate and the prediction the migration of the CO₂ plume in the injected formation and detect early-on potential leakage pathways. This can be used to make decisions regarding injection rates and injection/production schemes and strategies to improve CO₂ storage and reduce brine/CO₂ leakages. Future research to improve detection of the caprock discontinuities will include further study of more complex scenarios (e.g. the update of aquifers permeability). Also measurements of CO₂ saturations on the upper aquifer can be collected when CO₂ leakage has occurred, providing valuable information about to the system behavior. Simplifications to the model setup to achieve reasonable simulation times are a matter of great concern for the application of the proposed methodology. Future work will also need to investigate the importance of grid resolution, particularly in those instances where pressure observations are made in regions of the subsurface where CO₂ and brine phases coexist.

Acknowledgements

This research was supported by the U.S. Department of Energy, National Energy Technology Laboratory (DOE Grant FE0001830). The research conducted was also made possible with the support of Schlumberger Ltd., who kindly donated the reservoir simulation software suites PETREL and ECLIPSE, together with training opportunities to their use.

Appendix A. Categorical indicator Kriging simulator (CIKSIM)

We introduce the categorical indicator Kriging simulator (CIKSIM) algorithm developed and implemented to generate the ensemble of realization fields of permeability in the sealing formation. A more detailed description of CIKSIM is provided in [63]. This algorithm creates two types of facies: 1) sealing formation (areas with low permeability), and 2) inclusions (areas with high permeability) where CO₂ may leak.

Indicator Kriging (IK) is a geostatistical approach to geospatial modeling. IK is useful when the data are non-normally distributed, highly skewed, or essentially of categorical nature (e.g. types of vegetation). IK does not assume a normal distribution at unsampled locations; instead IK builds the cumulative distribution function (CDF) at each unsampled location with the information available from the neighborhood locations. IK needs a series of thresholds (or cutoff) values that comprise the data and uses them to build the CDF at the unsampled locations. For each threshold, IK transforms the neighborhood data to indicators and calculates their correlation based on an interpolation method such as simple Kriging (SK) [64] or ordinary Kriging (OK) [65]. With this information, IK estimates the probability of the variable at the unsampled location to be less or equal to the threshold. The final result of IK represents the uncertain distribution at the unsampled locations.

To sample a value at the unsampled locations, first, a conditional technique must be applied. IK uses an indicator transform. For a continuous random variable $X(u)$ at position u , the corresponding indicator transform can be defined as [66]:

$$\begin{aligned}
I(u, x_t) &= 1, \quad \text{if } X(u) \leq x_t \quad t = 1, \dots, n_t \\
&= 0 \quad \text{if } X(u) > x_t
\end{aligned}
\tag{A-1}$$

817 Similarly, for a discrete variable $K(u)$:

$$\begin{aligned}
I(u, k_t) &= 1, \quad \text{if } K(u) = k_t \quad t = 1, \dots, n_t \\
&= 0 \quad \text{otherwise}
\end{aligned}
\tag{A-2}$$

818 where: I is the indicator value, $X(u)$ or $K(u)$ is the value of the variable at position u , x_t and k_t are the
819 threshold or category value, and n_t is the total number of thresholds or categories. This indicator
820 transform has the characteristic that its expected value is equal to the cumulative probability of the
821 variable [67] and $I(u, x_t)$ is a random variable itself. That is to say, the IK method, instead of predicting
822 the $X(u)$ value, predicts $I(u, x_t)$ for the selected thresholds, and provides a least-squares estimate of the
823 conditional CDF at the threshold x_t [68]:

$$[I(u, x_t)]^* = E\{I(u, x_t|(n))\}^* = P^*\{X(u) \leq x_t|(n)\} \tag{A-3}$$

824 where: E indicates the expectancy of $I(u, x_t)$ conditioned to the information (n) available in the
825 neighborhood of location u , and P is the probability. Therefore $I(u, x_t)$ means the probability that
826 $X(u) \leq x_t$. The indicator value needs to be calculated at each threshold, which discretizes the interval of
827 variability of property $X(u)$. Depending on the value of $X(u)$, the indicator variable will take a value of 0
828 or 1 at each threshold, as Equation (A-1) states.

829 Once the indicator values are assigned at each threshold, the indicator values at the unsampled
830 locations can be estimated. To calculate the indicator values at the unsampled locations u_0 , a Kriging
831 algorithm is used. If SK [59] is used, for continuous and discrete variables the following equations are
832 written [69], respectively:

$$I^*(u_0, x_t) = w_0 + \sum_{j=1}^n w_j I(u_j, x_t) \tag{A-4}$$

$$I^*(u_0, k_t) = w_0 + \sum_{j=1}^n w_j I(u_j, k_t) \tag{A-5}$$

833 where: n is the number of sampled locations to calculate the indicator values and w are the SK weights
 834 corresponding to the thresholds. Since the indicator values represent the CDF, they must satisfy for a
 835 continuous variable:

$$I(u, x_1) \leq I(u, x_2), \quad \text{if } x_1 \leq x_2 \quad (\text{A-6})$$

836 and for a discrete variable:

$$\sum_{t=1}^{n_t} I(u, k_t) = 1 \quad (\text{A-7})$$

837 CIKSIM relies on a sequential conditional simulation method, in which unsampled locations on a
 838 prescribed grid are sequentially visited randomly until all locations are visited. The sequential simulation
 839 extends the conditioning to include all data available in the neighborhood of u , which includes the
 840 original data and values previously simulated. The sequential indicator simulation can be used for both
 841 continuous and discrete variables.

842 To generate the path to follow, a random number generator is used to create a sequence of
 843 random numbers corresponding to the total number of grid blocks. The sequence of the path is random to
 844 avoid artifacts resulting from a restricted search and a regular path. Based on the order, all unsampled
 845 locations are visited following the path. When an unsampled location is visited, a search neighborhood is
 846 applied where both the sampled points and previously simulated values are selected. To estimate the value
 847 of the unsampled location, SK is used. Therefore, for location $(n + 1)$ the probability of the variable
 848 $X(u_{n+1}) \leq x_t$ is:

$$[I(u_{n+1}, x_t)]_{SK}^* = CDF(u_{n+1}, x_t) + \mathbf{C}_{n+1,n}^T \mathbf{C}_{n,n}^{-1} [\mathbf{I}(u_n, x_t) - \mathbf{CDF}(u_n, x_t)] \quad (\text{A-8})$$

849 where: $[I(u_{n+1}, x_t)]_{SK}^*$ is the marginal probability of category x_t , $\mathbf{I}(u_n, x_t)$ is a column vector including
 850 the indicator value calculated at the locations previously simulated, $\mathbf{CDF}(u_n, x_t)$ is a column vector
 851 including the CDF value for threshold x_t calculated at the previous n locations, $\mathbf{C}_{n,n}$ is the covariance
 852 matrix for the n location previously simulated, $\mathbf{C}_{n+1,n}$ is the covariance matrix between the current

simulation location $n + 1$ and the n locations already simulated. Next, the simulated value is added to the data set and next location is simulated.

5 References

- [1] Birkholzer J, Q Zhou, C Tsang. Large-scale impact of CO₂ storage in deep saline aquifers: A sensitivity study on pressure response in stratified systems. *International Journal of Greenhouse Gas Control*. 3 (2009) 181-94, doi: 10.1016/j.ijggc.2008.08.002.
- [2] Kharaka Y, D Cole, S Hovorka, W Gunter, K Knauss, B Freifeld. Gas-water-rock interactions in Frio Formation following CO₂ injection: Implications for the storage of greenhouse gases in sedimentary basins. *Geology*. 34 (2006) 577-80.
- [3] Apps J, L Zheng, Y Zhang, T Xu, J Birkholzer. Evaluation of potential changes in groundwater quality in response to CO₂ leakage from deep geologic storage. *Transport in Porous Media*. 82 (2010) 215-46.
- [4] Metz B, Intergovernmental Panel on Climate Change. Working Group III. IPCC Special Report on Carbon Dioxide Capture and Storage. Cambridge University Press, for the Intergovernmental Panel on Climate Change, Cambridge, 2005.
- [5] DOE/NETL. Monitoring, Verification and Accounting of CO₂ Stored in Deep Geologic Formations. DOE/NETL-311/0815082009.
- [6] Seto CJ, GJ McRae. Reducing Risk in Basin Scale CO₂ Sequestration: A Framework for Integrated Monitoring Design. *Environ Sci Technol*. 45 (2011) 845-59, doi: 10.1021/es102240w.
- [7] Cole S, D Lumley, M Meadows, A Tura. Pressure and saturation inversion of 4D seismic data by rock physics forward modeling. 2002 SEG Annual Meeting2002.
- [8] Cole SP, DE Lumley, MA Meadows. Method of obtaining pore pressure and fluid saturation changes in subterranean reservoirs by forward modeling. Google Patents2006.
- [9] MacBeth C, M Floricich, J Soldo. Going quantitative with 4D seismic analysis. *Geophysical Prospecting*. 54 (2006) 303-17.
- [10] Chabora ER, SM Benson. Brine Displacement and Leakage Detection Using Pressure Measurements in Aquifers Overlying CO₂ Storage Reservoirs. *Energy Procedia*. 1 (2009) 2405-12, doi: <http://dx.doi.org/10.1016/j.egypro.2009.01.313>.
- [11] Zeidouni M, M Pooladi-Darvish, DW Keith. Leakage detection and characterization through pressure monitoring. *Energy Procedia*. 4 (2011) 3534-41, doi: <http://dx.doi.org/10.1016/j.egypro.2011.02.281>.
- [12] Nogue JP, JM Nordbotten, MA Celia. Detecting leakage of brine or CO₂ through abandoned wells in a geological sequestration operation using pressure monitoring wells. *Energy Procedia*. 4 (2011) 3620-7, doi: <http://dx.doi.org/10.1016/j.egypro.2011.02.292>.
- [13] Park YC, DG Huh, CH Park. A pressure-monitoring method to warn CO₂ leakage in geological storage sites. *Environ Earth Sci*. 67 (2012) 425-33, doi: DOI 10.1007/s12665-012-1667-2.
- [14] Sun AY, J-P Nicot. Inversion of pressure anomaly data for detecting leakage at geologic carbon sequestration sites. *Adv Water Resour*. 44 (2012) 20-9, doi: <http://dx.doi.org/10.1016/j.advwatres.2012.04.006>.

- 892 [15] Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng. 82 (1960) 35-
893 45.
- 894 [16] Hendricks Franssen H, W Kinzelbach. Ensemble Kalman filtering versus sequential self-calibration
895 for inverse modelling of dynamic groundwater flow systems. Journal of Hydrology. 365 (2009) 261-
896 74.
- 897 [17] Chen Y, D Zhang. Data assimilation for transient flow in geologic formations via ensemble Kalman
898 filter. Adv Water Resour. 29 (2006) 1107-22.
- 899 [18] Keppenne CL, MM Rienecker. Assimilation of temperature into an isopycnal ocean general
900 circulation model using a parallel ensemble Kalman filter. Journal of Marine Systems. 40 (2003) 363-
901 80.
- 902 [19] Nævdal G, T Mannseth, EH Vefring. Near-well reservoir monitoring through ensemble Kalman
903 filter. SPE/DOE Improved Oil Recovery Symposium 2002. pp. 13-7.
- 904 [20] Zhou H, L Li, H-J Hendricks Franssen, JJ Gómez-Hernández. Pattern Recognition in a Bimodal
905 Aquifer Using the Normal-Score Ensemble Kalman Filter. Math Geosci. 44 (2012) 169-85, doi:
906 10.1007/s11004-011-9372-3.
- 907 [21] Evensen G. Inverse methods and data assimilation in nonlinear ocean models. Physica D Nonlinear
908 phenomena. 77 (1994) 108-29.
- 909 [22] Evensen G. Data assimilation: the ensemble Kalman filter. Springer, 2009.
- 910 [23] Brusdal K, J Brankart, G Halberstadt, G Evensen, P Brasseur, P Van Leeuwen, et al. A
911 demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of
912 operational ocean forecasting systems. Journal of Marine Systems. 40 (2003) 253-89.
- 913 [24] Dunne S, D Entekhabi. Land surface state and flux estimation using the ensemble Kalman smoother
914 during the Southern Great Plains 1997 field experiment. Water Resour Res. 42 (2006).
- 915 [25] Ngodock HE, GA Jacobs, MS Chen. The representer method, the ensemble Kalman filter and the
916 ensemble Kalman smoother: A comparison study using a nonlinear reduced gravity ocean model.
917 Ocean Model. 12 (2006) 378-400, doi: DOI 10.1016/j.ocemod.2005.08.001.
- 918 [26] Hendricks Franssen H, W Kinzelbach. Real-time groundwater flow modeling with the Ensemble
919 Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water
920 Resour Res. 44 (2008).
- 921 [27] Li L, H Zhou, JJ Gómez-Hernández, H-J Hendricks Franssen. Jointly mapping hydraulic
922 conductivity and porosity by assimilating concentration data via ensemble Kalman filter. Journal of
923 Hydrology. 428-429 (2012) 152-69, doi: <http://dx.doi.org/10.1016/j.jhydrol.2012.01.037>.
- 924 [28] Skjervheim J, G Evensen. An ensemble smoother for assisted history matching. SPE Reservoir
925 Simulation Symposium 2011.
- 926 [29] Bailey RT, D Baù. Estimating geostatistical parameters and spatially-variable hydraulic conductivity
927 within a catchment system using an ensemble smoother. Hydrol Earth Syst Sci. 16 (2012) 287-304,
928 doi: 10.5194/hess-16-287-2012.
- 929 [30] Herrera GS, R Simuta-Champo. Optimal Design of Groundwater-Quality Sampling-Networks with
930 3D Selection of Sampling Locations Using an Ensemble Smoother. Journal of water resources
931 planning and management. 1 (2012) 169.
- 932 [31] Gu Y, DS Oliver. The ensemble Kalman filter for continuous updating of reservoir simulation
933 models. Transactions-American society of mechanical engineers journal of energy resources
934 technology. 128 (2006) 79.

- 935 [32] Goovaerts P. Geostatistics for natural resources evaluation. Oxford university press, 1997.
- 936 [33] Crestani E, M Camporese, D Baú, P Salandin. Ensemble Kalman filter versus ensemble smoother for
 937 assessing hydraulic conductivity via tracer test data assimilation. *Hydrol Earth Syst Sci.* 17 (2013)
 938 1517-31.
- 939 [34] Zhou H, JJ Gómez-Hernández, H-J Hendricks Franssen, L Li. An approach to handling non-
 940 Gaussianity of parameters and state variables in ensemble Kalman filtering. *Adv Water Resour.* 34
 941 (2011) 844-64, doi: 10.1016/j.advwatres.2011.04.014.
- 942 [35] Simon E, L Bertino. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled
 943 physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Science.* 5
 944 (2009) 495-510.
- 945 [36] Béal D, P Brasseur, J-M Brankart, Y Ourmières, J Verron. Characterization of mixing errors in a
 946 coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation
 947 using Gaussian anamorphosis. *Ocean Science.* 6 (2010) 247-62.
- 948 [37] Schöniger A, W Nowak, HJ Hendricks Franssen. Parameter estimation by ensemble Kalman filters
 949 with transformed data: Approach and application to hydraulic tomography. *Water Resour Res.* 48
 950 (2012) W04502, doi: 10.1029/2011wr010462.
- 951 [38] Schlumberger. Eclipse technical description, v. 2010.1, Report, Abingdon, U.K. (2010).
- 952 [39] Trangenstein JA, JB Bell. Mathematical structure of compositional reservoir simulation. *SIAM*
 953 *journal on scientific and statistical computing.* 10 (1989) 817-45.
- 954 [40] Van Genuchten MT. A Closed-form Equation for Predicting the Hydraulic Conductivity of
 955 Unsaturated Soils1. *Soil Science Society of America Journal.* 44 (1980) 892, doi:
 956 10.2136/sssaj1980.03615995004400050002x.
- 957 [41] Brooks R, A Corey. 1964, Hydraulic properties of porous media, Hydrology Paper No. 3. Colorado
 958 State University, Fort Collins, CO. (1964).
- 959 [42] Spycher N, K Pruess. CO₂-H₂O mixtures in the geological sequestration of CO₂. II. Partitioning in
 960 chloride brines at 12–100°C and up to 600 bar. *Geochimica et Cosmochimica Acta.* 69 (2005) 3309-
 961 20, doi: <http://dx.doi.org/10.1016/j.gca.2005.01.015>.
- 962 [43] Spycher N, K Pruess. A phase-partitioning model for CO₂-brine mixtures at elevated temperatures
 963 and pressures: application to CO₂-enhanced geothermal systems. *Transport in Porous Media.* 82
 964 (2010) 173-96.
- 965 [44] Redlich O, JNS Kwong. On the Thermodynamics of Solutions. V. An Equation of State. Fugacities
 966 of Gaseous Solutions. *Chemical Reviews.* 44 (1949) 233-44, doi: 10.1021/cr60137a013.
- 967 [45] Vesovic V, W Wakeham, G Olchoway, J Sengers, J Watson, J Millat. The transport properties of
 968 carbon dioxide. *Journal of Physical and Chemical Reference Data.* 19 (1990) 763-808.
- 969 [46] Fenghour A, WA Wakeham, V Vesovic. The viscosity of carbon dioxide. *Journal of Physical and*
 970 *Chemical Reference Data.* 27 (1998) 31-44.
- 971 [47] Vinsome P. Orthomin an Iterative Method for Solving Sparse Sets of Simultaneous Linear
 972 Equations. SPE Symposium on Numerical Simulation of Reservoir Performance. Society of
 973 Petroleum Engineers 1976.
- 974 [48] Bishop A. The principle of effective stress. *Teknisk Ukeblad.* 106 (1959) 859-63.
- 975 [49] Gray W, C Miller. Consistent thermodynamic formulations for multiscale hydrologic systems: Fluid
 976 pressures. *Water Resour Res.* 43 (2007).

- 977 [50] van Leeuwen PJ, G Evensen. Data Assimilation and Inverse Methods in Terms of a Probabilistic
 978 Formulation. *Mon Weather Rev.* 124 (1996) 2898-913, doi: 10.1175/1520-
 979 0493(1996)124<2898:daaimi>2.0.co;2.
- 980 [51] Wen X-H, W Chen. Real-time reservoir model updating using ensemble Kalman filter. *SPE*
 981 *reservoir simulation symposium*. Society of Petroleum Engineers2005.
- 982 [52] Nowak W. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator.
 983 *Water Resour Res.* 45 (2009).
- 984 [53] de Marsily G. Quantitative hydrogeology groundwater hydrology for engineers. Academic Press,
 985 Orlando, FL, 1986.
- 986 [54] Zhou Q, JT Birkholzer, E Mehnert, Y-F Lin, K Zhang. Modeling Basin- and Plume-Scale Processes
 987 of CO₂ Storage for Full-Scale Deployment. *Ground Water.* 48 (2009) 494-514, doi: 10.1111/j.1745-
 988 6584.2009.00657.x.
- 989 [55] Baù DA, AS Mayer. Optimal design of pump-and-treat systems under uncertain hydraulic
 990 conductivity and plume distribution. *J Contam Hydrol.* 100 (2008) 30-46.
- 991 [56] Hahn GJ. Statistical models in engineering. Wiley, New York, 1967.
- 992 [57] Evensen G, PJ van Leeuwen. An ensemble Kalman smoother for nonlinear dynamics. *Mon Weather*
 993 *Rev.* 128 (2000) 1852-67, doi: Doi 10.1175/1520-0493(2000)128<1852:Aeksfn>2.0.Co;2.
- 994 [58] Nævdal G, LM Johnsen, SI Aanonsen, EH Vefring. Reservoir monitoring and continuous model
 995 updating using ensemble Kalman filter. *SPE Journal.* 10 (2005) 66-74.
- 996 [59] Jafarpour B, D McLaughlin. Estimating channelized-reservoir permeabilities with the ensemble
 997 Kalman filter: the importance of ensemble design. *SPE Journal.* 14 (2009) 374-88.
- 998 [60] Saadatpoor E, SL Bryant, K Sepehrmoori. New trapping mechanism in carbon sequestration.
 999 *Transport in Porous Media.* 82 (2010) 3-17.
- 1000 [61] Flett M, R Gurton, I Taggart. The function of gas-water relative permeability hysteresis in the
 1001 sequestration of carbon dioxide in saline formations. *SPE Asia Pacific Oil and Gas Conference and*
 1002 *Exhibition*. Society of Petroleum Engineers2004.
- 1003 [62] Juanes R, EJ Spiteri, FM Orr, MJ Blunt. Impact of relative permeability hysteresis on geological
 1004 CO₂ storage. *Water Resour Res.* 42 (2006) n/a-n/a, doi: 10.1029/2005wr004806.
- 1005 [63] González-Nicolás Álvarez A. Methodologies to detect leakages from geological carbon storage sites.
 1006 Fort Collins, CO: Colorado State University; 2014.
- 1007 [64] Krige DG. A statistical approach to some mine valuations and allied problems on the Witwatersrand.
 1008 Johannesburg: M.Sc. Eng. Thesis of University of Witwatersrand; 1951.
- 1009 [65] Matheron G. Principles of geostatistics. *Economic Geology.* 58 (1963) 1246-66, doi:
 1010 10.2113/gsecongeo.58.8.1246.
- 1011 [66] Journel, A G., Alabert, F. Non-Gaussian data expansion in the Earth sciences. *Terra Nova.* 1 (1989)
 1012 123-34.
- 1013 [67] Olea RA. Geostatistics for Engineers and Earth Scientists. Kluwer Academic, 1999.
- 1014 [68] Deutsch CV, AG Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford
 1015 University Press, USA, 1997.
- 1016 [69] Kelkar M, G Perez, A Chopra. *Applied Geostatistics for Reservoir Characterization*. Society of
 1017 Petroleum Engineers, 2002.