Direct Comparison of Numerical Simulations and Experiments of CO$_2$ Injection and Migration in Geologic Media: Value of Local Data and Forecasting Capability

Lluís Saló-Salgado[1,2] · **Malin Haugen**[3] · **Kristoffer Eikehaug**[3] · **Martin Fernø**[3,5] · **Jan M. Nordbotten**[4,5] · **Ruben Juanes**[1,2]

Abstract

The accuracy and robustness of numerical models of geologic CO$_2$ sequestration are almost never quantified with respect to direct observations that provide a ground truth. Here, we conduct CO$_2$ injection experiments in meter-scale, quasi-2D tanks with porous media representing stratigraphic sections of the subsurface, and compare them to numerical simulations of those experiments. We evaluate (1) the value of prior knowledge of the system, expressed in terms of ex situ measurements of the tank sands' multiphase flow properties (local data), with respect to simulation accuracy; and (2) the forecasting capability of history-matched numerical models, when applied to different settings. We match three versions of a numerical simulation model—each with access to an increasing level of local data—to a CO$_2$ injection experiment in Tank 1 ($89.7 \times 47 \times 1.05$ cm). Matching is based on a quantitative comparison of CO$_2$ migration at different times from timelapse image analysis. Next, use the matched models to make a forecast of a different injection scenario in Tank 1 and, finally, a different injection scenario in Tank 2 ($2.86 \times 1.3 \times 0.019$ m), which represents an altogether different stratigraphic section. The simulation model can qualitatively match the observed free-phase and dissolved CO$_2$ plume migration and convective mixing. Quantitatively, simulations are accurate during the injection phase, but their concordance decreases with time. Using local data reduces the time required to history match, although the forecasting capability of matched models is similar. The sand–water–CO$_{2(g)}$ system is very sensitive to effective permeability and capillary pressure changes; where heterogeneous structures are present, accurate deterministic estimates of CO$_2$ migration are difficult to obtain.

Keywords CO$_2$ storage · Geologic carbon sequestration · Two-phase flow · Numerical simulations · History matching · FluidFlower

1 Introduction

CO$_2$ capture and subsequent geologic carbon sequestration (GCS) is a climate-change mitigation technology that can be deployed at scale to offset anthropogenic CO$_2$ emissions during the energy transition (Marcucci et al. 2017; European Academies Science Advisory

---

Council (EASAC) 2018; Celia 2021; Intergovernmental Panel on Climate Change (IPCC) 2022). In GCS, reservoir simulation, including coupled flow and geomechanics, is the primary tool used to assess and manage geologic hazards such as fault leakage (e.g., Caine et al. 1996; Ingram and Urai 1999; Nordbotten and Celia 2012; Zoback and Gorelick 2012; Juanes et al. 2012; Jung et al. 2014; Vilarrasa and Carrera 2015; Saló-Salgado et al. 2023) and induced seismicity (e.g., Cappa and Rutqvist 2011; Zoback and Gorelick 2012; Juanes et al. 2012; Ellsworth 2013; Verdon et al. 2013; Alghannam and Juanes 2020; Hager et al. 2021). In response to the inherent uncertainties associated with modeling and simulation of CO_2 storage (Nordbotten et al. 2012), building confidence in the forecasting capabilities of simulation models requires calibration (or, synonymously, history matching), a process that involves updating the reservoir model to match field observations as they become available (Oliver and Chen 2011; Doughty and Oldenburg 2020).

History matching is an ill-posed inverse problem (Oliver and Chen 2011). This means that multiple solutions (i.e., parameter combinations) exist that approximate the data equally well. Automated techniques such as Markov chain Monte Carlo, randomized maximum likelihood or ensemble-based methods can be used to quantify uncertainty in history-matched models, especially in combination with surrogate models to reduce forward model computational time (see Aanonsen et al. 2009; Oliver and Chen 2011; Jagalur-Mohan et al. 2018; Jin et al. 2019; Liu and Durlofsky 2020; Santoso et al. 2021; Landa-Marbán et al. 2023, forthcoming, and references therein). In practice, however, it may be difficult to ensure that the chosen simulation model provides the best possible forecast. This is due to different subsurface conditions, the inability to include all sources of uncertainty in the models, incomplete field data and limited time for history matching.

In the laboratory, intermediate-scale (~meter) experiments have been used to study the physics of petroleum displacement (e.g., Gaucher and Lindley 1960; Brock and Orr 1991; Cinar et al. 2006) and contaminant transport (e.g., Silliman and Simpson 1987; Wood et al. 1994; Lenhard et al. 1995; Fernández-García et al. 2004). Similar 2D and 3D flow rigs have recently been applied to CO_2 storage, providing a link between core-scale measurements and field observations:

Kneafsey and Pruess (2010) found the impact of convective dissolution to be significant, using a page-size Hele–Shaw cell and numerical simulations. Neufeld et al. (2010) studied the scaling of convective dissolution and found it to be an important mechanism in the long-term trapping of injected CO_2 in an idealized site. Wang et al. (2010) used a 3D setup to investigate the ability of electrical resistivity tomography to identify localized leaks. Trevisan et al. (2014, 2017) focused on the impact of structural and residual trapping. In homogeneous sands, they found that previous trapping models, such as the Land (1968) model, can approximate the residually trapped gas saturation ($R^2 > 0.6$). Studying an heterogeneous aquifer characterized by a log-normal distribution of six different sand facies, they report that trapping efficiency increased significantly due to structural trapping. A strong control of sand heterogeneity on upward migration of CO_2 was also found by Lassen et al. (2015). Krishnamurthy et al. (2019, 2022) devised a novel technique to automate the process of beadpack/sandpack deposition and generate realistic depositional fabrics; they concluded that grain-size contrast and bedform architecture significantly impact CO_2 trapping. Subsequently, Ni et al. (2023) presented modified invasion-percolation simulations and reported that bedform architecture can impact CO_2 saturation if enough grain-size contrast is present. Askar et al. (2021) used a ~8-m-long tank to test a framework for GCS monitoring of CO_2 leakage. These studies employed homogeneous glass beads or sands, or focused on heterogeneities and bedform architectures in the aquifer layer; structural complexity was minimal.

In this paper, we use quasi-2D, intermediate-scale experiments of CO_2 storage to evaluate, quantitatively, the forecasting capability of history-matched simulation models against well-defined spatial data. An attempt was made to recreate realistic basin geometries, including stacking of storage reservoirs, faults, caprock and overburden. We simulate each of the three presented experiments with three versions of a numerical model, each with increasing access to local petrophysical measurements. These different versions are denoted Model 1 (M_1), Model 2 (M_2) and Model 3 (M_3). This allows us to assess (1) the value of local information of the system, expressed in terms of sand petrophysical measurements, during history matching, and (2) transferability or forecasting capability of our matched simulation models, when tested against a different experiment. The term *concordance* is used to evaluate agreement between experiments and simulations (Oldenburg 2018).

2 Physical Experiments

The physical experiments of CO_2 injection are conducted using the *FluidFlower* rigs. These rigs are meter-scale, quasi-2D tanks with transparent Plexiglass panels designed and built in-house at the University of Bergen (Fig. 1). Here, we used two tanks, with dimensions $89.9 \times 47 \times 1.05$ cm and $2.86 \times 1.3 \times 0.019$ m (referred herein to as Tank 1 and Tank 2, respectively). Different geologic settings are constructed by pouring unconsolidated sands with desired grain sizes into the water-saturated rigs. The rigs have multiple ports which allow
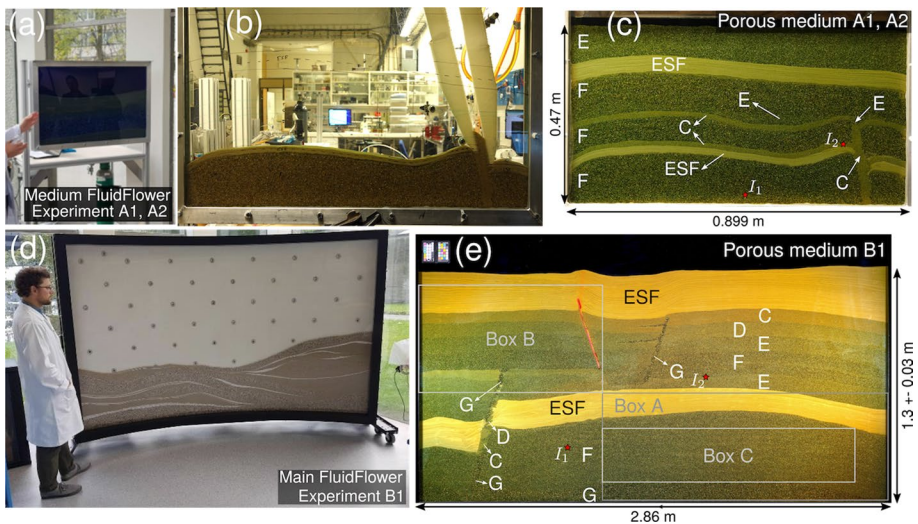

Fig. 1 Overview of the FluidFlower rigs and porous media used in the physical experiments. **a** Medium FluidFlower rig (Tank 1). **b** Snapshot during sand pouring to build the porous medium used in Experiments A1 and A2 in Tank 1 (Haugen et al. 2023, this issue). **c** Front view of porous medium in Tank 1, with lithologies in white and injector location shown with a red star. The length and height correspond to the porous medium. Note the fixed water table at the top. **d** Overview of the main FluidFlower rig (Tank 2), showing the back panel with sensor network. **e** Porous medium in Tank 2, used for Experiment B1, with lithologies in white. Location of injectors and Boxes A, B and C for analysis are shown with a red star and gray boxes, respectively. Length and variable height correspond to the porous medium

flushing out fluids after a given CO_2 injection, such that multiple injections can be conducted in the same setting. The location of the ports can be adjusted to accommodate different injection scenarios. A variety of techniques have been developed by UiB engineers in order to build complex structures such as folds and faults.

Below, we summarize the petrophysical measurements, experimental setup, geologic model/porous media construction and experimental schedule. Details on the conceptualization of the FluidFlower rigs and technical information are given in Fernø et al. (2023, this issue) and Eikehaug et al. (2023, this issue), while the full description of the physical experiment in Tank 1 and ex situ measurements are provided by Nordbotten et al. (2022); Haugen et al. (2023, this issue). Further details on the experiment in Tank 2, as well as results of the international benchmark study (IBS), are provided by Flemisch et al. (2023, this issue).

2.1 Sand Petrophysical Properties

Measurements on the employed Danish quartz sands were conducted using specialized equipment to determine average grain size (d), porosity (ϕ), permeability (k), capillary entry pressure (p_e) and drainage and imbibition saturation endpoints (denoted as connate water saturation, S_{wc}, and trapped gas saturation, S_{gt}). The methodology is described by Haugen et al. (2023, this issue), and obtained values are provided in Table 1. Sands C, D, E and F are very well sorted, sand G is well sorted, and sand ESF is moderately sorted (Haugen et al. 2023, this issue). We verified that Darcy's law is applicable in our system using the Reynolds number (R_e):

$$R_e = \frac{ud}{\nu} \tag{1}$$

where u is the fluid discharge per unit area, d the mean grain diameter, and ν the kinematic viscosity of the fluid. From our simulation results, matched to experimental observations, $\max(R_e) \leq 1$, which ensures the applicability of Darcy's law (e.g., Bear 1972).

Table 1 Petrophysical properties for used quartz sands, as obtained from local, ex situ measurements

Sand type	d (std) (mm)	ϕ (−)	k (D)	p_e (mbar)	S_{wc}	S_{gt}
ESF	0.2 (0.11)	0.435	44	15	0.32	0.14
C	0.66 (0.09)	0.435	473	3	0.14	0.1
D	1.05 (0.14)	0.44	1110	1	0.12	0.08
E	1.45 (0.19)	0.45	2005	–	0.12	0.06
F	1.77 (0.31)	0.44	4259	–	0.12	0.13
G	2.51 (0.63)	0.45	9580	–	0.1	0.06

Porosity and permeability are the average from two measurements for each sand, with a maximum difference between measurements of 0.02 (ϕ) and estimated 20% uncertainty (k). Measured gas column heights for sands E-G were 0, so p_e could not be directly measured. Experimental error in p_e, S_{wc} and S_{gt} was not quantified. A detailed description of the methodology and petrophysical values is provided by Nordbotten et al. (2022); Haugen et al. (2023, this issue)

2.2 Experimental Setup

The front and back panels of the FluidFlower are mounted on a portable aluminum frame, such that boundaries are closed on the sides and bottom (no flow). The top surface is open and in contact with fluctuating atmospheric pressure (Fig. 1). A fixed water table above the top of the porous medium was kept throughout the experiments conducted here. The experimental setup incorporates mass flow controllers to inject gaseous CO_2 at the desired rate, and a high-resolution digital camera with time-lapse function (Haugen et al. 2023, this issue).

Experiments were conducted in 2021 and 2022 in Bergen (Norway) at room temperature ($\approx 23\,°C$) and ambient atmospheric pressure. Temperature changes were minimized as much as possible, but maintaining a constant temperature was not possible in the available laboratory space. The fluids and sands were set in the FluidFlowers using the following procedure:

1. The silica sands are cleaned using an acid solution of water and HCl to remove carbonate impurities.
2. The FluidFlower rig is filled with deionized water.
3. Sands are manually poured into the rig using the open top boundary, in order to construct the desired porous medium.
4. A pH-sensitive, deionized-water solution containing bromothymol blue, methyl red, hydroxide and sodium ions is injected through multiple ports until the rig is fully saturated. This enables direct visualization of CO_2 gas (white), dissolved CO_2 (yellowish orange to red), and pure water (dark teal).
5. 5.0 purity (99.999%) CO_2 is injected as gaseous phase at the desired rate. CO_2 is injected through dedicated ports directly into the rig (Fig. 1).
6. After the injection phase, injection ports are closed and CO_2 migration continues.
7. Once the experiment is finished, the rig can be flushed with deionized water and the process can start again from step 4.

Full details on the fluids are given in Fernø et al. (2023, this issue) and Eikehaug et al. (2023, this issue). Below, we refer to the pH-sensitive solution in the rigs as "dyed water".

2.3 Porous Media Geometries

The geometries of the porous media used in this paper aim to recreate the trap systems observed in faulted, siliciclastic, petroleum-bearing basins around the world, given the geometrical constraints of the FluidFlowers and manual sand pouring (Fernø et al. 2023; Eikehaug et al. 2023, this issue). Features such as folds, faults and unconformities were built in both Tanks 1 and 2. The construction of faults, shown in Fig. 1b and detailed in Haugen et al. (2023, this issue), requires a minimum effective "fault-plane" thickness; hence, our fault structures are thicker than natural faults with the same displacement (Childs et al. 2009). Fine sands ($d \approx 0.2$ mm) are used to represent sealing or caprock formations.

The geometry in Tank 1 (Fig. 1c) contains three main high-permeability reservoirs (F sand). The bottom and middle F sand are separated by a seal (ESF sand), while the middle and top are separated by the C sand and connected through a higher permeability fault (refer to Sect. 2.1 for petrophysical properties). The fault separates the bottom section into

two compartments. The bottom and top F sand provide anticlinal traps for the CO_2 to accumulate in.

The geometry in Tank 2 (Fig. 1e) was specifically motivated by the structure of North Sea reservoirs and petroleum basins. From bottom to top, it contains two sections of decreasing-permeability reservoirs capped by two main sealing layers. A fault separates the bottom section into two compartments, while two faults separate the top section into three compartments. Each fault has different petrophysical properties: The bottom fault is a heterogeneous structure containing ESF, C, D, F and G sands, the top-left fault is an impermeable structure made of silicone and the top-right fault is a conduit structure containing G sand.

2.4 Experimental Injection Schedule

The injection schedules for experiments in Tanks 1 and 2 are provided in Table 2. Injection ports have an inner diameter of 1.8 mm.

3 Numerical Simulations

3.1 Model Setup

The isothermal simulations presented in this work were performed with the MATLAB Reservoir Simulation Toolbox, MRST (Krogstad et al. 2015; Lie 2019; Lie and Møyner 2021). Specifically, we used the black-oil module, which is based on fully implicit solvers with automatic differentiation, and assigned properties of water to the oleic phase, such that the gaseous phase (CO_2 only) can dissolve in it. Vaporization of water into the gas phase and chemical reactions are not considered, because they are not primary controls on fluid migration for our operational setup and analysis time.

Table 2 Schedules for the three CO_2 injection experiments simulated in this work

Experiment A1		A2		B1	
I_R (ml/min)	t (hh:mm:ss)	I_R	t	I_R	t
0.1 (I_1)	00:00:00	0.1 (I_1)	00:00:00	10.0 (I_1)	00:00:00
2.0	00:05:00	2.0	00:05:00	10.0	05:00:00
2.0	00:50:00	2.0	04:43:44	0.0	05:00:01
0.0	00:55:00	0.0	04:48:33	10.0 (I_2)	02:15:00
0.1 (I_2)	01:09:11	0.0	05:00:00	10.0	05:00:00
2.0	01:14:11			0.0	05:00:01
2.0	02:29:11			0.0	120:00:00
0.0	02:34:00				
0.0	48:00:00				

I_R is injection rate, while I_i denotes injector (port) number. A five-minute ramp-up and ramp-down was applied in Experiments A1 and A2 in Tank 1. Total duration of conducted experiments and simulations is 48 h (A1), 5 h (A2) and 120 h (B1). Location of injection wells is provided in Fig. 1

In addition to structural and dissolution trapping, we also considered residual trapping (Juanes et al. 2006) to be consistent with local measurements showing nonzero trapped gas saturation (Sect. 2.1). This is achieved through hysteretic relative permeability curves for the nonwetting (gas) phase (see Sect. 3.2). Our implementation in MRST follows ECLIPSE's technical description (Schlumberger 2014), and Killough's (1976) model is used to compute the scanning curves (Saló-Salgado et al. 2023, forthcoming). Physical diffusion was also included through the addition of a diffusive flux term with a scalar, constant coefficient in the computation of the total CO_2 flux (Bear 1972).

The simulator requires very small time-steps (seconds to minutes) due to the buoyancy of CO_2 at atmospheric conditions and high sand permeabilities (Table 1). Linear solver time was reduced by means of AMGCL (Demidov and Rossi 2018; Lie 2019), an external, pre-compiled linear solver. The greatest challenge was the convergence of the nonlinear solver, which required many iterations and time-step cuts. This is consistent with the groups working in the FluidFlower international benchmark study (Flemisch et al. 2023, this issue).

Next, we describe the computational grids for experiments in Tanks 1 and 2, PVT properties and boundary conditions. Petrophysical properties are specific of each model version and are detailed in Sect. 3.2.

3.1.1 Computational Grids

A front panel image of the porous medium was used to obtain layer contact coordinates through a vector graphics software (Fig. 2a). These contacts were then imported into MATLAB to generate the computational grids using the UPR module (Berge et al. 2019, 2021) (Fig. 2b, d). The grids were generated in 2D and then extruded to 3D (using a single cell layer) to account for thickness and volume. Note that, in Tank 1, where the porous
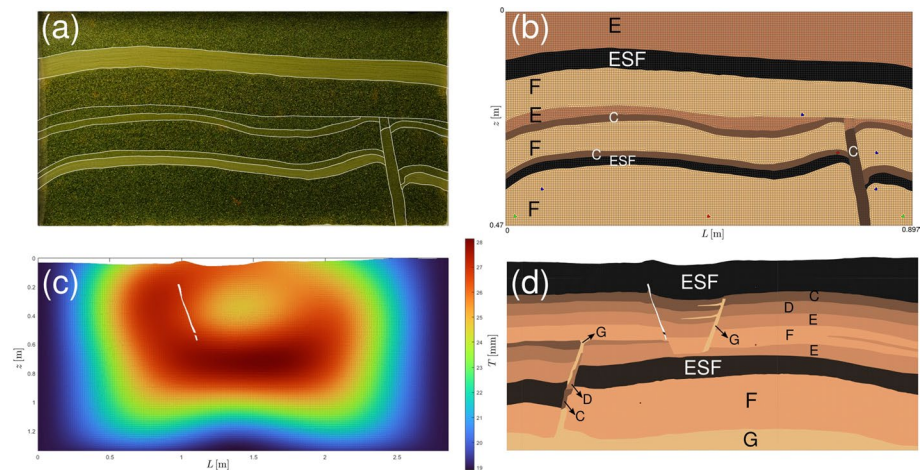

Fig. 2 Simulation grids overview. **a** Front-panel view of Tank 1, where the layer contacts have been highlighted in white. **b** Front view of simulation grid for experiments in Tank 1, with lithologies indicated and colored based on petrophysical properties (see Sect. 3.2). Location of injection wells is shown in red. **c** Thickness map of simulation grid for experiments in Tank 2. **d** Front view of simulation grid for experiments in Tank 2, with lithologies indicated and colored based on petrophysical properties. Location of injection wells is shown in red

medium has dimensions of $89.7 \times 47 \times 1.05$ cm, the thickness (space between the front and back panels) is constant (10.5 mm). Tank 2, which is significantly larger (porous medium dimensions $2.86 \times 1.3 \times 0.019$ m), has a thickness of 19 mm at the sides; however, it varies towards the middle due to forces exerted by the sand and water, to a maximum of 28 mm. A thickness map obtained after initial sand filling was used to generate our variable-thickness mesh via 2D interpolation (Fig. 2c). Also, the top surface of the porous medium is not flat (height $= 130 \pm 3$ cm).

Our composite Pebi grids (Heinemann et al. 1991) have a Cartesian background and are refined around face constraints (contacts and faults) as well as cell constraints (injection wells) (Berge et al. 2019, 2021). We generated multiple grids to test the finest grid we could afford to simulate Experiment B1 in Tank 2 with. Our grid has a cell size $h \approx 5$ mm and 151,402 cells (Fig. 2d). The grid used for Tank 1 has a similar cell size ($h \approx 4$ mm and 27,200 cells), which was chosen to reduce grid-size dependencies when applying our matched models to Experiment B1.

3.1.2 PVT Properties

Consistent with experimental conditions, our simulations are conducted at atmospheric conditions ($T = 25$ C), where the CO_2 is in gaseous state. We employed a thermodynamic model based on the formulations by Duan and Sun (2003) and Spycher et al. (2003); Spycher and Pruess (2005) to calculate the composition of each phase as a function of p, T. The implementation for a black-oil setup is described in Hassanzadeh et al. (2008) and references therein. Given the boundary conditions (Sect. 3.1.3) and dimensions of our experimental porous media, pore pressure changes (Δp) are very small in our simulations (max $\Delta p \ll 1$ bar). Hence, the fluid properties remain similar to surface conditions, where the water and CO_2 have, respectively, a density of 997 and 1.78 kg/m^3, and a viscosity of 0.9 and 0.015 cP. The maximum concentration of CO_2 in water is ≈ 1.5 kg/m^3.

3.1.3 Initial, Boundary and Operational Conditions

Our porous media are fully saturated in water at the beginning of CO_2 injection. No-flow boundary conditions were applied everywhere except at the top boundary, which is at constant pressure and includes a fixed water table a few cm above the top of the porous medium. Injection is carried out via wells completed in a single cell at the corresponding coordinates. The diameter of injection wells is 1.8 mm in both Tank 1 and Tank 2, which operate at a constant flow rate (see Sect. 2). The simulation injection schedule follows the experimental protocol, provided in Table 2. Note that injection rates in our simulations of Experiment A1 and A2 were slightly adjusted during the calibration procedure, as explained in Sects. 3.3 and 4.

3.2 Simulation Model

Three different model versions, denoted Model 1 (M_1), Model 2 (M_2) and Model 3 (M_3), are used throughout this study to evaluate the value of local data in forecasting subsurface CO_2 migration. Each successive model was constructed based on access to an increasing level of local data, with M_1 having access to the least data and M_3 having access to the most data. The model-specific parameters are limited to the following:

- Petrophysical properties (porosity, permeability, capillary pressure and relative permeability), which depend on available local data and are described in this section.
- The molecular diffusion coefficient (D). Models 1–3 were calibrated using the same value, $D = 10^{-9}$ m^2/s. Additionally, Model 3 was also calibrated with $D = 3 \times 10^{-9}$ m^2/s. Accordingly, where required we denote Model 3 as $M_{3,1}$ and $M_{3,3}$.
- Injection rate. Experiments in Tank 1 were conducted at a very low injection rate ($I_R = 2$ ml/min, see Table 2). Given that the mass flow controllers used in Tank 1 may be inaccurate for this rate, the injection rate was also modeled as an uncertain parameter. Model calibration was achieved with $I_R \in [1.6, 1.8]$ ml/min for all three models.

All other model characteristics, including the grid and numerical discretization, remain unchanged. Below, we describe the starting petrophysical values for each of our three simulation models. Note that the experimental geometry in Tank 1, used for matching, only contained sands ESF, C, E and F. Properties for sands D and G are also provided because they were required to simulate the experiment in Tank 2 (Fig. 1).

3.2.1 Model 1 (M_1)

For this model, local petrophysical data were limited to a measure of the average grain size (d; see Sect. 2.1 and Table 1). Hence, petrophysical properties were estimated from published data in similar silica sands. Porosity was selected from data in Beard and Weyl (1973) and Smits et al. (2010) for moderately to well-sorted sands. Permeability was obtained from fitting a Kozeny–Carman model to data in Beard and Weyl (1973) and Trevisan et al. (2014). The resulting equation has the form $k = \beta d^2 \phi^3$, where β equals 12,250 in our fit with d in mm and k in D. Obtained porosity and permeability values are provided in Table 3.

Capillary pressure curves were computed as described below:

1. Capillary pressure measurements in a similar system were obtained from the literature. In this case, Plug and Bruining (2007) measured capillary pressure curves on the unconsolidated quartz sand-CO_2-distilled water system at atmospheric conditions. We used their measurements on sand packs with an average particle size between 0.36 and 0.41 mm, which are closest to the C sand in our experiments (Fig. 3a).
2. A Brooks and Corey (1964) model of the form $p_c = p_e(S_w^*)^{-\frac{1}{\lambda}}$ was fitted to these data, where p_e is the nonwetting phase entry pressure at $S_w = 1$, $\lambda = 2.6$ and $S_w^* = \frac{S_w - S_{wc}}{1 - S_{wc}}$ is

Sand type	d (mm)	ϕ (−)	k (D)
ESF	0.2	0.37	25
C	0.66	0.38	290
D	1.05	0.40	930
E	1.45	0.39	1530
F	1.77	0.39	2280
G	2.51	0.42	5720

Table 3 Initial porosity and permeability for Model 1

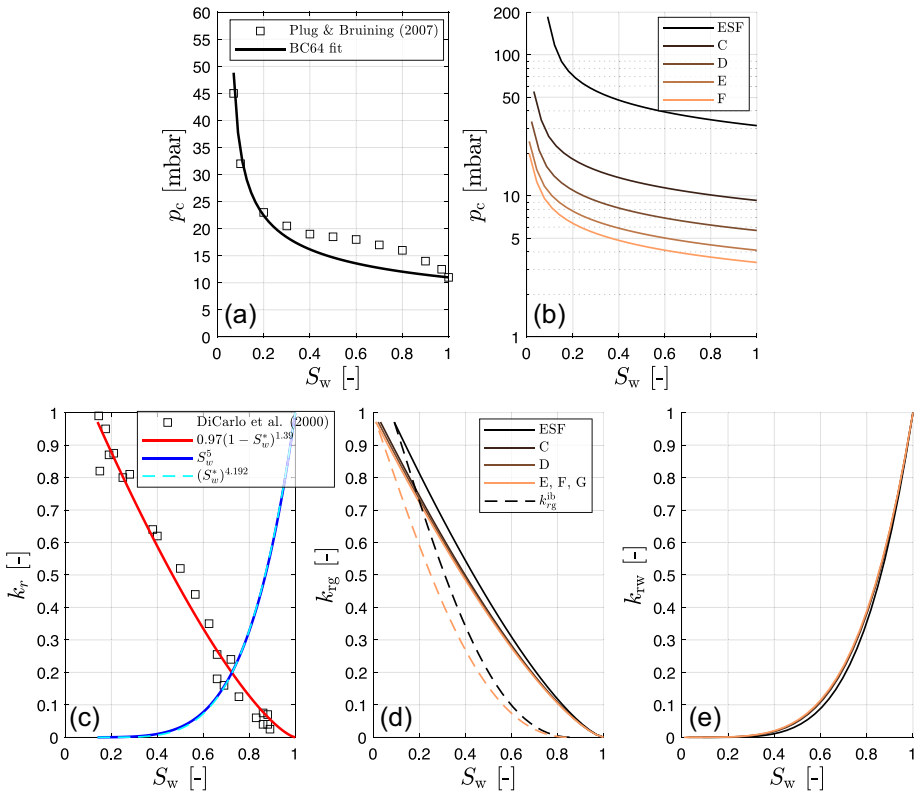See main text for estimation details

Fig. 3 Multiphase flow properties for Model 1. **a** Capillary pressure measurements and reference curve using a Brooks and Corey (1964) function. **b** Initial capillary pressure curves, computed from the reference curve using Leverett scaling (see main text). **c** Relative permeability data (squares and S_w^5 model) and our fitted Corey model. **d**, **e** Relative permeability of gas and water, respectively. The drainage curve is shown as a solid line, while the bounding imbibition curve is shown for sands ESF and G as a discontinuous line. No relative permeability hysteresis was considered for the water phase

the normalized water saturation with irreducible or connate water saturation S_{wc}. This fit led to our reference curve, p_{cr} (Fig. 3a).

3. The capillary pressure depends on the pore structure of each material, such that sands with different grain sizes require different p_c curves. The capillary pressure variation can be modeled by means of the dimensionless J-function proposed by Leverett (Leverett 1941; Saadatpoor et al. 2010): $J(S_w) = \frac{p_c}{\sigma \cos \theta} \sqrt{\frac{k}{\phi}}$, where σ is the surface tension and θ the contact angle. Assuming the same wettability and surface tension for different sand regions, and the same shape of the p_c curve, the capillary pressure for any given sand (p_{cs}) can be obtained from the reference curve as $p_{cs}(S_w) = p_{cr}(S_w)\sqrt{\frac{k_r \phi_s}{k_s \phi_r}}$ (Fig. 3b).

Drainage relative permeabilities were obtained from CO_2-water measurements by DiCarlo et al. (2000), who used water-wet sandpacks with 0.25 mm grain size. Specifically, we used the data reported in their Figs. 4 and 5 and fitted Corey-type functions (Corey 1954; Brooks and Corey 1964) of the form $k_{rw} = (S_w^*)^a$ and $k_{rg} = c(1 - S_w^*)^b$ (Fig. 3c). The fitted exponents a and b are 4.2 and 1.4, respectively, while c is 0.97. We assumed that the

difference in relative permeability of different sands is the result of different irreducible water saturation only (see Fig. 3d,e). For each of our sands, S_{wc} was obtained from Timur (1968) as $S_{wc} = 0.01 \times 3.5 \frac{\phi^{1.26}}{k^{0.35}} - 1$, where ϕ is in percent and k in mD. This model was used to compute S_{wc} for both the p_c and k_r curves.

In CO_2 storage, secondary imbibition occurs where the water displaces buoyant gas at the trailing edge of the CO_2 plume, disconnecting part of the CO_2 body into blobs and ganglia and rendering them immobile (Juanes et al. 2006, and references therein). This means that the maximum water saturation that can be achieved during imbibition equals $1 - S_{gt}$ (the trapped gas saturation). Here, we used measurements in sandpacks from Pentland et al. (2010) to determine S_{gt}. In particular, we fitted Land (1968)'s model with the form $S_{gt}^* = \frac{S_{gi}^*}{1 + C S_{gi}^*}$, where $S_g^* = \frac{S_g}{1 - S_{wc}} = 1 - S_w^*$, S_{gi} is the gas saturation at flow reversal, and C is Land's trapping coefficient with a value of 5.2 in our fit. Although Pentland et al. (2010) report that the best fit is achieved with the Aissaoui (1983) and Spiteri et al. (2008) models (cf. their Fig. 5), Land's model was chosen here, given that most relative permeability hysteresis models build on this one (see next paragraph).

Nonwetting phase trapping contributes to irreversibility of the relative permeability and capillary pressure curves (hysteresis). Here, we accounted for this mechanism in the gas relative permeability due to its importance in subsurface CO_2 migration (Juanes et al. 2006, and references therein). In particular, we used Land's (1968) model to compute the bounding imbibition curve (see Fig. 3d), where S_{gt} is obtained as described above, and Killough's (1976) model to characterize the scanning curves. In Killough's model, the scanning curves are reversible, such that the relative permeability at $S_g < S_{gi}$ no longer depends on the displacement type.

3.2.2 Model 2 (M_2)

This model had access to local, ex situ measurements of single-phase petrophysical properties, i.e., porosity and intrinsic permeability (see Sect. 2.1 and Table 1). Comparing with Table 3, it can be seen that our estimation for Model 1 above was correct to the order of magnitude, but resulted in smaller values: porosity $\in [85, 93]\%$ and permeability $\in [53, 84]\%$ of the local measurements.

Capillary pressures and relative permeabilities were obtained using the same procedure described above for Mdel 1. The slight differences with respect to the curves shown in Fig. 3b, d, e come from the porosity and permeability values used in the Leverett scaling and to determine S_{wc}, which were taken from Table 1 instead. The obtained curves for Model 2 are provided in Fig 4.

3.2.3 Model 3 (M_3)

This model was allowed access to all local, ex situ measurements (see Table 1). Initial porosity and permeability remain unchanged with respect to Model 2. Capillary pressure curves were obtained by scaling the reference curve described in Sect. 3.2.1 and shown in Fig. 3a using the measured entry pressure (Sect. 2.1). The scaling followed the model $p_{cs}(S_w) = p_{cr}(S_w) \frac{p_e}{p_{er}}$, where p_e is the measured entry pressure for each sand, and p_{er} is the reference curve entry pressure. The obtained curves are shown in Fig. 5a.
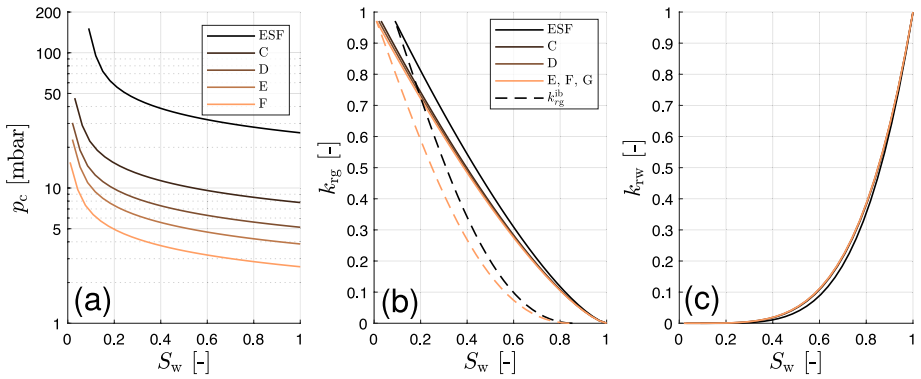
Fig. 4 Multiphase flow properties for Model 2. **b** Initial capillary pressure curves, computed from the reference curve using Leverett scaling (see main text). **b**, **c** Relative permeability of gas and water, respectively. The drainage curve is solid, while the bounding imbibition curve is shown for sands ESF and G as a discontinuous line. No relative permeability hysteresis was considered for the water phase
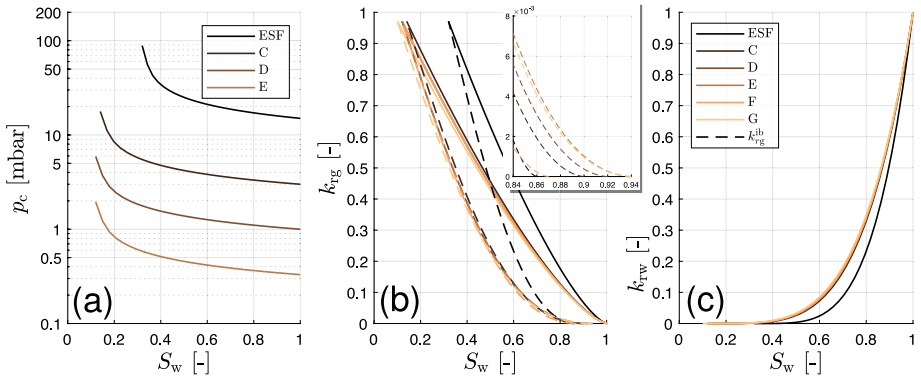

Fig. 5 Multiphase flow properties for Model 3. **b** Initial capillary pressure curves, computed according to the entry pressure determined experimentally (see Sect. 2.1). **b**, **c** Relative permeability of gas and water, respectively, according to the endpoints determined experimentally (Sect. 2.1). The drainage curves are solid, while the bounding imbibition curves are shown as a discontinuous line. The inset in **b** is a zoom view around the trapped gas saturation. No relative permeability hysteresis was considered for the water phase

Relative permeabilities were computed following the same procedure described for Model 1 above. In this case, however, each sand type was assigned the measured S_{wc} and S_{gt} values (see Table 1). This led to differences in both the drainage and imbibition curves, as shown in Fig. 5.

3.3 Model Calibration

Concordance between results obtained with each simulation model (M_1 to M_3) and the validation experiment in Tank 1 (A1, see Sect. 2.4) is quantitatively assessed by comparing the following quantities (see Fig. 6):
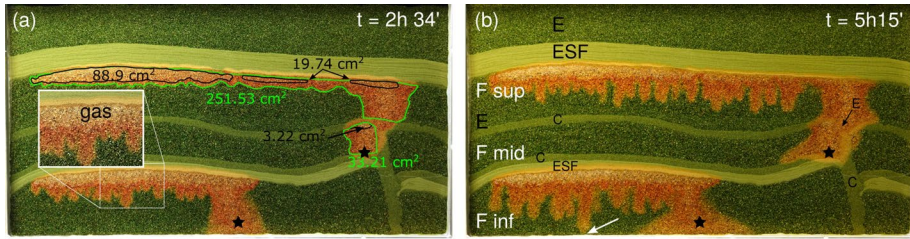
Fig. 6 Front panel view of Tank 1, showing quantities and times for history matching of numerical models to Experiment A1. **a** shows areas with gaseous CO_2 (free-phase, black contours) and dyed water with dissolved CO_2 (green contours) at the end of injection. Location of injection ports is shown with a star. **b** shows the time and location where the first finger touches the bottom of the tank (white arrow), as well as the different lithological units. Note the three F reservoirs labeled 'inf', 'mid' and 'sup', mentioned in the text and other figures

1. At $t = 55$ min (end of injection in port I_1): Areas occupied by free-phase CO_2, and dyed water with dissolved CO_2 in the bottom F reservoir.
2. At $t = 154$ min (end of injection in port I_2): Areas occupied by free-phase CO_2, and dyed water with dissolved CO_2, in the middle and top F reservoirs.
3. Time at which the first finger touches the tank bottom.
4. Time at which the first finger (sinking from the top F reservoir) touches the middle C sand.

Experimental values for points 1–2 were obtained by computing areas from time-lapse images using a vector-graphics software. Careful visual inspection of color-enhanced images was used to distinguish between free-phase CO_2 (white) and dyed water with dissolved CO_2 (yellowish orange to red), and to identify the times for points 3–4 above. Error in experimental values was estimated to be $\leq 5\%$, based on repeated measurements (points 1–2), and ~ 5 min, based on timelapse image comparison (points 3–4). In the simulation models, the threshold gas saturation and CO_2 concentration in water used to compute areas were $S_g > 10^{-3}$ and $C_{CO_2} > 15\%(C_{CO_2}^{max}) \approx 0.2$ [kg/m^3], respectively. The C value was chosen after a shape comparison of the region with dissolved CO_2. A smaller value of $C_{CO_2} > 0.05$ [kg/m^3] was selected to determine finger times for points 3 and 4 above. Figure 6 shows an overview of the experimental values for points 2 and 3, while Fig. 12 in Sect. 4.2 shows the full comparison with the history-matched/calibrated simulation models.

The experiment was conducted first. Afterwards, the process consisted of running Simulation models 1 to 3, in parallel, starting with the petrophysical properties described in Sect. 3.2. Given the number of uncertain variables (four petrophysical properties for each lithological unit, the diffusion coefficient and the injection rate) and the time required to complete a single simulation, a manual history matching method was employed. At the end of each run, quantities 1–4 above were compared and one or more properties were manually changed based on observed concordance and domain knowledge. During the first few runs, only quantities 1 and 2 above were compared. After obtaining a satisfactory areal match, petrophysical properties were further adjusted to match quantities 3 and 4.

4 Results

In Sect. 4.1, we present the results of the first simulation of Experiment A1 with each model and property values detailed in Sect. 3.2. Then, we detail the calibration of simulation models using Experiment A1 and assess the value of local data to history-match CO_2 storage simulation models (Sect. 4.2). Finally, we apply these matched models to Experiment A2, analog for a longer injection in the same geology (Sect. 4.3.1), and to Experiment B1, analog for a larger-scale injection in a different geologic setting (Sect. 4.3.2). We use simulations of Experiments A2 and B1 to assess the forecasting ability of simulation models in different conditions.

4.1 Initial Model Results

Figure 7 shows the comparison between Experiment A1 and the first run with each model, at times indicated in Sect. 3.3. Numerous differences are evident between the experiment and Models 1 and 2, while Model 3 is much closer to the experiment. In particular, models 1 and 2 overestimate the extent of CO_2-rich brine and underestimate the amount of gaseous CO_2 in all F reservoirs (refer to Fig. 6 for location). Model 3 approximates much better the areal extent of gaseous CO_2 in all regions, as well as the CO_2-rich brine in the middle and upper F reservoirs. Model 2 provides the closest finger migration times (points 3 and 4 in Sect. 3.3), although this was not evaluated in the first run, as discussed below.

Petrophysical properties for Models 1 and 2 were obtained from references in Sect. 3.2, which also used silica sands with similar grain sizes. However, despite the relatively homogeneous nature of our quartz sands, Model 3 is significantly more concordant. This result stems from natural sand variability and highlights the difficulty in establishing general,
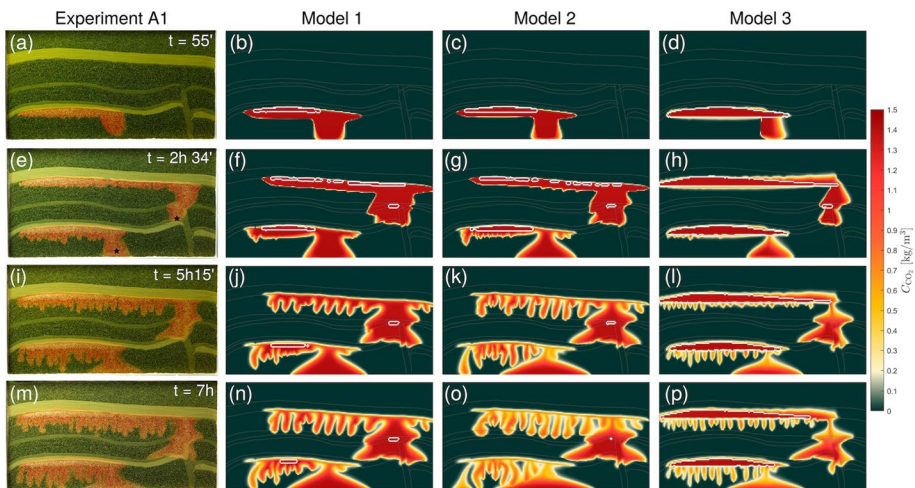

Fig. 7 Comparison between Experiment A1 in Tank 1 (left column) and first-run simulation results with Models 1–3. Color map in simulation plots refers to CO_2 concentration in water, according to color bar. The white contours in simulation plots indicate $S_g = 10^{-3}$. **a–d** End of injection in port 1. **e–h** End of injection in port 2. **i–l** Time at which the first finger touches the tank bottom. **m–p** Time at which the first finger touches the middle C sand

representative elementary volume-scale properties for porous media (see, for instance, Hommel et al. 2018; Schulz et al. 2019, for a discussion on intrinsic permeability). Additionally, results in Fig. 7 highlight the need for conducting sand/rock-specific measurements, even in the case of well-sorted, homogeneous sediments.

4.2 Manual History Matching and Value of Local Data

Figure 8 shows convergence of areas occupied by free gas (A_g) and water with dissolved CO_2 (A_d), according to Sect. 3.3. Each iteration corresponds to a successive model with manually updated parameters, and the different F sand regions evaluated in each panel (a) to (f) are provided in Fig. 6. With the exception of A_d in the lower F sand, Model 3 is accurate since the beginning, and all areas were satisfactorily matched after four iterations. Conversely, Models 1 and 2 were significantly off the experimental reference during the first few iterations. Model 2, however, was accurate after five iterations, while
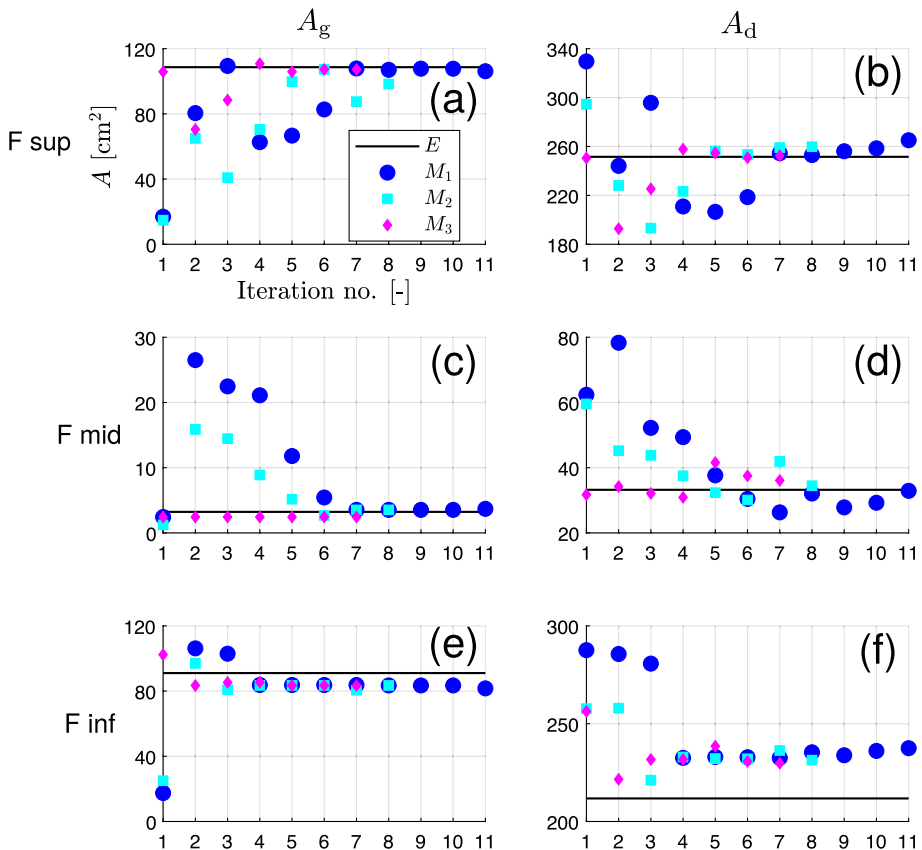
Fig. 8 Convergence of areas occupied by free gas (A_g, left column) and water with dissolved CO_2 (A_d, right column), during the calibration of Models 1–3 with Experiment A1. A_d includes area with gaseous CO_2 (see Fig. 6). Each iteration represents a new simulation run, and the experimental reference (E) is shown as a black line. Refer to Fig. 6 for region location, and to Sect. 3.3 for calibration procedure. **a**, **b**: Upper F sand. **c**, **d**: Middle F sand. **e**, **f**: Lower F sand

Fig. 9 Convergence of mean absolute error over the six areal quantities measured during the calibration process. The error is computed with respect to experimental values. See Fig. 8 for areas measured, and refer to Sect. 3.3 for calibration procedure
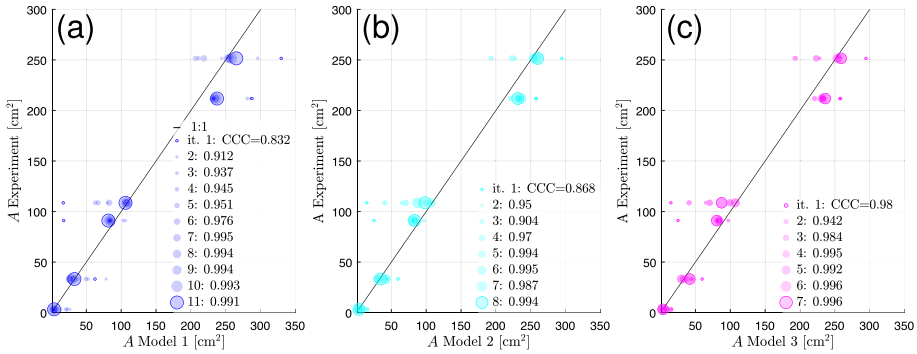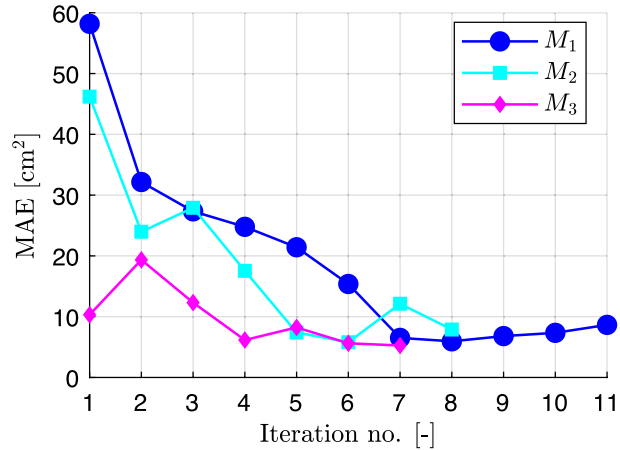

Fig. 10 Concordance between successive model iterations and the experiment, based on six areal measures evaluated during the calibration. Lin's CCC (Lin 1989) is shown in the key of each subplot, computed according to Eq. 2. **a** Model 1. **b** Model 2. **c** Model 3

Model 1 required seven iterations to give satisfactory areal estimates. The mean absolute error (MAE) over the six areal quantities presented in Fig. 8 is evaluated in Fig. 9, where it can be seen that, while all models are accurate towards the end (MAE $\in [5-10]\,\mathrm{cm}^2$), that required a sixfold improvement in Models 1 and 2, but only twofold in Model 3. As mentioned in Sect. 3.3, $C_{CO_2} > 15\%(C_{CO_2}^{max}) \approx 0.2\,[\mathrm{kg/m}^3]$ was used as threshold to determine areas. While the absolute values and error would change with a different C_{CO_2} threshold, we checked that the relative accuracy of our calibrated models does not with both $C_{CO_2} > 0.01$ and $0.1\,[\mathrm{kg/m}^3]$.

Agreement between simulations and experimental observations is readily seen in Fig. 10, where the 1:1 line indicates perfect concordance. The degree of concordance can be quantified by means of Lin's concordance correlation coefficient (CCC) (Lin 1989; Oldenburg 2018), which, for N-valued observation (x) and model (y) vectors (the six areal quantities), is computed as:

$$\text{CCC} = \frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2 + (\bar{x} - \bar{y})^2} \tag{2}$$

where $\bar{x}$ and $\bar{y}$ are the means, σ_x^2 and σ_y^2 the variances, and σ_{xy} the covariance, all calculated using $1/N$ normalization. Results in Fig. 10 show that model calibration results in very good concordance for all models (CCC ≥ 0.99).

Convergence of quantities 3 and 4 in Sect. 3.3, the times at which the first finger touches the rig bottom and the middle C sand, respectively, is provided in Fig. 11. These times were only evaluated after a satisfactory areal match for quantities in Fig. 8 was achieved. Therefore, areas no longer change much in the last few iterations in Fig. 8. In Fig. 11, it can be seen that Models 2 and 3, which incorporated local intrinsic permeability measurements, were significantly closer to our experimental reference than Model 1. Initially, however, we observed that sinking of gravity fingers in the experiment was faster than our model values by a factor of ≈ 2. A satisfactory match of all quantities evaluated was achieved after 11, 8, and 7 iterations for Models 1–3, respectively.

Overall, we find that Model 3, with access to local single-phase and multiphase flow properties, is closer to the experimental reference (i.e., more concordant) from the start. Model 1 started farthest and required significantly more effort for calibration. After the calibration process, all models achieve very good concordance (CCC ≥ 0.99), based on evaluated quantities (Fig. 10). The calibration shown in Figs. 8, 9, 10, 11 employs $D = 10^{-9}$ m^2/s in all model versions (M_1 to M_3). Injection rates (I_R) started at 2.0 ml/min for all three models and were 1.6 ml/min, 1.8 ml/min and 1.75 ml/min, respectively, at the end of the calibration. I_R is slightly different because the goal was to obtain the best match with each model, considering I_R to be an uncertain variable. In Sect. 4.3, the same I_R is used to make forecasts with all three models.

Table 4 compares the starting and final (matched) key petrophysical variables for each model. The models were successfully calibrated by adjusting intrinsic permeability and the capillary pressure curves (same shape, but scaled to higher or lower p_e only). It was found that CO_2 migration was most sensitive to the properties of the F sand were most of the CO_2 migration occurs, as well as the ESF seal, which structurally traps the CO_2 plume. In our matched models, p_e of ESF is about twice the measured value; this was required because the minimum saturation at which we can define p_e and ensure
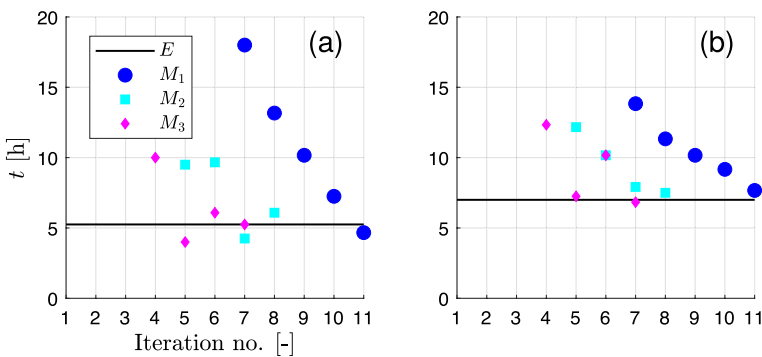

Fig. 11 Convergence of times at which the first finger touches the bottom of the rig (**a**) and the middle C sand (**b**), during the calibration of Models 1–3 with Experiment A1. Refer to Sect. 3.3 for calibration procedure

Table 4 Petrophysical properties for used quartz sands in Experiment A1

Sand type / model	ϕ (−)	k (D)	p_e (mbar)	S_{wc} (−)	S_{gt} (−)
ESF	0.435	44	15	0.32	0.14
M_1^i	0.37	25	31.4	0.09	0.1468
M_1^f	0.37	6	31.4	0.09	0.1468
M_2^i	0.435	44	25.6	0.09	0.1468
M_2^f	0.435	44	25.6	0.09	0.1468
M_3^i	0.435	44	15	0.32	0.14
M_3^f	0.435	15	30	0.32	0.14
C	0.435	473	3	0.14	0.1
M_1^i	0.38	293	9.3	0.03	0.1565
M_1^f	0.38	293, 27	4.6	0.03	0.1565
M_2^i	0.435	473	7.8	0.03	0.1565
M_2^f	0.435	473, 158	2.6	0.03	0.1565
M_3^i	0.435	473	3	0.14	0.1
M_3^f	0.435	473, 118	4.5	0.14	0.1
E	0.45	2005	–	0.12	0.06
M_1^i	0.39	1528	4.1	0.01	0.16
M_1^f	0.39	1528	0.5	0.01	0.16
M_2^i	0.45	2005	3.86	0.01	0.16
M_2^f	0.45	3008	0.58	0.01	0.16
M_3^i	0.45	2005	0.33	0.12	0.06
$M_{3,1}^f$	0.45	2406	0.33	0.12	0.06
$M_{3,3}^f$	0.45	3208	0.33	0.12	0.06
F	0.44	4259	–	0.12	0.13
M_1^i	0.39	2277	3.3	0.01	0.16
M_1^f	0.39	6540, 2907	0	0.01	0.16
M_2^i	0.44	4259	2.62	0.01	0.16
M_2^f	0.44	6814, 4259	0	0.01	0.16
M_3^i	0.44	4259	0	0.12	0.13
$M_{3,1}^f$	0.44	7240, 4685	0	0.12	0.13
$M_{3,3}^f$	0.44	9796, 4259	0	0.12	0.13

Methodology for local measurements is provided by Haugen et al. (2023), while starting property modeling is described in Sect. 3.2. For each sand, measured (first row), initial (superscript i) and final (superscript f) values for each of our models are shown. For sand C, the second permeability value refers to the fault, if different from the rest. For sand F, the second permeability value refers to the middle F layer, if different from the rest. For Model 3, where property values are different, $M_{3,1}$ refers to the calibration with $D = 10^{-9}\,\mathrm{m^2/s}$ and $M_{3,3}$ refers to $D = 3 \times 10^{-9}\,\mathrm{m^2/s}$

numerical convergence is $S_g \approx 10^{-4}$. Reality, however, is closer to a jump in p_c from 0 to p_e at an infinitesimally small S_g. Additionally, we found that concordance improved when using different values for the C and F sands in different model regions. In the case of the C sand, the explanation lies in the fault construction process, which may

reduce porosity with respect to "natural" sedimentation of stratigraphic layers (Haugen et al. 2023, this issue). The increase in F sand permeability was required to match finger migration times and is possibly compensating the absence of mechanical dispersion in the simulations. This is discussed in Sect. 5. Our calibrated values are within the same order of magnitude of the ex situ measurements (Table 4) and history-matched values for the porous medium in Tank 2 (Landa-Marbán et al. 2023, forthcoming).

Figure 12 shows gas saturation (S_g) and CO_2 concentration (C_{CO_2}) maps at times at which quantities 1–4 described in Sect. 3.3 are evaluated. Snapshots are provided for Model 3 only, since all three calibrated models were qualitatively very similar. It can be seen that CO_2 migration is successfully approximated by our numerical model. In detail, however, some differences are apparent: Firstly, sinking of CO_2-rich water from the bottom injector and horizontal migration along the bottom of the rig is faster in the model. This is due to the higher permeability that our numerical model requires in order to match the gravity fingering advance (cf. Table 4). Secondly, the experiment shows that denser, CO_2-rich water sinks with a rather compact front and closely spaced, wide fingers. Our model with constant $D = 10^{-9}$ m²/s approximates all gravity-driven migration of the CO_2-rich water through thinner fingers, with the CO_2-saturated region receding with S_g. To better represent fingering widths, we also matched Model 3 with $D = 3 \times 10^{-9}$ m²/s, used in Sect. 4.3.2.
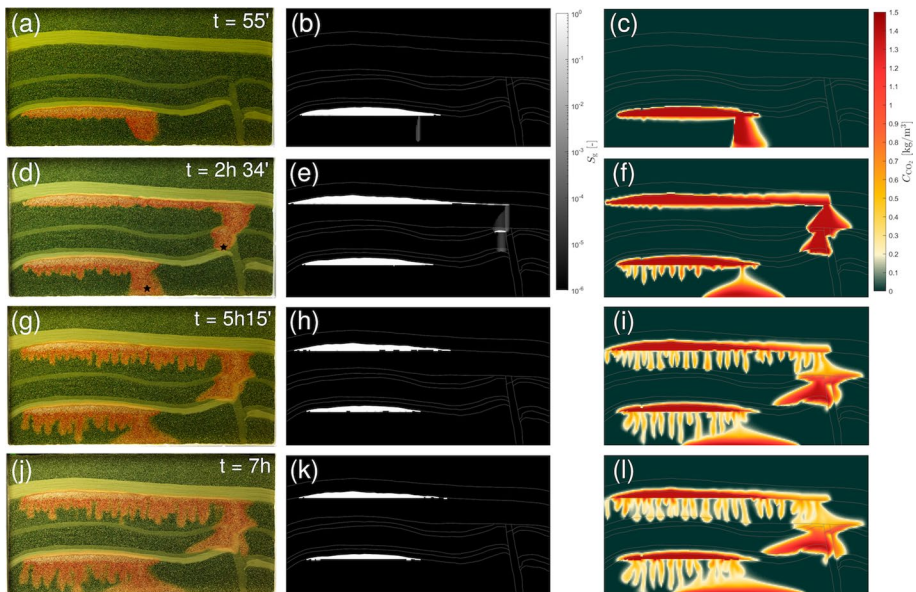
Fig. 12 Comparison between Experiment A1 in Tank 1 (left column) and simulation results with Model 3 after calibration (gas saturation shown in middle column, and CO_2 concentration shown in right column). Location of injection ports shown by black stars in **d**. $D = 10^{-9}$ m²/s. **a–c** End of injection in lower port. **d–f** End of injection in upper port. **g–i** Time at which the first finger touches the rig bottom. **j–l** Time at which the first finger touches the middle C layer

4.3 Transferability: Model Forecasts

A key question after history matching a flow simulation model is whether the physical description has actually been improved, or whether parameters have been modified to match a set of specific observations only. By applying the history-matched models to a different injection protocol (Experiment A2 in Tank 1; refer to Table 2) and subsequently to a different geometry (Experiment B1 in Tank 2), this can be assessed to some extent.

4.3.1 Analog for a Longer CO_2 Injection in the Same Geologic Setting

This case illustrates concordance of our history-matched models in a much longer injection in the same geology (Experiment A2). Before simulating this case, we observed that the trapped gas column against the fault in the experiment was different than what could be achieved with our previous p_e for Models 1–3 (Table 4). Because the capillary properties of the C sand in the fault were not directly involved in Experiment A1, we increased p_e in our calibrated models for that specific region ($p_e = 5$ mbar against the lower F sand, and 3.5 mbar against the middle F sand). All other parameters were taken from the values calibrated to match Experiment A1.

Evaluation was performed at the end of injection, at $t = 4$ h 48 min, with a single run with Models 1–3. I_R and D were set to the same value in all three models: 1.7 ml/min and 10^{-9} m/s^2, respectively. The experimental result is shown in Fig. 13a, while the simulation with Model 3 is depicted in Fig. 13b, c. We observe that the general distribution of CO_2 is close to the experimental truth. However, the experiment shows a compact sinking front of the CO_2-rich water without fingers; in our model, gravity fingering is apparent at this stage and fingers are close to the bottom of the rig. Additionally, CO_2-saturated brine touches the right boundary in the upper F reservoir, which does not occur in the experiment. This is due to capillary breach of the C sand above the middle F reservoir, as shown in Fig. 13b, and can be avoided by reducing the gas saturation value at which p_e is defined, or by increasing p_e.

The comparison of areal quantities is provided in Fig. 14 and demonstrates good-to-very-good concordance. Models 2 (MAE = 16 cm^2, CCC = 0.996) and 3 (MAE = 14.54 cm^2, CCC = 0.996) are similarly accurate and slightly better than Model 1 (MAE = 20.18 cm^2, CCC = 0.988), but there are no marked differences.

4.3.2 Analog for a Larger-Scale CO_2 Injection in a Different Geologic Setting

Finally, we compare the forecasting ability of our calibrated models against Experiment B1, conducted in a larger-scale, more complex geology (Fig. 1e) (Flemisch et al. 2023,
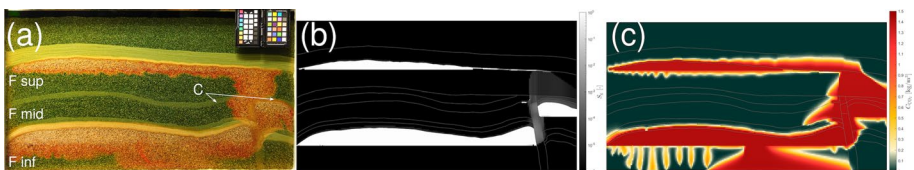

Fig. 13 Comparison between Experiment A2 in Tank 1 (**a**) and simulation results with Model 3 (**b**, **c**) at the end of the injection phase ($t = 4$ h 48 min)
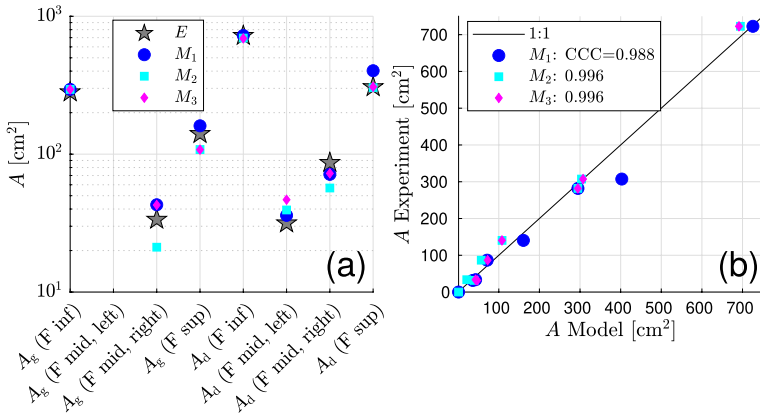
Fig. 14 a: Comparison of areas occupied by free gas (A_g) and water with dissolved CO_2 (A_d) for Experiment A2 in Tank 1. Experimental reference shown with a star (E). A_g (F mid, left) not shown because values are very close to 0. Refer to Fig. 6 or Fig. 13a for region location. **b** Concordance plot for each of the three models, using the same areal quantities as in **a**. Lin's CCC (Lin 1989) is shown in the key, according to Eq. 2

this issue). Similar to Sect. 4.3.1, our goal is to assess the forecasting ability of our calibrated models—without changing their properties. However, given that sand D controls migration in the lower fault (see Fig. 2d) and it was not present in our calibrated models, we allowed one change for Models 1 and 2, which did not have access to local p_c measurements. This means that we ran an initial simulation of this experiment with Models 1 and 2 and then adjusted the p_c curve of the D sand. The selected curve lies at $\approx \frac{1}{3}$ of the $p_c(S_w)$ shown in Fig. 3 and Fig. 4, respectively.

Next, we evaluate concordance of Models 1–3 by comparing them to the experimental truth after a single run. Evaluation is performed over the total duration of the experiment (120 h), which is simulated with the same I_R (10 ml/min) and D (10^{-9} m^2/s) in all three models (M_1, M_2, $M_{3,1}$). Additionally, a run with $D = 3 \times 10^{-9}$ m^2/s was completed with Model 3 ($M_{3,3}$) to better approximate finger widths, as noted in Sect. 4.2.

Gas saturation and CO_2 concentration maps at the end of injection with Model 1 are shown in Fig. 15a and Fig. 15b, respectively. The full visual comparison is provided in Fig. 16. We make the following observations:

- At the end of injection ($t = 5$ h), all three models forecast some migration of CO_2 into Box B. Models 2 (Fig. 16c) and 3 (Fig. 16d) underestimate the amount of CO_2, while Model 1 (Fig. 16b) overestimates the amount of CO_2 in the top C sand.
- Also at the end of injection, all models forecast faster sinking of the CO_2-charged water tongue arising from the lower injector. This is due to the higher F sand permeability required to match finger advance (see Sect. 4.2), particularly in Model 3 with $D = 3 \times 10^{-9}$ m^2/s.
- The speed at which CO_2-rich fingers sink is slightly faster in our models, compared to the experiment. As expected, Model 3, with a higher diffusion coefficient, displays thicker fingers, with closer widths to the experiment. Similar to our previous
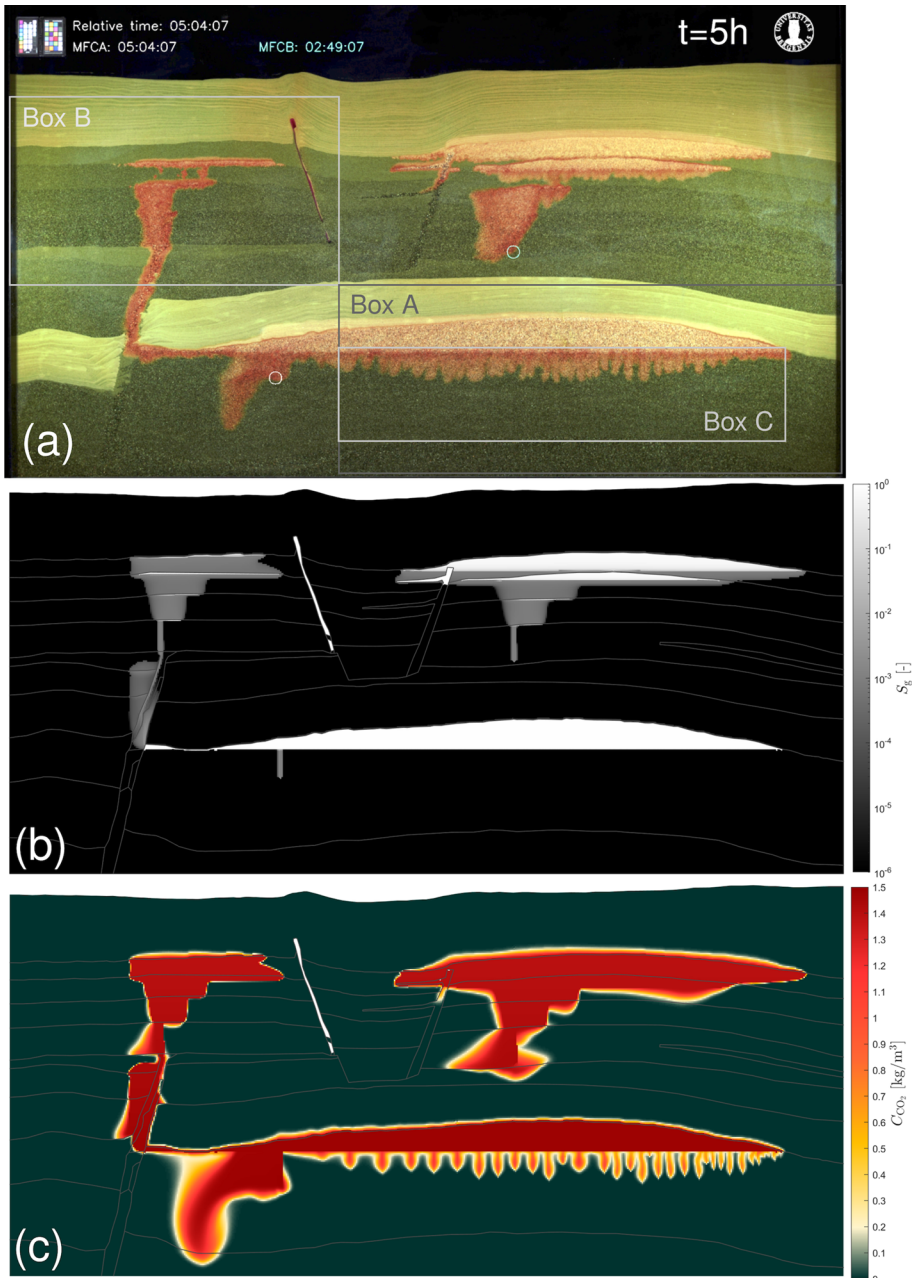
Fig. 15 Comparison between Experiment B1 in Tank 2 (**a**) and simulation Model 1 (**b**, **c**) at the end of injection ($t = 5$h). Circles in **a** denote the location of injection ports

observations, the numerical models cannot approximate the compact, CO_2-rich water front closely trailing the fingers.
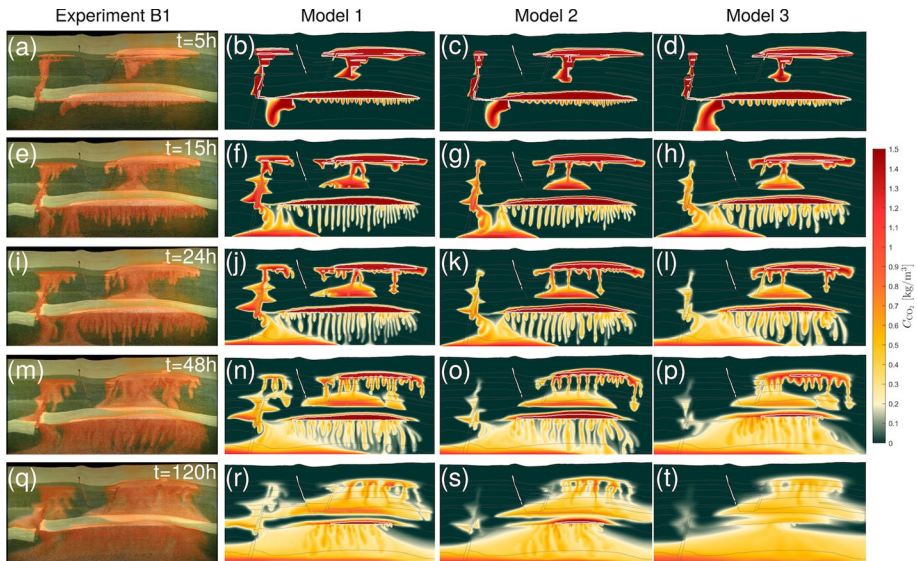
Fig. 16 Comparison between Experiment B1 in Tank 2 (leftmost column) and CO_2 concentration maps for simulation Models 1–3 (middle-left, middle-right and rightmost, respectively). $D = 10^{-9}$ m²/s (Model 1 and 2), $D = 3 \times 10^{-9}$ m²/s (Model 3). The white contours in simulation plots indicate $S_g = 10^{-3}$. **a–d** End of injection. **e–h** $t = 15$h. **i–l** $t = 24$h. **m–p** $t = 48$h. **q–t** $t = 120$h

- Dissolution of CO_2 is underestimated by Models 1 and 2, while it is closer, but overestimated, by Model 3.

Consistent with our approach described in Sect. 3.3, quantitative analysis is provided by means of areal quantities over time in Fig. 17. Experimental values were obtained via segmentation of timelapse images, and the data were reported on a 1×1 cm grid where 0 is pure water, 1 is water with dissolved CO_2, and 2 is gaseous CO_2. The segmentation procedure is explained in Nordbotten et al. (2023, this issue). We then obtained the areas of each phase within Box A and B to generate Fig. 17 (refer to Fig. 15a for box location).

In Box A, which contains the main F reservoir and ESF seal, we observe very good concordance (accurate areas) during injection. Afterwards, Model 3 with $D = 3 \times 10^{-9}$ m²/s continues to follow the experiment closely, whereas the others overestimate gaseous CO_2. Note that the PVT properties of our fluids are the same in all models; differences arise due to (1) higher sand F S_{wc} in Model 3, and higher sand F k in Model 2 and especially Model 3 ($D = 3 \times 10^{-9}$ m²/s), compared to Model 1, which allow greater convective mixing (Ennis-King and Paterson 2005) (Table 4); and (2) lower p_e and higher k of sand ESF in Model 2 (Table 4), which allows some CO_2 migration into the seal (Fig. 16). In Box B (Fig. 17d–f), Model 1 and Model 3 with $D = 10^{-9}$ m²/s are able to approximately track the experimental truth during injection. However, our models without dispersion cannot capture the areal increase of CO_2-rich water that occurs afterwards (cf. Fig. 16).

To put these results in perspective, Fig. 18 provides a comparison with results submitted by the international benchmark study (IBS) participants, as well as Experiment B1 (Flemisch et al. 2023, this issue). Figure 18 presents, for each datapoint, mean Wasserstein distances to experiments and forecasts (simulations by IBS participants). Specifically,
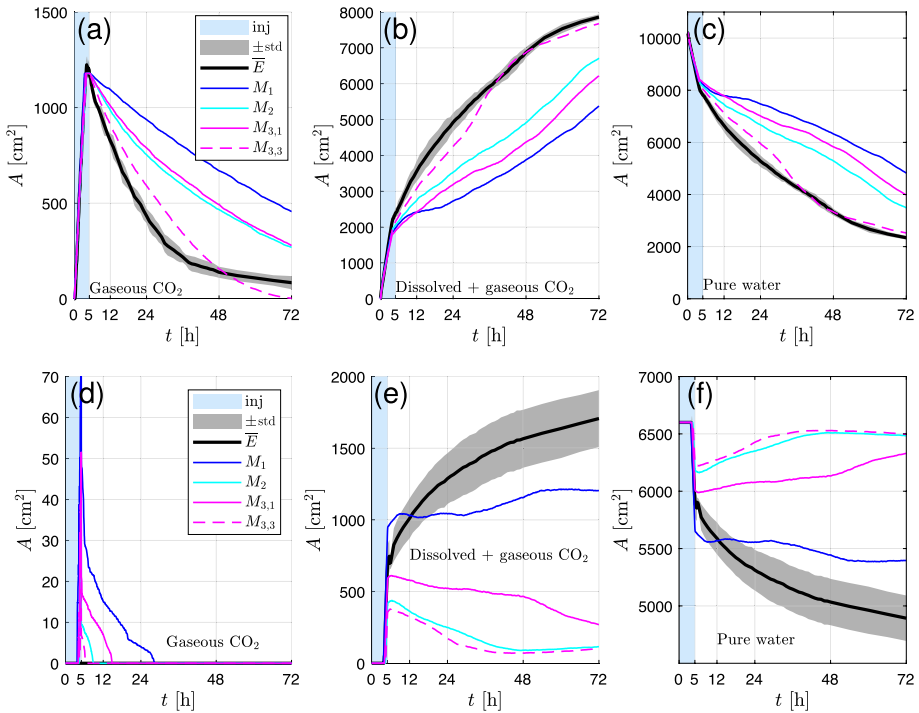
Fig. 17 Comparison of areas occupied by each phase during the first 72 h of case B1. Experimental mean (E) and standard deviation (std) obtained from four experimental runs with identical protocol, while the results for Models 1–3 are for a single run with each matched model. For M_3, two cases are shown: $D = 10^{-9}$ m^2/s ($M_{3,1}$) and $D = 3 \times 10^{-9}$ m^2/s ($M_{3,3}$). Top row shows areas in Box A, and bottom row shows areas in Box B. **a**, **d** Gaseous CO_2. **b**, **e** Dissolved CO_2 (includes area with gaseous CO_2). **c**, **f** Pure water

the Wasserstein metric (W) measures "the minimal effort required to reconfigure the probability mass of one distribution in order to recover the other distribution" (Panaretos and Zemel 2019). We expect $W \to 0$ for two samples from the same distribution, given enough values, and two samples to be more similar or concordant the closer W is to 0. To calculate distances shown in Fig. 18, the cell mass density in a 1×1 cm grid was estimated for all simulations and experiments and then normalized. Therefore, this metric provides a measure of the overall degree of agreement (i.e., in the whole domain). Resulting distances were dimensionalized using the total CO_2 mass in the system, such that the units are grams × centimeter, with values < 100 gr cm and < 50 gr cm representing good concordance and very good concordance, respectively. Details and code are provided by Flemisch et al. (2023, this issue). In Fig. 18, it can be seen that M_1–M_3 are comparable to or better than the best forecasts by IBS participants. M_1 and $M_{3,1}$, in particular, achieved very good concordance.

Further evaluation of simulation model concordance, including comparison with model results before calibration, mass quantities and error measures, is provided in Appendix A. From this analysis (Sect. 4.3 and Appendix A), we find that:
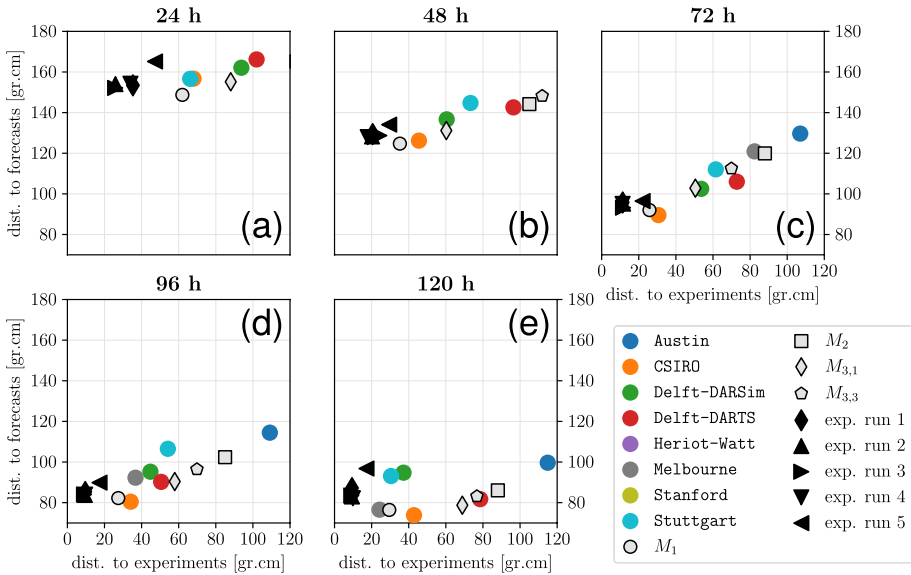
Fig. 18 Wasserstein distances to experiments and forecasts (simulations). Colored circles show forecasts by IBS groups, and results with calibrated Models 1–3 are presented with light gray markers. In each subplot, the vertical axis shows the mean distance between a given datapoint and the forecasts (considering the IBS participants only), while the horizontal axis shows the mean distance between a given datapoint and the experiments. Markers not present fall outside of the axes limits. **a** 24 h. **b** 48 h. **c** 72 h. **d** 96 h. **e** 120 h

- All matched models approximate well CO_2 migration and distribution in the domain, seal capacity, and onset of convective mixing. M_1 and $M_{3,1}$ are most concordant to experiments (Fig. 18).
- Calibrated models are able to accurately estimate specific quantities during the injection phase, yet they accumulate higher errors later on (Fig. 17 and Appendix B).
- Similar to Experiment A1, the calibration procedure significantly improved the concordance of M_1 and M_2 with the experiment (Figs. 19 and 16). In Box A, calibration also improved concordance for M_3 (Figs. 20 and 24). Overall, however, matched $M_{3,1}$ and $M_{3,3}$ are less concordant than their initial versions, which were already in very good agreement with the experiment (Figs. 21 and 18).

In summary, calibrated models are transferable to a different operational setting or geologic structure, as long as sediments and trap systems remain the same (Experiment A2 and Box A in Experiment B1). Where reservoir connectivity is provided by heterogeneous structures with uncertain properties, accurate deterministic estimates of CO_2 migration are unlikely; models calibrated elsewhere (Experiment A1) were not accurate in our test (Box B in Experiment B1). Given unlimited computational time, the forecasting capability of numerical models calibrated with published data appears similar to those having access to local measurements; the main value of local data lies in reducing the time required for history matching. Obtained results suggest that history matching worsened M_3 forecasts in a different setting (Experiment B1). Therefore, forecasts in a given geologic setting may benefit more from local measurements

and accurate physics, rather than history matching, unless historical data of the same setting are available. This is because CO_2-brine flow is very sensitive to variations in petrophysical properties such as capillary pressure, which will change in different areas, even if the geology is similar.

5 Discussion

In the FluidFlower, strong buoyancy and high permeability lead to persistent appearance and disappearance of fluid phases, as the gas migrates upward and dissolves in the water; coupled with other two-phase flow nonlinearities, these aspects make this problem difficult to solve numerically (e.g., Lie 2019). Comparison between the number of nonlinear iterations and the strength of different physical mechanisms (flow rates, buoyancy, capillarity and dissolution) is presented in Appendix B. A clear correlation can be seen between flow rates and number of iterations. However, buoyancy, capillarity and dissolution all appear to be playing a role, and it is not straightforward to discern which effect dominates; hence, this is a topic that requires further study. We note that difficulties with the convergence of the nonlinear solver have been reported by all participants in the international benchmark study (Flemisch et al. 2023, this issue). As hinted in Sect. 3.1, we addressed this by optimizing linear solver time, reducing the time-step length, increasing the number of time-step cuts and relaxing MRST's maximum normalized residual where required.

In a 2D isotropic medium and assuming uniform flow, the hydrodynamic dispersion coefficient ($\underline{\underline{D}}_{\mathrm{h}}$) can be modeled as $\underline{\underline{D}}_{\mathrm{h}} = \begin{bmatrix} \alpha_{\mathrm{L}}\overline{u} & 0 \\ 0 & \alpha_{\mathrm{T}}\overline{u} \end{bmatrix}$, where α_{L} and α_{T} are the longitudinal and transverse dispersivity, respectively, and $\overline{u}$ is the average Darcy velocity (Bear 1972). Assuming dispersivities $\geq 10^{-3} - 10^{-2}$ m (Garabedian et al. 1991; Gelhar et al. 1992; Schulze-Makuch 2005) and $\overline{u} \approx 3 \times 10^{-6}$ m/s (from our simulations), we get $\underline{\underline{D}}_{\mathrm{h}} \in [3 \times 10^{-9}, 3 \times 10^{-8}]$ m^2/s or larger; this means that $\underline{\underline{D}}_{\mathrm{h}} \geq D$ for the timescales considered (Riaz et al. 2004; Rezk et al. 2022). We also note that numerical dispersivity is on the order of the cell size ($h \approx 4$ mm in Tank 1, and ≈ 5 mm in Tank 2), so it is likely smaller than physical dispersivity. Numerical diffusion can be approximated as uh, which yields maximum values $\sim O(10^{-7}$ m^2/s) (water phase). However, using the mean of the 75$^{\text{th}}$ percentile flow velocity over all time-steps, we obtain $\sim O(10^{-9}$ m^2/s). Therefore, we estimate that numerical diffusion is lower than physical diffusion almost everywhere in our simulations. Previous work suggested that hydrodynamic dispersion in homogeneous sediments can be accounted for by increasing D (Riaz et al. 2004, 2006), as done here. However, our analysis shows that the spreading of CO_2-rich water during convective mixing can be loosely, but not accurately, represented by molecular diffusion. Given (1) the dominance of convective mixing on solubility trapping (Ennis-King and Paterson 2005; Neufeld et al. 2010; MacMinn and Juanes 2013; (2) heterogeneity of many natural reservoirs, which increases the importance of dispersion (Riaz et al. 2006; Bear 2018); and (3) the acceleration of CO_2 dissolution due to dispersion, as observed here and by others (e.g., Hidalgo and Carrera 2009), it is important to quantify the balance between diffusion and dispersion to estimate CO_2 trapping.

Our study of CO_2 injection and migration in unconsolidated sands at atmospheric p, T conditions captures the CO_2-water system dynamics at short to intermediate timescales: buoyancy-driven flow and structural trapping (Bachu et al. 1994; Bryant et al.

2008; Hesse and Woods 2010; Szulczewski et al. 2013), residual trapping (Juanes et al. 2006; Burnside and Naylor 2014) and convective mixing and dissolution trapping (Weir et al. 1996; Ennis-King and Paterson 2005; Riaz et al. 2006; Neufeld et al. 2010; Hidalgo et al. 2012; MacMinn and Juanes 2013; Szulczewski et al. 2013). Due to the very large sand permeability ($10^2 - 10^4$ D), convective mixing and dissolution dominate CO_2 trapping. With respect to values at ~ 1 km depth ($p \sim 100$ bar, $T \sim 40$ C), the dynamic viscosity and density of CO_2 are $\approx 1/3$ and 3×10^{-3}. Conversely, previous studies with similar setups used analogous fluids with density and viscosity ratios similar to supercritical CO_2-brine (Trevisan et al. 2017; Krishnamurthy et al. 2022). Dynamics observed in these systems are similar to ours, with vertical migration of CO_2 dominated by buoyancy and lateral spreading of CO_2 plumes with a main tongue at the top of the aquifer or high permeability layer. A quantitative scaling analysis of the FluidFlower (Tank 2) was performed by Kovscek et al. (2023, this issue), who showed that scaling of physical mechanisms to the field scale is possible. Compared to three CO_2 storage projects (Northern Lights, Sleipner and In Salah) the vertical dimension of the storage reservoir is exaggerated 2 to 3 times. Temporally, 1 h in the FluidFlower is equivalent to $\sim 100 - 400$ y in the field; thus, the experiment in Tank 2 (120 h) covers well the injection and post-injection periods. Similar to the FluidFlower, Kovscek et al. (2023, this issue) estimate the onset of convective mixing to occur during injection in high-permeability formations like the Utsira Sand (Sleipner). This analysis demonstrates that observations made in the FluidFlower can be used to describe field-scale fluid dynamics and quantify forecasting accuracy.

Our models retained some error at the end of the calibration phase, which is a known problem of manual history matching (Oliver and Chen 2011). Consistent with previous findings (e.g., Fisher and Jolley 2007), results show that Model 2 and 3, which had access to local data, achieved faster match to the experimental truth than Model 1 (Sect. 4.2). However, all models seem to have similar forecasting capability (Sect. 4.3). Subsurface heterogeneity and time constraints may explain why, in practice, it is critical to include local data to achieve history matching, and, especially, concordant forecasting (e.g., Gosselin et al. 2003; Fisher and Jolley 2007; Myers et al. 2007; Kam et al. 2015; Avansi et al. 2016). Calibration with Experiment A1 decreased overall concordance of Model 3 to Experiment B1 (but improved concordance in Box A), compared to forecasts with initial (measured) parameter values. We interpret this to be the result of fluid migration in Experiment A1 being controlled by different units than in Box B in Experiment B1. Therefore, local measurements are paramount, especially if historical data in the trap system of interest are not available.

Additionally, we did not quantify uncertainty in history-matched models due to the availability of a ground truth. In general, however, this is necessary to manage reservoir operations (e.g., Aanonsen et al. 2009; Oliver and Chen 2011; Jagalur-Mohan et al. 2018; Jin et al. 2019; Liu and Durlofsky 2020; Santoso et al. 2021, and references therein). It is also important to note that history-matched models may have grid-size dependencies (see Appendix C), which may require that the grid used to make forecasts, if different or encompassing additional regions, maintain a similar resolution. Finally, multiphase flow in poorly lithified sediments is non-unique (Haugen et al. 2023, this issue), which also contributes to uncertainty. Therefore, it seems prudent to adopt a probabilistic perspective when estimating subsurface CO_2 migration. This is consistent with results in Fig. 18 and Flemisch et al. (2023, this issue): in the highly resolved and geologically simple FluidFlower (compared to the subsurface), forecasts by different simulation groups show large spread.

6 Conclusions

We performed experiments (Sect. 2) and numerical simulations (Sect. 3) of CO_2 migration in poorly lithified, siliciclastic sediments at the meter scale. Three simulation model versions, with access to different levels of local data, were manually history-matched to the experiments (Sects. 4.1, 4.2), and then used to make forecasts (Sect. 4.3). The main findings are:

1. The time required to history match Model 3 (access to both single-phase and multiphase measurements) is lower than Model 2 (access to local single-phase measurements), which is lower than Model 1 (no access to local petrophysical measurements).
2. All simulation models achieve a satisfactory qualitative match throughout the experiments. Quantitatively, forecasting capability of Models 1–3 appears similar: in specific domain regions, models were close to the experimental truth during CO_2 injection, and accumulated larger errors afterwards, especially where heterogeneous structures control CO_2 migration.
3. Overall forecasts with Model 3 after calibration in a similar, but not identical, geologic setting were less accurate than forecasts made with measured values. This emphasizes the importance of local measurements and history matching in the same geologic setting.
4. The addition of a constant molecular diffusion coefficient allows matching convective finger widths to experimental observations. However, simulations without dispersion cannot approximate the compact, CO_2-rich sinking front closely trailing convective fingers in our experiments.

Simulation models were not always accurate. Given the degree of control in our study, it seems prudent to quantify uncertainty when assessing subsurface CO_2 migration in the field using numerical models. Obtained results suggest that confidence can be increased by obtaining local data, quantifying petrophysical parameter uncertainty, testing sensitivity to petrophysical parameters in different model regions, using historical data from the same setting and including post-injection data when history matching, and incorporating multiple scenarios of CO_2 migration, particularly where heterogeneous structures are at play.

Appendix A: Additional Analysis of Simulation Model Concordance with Experiment B1

A.1 Results with Initial Model Parameters

Figure 19 compares Experiment B1 and concentration maps from simulations with initial parameters, for each of the three model versions considered. Qualitatively, all models estimate the location of the two main gas plumes correctly, but it is clear that Models 1 and 2 are less concordant to the experiment than Model 3. This is particularly true in the upper left of the domain, where CO_2 migration is controlled by the heterogeneous fault. Similar to results presented in Sect. 4.1, Model 3 is already very close to the experiment, although the advance of convective fingers is slower.

Concordance between our initial models and the simulation is shown in Fig. 20 by means of the ratio between the model and experimental areas for different quantities in Box
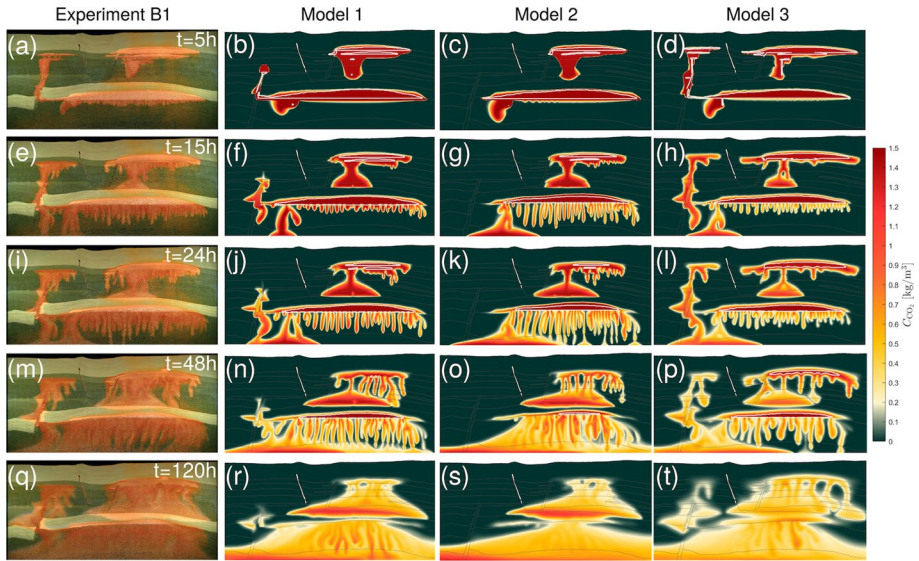
Fig. 19 Comparison between Experiment B1 in Tank 2 (leftmost column) and CO_2 concentration maps for simulation Models 1–3 (middle-left, middle-right and rightmost, respectively) with initial parameters. $D = 10^{-9}$ m^2/s (Model 1 and 2), $D = 3 \times 10^{-9}$ m^2/s (Model 3). **a–d** End of injection. **e–h** $t = 15$h. **i–l** $t = 24$h. **m–p** $t = 48$h. **q–t** $t = 120$h

A and Box B (see Fig. 15 for box location). Values below 1 indicate that the model underestimates the areal extent of a given quantity, while values above 1 indicate that the model overestimates it. During the first 48–72 h, all models except $M_{3,1}$ are reasonably close to the experiment in Box A. Afterwards, M_1, M_2 and $M_{3,3}$ forecast earlier dissolution of the CO_2 plume, while $M_{3,1}$ forecasts later dissolution. In Box B, concordance is relatively good for M_3 during the first 48 h, but model accuracy diminishes with time for all model versions.

Further comparison between our initial model results and experimental values are provided in Fig. 21, where we evaluate mean Wasserstein distances to the international benchmark study (IBS) participants' forecasts and experiments (Flemisch et al. 2023, this issue). Figure 21 is consistent with Fig. 19, where it can be seen that M_3 is already very close to the experiment and is similarly concordant or more concordant than the best of the IBS participants.

A.2 Calibrated Models

First, we provide the total mass of CO_2 in the computational domain in Fig. 22, and the mass in Boxes A and B in Fig. 23.

Next, in Table 5, the following measures are compared with quantities estimated from the experiment via segmentation of timelapse images (Nordbotten et al. 2023, this issue). These measures correspond to the sparse data requested to participants of the FluidFlower IBS (Flemisch et al. 2023, this issue):
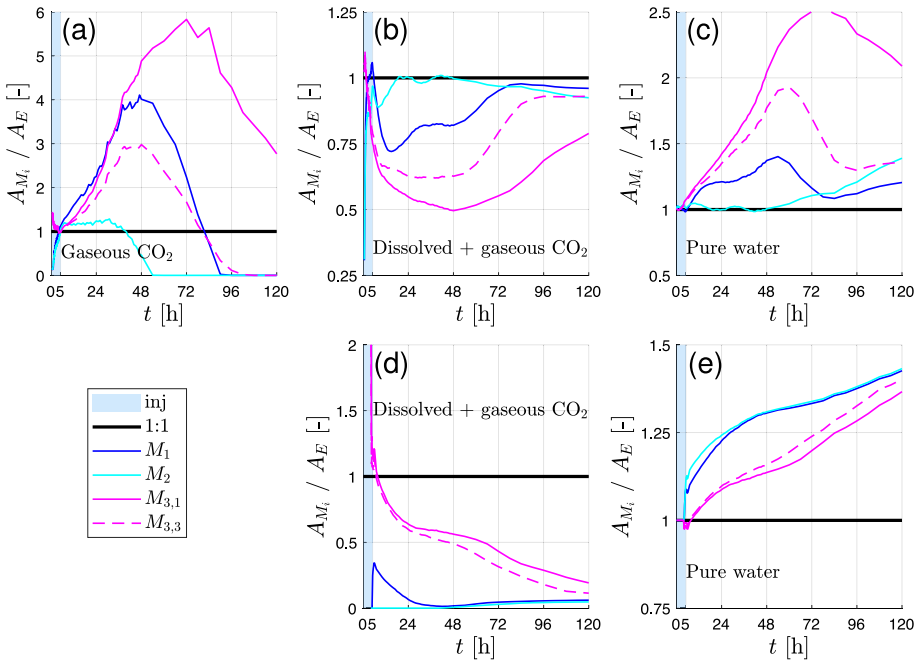
Fig. 20 Ratios between model (A_{M_i}) and experimental mean (A_E) areas occupied by each phase. Experimental mean was obtained from four experimental runs with identical protocol, while the results for Models 1–3 are for a single run. For M_3, two cases are shown: $D = 10^{-9}$ m²/s ($M_{3,1}$) and $D = 3 \times 10^{-9}$ m²/s ($M_{3,3}$).

Top row shows Box A, and bottom row shows Box B. Ratios for gaseous CO_2 in Box B are not computed because experimental values are 0. **a** Gaseous CO_2. **b**, **d** Dissolved CO_2 (includes area with gaseous CO_2). **c**, **e** Pure water

1. Time of maximum mobile free phase in Box A
2. Mass of mobile $CO_{2(g)}$, immobile $CO_{2(g)}$, dissolved CO_2, and CO_2 in the seal (in any phase), in Box A, 72 h after injection start (2a-d)
3. The same quantities as 2. for Box B (3a-d)
4. time at which m (defined below) exceeds 110% of the width of Box C
5. Total mass of CO_2 in the ESF seal, in Box A, at $t = 120$ h

Convective mixing in Box C (see Fig. 1e) is reported as the integral of the magnitude of the gradient in relative concentration of dissolved CO_2 (Flemisch et al. 2023):

$$m(t) = \int_C \left| \nabla \left(\frac{\chi^{w}_{CO_2}}{\chi^{w,max}_{CO_2}} \right) \right| d\mathbf{x} \tag{A1}$$

where $\chi^{w}_{CO_2}$ is the mass fraction of CO_2 in water, and the dissolution limit is $\chi^{w,max}_{CO_2}$. Note that quantity 4, based on m, cannot be provided with full accuracy based on experimental data, so an uncertain lower and upper bound is provided instead. Therefore, error is not computed in Table 5.

Relative error is evaluated with respect to the experimental mean ($\bar{E}$) as $\varepsilon_i(\%) = 100 \times \frac{|\bar{E}_i - M_{J,i}|}{\bar{E}_i}$, where i is a given measure and J refers to any of the Models 1–3. In
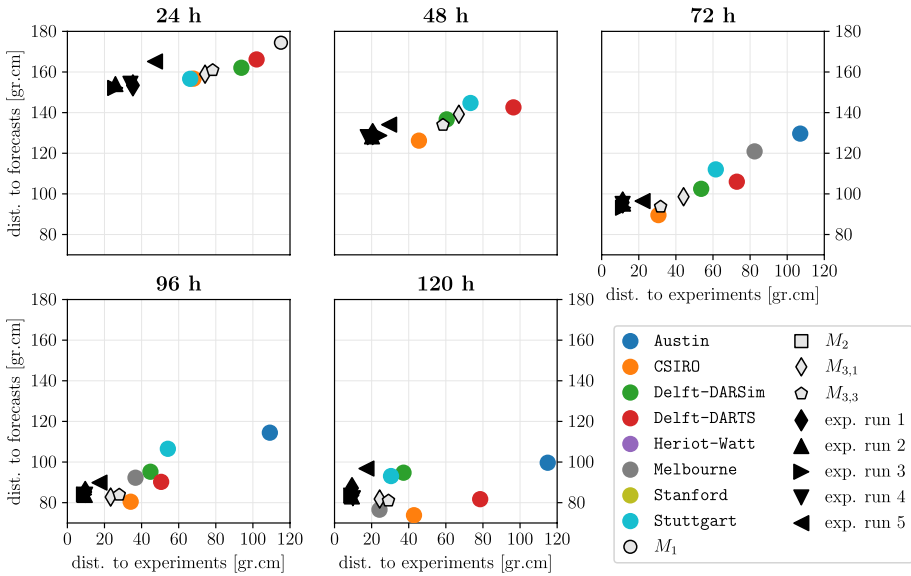
Fig. 21 Wasserstein distances to experiments and forecasts (simulations). Colored circles show forecasts by IBS groups, and results with initial Models 1–3 are presented with light gray markers. In each subplot, the vertical axis shows the mean distance between a given datapoint and the forecasts (considering the IBS participants only), while the horizontal axis shows the mean distance between a given datapoint and the experiments. Markers not present fall outside of the axes limits. See Sect. 4.3.2 for details
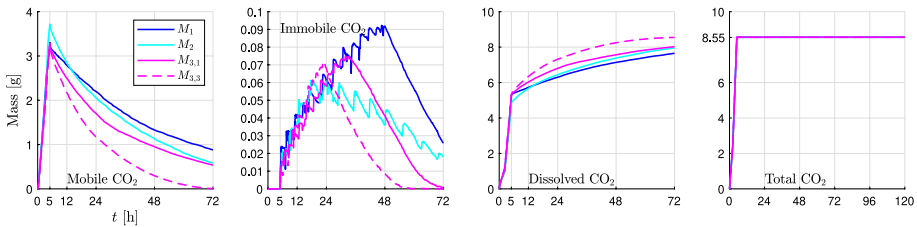
Fig. 22 Total mass of CO_2 for our simulations of Experiment B1 presented in Sect. 4.3.2. Results are provided for Models 1–3. For M_3, two cases are shown: $D = 10^{-9}$ m^2/s ($M_{3,1}$) and $D = 3 \times 10^{-9}$ m^2/s ($M_{3,3}$)

Table 5, it can be seen that all models accumulate some error in most of the quantities reported. The maximum errors are $\approx 140\%$ for Models 1–2 and $< 100\%$ for Model 3. Model 1 is more concordant in the uncertain region (Box B; see Sect. 4.3.2 as well), while Models 2 and 3 are more accurate in Box A, the region where the calibration performed with Experiment A1 is more meaningful. Overall, $M_{3,1}$ does marginally better.

We provide additional analysis in Fig. 24, which shows ratios between model and experimental areas, similar to Fig. 20. As shown in Sect. 4.3.2, $M_{3,3}$ is most concordant in Box A, while M_1 and $M_{3,1}$ do better in Box B. Compared to Fig. 20, the maximum ratio is reduced. In Box A ($t < 72$h), Models 1 and 2 are less accurate than in Fig. 20, but this is not representative of their concordance in the whole domain (Sect. 4.3.2).
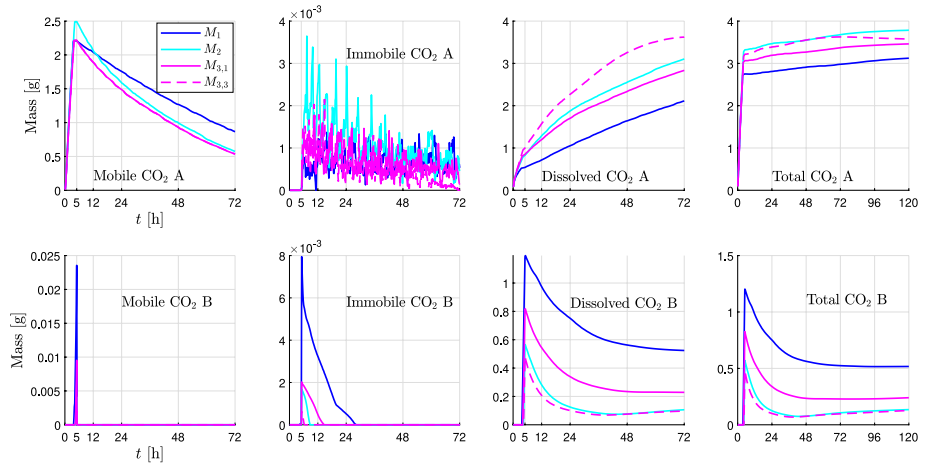
Fig. 23 Mass of CO_2 in Boxes A and B defined in Fig. 1e, for our simulations of Experiment B1 presented in Sect. 4.3.2. Results are provided for Models 1 to 3. For M_3, two cases are shown: $D = 10^{-9}$ m^2/s ($M_{3,1}$) and $D = 3 \times 10^{-9}$ m^2/s ($M_{3,3}$)

Appendix B: Nonlinear Solver Number of Iterations

According to fluid migration in the FluidFlower, flow dynamics are initially dominated by injection rates, then by buoyancy of the gas phase, and finally by capillarity and dissolution. In Fig. 25, we present, for the experiment in Tank 2, the relationship between the number of iterations, the maximum Darcy velocity (u) and the maximum concentration rate ($\dot{C}$), evaluated as dC/dt, as a function of time. Additionally, we estimated the maximum values of the dimensionless Reynolds (R_e, see Eq. 1), Capillary (C_a) and Bond (B_o) numbers during and after injection (e.g., Bear 1972):

$$C_a = \frac{\mu_\alpha u_\alpha}{\sigma} \tag{B1}$$

$$B_o = \frac{\Delta \rho g k}{\sigma} \tag{B2}$$

where μ is the dynamic viscosity, u the Darcy velocity, σ the interfacial tension, $\Delta\rho$ the density difference, g the gravity, k the permeability, and subscript α denotes a generic fluid phase. Max $B_o \sim O(10^{-3})$ and remains constant in our system. Max $C_a \sim O(10^{-6})$, $\sim O(10^{-7})$ for water and $\sim O(10^{-6})$, $\sim O(10^{-8})$ for gas (during and after injection, respectively), while max $R_e \sim O(10^{-2})$ for water and $\sim O(10^{-1})$, $\sim O(10^{-2})$ for gas (during and after injection, respectively).

From Fig. 25, a correlation between max $|u_{h,g}|$ and the number of iterations is apparent during injection. The number of iterations increases significantly after an injection port becomes active, and also when CO_2 spills out of the lower reservoir and starts migrating along the lower fault (see Fig. 15); this occurs at $t \approx 215$ min and $t \approx 250$ for M_1 and $M_{3,3}$, respectively. Peaks in $\dot{C}$ appear at the onset of injection, but we do not observe significant variations otherwise. Values from the dimensionless groups

Table 5 Sparse data comparison between Experiment B1 in Tank 2 and simulation results with Models 1–3

Measure	$\overline{E}$	$\sigma(E)$	M_1	ϵ_1 (%)	M_2	ϵ_2 (%)	$M_{3,1}$	$\epsilon_{3,1}$ (%)	$M_{3,3}$	$\epsilon_{3,3}$ (%)
1 (s)	14,880	720	17,700	19	17,160	15.3	17,280	16.1	18,000	21
2a (g)	0.36	0.13	0.87	140.7	0.57	59.7	0.54	48.7	0.005	98.8
2b (g)	0	0	0	0	0	0	0	0	0	0
2c (g)	3.5	0.08	2.11	39.7	3.1	11.3	2.8	19	3.6	3.5
2d (g)	–	–	0.43	n/a	0.97	n/a	0.74	n/a	0.7	n/a
3a (g)	0	0	0	0	0	0	0	0	0	0
3b (g)	0	0	0	0	0	0	0	0	0	0
3c (g)	0.55	0.32	0.52	4.7	0.1	80.7	0.23	58.4	0.1	82.4
3d (g)	n/a	n/a	0.002	n/a	0.006	n/a	0.002	n/a	0.004	n/a
4 (s)	[12,180, 17,990]	[438, 2261]	15,000	n/a	15,000	n/a	18,600	n/a	15,600	n/a
5 (g)	0.38	0.047	0.52	37.1	0.94	148.5	0.73	91.6	0.62	63.1
$\overline{\epsilon}$ (%)	n/a	n/a	n/a	30.1	n/a	39.4	n/a	29.2	n/a	33.6

Experimental mean and standard deviation were obtained from six experimental runs with identical protocol, while the results for Models 1–3 are for a single run with each matched model. For M_3, two cases are shown: $D = 10^{-9}$ m²/s ($M_{3,1}$) and $D = 3 \times 10^{-9}$ m²/s ($M_{3,3}$). Experimental quantity 4 is reported using a lower and upper bound due to high uncertainty, so errors are not computed. See main text for measure description
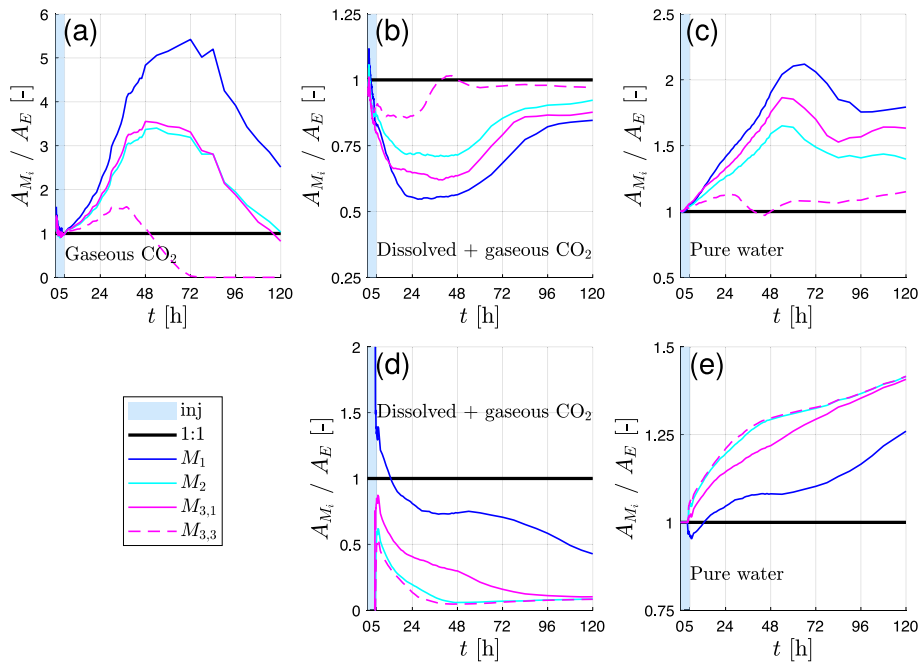
Fig. 24 Ratios between calibrated model (A_{M_i}) and experimental mean (A_E) areas occupied by each phase in case B1. Experimental mean was obtained from four experimental runs with identical protocol, while the results for Models 1–3 are for a single run with each model. For M_3, two cases are shown: $D = 10^{-9}$ m^2/s ($M_{3,1}$) and $D = 3 \times 10^{-9}$ m^2/s ($M_{3,3}$). Top row shows Box A, and bottom row shows Box B. Ratios for gaseous CO_2 in Box B are not computed because experimental values are 0. **a** Gaseous CO_2. **b**, **d** Dissolved CO_2 (includes area with gaseous CO_2). **c**, **e** Pure water

are indicative of high flow rates (R_e close to 1), relatively strong capillary forces, compared to viscous forces ($C_a \sim O(10^{-6})$ or smaller), and appreciable buoyancy. We identify that high flow rates and sudden appearance/disappearance of fluid phases challenge the nonlinear solver during injection. Buoyancy and capillarity forces, which are active throughout the simulation, also impact convergence, but it is not straightforward to identify if one exerts a greater control on the number of iterations. After injection, we observe difficulties between $t \approx 315$ and 1440 min in M_1, and $t \approx 720$ and 1440 in $M_{3,3}$. Our analysis does not reveal why, so this is a topic that warrants further study.

Appendix C: Comparison of Simulation Results with Multiple Grid Resolutions

This section provides two comparisons of concentration maps obtained with Model 3 after the calibration presented in Sect. 4.2:
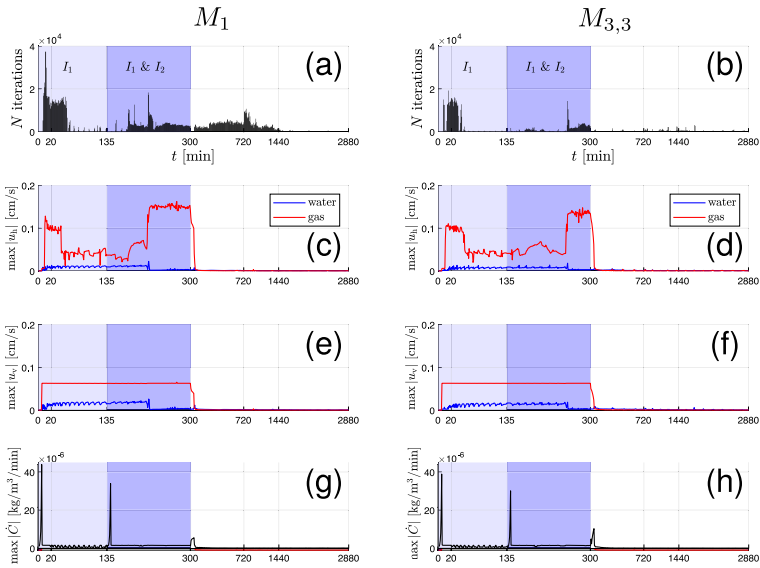
Fig. 25 Number of iterations and maximum values in the simulation domain for various quantities, as a function of time. Results are provided for M_1 (left column) and $M_{3,3}$ (right column). **a,b** Number of nonlinear solver iterations. **c,d** Horizontal Darcy velocity. **e,f** Vertical Darcy velocity. **g,h** Concentration rate ($\dot{C}$)
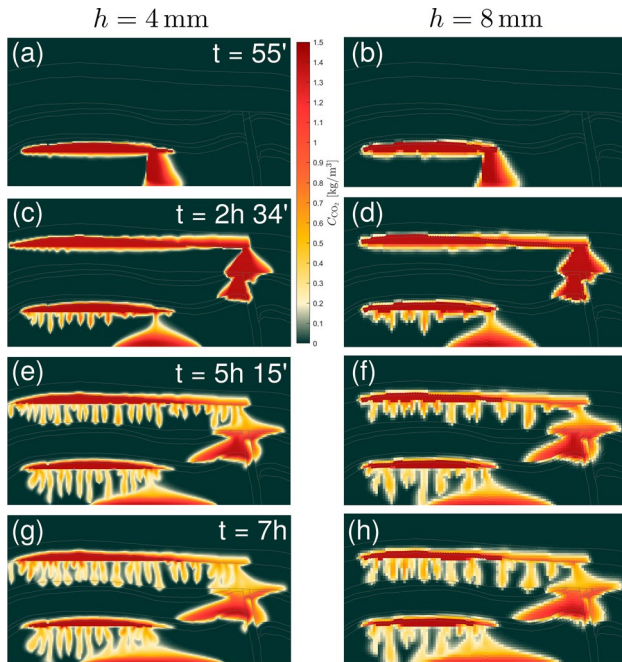
Fig. 26 Concentration maps from our simulations of Experiment A1 with Model 3. Results with two grids are shown: $h = 4\,\mathrm{mm}$ (**a, c, e, g**) and $h = 8\,\mathrm{mm}$ (**b, d, f, h**)
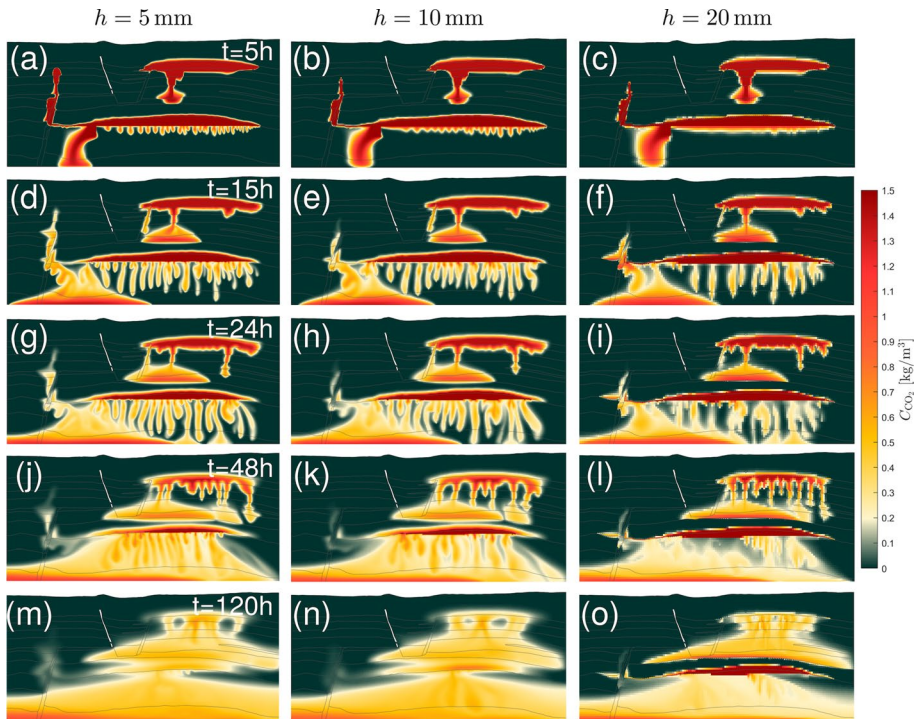
Fig. 27 Concentration maps from simulations of Experiment B1 with Model 3. Results with three grids are shown: $h = 5$ mm (**a, d, g, j, m**), $h = 10$ mm (**b, e, h, k, n**) and $h = 20$ mm (**c, f, i, l, o**)

1. For Experiment A1, we compare two grid sizes: $h = 4$ mm, as shown in the paper, and a coarser grid with $h = 8$ mm (Fig. 26).
2. For Experiment B1, we compare three grid sizes: $h = 5$ mm, used throughout the paper, and two coarser grids with $h = 10$ mm and $h = 20$ mm, respectively (Fig. 27). Note that, in the three simulations in Fig. 27, a total of 8.13 g of CO_2 was injected; this is slightly smaller than the 8.55 g actually injected in the experiment and in our simulations in the rest of the paper.

It can be seen that, for the calibrated parameter set (Table 4), the coarser models maintain a general agreement with the finer ones (and the experimental solution). However, some differences are clear even in this qualitative comparison, including (1) smaller extent of the CO_2 plume, (2) lower dissolution, (3) lower number of fingers and finger widths, and (4) different CO_2-rich finger sinking speed. Therefore, the calibration process is somewhat cell-size dependent, which has implications for applying history matched models from e.g., pilot tests to field-scale CO_2 storage projects.

Data Availability The dataset for Experiment B1 and data analysis scripts can be obtained from https://github.com/fluidflower. The remaining experimental and simulation data are available from the corresponding authors on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aanonsen, S.I., Nœvdal, G., Oliver, D.S., Reynolds, A.C., Vallès, B.: The ensemble Kalman filter in reservoir engineering—a review. SPE J. **14**(03), 393–412 (2009). https://doi.org/10.2118/117274-PA

Aissaoui, A.: Etude théorique et expérimentale de l'hystérésis des pressions capillaries et des perméabilitiés relatives en vue du stockage souterrain de gaz. Ph. D. thesis, Ecole des Mines de Paris, Paris (1983)

Alghannam, M., Juanes, R.: Understanding rate effects in injection-induced earthquakes. Nat. Commun. **11**(1), 3053 (2020). https://doi.org/10.1038/s41467-020-16860-y

Askar, A.H., Illangasekare, T.H., Ilie, A.M.C.: Monitoring brine leakage from deep geologic formations storing carbon dioxide: design framework validation using intermediate-scale experiment. Water Resources Res. **57**(12), 2021WR031005 (2021). https://doi.org/10.1029/2021WR031005

Avansi, G.D., Maschio, C., Schiozer, D.J.: Simultaneous history-matching approach by use of reservoir-characterization and reservoir-simulation studies. SPE Reserv. Eval. Eng. **19**(04), 694–712 (2016). https://doi.org/10.2118/179740-PA

Bachu, S., Gunter, W., Perkins, E.: Aquifer disposal of CO_2: hydrodynamic and mineral trapping. Energy Convers. Manag. **35**(4), 269–279 (1994). https://doi.org/10.1016/0196-8904(94)90060-4

Bear, J.: Dynamics of Fluids in Porous Media. Dover, Mineola (1972)

Bear, J.: Modeling Phenomena of Flow and Transport in Porous Media, vol. 31. Springer, Berlin (2018)

Beard, D., Weyl, P.: Influence of texture on porosity and permeability of unconsolidated sand. AAPG Bull. **57**(2), 349–369 (1973). https://doi.org/10.1306/819A4272-16C5-11D7-8645000102C1865D

Berge, R.L., Klemetsdal, Ø.S., Lie, K.A.: Unstructured Voronoi grids conforming to lower dimensional objects. Comput. Geosci. **23**, 169–188 (2019). https://doi.org/10.1007/s10596-018-9790-0

Berge, R.L., Klemetsdal, Ø.S., Skaflestad, B., Lie, K.A.: The UPR module. https://github.com/rbe051/UPR (2021)

Brock, D., Orr, F.: Flow visualization of viscous fingering in heterogeneous porous media. SPE Annual Technical Conference and Exhibition (SPE-22614-MS) https://doi.org/10.2118/22614-MS 1991

Brooks, R.H., Corey, A.T.: Hydraulic properties of porous media. Technical report, Colorado State University, Fort Collins, CO (1964)

Bryant, S.L., Lakshminarasimhan, S., Pope, G.A.: Buoyancy-dominated multiphase flow and its effect on geological sequestration of CO_2. SPE J. **13**(04), 447–454 (2008). https://doi.org/10.2118/99938-PA

Burnside, N., Naylor, M.: Review and implications of relative permeability of CO_2/brine systems and residual trapping of CO_2. Int. J. Greenh. Gas Control **23**, 1–11 (2014). https://doi.org/10.1016/j.ijggc.2014.01.013

Caine, J.S., Evans, J.P., Forster, C.B.: Fault zone architecture and permeability structure. Geology **24**(11), 1025–1028 (1996). https://doi.org/10.1130/0091-7613(1996)024<1025:FZAAPS>2.3.CO;2

Cappa, F., Rutqvist, J.: Impact of CO_2 geological sequestration on the nucleation of earthquakes. Geophys. Res. Lett. **38**(17), L17313 (2011). https://doi.org/10.1029/2011GL048487

Celia, M.A.: The role of subsurface engineering in the net-zero energy transition. In: IOP Conference Series: Earth and Environmental Science, vol. 861, IOP Publishing (2021)

Childs, C., Manzocchi, T., Walsh, J.J., Bonson, C.G., Nicol, A., Schöpfer, M.P.: A geometric model of fault zone and fault rock thickness variations. J. Struct. Geol. **31**(2), 117–127 (2009). https://doi.org/10.1016/j.jsg.2008.08.009

Cinar, Y., Jessen, K., Berenblyum, R., Juanes, R., Orr, F.M.: An experimental and numerical investigation of crossflow effects in two-phase displacements. SPE J. **11**(02), 216–226 (2006). https://doi.org/10.2118/90568-PA

Corey, A.T.: The interrelation between gas and oil relative permeabilities. Prod. Mon. **19**, 38–41 (1954)

Demidov, D., Rossi, R.: Subdomain deflation combined with local AMG: a case study using AMGCL library. arXiv preprint. https://doi.org/10.48550/arXiv.1710.03940 (2018)

DiCarlo, D.A., Sahni, A., Blunt, M.J.: Three-phase relative permeability of water-wet, oil-wet, and mixed-wet sandpacks. SPE J. **5**(01), 82–91 (2000). https://doi.org/10.2118/60767-PA

Doughty, C., Oldenburg, C.M.: CO_2 plume evolution in a depleted natural gas reservoir: modeling of conformance uncertainty reduction over time. Int. J. Greenh. Gas Control **97**, 103026 (2020). https://doi.org/10.1016/j.ijggc.2020.103026

Duan, Z., Sun, R.: An improved model calculating CO_2 solubility in pure water and aqueous NaCl solutions from 273 to 533 K and from 0 to 2000 bar. Chem. Geol. **193**(3–4), 257–271 (2003). https://doi.org/10.1016/S0009-2541(02)00263-2

Eikehaug, K., Haugen, M., Folkvord, O., Benali, B., Larsen, E.L., Tinkova, A., Rotevatn, A., Nordbotten, J.M., Fernø, M.: Engineering meter-scale porous media flow experiments for quantitative studies of geological carbon sequestration. Transp. Porous Media (this issue). Preprint available on arXiv through https://doi.org/10.48550/arXiv.2301.09304 (2023)

Ellsworth, W.L.: Injection-induced earthquakes. Science **341**(6142), 1225942 (2013). https://doi.org/10.1126/science.1225942

Ennis-King, J., Paterson, L.: Role of convective mixing in the long-term storage of carbon dioxide in deep saline formations. SPE J. **10**(03), 349–356 (2005). https://doi.org/10.2118/84344-PA

European Academies Science Advisory Council (EASAC). Policy Report 35, 2018. Negative emission technologies: what role in meeting Paris Agreement targets? https://easac.eu/publications/details/easac_net/. Accessed on 20 Jan 2023

Fernández-García, D., Illangasekare, T.H., Rajaram, H.: Conservative and sorptive forced-gradient and uniform flow tracer tests in a three-dimensional laboratory test aquifer. Water Resour. Res. **40**(10), W10103 (2004). https://doi.org/10.1029/2004WR003112

Fernø, M., Haugen, M., Eikehaug, K., Folkvord, O., Benali, B., Both, J.W., Storvik, E., Nixon, C.W., Gawthorpe, R.L., Nordbotten, J.M.: Room-scale CO_2 injections in a physical reservoir model with faults. Transp. Porous Media (this issue). Preprint available on arXiv through https://doi.org/10.48550/arXiv.2301.06397 (2023)

Fisher, Q., Jolley, S.: Treatment of faults in production simulation models. Geol. Soc. Lond. Spec. Publ. **292**, 219–233 (2007). https://doi.org/10.1144/SP292.13

Flemisch, B., Nordbotten, J.M., Fernø, M., Juanes, R., Class, H., Doster, F., Ennis-King, J., Franc, J., Geiger, S., Gläser, D., Green, C., Gunning, J., Hajibeygi, H., Jackson, S.J., Jammoul, M., Karra, S., Li, J., Matthäi, S.K., Miller, T., Shao, Q., Spurin, C., Stauffer, P., Tchelepi, H., Tian, X., Viswanathan, H., Voskov, D., Wang, Y., Wapperom, M., Wheeler, M.F., Wilkins, A., Youssef, A.A., Zhang, Z.: The FluidFlower international benchmark study: process, modeling results, and comparison to experimental data. Transp. Porous Media (this issue). Preprint available on arXiv through https://doi.org/10.48550/arXiv.2302.10986 (2023)

Garabedian, S.P., LeBlanc, D.R., Gelhar, L.W., Celia, M.A.: Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 2. Analysis of spatial moments for a nonreactive tracer. Water Resources Res. **27**(5), 911–924 (1991). https://doi.org/10.1029/91WR00242

Gaucher, D., Lindley, D.: Waterflood performance in a stratified, five-spot reservoir—a scaled-model study. Trans. AIME **219**(01), 208–215 (1960). https://doi.org/10.2118/1311-G

Gelhar, L.W., Welty, C., Rehfeldt, K.R.: A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. **28**(7), 1955–1974 (1992). https://doi.org/10.1029/92WR00607

Gosselin, O., Aanonsen, S., Aavatsmark, I., Cominelli, A., Gonard, R., Kolasinski, M., Ferdinandi, F., Kovacic, L., Neylon, K.: History matching using time-lapse seismic (HUTS). In: SPE Annual Technical Conference and Exhibition (SPE 84464). https://doi.org/10.2118/84464-MS (2003)

Hager, B.H., Dieterich, J., Frohlich, C., Juanes, R., Mantica, S., Shaw, J.H., Bottazzi, F., Caresani, F., Castineira, D., Cominelli, A., Meda, M., Osculati, L., Petroselli, S., Plesch, A.: A process-based approach to understanding and managing triggered seismicity. Nature **595**(7869), 684–689 (2021). https://doi.org/10.1038/s41586-021-03668-z

Hassanzadeh, H., Pooladi-Darvish, M., Elsharkawy, A.M., Keith, D.W., Leonenko, Y.: Predicting PVT data for CO_2-brine mixtures for black-oil simulation of CO_2 geological storage. Int. J. Greenh. Gas Control **2**(1), 65–77 (2008). https://doi.org/10.1016/S1750-5836(07)00010-2

Haugen, M., Saló-Salgado, L., Eikehaug, K., Benali, B., Both, J.W., Storvik, E., Folkvord, O., Juanes, R., Nordbotten, J.M., Fernø, M.: Physical variability in meter-scale laboratory CO_2 injections in faulted geometries. Transp. Porous Media (this issue). Preprint available on arXiv through https://doi.org/10.48550/arXiv.2301.07347 (2023)

Heinemann, Z., Brand, C., Munka, M., Chen, Y.: Modeling reservoir geometry with irregular grids. SPE Reserv. Eng. **6**(2), 225–232 (1991). https://doi.org/10.2118/18412-PA

Hesse, M., Woods, A.: Buoyant dispersal of CO_2 during geological storage. Geophys. Res. Lett. **37**(1), L01403 (2010). https://doi.org/10.1029/2009GL041128

Hidalgo, J.J., Carrera, J.: Effect of dispersion on the onset of convection during CO_2 sequestration. J. Fluid Mech. **640**, 441–452 (2009). https://doi.org/10.1017/S0022112009991480

Hidalgo, J.J., Fe, J., Cueto-Felgueroso, L., Juanes, R.: Scaling of convective mixing in porous media. Phys. Rev. Lett. **109**(26), 264503 (2012). https://doi.org/10.1103/PhysRevLett.109.264503

Hommel, J., Coltman, E., Class, H.: Porosity–permeability relations for evolving pore space: a review with a focus on (bio-)geochemically altered porous media. Transp. Porous Media **124**(2), 589–629 (2018). https://doi.org/10.1007/s11242-018-1086-2

Ingram, G.M., Urai, J.L.: Top-seal leakage through faults and fractures: the role of Mudrock properties. Geol. Soc. Lond. Spec. Publ. **158**, 125–135 (1999). https://doi.org/10.1144/GSL.SP.1999.158.01.10

Intergovernmental Panel on Climate Change (IPCC): Climate change 2022: mitigation of climate change. In: Shukla, P.R., Skea, J., Slade, R., Al Khourdajie, A., van Diemen, R., McCollum, D., Pathak, M., Some, S., Vyas, P., Fradera, R., Belkacemi, M., Hasija, A., Lisboa, G., Luz, S., Malley, J. (eds.) Contribution of Working Group III to the Sixth Assessment Report of the IPCC. Cambridge University Press, Cambridge (2022). https://doi.org/10.1017/9781009157926

Jagalur-Mohan, J., Jha, B., Wang, Z., Juanes, R., Marzouk, Y.: Inferring fault frictional and reservoir hydraulic properties from injection-induced seismicity. Geophys. Res. Lett. **45**(3), 1313–1320 (2018). https://doi.org/10.1002/2017GL075925

Jin, L., Lu, H., Wen, G.: Fast uncertainty quantification of reservoir simulation with variational U-Net (2019) https://doi.org/10.48550/arXiv.1907.00718, arXiv:1907.00718

Juanes, R., Hager, B.H., Herzog, H.J.: No geologic evidence that seismicity causes fault leakage that would render large-scale carbon capture and storage unsuccessful. Proc. Natl. Acad. Sci. **109**(52), E3623–E3623 (2012). https://doi.org/10.1073/pnas.1215026109

Juanes, R., Spiteri, E., Orr, F., Jr., Blunt, M.: Impact of relative permeability hysteresis on geological CO_2 storage. Water Resour. Res. **42**(12), W12418 (2006). https://doi.org/10.1029/2005WR004806

Jung, N.H., Han, W.S., Watson, Z., Graham, J.P., Kim, K.Y.: Fault-controlled CO_2 leakage from natural reservoirs in the Colorado Plateau, East-Central Utah. Earth Planet. Sci. Lett. **403**, 358–367 (2014). https://doi.org/10.1016/j.epsl.2014.07.012

Kam, P., Nadeem, M., Novlesky, A., Kumar, A., Omatsone, E.N.: Reservoir characterization and history matching of the Horn River shale: an integrated geoscience and reservoir-simulation approach. J. Can. Pet. Technol. **54**(06), 475–488 (2015). https://doi.org/10.2118/171611-PA

Killough, J.E.: Reservoir simulation with history-dependent saturation functions. SPE J. **16**(01), 37–48 (1976). https://doi.org/10.2118/5106-PA

Kneafsey, T.J., Pruess, K.: Laboratory flow experiments for visualizing carbon dioxide-induced, density-driven brine convection. Transp. Porous Media **82**, 123–139 (2010). https://doi.org/10.1007/s11242-009-9482-2

Kovscek, A.R., Nordbotten, J.M., Ferno, M.A.: Scaling up FluidFlower results for carbon dioxide storage in geological media. Transport in Porous Media (this issue). Preprint available on arXiv through https://doi.org/10.48550/arXiv.2301.09853 (2023)

Krishnamurthy, P.G., DiCarlo, D., Meckel, T.: Geologic heterogeneity controls on trapping and migration of CO_2. Geophys. Res. Lett. **49**(16), e2022GL099104 (2022). https://doi.org/10.1029/2022GL099104

Krishnamurthy, P.G., Meckel, T.A., DiCarlo, D.: Mimicking geologic depositional fabrics for multiphase flow experiments. Water Resour. Res. **55**(11), 9623–9638 (2019). https://doi.org/10.1029/2019WR025664

Krogstad, S., Lie, K., Møyner, O., Nilsen, H.M., Raynaud, X., Skaflestad, B.: MRST-AD—an open-source framework for rapid prototyping and evaluation of reservoir simulation problems. In: SPE Reservoir Simulation Conference, pp. 1–20. https://doi.org/SPE-173317-MS (2015)

Land, C.S.: Calculation of imbibition relative permeability for two-and three-phase flow from rock properties. Soc. Petrol. Eng. J. **8**(02), 149–156 (1968). https://doi.org/10.2118/1942-PA

Landa-Marbán, D., Sandve, T.H., Both, J.W., Nordbotten, J.M., Gasda, S.E.: Performance of an open-source image-based history matching framework for CO_2 storage (2023, forthcoming)

Lassen, R.N., Plampin, M.R., Sakaki, T., Illangasekare, T.H., Gudbjerg, J., Sonnenborg, T.O., Jensen, K.H.: Effects of geologic heterogeneity on migration of gaseous CO_2 using laboratory and modeling investigations. Int. J. Greenh. Gas Control **43**, 213–224 (2015). https://doi.org/10.1016/j.ijggc.2015.10.015

Lenhard, R., Oostrom, M., Simmons, C., White, M.: Investigation of density-dependent gas advection of trichloroethylene: experiment and a model validation exercise. J. Contam. Hydrol. **19**(1), 47–67 (1995). https://doi.org/10.1016/0169-7722(94)00055-M

Leverett, M.C.: Capillary behavior in porous solids. Trans. AIME **142**(01), 152–169 (1941). https://doi.org/10.2118/941152-G

Lie, K.A.: An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST). Cambridge University Press, Cambridge (2019)

Lie, K.A., Møyner, O. (eds.): Advanced Modeling with the MATLAB Reservoir Simulation Toolbox. Cambridge University Press, Cambridge (2021)

Lin, L.I-K.: A concordance correlation coefficient to evaluate reproducibility. Biometrics **45**(1), 255–268 (1989). https://doi.org/10.2307/2532051

Liu, Y., Durlofsky, L.J.: Multilevel strategies and geological parameterizations for history matching complex reservoir models. SPE J. **25**(01), 081–104 (2020). https://doi.org/10.2118/193895-PA

MacMinn, C.W., Juanes, R.: Buoyant currents arrested by convective dissolution. Geophys. Res. Lett. **40**(10), 2017–2022 (2013). https://doi.org/10.1002/grl.50473

Marcucci, A., Kypreos, S., Panos, E.: The road to achieving the long-term Paris targets: energy transition and the role of direct air capture. Clim. Change **144**(2), 181–193 (2017). https://doi.org/10.1007/s10584-017-2051-8

Myers, R.D., Allgood, A., Hjellbakk, A., Vrolijk, P., Briedis, N.: Testing fault transmissibility predictions in a structurally dominated reservoir: Ringhorne field, Norway. Geol. Soc. Lond. **292**(1), 271–294 (2007). https://doi.org/10.1144/SP292.16

Neufeld, J.A., Hesse, M.A., Riaz, A., Hallworth, M.A., Tchelepi, H.A., Huppert, H.E.: Convective dissolution of carbon dioxide in saline aquifers. Geophys. Res. Lett. **37**(22), L22404 (2010). https://doi.org/10.1029/2010GL044728

Ni, H., Bakhshian, S., Meckel, T.: Effects of grain size and small-scale bedform architecture on CO_2 saturation from buoyancy-driven flow. Sci. Rep. **13**, 2474 (2023). https://doi.org/10.1038/s41598-023-29360-y

Nordbotten, J.M., Benali, B., Both, J.W., Brattekås, B., Storvik, E., Fernø, M.: DarSIA: an open-source python toolbox for two-scale image processing for porous media. Transp. Porous Media (this issue). Preprint available on arXiv through https://doi.org/10.48550/arXiv.2301.05455 (2023)

Nordbotten, J.M., Celia, M.A.: Geological Storage of CO_2: Modeling Approaches for Large-scale Simulation, p. 242. Wiley, Hoboken (2012)

Nordbotten, J.M., Fernø, M., Flemisch, B., Juanes, R., Jørgensen, M.: Final benchmark description. FluidFlower Int. Benchmark Study (2022). https://doi.org/10.5281/zenodo.6807102

Nordbotten, J.M., Flemisch, B., Gasda, S.E., Nilsen, H., Fan, Y., Pickup, G.E., Wiese, B., Celia, M.A., Dahle, H.K., Eigestad, G.T., Pruess, K.: Uncertainties in practical simulation of CO_2 storage. Int. J. Greenh. Gas Control **9**, 234–242 (2012). https://doi.org/10.1016/j.ijggc.2012.03.007

Oldenburg, C.M.: Are we all in concordance with the meaning of the word conformance, and is our definition in conformity with standard definitions? Greenh. Gases Sci. Technol. **8**(2), 210–214 (2018). https://doi.org/10.1002/ghg.1773

Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. Comput. Geosci. **15**(1), 185–221 (2011). https://doi.org/10.1007/s10596-010-9194-2

Panaretos, V.M., Zemel, Y.: Statistical aspects of Wasserstein distances. Ann. Rev. Stat. Appl. **6**, 405–431 (2019). https://doi.org/10.1146/annurev-statistics-030718-104938

Pentland, C.H., Itsekiri, E., Al Mansoori, S.K., Iglauer, S., Bijeljic, B., Blunt, M.J.: Measurement of nonwetting-phase trapping in sandpacks. SPE J. **15**(02), 274–281 (2010). https://doi.org/10.2118/115697-PA

Plug, W.J., Bruining, J.: Capillary pressure for the sand-CO_2-water system under various pressure conditions, application to CO_2 sequestration. Adv. Water Resources **30**(11), 2339–2353 (2007). https://doi.org/10.1016/j.advwatres.2007.05.010

Rezk, M.G., Foroozesh, J., Abdulrahman, A., Gholinezhad, J.: CO_2 diffusion and dispersion in porous media: review of advances in experimental measurements and mathematical models. Energy Fuels **36**(1), 133–155 (2022). https://doi.org/10.1021/acs.energyfuels.1c03552

Riaz, A., Hesse, M., Tchelepi, H., Orr, F.: Onset of convection in a gravitationally unstable diffusive boundary layer in porous media. J. Fluid Mech. **548**, 87–111 (2006). https://doi.org/10.1017/S0022112005007494

Riaz, A., Pankiewitz, C., Meiburg, E.: Linear stability of radial displacements in porous media: influence of velocity-induced dispersion and concentration-dependent diffusion. Phys. Fluids **16**(10), 3592–3598 (2004). https://doi.org/10.1063/1.1775431

Saadatpoor, E., Bryant, S.L., Sepehrnoori, K.: New trapping mechanism in carbon sequestration. Transp. Porous Media **82**(1), 3–17 (2010). https://doi.org/10.1007/s11242-009-9446-6

Saló-Salgado, L., Davis, J.S., Juanes, R.: Fault permeability from stochastic modeling of clay smears. Geology **51**(1), 91–95 (2023). https://doi.org/10.1130/G50739.1

Saló-Salgado, L., Møyner, O., Lie, K.A., Juanes, R.: Three-dimensional simulation of geologic carbon sequestration using MRST-ad-blackoil (2023, forthcoming)

Santoso, R.K., He, X., Alsinan, M., Kwak, H., Hoteit, H.: Bayesian long-short term memory for history matching in reservoir simulations. In: SPE Reservoir Simulation Conference (SPE-203976-MS). https://doi.org/10.2118/203976-MS (2021)

Schlumberger: ECLIPSE Technical description (2014.1 ed.). Schlumberger Ltd (2014)

Schulz, R., Ray, N., Zech, S., Rupp, A., Knabner, P.: Beyond Kozeny–Carman: predicting the permeability in porous media. Transp. Porous Media **130**(2), 487–512 (2019). https://doi.org/10.1007/s11242-019-01321-y

Schulze-Makuch, D.: Longitudinal dispersivity data and implications for scaling behavior. Groundwater **43**(3), 443–456 (2005). https://doi.org/10.1111/j.1745-6584.2005.0051.x

Silliman, S., Simpson, E.: Laboratory evidence of the scale effect in dispersion of solutes in porous media. Water Resour. Res. **23**(8), 1667–1673 (1987). https://doi.org/10.1029/WR023i008p01667

Smits, K.M., Sakaki, T., Limsuwat, A., Illangasekare, T.H.: Thermal conductivity of sands under varying moisture and porosity in drainage-wetting cycles. Vadose Zone J. **9**(1), 172–180 (2010). https://doi.org/10.2136/vzj2009.0095

Spiteri, E.J., Juanes, R., Blunt, M.J., Orr, F.M.: A new model of trapping and relative permeability hysteresis for all wettability characteristics. SPE J. **13**(03), 277–288 (2008). https://doi.org/10.2118/96448-PA

Spycher, N., Pruess, K.: CO_2-H_2O mixtures in the geological sequestration of CO_2. II. Partitioning in chloride brines at 12–100 C and up to 600 bar. Geochimica et Cosmochimica Acta **69**(13), 3309–3320 (2005). https://doi.org/10.1016/j.gca.2005.01.015

Spycher, N., Pruess, K., Ennis-King, J.: CO_2-H_2O mixtures in the geological sequestration of CO_2. I. Assessment and calculation of mutual solubilities from 12 to 100 C and up to 600 bar. Geochimica et Cosmochimica Acta **67**(16), 3015–3031 (2003). https://doi.org/10.1016/S0016-7037(03)00273-4

Szulczewski, M., Hesse, M., Juanes, R.: Carbon dioxide dissolution in structural and stratigraphic traps. J. Fluid Mech. **736**, 287–315 (2013). https://doi.org/10.1017/jfm.2013.511

Timur, A.: An investigation of permeability, porosity, & residual water saturation relationships for sandstone reservoirs. Log Anal. **9**(04), 8–17 (1968)

Trevisan, L., Cihan, A., Fagerlund, F., Agartan, E., Mori, H., Birkholzer, J.T., Zhou, Q., Illangasekare, T.H.: Investigation of mechanisms of supercritical CO_2 trapping in deep saline reservoirs using surrogate fluids at ambient laboratory conditions. Int. J. Greenh. Gas Control **29**, 35–49 (2014). https://doi.org/10.1016/j.ijggc.2014.07.012

Trevisan, L., Pini, R., Cihan, A., Birkholzer, J.T., Zhou, Q., González-Nicolás, A., Illangasekare, T.H.: Imaging and quantification of spreading and trapping of carbon dioxide in saline aquifers using meter-scale laboratory experiments. Water Resour. Res. **53**(1), 485–502 (2017). https://doi.org/10.1002/2016WR019749

Verdon, J.P., Kendall, J.M., Stork, A.L., Chadwick, R.A., White, D.J., Bissell, R.C.: Comparison of geomechanical deformation induced by megatonne-scale CO_2 storage at Sleipner, Weyburn, and In Salah. Proc. Natl. Acad. Sci. **110**(30), E2762–E2771 (2013). https://doi.org/10.1073/pnas.1302156110

Vilarrasa, V., Carrera, J.: Geologic carbon storage is unlikely to trigger large earthquakes and reactivate faults through which CO_2 could leak. Proc. Natl. Acad. Sci. **112**(19), 5938–5943 (2015). https://doi.org/10.1073/pnas.1413284112

Wang, S., Lee, M., Park, M.K., Kim, J.M.: Box experiments on monitoring the CO_2 migration in a homogeneous medium using electrical resistivity survey. Geosci. J. **14**, 77–85 (2010). https://doi.org/10.1007/s12303-010-0009-1

Weir, G.J., White, S.P., Kissling, W.M.: Reservoir storage and containment of greenhouse gases. Transp. Porous Media **23**, 37–60 (1996). https://doi.org/10.1007/BF00145265

Wood, B.D., Dawson, C.N., Szecsody, J.E., Streile, G.P.: Modeling contaminant transport and biodegradation in a layered porous media system. Water Resour. Res. **30**(6), 1833–1845 (1994). https://doi.org/10.1029/94WR00434

Zoback, M.D., Gorelick, S.M.: Earthquake triggering and large-scale geologic storage of carbon dioxide. Proc. Natl. Acad. Sci. **109**(26), 10164–10168 (2012). https://doi.org/10.1073/pnas.1202473109

Authors and Affiliations

Lluís Saló-Salgado[1,2] · Malin Haugen[3] · Kristoffer Eikehaug[3] · Martin Fernø[3,5] · Jan M. Nordbotten[4,5] · Ruben Juanes[1,2]

✉ Martin Fernø
  Martin.Ferno@uib.no

✉ Jan M. Nordbotten
  Jan.Nordbotten@uib.no

✉ Ruben Juanes
  juanes@mit.edu

  Lluís Saló-Salgado
  lsalo@mit.edu

  Malin Haugen
  Malin.Haugen@uib.no

  Kristoffer Eikehaug
  Kristoffer.Eikehaug@uib.no

[1] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge 02139, MA, USA

[2] Earth Resources Laboratory, Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge 02139, MA, USA

[3] Department of Physics and Technology, University of Bergen, 5020 Bergen, Norway

[4] Department of Mathematics, Center for Modeling of Coupled Subsurface Dynamics, University of Bergen, 5020 Bergen, Norway

[5] Norwegian Research Center, Postboks 22 Nygårdstangen, 5020 Bergen, Norway