




A History Matching Study for the FluidFlower Benchmark Project

Xiaoming Tian^{1,2} · Michiel Wapperom¹ · James Gunning³ · Samuel Jackson³ · Andy Wilkins⁴ · Chris Green³ · Jonathan Ennis-King³ · Denis Voskov^{1,5} 

Received: 9 January 2023 / Accepted: 11 December 2023 / Published online: 16 January 2024
© The Author(s) 2024, corrected publication 2024

Abstract

In this study, we conduct a comprehensive history matching study for the FluidFlower benchmark model. This benchmark was prepared and organized by the University of Bergen, the University of Stuttgart, and Massachusetts Institute of Technology, for promoting understanding of the complex physics of geological carbon storage (GCS) through in-house experiments and numerical simulations. This paper synthesizes the experiences of history matching the benchmark data encountered by the Delft-DARTS and CSIRO participants. History matching is first performed based on a low-dimensional-zonated structured model using a simple Poisson-like solver. The permeability of six facies and two faults is inferred in this stage to match the digitized concentration data. The history matching is then further enhanced to richer spatial and physical models to capture the spatial variation of permeability and buoyancy effects, using an unstructured grid. Efficient adjoint methods are used to evaluate the gradient used in the optimization of data misfits or equivalent Bayesian log-likelihoods. With efficient optimization methods available for both maximum a posteriori model inference and Randomized Maximum Likelihood methods for model uncertainty, we perform history matching using both binary and continuous concentration observations. The results show that the tracer plumes in the enriched model match the experimental plumes more accurately compared with the results from the parsimonious-zonated model. The history matching results based on the concentration observations provide more similar plume shapes compared with the case based on the binary observations. The permeability difference between the model before and after history matching reveals that the tracer plume zone and the high permeable zone are the regions of high sensitivity in terms of data misfit between the model response and observations. Surprisingly, CO₂ concentration plume forecasts based on these history-matched models were not especially sensitive to the improvements observed in the enhanced model.

Keywords FluidFlower · CCS · History matching · DARTS · Adjoint gradients

✉ Denis Voskov
D.V.Voskov@tudelft.nl

¹ Department of Geoscience and Engineering, TU Delft, Delft, The Netherlands

² Guangzhou Institute of Energy Conversion, Chinese Academy of Sciences, Guangzhou, China

³ CSIRO Energy, Bayview Avenue, Clayton, VIC, Australia

⁴ CSIRO Mineral Resources, Pullenvale, QLD, Australia

⁵ Department of Energy Science and Engineering, Stanford University, CA, USA

1 Introduction

Escalating greenhouse gas concentrations in the atmosphere and their effect on climate have become an urgent and global concern. Extensive CO₂ emissions are a pervasive aspect of modern industrialised society and its dependence on fossil fuels. Large reductions in CO₂ emissions are mandated by the most recent IPCC reports, and virtually all the conceivable pathways require engineered removal of new carbon pollution sources (IPCC 2022; IEA 2020a). Carbon capture and storage (CSS) has been proven to be one of the most promising solutions to this environmental issue. Typically, CCS can reduce 85–90% CO₂ emissions from large point emission sources (Leung et al. 2014), for example, power plants, cement kiln plants, etc. It is also very likely that carbon-removal technologies—direct air capture or bioenergy with CCS—will be required because emissions of CO₂ cannot otherwise be reduced in time to avoid dangerous global warming (IEA 2020b). The IEA models require a rapid scale-up of CCS, from 40 Mt p.a. now to a Gt p.a. by 2030 (IEA 2020c).

There are several ways to store CO₂ into the geological formations (Chiaramonte et al. 2011; White et al. 2003; Lyu et al. 2021b; House et al. 2006; Schaefer et al. 2010). Of these possibilities, CO₂ storage in saline aquifers is considered to have the largest potential for the storage of CO₂. This is because saline aquifers can be widely found in both onshore and offshore areas. The usual four CO₂ trapping mechanisms, with varying characteristic time scales, are expected to apply: structural, residual, solubility, and mineral trapping, in increasing order of time scale.

To understand and quantify the flow dynamics of CO₂ storage process, numerical simulation is essential for modelling the process based on the governing equations and equations of state. However, the numerical simulation of CO₂ storage is challenging because of the complexity of the fluid thermodynamics and the unfolding of multiscale physics.

To assist in these challenges, the FluidFlower benchmark project (Nordbotten et al. 2022) has been constructed to accelerate progress in the field of CO₂ storage modelling in geological settings. This project has closely inspired the 11th Society of Petroleum Engineers Comparative Solution Project (SPE 2023). The central component of the benchmark is an experimental rig emulating a typical geological scenario for CO₂ injection, constructed by the University of Bergen. Several international research groups were then invited to perform numerical simulation of this benchmark, where the dominant processes are associated with multiphase flows, capillarity, dissolution, and convective mixing. The benchmark experiments furnish tracer images and pressure observations to serve as calibration data for the spatial model of petrophysical properties. Details of these experiments are described at length in Nordbotten et al. (2022). The efforts in history-matching following in this paper were motivated by the natural and reasonable belief that this would lead to the most accurate forecasts of the CO₂ injection. Hence, this study focuses on the history matching of the tracer and pressure tests of this benchmark, but we consider also the issue of forecasting uncertainty under CO₂ injection.

Inversion of geophysical data very often reduces to an optimization problem where some misfit measure between the observations and model predictions is minimized. Depending on whether model gradients are available, optimization methods are then classified into gradient-free and gradient-based methods. For problems with challenging nonconvexity or multimodality in the misfit function, gradient methods may be of

limited use, as they are vulnerable to trapping in inferior local minima. In such cases gradient-free global methods like simulated annealing (Kirkpatrick et al. 1983), genetic algorithms (Holland 1984), or particle swarm algorithm (Eberhart and Kennedy 1995) are popular. Their efficiency is always vastly inferior to gradient methods, if applied to a problem that has a single or dominating global optimum. For such problems, where gradients can be obtained, the most efficient methods are usually variants of the Gauss–Newton method if the second-order (Hessian) information can be obtained or well approximated, or varieties of conjugate-gradient or sparse quasi-Newton methods (e.g. l-BFGS) where only the gradient is available, as is typical in high dimensional model settings (Nocedal and Wright 2006). Even in non-convex settings, gradient-based methods are very often the techniques of choice, if combined with graduation or continuation strategies for smoothing out the nonconvexity in the early phases of the optimisation. This paper uses exclusively gradient based techniques, which approach Newton-like (quadratic) efficiency when the model dimensionality is very low and the Hessian is well approximated via use of the explicit Jacobian, as per the Marquardt method, but reduce to linear-only efficiency when only the adjoint-based gradient is used in the high-dimension cases.

When the model dimensionality is high and the forward physics is generated by a set of PDEs that are differentiable in the model parameters, which is typical of geo-energy systems, the most efficient way to compute gradients is the adjoint method. It allows for the efficient calculation of the gradient by reusing the solution of the governing equations, since the solution of the adjoint problem is usually very closely related to the forward problem, often by a time-reversal procedure. In the field of geo-energy reservoir engineering, adjoint methods have been applied to perform history matching (Chen et al. 1974; Oliver et al. 2008), petroleum recovery processes (Mehos and Ramirez 1989; Fathi and Ramirez 1984; Ramirez et al. 1984), thermal recovery processes (Wei et al. 1993) and so on. Later, with the advent of “smart well” and “smart field” concepts, adjoint-based optimization and optimal control ideas have been widely studied and applied in the field of reservoir engineering and reservoir management (Brouwer and Jansen 2004; Sarma et al. 2005, 2006; Volkov and Voskov 2016).

In this study, we will utilize the open-source Delft Advanced Research Terra Simulator (open-DARTS) framework (Voskov et al. 2023) designed for modelling energy transition applications (Khait and Voskov 2018b; Wang et al. 2020; Lyu and Voskov 2023). This implementation features an efficient forward reservoir simulation but also inverse modelling capability based on the adjoint gradient method. The adjoint implementation for multi-component, multi-phase system in petroleum-related reservoir simulation problems are described in Tian et al. (2021), and both the forward and inverse modelling capabilities have been extended into generic energy transition applications (Tian and Voskov 2023). This extension allows for the incorporation of various effects, such as energy transfer, gravity, diffusion, and chemical reaction, through corresponding operators.

This paper presents a two-stage approach for history matching tracer test image data, using first a simple Poisson-like solver in a very parsimonious zoned model, then escalating to an adjoint-based framework for a much richer model with ≈ 72000 transmissibility parameters. The simple zoned model uses structured grids, but we move to unstructured grids for the high dimensional model. The adjoint method uses operator-based linearization (OBL) for the development of the misfit gradient, and its efficiency makes possible the modelling of heterogeneity and computation of uncertainties in the history matching process. To estimate the uncertainty of the model inference, we use the standard Randomized

Maximum Likelihood (RML) method (Oliver et al. 2008), and a spatial model of heterogeneity in lognormal permeability is added to the geological model.

The structure of this paper is as follows. In Sect. 2, we give a short description of the FluidFlower benchmark setup. In Sect. 3, history matching is performed using a simple zoned model using buoyancy-free Poisson-like physics. To further improve the history matching results and account for permeability uncertainty and buoyancy effects, enhanced history matching using a two-stage approach with richer physics is described in Sect. 4. Section 5 discusses the results of this enhanced method, including uncertainty estimates and forecasting behaviour. Discussion and conclusions then follow in the usual way.

2 FluidFlower Benchmark Description

The FluidFlower experimental rig (Nordbotten et al. 2022) comprises an engineered heterogeneous sand pack assembled within a thin (25 mm) vertical “filled” Hele-Shaw cell, about 2.8 m wide by 1.3 m high. The layered heterogeneous structure is filled using sands filtered into different grain sizes: the sand is placed in a layered fashion between the front glass panel and a sealed back panel with perforations for ports that can be used as injectors, producers, or pressure gauge locations. A sketch of the experimental rig with colored layers is shown in Fig. 1.

Experimentally, the sand facies have been sieved into groups labelled ESF, C, D, E, F, G, and there are 3 “fault” regions manufactured in the model.

Permeability and porosity measurements for the different types of sand have been performed in independent sand-pack experiments (Nordbotten et al. 2022). We expect that the in-situ values of sand layer parameters may vary due to non-uniform sand distribution and boundary effects. The tracer tests have generated tracer data in the form of time-stamped camera images. These images are digitized and mapped on to grid representations

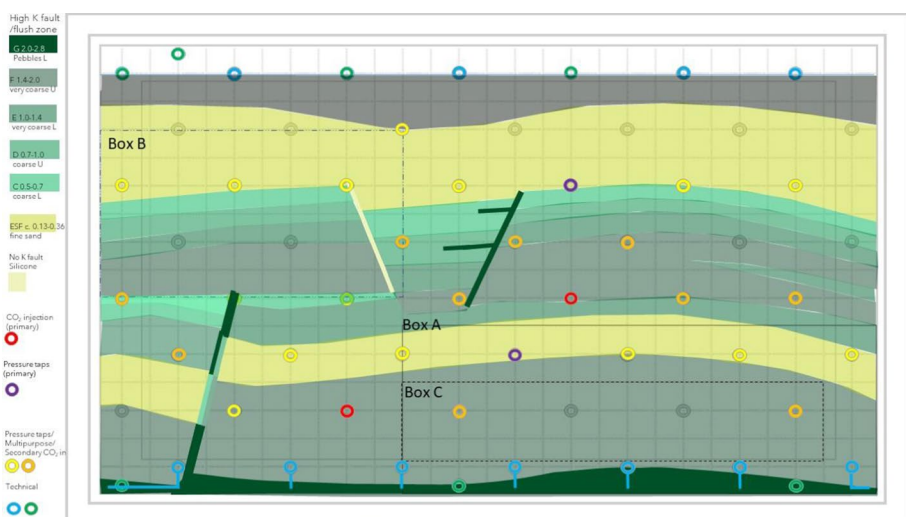


Fig. 1 The sketch of the experimental rig geometry following FluidFlower benchmark description

as needed. Pressure data are measured at gauges attached to ports in the rig, but the gauges are some distance from the ports, on narrow feed plumbing.

Since the principle objective of the FluidFlower comparison project is to predict the flow of CO₂ in a controlled study, and because this flow will depend strongly on the flow properties of the media to leading order, it is sensible to use all experimental information to estimate permeability and porosity. Apart from the sand-pack experiments, the richest source of such information is the history matching of the tracer experiments. Section 3 describes a simple and fast way to invert the tracer data for leading-order estimates, and this is augmented to richer models and full-physics simulation in Sect. 4.

3 History Matching Based on Zonated Model and Simple Poisson-Like Solver

In this section, we describe a fast, parsimonious, but approximate baseline history match using a simple single-phase flow model. The closely similar densities of the tracers, together with the extremely rapid pressure communication in the system, are exploited to write down a fast approximate forward model which can be exploited for inversion.

3.1 Governing Forward Model

For the tracer tests, flow is presumed to be single phase, isotropic, the density-independent of tracer, and governed by the averaged thickness (2D model) diffusion equation

$$c_t \phi h \frac{\partial p}{\partial t} - \nabla \cdot (k_w / \mu_w) h(\mathbf{r}) \nabla p = Q(\mathbf{r}, t) \quad (1)$$

where p is the pressure, h the cell thickness, ϕ the porosity, k_w the endpoint permeability of the media for water, μ_w the water viscosity, c_t the total (rock plus water) compressibility, and Q the volume rate of water injection at the ports as a function of time. The porosity ϕ and thickness h are taken as “known”—from sample measurement and thickness measurement bilinear interpolations. Given the experimental care in the sieving process and sand-pack porosity measurements, and the relative insensitivity of unconsolidated porosity values, it was deemed unnecessary to model porosity variation with extra parameters. The sample measurements for (endpoint) permeability, for the labelled facies, are notated $k_{w,l}$ in the below, for facies l .

The digitized and rasterized model of the facies is shown in Fig. 2, with labels 1, ... 9 for the facies {1=ESF, 2=C, 3=D, 4=E, 5=F, 6=G, 7=Fault1, 8=Fault2, 9=Fault3}. Fault 2 (facies 8) is designed to be impermeable, and no inferences for its properties are performed in the below. The digitization process uses a linear mapping between intensity and tracer concentration, accounting for labelled porosity and local thickness, and calibrated using known injection volumes and mass conservation. The computed values map closely to the interval [0, 1], but are clipped before use in inversion.

From the digitized and interpreted images of facies labels $F_i \in \mathcal{L} = \{1, 2, \dots, 9\}$, in voxel i , for facies {ESF, C, D, E, F, G, Fault1, Fault3}, the absolute permeabilities are modelled by $k_{w,i} = u_{F_i} k_{F_i}$ in gridblock i , where the model is $\mathbf{u} = \{u_l\}$, $l \in \mathcal{L}$. The supplied “core” sample absolute permeabilities k_l are taken as per Table 1.

The model parameters u_l , $l \in \mathcal{L}$ are expected to be $O(1)$ correction factors, and the rescaling is designed to make the inversion as well-scaled as practically reasonable. The

Table 1 Endpoint permeabilities and fixed porosities for modelling

Facies	Index l	Permeability [D] (k_l)	Porosity ϕ_l
ESF	1	39	0.44
C	2	293	0.43
D	3	424	0.44
E	4	708	0.45
F	5	258	0.43
G	6	488	0.46
Fault1	7	488	0.46
Fault2	8	0	0
Fault3	9	488	0.46

The experimental values are taken from sand-pack experiments documented in Nordbotten et al. (2022)

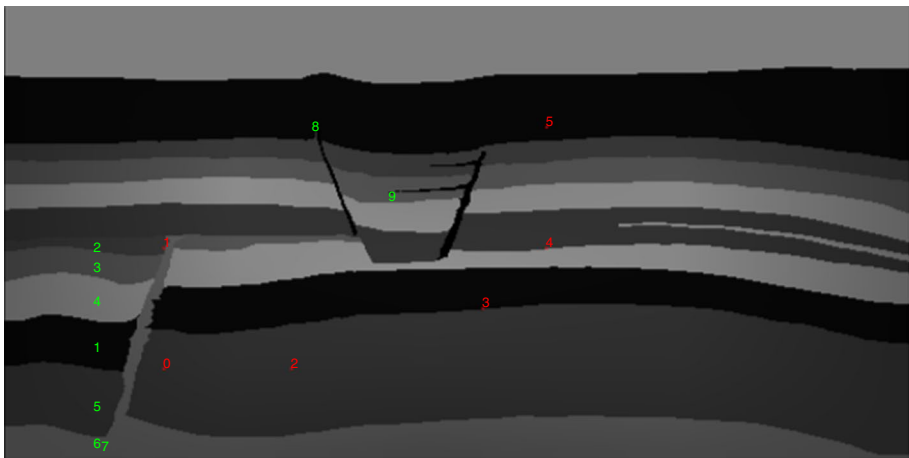


Fig. 2 Digitised and interpreted facies-label (green codes) image used as the basis for the parameterization. The labelled ports (red) 0, 1, ... 5 are used in the inversion, corresponding to ports {5_3(0),5_7(1),9_3(2),15_5(3),17_7(4),17_11(5)}

characteristic time scale for diffusion $t = \frac{\phi \mu c_l L^2}{k_{sc}}$, for the experiment box size L , is very short, typically $t \approx 5$ ms using scaling $\phi = .45$, $\mu_w = 10^{-3}$ Pa.s, $c_l \approx 0.7310^{-9}$ Pa $^{-1}$, $L = 1$ m, $k_{sc} = 6.9874 \times 10^{-10}$ m 2 . All pressures are rescaled to dimensionless pressure using a scaling $P_{sc} = Q_{sc} \mu_w / (k_{sc} h_0)$, with additional chosen constants $h_0 = 0.025$ m (typical cell width), $Q_{sc} = 6.25 \times 10^{-7}$ m 3 /s (the water injection rate) yielding scaling pressure $P_{sc} = 35.778$ Pa.

For this reason, the forward modelling has approximated the response as steady state on the time scale of the experiment, with the pressure corresponding to the steady-state solution of Eq. (1) at each instant of time. During the experiment, the rates Q are sustained as constant over 30-min interval chunks, so the overwhelming majority of the data are collected with the system equilibrated in terms of pressure. In the experimental data, we see the measured pressures respond virtually instantly to the applied Q when it changes.

The implication of the assumption of constant density is that the fluid configuration remains fixed as long as there is no water injection. The triple-injection tracer experiment was split into 3 successive injections/tranches, each of three 30 min constant-rate injections with rest intervals:

(Port17_7)	5.19 – 5.49 pm(clear)	6.20 – 6.50 pm(clear)	7.20 – 7.50 pm(clear)	(Day1)
(Port9_3)	8.52 – 9.22 pm(clear)	9.52 – 10.22 pm(clear)	10.52 – 11.22 pm(clear)	(Day1)
(Port9_3)	2.48 – 3.18 pm(blue)	3.48 – 4.18 pm(blue)	4.48 – 5.18 pm(blue)	(Day2)

The upshot of this is that for the first 6 h of the tracer experiment, about 3 h are injecting, and 3 h waiting. After Day 1, there is a ≈ 16 h wait till the third tranche of tests on Day 2, during which buoyancy rise in the tracer becomes evident in the comparison of the images at the end of tranche 2 and the start of tranche 3. Since the gravity effects are slow, we expect this effect in the first two tranches to be relatively weak, but the effect is clear in the 3rd tranche images, especially around the 9_3 port injection. The Poisson-like model neglects this buoyancy effect, but for small buoyant drift we expect the residual error at the tracer front to end up symmetrically distributed around the prediction of the constant-density model, so we believe the approximation is reasonable, especially for profiles formed around newly-injected ports. Roughly speaking, we expect the approximation to increase the variance of the predictive model to the leading order, but not the bias. Similarly, dispersion is treated as a minor effect, and probably confounded with 3D effects and heterogeneity. During injection, the Peclet number is estimated from the tracer velocity and grain size to be $Pe \approx 50$, so advection should be more dominant, but dispersion will smear the front and the sharp modelled tracer front from the Poisson model should fit the mildly dispersed experimental front at a middle value corresponding to an average velocity.

Thus, on the basis that time-stepping the diffusion equation was unnecessary, we solve the steady-state Poisson equation with Q at two different places (ports 9_3, 17_7), corresponding to the experiment performed. The resulting fixed pressure fields and velocities are then steady over the period where the injection rate is held steady, and are sufficient to compute the advection of the tracers over this interval and the observed pressure fields. The model was based on a 5 mm grid ($N_x = 568$, $N_y = 300$), with upper boundary condition set as fixed (atmospheric) pressure in the water column, and no-flow Neumann conditions on the left, right, and lower edges of the model. A typical pressure solution for injection at port 17_7 is shown in Fig. 3.

To model the tracer movement, tracers were advected along streamlines using the fixed velocities computed from the Poisson solves, continuing for the correct time corresponding to each tracer color. The tracer advection was implemented using the upwinding scheme described in Vreugdenhil and Koren (1993). The time-stepping in the tracer computation was adjusted so the number of time steps divided the total tracer time exactly, which has the merit that any computed quantities from the tracer image are a smooth differentiable function of the parameters in the PDE.

3.2 History Matching

For inversion, the data available consisted of the injection port and monitor port pressures during the injections, and colour tracer images at the end of the injection periods. The cross-port pressures were conspicuously noisy and clearly close to the noise floor of the instruments. Significant drift was evident in these measurements, and even the stable values showed the curious property of the amplitude not diminishing in a

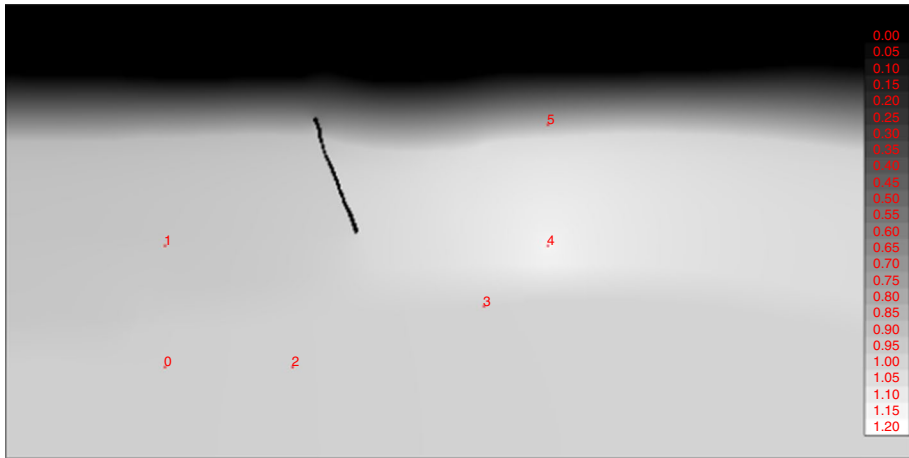


Fig. 3 Typical (dimensionless) pressure solution of steady-state Poisson problem for injection at 17_7 port (indexed 4)

consistent way with the distance from the injection port. This has implications in the inversion if this data is weighted very heavily. The injection pressures were strong signals but measured some way away (20 cm or more) from the actual injection face and subject to unknown frictional and other losses in the feed plumbing. This makes them not very useful for inversion. By contrast, the tracer images were very clearly interpretable, rich in spatial content, and not obviously contaminated by an experimental artifact of any significant kind.

The cell is initially filled with a blue tracer. Injections were modelled at $Q = 2250$ ml/h (6.25×10^{-7} m³/s) for 3×30 min at port 17_7 (with clear water), then 3×30 min at port 9_3 (more clear water), then 3×30 min at 9_3 with blue tracer. The forward modelling operation, which implements standard tracer advection under an upwinding scheme, is expressed below as a function $f_i(\mathbf{u})$ which generates concentration profiles, which are very close to unity inside the swept region, and fall rapidly to zero at the tracer front. The tracer advection is stepped forward for precisely the number of time steps needed for the injection, and numerical integrations of the total tracer mass over the modelling grid at the end of the simulation agree very closely with the mass known to be injected from Q in the tracer source. Under the assumptions of the single-phase PDE and the fast equilibration time, the experimental 30 min wait time between injections does not need to be modelled, as nothing happens in the Poisson model if the sources are switched off since the velocities are then instantly zero and no advection occurs. The modelled tracer positions at the end of each 30 min injection period are compared to digital image experimental data for inversion.

The inversion of this data was couched as a Bayesian inverse problem with a likelihood $P(\mathbf{d}_{\text{obs}} | \mathbf{u})$ formed as a joint probability using pressure and tracer data $\mathbf{d}_{\text{obs}} = \{\mathbf{P}_{\text{obs}}, \mathbf{c}_{\text{tracer}}\}$. The model was taken to be multiplier modifiers of the permeability parameters, per facies, and applied in a “paint-by-numbers” fashion over the labelled facies model. The thickness and porosity data were considered to be sufficiently precise and experimentally stable to be fixed for the purposes of model prediction and inversion. The Bayesian framework was completed with the provision of a weak prior for the modifier parameters, of Gaussian form $P(u_i) \sim N(1, \sigma^2)$ with $\sigma = 5$ for each parameter.

The associated prior covariance is $C_p = \text{diag} \{ \sigma^2 \}$. The model point estimate at the global maximum of the posterior probability is referred to as the MAP (maximum a posteriori) inversion.

The inversion is performed using a Levenberg–Marquardt routine (Nocedal and Wright 2006; Madsen et al. 2004), which requires the Jacobian $J = \{ \partial \mathbf{f} / \partial \mathbf{u} \}$ of the forward response with respect to the unknown model parameters. Since the model dimensionality was very low and the forward model speed very fast (measured in seconds), this was computed using simple forward differences.

The negative log posterior $E(\mathbf{u}) \sim -\log(P(\mathbf{d}_{\text{obs}} | \mathbf{u})P(\mathbf{u}))$ used in the optimization step was written as a standard l_2 misfit energy

$$E(\mathbf{u}) = E_{\text{pressure}}(\mathbf{u}) + E_{\text{tracer}}(\mathbf{u}) + E_p(\mathbf{u})$$

where the cross-port pressure misfit, accumulated over only stable average measurements at ports p5.3_1, p5.7_1, p9.3_1, p15.5_1, p17.7_1, p17.11_1 is

$$E_{\text{pressure}}(\mathbf{u}) = \frac{1}{2} \lambda_p \| \mathbf{P}_{\text{obs}} - \mathbf{f}_p(\mathbf{u}) \|_2^2$$

and the trace image mismatch is written as

$$E_{\text{tracer}}(\mathbf{u}) = \frac{1}{2} \lambda_t \| \mathbf{c}_{\text{tracer}} - \mathbf{f}_t(\mathbf{u}) \|_2^2$$

The weights λ_p, λ_t are adjusted so the tracer data is dominant in the likelihood as this data is much more abundant and artifact-free. The prior Bayesian term amounts to

$$E_p(\mathbf{u}) = -\log(P(\mathbf{u})) \sim \frac{1}{2} \sum_{i \in \mathcal{L}} (u_i - 1)^2 / \sigma_p^2$$

and has a very benign influence on the inversion, except that the likelihood term from the tracer image is expected to be sensitive to permeability ratios only, i.e. has a null space associated with a global scalar multiplier. If no pressure data are used, the weak prior will have the effect of producing a MAP point as the nearest point in the likelihood null space to the prior mean point $\mathbf{u} = \mathbf{1}$, i.e. the core measurements. In dimensionless units, the pressure data \mathbf{P}_{obs} are $O(1)$ numbers, but rather noisy, so setting $\lambda_p = 1$ seems appropriate. The tracer data \mathbf{c} are processed from the digital images to have concentration values ranging over $0 < y < 1$. Since the associated l_2 norm has a very large number of voxels, λ_t is scaled such that the tracer misfit energy is $E_{\text{tracer}}(\mathbf{u}) = 1000/2$ for a model that produces no tracer concentration ($\mathbf{f}_t(\mathbf{u}) = \mathbf{0}$), i.e. the information content is equivalent to 1000 measurements. In practice since the volume in which experimental and forward-modelled concentrations differ is only a small fraction of the image, the misfit energy from this term ends up being $O(10)$, perhaps equivalent to putting 10-fold the emphasis on the tracer images as the crosswell pressure data.

One possible approach was to assert that the inconsistencies in the crosswell pressure data were too problematic to warrant their inclusion, and that the inversion should be performed on the basis of the 3-injection tracer data alone. It was also considered reasonable to merge the parameters for regions 5 and 6, since region 6 is at the edge of the modelling region and will have a more fragile permeability inference.

The corresponding parameter inferences are as per Table 2 (Model A). The table shows the dimensionless scaled model inference and corresponding actual unitized values. The final column is the dimensionless uncertainty estimate $\sigma_i^2 = H_{ii}^{-1}$ for

Table 2 Inversion results from 3-tracer inversion

Facies	Index l	Model A			Model B		
		u_l	Permeability[D] ($u_l k_l$)	σ_l	u_l	Permeability[D] ($u_l k_l$)	σ_l
ESF	1	0.67	26.2	0.235	0.89	34.6	0.24
C	2	0.60	176	0.65	1.08	317	0.65
D	3	0.73	309	0.78	0.78	332	0.78
E	4	0.96	678	1.0	1.2	840	1.0
F	5	4.18	1080	0.61	5.7	1477	0.61
G	6	4.18	1080	0.61	2.43	1186	0.84
Fault1	7	2.74	1340	0.84	4.27	2082	0.84
Fault2	8	–	0	–	–	0	–
Fault3	9	3.23	1575	0.84	2.41	1176	0.84

Model A: regions 5 and 6 merged to a common parameter, no crosswell data. Model B: model with facies 5,6 independent, including crosswell data

each parameter formed from the inverse Hessian matrix at the final optimum, where $H = J^T J + C_p^{-1}$. The final inversion forward model images and associated data snapshots are depicted in Fig. 4.

A more optimistic approach was to include the crosswell data, but down-weighted in the sense previously described, and allow independence in zone 5 and 6 scaled permeabilities. The result of this inversion is shown in Table 2 (Model B). One sees that the parameters differ

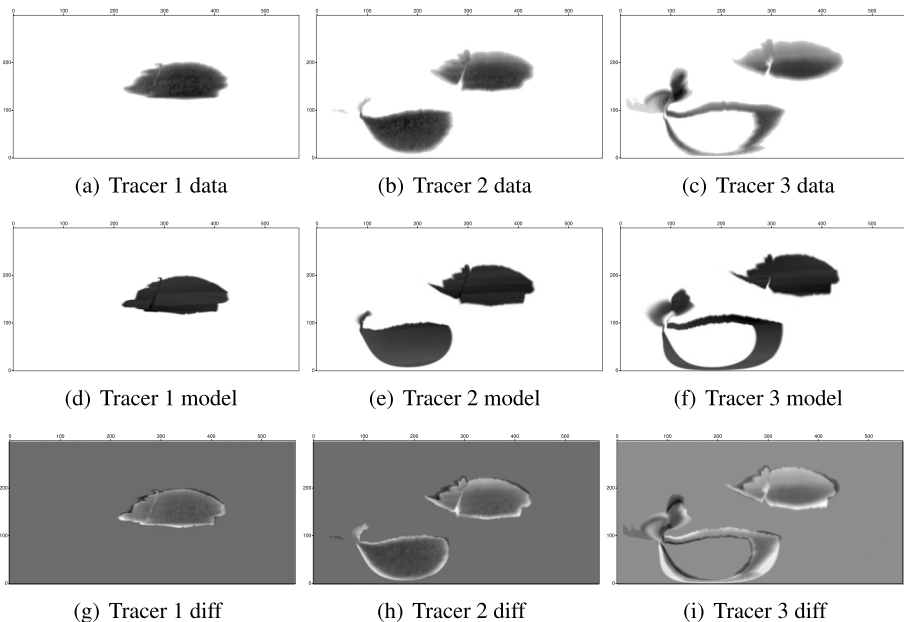


Fig. 4 Tracer data image and associated MAP forward models. Greyscale is [0,1]=[white,black]

from the previous inversion but within the estimated standard deviation associated with the estimated “statistical power” of the data embedded in the likelihood weightings. Since permeability is usually a sensitive parameter, the differences between these inverted values and the core estimates could be due to many factors, including significant experimental variability in the core experiments, modelling error, inversion sensitivity, or other reasons.

4 History Matching of FluidFlower Benchmark Using Two-Stage Approach

The preceding model fits show clear systematic structure in the tracer misfit residuals due to the simplistic physics, and neglect of zone-internal heterogeneity. This can only be addressed by escalating the forward physics to a model that includes buoyancy, but also with adjoint capabilities to manage the increased dimensionality that will inevitably follow upon an increased spatial resolution in the model. The capabilities of the open-DARTS code meet these needs very well, but it is helpful to rehearse the basic conservation and flow laws that are implemented in this code.

4.1 Governing Equations for Forward Simulation

The mass conservation equations describe a flow dynamic system bounded in the domain with volume Ω and surface Γ . The conservation equation can be written as:

$$\frac{\partial}{\partial t} \int_{\Omega} M^c d\Omega + \int_{\Gamma} \mathbf{F}^c \cdot \mathbf{n} d\Gamma = \int_{\Omega} Q^c d\Omega, \tag{2}$$

where M^c is the accumulation term for the c^{th} component ($c = 1, \dots, n_c$, index of the mass components [e.g., water, CO₂]), \mathbf{F}^c is the flux term of the c^{th} component, \mathbf{n} is the unit normal direction pointing outward to the domain boundary, and Q^c is the source/sink term of the c^{th} component.

For the accumulation term M^c for a given component c , it can be written as:

$$M^c = \phi \sum_{j=1}^{n_p} x_{cj} \rho_j s_j, \quad c = 1, \dots, n_c, \tag{3}$$

where ϕ is porosity, s_j is the saturation of phase j , ρ_j is the density [kmol/m³] of phase j , and x_{cj} is the molar fraction of component c in the j th phase.

The mass flux \mathbf{F}^c for a given component c is written as:

$$\mathbf{F}^c = \sum_{j=1}^{n_p} x_{cj} \rho_j \mathbf{v}_j, \quad c = 1, \dots, n_c. \tag{4}$$

Here \mathbf{v} is the velocity and follows Darcy’s law including gravity effects for the given phase j :

$$\mathbf{v}_j = -\mathbf{K} \frac{k_{rj}}{\mu_j} (\nabla p_j - \gamma_j \nabla z), \tag{5}$$

where \mathbf{K} is the permeability tensor [mD], k_{rj} is the relative permeability of phase j , μ_j is the viscosity of phase j [mPa · s], p_j is the pressure of phase j [bar], $\gamma_j = \rho_j \bar{g}$ is the specific weight [N/m³] (where \bar{g} is the gravity acceleration), z is the depth vector [m]. The source/sink term Q^c mainly includes terms related to chemical reactions, which are absent in this study.

Based on the finite volume method using two-point flux approximation, the discretized form of Eq. (2) for the i th reservoir gridblock can be written as:

$$g_i^c = V_i \left(M_i^c(\omega_i) - M_i^c(\omega_{i(k-1)}) \right) - \Delta t \left(\sum_l a_l F_l^c(\omega) + V_i Q_i^c(\omega) \right) = 0, \quad c = 1, \dots, n_c. \quad (6)$$

Equation (6) shows the residual form of discretized Eq. (2). We introduce V_i is the volume of the i th gridblock and ω_i is the state variables at the current time step. In addition, $\omega_{i(k-1)}$ is the state variables at the previous time step and Δt is the time step. The term a_l is the contact area of the interface l between neighbouring elements.

The resulting highly nonlinear system of equations is linearized using the Operator-Based Linearization approach (Voskov 2017; Khait and Voskov 2018a; Lyu et al. 2021a), and a modified Newton–Raphson method is employed at each nonlinear iteration. For modelling of tracer experiments, we used the same CO₂-brine physics used in the benchmark with the concentration of CO₂ in the injected stream below the solubility limit and a linear dependency of density on CO₂ concentration.

4.2 Initial Model Tuning

Based on interpreted high-resolution images of the rig, we built an unstructured mesh grid model, shown in Fig. 5.

Following the experimental description of unconsolidated layering in the rig, the mesh grid model is divided into multiple zones/layers filled with nine different types of sand in total. Different layers are assigned to different permeability and porosity values. The permeability anisotropy is modelled using a vertical/horizontal anisotropy ratio, per zone. This anisotropy ratio models the laminations clearly visible in high-resolution images, mostly related to sorting of fine-grain sands. Within the same layer, the same type of sand is maintained, and the petrophysical values of the grid cells are kept constant.

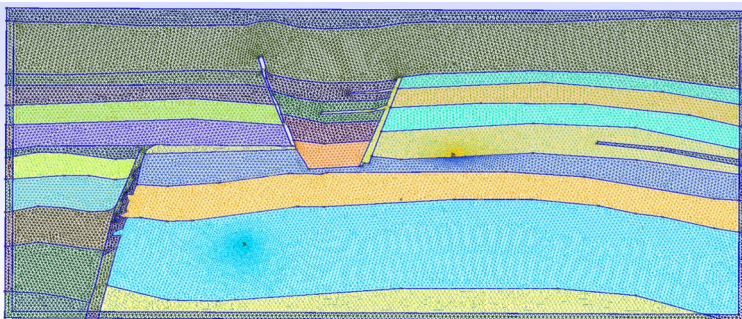


Fig. 5 The unstructured mesh grid of FluidFlower model

The petrophysical properties of the layers in Fig. 1 including permeability are initially estimated from the sand-pack experiments. These measurements will differ from the in-situ properties due to experimental variations in sand deposition/assembly, loading, and boundary effects. For the fitting procedure, we keep the porosity unchanged, while the permeability and the anisotropy of the layers are taken as the free parameters. A significant decision was taken to also introduce the density of the clear water as an additional parameter. The optimization was again based on Gauss–Newton methods in low dimensions from finite differences of efficient forward simulations. Specifically, the objective function is defined as:

$$E(\mathbf{u}) = \|\mathbf{G}(\mathbf{u}) - \mathbf{d}_{\text{obs}}\|_2^2 \quad (7)$$

where E is the objective function, \mathbf{u} are the model variables, G is the model response, and \mathbf{d}_{obs} is the observation data.

Instead of using the original tracer concentration of the experimental images as the observations, we process the experimental tracer images into several binary images, corresponding to the end of each 3×30 minute tracer injection group; see Fig. 6. The first row of this figure demonstrates the experimental images of the tracer plumes at the 3rd, 6th, and 9th “macro” time steps. The second row shows the associated digitized binary map of each time step. The value of the red colour is set as 1, while the blue colour is 0. A threshold of $\epsilon_t = 7 \times 10^{-5}$ tracer concentration (mole fraction and unitless) is used to binarise the response $G(\mathbf{u})$ to either 0 or 1. Although the objective function defined by the binarised model response and observations is mathematically non-differentiable, empirically, the numerical implementation is robust to this discontinuity. These binary maps delineate the boundaries of the tracer plumes and will be utilized as the observations in the initial stage of model tuning. This simple image processing technique will be later compared with the more accurate image recognition shown in Fig. 4.

The initial model parameters are obtained from Table 2 (Model A, except anisotropy). They are the permeability of layers C, D, E, F, Fault1, Fault3, G, and the anisotropy factor of the layer ESF in the z -direction. These eight parameters are updated in the history matching iterations. The optimal values of these eight parameters are shown in Table 3. Note that Table 3 also includes placeholder information for the water layer W.

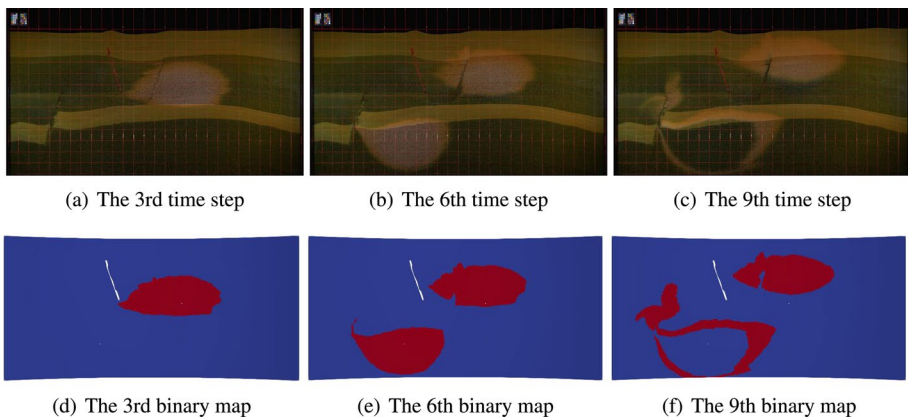
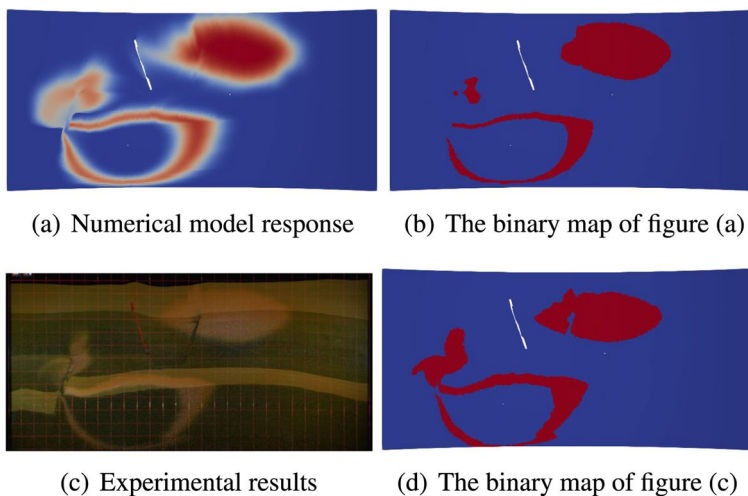


Fig. 6 The experimental images of tracer plumes and the associated digitized binary maps

Table 3 The petrophysical properties of the model layers

Facies	Porosity[-]	Permeability[D]	Anisotropy[x, z]
ESF	0.43	34.6	[1, 0.316]
C	0.44	302	[1, 1]
D	0.44	1016	[1, 1]
E	0.45	549	[1, 1]
F	0.45	1976	[1, 1]
G	0.44	1743	[1, 1]
Fault1	0.44	2554	[1, 1]
Fault3	0.44	739	[1, 1]
W	0.44	10000	[1, 1]

Figure 7 shows the corresponding forward model of the last time step (i.e. the 9th time step in Fig. 6) based on the tuned petrophysical properties of Table 3. In Fig. 7, the left figure is the tracer concentration map of the model response. The middle figure is the binary plot based on a given threshold of the tracer concentration (i.e. red colour if concentration $> \epsilon_t$; otherwise, blue). The right figure is the experimental result of the tracer test at the last time step. It is evident that the position of the simulated tracer plume (the left and middle figure) is systematically shifted from the experimental results. Assuming the forward physics is adequate, and that the optimization has found a global minimum, this indicates that the spatial model needs to be enriched in order to better match the experimental results. In the next section, we enlarge the model with additional parameters to address the underfitting problem. Spatial variation of the petrophysical properties will also be introduced as an attempt to model the variation caused by the manual sand deposition process in the experimental rig.

**Fig. 7** The tracer test results at the last time step

4.3 The Generation of the Prior Ensemble

In the previous sections, nine types of sand are “painted” into the model to represent the different layers and faults. This forms homogeneous petrophysical properties within a single layer. However, since the sand is manually filled into the rig, it is difficult to maintain homogeneous properties within the same layer. Moreover, compaction is observed in the course of pre-injection flushing, environmental temperature fluctuations, water injection, etc, which is certain to be non-uniform, and the tracer profiles exhibit visual features probably caused by internal heterogeneity. We concluded that it is necessary to include spatial heterogeneity into the model.

To this end, a high dimensional permeability field \mathbf{K} is introduced to the model, equipped with a prior distribution $\log(\mathbf{K}) \sim N(\log(\mathbf{K}_t), C_M)$, where \mathbf{K}_t is the spatially mapped tuned permeability from Table 3, and $C_M = \text{diag}\{\sigma^2\}$, with $\sigma = 0.02$. Samples may be drawn from this model in many standard ways, such as sequential Gaussian simulation, as depicted in Fig. 8.

4.4 History Matching Framework Using Randomized Maximum Likelihood (RML)

The enriched model is now equipped with some regularization apparatus, which behaves in a very similar way to explicit Bayesian prior declarations. Further, the high dimensionality of this model ($n = 72262$ parameters) means that dense Jacobian methods are no longer possible, and efficient gradient methods must be sought. The new objective function reads

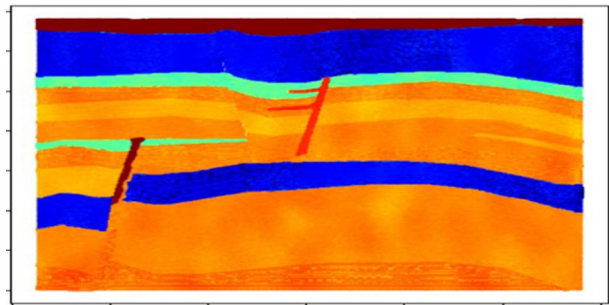
$$E(\mathbf{u}) = \|\mathbf{G}(\mathbf{u}) - \mathbf{d}_{\text{obs}}\|_2^2 + R(\mathbf{u}), \quad (8)$$

where R is the regularization term. Posterior samples of this model, associated with the Gibbs distribution $P(\mathbf{u}|\mathbf{d}_{\text{obs}}) \sim \exp(-E(\mathbf{u})/2)$, are generated using the Randomized Maximum Likelihood (RML) technique, briefly summarised in the Appendix.

In the gradient-based optimization method, we take all n transmissibilities of the grid cell interfaces as the model variables to do the history matching. In this study, we use the adjoint method in the DARTS framework to evaluate the gradient. The idea of the adjoint method is to introduce a Lagrange multiplier λ^T to combine the objective function E with the governing equation \mathbf{g} of the forward simulation. The new augmented objective function \bar{E} can be written as:

$$\bar{E}(\boldsymbol{\omega}, \mathbf{u}, \lambda) = E(\boldsymbol{\omega}, \mathbf{u}) + \lambda^T \mathbf{g}(\boldsymbol{\omega}, \mathbf{u}), \quad (9)$$

Fig. 8 A sample from the prior distribution for \mathbf{K} with spatial lognormal variation



where $\boldsymbol{\omega}$ is the state variable. The term λ^T is the transposed form of the Lagrange multiplier. Following Tian et al. (2021) and Tian and Voskov (2022), the discretized adjoint method can then be written as:

$$\lambda_{k+1}^T \frac{\partial \mathbf{g}_{k+1}}{\partial \boldsymbol{\omega}_k} + \lambda_k^T \frac{\partial \mathbf{g}_k}{\partial \boldsymbol{\omega}_k} + \frac{\partial \mathbf{j}_k}{\partial \boldsymbol{\omega}_k} = \mathbf{0}, \quad (10)$$

$$\lambda_K^T \frac{\partial \mathbf{g}_K}{\partial \boldsymbol{\omega}_K} + \frac{\partial \mathbf{j}_K}{\partial \boldsymbol{\omega}_K} = \mathbf{0}, \quad (11)$$

$$\frac{\partial \bar{\mathcal{E}}}{\partial \mathbf{u}} = \sum_{k=1}^K \left(\lambda_k^T \frac{\partial \mathbf{g}_k}{\partial \mathbf{u}_k} + \frac{\partial \mathbf{j}_k}{\partial \mathbf{u}_k} \right), \quad (12)$$

where K is the total number of the simulation time steps, \mathbf{j}_k is the misfit between the model response and the observation data at the given simulation time step k . Therefore $E(\boldsymbol{\omega}, \mathbf{u})$ can also be defined as:

$$E(\boldsymbol{\omega}, \mathbf{u}) = \sum_{k=1}^K \mathbf{j}_k + R(\mathbf{u}). \quad (13)$$

The Lagrange multiplier λ^T can be solved from Eqs. (11) and (10) recursively in a backward manner. Finally, the gradient $\frac{\partial \bar{\mathcal{E}}}{\partial \mathbf{u}}$ can be obtained from Eq. (12).

5 History Matching Formulation and Results

The history matching approach described below commences with binarized tracer images. Particular reformulations of the likelihood function proved to be efficacious in handling the data in this form, involving the *hinge* loss. We sketch this below, before surveying the inversion results generated using this measure.

5.1 Hinge Loss Function

As per the approach in Sect. 4.2, instead of using the original tracer plumes of the images as the observations, we digitized the original images into binary maps, e.g. the second row of Fig. 6. Based on the information of the binary maps, we replaced the misfit term of Eq. (8) (i.e. the first term at the right-hand side in this equation) with the hinge loss function. The updated objective function reads:

$$E(\mathbf{u}) = \|H(\mathbf{u})\|_2^2 + R(\mathbf{u}), \quad (14)$$

where $H(\mathbf{u})$ is the hinge loss function. If the given cell is located at the red region in Fig. 6, the hinge loss function is defined as:

$$H(\mathbf{u}) = \begin{cases} 0 & \text{if } G(\mathbf{u}) > \epsilon_t \\ G(\mathbf{u}) - \mathbf{d}_{\text{obs}} & \text{if } G(\mathbf{u}) \leq \epsilon_t \end{cases}. \quad (15)$$

Similarly, when the given cell is located at the blue region in Fig. 6, the hinge loss function is defined as:

$$H(\mathbf{u}) = \begin{cases} 0 & \text{if } G(\mathbf{u}) < \epsilon_t \\ G(\mathbf{u}) - \mathbf{d}_{\text{obs}} & \text{if } G(\mathbf{u}) \geq \epsilon_t \end{cases}. \quad (16)$$

5.2 Regularization and Randomized Maximum Likelihood

The regularization term $R(\mathbf{u})$ in Eq. (8) is defined as:

$$R(\mathbf{u}) = \alpha_R (\mathbf{u} - \mathbf{u}_{\text{ref}})^T C_M^{-1} (\mathbf{u} - \mathbf{u}_{\text{ref}}), \quad (17)$$

where α_R is the scaling coefficient, and \mathbf{u}_{ref} is the reference of the model variables (i.e. transmissibility at the grid cell interfaces). The covariance matrix C_M is modelled here as a diagonal, since the experimental rig is artificially filled, with few of the usual spatial correlations associated with deposition or consolidation.

In the Randomized Maximum Likelihood (RML) sampling method, \mathbf{u}_{ref} is a realization from the prior. The details of the generation of these priors are described in Sect. 4.3. Using the digitized binary observations in the RML method, we implemented multiple history matchings using sampled reference models \mathbf{u}_{ref} . Figure 9 demonstrates two samples of the history matching results of the tracer concentration at the last time step. These two samples correspond to the realizations of P10 and P90 of the accumulated mobile CO_2 in Box B, which will be demonstrated in Fig. 11. Their associated changes of the permeability after history matching are shown in Fig. 10.

Compared with the tracer concentration result in Fig. 7, the results in Fig. 9 match the experimental plume observations more accurately, especially for the top right plume. The total misfit error of the inferred model ensemble is reduced by around 39%. Figure 10 illustrates two RML samples of the permeability difference between the model before and after history matching. These two samples correspond to the realizations from Fig. 9. It is clear from Fig. 10 that the permeabilities in the tracer plume zone and the high permeable zone have more permeability adjustment compared with the rest of the region; large parts of the permeability of the sands are barely adjusted in Fig. 10 because the tracer does not reach them. Specifically, mathematically speaking, regions where the tracer does not reach exhibit zero gradient with respect to that particular zone, implying that no useful information can be gathered from those regions in the course of calibration. Thus, this calibration result is likely not transferable to another injection scenario. In this sense, we believe conducting more tracer tests at various injection locations would help capture sensitivity across the entire domain. However, in reality, multiple field experiments or geophysical tests can be very expensive or sometimes infeasible, making it difficult to cover the entire domain. Therefore, the primary concern should be whether the calibrated model can predict results to an acceptable degree. While a more detailed characterization would result in more accurate predictions, we must strike a balance between the accuracy of our predictions and the cost of calibration. In this FluidFlow project, the aim was to simulate a practical CO_2 storage project, in which a tracer test has been conducted, albeit with limited tracer test scenario and data collection. Our subsequent analysis suggests that our predictive model is highly effective in forecasting the CO_2 plume, as illustrated in our co-authored publication "The FluidFlow International Benchmark Study: Process, Modelling Results, and Comparison to Experimental Data."

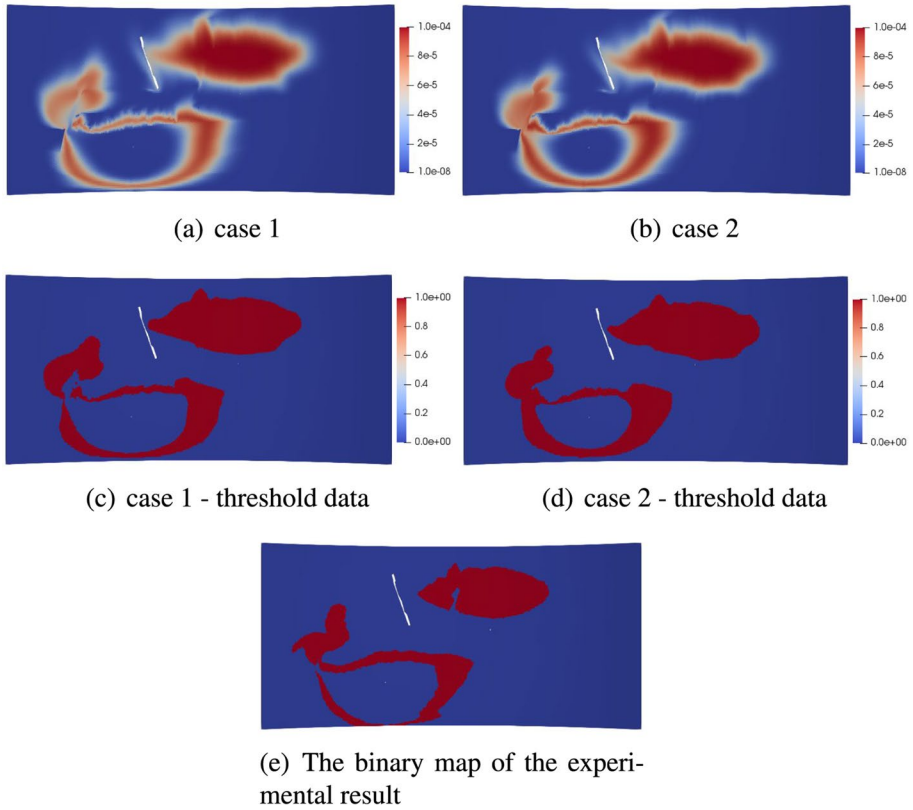


Fig. 9 Figure **a** and **b** represent two samples of the history matching results based on the binary observations. Figure **c** and **d** are their associated threshold data. Figure **e** is the binary map of experimental results. These figures demonstrate tracer test results at the last time step

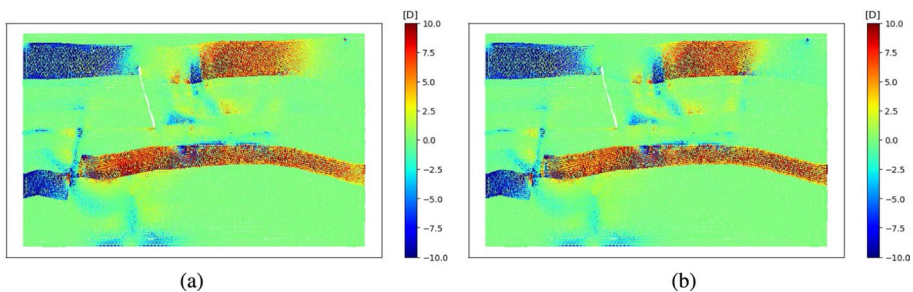


Fig. 10 Two samples of the changes of the permeability distribution after the history matching based on the binary observations

Utilizing the inferred model ensemble, we performed numerical simulations using 100 realizations of CO_2 injection, all with identical CO_2 injection rates and thermodynamic conditions. The ensemble outcomes of the CO_2 plumes at the last time step (after 120 h of CO_2 injection), as illustrated in Fig. 11, illustrates the variability associated

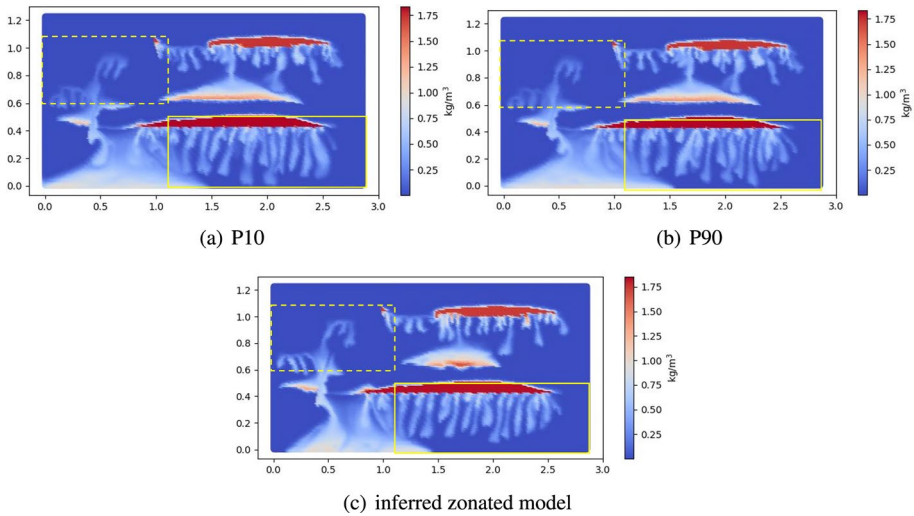


Fig. 11 The forward predictions of CO₂ plumes at the last time step (after 120 h of CO₂ injection). Figure **a** and **b** show the CO₂ plumes of the 10th and 90th percentiles of the entire ensemble of 100 inferred models calibrated to binary observations, respectively. Figure **c** shows the forward predictions of plumes based on the inferred zoned model. The solid and dashed boxes represent Box A and B, respectively

with the fingering phenomenon. The majority of uncertainty is concentrated in the areas exhibiting fingering.

For comparison, we present profiles depicting the temporal evolution of mobile CO₂ mass within Box A and B from Fig. 1. The variations in CO₂ quantities among different realizations within Box A and B, as illustrated in Fig. 12, emerge naturally from the posterior uncertainty in permeability. In Fig. 12, we observe that the zonal matched model overpredicts the mass of CO₂ in Box B. This is due to the zonal model’s assumption of a

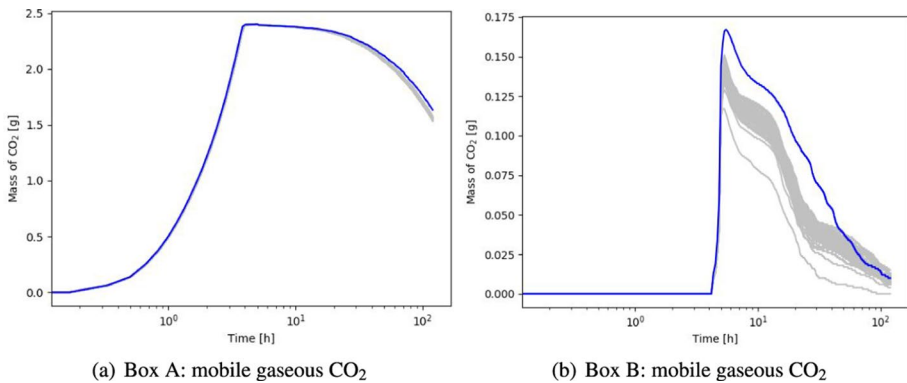


Fig. 12 The mobile gaseous CO₂ of Box A (left figure) and Box B (right figure) in Box B based on the inferred models calibrated to binary observations. The blue and grey curves represent the results based on the inferred zoned model and the inferred model ensemble, respectively

uniform thickness for the experimental rig, potentially introducing inaccuracies in estimating the model's total volume.

5.3 Regularization and RML Based on the Concentration Interpreted from Images

An alternative strategy of history matching was also conducted based on the concentration data interpreted from experimental images, as per the Tracer 1, 2, and 3 data in Fig. 4. In this case, we use the original objective function Eq. (8) to infer the model, instead of applying the hinge loss to the binary observations in Eq. (14). Based on the same realization ensemble \mathbf{u}_{ref} and RML method, the history matching results of the tracer concentration and the changes of permeability are shown in Figs. 13 and 14.

The error of the inferred model ensemble is reduced by around 16%. The misfits associated with the parsimonious layer models are not discussed here. Figure 13 has more similar plume shapes to the experimental images compared with the results in Fig. 9. Figure 14 shows more intensive adjustments of permeability by the optimizer in comparison to Fig. 10. This disparity can be attributed to the digitised concentration observations derived from the image, which provide additional "concentration gradient"

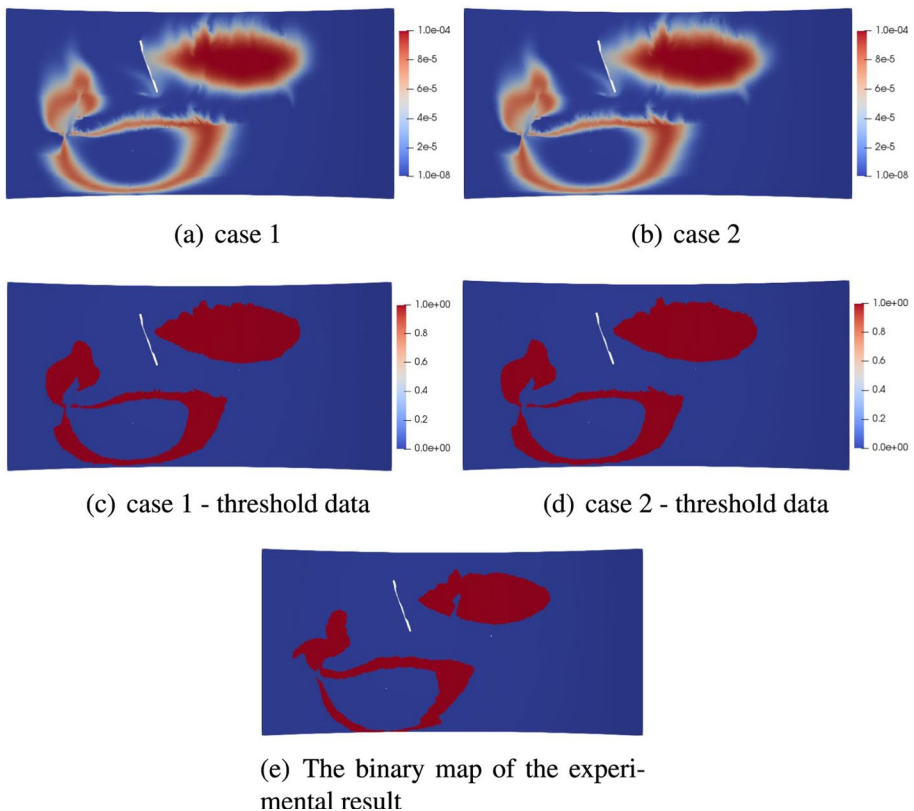


Fig. 13 Figure a and b represent two samples of the history matching results based on the tracer concentration observations. Figure c and d are their associated threshold data. Figure e is the binary map of the experimental results. These figures depict tracer test results at the last time step

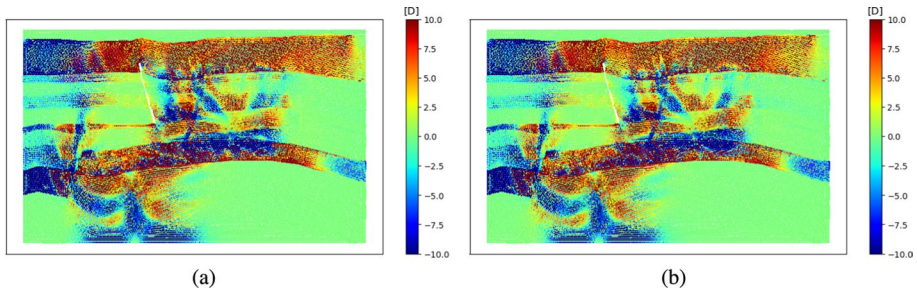


Fig. 14 Two samples of the changes of the permeability distribution after the history matching based on the tracer concentration observations

information to the optimizer. Conversely, the binary observation and the implementation of the Hinge loss function tend to attenuate this "gradient" information.

We also conducted the simulations of CO₂ injection based on the ensemble of 100 realizations of inferred models. The obtained results of the 10th and 90th percentiles of the entire ensemble are presented in Fig. 15. The CO₂ mass within Box A and B, as observed across the 100 realizations, is shown in Fig. 16. The curves in Figs. 16 and 12 show that the CO₂ mass within Box A remains relatively insensitive to variations in model calibration. This is primarily due to its substantial coverage of the CO₂ injection formation, where CO₂ concentrations significantly outweigh those in regions exhibiting fingering patterns characterized by lower CO₂ concentration. In essence, the concentration within the injected formation exerts a dominant influence on the mass of CO₂, rendering the mass of CO₂ within Box A rather indifferent to adjustments in the model calibration. While there is an improvement in the history matching results when compared to those of Fig. 9, the prediction outcomes of CO₂ injection do not exhibit significant enhancements. This observation suggests that the model predictions are not excessively sensitive to the selection of observations in the inversion. The ensemble results appear to be a reasonable quantification of the uncertainty associated with the fingering phenomenon.

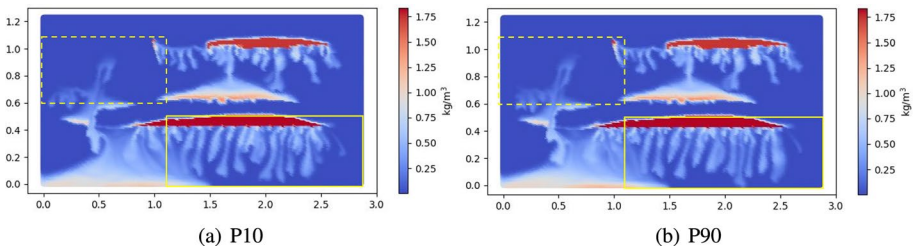


Fig. 15 The forward predictions of CO₂ plumes at the last time step (after 120 h of CO₂ injection). Figures **a** and **b** show the CO₂ plumes of the 10th and 90th percentiles of the entire ensemble of 100 inferred models calibrated to concentration observations, respectively. The solid and dashed boxes represent Box A and B, respectively

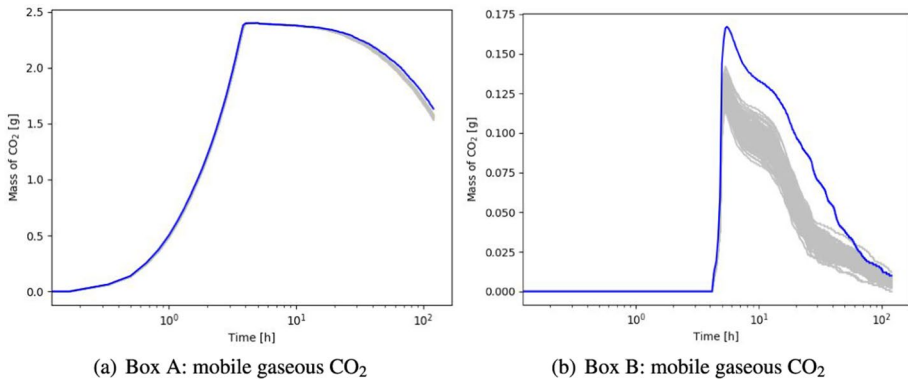


Fig. 16 The mobile gaseous CO₂ of Box A (left figure) and Box B (right figure) based on the inferred models calibrated to concentration observations. The blue and grey curves represent the results based on the zoned model and the inferred model ensemble, respectively

6 Discussion

In this study, we have performed history matching of the FluidFlower benchmark using a two-stage approach. In the first stage of history matching, a zoned model and a simple Poisson-like solver were utilized to infer the petrophysical parameters for each layer. The inferred model parameters are used in the second stage of history matching, which achieved notable improvements through the use of multiple high-dimensional spatial model realizations and adjoint methods. The misfits for both binary and concentration observations were reduced by 39% and 16%, respectively.

Naturally, the computational time for the simple-physics zoned model is much lower compared to the high-dimensional spatial model. Running a single-zoned model only takes approximately 5 seconds due to the assumption of steady-state pressure, and the Marquardt method converges in $O(10)$ iterations; for comparison, the high-dimensional spatial model forward model takes around 70 seconds on CPU, and the optimization typically around 100 times of iterations.

However, the simple zoned steady-state model does not account for the geological uncertainty within the layers associated with the dynamic changes in tracer behavior. The high-dimensional spatial model is better suited to quantifying this uncertainty, and more realistically models the complexity of real geological reservoirs and fluid dynamics. In the simulations of CO₂ injection using the inferred ensemble of high-dimensional spatial models, variability in the fingering phenomenon emerges from permeability uncertainty in the posterior. The percentile plot of CO₂ plumes of the entire ensemble show that the majority of uncertainty is concentrated in regions where fingering occurs.

In comparing the CO₂ injection simulations based on the model ensemble calibrated to binary and concentration observations, we observed an improvement in the history matching results when using the concentration observation. Somewhat paradoxically, the flow prediction outcomes of CO₂ injection did not exhibit significant enhancements. These findings indicate that the model predictions are not excessively sensitive to the selection of observations during the model inference process, at least within the context of the FluidFlower project.

The FluidFlower benchmark employs a relatively modest and small-scale experimental configuration. Nevertheless, in the context of real field-scale CO₂ storage projects, it is important to assess the transferability of the findings from this FluidFlower study. The geological and physical conditions in field-scale projects may differ substantially from the experimental setup used in our study. A thorough program of numerical simulations, as detailed in the simulation study for the FluidFlower benchmark (Wapperom et al. 2023), will highlight the importance of physical conditions when extending the FluidFlower methodology to actual CO₂ storage projects. A crucial issue is the substantial difference in physical properties such as density and viscosity between surface and typical reservoir conditions. Surface conditions introduce very strong buoyancy and marked nonlinearity into the system, thus posing computational challenges for both linear and nonlinear solvers. One symptom of this is the inability of the CPR preconditioner to operate effectively for the FluidFlower problem. In more representative reservoir conditions, smaller variations in physical properties result in comparatively fewer computational difficulties.

The methodology of this paper may be applied for practical scenarios resembling the shallow FluidFlower experimental conditions. The use of an unstructured mesh for discretization provides high flexibility in representing diverse features encountered in practical field-scale settings. Further, it is realistic to expect that real geological scenarios introduce increased uncertainty, such as channels, faults, and other complexities. A comprehensive approach to uncertainty quantification via simulation, based on a sizable ensemble of both geological scenarios and realizations, is a necessity. For a wider range of models and scenarios, the history matching procedure will naturally be more computationally intensive. These challenges underscore the importance of an efficient approach to history matching in real field-scale applications. To mitigate the computational burden in cases where available computational resources are more limited, one may consider implementing proxy methods or dimensionality reduction approaches to optimize computational efficiency.

7 Conclusion

A comprehensive history matching for the FluidFlower experimental results has been described in this study. Several stages of model tuning and inference have been implemented to improve the history matching results and estimate a spatial uncertainty range in permeability from the FluidFlower tracer tests. The pressure data were of limited use in this experiment due to noise and unknown frictional losses, but since the salient issues were unrepresentative of field measurements, we do not believe this reduces the importance of pressure data in the context of field-scale CO₂ injection. For example, in the case of the Ketzin pilot site's onshore CO₂ storage project, pressure and temperature are continuously monitored by sensors over the entire project lifetime (Martens et al. 2013). Furthermore, the mining authority has mandated an operating pressure threshold for this pilot project. Their monitoring data reveals a positive correlation between the injection rate and reservoir pressure.

Compared with the pressure data collected in this experiment, the tracer images were of high quality and a rigorous test of model choices and fitting. The history matching was first conducted utilizing a fast single-phase physics model without buoyancy effects, using a simple Poisson-like solver. A structured grid and digitized concentration data were utilized in this stage to infer the permeability of six facies types and two faults. The numerical

framework of this approach is a simple finite difference code with upwind tracer advection, using Marquardt methods for gradient-based optimisation.

To improve the model fit and address the issues around buoyancy and heterogeneity, enhancements in two stages were added, first for physical modelling, and later, significant model inflation to capture spatial heterogeneity. For the physics enhancements, the forward physics was switched to an unstructured grid with a facility for modelling buoyancy. Eight petrophysical parameters (including permeability and anisotropy) for model variables and simplified binary tracer images were used. Inversions were performed using the Gauss–Newton method with explicit Jacobians obtained by finite differences. This low-dimensional inversion was then taken as the reference model of the next stage of history matching using the adjoint-based gradient method as the workhorse in a Randomized Maximum Likelihood (RML) approach, with a richer and high-dimensional spatial model to account for heterogeneity. The coarse-scale morphology of the tracer images was remarkably well predicted by the low dimensional models, but fine detail could only be well reproduced by significant inflation of the spatial model space.

The high-dimensional spatial model is naturally better suited to representing typical geological heterogeneities, and has enough freedom to fit the finer spatial detail in the tracer data. We found that forward predictions of CO₂ distribution and fingering patterns are sensitive to this heterogeneity, but the uncertainties of these forecasts were not sensitive to whether the tracer data were transformed in the history matching process. The forward prediction of CO₂ mass within a specified area (e.g., Box A and B) demonstrates differing outcomes when accounting for the influence of buoyancy, despite the CO₂ plume contour exhibiting minimal sensitivity to the buoyancy effect.

We have demonstrated the cost and value of an escalating range of history matching models on the FluidFlower CO₂ storage benchmark model. The simple parsimonious models are fast and efficient but do not fit all the details of the experimental data due to simplifications in the physics and lack of resolution in the spatial models. The heterogeneous models are a successful demonstration of an adjoint method derived from operator-based linearization of multiphase physics in a very high dimensional model, and persuasively capture effects related to buoyancy and heterogeneity. In the predictive regime, the RML ensemble method produces a very useful uncertainty forecast of CO₂ injection behaviour. This is particularly evident in the fingering regions, which are strongly influenced by permeability fluctuations.

Appendix

To estimate the posterior uncertainty of the model after history matching, it suffices to generate an ensemble of approximate samples from the posterior PDF. From Bayesian extensions to the usual theory of parametric bootstrapping, it can be shown that a sample may be generated by solving

$$\mathbf{u}_c = \arg \min_{\mathbf{u}} E(\mathbf{u}) \quad (\text{A1})$$

where $E(\mathbf{u})$ is defined by

$$E(\mathbf{u}) = \frac{1}{2}(\mathbf{u} - \mathbf{u}_{\text{ref}})^T C_M^{-1}(\mathbf{u} - \mathbf{u}_{\text{ref}}) + \frac{1}{2}(\mathbf{G}(\mathbf{u}) - \mathbf{d}_{\text{obs}} + \boldsymbol{\epsilon})^T C_D^{-1}(\mathbf{G}(\mathbf{u}) - \mathbf{d}_{\text{obs}} + \boldsymbol{\epsilon}), \quad (\text{A2})$$

where C_D^{-1} is the inverse of the covariance matrix of the combined measurement and forward modelling error process, $\epsilon \sim N(0, C_D)$ is a “bootstrap” sample from that noise distribution, and $\mathbf{u}_{\text{ref}} \sim N(\bar{\mathbf{u}}, C_M)$ is a sample from the geostatistical prior. Note that in this study, the experimental measurement error data is not available, so the approach has been simplified by setting $\epsilon = \mathbf{0}$; most of the model uncertainty is controlled by the prior rather than experimental noise in this sort of problem. Finally, we get the modified objective function shown in Eq. (8) (omitting the coefficient $\frac{1}{2}$). A single sample of the posterior PDF can be obtained by solving Eq. (A1).

Acknowledgements Author Xiaoming Tian acknowledges the financial support from China Scholarship Council (CSC). The open-DARTS project is supported by the Netherlands eScience Center under Grant Number NLESC.OEC.2021.026. We also want to acknowledge Artur Palha and Luisa Orozco for their contribution to the open-DARTS development.

Funding The authors have not disclosed any funding.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. All authors contributed to the study conception and design. The first draft of the manuscript was written by Xiaoming Tian. The unstructured model was built by Michiel Wapperom. Section 3 was prepared by all the authors from CSIRO. James Gunning and Denis Voskov reviewed the manuscript. All authors read and approved the final manuscript. The experimental datasets used in this study are available in the repository *Final Benchmark Description: FluidFlower International Benchmark Study* (Nordbotten et al. 2022).

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

References

- Brouwer, D., Jansen, J.: Dynamic optimization of waterflooding with smart wells using optimal control theory. *SPE J.* **9**(04), 391–402 (2004). <https://doi.org/10.2118/78278-pa>
- Chen, W., Gavalas, G., Seinfeld, J., Wasserman, M.: A new algorithm for automatic history matching. *Soc. Pet. Eng. J.* **14**(06), 593–608 (1974). <https://doi.org/10.2118/4545-pa>
- Chiaramonte, L., Zoback, M., Friedmann, J., Stamp, V., Zahm, C.: Fracture characterization and fluid flow simulation with geomechanical constraints for a CO₂-EOR and sequestration project teapot dome oil field, Wyoming, USA. *Energy Procedia* **4**, 3973–3980 (2011). <https://doi.org/10.1016/j.egypro.2011.02.337>
- Eberhart, R., Kennedy, J.: A new optimizer using particle swarm theory. In: MHS’95. Proceedings of the sixth international symposium on micro machine and human science. IEEE, (1995). <https://doi.org/10.1109/mhs.1995.494215>
- Fathi, Z., Ramirez, F.W.: Optimal injection policies for enhanced oil recovery: part 2-surfactant flooding. *Soc. Pet. Eng. J.* **24**(03), 333–341 (1984). <https://doi.org/10.2118/12814-pa>
- Holland, J. H.: Genetic algorithms and adaptation. In: Adaptive control of Ill-defined systems, pp. 317–333. Springer (1984). https://doi.org/10.1007/978-1-4684-8941-5_21
- House, K.Z., Schrag, D.P., Harvey, C.F., Lackner, K.S.: Permanent carbon dioxide storage in deep-sea sediments. *Proc. Natl. Acad. Sci.* **103**(33), 12291–12295 (2006). <https://doi.org/10.1073/pnas.0605318103>
- IEA. Energy technology perspectives, special report on carbon capture utilisation and storage: ccus in clean energy transitions. IEA, (2020a). <https://www.iea.org/reports/ccus-in-clean-energy-transitions>

- IEA. CCUS in the transition to net-zero emissions. IEA, (2020b). <https://www.iea.org/reports/ccus-in-clean-energy-transitions/ccus-in-the-transition-to-net-zero-emissions>
- IEA. Carbon capture, utilisation and storage. IEA, (2020c). <https://www.iea.org/fuels-and-technologies/carbon-capture-utilisation-and-storage>
- IPCC, IPCC: Climate Change 2022: mitigation of climate change. Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change, p. 2022. Cambridge University Press, New York (2022)
- Khait, M., Voskov, D.: Operator-based linearization for efficient modeling of geothermal processes. *Geothermics* **74**, 7–18 (2018). <https://doi.org/10.1016/j.geothermics.2018.01.012>
- Khait, M., Voskov, D.: Adaptive parameterization for solving of thermal/compositional nonlinear flow and transport with buoyancy. *SPE J.* **23**(02), 522–534 (2018). <https://doi.org/10.2118/182685-pa>
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *Science* **220**(4598), 671–680 (1983). <https://doi.org/10.1126/science.220.4598.671>
- Leung, D.Y., Caramanna, G., Maroto-Valer, M.M.: An overview of current status of carbon dioxide capture and storage technologies. *Renew. Sustain. Energy Rev.* **39**, 426–443 (2014). <https://doi.org/10.1016/j.rser.2014.07.093>
- Lyu, X., Voskov, D.: Advanced modeling of enhanced co2 dissolution trapping in saline aquifers. *Int. J. Greenh. Gas Control* **127**, 103907 (2023). <https://doi.org/10.1016/j.ijggc.2023.103907>
- Lyu, X., Khait, M., Voskov, D.: Operator-based linearization approach for modeling of multiphase flow with buoyancy and capillarity. *SPE J.* **26**(04), 1858–1875 (2021). <https://doi.org/10.2118/205378-pa>
- Lyu, X., Voskov, D., Rossen, W.R.: Numerical investigations of foam-assisted CO2 storage in saline aquifers. *Int. J. Greenh. Gas Control* **108**, 103314 (2021). <https://doi.org/10.1016/j.ijggc.2021.103314>
- Madsen, K., Nielsen, H.B., Tingleff, O.: Methods for non-linear least squares problems, 2nd edn. Informatics and Mathematical Modelling Technical University of Denmark, Lyngby (2004)
- Martens, S., Liebscher, A., Möller, F., Hennings, J., Kempka, T., Luth, S., Norden, B., Prevedel, B., Szzybalski, A., Zimmer, M., Kuhn, M.: Group, K: CO2 storage at the ketzin pilot site, germany: Fourth year of injection, monitoring, modelling and verification. *Energy Procedia* **37**, 6434–6443 (2013). <https://doi.org/10.1016/j.egypro.2013.06.573>
- Mehos, G.J., Ramirez, W.: Use of optimal control theory to optimize carbon dioxide miscible-flooding enhanced oil recovery. *J. Pet. Sci. Eng.* **2**(4), 247–260 (1989). [https://doi.org/10.1016/0920-4105\(89\)90002-8](https://doi.org/10.1016/0920-4105(89)90002-8)
- Nocedal, J., Wright, S.J.: Numerical optimization. Springer, New York (2006). <https://doi.org/10.1007/978-0-387-40065-5>
- Nordbotten, J.M., Fernø, M., Flemisch, B., Juanes, R., Jørgensen, M.: Final benchmark description: fluidflower international benchmark study. (2022). <https://doi.org/10.5281/ZENODO.6807102>
- Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse theory for petroleum reservoir characterization and history matching. Cambridge University Press, Cambridge (2008). <https://doi.org/10.1017/CBO9780511535642>
- Ramirez, W.F., Fathi, Z., Cagnol, J.L.: Optimal injection policies for enhanced oil recovery: part 1 theory and computational strategies. *Soc. Pet. Eng. J.* **24**(03), 328–332 (1984). <https://doi.org/10.2118/11285-pa>
- Sarma, P., Aziz, K., Durlofsky, L.: Implementation of adjoint solution for optimal control of smart wells. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, (2005). <https://doi.org/10.2118/92864-ms>
- Sarma, P., Durlofsky, L.J., Aziz, K., Chen, W.H.: Efficient real-time reservoir management using adjoint-based optimal control and model updating. *Comput. Geosci.* **10**(1), 3–36 (2006). <https://doi.org/10.1007/s10596-005-9009-z>
- Schaeff, H., McGrail, B., Owen, A.: Carbonate mineralization of volcanic province basalts. *Int. J. Greenh. Gas Control* **4**(2), 249–261 (2010). <https://doi.org/10.1016/j.ijggc.2009.10.009>
- Tian, X., Voskov, D.: Efficient application of stochastic discrete well affinity (DiWA) proxy model with adjoint gradients for production forecast. *J. Petrol. Sci. Eng.* **210**, 109911 (2022). <https://doi.org/10.1016/j.petrol.2021.109911>
- Tian, X., Voskov, D.: Efficient inverse modeling framework for energy transition applications using operator-based linearization and adjoint gradients. (2023). <https://doi.org/10.2118/212169-MS>
- Tian, X., Blinovs, A., Khait, M., Voskov, D.: Discrete well affinity (diwa) data-driven proxy model for production forecast. *SPE J.* (2021). <https://doi.org/10.2118/205489-pa>
- The 11th society of petroleum engineers comparative solution project. <https://www.spe.org/en/csp/>, (2023). Accessed 25 April (2023)

- Volkov, O., Voskov, D.: Effect of time stepping strategy on adjoint-based production optimization. *Comput. Geosci.* **20**(3), 707–722 (2016). <https://doi.org/10.1007/s10596-015-9528-1>
- Voskov, D.: Operator-based linearization approach for modeling of multiphase multi-component flow in porous media. *J. Comput. Phys.* **337**, 275–288 (2017)
- Voskov, D., Saifullin, I., Wapperom, M., Tian, X., Palha, A., Orozco, L., Novikov, A.: Open delft advanced research terra simulator (open-DARTS), (2023). URL <https://doi.org/10.5281/zenodo.8046983>
- Vreugdenhil, C., Koren, B.: A robust upwind discretization method for advection, diffusion and source terms, pp. 117–138. *Notes on Numerical Fluid Mechanics*. Vieweg, Germany, (1993)
- Wang, Y., Voskov, D., Khait, M., Bruhn, D.: An efficient numerical simulator for geothermal simulation: a benchmark study. *Appl. Energy* (2020). <https://doi.org/10.1016/j.apenergy.2020.114693>
- Wapperom, M., Tian, X., Novikov, A., Voskov, D.: FluidFlow benchmark: lessons learned from the perspective of subsurface simulation. *Transp. Porous Media* (2023). <https://doi.org/10.1007/s11242-023-01984-8>
- Wei, L., Ramirez, W.F., Qi, Y.F.: Optimal control of steamflooding. *SPE Adv. Technol. Ser.* **1**(02), 73–82 (1993). <https://doi.org/10.2118/21619-pa>
- White, C.M., Strazisar, B.R., Granite, E.J., Hoffman, J.S., Pennline, H.W.: Separation and capture of CO₂ from large stationary sources and sequestration in geological formations-coalbeds and deep saline aquifers. *J. Air Waste Manag. Assoc.* **53**(6), 645–715 (2003). <https://doi.org/10.1080/10473289.2003.10466206>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.