

Performance of an open-source image-based history matching framework for CO₂ storage

D. Landa-Marbán¹ T.H. Sandve¹ J.W. Both² J.M. Nordbotten^{1,2}
S.E. Gasda^{1,3}

¹ Division of Energy and Technology, NORCE Research AS, Nygårdsgaten 112, Bergen 5008, Norway.

² Center for Modeling of Coupled Subsurface Dynamics, Department of Mathematics, University of Bergen, Allégaten 41, Bergen 5007, Norway.

³ Department of Physics and Technology, University of Bergen, Allégaten 41, Bergen 5007, Norway.

Corresponding author: David Landa-Marbán (E-mail address: dmar@norceresearch.no).

Abstract

We present a history matching (HM) workflow applied to the International FluidFlower benchmark study dataset, which features high-resolution images of CO₂ storage in a meter-scale, geologically complex reservoir. The dataset provides dense spatial and temporal observations of fluid displacement, offering a rare opportunity to validate and enhance HM techniques for geological carbon storage (GCS). The combination of detailed experimental data and direct visual observation of flow behavior at this scale is novel and valuable. This study explores the potential and limitations of using experimental data to calibrate standard models for GCS simulation. By leveraging high-resolution images and resulting interpretations of fluid phase distributions, we adjust uncertain parameters and reduce the mismatch between simulation results and observed data. Simulations are performed using the open-source OPM Flow simulator, while the open-source Everest decision-making tool is employed to conduct the HM. After the HM process, the final simulation results show good agreement with the experimental CO₂ storage data. This suggests that the system can be effectively described using standard flow equations, conventional saturation functions, and typical PVT properties for CO₂-brine mixtures. Our results demonstrate that the Wasserstein distance is a particularly effective metric for matching multi-phase, multi-component flow data. The entire workflow is implemented in a Python package named **pofff** (Python OPM Flow FluidFlower), which organizes

all functionality through a single input file. This design ensures reproducibility and facilitates future extensions of the study.

Keywords Data assimilation · FluidFlower · Geological carbon storage · Modeling calibration · Numerical simulations · Open-source software · OPM Flow

Highlights

- The performance of OPM Flow enables rapid history matching studies on the FluidFlower (a single run completing in ca. 2 minutes)
- Calibrated parameters are sensitive to grid type and resolution, i.e., parameter validity is restricted to the specific input grid used
- The workflow for conducting history matching studies is provided in a user-friendly Python package named `pofff`

1 Introduction

Geological carbon storage (GCS) is considered a crucial strategy for reducing emissions and achieving net-zero targets, supported by more than 30 years of pilot and commercial operations worldwide ([Global CCS Institute \[2021\]](#)). In recent years, interest in GCS has grown significantly, with several major projects either underway or in development. Notable examples include Quest in Canada, Porthos in the Netherlands, and Northern Lights in Norway ([Furre et al. \[2019\]](#)). The collective experience from these diverse projects, spanning a range of geological and technical settings, has been complemented by multiple decades of research and insights from controlled laboratory experiments and modeling studies. Together, these efforts have contributed to a robust and evolving knowledge base on GCS processes. This understanding has been incorporated into both commercial and research-oriented simulation tools designed to model CO₂ migration and trapping within storage formations ([Rasmussen et al. \[2021\]](#), [Lie \[2019\]](#), [Koch et al. \[2021\]](#), [Voskov et al. \[2023\]](#)). These models are used in conjunction with detailed geological characterization and monitoring data throughout all stages of project development and operation. Applications include estimating storage capacity and injectivity, optimizing injection strategies, and supporting risk assessment and management.

Predicting GCS performance in realistic geological settings remains a significant challenge. This difficulty arises primarily from two sources of uncertainty. First, reservoir models are typically constructed using sparse well data and relatively low-resolution seismic surveys, which introduce substantial uncertainty into model parameters. Second, simulation tools often rely on simplified

physical approximations to maintain computational efficiency, leading to modeling errors. As an example, a common simplification is to neglect capillary pressure at the field scale (Ni et al. [2025]). Measured data play a crucial role in reducing both types of uncertainty and enhancing predictive accuracy. History matching (HM) algorithms, reducing the mismatch between collected data and simulation results, offer a systematic approach to calibrating model parameters by treating the reservoir simulator as a black box. In addition to parameter estimation, the resulting optimized simulation results and minimal mismatch can reveal missing or poorly represented physical processes within the underlying mathematical model.

Several field studies have shown that matching field data often requires a combination of parameter calibration and adjustments to the mathematical model itself. A prominent example is the Sleipner benchmark dataset (Equinor [2020]), where time-lapse seismic data have been used to calibrate formation properties, reservoir structure, geothermal gradients, and fluid composition, with varying levels of success (IEAGHG [2021]). These studies highlight the dominant role of gravity in the Utsira formation and suggest that HM alone is insufficient to address the limitations of standard simulators in capturing the strong gravity-driven segregation and migration patterns observed in the seismic data. Other notable examples include the In Salah project (Ringrose et al. [2013]), where the assumption of linearly elastic deformation failed to reproduce observed surface uplift suggesting the use of nonlinear stress-strain relationships (Rinaldi and Rutqvist [2017]), and the Tubåen injection at Snøhvit, where matching downhole pressure data required incorporating additional geological heterogeneity, such as sub-seismic faults (Hansen et al. [2013]), and physical processes like salt precipitation (Grude et al. [2014]).

Solubility trapping is a key mechanism in GCS, where the dissolution of CO_2 into formation brine is enhanced by density-driven convection. As CO_2 dissolves, the brine becomes denser, leading to gravitational instabilities that cause the CO_2 -rich brine to sink in finger-like patterns (Ennis-King and Paterson [2003]). This convective mixing process has been extensively studied in controlled laboratory settings, including Hele-Shaw cell experiments conducted under relevant temperature and pressure conditions (Faisal et al. [2015], Amarasinghe et al. [2020]), as well as in intermediate-scale porous media experiments (Agartan et al. [2020]) and numerical simulations (Neufeld et al. [2010], Elenius and Gasda [2013]). These studies have primarily focused on the dissolution process in both homogeneous and heterogeneous media, often in isolation from other migration and trapping mechanisms. While experimental results suggest that convective mixing should occur at the field scale, direct observation of this phenomenon in subsurface reservoirs remains elusive. At the Sleipner site, gravimetric data have been used to estimate an upper bound on dissolution rates due to convective mixing. Subsequent simulation studies have supported the plausibility of these estimates by comparing modeled results with seismic observations (Mykkeltvedt and Nordbotten [2012]).

Experimental studies have played a key role in constraining model assumptions and parameters under idealized conditions. However, verifying and calibrating simulation models that account for

convective mixing alongside other CO₂ storage processes in realistic geological settings remains a significant challenge. This difficulty stems from the limited availability of data that capture convective mixing in conjunction with CO₂ migration influenced by factors such as injection dynamics, formation and petrophysical properties, residual trapping, permeability and capillary barriers, and fault-related fluid flow. Recent advances have addressed this gap through the FluidFlower experimental system, a meter-scale CO₂ injection rig designed to replicate dominating physical processes during geological CO₂ storage (Fernø et al. [2024]). At this scale, model parameters and injection conditions can be precisely measured and controlled. The facility enables direct observation of CO₂ migration and dissolution using high-resolution visual imaging, resulting in a spatially dense, time-lapse dataset that captures the dynamic behavior of CO₂ in porous media. Accompanied with calibrated image analysis tools (Nordbotten et al. [2024a]), quantitative datasets can be generated enabling comparison against simulation results. In this spirit, the double-blind, community-wide FluidFlower Validation Benchmark Study has been organized (Nordbotten et al. [2022]), comprising of five repeated multi-day experiments of CO₂ storage in a meter-scale, layered, heterogeneous sand geometry. Nine international modeling groups participated aiming at replicating the experiments without being able to review the experimental results before submitting their results. Finally, quantitative comparisons were drawn between the modeling and experimental results (Flemisch et al. [2024]) and the accompanying image data, processing and data comparison scripts have been published.

The International FluidFlower benchmark study dataset (Eikehaug et al. [2023]) offers a unique opportunity to investigate a key open question in GCS: how does convective mixing interact with other multi-phase flow processes, and can standard simulators accurately capture this interplay? Specifically, the study examines whether physical mechanisms that are currently missing from existing models need to be incorporated, or if improved predictability can be achieved through the careful calibration of model inputs using experimental data. To address this question, we apply a HM workflow to time-lapse images from the series of CO₂ injection experiments associated to the FluidFlower benchmark. On the one hand, this approach enables direct comparison between simulated and observed flow behavior, providing insight into the adequacy of current modeling frameworks. On the other hand, through the systematic approach and with full data visibility, this study directly complements the community-wide benchmark initiative (Flemisch et al. [2024]), providing the possibility to provide the first "optimal" modeling fit of the dataset. Furthermore, this also clarifies whether parameter tuning is a key factor in aligning the model with the experimental data, especially in contrast to the double-blind benchmark participants.

Regarding numerical simulations, reproducibility remains a significant challenge in research published in academic journals. For example, a recent study by Riehl et al. [2025] examined 11,879 simulation studies and 672 associated repositories in the field of transportation modeling, finding that fewer than 2 percent provided supplementary materials to support their findings – a trend we

must also assume conceptually applies to the field of reservoir modeling. This highlights a broader issue across scientific disciplines, including reservoir engineering. As a key contribution of this work, we introduce an open-source workflow named `pofff` (Python OPM Flow FluidFlower), which is hosted in a public repository (Landa-Marban [2025]). This tool adopts the workflow and methods from `pyopmspe11`, a Python framework using OPM Flow for the SPE11 benchmark project (Landa-Marban and Sandve [2025]), aiming at multiphase flow simulations of FluidFlower experiments. The repository contains the complete model setup, simulation tools, and HM framework, enabling researchers to reproduce the results presented in this study. Moreover, the workflow is designed to be adaptable for future experiments conducted in the FluidFlower facility (Eikehaug et al. [2024]) or similar experimental rigs. This resource aims to support ongoing research efforts and serve as a practical training tool for the broader scientific community.

The remainder of the paper is organized as follows. Section 2 outlines the computational workflow, including a description of the experimental dataset, the mathematical model used for flow simulations, and the algorithm employed for HM. Section 3 presents the HM results and compares them with previously published outcomes from the FluidFlower benchmark study. Section 4 offers concluding remarks and discusses directions for future research. The Appendix includes an example input configuration file for the `pofff` workflow, as well as illustrative model sensitivity results.

2 Methodology

In this section we describe the followed steps from the design of the FluidFlower benchmark experiments to the development of the Python OPM Flow FluidFlower `pofff` simulation tool.

2.1 Experimental FluidFlower data

This subsection summarizes the laboratory CO₂ storage experiments conducted as part of the FluidFlower benchmark. A detailed description of these, allowing to setup associated simulations, is provided in Nordbotten et al. [2022], and the experimental data is presented in Ferno et al. [2024].

The FluidFlower is a meter-scale CO₂ injection rig featuring a complex geometric design and transparent glass walls, enabling direct observation and high-resolution data acquisition of CO₂ migration. The experimental setup of the FluidFlower is described in more detail by Ferno et al. [2024], Haugen et al. [2024], and Eikehaug et al. [2024]. The geometry of the FluidFlower rig, used in this study, is illustrated in Figure 1.

The experimental domain has a length of 2.8 meters and a height of 1.5 meters. The initial thickness (before running the CO₂ experiments) varies from 0.019 meters at the sides to a maximum of 0.028 meters at the center (see Figure 3 in Nordbotten et al. [2022]). The injection ports, indicated by red circles in the figure, have a radius of 0.0009 meters. The density of the water used

Table 1: Measured sand properties (Nordbotten et al. [2022])

Id	Sand	Grain size [mm]	k [D]	ϕ [-]	$s_{r,w}$ [-]	$k_{r,g}^e$ [-]	$s_{r,g}$ [-]	$k_{r,w}^e$	p_e [Pa]
1	ESF	0.20±0.11	44	0.43	0.32	0.09	0.14	0.71	1471.5
2	C	0.66±0.09	473	0.44	0.14	0.05	0.10	0.93	294.3
3	D	1.05±0.14	1110	0.44	0.12	0.02	0.08	0.95	98.1
4	E	1.45±0.19	2005	0.45	0.12	0.10	0.06	0.93	-
5	F	1.77±0.31	4259	0.45	0.12	0.11	0.13	0.72	-
6	G	2.51±0.63	9580	0.44	0.10	0.16	0.06	0.75	-

corresponds to the residual water saturation, below which the water phase is immobile. Similarly, $s_{r,g}$ is the residual gas saturation, defining the threshold below which the gas phase is immobile. The terms $k_{r,g}^e$ and $k_{r,w}^e$ indicate the endpoint relative permeabilities of gas and water, respectively. Finally, p_e refers to the capillary entry pressure.

As shown in Table 1, entry pressure measurements are not available for Sands E, F, and G. This is due to limitations in the measurement technique. For reference, the entry pressure of 98.1 Pa reported for Sand D corresponds to a measured gas column height of only 0.01 meters. Given the sensitivity of fluid distribution and composition to entry pressure values, these parameters play a critical role in the simulation and must be estimated through the HM process.

2.1.2 CO₂ experiments and data acquisition

Figure 2 displays a subset of the experimental data from the CO₂ injection study, which is described in full detail by Fernø et al. [2024]. Among the five experimental runs (C1-C5) conducted under nearly identical operational conditions, the results presented here correspond to run C2 (Fernø et al. [2024]). In this run, CO₂ was injected at a rate of 10 ml/min under standard conditions (equivalent to 1.67×10^{-7} m³/s). Injection into the lower well (see Figure 1 for location) lasted for 5 hours and 5 minutes. Injection into the upper well began 2 hours and 15 minutes after the lower well injection started and continued for 2 hours and 50 minutes. Both wells were shut down simultaneously after 5 hours and 5 minutes of total injection time. The final image in the dataset was captured five days after the start of the experiment, marking the end of the experiment.

High-resolution photographs of the CO₂ migration define the data acquisition. The optical images have been processed using DarSIA (Nordbotten et al. [2024a]), an open-source Python-based image analysis toolbox for Darcy-scale images. This tool enables the generation of segmented images, which are used to compare the simulation results submitted by various research groups to the five experimental runs as discussed in Flemisch et al. [2024]. The segmentation identifies the spatial distribution of the different fluid phases: gaseous CO₂, dissolved CO₂, and pure water. This segmented data also build the foundation for the history matching presented herein.

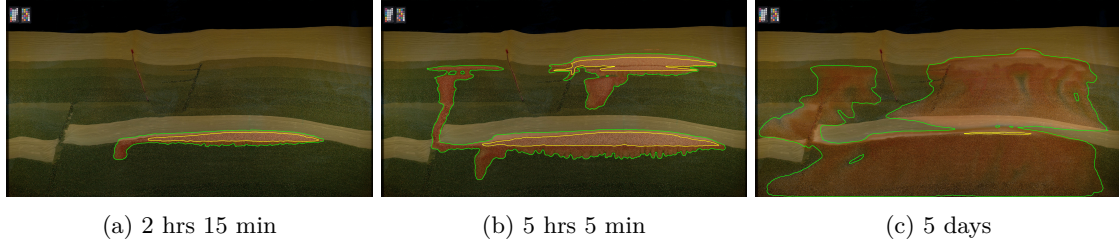


Figure 2: Photographs from the CO₂ injection experiment (C1) showing contours of dissolved CO₂ (green) and gaseous CO₂ (yellow). a) Injection into the lower well only. b) Injection into both the lower and upper wells. c) Post-injection phase after both wells were shut down

2.2 Reservoir simulator

The numerical simulations corresponding to the CO₂ injection experiments are performed using the open-source simulator OPM Flow ([Rasmussen et al. \[2021\]](#)). OPM Flow is a reservoir simulator that supports industry-standard input and output formats, making it suitable for both research and practical applications. In this study, we briefly describe the mathematical model for GCS as implemented in OPM Flow, and provide references for readers seeking more detailed information. It is worth noting that this implementation has demonstrated strong performance in recent benchmark studies, including the 11th Society of Petroleum Engineers Comparative Solution Project ([Nordbotten et al. \[2025\]](#)), which was inspired by the FluidFlower experimental setup.

2.2.1 Mathematical model

The OPM Flow simulator includes a dedicated module for CO₂ storage applications. When this option is enabled, the simulator internally computes fluid properties such as density, viscosity, and the solubility limit of CO₂ in brine. These properties are calculated as functions of pressure, temperature, and composition using analytical correlations and models from the literature, rather than relying on interpolation from tabulated data. Internally, these compositional properties are converted to their black-oil equivalents, allowing the simulator to retain the computational efficiency and robustness of a black-oil formulation while achieving the accuracy typically associated with compositional models. A detailed description of the CO₂ storage module in OPM Flow is provided in [Sandve et al. \[2021\]](#). In this study, we adopt a simplified version of the CO₂ storage model tailored to the FluidFlower experiments. The formulation and notation used follow the conventions presented in [Nordbotten et al. \[2024b\]](#).

The pore space in the reservoir is occupied by a two-component, two-phase fluid system, consisting of water and CO₂ as the components ($i \in \{\text{H}_2\text{O}, \text{CO}_2\}$), and liquid and gas as the phases

($\alpha \in \{l, g\}$). The two-phase extended Darcy's law for phase α is written as:

$$\mathbf{u}_\alpha = -\frac{k_{r,\alpha}\mathbf{k}}{\mu_\alpha}(\nabla p_\alpha - \rho_\alpha\mathbf{g}), \quad (1)$$

where $\mathbf{u}_\alpha$ is the flux discharge per unit area [m/s], $k_{r,\alpha}$ the relative permeability [-], μ_α the dynamic viscosity [Pa·s], p_α the pressure [Pa], ρ_α the density [kg/m³], $\mathbf{g}$ the gravity (9.81 [m/s²]), and $\mathbf{k}$ a symmetric tensor of rank 2 for the rock permeability [m²]. The component mass conservation is written as:

$$\sum_{\alpha=l,g} \left\{ \frac{\partial}{\partial t} [\rho_\alpha \phi s_\alpha \chi_\alpha^i] + \nabla \cdot \rho_\alpha [\mathbf{u}_\alpha \chi_\alpha^i - (s_\alpha \phi D_\alpha + E \|\mathbf{u}_\alpha\|) \nabla \chi_\alpha^i] \right\} = 0, \quad (2)$$

where ϕ is the porosity [-], s_α the saturation [-], χ_α^i the component mass fraction in phase α [-], D_α the molecular diffusion [m²/s], E the isotropic dispersion coefficient [m], and $\|\cdot\|$ denotes the Euclidean norm. The phase partitioning of each component is defined according to [Spycher et al. \[2003\]](#). Under room temperature and pressure conditions, water vaporization is minimal. Therefore, in the simulations presented in this study, vaporization of water into the gas phase ($\chi_g^{\text{H}_2\text{O}}$) is neglected. The phase saturations, pressures, and components fulfill the following conditions:

$$p_g - p_l = p_c, \quad s_l + s_g = 1, \quad \text{and} \quad \sum_{i=\text{H}_2\text{O}, \text{CO}_2} \chi_\alpha^i = 1 \quad \alpha \in \{l, g\}, \quad (3)$$

where $p_c(s_l)$ is the capillary pressure [Pa]. In this study, we use the Brooks-Corey functions, where these relationships are given as a function of the effective saturations s_α^* :

$$s_\alpha^* = \max \left(\frac{s_\alpha - s_{\alpha,i}}{1 - s_{\alpha,i}}, 0 \right), \quad k_{r,\alpha} = (s_\alpha^*)^{n_\alpha}, \quad \text{and} \quad p_c = p_e (s_l^*)^{-\frac{1}{n}}, \quad (4)$$

where $s_{\alpha,i}$ is the saturation below which the phase is immobile [-], p_e the entry pressure [Pa], and n_α and n fitting coefficients [-]. Although OPM Flow supports hysteresis modeling, it is intentionally omitted in the simulations presented in this study as a simplification choice to reduce model complexity.

The PVT (pressure, volume, temperature) properties such as densities and viscosities as a function of pressure, temperature, and composition are computed internally by using analytical correlations and models from the literature ([Bell et al. \[2014\]](#)). We refer to the OPM Flow manual for details on these models ([Goodfield et al. \[2025\]](#)).

2.2.2 Spatial characterization

A grid representing the different facies in the FluidFlower rig is required as input to the simulator. Cartesian grids offer computational efficiency due to the use of the two-point flux approximation in the discretization scheme (Lie [2019]). However, accurately representing the interfaces between sand layers with Cartesian grids requires a very fine spatial resolution, which can increase computational cost. Unstructured grids provide greater flexibility by allowing the grid to conform to complex geological features such as sand layers and faults. This adaptability comes with increased complexity in the discretization matrix and may introduce grid-orientation effects when using two-point flux approximation (Klemetsdal et al. [2017]). Corner-point grids are widely used in subsurface simulations because they balance geometric flexibility with computational efficiency (Ponting [1989]). These grids are defined by vertical pillars and horizontal lines connecting the pillars. For the simulations presented in this study, we use the corner-point grid shown in Figure 3, which allows for accurate representation of geological structures while maintaining compatibility with the numerical methods employed. We refer to Holme et al. [2025] for an extensive study on grid-orientation effects and discretization methods on the FluidFlower geometry.

The computational grid in Figure 3 comprises 140 elements in the horizontal dimension, each with a uniform size of 2 cm, and 69 elements in the vertical dimension, where cell height varies from a minimum of 4.7×10^{-3} cm to a maximum of 3.11 cm, with an average value of 1.74 cm. This configuration results in a total of 9,960 grid cells, of which 9,627 are active. The remaining 333 cells are inactive: 18 correspond to the sealed fault (see Figure 1), while 315 are located in the lower-right corner of the grid, where the horizon of facies G coincides with facies F. The grid itself is defined with a constant thickness equal to the minimum value from the measured thickness map. Variations are incorporated indirectly by adjusting the pore volumes and transmissibilities of the cells.

2.3 Modeling decisions and comparison metrics

This study is closely aligned with the special issue on the “FluidFlower validation benchmark study” (Nordbotten et al. [2024c]), which includes 14 research papers that contribute valuable knowledge relevant to the current work. This work in particular complements and reinforces the collective understanding of the benchmark, offering additional validation and perspective. Additionally, results from the 11th SPE Comparative Solution Project (Nordbotten et al. [2025]), a benchmark study inspired by FluidFlower, are also pertinent to this research. This subsection highlights key findings from both contributions that informed our modeling decisions, particularly regarding grid type and resolution.

As part of the later discussion, our history matched results will be put into direct comparison with not only the experimental data itself, but also the other modeling contributions to the model

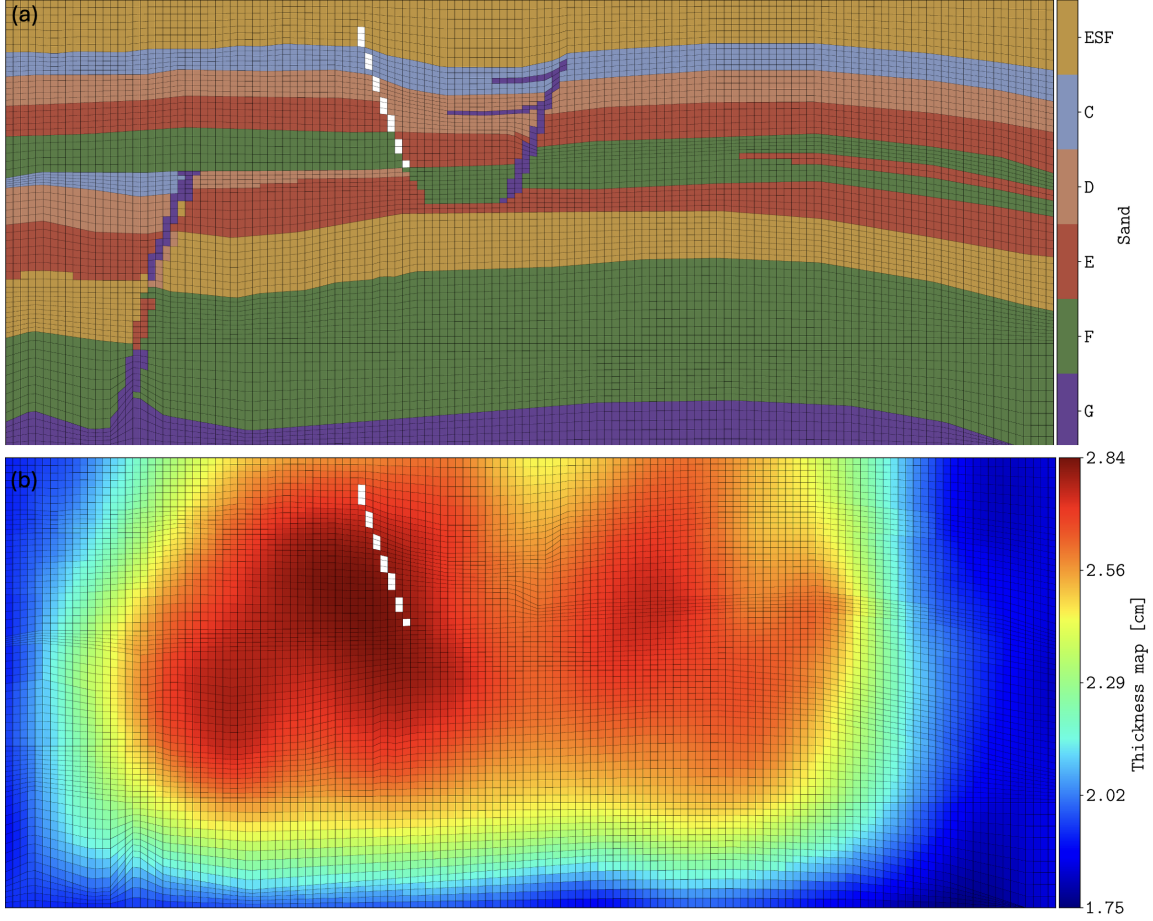


Figure 3: (a) Corner-point grid representation of the FluidFlower geological model. (b) Measured thickness map of the FluidFlower

validation benchmark (Flemisch et al. [2024]). One of the metrics employed in this comparison is the Wasserstein distance, also referred to as the Earth mover's distance or optimal transport distance (Villani [2008]). For two distributions $\rho_A : \Omega \rightarrow \mathbb{R}_+$ and $\rho_B : \Omega \rightarrow \mathbb{R}_+$ defined over the space Ω and of same measure, i.e., $\int_{\Omega} \rho_A dV = \int_{\Omega} \rho_B dV$, their Wasserstein distance is defined through a nonlinear minimization problem originating from optimal transport, also called Beckmann problem (Santambrogio [2015]):

$$W^1(\rho_A, \rho_B) := \min \left\{ \int_{\Omega} |\mathbf{q}| dV \mid \nabla \cdot \mathbf{q} = \rho_A - \rho_B, \mathbf{q} \cdot \mathbf{n}|_{\partial\Omega} = 0 \right\}. \quad (5)$$

In words, this metric determines the weakest flux field that transports the one distribution into

another. The use of the L^1 norm localizes the transport, resulting in a shortest path. With this, it provides an objective measure for data with inherent transportation character and is convenient for model-experiment comparisons in flow in porous media (Both et al. [2024]). Among all nine submitted simulation results, the results from CSIRO show the closest agreement with the experimental data in terms of the Wasserstein distance (Green et al. [2024]); their results will be later explicitly highlighted in the discussion and plots of the HM results.

In an associated study, Saló-Salgado et al. [2024] manually calibrated model parameters using experimental data from a smaller-scale setup and used these parameters to simulate the FluidFlow system. One key finding from this study is the high sensitivity of the system to variations in permeability and capillary pressure. To evaluate model performance, the Wasserstein distance was computed relative to the experimental results. Among the tested configurations, Model 1 showed the closest agreement with the experimental data; again, the corresponding results will be explicitly highlighted in the discussion and plots of the HM results. In our paper, Model 1 is referred to as MIT.M1.

Several studies of the FluidFlow special issue discuss the sensitivity of simulation results to model choices and parameter variations. One study investigates the roles of hysteresis and molecular diffusion, finding that hysteresis contributes minimally, whereas enhanced diffusion significantly improves dissolution trapping (Wang et al. [2024]). Another contribution compares discretization techniques, evaluating Cartesian versus unstructured grids and contrasting two-point with multi-point flux approximations (Wapperom et al. [2024]). A further contribution uses ensemble simulations to assess uncertainties in rock and gas relative permeability, revealing a notable sensitivity to the image segmentation threshold, particularly when comparing the assigned value in the FluidFlow benchmark study of 0.1 kg/m^3 (CO_2 concentration in the liquid phase) with an alternative of 0.05 kg/m^3 (Jammoul et al. [2024]). Additionally, permeability field heterogeneity is analyzed through HM at each grid cell, incorporating both tracer and CO_2 concentration data (Tian et al. [2024]).

One of the key findings from the SPE11 benchmark study is that human errors appear to be the primary source of variability in the simulation results (Nordbotten et al. [2025]). This underscores the importance of open-source code, which enables independent validation and verification by the broader research community. Another relevant contribution to the present study is the comparison of grid structures and convergence criteria, provided by the OPM team. Figure 4 presents spatial maps of dissolved CO_2 after two days of simulation. For details on the simulation setup, configuration files, and additional results, refer to the documentation of the `pyopmspe11` tool (Landa-Marbán and Sandve [2025]).

Figure 4a, corresponding to SPE11 simulation results close to the considered FluidFlow setup, illustrates that in simulations using 1 cm Cartesian grids, convective mixing ceases in the lower storage unit, an artifact attributed to this grid configuration (Flemisch et al. [2024]). Figure 4b

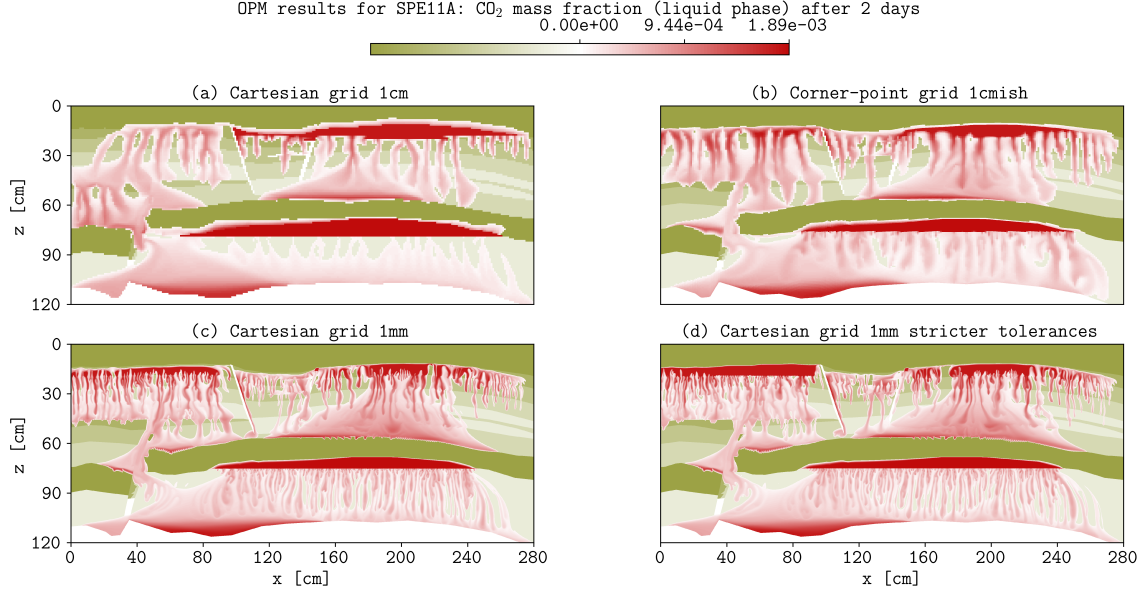


Figure 4: Spatial distribution of dissolved CO₂ from the OPM simulation results in case SPE11a, which closely resembles the FluidFlower experimental system

presents results obtained with a corner-point grid, which show a better agreement than Figure 4a (Cartesian grid) with the fine-scale simulations depicted in Figures 4c and 4d. The difference between these two fine-scale cases lies in the solver tolerances, with 4d having stricter tolerances (refer to the online documentation of the `pyopmspe11` tool (Landa-Marban and Sandve [2025]) for the values of these tolerances). Tightening the solver tolerances improves mass conservation, where in this system the theoretical injected mass is 4.59×10^{-3} kg, while the injected mass in Figure 4c is 4.43×10^{-3} kg (3.5% error), and in Figure 4d is 4.58×10^{-3} kg (0.2% error). However, this improvement comes at the cost of increased simulation time, rising from approximately 17 days in Figure 4c to around 55 days in Figure 4d.

Overall, the results presented in Figure 4 demonstrate that simulation outcomes are highly sensitive to grid type and resolution, as well as to solver tolerances. While high-resolution grids combined with strict solver settings are desirable for accuracy, they are impractical for HM studies, which typically require numerous simulation runs. Therefore, for our HM analysis, we adopt the coarser corner-point grid shown in Figure 3, with a resolution of approximately 2 cm (2 cm in the x-direction and around 2 cm in the z direction). This grid is sufficiently coarse to avoid mass conservation issues when using default solver tolerances. The choice of this grid is motivated by three factors: it adequately captures the heterogeneity of the fault zone, it offers acceptable simulation times (approximately 8 minutes per run in serial, 2 minutes in parallel using eight cores), and it

aligns with the resolution used in the FluidFlower benchmark study for computing the Wasserstein distance, which was based on a 2 cm Cartesian grid. Note that since the simulation results are sensitive to grid size, the history-matched parameters will be grid-dependent and cannot be directly applied to other grids and configurations.

2.4 History matching

Data assimilation refers to the process of integrating measured data into a model to enhance its predictive accuracy. HM is a specific form of data assimilation in which reservoir parameters are adjusted to align simulation outputs with observed data. This approach typically involves generating an ensemble of reservoir models by sampling from distributions of uncertain parameters. The ensemble framework helps quantify and reduce uncertainty in both model parameters and measured data. Widely used algorithms for HM include the Ensemble Kalman Filter (Evensen [2003]) and the Ensemble Smoother (Emerick and Reynolds [2013]), which update model states and parameters based on observed measurements while preserving computational efficiency. Refer to Tian et al. [2024] for an ensemble-base HM for the FluidFlower system.

An alternative approach to Ensemble Kalman Filter and Ensemble Smoother that has gained increasing attention is the differential evolution method, a population-based optimization algorithm (Storn and Price [1997]). The differential evolution method is designed to iteratively improve a population of candidate solutions and is particularly well suited for complex optimization problems. It performs effectively in scenarios involving non-differentiable, non-linear, and multimodal objective functions, making it a robust choice for reservoir HM where the solution space is often irregular and high-dimensional. For our HM study, we use the differential evolution algorithm available via the SciPy library (Virtanen et al. [2020]) and is integrated using the Everest decision-making tool.

The HM algorithm requires a metric to quantify the mismatch between observed data and simulation results and guide the iterative improvement of the tuning parameters. For this purpose, we utilize the Wasserstein distance, which also built the foundation for the dense data comparison of the benchmark (Flemisch et al. [2024]); we leverage the same Python implementation of its computation provided in the associated data repositories, also used in Saló-Salgado et al. [2024]. The comparison is performed on spatial maps at 24, 48, 72, 96, and 120 hours, matching the time points reported in Flemisch et al. [2024]. The total cost, provided to the HM algorithm consists of the sum of the five distinct Wasserstein distance values.

2.5 Integration

In this subsection, we describe the followed steps to conduct the HM. In order to assess the efficacy, we evaluate the distance to the experimental benchmark data and compare it against the other submissions to the model validation benchmark (Flemisch et al. [2024]).

As previously explained, the FluidFlower system presents significant modeling challenges due to its high sensitivity to input parameters, grid configuration, and simulation settings. In this study, the parameters considered for HM include permeability, entry pressure, and residual saturations for each of the six sand layers, resulting in a total of 24 parameters. The temperature is set to 20°C, and the pressure on the top boundary (at 1.2 m) is set to 104,900 Pa to match the pressure in the sensors. Mechanical dispersion is neglected to reduce the parameter space, which is a justified model simplification due to the coarse grid resolution. In addition, porosity values are not included in the HM process; instead, they are set equal to those used in Model MIT_M1 from [Saló-Salgado et al. \[2024\]](#), which are selected from published data in similar silica sands. This decision is based on the observation that simulation results are extremely sensitive to porosity variations, which posed difficulties for the considered HM algorithm during preliminary testing.

To illustrate the scale of the HM problem, consider a simplified scenario where each of the 24 parameters (permeability, entry pressure, and residual saturations in the six sands) is sampled at two values (minimum and maximum), resulting in $2^{24} = 16,777,216$ possible combinations. Expanding this to three values per parameter (e.g., minimum, middle, and maximum) increases the number of combinations to over 10^{11} . In practice, the parameter space is even larger, e.g., this study considers 50 possible values per permeability interval, 20 for entry pressures, and between 15 and 4 for residual saturations (see Table 2), resulting in 6×10^{28} possible combinations. Given this vast search space, the success of HM depends not only on the choice of algorithm but also on additional factors such as the initial parameter values and the selected random seed, which introduces an element of stochasticity and, to some extent, luck ([Rescher \[2021\]](#)).

Table 2 summarizes the initial parameter range and number of samples. These parameter ranges and number of samples are set based on our preliminary testing. To conduct the HM, we adopt a sequential approach commonly used in the oil and gas industry, where different groups of parameters are adjusted in successive steps ([Yin et al. \[2011\]](#)). In the first iteration, the parameters subject to HM are the permeabilities, residual saturations, and entry pressures for sands two through five. This choice is motivated by the fact that these formations constitute the primary storage sands, and simulation outcomes are highly sensitive to variations in these parameters. In the second iteration, the same set of parameters is adjusted, but only for sands one and six, which were not included in the first iteration. Since $s_{r,g}$ for sand five was 0.05 after the first iteration, then the value for $s_{r,g}$ for sand six was set to 0 and not HM. In the final iteration, a second adjustment of permeability, residual saturations, and entry pressure for sand five is performed. The rationale for revisiting sand five in the last step is that since CO₂ is injected into the lower unit, which consists primarily of sand five, its properties play a critical role in controlling upward leakage and density driven fingers. In addition, a constraint is imposed throughout the HM workflow to ensure monotonicity in the parameter values, e.g., permeability must increase with increasing grain size. The configuration files containing all necessary details to reproduce this study are available

Table 2: Initial model parameters [min, max, # equidistance samples]

Id	Sand	k [D]	$s_{r,w}$ [-]	$s_{r,g}$ [-]	p_e [Pa]
1	ESF	[50, 150, 50]	[0.32, 0.35, 4]	[0.20, 0.35, 15]	[1000, 2000, 20]
2	C	[50, 200, 50]	[0.05, 0.35, 10]	[0.05, 0.35, 10]	[500, 1000, 20]
3	D	[100, 500, 50]	[0.05, 0.32, 10]	[0.05, 0.32, 10]	[150, 600, 20]
4	E	[300, 1300, 50]	[0.05, 0.3, 10]	[0.05, 0.30, 10]	[100, 400, 20]
5	F	[1300, 3000, 50]	[0.05, 0.28, 10]	[0.05, 0.28, 10]	[50, 200, 20]
6	G	[2500, 5000, 50]	[0.05, 0.26, 10]	⁻¹	[0, 170, 20]

¹ Set to 0.

at <https://github.com/cssr-tools/pofff/paper>. The configuration file used for the sequential HM approach (first iteration) is provided in Appendix A. Simulations were executed on a local server equipped with 144 CPUs, and approximately 600,000 runs were performed.

3 Results and Discussion

In this section, we present the results of the image-based HM framework, when applied to the FluidFlower benchmark data to tune various uncertain material parameters as elaborated in the previous section. Table 3 summarizes the resulting values of the HM parameters.

Table 3: Final model parameters after the HM

Id	Sand	k [D]	ϕ [-] ¹	$s_{r,w}$ [-]	$s_{r,g}$ [-]	p_e [Pa]
1	ESF	62	0.37	0.34	0.25	1900
2	C	152	0.38	0.32	0.20	950
3	D	428	0.40	0.29	0.08	185
4	E	1120	0.39	0.27	0.06	175
5	F	2014	0.39	0.26	0.01	170
6	G	2500	0.42	0.17	0	163

¹ Porosity values from model M1 in Saló-Salgado et al. [2024].

Although the HM values are generally of the same order of magnitude as the measured values in Table 1 (used by CSIRO) and the ones for MIT_M1 (see Saló-Salgado et al. [2024]), some differences are observed. Notably, the entry pressure values obtained from the HM are consistently higher. This is expected, as coarse grids in numerical models lead to overestimation of entry pressure, since simulations in coarse grids underestimate CO₂ dissolution rates (Suriano et al. [2022]), while higher entry pressures leads to higher CO₂ dissolution rates (Martinez and Hesse [2016]). In addition, the CSIRO simulations used a constant thickness of 25 mm and the MIT_M1 simulations employed a different thickness map derived from initial experimental data (see Fig. 2c in Saló-Salgado et al.

[2024]), whereas our simulations used the final measured thickness map (Fig. 3b). The relative permeability and capillary pressure models also differ between the CSIRO (see Eqs. 8-11 in Green et al. [2024]), MIT_M1 (see Subsection 3.2.1 in Saló-Salgado et al. [2024]), and our simulations (Eq. 4). The differences between thickness maps and saturation function models, combined with the use of distinct grid types and resolutions across the groups, contribute to the variation in parameter values.

Figure 5 presents spatial maps comparing the simulation results, using the history-matched parameter set, and the experimental data. Figures 5a-f show good agreement between the simulation results and the experimental data. However, the segmented simulation maps in Figures 5g-i exclude certain CO₂ regions, particularly in the upper-left area, resulting in the simulated CO₂ reaching the upper-left boundary. This discrepancy is related to the segmentation thresholds used in the benchmark study, where a gas saturation threshold of 0.01 and a CO₂ concentration threshold of 0.1 kg/m³ were applied. These thresholds introduce additional sensitivity into the analysis, as also discussed by Jammoul et al. [2024]. In Appendix B, simulation results using a lower threshold of 0.05 for the CO₂ concentration are shown. To mitigate the impact of threshold selection, ongoing research is focused on improving the image processing workflow utilizing regression instead of segmentation (Folkvord et al. [2025]). The goal is to generate continuous maps of CO₂ mass directly from experimental images, enabling more robust comparisons with simulation results.

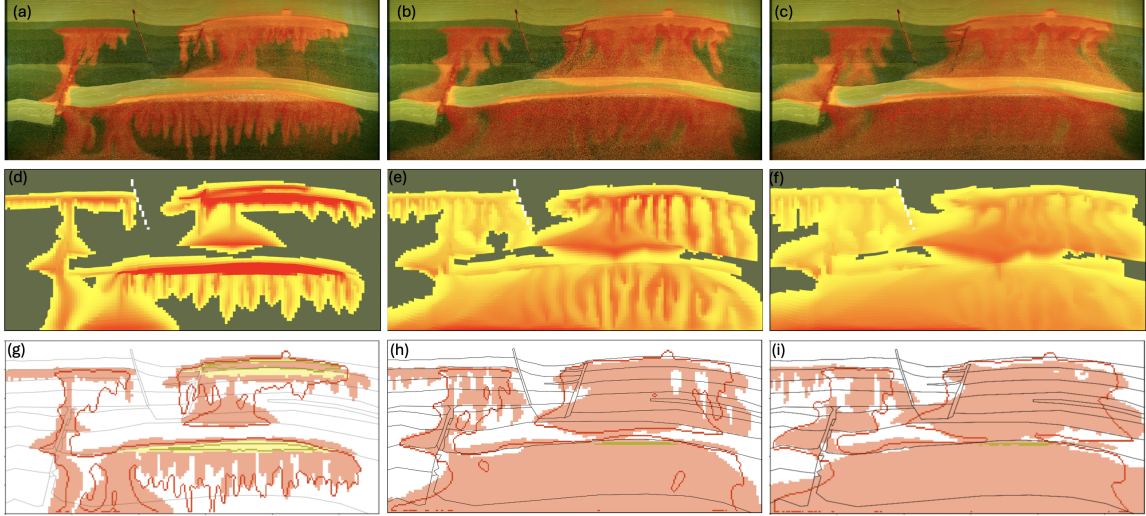


Figure 5: (a-c) Photographs from the FluidFlower experiment C2 at one, three, and five days, respectively. (d-e) Spatial maps of CO₂ concentration from the corresponding simulation results. (f-g) Contour plots comparing experimental observations and segmented simulation outputs at the same time intervals

To further evaluate the quality of the HM simulation results and to place them in context with

the FluidFlower benchmark figures (Flemisch et al. [2024]), we compute the Wasserstein distance between all experimental datasets (experiments C1 through C5) and the corresponding simulation outputs. As the HM workflow explicitly aims at minimizing the Wasserstein distance, it is expected that the HM yields to the overall best agreement in this metric. In addition to dense data comparisons, the benchmark study also included time-series and sparse data corresponding to effective quantities in pre-defined boxes (Nordbotten et al. [2022], Flemisch et al. [2024]). As these were not included in the HM workflow, the comparison of the time-series and sparse data is of high relevance. For comparison, we highlight the individual result from CSIRO, being the participant with the closest results to the experimental data, and include the result MIT_M1 from Saló-Salgado et al. [2024], where model parameters were set from published data based on the measurements of the average grain sizes, and the capillary pressure curve for sand D was manually HM. Our HM result is labeled CSSR, representing the research center that supported this study (see Acknowledgments).

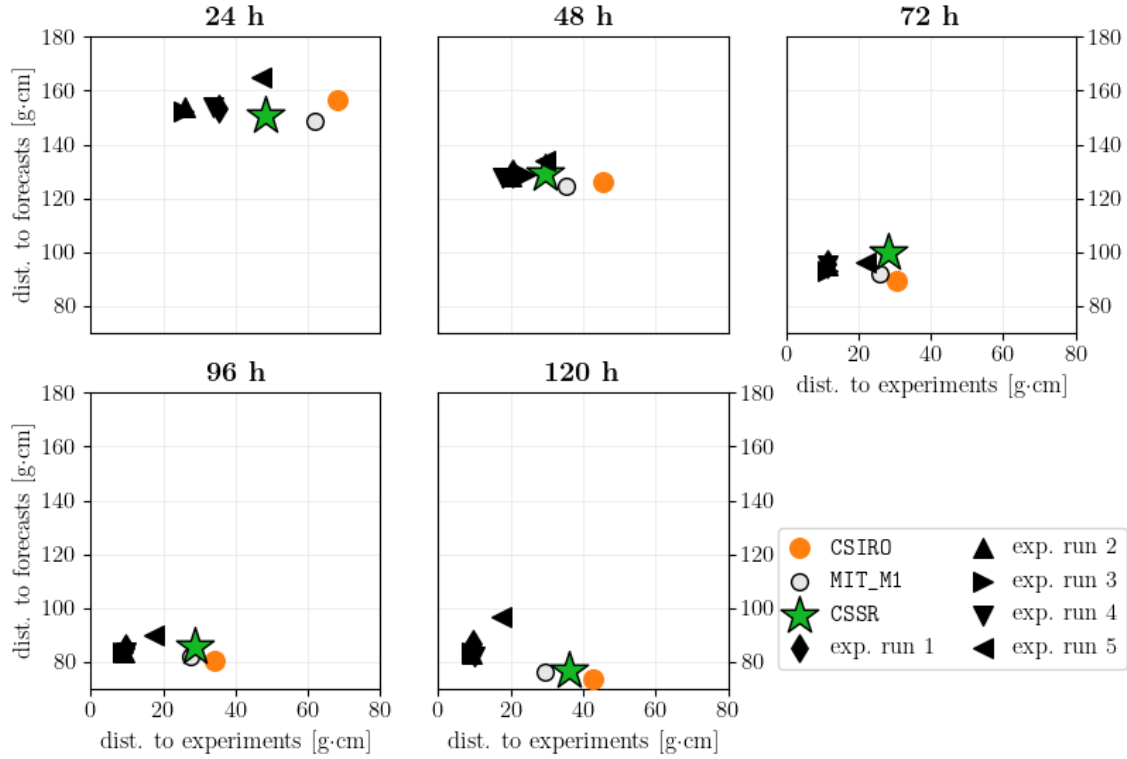


Figure 6: Wasserstein distances computed between simulation results and experimental data (experiments C1-C5), as well as between simulation forecasts. Forecast comparisons include only results submitted by participants in the FluidFlower benchmark study

Figure 6 displays the spreading of the Wasserstein distance and shows that all experimental

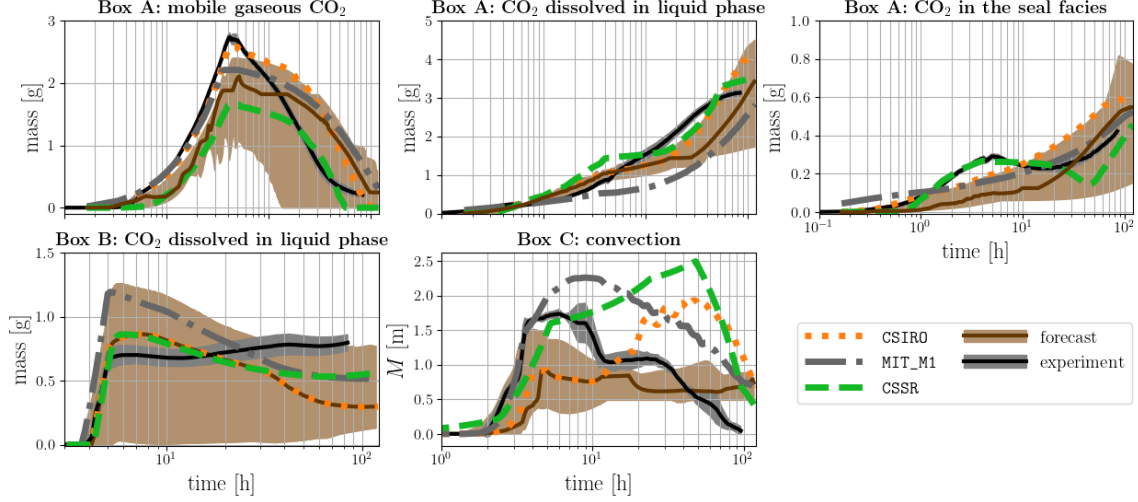


Figure 7: Comparison between simulation forecasts and experimental observations for the temporal evolution of box quantities. The brown line represents the median of the simulation results submitted by benchmark participants, with the shaded pale brown region indicating the interquartile range (from the first to the third quartile). The black line shows the mean of the experimental measurements, and the surrounding gray region reflects the variability expressed as one standard deviation

results yield Wasserstein distances below 50 g·cm, with the closest distance being approximately 10 g·cm. The results from CSIRO demonstrate excellent agreement with the experimental data under this metric, especially considering that benchmark participants did not have access to the CO₂ experimental data prior to submission. The calibrated model from [Saló-Salgado et al. \[2024\]](#) (MIT_M1) also performs well, supporting the available published data for the different grain sizes and the effectiveness of manual parameter tuning. As expected, our history-matched result (CSSR) shows similarly close agreement with the experimental data, highlighting the quality of the HM approach used in this study.

The time-series data is displayed in Figure 7. For the temporal evolution of these box quantities, the results from CSIRO, MIT_M1, and CSSR follow similar trends across all metrics, with the exception of box c, which captures convective behavior, defined as the integral of the magnitude of the gradient in relative concentration of dissolved CO₂. This quantity is particularly challenging to capture, as previously discussed in [Flemisch et al. \[2024\]](#).

Finally, Figure 8 displays the comparison of the sparse data. It shows that the closest match to the experimental mean varies depending on the specific sparse data quantity. For example, CSSR performs best for quantity 2, CSIRO for quantity 3a, and MIT_M1 for quantity 5. To quantify these comparisons, we compute the error with respect to the experimental data for each sparse quantity

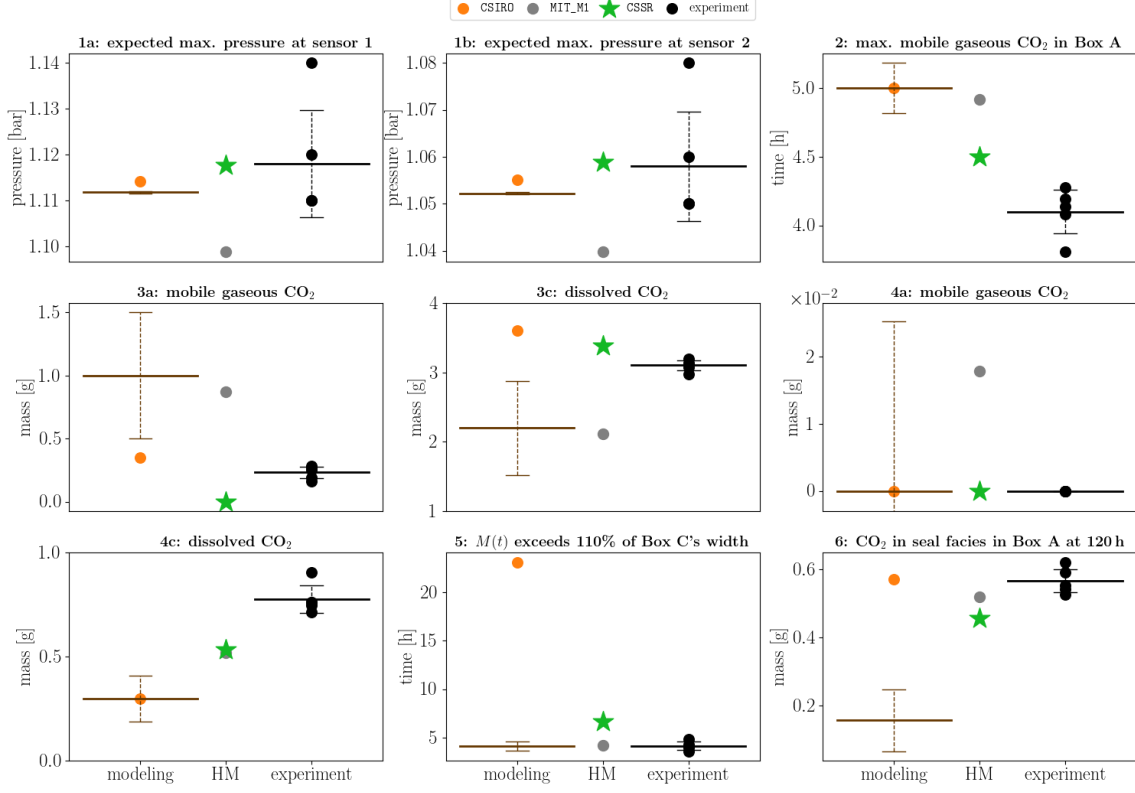


Figure 8: Comparison of sparse data reported by benchmark participants with experimental observations. For the simulation results, the horizontal brown line represents the median, while the dashed vertical brown line indicates $\pm$ the median of all reported P50 values for the standard deviation. The middle x-axis (HM) displays the results from [Saló-Salgado et al. \[2024\]](#) (MIT_M1) and from this study (CSSR). For the experimental data, black circles represent individual measurements from each run, the horizontal black line shows the mean, and the dashed vertical black line denotes $\pm$ one standard deviation

across all datasets as follows:

$$\text{error}_g = \frac{100}{6} \sum_{i \in \{2, 3a, 3c, 3d, 4c, 6\}} \left| \frac{g_i - \text{exp}_i}{\text{exp}_i} \right|, \quad g \in \{\text{CSIRO}, \text{MIT_M1}, \text{CSSR}\} \quad (6)$$

where the quantity 2 is the time of max CO₂ mobile free phase in box a [s], 3a the CO₂ mobile free phase in box a at 72h [g], 3c the CO₂ dissolved in water in box a at 72h [g], 3d the CO₂ in the seal in box a at 72h [g], 4c the CO₂ dissolved in water in box b at 72h [g], and 6 the total mass of CO₂ in the top seal facies [g]. The results are scaled by a factor of 100/6 to bring the values into the same order of magnitude as the Wasserstein distance, see Table 4. Note that pressure values

(quantities 1a and 1b) are excluded from this analysis, as our simulations impose a top boundary condition that matches the experimental mean. Additionally, box C (quantity 5) is omitted due to the modeling challenges previously discussed in [Flemisch et al. \[2024\]](#).

Table 4: Sparse data and Wasserstein distance (WD) comparison between experimental and simulation results

Data source	# cells	2 (s)	3a (g)	3c (g)	3d (g)	4c (g)	6 (g)	error _g ¹	WD (g·cm)
Experiment	-	14800	0.23	3.10	0.38	0.78	0.57	-	18.32
MIT_M1	151,402	17700	0.87	2.11	0.43	0.52	0.52	63.60	35.98
CSIRO	44,284	18000	0.35	3.60	0.57	0.30	0.57	33.38	44.21
CSSR	9,627	16200	0	3.38	0.28	0.53	0.46	32.85	34.06

¹ See Eq. 6.

From Table 4, we observe that the results obtained using the HM presented in this paper, CSSR, yield the lowest errors for both the sparse data and the Wasserstein distance metrics. It is noteworthy that the HM resulted in a 0 value for 3a (CO₂ mobile free phase in box a at 72h [g]), compared to the experimental value of 0.23. While this still deviates from the observed data, it represents a closer match than the MIT_M1 prediction of 0.87. Table 4 also shows the number of cells for each of the three simulation studies, with our grid having the less number of cells. The number of cells (grid size) for the MIT_M1 was selected after doing a grid refinement study, where the grid size of 5 mm was needed to get accurate results ([Saló-Salgado et al. \[2024\]](#)). Additionally, the simulation time (in serial) for a single run using our coarse grid is approximately 8 minutes, which can be even less than two minutes running the simulator in parallel (the simulation times for CSIRO and MIT_M1 are not available). This observation brings us back to the broader discussion on the trade-off between accuracy and computational efficiency, a balance that remains difficult to quantify, particularly when considering how to weight the errors in Table 4 against the corresponding run times/number of cells.

4 Conclusion and Outlook

In this work, we presented an image-based history matching workflow applied to the International FluidFlower benchmark data. In summary, the exercise enabled additional validation of the modeling of CO₂ at laboratory conditions.

HM relies on the ability to perform hundreds of simulations efficiently and reliably. Simulating the CO₂ injection experiment under room conditions using Darcy-scale permeability (as opposed to the more common milliDarcy range in reservoir applications) presents challenges for traditional reservoir simulators. Achieving sufficiently accurate solutions requires small time steps, which significantly increases simulation time when using fine grids on the order of millimeters. This, in

turn, limits both the total number of simulations and the number of iterations feasible within the HM workflow. To address this, we designed and tuned a coarse corner-point grid with a cell size of approximately 2 cm. This adjustment reduced the simulation time to around 8 minutes per run (2 minutes in parallel using eight cores), enabling the execution of hundreds of thousands of simulations on local servers within a few weeks. For this study, we utilized a local server equipped with 144 CPUs. After the HM process, the final simulation results show good agreement with the experimental CO₂ injection data. It would be desirable to further explore the predictive capability of the updated effective model against forecast data. Unfortunately, this validation is not possible given the lack of additional experimental scenarios in the original benchmark study. Nonetheless, the successful results of the present study reinforces the findings presented in the evaluation of the FluidFlower benchmark (Flemisch et al. [2024]), which suggest that the system can be effectively described using standard flow equations, conventional saturation functions, and typical PVT properties for CO₂-brine mixtures.

The developed Python package for OPM Flow and FluidFlower, `pofff`, enables full reproducibility of the results presented in this study. Moreover, the input configuration file provides a flexible framework for exploring alternative approaches to HM using FluidFlower data. This includes testing different parameter distributions, applying various strategies for parameter splitting across HM steps, adjusting simulation settings to reduce computational time, and evaluating alternative models for saturation functions. It is important to note that simulation results are highly sensitive to changes in grid resolution, model formulations (such as the exponents in the relative permeability and capillary pressure relationships), and threshold values, including tolerances used for simulation segmentation. The history-matched parameters are thus not directly applicable to other setups. Given the vast space of possible model configurations (e.g., grid type and size, saturation models, and parameter values), `pofff` includes functionality for qualitative comparison of simulation outputs against both experimental data and the final results presented in this study (CSSR, see Table 4). Users are encouraged to contribute their configuration files to a dedicated folder within `pofff`, facilitating collaborative efforts toward improved matches and model refinement.

Author Contributions All authors contributed to the study conception and design. Code implementation, numerical simulations, and analysis were performed by David Landa-Marban, Tor H. Sandve, and Jakub W. Both. The first draft of the manuscript was written by David Landa-Marban. All authors read and approved the final manuscript.

Data Availability The simulator OPM Flow can be obtained at <https://opm-project.org>. Link to complete codes for the numerical studies are available in <https://github.com/cssr-tools/pofff>. The optimization tool Everest is available at <https://github.com/equinor/ert>. The Python image analysis toolbox DarSIA used for obtaining the segmentation and continuous CO₂

data can be accessed via <https://github.com/pmgbergen/DarSIA>; corresponding runscripts analyzing the FluidFlower dataset are available at https://github.com/pmgbergen/fluidflower_benchmark_analysis. The plotting tool for the simulation results is available at <https://github.com/cssr-tools/plopm>.

Funding The authors acknowledge funding from Centre of Sustainable Subsurface Resources (CSSR), grant nr. 331841, supported by the Research Council of Norway, research partners NORCE Research AS and the University of Bergen, and user partners Equinor ASA, Harbour Energy, Sumitomo Corporation, Earth Science Analytics, GCE Ocean Technology, and SLB Scandinavia. JWB was in part supported by the Research Council of Norway through the project Unlocking maximal geological CO₂ storage through experimentally validated mathematical modeling of dissolution and convective mixing (project nr. 355188).

Appendix A Configuration file format

Configuration files ease the setting and reproducibility of simulation studies. In `pofff`, the widely-used TOML format is adopted, see Figure 9 for the initial configuration file run in this study.

In Figure 9 comments are added to describe the different available settings, and we refer the reader to the `pofff` online documentation for an extended description. We remark that this approach (using configuration files to set up numerical studies) easily allows for further extension of the work.

Appendix B Lower threshold for CO₂ concentration

To illustrate the sensitivity of the segmentation threshold in the numerical simulations of CO₂ concentration, Figures 10 and 11 present the spatial distribution maps and corresponding Wasserstein distances obtained using a threshold of 0.05 kg/m³, in contrast to the benchmark value of 0.1 kg/m³. See Table 5 for the computed errors for the sparse data and Wasserstein distances, and Table 6 for the simulation parameters used to generate these results.

Table 5: Sparse data and Wasserstein distance (WD) comparison between experimental and simulation results (threshold of 0.05 kg/m³ for the CO₂ concentration)

Data source	# cells	2 (s)	3a (g)	3c (g)	3d (g)	4c (g)	6 (g)	error _g	WD (g·cm)
Experiment	-	14800	0.23	3.10	0.38	0.78	0.57	-	18.32
CSIRO	151,402	18000	0.35	3.60	0.57	0.30	0.57	33.38	39.52
MIT_M1	44,284	17700	0.87	2.11	0.43	0.52	0.52	63.60	36.31
CSSR	9,627	15600	0	3.40	0.31	0.44	0.50	31.61	28.32

```

1 # Set the full path to the flow executable and flags
2 flow="flow --enable-opm-rst-file=true --enable-tuning=true --linear-solver=cpr.trueimpes"
3
4 # Set the model parameters
5 grid="corner-point" # Type of grid (cartesian, tensor, or corner-point (cp))
6 thickness="final" # Thickness maps (measured 'initial', 'final', or a real positive value)
7 mult.thickness=1 # Thickness multiplier (a real positive value)
8 x=[140] # If cartesian, number of x cells; otherwise, variable array of x-refinement
9 # cartesian, number of z cells; tensor, variable array of refinement; cp, fix array of refinement (18 entries)
10 z=[7,5,5,5,5,5,5,8,10,9,5]
11 temperature=[20, 20] # Temperature bottom and top rig [C]
12 pressure=104900 # Pressure at the datum [Pa]
13 diffusion=[1e-9, 2e-8] # Diffusion (in liquid and gas) [m^2/s]
14 sources=[[0.9, 0.005, 0.3], [1.7, 0.005, 0.7]] # Source positions: x, y, and z coordinates [m], source 1 to 2
15
16 # Set the saturation functions
17 krw="(max(0, (sw - swi) / (1 - swi))) ** nkrw" # Wetting rel perm saturation function [-]
18 krn="(max(0, (1 - sw - sni) / (1 - sni))) ** nkrn" # Non-wetting rel perm saturation function [-]
19 cap="pen * ((sw-swi) / (1-swi)) ** ((-1.0 / npen))" # Capillary pressure saturation function [Pa]
20
21 # Facie properties (ESF, C, D, E, F, G -> facie1, facie2, ..., facie6)
22 facie1={"PERMX1}=50E3, "PERMZ1}=50E3, "PORO1}=0.37, "DISPERC1}=0, "SWI1}=0.32, "SNI1}=0.1, "PEN1}=1500, "NKRW1}=2, "NKRN1}=2, "NPE1}=2, "
    THRE1}=5e-2, "NPNT1}=100}
23 facie2={"PERMX2}=100E3, "PERMZ2}=100E3, "PORO2}=0.38, "DISPERC2}=0, "SWI2}=0.14, "SNI2}=0.1, "PEN2}=800, "NKRW2}=2, "NKRN2}=2, "NPE2}=2, "
    THRE2}=5e-2, "NPNT2}=100}
24 facie3={"PERMX3}=300E3, "PERMZ3}=300E3, "PORO3}=0.40, "DISPERC3}=0, "SWI3}=0.12, "SNI3}=0.1, "PEN3}=200, "NKRW3}=2, "NKRN3}=2, "NPE3}=2, "
    THRE3}=5e-2, "NPNT3}=100}
25 facie4={"PERMX4}=800E3, "PERMZ4}=800E3, "PORO4}=0.39, "DISPERC4}=0, "SWI4}=0.12, "SNI4}=0.1, "PEN4}=150, "NKRW4}=2, "NKRN4}=2, "NPE4}=2, "
    THRE4}=5e-2, "NPNT4}=100}
26 facie5={"PERMX5}=1500E3, "PERMZ5}=1500E3, "PORO5}=0.39, "DISPERC5}=0, "SWI5}=0.12, "SNI5}=0.1, "PEN5}=100, "NKRW5}=2, "NKRN5}=2, "NPE5}=2, "
    THRE5}=5e-2, "NPNT5}=100}
27 facie6={"PERMX6}=3000E3, "PERMZ6}=3000E3, "PORO6}=0.42, "DISPERC6}=0, "SWI6}=0.1, "SNI6}=0.1, "PEN6}=1, "NKRW6}=2, "NKRN6}=2, "NPE6}=2, "
    THRE6}=5e-2, "NPNT6}=100}
28
29 # Schedule: 1) injection time [s], 2) time step size to write results [s], 3) injection rate [kg/s] (source1),
30 # 4) injection rate [kg/s] (source2). If --enable-tuning=true, then the TUNING values [days]
31 inj=[
32   [ 8100, 8100, 3E-7, 0, '1e-2 3e-4 1e-20 1e-20 1.6 0.2 0.65 1.1'],
33   [10200, 10200, 3E-7, 3E-7, '1e-2 1e-4 1e-20 1e-20 1.6 0.2 0.65 1.1'],
34   [ 68100, 68100, 0, 0, '1e-2 1e-3 1e-20 1e-20 1.6 0.2 0.65 1.1'],
35   [345600, 86400, 0, 0, '1e-2 1e-2 1e-20 1e-20 1.6 0.2 0.65 1.1']]
36
37 # Everest
38 min.realizations.success=0 # Minimum number of simulations to be regarded as a success.
39 max.function.evaluations=200000 # Maximum number of simulations
40 random.seed=7 # Set a specific seed for reproducibility; a value of 0 means no seed
41 cores=50 # Maximum number of simulations running in parallel
42 maxtime=3600 # Maximum runtime in seconds of a realization; a value of 0 means unlimited runtime
43 delete=true # Delete large files?
44 monotonic=false # Only consider monotonic values, e.g. increasing entry pressure with decreasing sand size
45 popsize=200 # Population size
46 PERM2=[ 100E3, 50E3, 200E3, 50] # Initial value, min, max, and size of interval
47 PERM3=[ 300E3, 100E3, 500E3, 50]
48 PERM4=[ 800E3, 300E3, 1300E3, 50]
49 PERM5=[1500E3, 800E3, 3000E3, 50]
50 SWI2=[ 0.14, 0.05, 0.35, 10]
51 SWI3=[ 0.12, 0.05, 0.32, 10]
52 SWI4=[ 0.12, 0.05, 0.30, 10]
53 SWI5=[ 0.12, 0.05, 0.28, 10]
54 SNI2=[ 0.10, 0.05, 0.35, 10]
55 SNI3=[ 0.10, 0.05, 0.32, 10]
56 SNI4=[ 0.10, 0.05, 0.30, 10]
57 SNI5=[ 0.10, 0.05, 0.28, 10]
58 PEN2=[ 800, 500, 1000, 20]
59 PEN3=[ 200, 150, 600, 20]
60 PEN4=[ 150, 100, 400, 20]
61 PEN5=[ 100, 50, 200, 20]

```

Figure 9: Example of configuration file for **pofff**. This file can be run from the terminal as: **pofff -i input.file.toml -o output.folder -m everest -t 24,48,72,96,120**, where **-t** sets the times in days for the experimental images to HM

Appendix C Computational details

This appendix provides a detailed description of the numerical study setup. The configuration files, scripts, and supplementary technical information required to reproduce the results are available in the **pofff** repository and its online documentation. The procedures outlined here are based on the Ubuntu local server configuration employed in our work. For users operating on macOS or Windows (via the Windows Subsystem for Linux), supplementary instructions are available in the official **pofff** online documentation to facilitate replication of the simulations across different platforms.

The simulations were executed on an Intel® Xeon® Platinum 8354H processors, featuring a

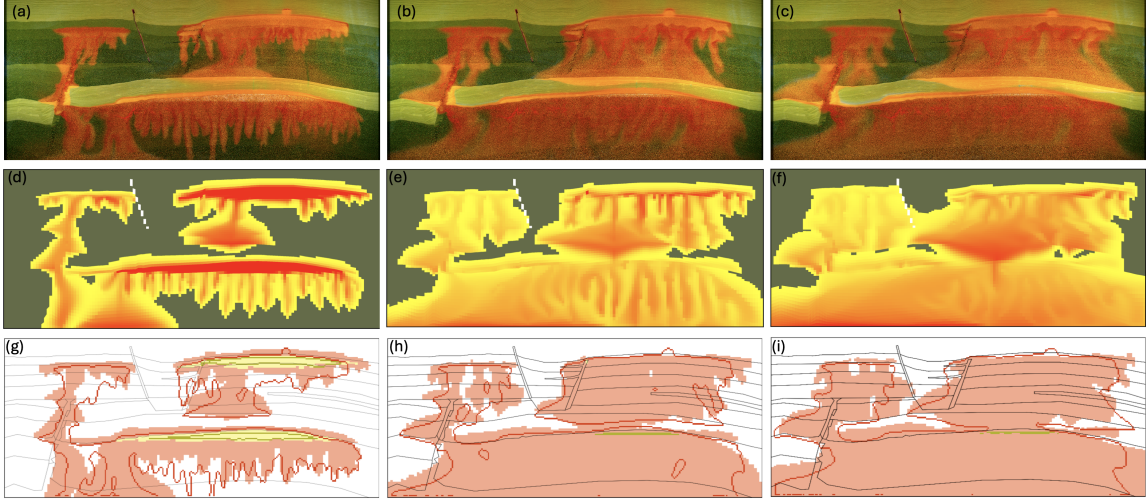


Figure 10: (a-c) Photographs from the FluidFlow experiment C2 at one, three, and five days, respectively. (d-e) Spatial maps of CO_2 concentration from the corresponding simulation results. (f-g) Contour plots comparing experimental observations and segmented simulation outputs (threshold of 0.05 kg/m^3 for the CO_2 concentration) at the same time intervals

Table 6: Model parameters

Id	Sand	k [D]	ϕ [-]	$s_{r,w}$ [-]	$s_{r,g}$ [-]	p_e [Pa]
1	ESF	62	0.37	0.34	0.25	1900
2	C	152	0.38	0.32	0.20	600
3	D	690	0.40	0.30	0.19	305
4	E	720	0.39	0.28	0.06	175
5	F	2000	0.39	0.25	0.01	170
6	G	2500	0.42	0.17	0	55

3.1 GHz base clock, x86_64 architecture, and a total of 144 logical CPUs (4 sockets $\times$ 18 cores per socket, with 2 threads per core). The system is equipped with 99 MB of L3 cache, 1583 GB RAM, and an HDD-based storage solution. It operated on Ubuntu 24.04.3 LTS (Noble), with Open MPI version 4.1.6 installed. Although GPU acceleration was not utilized in the simulations, it is worth noting that OPM Flow provides GPU support. Readers interested in leveraging this capability are encouraged to consult the OPM Flow manual ([Goodfield et al. \[2025\]](#)) for further details.

The numerical experiments were conducted using OPM Flow version 2025.10. The simulations employed the default Python distribution in Ubuntu 24.04 LTS, corresponding to Python 3.12.3. The `pofff` package was also used in version 2025.10. A complete specification of Python dependencies and their exact versions is provided in the `pyproject.toml` file included with the `pofff` v2025.10 release. For reproducibility, the `pofff` repository contains a continuous integration script (`CI.yml`)

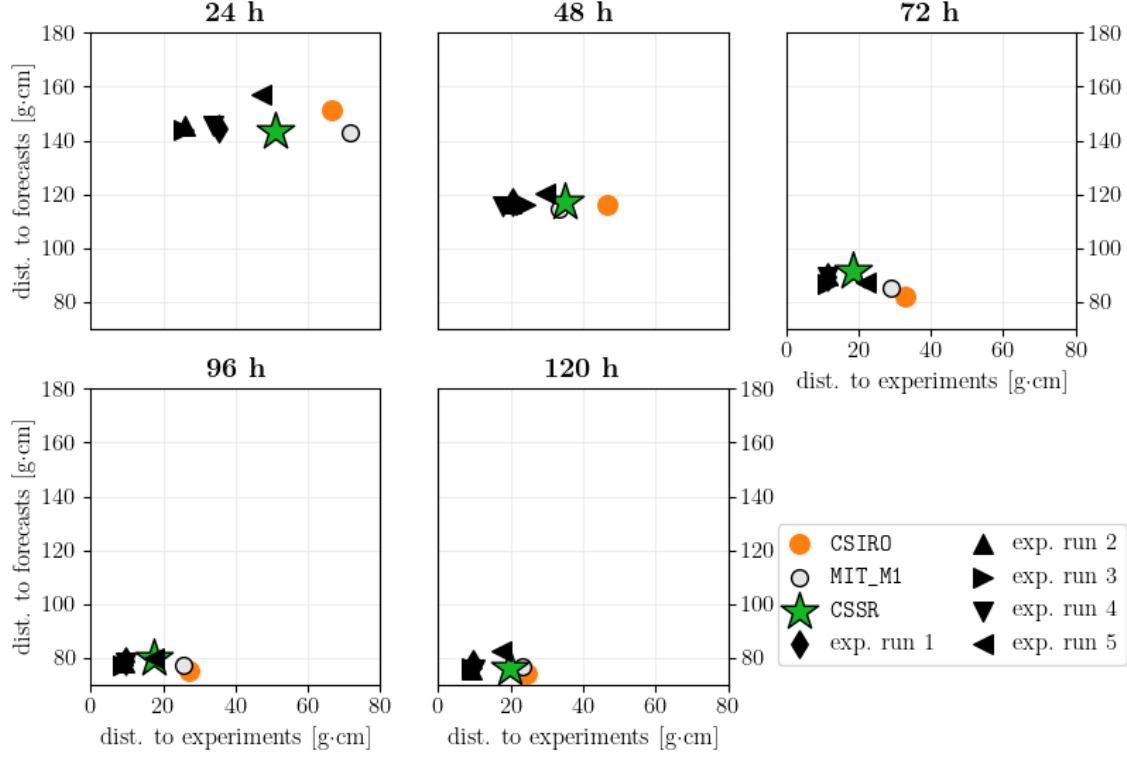


Figure 11: Wasserstein distances computed between simulation results and experimental data (experiments C1-C5), as well as between simulation forecasts (threshold of 0.05 kg/m^3 for the CO_2 concentration)

located in `.github/workflows`. This script is configured to run on the latest available Ubuntu release (at the time of writing, Ubuntu 24.04 LTS) with Python 3.12. Following the steps in this workflow allows users to install the OPM Flow binaries, create a dedicated Python virtual environment, install `pofff` together with the required Python libraries, and execute the test cases to verify that the installation has been successfully completed.

We profile the integrated workflow in three stages: preprocessing, history matching (HM), and postprocessing. The preprocessing stage consists of generating input files for both OPM Flow and the history matching tool. The HM stage involves executing a sequence of tasks, including OPM Flow simulations, for which profiling results are reported separately for the simulations and the remaining HM tasks. The postprocessing stage comprises the generation of figures and tables. For this profiling study, we use the configuration file from the final HM iteration presented in this paper, but restrict the run to 64 simulations to ensure computational feasibility. The case is executed on a single core, with reporting times summarized in Table 7, while scalability with respect to the

number of cores is illustrated in Figure 12.

Table 7: Execution time profiling of the integrated `pofff` workflow

	Preprocessing	History matching		Postprocessing	Total
		OPM Flow	Rest		
Wall time (s)	8.35	27,491.92 ¹	1,928.21	487.14	29,915.63

¹ 432.99±7.60 (median and interquantile range of the 64 OPM Flow runs).

Complementing the sensitivity studies reported in the FluidFlower benchmark special issue (Wang et al. [2024] on hysteresis and molecular diffusion, Wapperom et al. [2024] on discretization techniques and types of grids, Jammoul et al. [2024] on rock and gas relative permeability, and image segmentation threshold), as well as the OPM team’s results for the SPE11 benchmark (Nordbotten et al. [2025], Landa-Marban and Sandve [2025]) on types of grids, grid resolution, and solver tolerances, we investigate sensitivities in the HM related to the random seed and the operating system. The base case employs a random seed of 7, executed on the Ubuntu machine described earlier. This configuration corresponds to the case presented in Table 7 and Figure 12. To assess sensitivity, we vary one factor at a time: the random seed is changed to 11, and the operating system is switched to macOS (Tahoe 26.1, Apple M2 Pro chip). Table 8 reports the Wasserstein distance for each case. These results highlight the sensitivity of the HM outcomes to choices of random seed and operating system.

Table 8: Sensitivity of the history matching to setup choices

	Base case	Random seed	Operating system
Wasserstein distance [g·cm]	32.37	32.71	33.19

Regarding the accuracy-time trade-off analysis, we adopt the CO₂ mass in the system as the evaluation metric. Let CO₂^{inj} denote the total injected CO₂ mass (8.5 grams), and CO₂^{sim} the CO₂ mass at the end of the simulation. The accuracy is then defined as:

$$\text{Accuracy} = 100 \times \left(1 - \frac{|\text{CO}_2^{\text{sim}} - \text{CO}_2^{\text{inj}}|}{\text{CO}_2^{\text{inj}}} \right). \quad (7)$$

Table 9 reports the accuracy metric evaluated under two convergence (CNV) tolerance settings, spanning from the relaxed default values in OPM Flow (`--tolerance-cnv=0.01` and `--tolerance-cnv-relaxed=1`) to tightened tolerances (`--tolerance-cnv=0.001` and `--tolerance-cnv-relaxed=0.001`).

Table 9 indicates that adopting the relaxed CNV tolerance provides a suitable balance for the simulations presented in this study. As described in the online documentation of `pyopmspe11` (Landa-Marban and Sandve [2025]) for the SPE11a case with a 1 mm grid, using the default

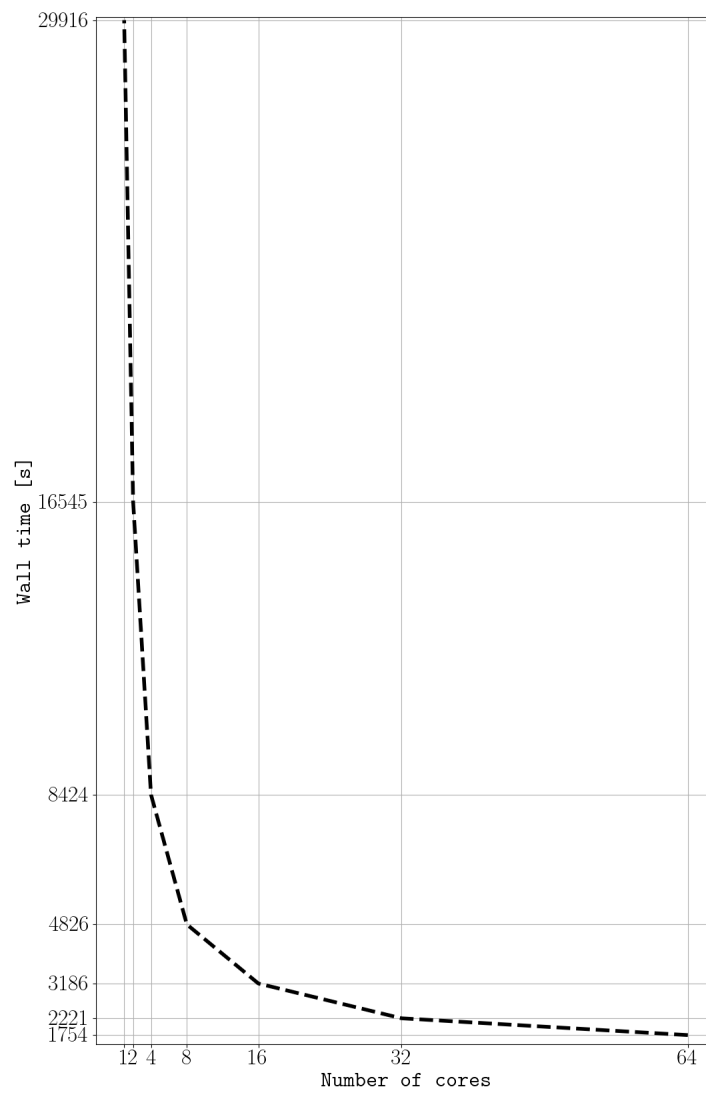


Figure 12: Scaling curve of wall time as a function of number of cores

Table 9: Accuracy-time trade-off for CO₂ mass

Case	OPM Flow parameters	Accuracy (%)	Wall time (s)
Relax ¹	--tolerance-cnv=0.01 --tolerance-cnv-relaxed=1	99.987	417.52
Tight	--tolerance-cnv=0.001 --tolerance-cnv-relaxed=0.001	99.998	570.58

¹ Default values in OPM Flow.

relax tolerances yields a mass accuracy of 96.5% with a simulation time of 17 days. By contrast, enforcing tighter tolerances improves the accuracy to 99.8%, but at the expense of a substantially longer runtime of 55 days.

References

- Global CCS Institute. Global Status of CCS 2021: CCS Accelerating to Net Zero., 2021. URL <https://www.globalccsinstitute.com/publications/global-status-of-ccs-2021/>. Accessed: 2025-11-13.
- A.-K. Furre, R. Meneguolo, P. Ringrose, and S. Kassold. Building confidence in CCS: From Sleipner to the Northern Lights Project. *First Break*, 37(7):81–87, 2019. doi: 10.3997/1365-2397.n0038.
- Atgeirr Flø Rasmussen, Tor Harald Sandve, Kai Bao, Andreas Lauser, Joakim Hove, Bård Skaflestad, Robert Klöfkorn, Markus Blatt, Alf Birger Rustad, Ove Sævareid, Knut-Andreas Lie, and Andreas Thune. The Open Porous Media Flow reservoir simulator. *Computers & Mathematics with Applications*, 81:159–185, 2021. ISSN 0898-1221. doi: 10.1016/j.camwa.2020.05.014. URL <https://www.sciencedirect.com/science/article/pii/S0898122120302182>.
- Knut-Andreas Lie. *An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST)*. Cambridge University Press, Norway, 2019. doi: 10.1017/9781108591416.
- Timo Koch, Dennis Gläser, Kilian Weishaupt, Sina Ackermann, Martin Beck, Beatrix Becker, Samuel Burbulla, Holger Class, Edward Coltman, Simon Emmert, Thomas Fetzner, Christoph Grüninger, Katharina Heck, Johannes Hommel, Theresa Kurz, Melanie Lipp, Farid Mohammadi, Samuel Scherrer, Martin Schneider, Gabriele Seitz, Leopold Stadler, Martin Utz, Felix Weinhardt, and Bernd Flemisch. DuMux 3 – an open-source simulator for solving flow and transport problems in porous media with a focus on model coupling. *Computers & Mathematics with Applications*, 81:423–443, 2021. ISSN 0898-1221. doi: <https://doi.org/10.1016/j.camwa.2020.02.012>. URL <https://www.sciencedirect.com/science/article/pii/S0898122120300791>. Development and Application of Open-source Software for Problems with Numerical PDEs.

- Denis Voskov, Ilshat Saifullin, Michiel Wapperom, Xiaoming Tian, Artur Palha, Luisa Orozco, and Aleks Novikov. open Delft Advanced Research Terra Simulator (open-DARTS), June 2023.
- Hailun Ni, Boxiao Li, Nihal Darraj, Bo Ren, Catrin Harris, Prasanna G. Krishnamurthy, Idris Bukar, Steffen Berg, Jeroen Snippe, Philip Ringrose, T.A. Meckel, Samuel Krevor, and Sally Benson. The impact of capillary heterogeneity on co2 flow and trapping across scales. *Earth-Science Reviews*, 270:105257, 2025. ISSN 0012-8252. doi: <https://doi.org/10.1016/j.earscirev.2025.105257>. URL <https://www.sciencedirect.com/science/article/pii/S0012825225002181>.
- Equinor. Sleipner 2019 Benchmark Model, 2020.
- IEAGHG. IEAGHG Webinar Sleipner Benchmark study. <https://www.youtube.com/watch?v=d2J05cxRpiQ>, 2021. Accessed: 2025-09-19.
- P.S. Ringrose, A.S. Mathieson, I.W. Wright, F. Selama, O. Hansen, R. Bissell, N. Saoula, and J. Midgley. The In Salah CO2 Storage Project: Lessons Learned and Knowledge Transfer. *Energy Procedia*, 37:6226–6236, 2013. ISSN 1876-6102. doi: 10.1016/j.egypro.2013.06.551. URL <https://www.sciencedirect.com/science/article/pii/S1876610213007947>.
- Antonio P. Rinaldi and Jonny Rutqvist. Modeling Ground Surface Uplift During CO2 Sequestration: The Case of in Salah, Algeria. *Energy Procedia*, 114:3247–3256, 2017. ISSN 1876-6102. doi: <https://doi.org/10.1016/j.egypro.2017.03.1456>. URL <https://www.sciencedirect.com/science/article/pii/S1876610217316508>. 13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland.
- Olav Hansen, Douglas Gilding, Bamshad Nazarian, Bård Osdal, Philip Ringrose, Jan-Boye Kristoffersen, Ola Eiken, and Hilde Hansen. Snøhvit: The History of Injecting and Storing 1 Mt CO2 in the Fluvial Tubåen Fm. *Energy Procedia*, 37:3565–3573, 2013. ISSN 1876-6102. doi: <https://doi.org/10.1016/j.egypro.2013.06.249>. URL <https://www.sciencedirect.com/science/article/pii/S187661021300492X>. GHGT-11 Proceedings of the 11th International Conference on Greenhouse Gas Control Technologies, 18-22 November 2012, Kyoto, Japan.
- S. Grude, M. Landrø, and J. Dvorkin. Pressure effects caused by CO2 injection in the Tubåen Fm., the Snøhvit field. *International Journal of Greenhouse Gas Control*, 27:178–187, 2014. ISSN 1750-5836. doi: <https://doi.org/10.1016/j.ijggc.2014.05.013>. URL <https://www.sciencedirect.com/science/article/pii/S1750583614001522>.
- Jonathan Ennis-King and Lincoln Paterson. Role of convective mixing in the long-term storage of carbon dioxide in deep saline formations. In *SPE annual technical conference and exhibition*. OnePetro, 2003. doi: <https://doi.org/10.2118/84344-MS>.

- Titly Farhana Faisal, Sylvie Chevalier, Yves Bernabe, Ruben Juanes, and Mohamed Sassi. Quantitative and qualitative study of density driven CO₂ mass transfer in a vertical Hele-Shaw cell. *International Journal of Heat and Mass Transfer*, 81:901–914, 2015. doi: <https://doi.org/10.1016/j.ijheatmasstransfer.2014.11.017>.
- Widuramina Amarasinghe, Ingebret Fjelde, Jan-Åge Rydland, and Ying Guo. Effects of permeability on CO₂ dissolution and convection at reservoir temperature and pressure conditions: A visualization study. *International Journal of Greenhouse Gas Control*, 99:103082, 2020. doi: <https://doi.org/10.1016/j.ijggc.2020.103082>.
- Elif Agartan, Tissa H. Illangasekare, Javier Vargas-Johnson, Abdullah Cihan, and Jens Birkholzer. Experimental investigation of assessment of the contribution of heterogeneous semi-confining shale layers on mixing and trapping of dissolved CO₂ in deep geologic formations. *International Journal of Greenhouse Gas Control*, 93:102888, 2020. ISSN 1750-5836. doi: <https://doi.org/10.1016/j.ijggc.2019.102888>. URL <https://www.sciencedirect.com/science/article/pii/S1750583619303925>.
- Jerome A Neufeld, Marc A Hesse, Amir Riaz, Mark A Hallworth, Hamdi A Tchelepi, and Herbert E Huppert. Convective dissolution of carbon dioxide in saline aquifers. *Geophysical research letters*, 37(22), 2010. doi: <https://doi.org/10.1029/2010GL044728>.
- Maria T Elenius and Sarah E Gasda. Convective mixing in formations with horizontal barriers. *Advances in water resources*, 62:499–510, 2013. doi: <https://doi.org/10.1016/j.advwatres.2013.10.010>.
- Trine S. Mykkeltvedt and Jan M. Nordbotten. Estimating effective rates of convective mixing from commercial-scale injection. *Environmental Earth Sciences*, 67(2):527–535, 2012. doi: [10.1007/s12665-012-1674-3](https://doi.org/10.1007/s12665-012-1674-3). URL <https://doi.org/10.1007/s12665-012-1674-3>.
- M. A. Fernø, M. Haugen, K. Eikehaug, O. Folkvord, B. Benali, J. W. Both, E. Storvik, C. W. Nixon, R. L. Gawthrope, and J. M. Nordbotten. Room-Scale CO₂ Injections in a Physical Reservoir Model with Faults. *Transport in Porous Media*, 151(5):913–937, 2024. doi: [10.1007/s11242-023-02013-4](https://doi.org/10.1007/s11242-023-02013-4). URL <https://doi.org/10.1007/s11242-023-02013-4>.
- Jan Martin Nordbotten, Benyamine Benali, Jakub Wiktor Both, Bergit Brattekkås, Erlend Storvik, and Martin A. Fernø. DarSIA: An Open-Source Python Toolbox for Two-Scale Image Processing of Dynamics in Porous Media. *Transport in Porous Media*, 151(5):939–973, 2024a. doi: [10.1007/s11242-023-02000-9](https://doi.org/10.1007/s11242-023-02000-9). URL <https://doi.org/10.1007/s11242-023-02000-9>.
- Jan M. Nordbotten, Martin A. Fernø, Bernd Flemisch, Ruben Juanes, and Magne Jørgensen. *Final Benchmark Description: FluidFlower International Benchmark Study*, jul 2022.

- Bernd Flemisch, Jan M. Nordbotten, Martin Fernø, Ruben Juanes, Jakub W. Both, Holger Class, Mojdeh Delshad, Florian Doster, Jonathan Ennis-King, Jacques Franc, Sebastian Geiger, Dennis Gläser, Christopher Green, James Gunning, Hadi Hajibeygi, Samuel J. Jackson, Mohamad Jammoul, Satish Karra, Jiawei Li, Stephan K. Matthäi, Terry Miller, Qi Shao, Catherine Spurin, Philip Stauffer, Hamdi Tchelepi, Xiaoming Tian, Hari Viswanathan, Denis Voskov, Yuhang Wang, Michiel Wapperom, Mary F. Wheeler, Andrew Wilkins, AbdAllah A. Youssef, and Ziliang Zhang. The FluidFlower Validation Benchmark Study for the Storage of CO₂. *Transport in Porous Media*, 151(5):865–912, 2024. doi: 10.1007/s11242-023-01977-7. URL <https://doi.org/10.1007/s11242-023-01977-7>.
- Kristoffer Eikehaug, Emil Bang Larsen, Malin Haugen, Olav Folkvord, Benyamine Benali, Jakub Both, Jan Martin Nordbotten, and Martin Fernø. The International Fluidflower benchmark study dataset, January 2023.
- Kevin Riehl, Anastasios Kouvelas, and Michail A. Makridis. Revisiting reproducibility in transportation simulation studies. *European Transport Research Review*, 17(1):22, 2025. doi: 10.1186/s12544-025-00718-9. URL <https://doi.org/10.1186/s12544-025-00718-9>.
- David Landa-Marbán. pofff: An open-source image-based history-matching framework for the Fluidflower Benchmark study using OPM Flow, 2025. URL <https://github.com/cssr-tools/pofff>.
- David Landa-Marbán and Tor H. Sandve. pyopmspe11: A Python framework using OPM Flow for the SPE11 benchmark project. *Journal of Open Source Software*, 10(105):7357, 2025. doi: 10.21105/joss.07357. URL <https://doi.org/10.21105/joss.07357>.
- Kristoffer Eikehaug, Malin Haugen, Olav Folkvord, Benyamine Benali, Emil Bang Larsen, Alina Tinkova, Atle Rotevatn, Jan Martin Nordbotten, and Martin A. Fernø. Engineering Meter-scale Porous Media Flow Experiments for Quantitative Studies of Geological Carbon Sequestration. *Transport in Porous Media*, 151(5):1143–1167, 2024. doi: 10.1007/s11242-023-02025-0. URL <https://doi.org/10.1007/s11242-023-02025-0>.
- Malin Haugen, Lluís Saló-Salgado, Kristoffer Eikehaug, Benyamine Benali, Jakub W. Both, Erlend Storvik, Olav Folkvord, Ruben Juanes, Jan Martin Nordbotten, and Martin A. Fernø. Physical Variability in Meter-Scale Laboratory CO₂ Injections in Faulted Geometries. *Transport in Porous Media*, 151(5):1169–1197, 2024. doi: 10.1007/s11242-023-02047-8. URL <https://doi.org/10.1007/s11242-023-02047-8>.
- Jan M. Nordbotten, Martin A. Fernø, Bernd Flemisch, Anthony R. Kovscek, Knut-Andreas Lie, Jakub W. Both, Olav Møyner, Tor Harald Sandve, Etienne Ahusborde, Sebastian Bauer, Zhangxing Chen, Holger Class, Chaojie Di, Didier Ding, David Element, Abbas Firoozabadi, Eric Flau-

- raud, Jacques Franc, Firdovsi Gasanzade, Yousef Ghomian, Marie Ann Giddins, Christopher Green, Bruno R. B. Fernandes, George Hadjisotiriou, Glenn Hammond, Hai Huang, Dickson Kachuma, Michel Kern, Timo Koch, Prasanna Krishnamurthy, Kjetil Olsen Lye, David Landa-Marbán, Michael Nole, Paolo Orsini, Nicolas Ruby, Pablo Salinas, Mohammad Sayyafzadeh, Jakub Solovský, Jakob Torben, Adam Turner, Denis V. Voskov, Kai Wendel, and AbdAllah A. Youssef. Benchmarking CO₂ Storage Simulations: Results from the 11th Society of Petroleum Engineers Comparative Solution Project, 2025. URL <https://arxiv.org/abs/2507.15861>.
- T.H. Sandve, S.E. Gasda, A. Rasmussen, and A.B. Rustad. Convective dissolution in field scale CO₂ storage simulations using the OPM Flow simulator. In N. A. Røkke and K. Knuutila, editors, *TCCS-11. CO₂ Capture, Transport and Storage. Trondheim 22nd–23rd June 2021. Short Papers from the 11th International Trondheim CCS Conference*, pages 113–119. SINTEF Academic Press, Oslo, 2021. <https://hdl.handle.net/11250/2780667>.
- Jan M. Nordbotten, Martin A. Fernø, Bernd Flemisch, Anthony R. Kovscek, and Knut-Andreas Lie. The 11th Society of Petroleum Engineers Comparative Solution Project: Problem Definition. *SPE Journal*, 29(05):2507–2524, 05 2024b. ISSN 1086-055X. doi: 10.2118/218015-PA.
- Nicolas Spycher, Karsten Pruess, and Jonathan Ennis-King. CO₂-H₂O mixtures in the geological sequestration of co₂. 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. ISSN 0016-7037. doi: [https://doi.org/10.1016/S0016-7037\(03\)00273-4](https://doi.org/10.1016/S0016-7037(03)00273-4). URL <https://www.sciencedirect.com/science/article/pii/S0016703703002734>.
- Ian H. Bell, Jorrit Wronski, Sylvain Quoilin, and Vincent Lemort. Pure and Pseudo-pure Fluid Thermophysical Property Evaluation and the Open-Source Thermophysical Property Library CoolProp. *Industrial & Engineering Chemistry Research*, 53(6):2498–2508, 2014. doi: 10.1021/ie4033999.
- Matthew Goodfield, David Baxendale, Tobias Meyer Andersen, Jostein Alvestad, Kai Bao, Markus Blatt, Joshua Charles Bowden, Artur Castiel Reis de Souza, Paul Egberts, Joakim Hove, Negar Khoshnevis Garga, Håkon Hægland, Vegard Kippe, Peter Kirkham, Robert Klöfkorn, Stein Krogstad, Arne Morten Kvarving, David Landa-Marbán, Andreas Lauser, Kjetil Olsen Lye, Cintia Goncalves Machado, Lisa Julia Nebel, Halvor Møll Nilsen, Atgeirr Flø Rasmussen, Antonella Ritorto, Alf Birger Rustad, Tor Harald Sandve, Erik Hide Sæternes Ove Sævareid, Bård Skaflestad, Torbjørn Skille, Jakob Torben, Michal Tóth, Sverre Tveit, and Pieter J. Verveer. *OPM Flow Reference Manual (2025-04)*. Open Porous Media initiative, Norway, 2025. URL https://opm-project.org/?page_id=955.
- Ø. S. Klemetsdal, R. L. Berge, K.-A. Lie, H. M. Nilsen, and O. Møyner. Unstructured Gridding and Consistent Discretizations for Reservoirs with Faults and Complex Wells. volume SPE Reservoir

- Simulation Conference of *SPE Reservoir Simulation Conference*, page D031S009R005, 02 2017. doi: 10.2118/182666-MS.
- D.K. Ponting. Corner Point Geometry in Reservoir Simulation. volume ECMOR I - 1st European Conference on the Mathematics of Oil Recovery of *SPE Reservoir Simulation Conference*, pages cp-234-00003, 07 1989. doi: 10.3997/2214-4609.201411305.
- K. Holme, K.-A. Lie, O. Møyner, and A. Johansson. Grid-Orientation Effects in the 11th SPE Comparative Solution Project Using Unstructured Grids and Consistent Discretizations. volume SPE Reservoir Simulation Conference of *SPE Reservoir Simulation Conference*, page D021S007R001, 03 2025. doi: 10.2118/223885-MS.
- Jan M Nordbotten, Martin A Fernø, Bernd Flemisch, and Ruben Juanes. FluidFlower: A Meter-Scale Experimental Laboratory for Geological CO₂ Storage. *Transport in Porous Media*, 151(5): 859–863, 2024c. doi: 10.1007/s11242-024-02067-y.
- Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, Germany, 2008. doi: 10.1007/978-3-540-71050-9.
- Filippo Santambrogio. *Optimal transport for applied mathematicians*. Springer, Switzerland, 2015. doi: 10.1007/978-3-319-20828-2.
- J. W. Both, B. Brattekkås, E. Keilegavlen, M. Fernø, and J. M. Nordbotten. High-fidelity experimental model verification for flow in fractured porous media. *InterPore Journal*, 1(3):IPJ271124-6, 2024. doi: 10.69631/ipj.v1i3nr31.
- Christopher Green, Samuel J. Jackson, James Gunning, Andy Wilkins, and Jonathan Ennis-King. Modelling the FluidFlower: Insights from Characterisation and Numerical Predictions. *Transport in Porous Media*, 151(5):1093–1111, 2024. doi: 10.1007/s11242-023-01969-7. URL <https://doi.org/10.1007/s11242-023-01969-7>.
- Lluís Saló-Salgado, Malin Haugen, Kristoffer Eikehaug, Martin Fernø, Jan M. Nordbotten, and Ruben Juanes. Direct Comparison of Numerical Simulations and Experiments of CO₂ Injection and Migration in Geologic Media: Value of Local Data and Forecasting Capability. *Transport in Porous Media*, 151(5):1199–1240, 2024. doi: 10.1007/s11242-023-01972-y. URL <https://doi.org/10.1007/s11242-023-01972-y>.
- Yuhang Wang, Ziliang Zhang, Cornelis Vuik, and Hadi Hajibeygi. Simulation of CO₂ Storage Using a Parameterization Method for Essential Trapping Physics: FluidFlower Benchmark Study. *Transport in Porous Media*, 151(5):1053–1070, 2024. doi: 10.1007/s11242-023-01987-5. URL <https://doi.org/10.1007/s11242-023-01987-5>.

- Michiel Wapperom, Xiaoming Tian, Aleks Novikov, and Denis Voskov. FluidFlower Benchmark: Lessons Learned from the Perspective of Subsurface Simulation. *Transport in Porous Media*, 151(5):1033–1052, 2024. doi: 10.1007/s11242-023-01984-8. URL <https://doi.org/10.1007/s11242-023-01984-8>.
- Mohamad Jammoul, Mojdeh Delshad, and Mary F. Wheeler. Numerical Modeling of CO₂ Storage: Applications to the FluidFlower Experimental Setup. *Transport in Porous Media*, 151(5):1071–1091, 2024. doi: 10.1007/s11242-023-01996-4. URL <https://doi.org/10.1007/s11242-023-01996-4>.
- Xiaoming Tian, Michiel Wapperom, James Gunning, Samuel Jackson, Andy Wilkins, Chris Green, Jonathan Ennis-King, and Denis Voskov. Correction: A History Matching Study for the FluidFlower Benchmark Project. *Transport in Porous Media*, 151(5):1141–1141, 2024. doi: 10.1007/s11242-024-02065-0. URL <https://doi.org/10.1007/s11242-024-02065-0>.
- Geir Evensen. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. doi: 10.1007/s10236-003-0036-9. URL <https://doi.org/10.1007/s10236-003-0036-9>.
- Alexandre A Emerick and Albert C Reynolds. Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55:3–15, 2013.
- Rainer Storn and Kenneth Price. Differential evolution –a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997. doi: 10.1023/A:1008202821328. URL <https://doi.org/10.1023/A:1008202821328>.
- Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T. Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, JoséVinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez,

- Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, Yoshiki Vázquez-Baeza, and SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2. URL <https://doi.org/10.1038/s41592-019-0686-2>.
- Nicholas Rescher. *Luck Theory: A Philosophical Introduction to the Mathematics of Luck*, volume 20 of *Logic, Argumentation & Reasoning*. Springer, Switzerland, 2021. doi: 10.1007/978-3-030-63780-4.
- Jichao Yin, Han-Young Park, Akhil Datta-Gupta, Michael J. King, and Manoj K. Choudhary. A hierarchical streamline-assisted history matching approach with global and local parameter updates. *Journal of Petroleum Science and Engineering*, 80(1):116–130, 2011. ISSN 0920-4105. doi: <https://doi.org/10.1016/j.petrol.2011.10.014>. URL <https://www.sciencedirect.com/science/article/pii/S0920410511002701>.
- Alessandro Suriano, Costanzo Peter, Christoforos Benetatos, and Francesca Verga. Gridding Effects on CO₂ Trapping in Deep Saline Aquifers. *Sustainability*, 14(22), 2022. ISSN 2071-1050. doi: 10.3390/su142215049. URL <https://www.mdpi.com/2071-1050/14/22/15049>.
- M. J. Martinez and M. A. Hesse. Two-phase convective CO₂ dissolution in saline aquifers. *Water Resources Research*, 52(1):585–599, 2016. doi: 10.1002/2015WR017085.
- OP Folkvord, JW Both, K Eikehaug, JM Nordbotten, and M Fernø. Laboratory Evaluation of Physical Variability of Multiphase Flow During CO₂ Sequestration. In *World CCUS Conference 2025*, volume 2025, pages 1–4. European Association of Geoscientists & Engineers, 2025.