TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Seabra, G. S., Mücke, N. T., Silva, V. L. S., Voskov, D., & Vossepoel, F. C. (2024). AI enhanced data assimilation and uncertainty quantification applied to Geological Carbon Storage. *International Journal of Greenhouse Gas Control*, *136*, Article 104190. https://doi.org/10.1016/j.ijggc.2024.104190

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

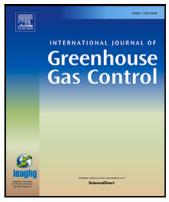
International Journal of Greenhouse Gas Control 136 (2024) 104190



Contents lists available at ScienceDirect

International Journal of Greenhouse Gas Control

journal homepage: www.elsevier.com/locate/ijggc



AI enhanced data assimilation and uncertainty quantification applied to Geological Carbon Storage

Gabriel Serrão Seabra [a,b,*], Nikolaj T. Mücke [c,d], Vinicius Luiz Santos Silva [b,e], Denis Voskov [a,f], Femke C. Vossepoel [a]

[a] Faculty of Civil Engineering and Geosciences, TU Delft, Stevinweg 1, 2628 CN Delft, Netherlands
[b] Petroleo Brasileiro S.A. (Petrobras), Rio de Janeiro, Brazil
[c] Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, Netherlands
[d] Mathematical Institute, Utrecht University, Utrecht, Netherlands
[e] Department of Earth Science and Engineering, Imperial College London, London, United Kingdom
[f] Department of Energy Resources Engineering, Stanford University, CA, USA

ARTICLE INFO

Keywords:
Geological Carbon Storage (GCS)
Uncertainty quantification
Data assimilation
Machine learning

ABSTRACT

This study investigates the integration of machine learning (ML) and data assimilation (DA) techniques, focusing on implementing surrogate models for Geological Carbon Storage (GCS) projects while maintaining the high fidelity physical results in posterior states. Initially, we evaluate the surrogate modeling capability of two distinct machine learning models, Fourier Neural Operators (FNOs) and Transformer UNet (T-UNet), in the context of CO_2 injection simulations within channelized reservoirs. We introduce the Surrogate-based hybrid ESMDA (SH-ESMDA), an adaptation of the traditional Ensemble Smoother with Multiple Data Assimilation (ESMDA). This method uses FNOs and T-UNet as surrogate models and has the potential to make the standard ESMDA process at least 50% faster or more, depending on the number of assimilation steps. Additionally, we introduce Surrogate-based Hybrid RML (SH-RML), a variational data assimilation approach that relies on the randomized maximum likelihood (RML) where both the FNO and the T-UNet enable the computation of gradients for the optimization of the objective function, and a high-fidelity model is employed for the computation of the posterior states. Our comparative analyses show that SH-RML offers a better uncertainty quantification when compared to the conventional ESMDA for the case study.

1. Introduction

Geological Carbon Storage (GCS) is a vital component of Carbon Capture, Utilization, and Storage (CCUS) to mitigate greenhouse gas emissions and achieve climate targets (Ringrose and Meckel, 2019). According to the International Energy Agency (IEA), project developers aim to bring more than 200 new capture and storage facilities into operation worldwide by 2030, with the capacity to handle over 220 Mton of CO_2 annually (IEA, 2022). To put this ambitious goal into context, one of the largest CCUS projects to date, the water alternating gas (WAG) injection project in the Brazilian Pre-Salt, has injected only 20 Mton of CO_2 into the largest Brazilian four carbonate reservoirs (Búzios, Mero, Sapinhoá, and Tupi) over a decade (Nunes et al., 2022), representing less than 10% of the IEA target. Many CO_2 storage sites are situated in geologically complex formations, such as fractured carbonate rocks or channelized reservoirs (Burchette, 2012; March et al., 2018). As the number of carbon storage projects grows, it becomes crucial for companies to develop efficient forecasts and uncertainty quantification (UQ), as these studies play an important role in securing support from investors, regulators, and society during project approval and implementation.

The ambitious scale of expansion in Carbon Capture, Utilization, and Storage (CCUS) projects necessitates addressing the associated risks, particularly those linked to pressure management in geological storage sites. Effective pressure management in Geological Carbon Storage (GCS) projects is crucial to optimizing storage capacity and mitigating significant risks such as induced seismicity and caprock failure. Li and Liu (2016) identified that most GCS project risks originate from pressure management issues at storage sites, which can induce seismicity and lead to caprock failure at weak points when CO_2 injection pressures exceed the mechanical stability thresholds of the geological formations. These scenarios can result in microseismic events or mechanical failures that compromise the caprock's integrity and

* Corresponding author at: Faculty of Civil Engineering and Geosciences, TU Delft, Stevinweg 1, 2628 CN Delft, Netherlands.
E-mail addresses: g.serraoseabra@tudelft.nl (G.S. Seabra), nikolaj.mucke@cwi.nl (N.T. Mücke), v.santos-silva19@imperial.ac.uk (V.L.S. Silva), voskov@tudelft.nl (D. Voskov), f.c.vossepoel@tudelft.nl (F.C. Vossepoel).

https://doi.org/10.1016/j.ijggc.2024.104190
Received 13 February 2024; Received in revised form 6 May 2024; Accepted 25 June 2024
Available online 8 July 2024
1750-5836/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the overall containment security (Zoback and Gorelick, 2012; Rutqvist, 2012). Furthermore, the CO_2 injection process can alter subsurface stress states, potentially triggering seismic events ranging from microseismicity to significant earthquakes (magnitude 5+ events) (White and Foxall, 2016). Researchers have proposed several solutions to improve pressure management at storage sites. For instance, Machado et al. (2023) suggests the application of horizontal wells to achieve a better distribution of the CO_2 plume and a smaller increase in reservoir pressure.

Robust UQ and forecasts in GCS projects typically rely on several key components. These encompass robust geological models representing reservoir complexities accurately, high-fidelity reservoir simulators capturing intricate CO_2 injection dynamics and data assimilation (DA) techniques to combine these elements with field observations. It is important to note that effective uncertainty quantification can still be achieved even without data assimilation, depending on the specific requirements of the project and the available data.

DA techniques contain ensemble methods (e.g., Ensemble Kalman Filters and Ensemble Smoothers), variational methods (e.g., Randomized Maximum Likelihood and 4D-Var), and fully nonlinear DA methods (e.g., Particle Filters and Markov Chain Monte Carlo) (Evensen et al., 2022; Tarantola, 2005). These methods leverage data, prior knowledge, and physics-based models to predict reservoir behavior under uncertain conditions. While ensemble methods are computationally efficient and flexible, variational approaches can offer better convergence but require gradient computations. Fully nonlinear methods can offer high accuracy for systems characterized by nonlinearities, such as the CO_2 injection. However, the computational intensity of these methods can be prohibitive due to the substantial resources and time they require. In the context of CO_2 injection and DA, Tarrahi et al. (2015) integrated microseismic monitoring data of CO_2 injection with coupled flow and geomechanical models using the Ensemble Kalman Filter (EnKF), enabling the conditioning of heterogeneous rock permeability and geomechanical property distributions on microseismic data. Similarly, Li et al. (2017) employed one-step ahead smoothing for joint state-parameter estimation, crucial for addressing nonlinearities in CO_2 storage aquifers. Tadjer and Bratvold (2021) introduced a Bayesian Evidential Learning (BEL) for managing uncertainties in geological CO_2 storage, integrating Monte Carlo simulations with ensemble smoother direct forecasting to enhance decision support systems. Utilizing the Utsira saline aquifer in Norway as an example, the BEL approach improves the predictability of CO_2 sequestration by effectively managing uncertainties and potential leakages. These studies emphasize the significance of assimilating diverse data types into reservoir models, refining the understanding of subsurface properties, and optimizing CO_2 injection strategies, thereby contributing to the realization of effective GCS projects in mitigating climate change.

Specialized softwares are often employed to model the heterogeneities in complex reservoirs, such as channelized formations. Among these, Alluvsim is an open-source option that generates multiple geological models with key features like channel size, curvature, and shifts, using streamlines as building blocks to mimic natural deposition processes (Pyrcz et al., 2009). These detailed models reflect the heterogeneities commonly encountered in GCS projects. For high-fidelity reservoir simulation, numerical simulators designed to handle multiphase, multicomponent flow and transport in porous media are utilized. Examples of such simulators include CMG GEM (Computer Modelling Group Ltd., 2023), SLB Eclipse (Schlumberger, 2023), and open-source options like DuMux (DuMux Development Team, 2023) and GEOSX (GEOSX Development Team, 2023). Delft Advanced Research Terra Simulator (DARTS), recently released as an open-source reservoir simulator for energy transition applications, efficiently simulates CO_2 injection using advanced numerical techniques like the operator-based linearization approach (Lyu and Voskov, 2023; Khait and Voskov, 2017). However, integrating these simulators into DA frameworks can be challenging due to high computational costs.

The choice of Alluvsim and DARTS as our primary tools stems from their proven efficacy in handling complex geological formations and fluid dynamics simulations, respectively. Alluvsim's capability to accurately simulate channelized reservoirs (Delottier et al., 2023), coupled with DARTS's optimized computational efficiency (Khait and Voskov, 2017), makes them particularly suitable for our study's objectives. This combination allows us to model the intricate interactions and variabilities within the reservoirs with a high degree of fidelity, crucial for reliable CO_2 injection simulations.

Recently, researchers have actively been exploring innovative strategies to merge machine learning (ML) and DA (Buizza et al., 2022; Silva et al., 2023). Buizza et al. (2022) provides a high-level overview of techniques for integrating DA and ML, called "Data Learning" for improving DA in several fields. Their key focus is on approaches that leverage the strengths of ML's ability to uncover complex patterns in data and data assimilation's incorporation of physical models and dynamical constraints. Similarly, Cheng et al. (2023) explores how mixing ML and DA can make research on DA robust. The study sorts these methods into two main groups. The first group, called "DA using ML" looks at how ML can help solve problems in data assimilation. This includes fixing errors in DA models by adding ML, using it to estimate unknown variables in DA, and defining error rates using ML methods. The study also talks about how neural networks can help in learning DA systems from start to finish. The second group, "ML improved by DA/UQ" focuses on how DA and UQ can improve ML models. This covers topics like using Bayesian neural networks for uncertainty analysis in ML, fixing errors in simplified ML models with real-time data, and using DA to identify key equations from noisy or incomplete data. Brajard et al. (2021) recently proposed an innovative approach that combines DA and ML to infer unresolved scale parameterization in models, helping overcome limitations from sparse and noisy observational data.

In the domain of deep learning for efficient surrogate modeling, UNets have long demonstrated their efficacy, particularly in tackling subsurface problems (Wen et al., 2021a; Zhang et al., 2021; Pintea et al., 2021). Originating from biomedical image segmentation, UNets excel at capturing local features through specialized convolutional layers. The architecture comprises an encoder and a decoder connected by a "highway" system of channel concatenation, facilitating the transfer of multiscale spatial information. This has enabled outstanding predictive accuracy in diverse applications (Ronneberger et al., 2015; Taccari et al., 2022). More recently, advancements have been made by integrating UNets with transformers. Li et al. (2023) explored this method for robust medical images segmentation and AlSalmi and Elsheikh (2023) applied an attention UNet for seismic segmentation. On the other hand, Fourier Neural Operators (FNOs) have recently emerged as a promising method to build surrogate models for reservoirs submitted to CO_2 injection. Employing Fourier basis functions, FNOs efficiently capture multiscale interactions and offer a novel way to overcome traditional limitations in surrogate modeling (Li et al., 2020; Wen et al., 2022; Witte et al., 2023).

Recently, Tang et al. (2022), Wen et al. (2021b), Sun and Durlofsky (2019), Agogo et al. (2022) developed surrogate models for CCUS DA, aiming to replace physics-based numerical models. However, these models often require a substantial amount of training data from high-fidelity simulations, posing practical challenges for real-world CCUS projects with limited computational resources. While deep learning has improved surrogate model accuracy, Dong et al. (2021) shows that it may struggle to capture subsurface complexities fully.

Hybrid models that combine physics-based and ML approaches have been explored by Tang and Durlofsky (2022), Korondi et al. (2020), de Brito and Durlofsky (2020) to mitigate the specific limitations inherent to both physics-based and ML models, with the goal of forging a more balanced and comprehensive tool. ML models, while proficient at identifying patterns and correlations within large datasets, may lack the capability to infer the underlying physical processes governing the

system. This limitation can lead to potential inaccuracies in predictions under unseen conditions or parameter ranges, especially when dealing with the intricate geological variations and non-linear fluid dynamics inherent in subsurface environments. On the other hand, physics-based models offer reliable insights into these underlying processes but may struggle with computational efficiency. However, significant challenges arise when incorporating ML surrogate models in DA due to the different parameterizations employed by ML surrogates and physics-based, high-fidelity simulators. This misalignment can impede the integration, limiting the applicability and efficacy of the resulting UQ in real-world GCS projects.

In our methodology, we initially employ a standard ESMDA approach utilizing DARTS for high-fidelity simulations of channelized reservoirs built with Alluvsim. This scenario poses a significant challenge for DA due to its highly nonlinear nature and the non-Gaussian distribution of parameters. To improve upon the standard ESMDA methodology, we evaluate two ML surrogate models for comparison: one rooted in the Fourier Neural Operators and the other adopting a Transformer UNet (T-UNet) architecture, which to our knowledge is the first application of these techniques to GCS subsurface problems. Our observations reveal that FNOs show a slight advantage over T-UNet, particularly for small datasets. Subsequently, we develop two hybrid techniques to integrate DA with these ML surrogates. The first, termed Surrogate-based hybrid ESMDA (SH-ESMDA), incorporates the ML surrogates models, expediting the ESMDA process by around 50% or more, and thereby facilitating quicker uncertainty evaluations. For the second technique, known as Surrogate-based Hybrid RML (SH-RML), we use the ML surrogates models specifically for gradient calculations within a variational framework and compute the posterior curves with high-fidelity physics simulator DARTS. The SH-RML achieves better history matching than ESMDA and SH-ESMDA.

In summary, our contributions are as follows:

- We train and test two different types of novel ML surrogates in a channelized reservoir setting for CO_2 storage, a FNO and a T-UNet.
- We introduce two novel hybrid methods, SH-ESMDA and SH-RML, that incorporate ML into both ensemble and variational DA techniques. The first, significantly expedites the DA process and the second allows one to perform variational DA.
- Both proposed methods ensure that posterior high-fidelity physics solution is respected.
- The proposed methods are versatile and can be adapted to various physical systems beyond CO_2 sequestration.

The paper is organized as follows: Section 2 discusses the creation of geological models, followed by CO_2 injection simulations with DARTS. In Section 3, we delve into the ML models used to build the surrogate models. Section 4 and Section 5 describes the DA methods discussed in the paper, ESMDA, RML and the hybrid methods SH-ESMDA and SH-RML. Finally, Section 6 presents the results and implications of these methods for enhancing DA in CO_2 storage projects.

2. Overview of the reservoir model for CO_2 injection

2.1. Geological modeling using Alluvsim

We use Alluvsim, a specialized algorithm to simulate channelized reservoirs (Pyrcz et al., 2009; Delottier et al., 2023). This tool allows us to manipulate different geological variables that impact channel features. Following the guidelines by Pyrcz et al. (2009), we create multiple geological models by altering essential parameters within certain limits.

We consider as variables the likelihood of channel shifts, known as avulsion probability, and vertical sediment build-up, or aggradation levels, within statistically defined ranges. Parameters such as channel orientation, thickness, and geometric aspects like the width-to-thickness ratio are also modeled using various distributions to mimic natural variability. We similarly vary levee width to represent lateral sediment deposition and channel sinuosity to capture meandering behavior. Properties are distributed across different facies, to accurately represent variations in rock quality. This allows a comprehensive evaluation of the reservoir's attributes while accounting for uncertainty.

By randomly selecting values for the aforementioned parameters, we produce multiple realizations that capture the variability in the properties of channelized reservoirs. Fig. 1 showcases the permeability distributions of six randomly sampled models from this dataset. Each model in the dataset has a grid dimension of $32 \times 32 \times 1$ with a spatial discretization of $192 \times 192 \times 10$ m. This grid resolution was carefully chosen, considering computational efficiency and memory requirements.

The histograms in Fig. 2 clarify the permeability distribution of a model sample from our dataset. Note that the distributions are not Gaussian. These diverse geological models define the permeability distribution for subsequent CO_2 simulations which will, in turn, be used for the training of the ML methods.

It is worth mentioning that a 2D horizontal model was employed to simplify the study, focusing on the areal permeability distribution within channelized reservoirs. While acknowledging the influence of gravity in three-dimensional settings, this approach enables the examination of non-Gaussian and highly heterogeneous permeability patterns typical of natural depositional environments. The decision to use a 2D model was made to highlight how lateral variations in permeability impact CO_2 flow dynamics.

2.2. CO_2 injection simulation using DARTS

Delft Advanced Research Terra Simulator (DARTS) is used to perform reservoir simulations of CO_2 injection into channelized reservoirs. DARTS is engineered for handling complex flow through porous media and is optimized for computational speed via techniques such as operator-based linearization (OBL) yielding a high-fidelity, physics-based representation (Khait and Voskov, 2017; Pour et al., 2023; Chen and Voskov, 2020). Detailed descriptions of the physics, analytical validations, and example models of DARTS for GCS modeling are documented in Wapperom et al. (2023), Lyu and Voskov (2023) and openly available in the software open repository (DARTS, 2024).

For the CO_2 flow simulation in this study, the gas phase viscosity is assumed to be solely that of CO_2, calculated using the model from Fenghour et al. (1998). The densities of the CO_2 and CH_4 mixture are computed using the Peng-Robinson equation of state, with properties determined by their composition. In terms of dissolution, all components are modeled simultaneously using a hybrid EOS with fugacity-activity formulation (Wapperom et al., 2023). A power-law model (Corey) is applied to describe relative permeability in the porous media, with exponent values $n_w = 2.0$ and $n_g = 1.5$, and terminal saturation points $S_{wc} = 0.25$ and $S_{gc} = 0.1$. Although aspects like diffusion and hysteresis can influence CO_2 migration and trapping, they are neglected to align the simulation objectives with broader reservoir behavior. Capillary pressure effects are also neglected, as they are considered to have minimal impact on the 2D model utilized. We assume this model captures the multiphase flow characteristics pertinent to CO_2 migration and trapping in porous media relevant this the present study.

In DARTS, the geological model is discredited as a computational grid. Within this grid, one well is placed in the center of the reservoir for CO_2 injection. The model simulates the injected CO_2 plume and conducts equilibrium flash calculations to determine phase partitioning.

Our simulation includes CO_2, CH_4, and H_2O as components, with initial conditions featuring a uniform gas saturation of 20% and a composition of 2% CO_2 and 98% CH_4. CO_2 is injected via one well

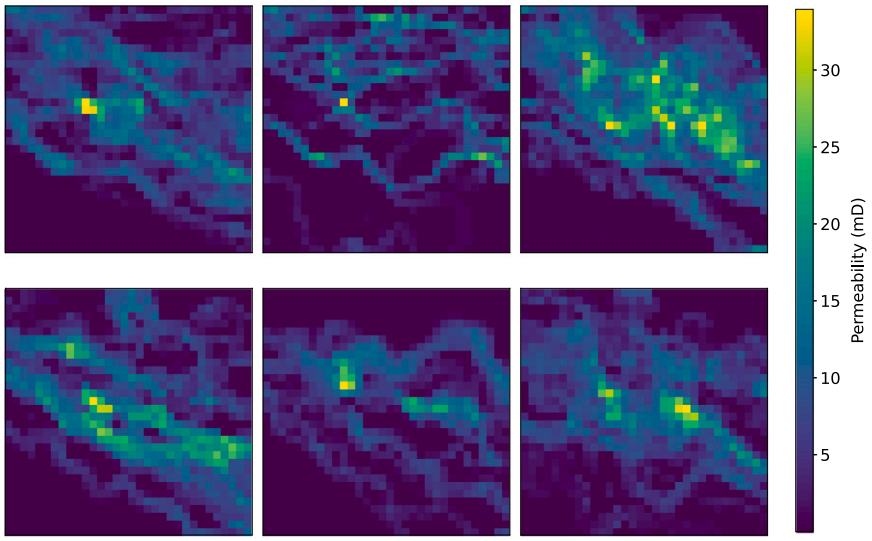


Fig. 1. Permeability maps of six models from the dataset.

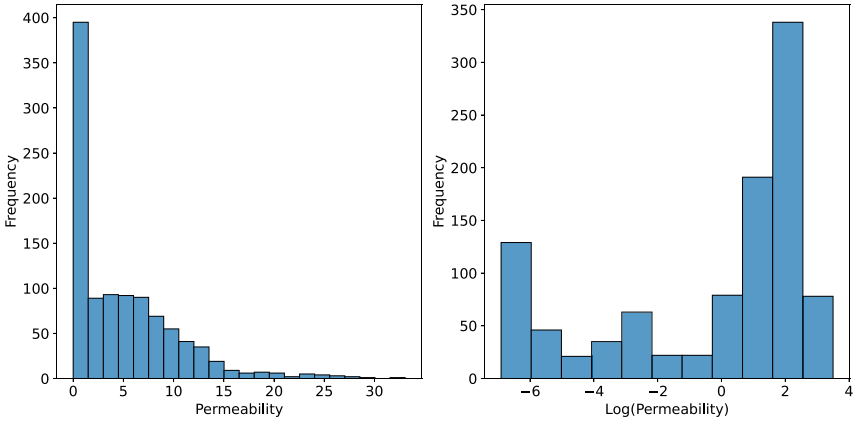


Fig. 2. Histogram of permeability (left) and log of permeability (right) for one sample.

in the center of the reservoir at a time-varying prescribed gas rate. In our simulations, the time frame comprises 61 time steps, each lasting 30 days. Although this is a shorter window than what is typically encountered in real-world CCUS projects, this duration is sufficient to induce overpressure in the reservoir, which is an aspect we aim to investigate as it can impact the further CO_2 distribution in the reservoir. Fig. 3 illustrates our simulation results. The top row shows pressure distributions, and the bottom row displays CO_2 molar fractions, captured at the initial, intermediate, and final time steps.

A key observation is the differential progression of the pressure and CO_2 fronts. Pressure changes, governed by diffusion phenomena propagate more rapidly through the reservoir. In contrast, the CO_2 front advances more slowly, influenced by complex transport mechanisms. This differential movement highlights the potential value of monitoring pressure as an early warning system for subsurface changes, even before significant CO_2 migration occurs. Because of this, we choose pressure as a critical variable for monitoring in our subsequent DA studies.

By applying DARTS to the Alluvsim realizations, we generate a comprehensive suite of high-fidelity simulations that form the basis for our subsequent ML model training and DA study. It is important to note that the combination of geological complexities modeled by Alluvsim and the fluid dynamics simulated by DARTS creates a highly nonlinear problem. Coupled with the non-Gaussian distribution parameters, as evident from Fig. 2, this poses a substantial challenge for traditional DA methods. This complexity further underscores the need for advanced approaches, including ML-based techniques, to accurately perform DA.

3. Neural network as surrogate forward models

A surrogate model is a model that replaces the high-fidelity model for the simulation of fluid behavior in a reservoir. The surrogate model should be computationally cheap to evaluate without significantly sacrificing accuracy. To achieve this, we construct a surrogate model based on the output of the high-fidelity model. This is done in two stages: an *offline stage*, in which the surrogate model is trained, and an *online stage*, where the surrogate model is used in place of the high-fidelity model. Conventionally, surrogate models achieve speed-ups by lowering the dimension of the state and corresponding equations to be solved, such as in proper orthogonal decomposition (Quarteroni et al., 2015; Hesthaven et al., 2016). However, due to large Kolmogorov N-widths for highly nonlinear and/or hyperbolic problems (Ohlberger and Rave, 2015), such linear approaches have lately been replaced with equation-free methods. Here, one replaces the equations with a function that directly gives the output of interest. For this to work, one must typically use highly complicated functions to make up for the lack of physical knowledge in the online stage. The offline stage is typically significantly more expensive and requires more training data. Neural networks are immensely popular in this approach due to their

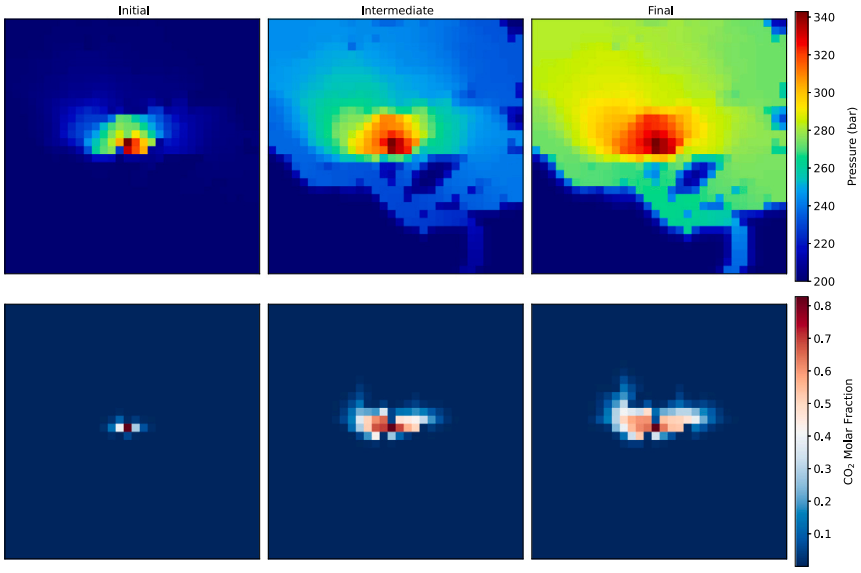


Fig. 3. Simulation results using DARTS. Top row: Pressure distributions; Bottom row: CO_2 molar fractions.

capability of approximating highly nonlinear functions (Mücke et al., 2021; Li et al., 2020; Geneva and Zabaras, 2022).

The general setup for the offline stage is to first generate high-fidelity solutions and then train the neural networks on these solutions. In this paper, we specifically make use of the neural network architectures T-UNet and FNOs.

3.1. Forward model

The forward model, that is, the simulator used to describe the behavior of fluid in a reservoir, maps input parameters to a state trajectory. In our case, it maps permeability and porosity to the state trajectory of pressure and fluid flow. Injection rate is the control that determines the behavior of the fluid. Let $\boldsymbol{K} \in \mathbb{R}^{N_x \times N_y}$ be the permeability, $\boldsymbol{\phi} \in \mathbb{R}^{N_x \times N_y}$ the porosity, $\boldsymbol{q} \in \mathbb{R}^{N_t}$ the injection rate, and $\boldsymbol{m} = (K, \phi, q)$. We then define the space of parameters $\boldsymbol{m} \in M$. The state trajectories consist of pressure, $\boldsymbol{p} \in \mathbb{R}^{N_x \times N_y \times N_t}$, and CO_2 molar fraction, $\boldsymbol{f} \in \mathbb{R}^{N_x \times N_y \times N_t}$. We furthermore define the state trajectory space, $\boldsymbol{d} \in V$ and find that $\boldsymbol{d} = (\boldsymbol{p}, \boldsymbol{f})$. With this, the forward map, G, is defined by:

$$G : M \to V, \quad \boldsymbol{m} \mapsto \boldsymbol{d}. \tag{1}$$

The surrogate forward model, $\hat{G}$, approximates G:

$$\hat{G} : M \to V, \quad \boldsymbol{m} \mapsto \boldsymbol{d}, \quad \hat{G}(\boldsymbol{m}) \approx G(\boldsymbol{m}). \tag{2}$$

While G maps $\boldsymbol{m}$ to $\boldsymbol{d}$ implicitly by solving a set of PDEs, $\hat{G}$ directly maps $\boldsymbol{m}$ to $\boldsymbol{d}$. $\hat{G}$ is parametrized by a family of neural networks. As $\hat{G}$ is a neural network, it consists of a set of weights, $\boldsymbol{\theta}$. The weights are fitted in the offline stage.

The training of $\hat{G}$ is performed in the offline stage by first generating N_s training samples by using the high-fidelity forward model:

$$M_{\text{train}} = \{\boldsymbol{m}_i\}_{i=1}^{N_s}, \quad V_{\text{train}} = \{G(\boldsymbol{m}_i)\}_{i=1}^{N_s} = \{\boldsymbol{d}_i\}_{i=1}^{N_s}, \quad S_{\text{train}} = (M_{\text{train}}, V_{\text{train}}). \tag{3}$$

$\hat{G}$ is then trained by minimizing a loss function with respect to the weights of $\hat{G}$, θ:

$$L(\hat{G}, S_{\text{train}}) = \frac{1}{N_s} \sum_{i=1}^{N_s} l(\hat{G}(\boldsymbol{m}_i), \boldsymbol{d}_i) + \lambda \|\boldsymbol{\theta}\|_2^2, \quad (\boldsymbol{m}_i, \boldsymbol{d}_i) \in S_{\text{train}}, \tag{4}$$

where l is some loss, typically the l^p or L^p norm of the residual, λ is a hyperparameter, and $\|\cdot\|_2^2$ is the squared l^2 norm and serves as a regularization term. The minimization of L is performed by stochastic gradient descent (SGD) with respect to $\boldsymbol{\theta}$. The specific SGD algorithm is often chosen to be the Adam optimizer (Kingma and Ba, 2014) or other variations thereof.

Below, we will present the particular neural network architectures we use in this paper.

3.2. Transformer UNet

The UNet architecture was introduced in Ronneberger et al. (2015). The idea is to reduce the dimension of the input data down to a bottleneck via a series of convolutional layers and then increase the dimension back to the original shape via upscaling convolutional layers. The bottleneck layers serves as a low-dimensional representation of the data that is rich in feature information. In the upscaling part of the network, the intermediate layers from the downscaling part are concatenated to the convolutions to provide context.

Originally developed and primarily utilized in the medical imaging field, the T-UNet architecture is being adapted in this work for GCS projects. To the best of our knowledge, this constitutes the first endeavor to apply the T-UNet architecture in the realm of GCS.

The choice of the T-UNet architecture, a variant of the well-established UNet model, is grounded in its substantial application and success in various fields, particularly in subsurface applications (Ronneberger et al., 2015; AlSalmi and Elsheikh, 2023).

The UNet type of architecture allows one to add information to the predictions on various spatial levels. Specifically, we utilize this to inform the forward model with the injection rate in the bottleneck layers. By adding this information in the bottleneck layers, we effectively affect the rich feature encodings with additional information in an efficient manner. This makes the computations cheaper and makes it easier for the network to learn the relations between the input parameters and the output.

In the proposed architecture, we input the spatially distributed parameters, porosity and permeability, as a two-channel "image". Then we copy that N_t times and concatenate a channel consisting only of the time. This way we can compute the bottleneck encoding of the space for each time step as a batch consisting of 3D tensors (channels, height, width) rather than a single 4D tensor (channels, time steps, height, width), which enables us to use 2D convolutional layers instead of 3D convolutional layers. Since 3D convolutional layers are significantly

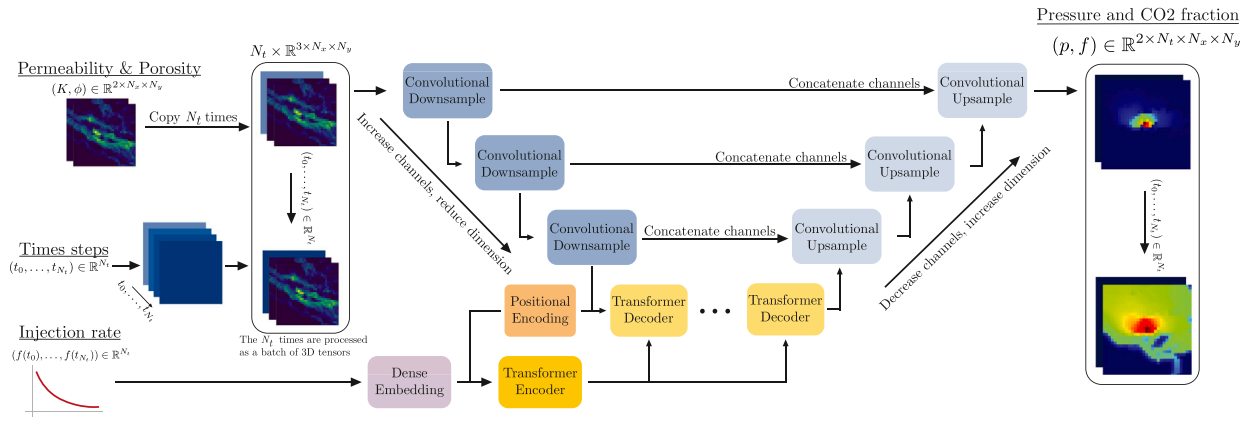


Fig. 4. T-UNet architecture.

more memory and compute-heavy, this gives us significantly more efficiency.

For the conditioning of the bottleneck layer, we use cross-attention in the shape of the transformer architecture (Vaswani et al., 2017). The transformer has been shown to provide state-of-the-art performance on multiple types of data and seems to be the superior choice for multi-modal data (Rombach et al., 2022; Xu et al., 2023). Unfortunately, the attention mechanism is very compute and memory intensive and scales poorly with the data dimension. By utilizing the transformer in the bottleneck layers, however, this problem is circumvented. For a description of the transformer neural network, see Vaswani et al. (2017).

As mentioned, we utilize transformers to condition the forward model on the injection rate. We do this by first embedding the injection rate time series through dense layers, such that the dimensions match the encoded spatial dimensions. Then, we employ positional encoding which receives information from a dense embedding that originates from the gas injection rates. The encoded positions are then passed to the transformer decoder layers, effectively providing a richer context for each time step. This ensures that the transformer is not only aware of the feature information but also the sequence in which they occur. The embedding time series is passed through transformer encoder layers after which it is combined with the spatial data through transformer decoders. For a visualization of the full T-UNet, see Fig. 4

3.3. Fourier neural operators

FNOs were introduced in Li et al. (2020) for various parametric PDE problems. In contrast to conventional neural networks, FNOs learn operators between function spaces instead of Euclidean spaces. This makes FNOs resolution invariant. The general idea is to make use of the Fourier transform, followed by a series of operations in Fourier space, after which the data is transformed back to physical space. Hence, a single Fourier layer is given by:

$$a^{n+1}(x) = \sigma\left(W a^n(x) + \mathcal{F}^{-1}(R \cdot (\mathcal{F} a^n))(x)\right), \tag{5}$$

where $a^n(x)$ is the output of the nth layer, W and R are affine transformations consisting of trainable weights, $\mathcal{F}$ is the Fourier transform, and σ is an activation function. Before applying R the number of modes is truncated to a pre-defined number of modes, k. The Fourier transform is in practice approximated by a discrete Fourier transform. W is typically a standard convolutional layer with a kernel of size one. While truncating the number of Fourier modes removes high-frequency information, the convolutional layer, W, compensates for that. For visualization of the FNO layer, see Fig. 5

The FNO layers are preceded by a projection layer, that maps the number of input channels to the desired number of hidden channels. Similarly, the FNO layers are superseded by another projection layer that maps the number of hidden channels to the number of output channels.

For our specific application, we use the FNO to map parameters, $\boldsymbol{m}$, to corresponding state trajectories, $\boldsymbol{d}$. To capture the 3D structure of the data, we use 3D Fourier transform and 3D convolutions. The injection rate is encoded to have a 3D structure. The rate at each time step is copied onto all discrete spatial points:

$$\boldsymbol{q}_{\text{enc}} = [\boldsymbol{q}_0 \mathbf{1}_{N_x\times N_y}, \ldots, \boldsymbol{q}_{N_t} \mathbf{1}_{N_x\times N_y}], \tag{6}$$

where $\mathbf{1}_{N_x\times N_y} \in \mathbb{R}^{N_x\times N_y}$ is a matrix consisting of ones. The subscript $()_{\text{enc}}$ signifies that the quantities are encoded to fit the 3D tensor format. Similarly, the spatial points coordinates, (x, y) are encoded and copied along the temporal dimension. The time steps are treated in the same way as the injection rate. Lastly, the porosity and permeability are also copied along the temporal dimension. Hence, the input to the FNO is:

$$(K, \phi, q, x, y, t)_{\text{enc}} = (K_{\text{enc}}, \phi_{\text{enc}}, q_{\text{enc}}, x_{\text{enc}}, y_{\text{enc}}, t_{\text{enc}}) \in \mathbb{R}^{N_c\times N_x\times N_y\times N_t}, \tag{7}$$

where N_c is the number of channels. In our case, $N_c = 6$ – permeability, porosity, injection rate, x, y, and time step.

For the training of the FNO, we use the squared L^2-norm. This is an unusual choice for neural networks but a very common metric for PDEs. As neural networks typically map tensors to tensors, the l^2-norm is the most frequent choice. However, since FNOs map functions to functions, we can make use of the squared L^2-norm, which is also a much more appropriate choice when dealing with PDEs. It is worth noting the nuanced difference between the L^2-norm and the l^2-norm, particularly when it comes to implementation. In both cases, you will need to discretize the integral for computational purposes. However, the key distinction lies in the underlying space over which the norms are computed. When using the L^2-norm, one is essentially approximating the integral over a function space, aiming to capture the "true" behavior of the function. On the other hand, the l^2-norm is computed over an Euclidean space, essentially summing up the squared differences in a point-wise manner. Hence, the loss function for the training is:

$$L(\hat{G}, S_{\text{train}}) = \frac{1}{N_s}\sum_{i=1}^{N_s} \|\hat{G}(\boldsymbol{m}_i) - \boldsymbol{d}_i\|_{L^2}^2 + \lambda\|\boldsymbol{\theta}\|_2^2, \quad (\boldsymbol{m}_i, \boldsymbol{d}_i) \in S_{\text{train}}. \tag{8}$$

Both FNO and T-UNet show promise in approximating the high-fidelity forward model, G. T-UNet capitalizes on the strength of 2D convolutional layers and transformer architectures to efficiently incorporate temporal information and spatial parameters. Its design allows for an efficient encoding of multi-dimensional data, making it computationally more lightweight. On the other hand, FNOs offer resolution invariance and the capability to operate directly in function spaces, making them highly suitable for parametric PDE problems. However, one caveat with FNOs is their higher memory consumption. This is

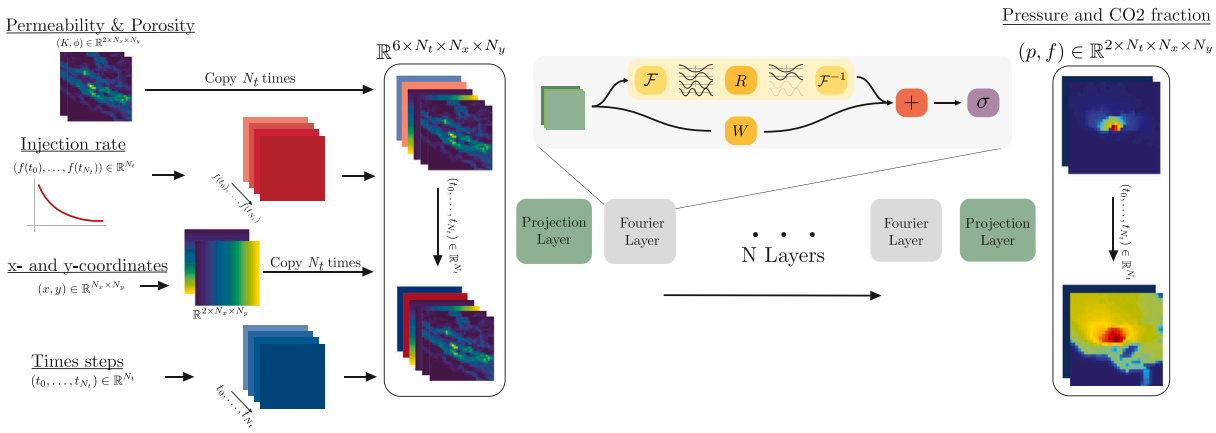


Fig. 5. FNO architecture.

largely due to their need to include time as an additional channel and their use of 3D Fourier transforms and convolutions, which significantly increase the size of input tensors.

The code for both the T-UNet and FNOs architectures is publicly accessible. The repository can be found at https://github.com/nmucke/subsurface-DA-with-generative-models.

4. Data assimilation

4.1. Ensemble Smoother with Multiple Data Assimilation (ESMDA)

Ensemble-based DA techniques are computationally efficient, as they lend themselves well to parallelization. Additionally, these methods offer a level of flexibility by requiring minimal alterations to the existing forward model code and avoiding the computation of adjoints gradients (Evensen et al., 2022). Among various ensemble-based DA techniques, the Ensemble Smoother (ES) serves as an effective method but it has limited ability to provide adequate data matches in complex problems, such as reservoir simulations, due to the application of a single Gauss–Newton correction for conditioning the ensemble to all available data (Emerick and Reynolds, 2013b). To address this, Emerick and Reynolds (2013a) introduced the ESMDA, an iterative version of ES, allowing for improved data matches by assimilating the same data multiple times with an inflated covariance matrix of measurement errors, which enables a more robust approach to updating the model. In spite of its inherent assumptions of Gaussianity, ESMDA is also applicable to weakly non-Gaussian problems. The method is easy to implement, leading to broad application in various scenarios.

Following the notation of Emerick and Reynolds (2013b), we can describe the analysis for each ensemble member's model parameter $\boldsymbol{m}_j^a$ as follows:

$$\boldsymbol{m}_j^a = \boldsymbol{m}_j^f + C_{MD}^f \left(C_{DD}^f + \alpha_i C_D\right)^{-1} \left(\boldsymbol{d}_j - G(\boldsymbol{m}_j)^f\right), \quad \text{for } j = 1, 2, \dots, N_e. \tag{9}$$

Here N_e is the total number of ensemble members, $\boldsymbol{m}_j^a$ and $\boldsymbol{m}_j^f$ represent the analyzed and forecasted parameters of the jth ensemble member, respectively, C_{MD}^f and C_{DD}^f are the cross- and auto-covariance matrices of the model parameters and data in the forecast step, α is a scaling factor, and $\boldsymbol{d}_j$ and $G(\boldsymbol{m}_j)^f$ are the perturbed and forecasted observations for the jth ensemble member. Besides, $G()$ represents the forward model. For more details, see Eq. (10):

$$\boldsymbol{d}_j = \boldsymbol{d}_{\text{obs}} + \sqrt{\alpha} C_D^{1/2} z, \tag{10}$$

In this equation, $\boldsymbol{d}_{\text{obs}}$ is the observed data, C_D is the measurement error covariance, and z is sampled from a standard normal distribution with zero mean and an identity matrix as the covariance. The factor $\sqrt{\alpha}$ scales the perturbations.

The ESMDA algorithm, summarized below, iteratively employs these equations to update each ensemble member. In this study, the model parameter $\boldsymbol{m}$ is defined as the permeability of the medium, and $\boldsymbol{d}_{\text{obs}}$ represents pressure values obtained from specific monitoring points. The ESMDA methodology serves as the basis for a hybrid method that combines DA with surrogate modeling, with the objective of history matching in GCS studies.

Algorithm 1: ESMDA Algorithm

Input : Initial ensemble, observed data $\boldsymbol{d}_{\text{obs}}$, and measurement error covariance C_D

Determine N_a and α for $i = 1, \dots, N_a$

for $i = 1$ *to* N_a **do**

  Compute predicted data $G(\boldsymbol{m}_j)^f$ for each ensemble member

  Generate perturbed observations $\boldsymbol{d}_j$ using Equation (10)

  Update ensemble members $\boldsymbol{m}^a$ using Equation (9)

end for

Output: Compute the final posterior results with DARTS with optimized parameters $\boldsymbol{m}^j$

.

4.2. Randomized Maximum Likelihood (RML)

Randomized Maximum Likelihood is a variational DA method for approximating the posterior pdf with a method introduced in Oliver et al. (1996). As a gradient-based method, it provides the advantage of better convergence and accuracy than an ensemble-based method, within a specified solution space. These benefits come at the cost of computational complexities. The need for adjoint models can be prohibitive and the computational efforts linked to the linearization of the model may introduce errors (García-Pintado and Paul, 2018).

RML employs a set of cost functions, often denoted as $J(\boldsymbol{m}_j)$. These cost functions aim to minimize the discrepancy between the ensemble's forecasted model states and the observed data, as well as the a priori model information. By optimizing these cost functions, RML generates multiple models that are consistent with the available observations, thus aiding in robust UQ. The core idea of RML is to use a set of cost functions, $J(\boldsymbol{m}_j)$, defined as:

$$J(\boldsymbol{m}_j) = (\boldsymbol{m}_j - \boldsymbol{m}_j^{\text{prior}})^T C_{MM}^{-1} (\boldsymbol{m}_j - \boldsymbol{m}_j^{\text{prior}}) + (G(\boldsymbol{m}_j) - \boldsymbol{d}_j)^T C_{DD}^{-1} (G(\boldsymbol{m}_j) - \boldsymbol{d}_j), \tag{11}$$

where $\boldsymbol{m}_j^{\text{prior}}$ and $\boldsymbol{d}_j$ represent the prior model parameters and the perturbed observed data for the j^{th} ensemble member, respectively. These cost functions are designed to produce an ensemble of models that are

coherent with the observed data, thereby assisting in robust DA. For the minimization of the RML cost function $J(\boldsymbol{m}_j)$, we employ the Adam optimizer due to its effectiveness and computational efficiency (Kingma and Ba, 2014).

Applying RML in practice can be computationally intensive as each member of the ensemble necessitates a separate optimization process. Gradient-based optimization methods are commonly used for this purpose, and when adjoints are available for the forward model, they can be applied and increase efficiency. Algorithm 2 describes RML. It takes an initial set of model realizations and observations as inputs. It then computes the relevant covariance matrices and generates variations of the prior model and observed data for each ensemble member. Each ensemble member is then updated by optimizing its respective cost function.

Algorithm 2: RML Algorithm

Input : Initial model set $\boldsymbol{m}_{\text{prior}}$ and perturbed observed data $\boldsymbol{d}_j$
Compute covariance matrices C_{MM} and C_{DD}
Generate variations of the prior model $\boldsymbol{m}_j^{\text{prior}}$ and perturbed observed data $\boldsymbol{d}_j$
for $j = 1, \ldots, N_e$ **do**
  Compute $J(\boldsymbol{m}_j)$ using Equation (11)
  Optimize $J(\boldsymbol{m}_j)$ using gradient-based optimization (e.g. Adam)
  Compute the final posterior results with DARTS with optimized parameters $\boldsymbol{m}^j$
end for
Output: Posterior history matched states for entire ensemble

5. Hybrid data assimilation

5.1. Surrogate-based hybrid ESMDA (SH-ESMDA)

ML surrogate models offer computational efficiency but may compromise robustness when used independently for DA. Conversely, ensemble-based DA methods like ESMDA are known for their accuracy but often come at a high computational cost. To preserve the trade-offs between efficiency and robustness, we introduce a surrogate-based hybrid approach called SH-ESMDA, which is within the "Data Learning" paradigm introduced by Buizza et al. (2022). A key feature facilitating this integration is the use of the same parameters as input for both the ML algorithm and the reservoir simulator, allowing the surrogate model to serve as a direct substitute for the forward model in the intermediate steps of ESMDA. It is important to note that the surrogate model is only used in the intermediate steps of the ESMDA process. This ensures that the high-fidelity forward model is leveraged for both the initial and the computation of the posterior step, resulting in a more robust and accurate assimilation process. The surrogate model for SH-ESMDA requires training with only the number of forward simulations typically necessary for running the prior in standard approaches. This feature ensures no additional computational burden beyond conventional methods.

The development of SH-ESMDA has the primary objective of accelerating the ESMDA procedure, not achieving more accurate history-matching results than conventional methods. This is due to the fact that in this scheme, ESMDA is still the core DA method, so this hybrid approach also will keep its limitations in terms of DA. The main gain is the acceleration of the process, which potentially enables more iterations, which might otherwise be computationally prohibitive if relying solely on high-fidelity simulations. To achieve this acceleration, the following steps are proposed:

1. **Prior Dataset Generation**: Generate a prior dataset consisting of channelized permeability models with Alluvsim. Subsequently, perform CO_2 injection simulations on these models using the DARTS simulator.
2. **Surrogate Model Training**: Train a surrogate model, such as a FNO or T-UNet, on the generated dataset. This training can be conducted as an offline stage, allowing the pretrained model to be reused in multiple subsequent Hybrid-ESMDA-Surrogate runs, thereby obviating the need for repetitive training and enhancing computational efficiency.
3. **Initial Analysis Step**: Compute the first analysis step with ES-MDA using prior forecasts.
4. **Intermediate ESMDA Steps**: Employ the trained surrogate model as a substitute for DARTS in the intermediate steps of the ESMDA process. This offers a computationally efficient approximation to the solution.
5. **Posterior**: Incorporate simulations from DARTS to compute the posterior to refine the solution and compute the final posterior states.

The key innovation is leveraging the efficiency of the trained surrogate model to handle the computationally intensive ESMDA iterations. The final step with DARTS simulations acts as a physics-based regularizer to enhance robustness. Another significant advantage of a SH-ESMDA is the capability to pre-train the surrogate model in an offline stage. Once trained, this model can be reused across multiple SH-ESMDA runs without the need for retraining, thereby providing an additional layer of computational efficiency. This feature is particularly beneficial when dealing with a series of similar scenarios, as it eliminates the need to undergo the training process before each new ESMDA run. Consequently, this enables more frequent and rapid iterations, further enhancing the overall efficiency of the DA process. This SH-ESMDA approach is summarized in the algorithm below:

Algorithm 3: Hybrid-ESMDA-Surrogate Algorithm

Input : Initial ensemble, observed data $\boldsymbol{d}_{\text{obs}}$, measurement error covariance C_D, and trained surrogate model
Determine N_a and α for $i = 1, \ldots, N_a$
for $i = 1$ *to* N_a **do**
  if $i = 1$ **then**
    Compute the prior using DARTS
  else
    Use surrogate model to compute predicted data $\hat{G}(\boldsymbol{m}_j)^f$ for each ensemble member
  end if
  Generate perturbed observations $\boldsymbol{d}_j$ using Equation (10)
  Update ensemble members $\boldsymbol{m}^a$ using Equation (9)
end for
Output: Compute the final posterior results with DARTS with optimized parameters $\boldsymbol{m}^a$

5.2. Surrogate-based hybrid RML (SH-RML)

In a similar manner to SH-ESMDA, we introduce a surrogate-based hybrid (SH-RML), a method that integrates ML surrogates into the RML variational framework. One of the critical features that make this integration possible is the use of consistent permeability parameterization for both the surrogate and the high-fidelity DARTS model. This uniformity allows for seamless transitions between the surrogate and the physics-based models during the optimization process. SH-RML employs a streamlined approach similar to SH-ESMDA, where the surrogate model training occurs in the same fashion. It requires only as many forward simulations as are typically necessary in the prior phase of conventional methods. This ensures the method's computational efficiency by aligning with the standard simulation demands, and avoids the need for additional computational resources.

A particular feature of this method is its ability to allow variational DA, even in cases where simulators lack adjoint capabilities. This capability is achieved through the computation of gradients using automatic

differentiation from a neural network. The primary objective of the SH-RML approach is to facilitate the optimization of the RML cost function $J(\boldsymbol{m}_j)$. The algorithmic flow of the SH-RML is as follows:

1. **Prior Dataset Generation**: Generate a prior dataset consisting of channelized permeability models with Alluvsim. Subsequently, perform CO_2 injection simulations on these models using the DARTS simulator.
2. **Surrogate Model Training**: Train a surrogate model, such as a FNO or T-UNet, on this dataset. This training can be conducted as an offline stage, allowing the pre-trained model to be reused in multiple subsequent SH-RML runs.
3. **Parameter Initialization**: Initialize the permeability parameters $\boldsymbol{m}_j$ for each ensemble member within the RML framework.
4. **Initial Optimization**: Use gradients derived from the surrogate model to perform initial optimization of the cost function $J(\boldsymbol{m}_j)$.
5. **Posterior Parameter Computation**: Compute the posterior parameters after the surrogate-based optimization using RML.
6. **High-Fidelity Refinement**: Apply these posterior parameters to the DARTS model, running high-fidelity simulations to refine the solution and compute the final posterior states.

The initial steps of the RML optimization process are accelerated by leveraging the surrogate model, which offers both efficiency and automatic differentiation capabilities. The final steps employ the DARTS model to ensure high-fidelity, physics-based solutions as summarized in the algorithm below:

Algorithm 4: Hybrid-RML-Surrogate Algorithm

Input : Initial model set $\boldsymbol{m}_{\text{prior}}$, perturbed observed data $\boldsymbol{d}_j$, and trained surrogate model

Compute covariance matrices C_{MM} and C_{DD}

Generate variations of the prior model $\boldsymbol{m}_j^{\text{prior}}$ and perturbed observed data $\boldsymbol{d}_j$

for $j = 1, \ldots, N_e$ **do**

  Use surrogate to compute initial $J(\boldsymbol{m}_j)$ and gradients

  Perform optimization of $J(\boldsymbol{m}_j)$ using gradient-based methods

end for

Output: Compute the final posterior results with DARTS with optimized parameters $\boldsymbol{m}_j$

One of the key innovations in SH-RML is the utilization of automatic differentiation capabilities provided by the surrogate model. This eliminates the need for manually deriving computationally expensive adjoint models, which are required in traditional variational DA methods. As a result, the SH-RML offers an efficient approach to DA in complex, nonlinear systems. However, it is important to acknowledge the inherent limitations due to the intrinsically ill-posed nature of the problem, affecting the overall results.

In summary, the proposed SH-ESMDA and SH-RML frameworks offer solutions for enhancing DA techniques in applications such as GCS. By adding the computational advantages of ML surrogates with the reliability of physics-based models, these hybrid methods pave the way for efficient and accurate DA and, ultimately, a better understanding and quantification of the model and data uncertainties.

6. Results

In this section, we present a comprehensive evaluation of surrogate models and history-matching methods for GCS applications. We begin with the training and evaluation of FNO and T-UNet as surrogate models. This is followed by an assessment of the ESMDA method for history matching. Subsequently, we evaluate the SH-ESMDA. Finally, we discuss the SH-RML results.

6.1. Training and evaluation of the FNO and T-UNet

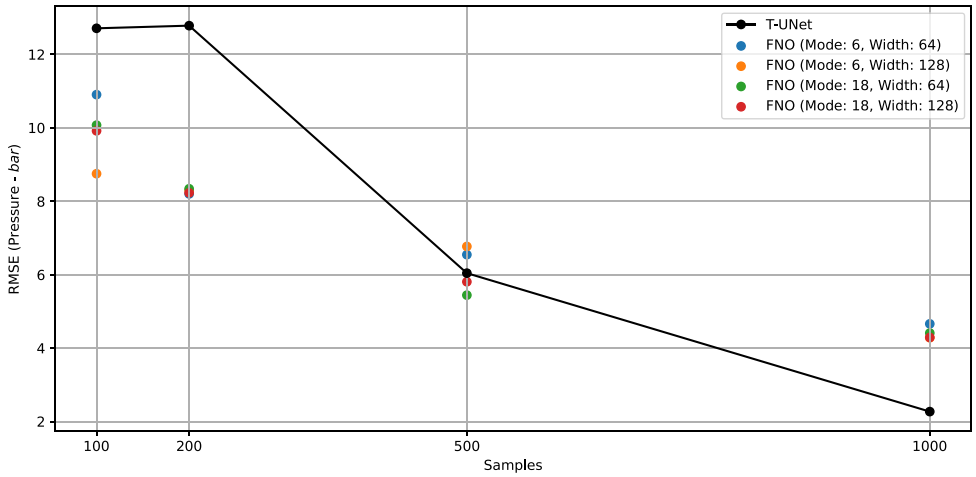
In this Section, the focus is on training the T-UNet and the FNO to serve as surrogate models that approximate the high-fidelity forward model G. As described in Section 3, these models are trained on a dataset derived from high-fidelity reservoir simulations generated with the DARTS simulator. This dataset includes critical unknowns such as permeability, porosity, and gas injection rates, along with corresponding state variables, pressure and CO_2 molar fraction. Given the developing state of understanding of optimal FNO configurations, we particularly investigated the impact of varying Fourier modes, considering both modes 18 and 6, and hidden channel widths of 128 and 64. This nuanced examination aims to contribute to the open question of how to best configure FNOs for subsurface modeling tasks.

To quantify the performance of the T-UNet and FNO surrogate models, we RMSE as a metric, focusing on both pressure and CO_2 molar fraction. Specifically, we consider training plus test sample sizes of 100, 200, 500, and 1000 to understand how the size of the training set affects the model's performance. Training and test data are split into 80% and 20%, respectively. The RMSE values presented as a function of the number of training samples in Fig. 6, and detailed in the Appendix, indicate that both neural network architectures yield accurate approximations of the high-fidelity forward model G. However, the FNO shows a slight advantage when the number of training samples is limited, particularly at the 100-sample size. This is a significant observation for subsequent data assimilation studies, as it suggests that FNO may require fewer training samples than the T-UNet, thus alleviating the need for additional high-fidelity simulations for neural network training. For context, it is important to note that the pressure range in the reservoir simulations is between 200 and 320 bars, and the CO_2 molar fraction varies from 0 to 1. In this range, the RMSE values indicate that the approximation errors are significantly small. However, these are still approximations and, although they are highly accurate, they cannot entirely replicate the high-quality DARTS simulations.

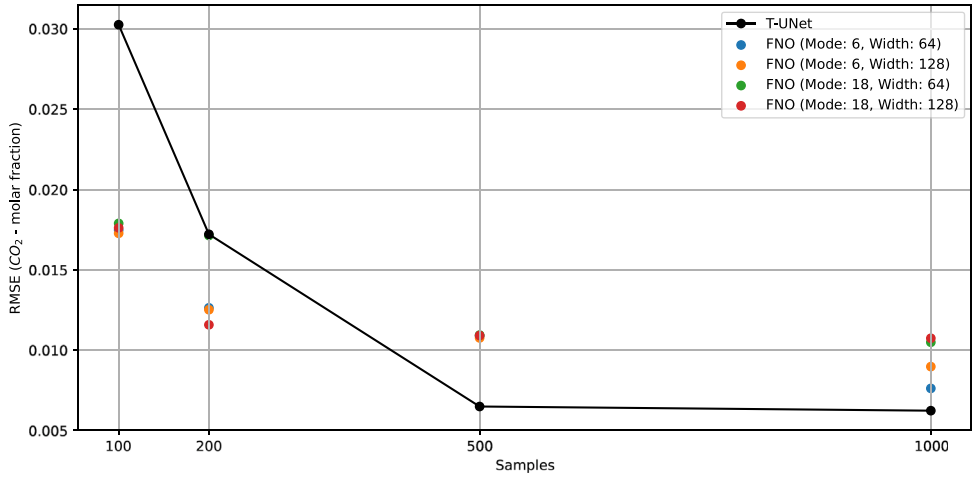
We also examine the capability of FNO and T-UNet to represent the spatial distributions of pressure and CO_2 molar fraction. Figs. 7 and 8 present the contour maps of CO_2 distribution for a specific test case (not part of the training dataset). At the final time step, we compare the true state variables against those predicted by FNO and T-UNet. Those are responses for the models with a sample size of 1000. Both models capture the overall behavior of the reservoir and what is particularly noteworthy is the ability of both models to approximate the shape of the CO_2 plume. We analyze the models' time evolution in addition to comparing their spatial distributions of pressure and CO_2 molar fraction at a specific point in the grid. Fig. 9 illustrates the temporal variations at the injection point located at the grid position (16,16). These figures show the predictive capabilities of both FNO and T-UNet in capturing dynamic behavior, as compared to the high-fidelity DARTS simulations. One observation that stands out is the relative smoothness in the time evolution generated by the FNO model, especially when compared to the more fluctuating curves from the T-UNet model. This difference is intriguing, and it prompts further discussion.

The FNO's less noisy, and thus more physically plausible, representation of the temporal evolution at the injection point may be attributed to the way it handles time steps as additional channels of data. This treatment allows the FNO to account for interdependencies across time, thus yielding a smoother, more coherent dynamic response. On the other hand, the T-UNet model generates time evolution independently from each other through integration with the transformer, as discussed in Section 3. This approach seems to result in a more noisy representation of the pressure, possibly because the temporal dependencies are not as explicitly captured as in the FNO model.

The difference in noise levels between the FNO and T-UNet models might not just be a matter of the numerical accuracy of the methods, but could also have implications for their respective usefulness in

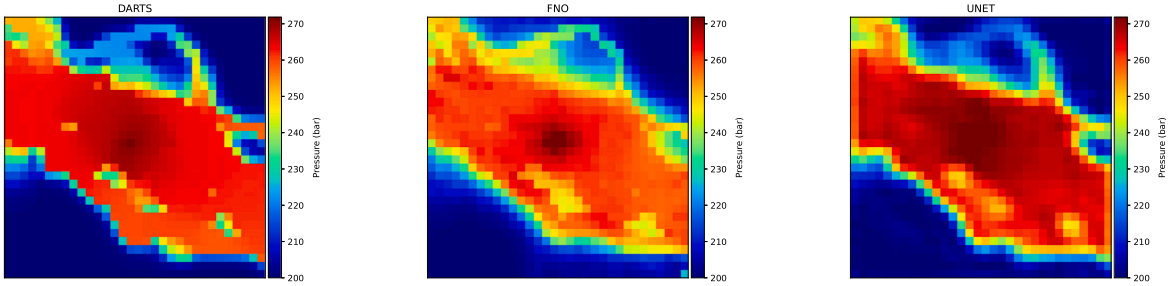


(a) Test RMSE Metrics for Pressure



(b) Test RMSE Metrics for CO_2 molar fraction

Fig. 6. Test RMSE metrics for FNO and T-UNet.



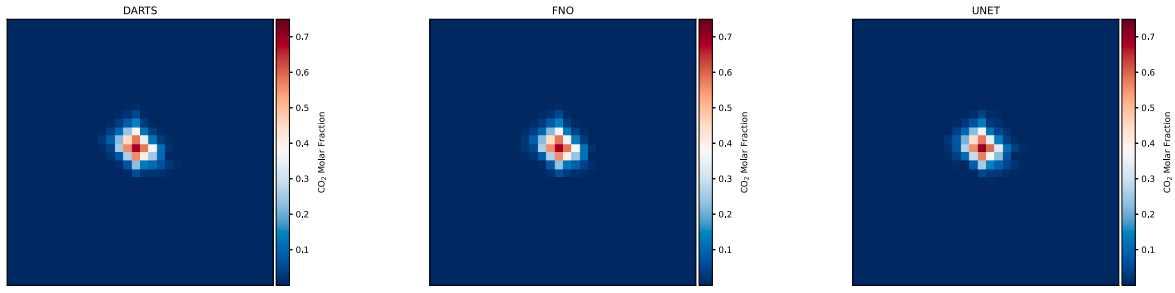
(a) DARTS Pressure (b) FNO Predicted Pressure (c) T-UNet Predicted Pressure

Fig. 7. Pressure distributions for a test case.

subsurface modeling and DA studies. The smoother temporal response of the FNO model may make it more suitable for cases that require a higher accuracy for the physical representation of the states. In summary, the ability of these models to adequately represent both the spatial and temporal complexities of the reservoir suggests their robustness and reliability for further analysis and DA studies. However, the observed differences in dynamic behavior between the FNO and T-UNet models could be a crucial factor in deciding which model to employ for specific applications.

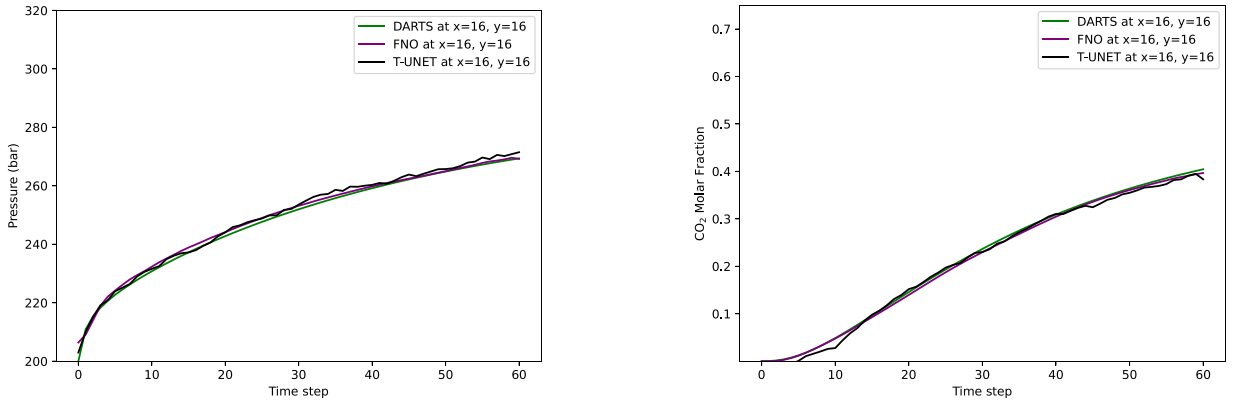
6.2. ESMDA history matching

ESMDA is employed to conduct history matching on our reservoir model described in Section 2. The primary objective is to evaluate ESMDA's efficiency and accuracy in modeling complex reservoir systems for this GCS application. The task at hand poses challenges due to the nonlinearity of the problem and the non-Gaussian nature of the reservoir's permeability field. To start our assessment, we utilized a distinct reference permeability model, generated outside our prior



(a) DARTS CO_2 molar fraction (b) FNO Predicted CO_2 molar fraction (c) T-UNet Predicted CO_2 molar fraction

Fig. 8. CO_2 distributions for a test case.



(a) Time evolution of Pressure at grid (16,16) (b) Time evolution of CO_2 molar fraction at grid (16,16)

Fig. 9. Comparison of time evolution of states at grid (16,16) for DARTS, FNO and the T-UNet.

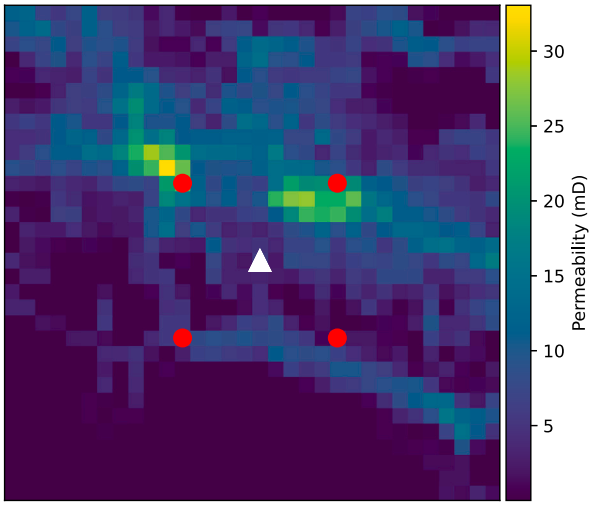


Fig. 10. Reference permeability model showcasing the central injection well (triangle) and peripheral monitoring points (circles). Monitoring points are numbered from 1 to 4, starting from the top-left corner and proceeding clockwise.

distribution, to produce synthetic observed data. This model, displayed in Fig. 10, incorporated a central injection well, surrounded by four pressure monitoring points.

The ESMDA approach utilizes an ensemble of 100 prior permeability maps to represent the uncertainty before DA. These maps are drawn from geological realizations using the Alluivsim algorithm. Over 4 ES-MDA steps, the results indicate that the posterior pressure distribution at each of the four monitoring points is closer to the observed data compared to the prior. Fig. 11 displays this improvement, comparing pressures from the prior, the reference model, and the posterior. We further analyze the sensitivity of the method to the number of steps in ESMDA at the monitoring pressure points for both the prior and posterior models by varying the number of ESMDA steps: 4, 8, 16, and 32. Fig. 12 reveals that increasing the number of iterations does not substantially improve the quality of history matching.

Table 1
Uncertainty reduction in P10-P90 pressure range at monitoring points for the ESMDA history matching with 4 iterations.

Location	Prior (bar)		Posterior (bar)	
	P10-P90	Difference	P10-P90	Difference
1	213.9–268.8	54.9	212.6–261.1	48.5
2	202.7–261.2	58.4	211.0–259.9	48.9
3	209.9–265.1	55.2	211.5–260.2	48.7
4	208.7–264.0	55.3	208.8–259.0	50.2

In terms of uncertainty quantification, ESMDA narrowed the range of uncertainty in pressure estimations. This is visualized in Fig. 13, where the histograms demonstrate a post-ESMDA variance reduction of the width of the pressure distribution at the final injection step, as compared to the broader spread of the prior. The observed pressure values, indicated by the red dashed line, are now more centrally situated within the tightened posterior distribution, signifying how the observations reduce the uncertainty of pressure buildup within the reservoir. The true pressures observed at the monitoring points are as follows: 255.1 bar at Monitoring Point 1, 252.0 bar at Monitoring Point 2, 255.0 bar at Monitoring Point 3, and 251.4 bar at Monitoring Point 4. The P10-P90 pressure range, which serves as an indicator of uncertainty, is reduced after the history matching process, as quantified in Table 1.

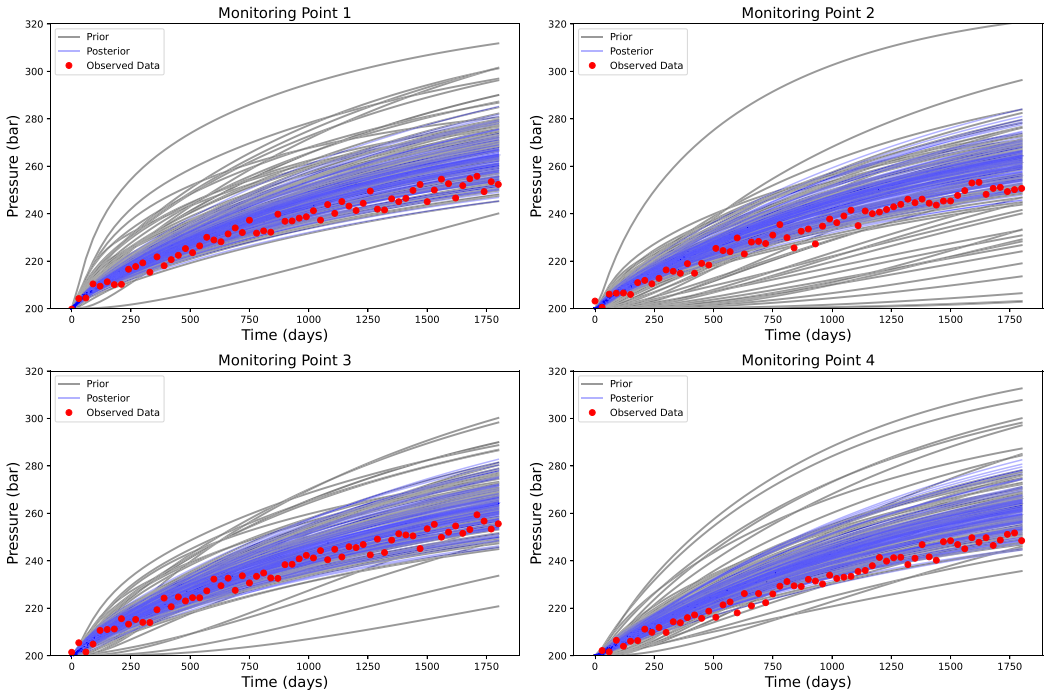


Fig. 11. Comparison between the prior, reference model, and posterior pressures at each monitoring point for the ESMDA history matching. Red dots represent a realization of perturbed observed data.

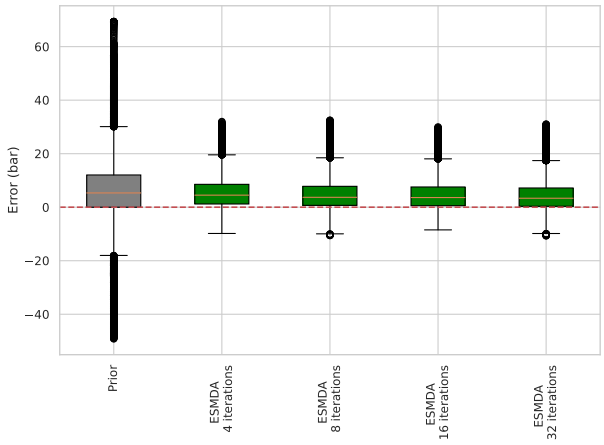


Fig. 12. Comparison of monitoring pressure absolute error across different ESMDA iterations and the prior. Red dots represent a realization of perturbed observed data.

Although ESMDA reduces errors related to measured pressure in comparison to the prior, it significantly overestimates reservoir permeability in comparison to prior permeability distributions, as illustrated in Fig. 14. While undesirable, the discrepancy can be explained by the fact that history matching is an ill-posed problem, allowing for multiple solutions that can satisfactorily fit the data.

Given the inherent assumptions of ESMDA, namely, its reliance on Gaussian distributions and its better suitability for linear problems—none of which are present in our case—it is crucial to recognize its limitations in addressing the ill-posed problems we encounter. Our subsequent results delve into hybrid methods, as detailed in Section 5. These techniques serve dual purposes: one aims to accelerate the computational process, while the other focuses on enhancing the accuracy of history matching.

Table 2
Uncertainty reduction in P10-P90 pressure range at monitoring points for the SH-ESMDA history matching with 4 iterations.

Location	Prior (bar)		Posterior (bar)	
	P10-P90	Difference	P10-P90	Difference
1	213.9–268.8	54.9	212.96–262.1	49.14
2	202.7–261.2	58.4	210.93–260.8	49.87
3	209.9–265.1	55.2	211.65–261.0	49.39
4	208.7–264.0	55.3	209.10–260.1	50.97

6.3. Results for SH-ESMDA

In order to accelerate the computational efficiency of the standard ESMDA, we employ ML surrogates in the form of FNO and T-UNet following the algorithm outlined in Section 5.1. For this study, we employ 100 samples and conduct history matching for pressure at four monitoring points over four iterations. Figs. 15 and 16 display the results of these history-matching exercises. In this context, it should be highlighted that the larger-scale training experiments involving additional models were designed for performance benchmarking of SH-ESMDA method. These figures reveal the performance when FNO and T-UNet are used as surrogate models for the intermediate steps in the ESMDA algorithm, respectively. As can be observed, the outcomes produced by both FNO and T-UNet are remarkably similar to each other and closely align with those obtained using the standard ESMDA methodology with 4 steps.

Fig. 17 presents the histograms of pressure at the last time step of injection for SH-ESMDA using the FNO surrogate model with 4 iterations, illustrating a variance reduction of the pressure distributions around the observed values at each monitoring point, results similar to what was achieved with ESMDA. This is quantified further in Table 2, which presents the P10-P90 pressure range before and after applying SH-ESMDA with FNO and 4 iterations. The reduction in uncertainty

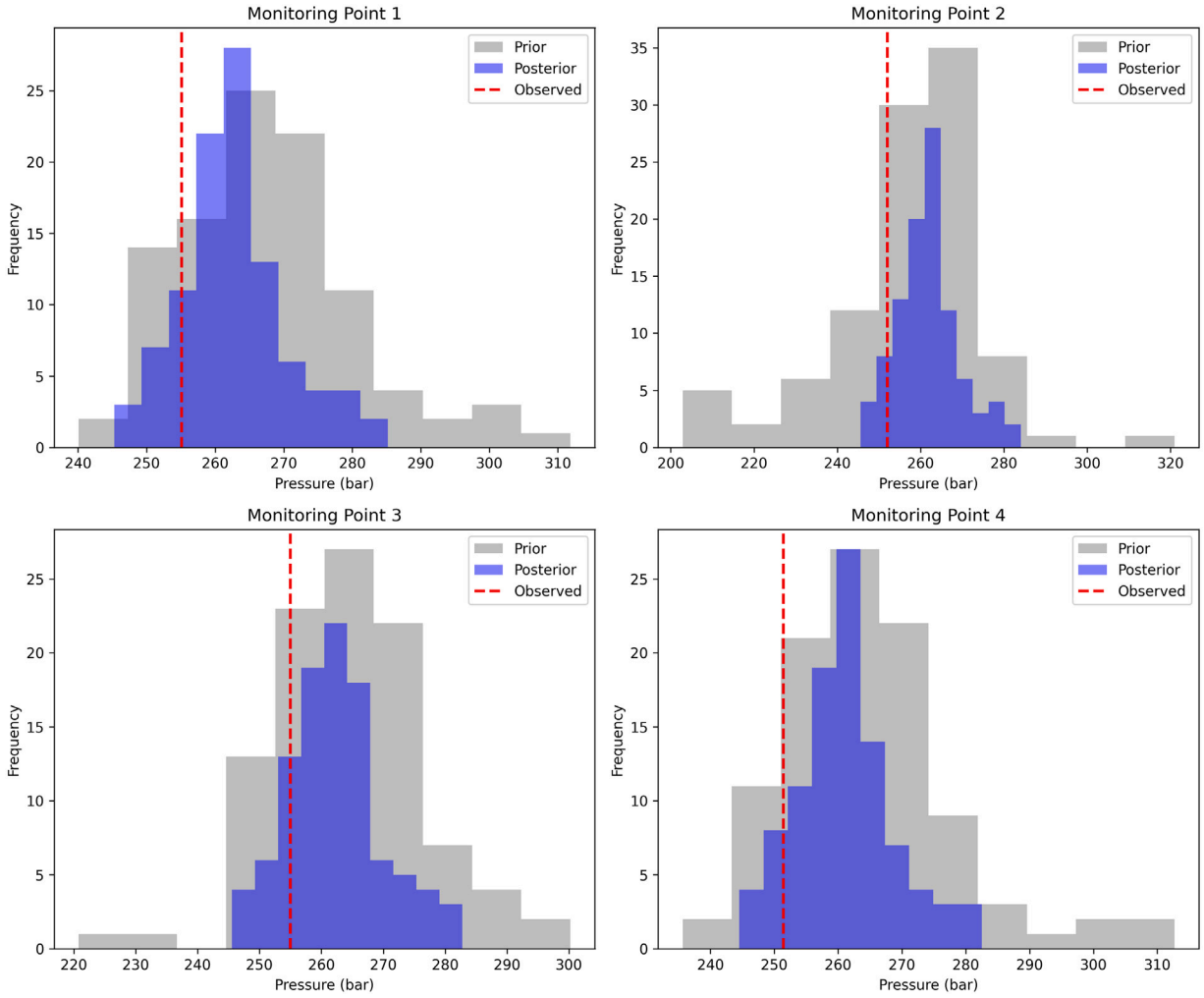


Fig. 13. Histograms of prior and posterior pressure distributions at the last time step of injection for each monitoring point for the ESMDA history matching with 4 iterations. The observed reference pressures are marked with a red dashed line.

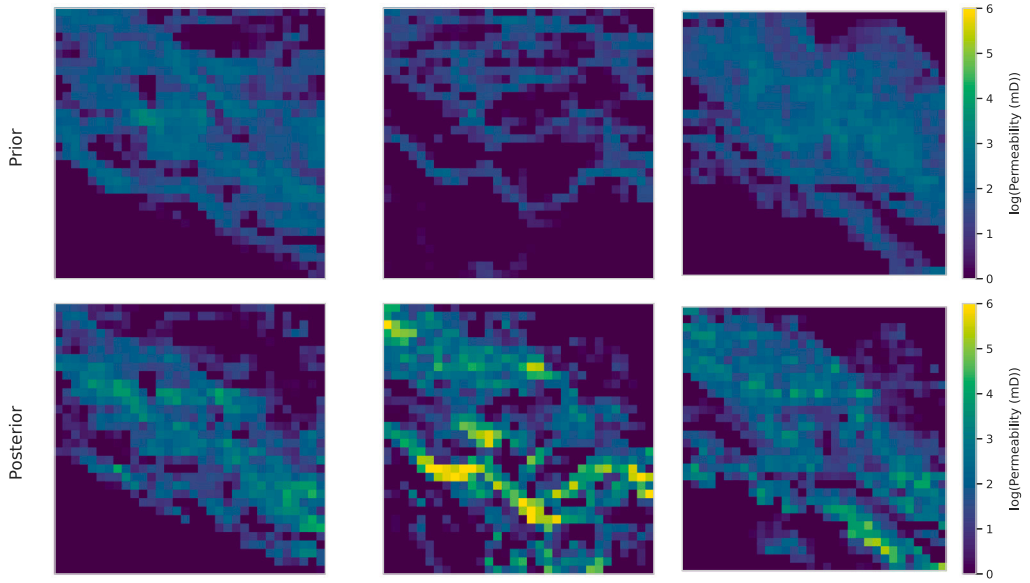


Fig. 14. Comparison between prior and posterior permeabilities across three different samples for the ESMDA.

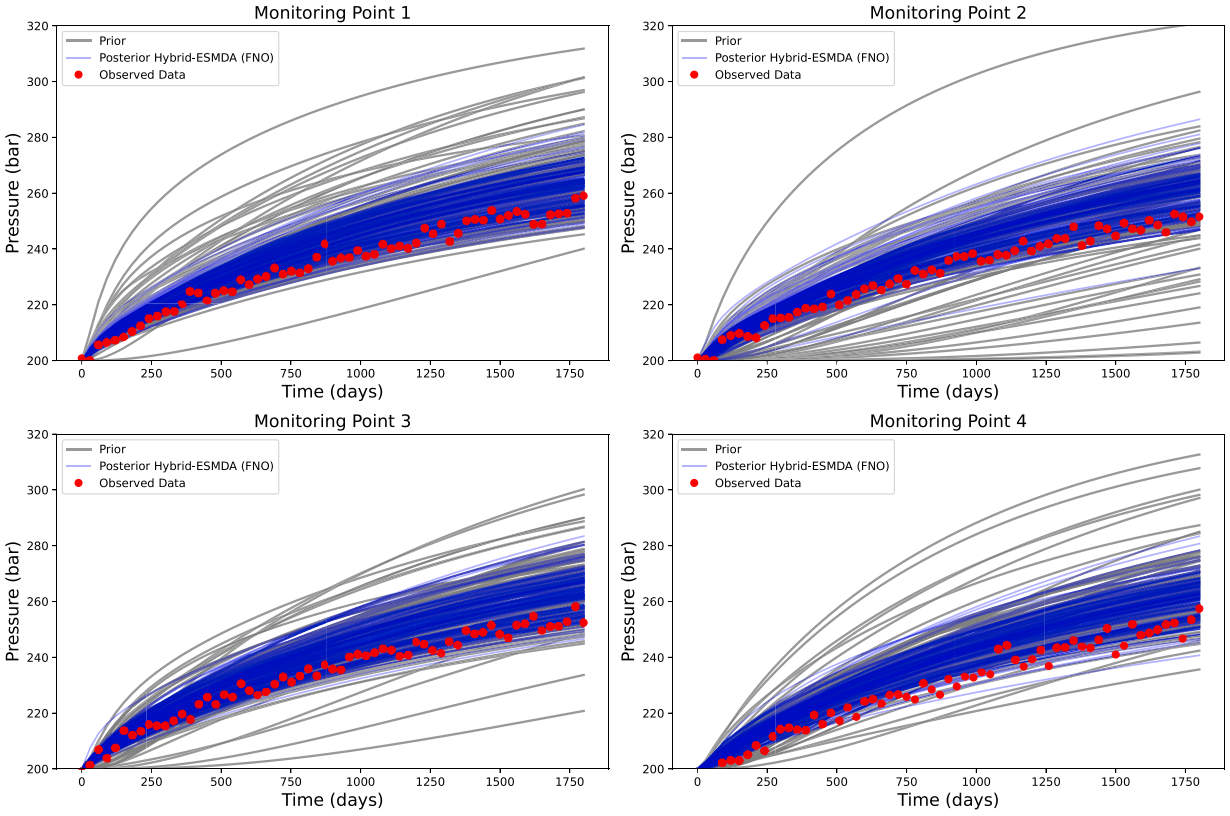


Fig. 15. Comparison between the prior, reference model, and posterior pressures at each monitoring point for the Hybrid-ESMDA-Surrogate history matching using FNO. Red dots represent a realization of perturbed observed data.

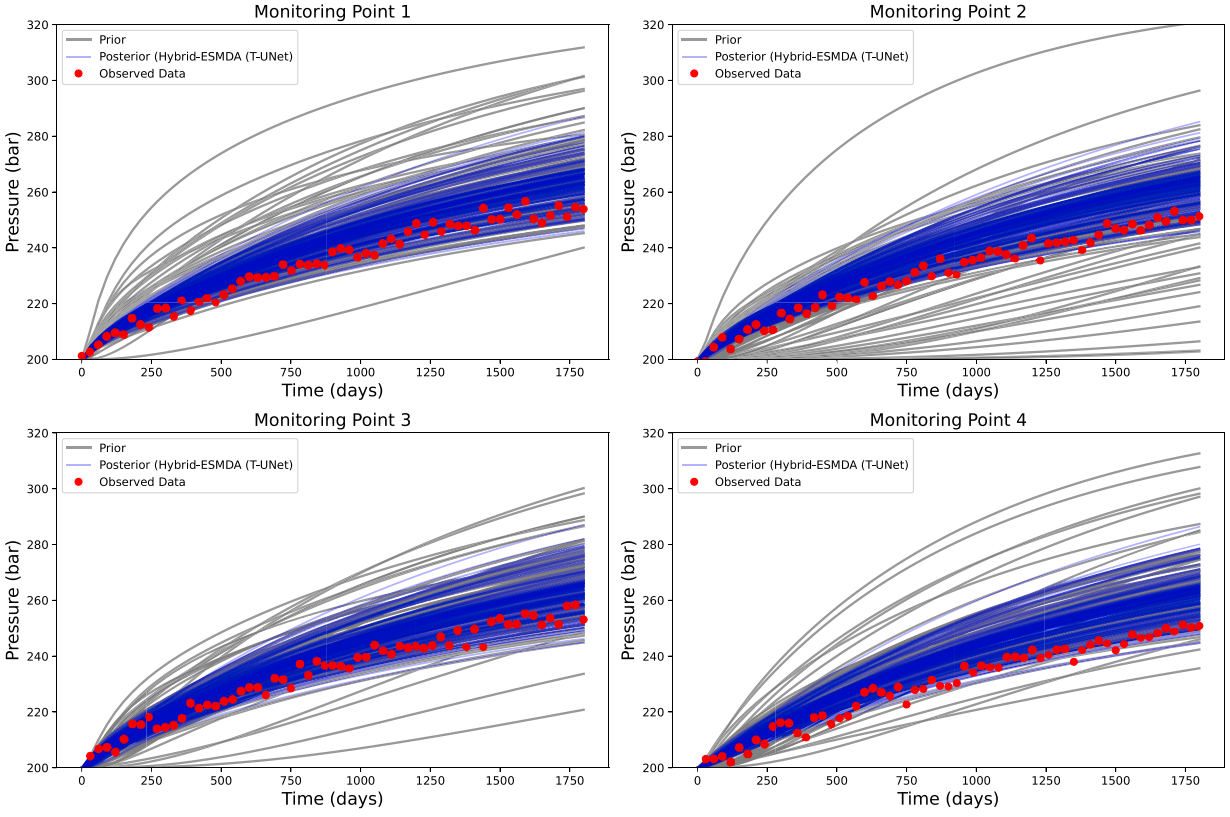


Fig. 16. Comparison between the prior, reference model, and posterior pressures at each monitoring point for the SH-ESMDA history matching using T-Unet. Red dots represent a realization of perturbed observed data.

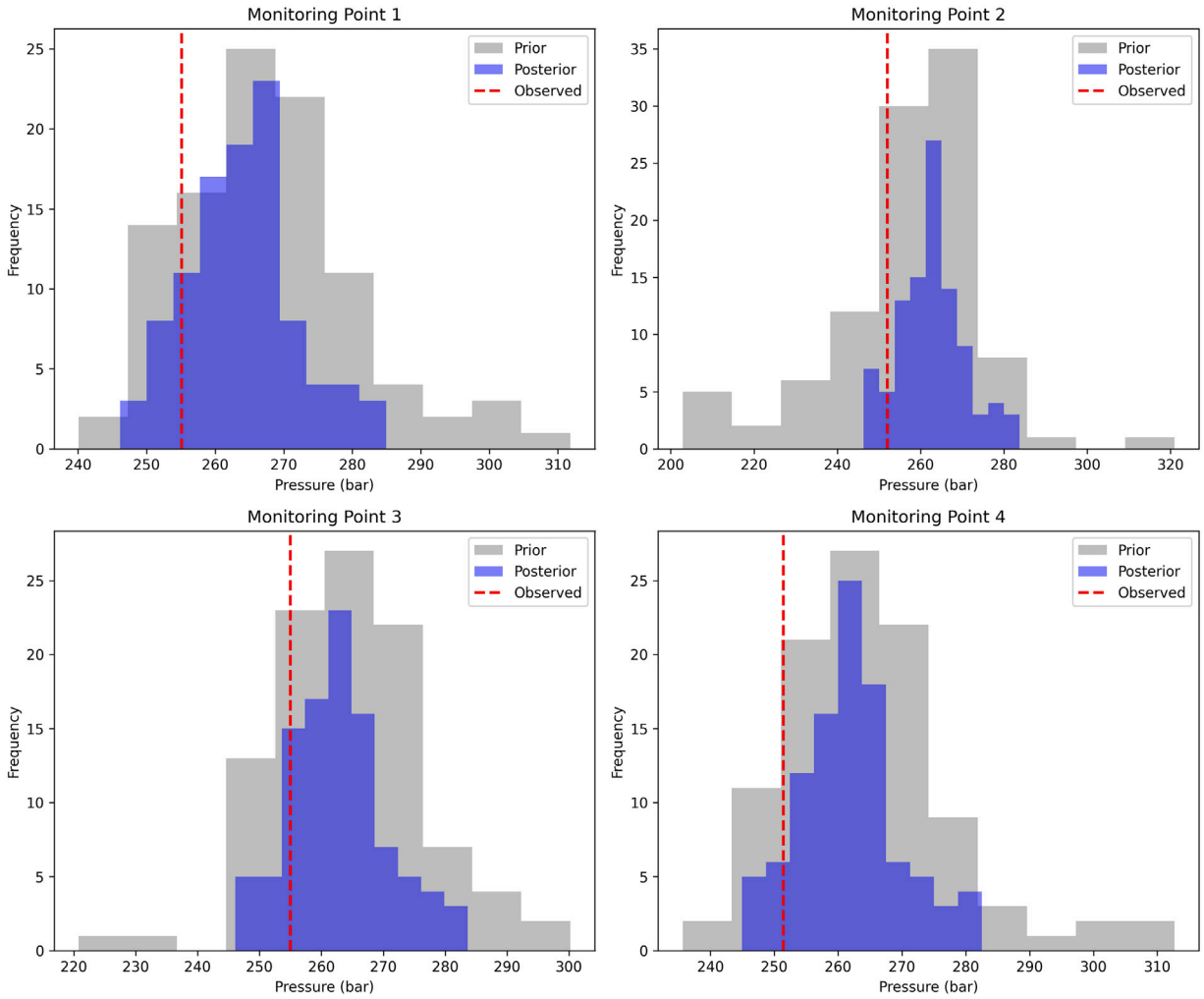


Fig. 17. Histograms of prior and posterior pressure distributions at the last time step of injection for each monitoring point for the SH-ESMDA history matching with 4 iterations. The observed reference pressures are marked with a red dashed line.

ranges from 5.8 bars at Monitoring Point 1 to 3.9 bars at Monitoring Point 4.

Fig. 18 presents box plots that offer a comparison of the monitoring pressure errors for both the standard ESMDA and the enhanced SH-ESMDA methodology, employing FNO and T-UNet as surrogates. The SH-ESMDA methods maintain a level of accuracy that is comparable to that of the standard ESMDA, while significantly reducing the computational time required. When the T-UNet model is employed as the surrogate, the errors observed are marginally higher compared to when the FNO is used. It is also crucial to highlight that, consistent with the observations made for the standard ESMDA, increasing the number of iterations does not result in a substantial reduction in errors.

The gains in computational efficiency are enumerated in Table 3. As the table illustrates, both versions of the SH-ESMDA method—utilizing either FNO or T-UNet as surrogates significantly reduce computation time. Specifically, there is a minimum speedup of around 50%. In all experiments, the ESMDA involved 4 steps. The computational times reported in Table 3 do not include the time required to train the FNO and T-UNet models. The exclusion is justified for two main reasons. First, the training of these neural network models is generally considered a one-time computational expense. For the configuration used in our hybrid methods—comprising 100 samples for both FNO and T-UNet, along with 6 models and a width of 64—both models can be trained in less than 30 min. Once trained, these models can be reused for multiple iterations or different scenarios without the need for retraining. This is particularly beneficial in applications where the same or similar systems are studied multiple times. Second, the training can be performed offline and in parallel, taking advantage of high-performance computing resources. This minimizes its impact

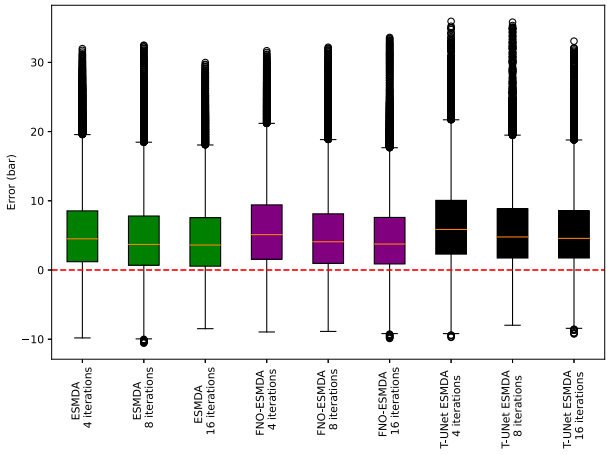


Fig. 18. Box plots comparing monitoring pressure errors between standard ESMDA (green) and SH-ESMDA using FNO (purple) and T-UNet (black) surrogates for 4, 8 and 16 iterations.

Table 3
Comparison of computational times for different methods, considering 100 ensemble members.

Total number of steps	Methods		
	ESMDA (min)	Hybrid-ESMDA-Surrogate (FNO) (min)	Hybrid-ESMDA-Surrogate (T-UNet) (min)
4	73	39	38
8	149	46	44
16	303	58	55

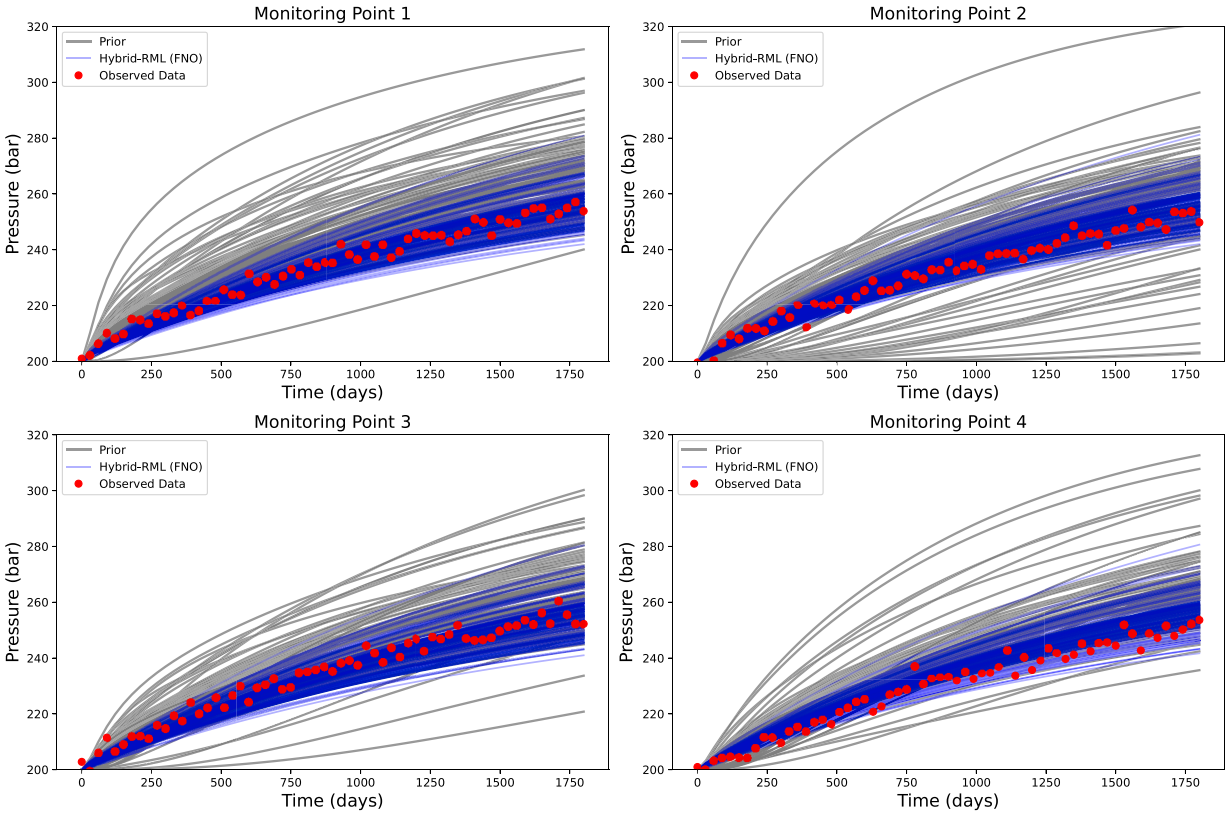


Fig. 19. Comparison between the prior, reference model, and posterior pressures at each monitoring point for the Hybrid-RML-Surrogate history matching using FNO. Red dots represent a realization of perturbed observed data.

on the overall computational efficiency when amortized over multiple applications.

The posterior permeability distributions obtained through both the SH-ESMDA schemes employing FNO and T-UNet closely align with those achieved using the standard ESMDA approach. Given that ESMDA serves as the core method for DA in these hybrid algorithms, the permeability distributions inherently reflect the strengths and limitations of ESMDA itself. Therefore, the hybrid schemes do not necessarily bring about a qualitative shift in the outcome; rather, their primary advantage lies in computational speedup. We acknowledge that the hybrid methods will inherit the constraints of ESMDA, and its inadequacies in encapsulating the uncertainties inherent in the system. For brevity and to avoid redundancy, we refrain from presenting additional figures showcasing the similarity in the posterior permeability distributions across the different methods, as they would closely mirror the results already discussed for ESMDA.

In summary, SH-ESMDA successfully combines the computational efficiency of ML with the reliability of DARTS like ESMDA. The approach reduces the time needed for each computational step, allowing for less time for the same amount of iterations. However, the SH-ESMDA inherits the fundamental weaknesses of the core ESMDA algorithm. Therefore, although the method provides substantial acceleration, the accuracy of the history matching is still constrained by the inherent limitations of ESMDA. This makes this SH-ESMDA scheme an important tool for accelerating DA studies, especially for complex reservoir models.

6.4. Results for SH-RML

As elaborated in Section 5.2, SH-RML offers a combination of factors that make it an effective choice for accurate DA for complex problems. Here, we explore further its comparative results over other methods like ESMDA. Our findings show that the SH-RML improves history matching and provides more accurate posterior permeabilities estimates. For a comprehensive perspective on the computational requirements of SH-RML, it is pertinent to discuss the computational time in the SH-RML. Both FNO and T-UNet models were subjected to 200 steps each for gradient evaluation in the optimization process for each one of the 100 prior models, and the full reservoir simulations with DARTS were also taken into account for the prior and the posterior curves. For SH-RML, it should be noted that the extensive training experiments, which included up to 1000 models, were primarily focused on performance benchmarking. These experiments are not a requirement for the framework but serve to showcase its adaptability and effectiveness in diverse application scenarios. The overall history-matching process requires approximately 8 h, which is justified by the enhanced accuracy and reliability of the estimates obtained. Nonetheless, the optimization time could potentially be reduced by employing a more tailored approach to

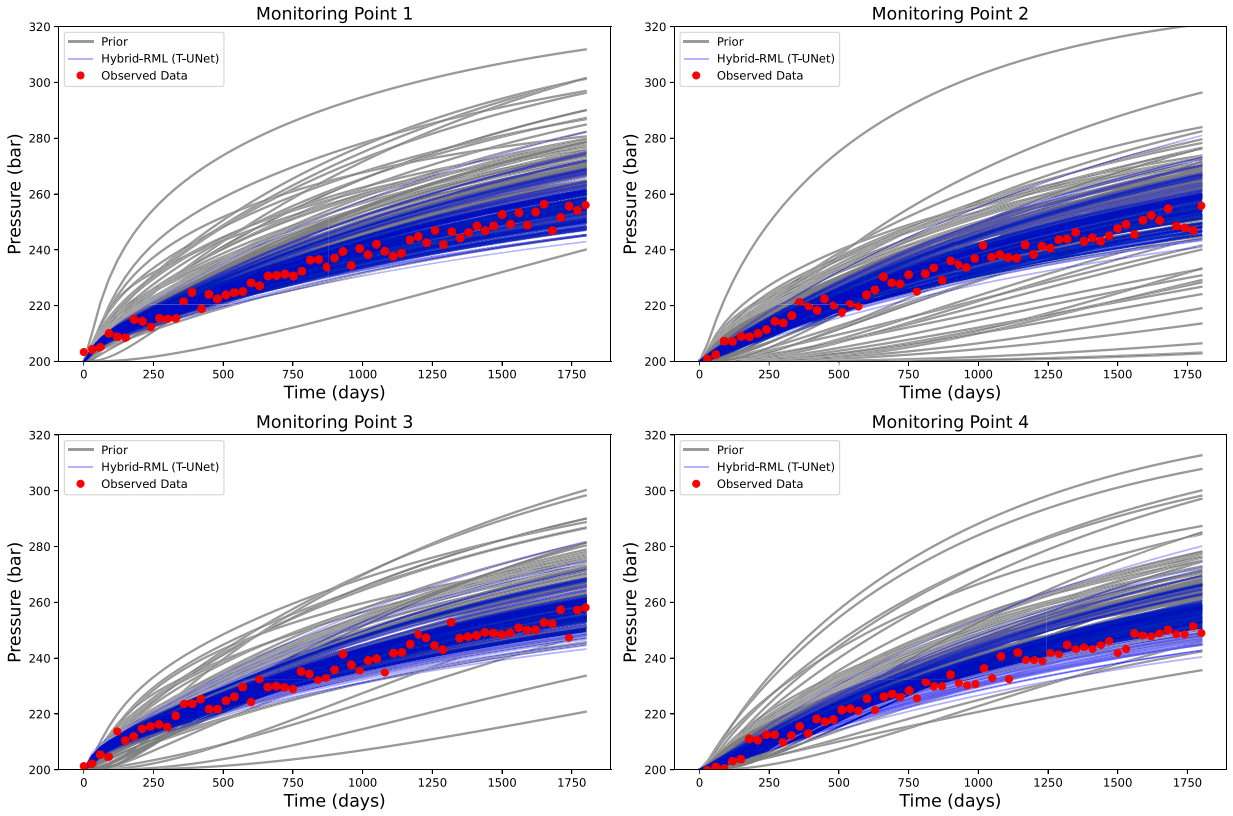


Fig. 20. Comparison between the prior, reference model, and posterior pressures at each monitoring point for the Hybrid-RML-Surrogate history matching using T-UNet. Red dots represent a realization of perturbed observed data.

Table 4
Uncertainty reduction in P10-P90 pressure range at monitoring points for SH-RML with FNO.

Location	Prior (bar)		Posterior (bar)	
	P10-P90	Difference	P10-P90	Difference
1	213.9–268.8	54.9	209.0–253.9	44.9
2	202.7–261.2	58.4	208.7–253.7	45.0
3	209.9–265.1	55.2	208.2–253.5	45.3
4	208.7–264.0	55.3	208.2–253.5	45.3

the cost function optimization without significantly compromising the quality of the results.

For SH-RML, we noticed an improvement in history matching with posterior pressure closer to the values of the observations. Figs. 19 and 20 show the pressure matching results at each monitoring point for SH-RML with FNO, and SH-RML with T-UNet, respectively. While surrogates are employed for gradient evaluations, the final posterior pressure curves are computed using the full reservoir simulator (DARTS). This allows SH-RML to leverage the efficiency of surrogates during optimization while still generating high-fidelity pressure forecasts with the simulator for final uncertainty quantification. Fig. 21 displays box plots comparing monitoring pressure errors between the standard ESMDA and SH-RML methods. The results from SH-RML exhibit a better balance around zero when compared to ESMDA. Additionally, Fig. 23 demonstrates that the posterior permeabilities derived from SH-RML, employing FNO are consistent with the prior. In contrast to ESMDA, RML provides permeabilities within a similar range as the prior. For conciseness and to avoid repetition, we do not include additional figures showcasing the posterior permeabilities using the T-Unet in SH-RML. The outcomes closely mirror those already presented for the FNO version, as well as the enhanced uncertainty quantification compared to ESMDA.

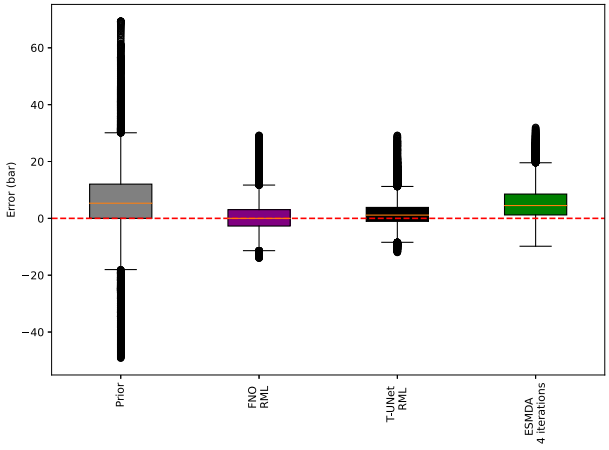


Fig. 21. Box plots comparing monitoring pressure errors between standard RML and Hybrid-RML-Surrogate using FNO and T-UNet surrogates.

Fig. 22 presents the histograms of pressure at the last time step of injection for SH-RML using the FNO surrogate model. These histograms demonstrate a better reduction in uncertainty than what was observed with both ESMDA and SH-ESMDA. This is quantitatively supported by the data in Table 4, which shows the P10-P90 pressure range reduction across the monitoring points. To maintain conciseness and since the outcomes between the FNO and T-UNet models are notably similar, we focus on presenting results from the FNO model only.

In summary, the results of this case study demonstrate that SH-RML offers improvements over ESMDA for history matching and uncertainty quantification. By leveraging the gradients of ML surrogates for optimization of the cost functions, SH-RML achieves better pressure

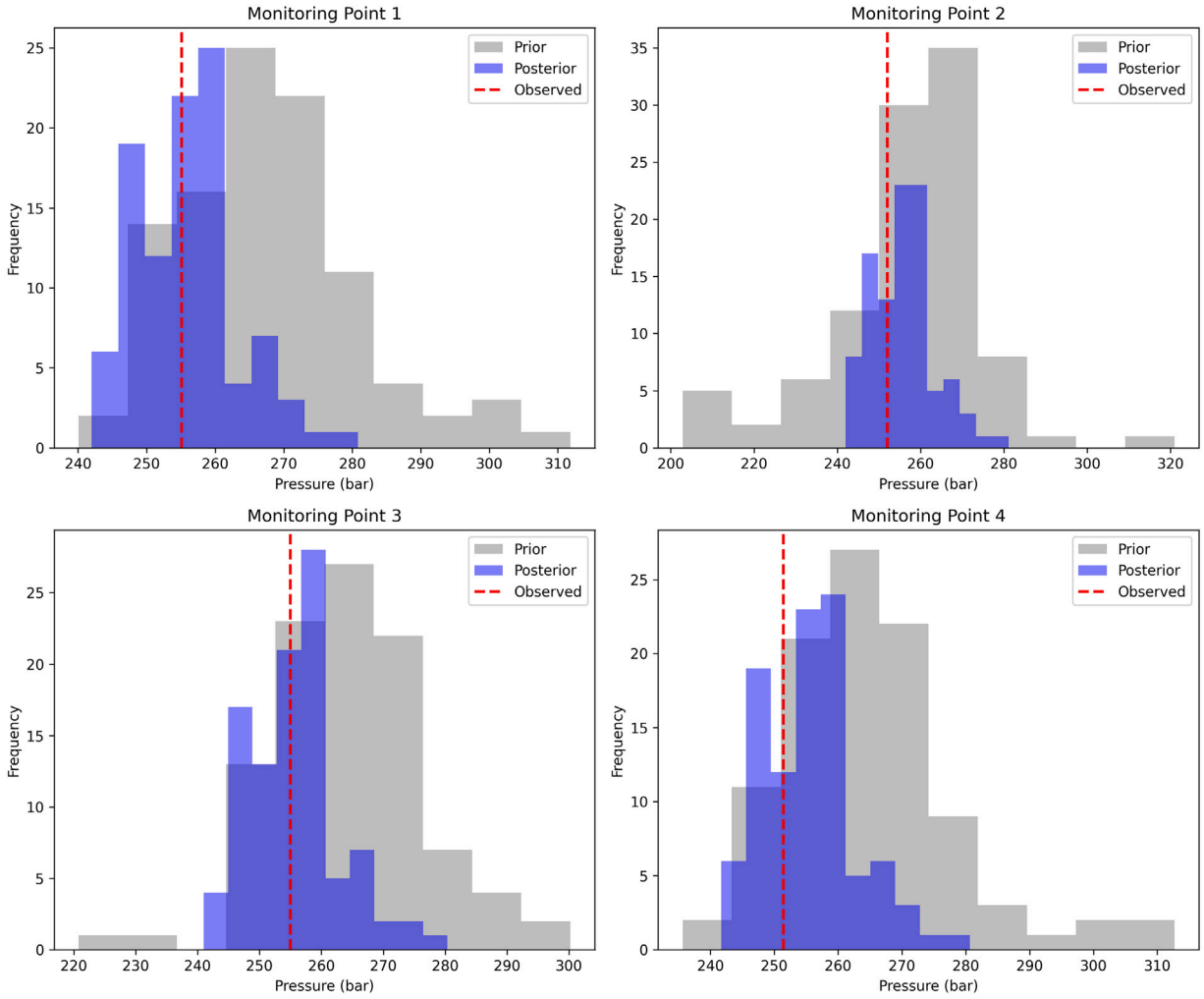


Fig. 22. Histograms of prior and posterior pressure distributions at the last time step of injection for each monitoring point for SH-RML with the FNO surrogate model.

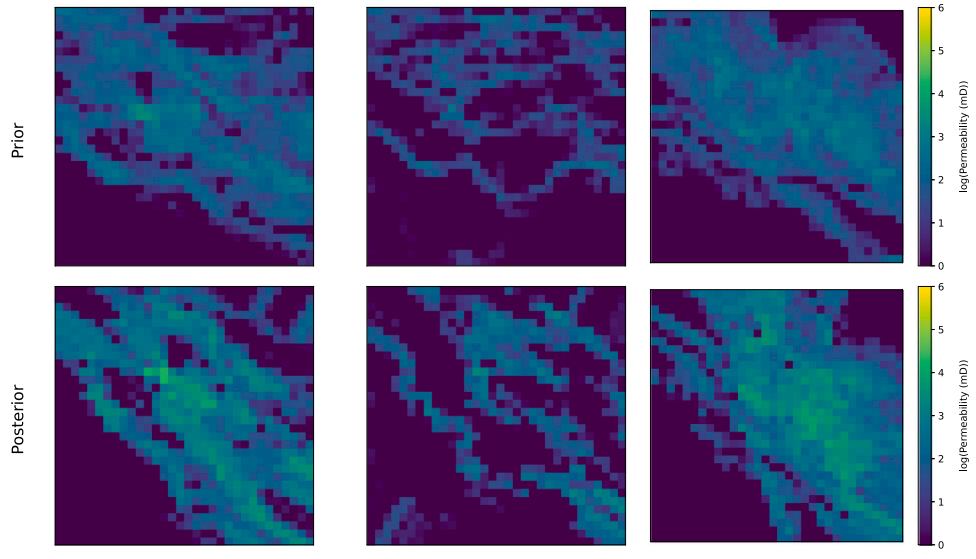


Fig. 23. Posterior permeabilities for Hybrid-RML-Surrogate with FNO surrogates.

matching while also providing more accurate estimates of posterior permeability distributions. The integration of surrogates enables the computations of gradient evaluations, enabling thorough optimization of the RML objective.

Developing on these results, our study contributes to and extends the current understanding of machine learning integration with data assimilation in geological carbon storage. It aligns with the data learning methods framework proposed by Buizza et al. (2022), contributing to the growing body of work on integrating machine learning with data assimilation. Similar to the approaches by Tang et al. (2022) and Wen et al. (2021b), our methods apply machine learning models to enhance data assimilation processes. However, our approach distinguishes itself by ensuring that the posterior is physics consistent, addressing a common limitation in previous studies, which rely heavily on surrogate models for data assimilation without adequately ensuring physical consistency (Tang et al., 2022; Wen et al., 2021b). In addition, our work demonstrates the feasibility of employing a combination of transformers with UNets in CO_2 storage problems, a method similar to that proposed by Li et al. (2023). This novel application within the field of geological carbon storage underscores the adaptability and potential of advanced neural network architectures in handling complex spatial and temporal data patterns typical of subsurface environments. Furthermore, while employing FNOs, we demonstrated the potential success of overcoming the challenge of data demand for training these models, which has been a notable issue in related works (Wen et al., 2021a; Tang et al., 2022; Sun and Durlofsky, 2019). Our approach not only reduces the amount of data required but also enhances the computational efficiency of the data assimilation framework, thereby offering a robust solution to one of the major hurdles in applying deep learning techniques to geoscience applications.

7. Discussion

This study presents advancements in combining DA and ML to enhance history matching in CO_2 storage projects, addressing key challenges and limitations inherent in existing methodologies. Our innovative hybrid frameworks, namely SH-ESMDA and SH-RML, optimize both computational efficiency and accuracy in DA studies. The integration of ML surrogates enabled the application of ensemble-based DA methods like ESMDA efficiently to complex subsurface systems. Specifically, the SH-ESMDA method has accelerated ESMDA computations while maintaining the physical consistency of posterior physical responses. The SH-RML method has excelled in achieving superior history matching compared to standard ESMDA, attributed to the better approximation of the gradients for non-linear and non-Gaussian problems. This method leveraged the automatic differentiation inherent in ML models, enabling gradient-based optimization and overcoming the challenges posed by the absence of adjoint models in reservoir simulators.

A consideration during this study was the choice of grid resolution. In the preliminary stages, models with resolutions of 256×256, 128×128, and 64×64 were examined. The choice of a 32×32 resolution was influenced by extended computational times required for numerous runs at higher resolutions and substantial memory requirements of the FNO. Although subsequent updates to the computational tools have mitigated these challenges, allowing for exploration at higher resolutions, a substantial portion of the work presented was conducted using the initial setup and resolutions due to the constraints at the time. The chosen resolution effectively captures the variability of the permeability of channelized reservoirs, offering a balance between computational demand and model fidelity.

Additionally, while we recognize that vertical gradients and gravity-driven flow are crucial in more comprehensive CO_2 storage models, the 2D horizontal model allows the examination of the areal and non-Gaussian characteristics of permeability distributions typical of geological formations used for CO_2 storage. These formations often feature highly heterogeneous and non-Gaussian permeability distributions due to natural depositional processes, which are effectively represented in the simplified 2D channelized models. By focusing on horizontal pressure distribution within the reservoir, we address the primary dynamics of CO_2 flow, where lateral variations in permeability significantly influence the migration and trapping of CO_2. This approach facilitates the analysis of horizontal flow behaviors, providing essential insights into lateral connectivity and barrier effects on CO_2 plume behavior without the complicating factors of vertical segregation. The 2D model aligns with the goals of developing a new computational framework and sets a clear path for future 3D extensions, which can incorporate vertical dynamics and gravity effects, enhancing the framework's applicability to real CO_2 storage projects. The insights gained from these 2D horizontal simulations provide a baseline for the development of more complex models, ensuring that the underlying data assimilation and machine learning techniques are robust and well-understood before adding additional layers of complexity.

Reflecting on the characteristics and applications of each method, the SH-ESMDA is optimal for scenarios prioritizing computational speed. At the same time, the SH-RML is particularly suited for problems necessitating more accurate gradient computations, even when forward models lack adjoints. The unified parameterization of permeability between ML and physics-based models has facilitated integration, ensuring physical consistency and reliability in the posterior solutions computed using the high-fidelity DARTS simulator. However, it is important to acknowledge some limitations of the methods. The SH-ESMDA inherits the fundamental constraints from ESMDA in capturing uncertainties, and the accuracy of both hybrid methods is contingent on the precision of the ML surrogates, which may encounter challenges in extrapolation. Addressing these limitations could potentially be achieved by incorporating more robust and nonlinear DA methods, enhancing the reliability of uncertainty quantification. A limitation of the SH-RML is the dependency on the accuracy of the ML surrogates for gradient computation of specific points of the reservoir. If the surrogates are not well-trained or accurate, the gradients derived from them could be unreliable, leading to suboptimal or incorrect solutions and impacting the method's overall effectiveness.

The implications of this uncertainty reduction relate to the core objectives of this study since a reduction in the uncertainties of the pressure buildup in the reservoir is important to ensure operations that preserve the integrity of caprock to prevent CO_2 leakage and mitigate the risk of induced seismicity and fault reactivation.

Our comparative assessment between FNO and T-UNet has demonstrated that both architectures exhibit low RMSE compared to DARTS simulations, with FNO having a slight edge, particularly when training data is limited. This insight is essential for scientists to choose the most suitable ML architectures effectively, especially when considering data availability constraints. Beyond CO_2 storage, the versatility of the developed methodologies is evident, with potential adaptability to other applications including geothermal energy and nuclear waste disposal. The advancements made in this study are poised to catalyze the broader adoption and application of hybrid DA-ML methodologies across diverse scientific domains, marking a significant step forward in the field.

To provide a comprehensive overview of the methodologies assessed in this study, we have included a comparative summary in Table 5.

8. Conclusion

This paper introduced novel frameworks for enhancing uncertainty quantification (UQ) in Geological Carbon Storage (GCS) projects through the integration of machine learning (ML) and data assimilation (DA) techniques. We evaluated two neural network architectures, the Fourier Meural Operators (FNOs) and Transformer-UNet (T-UNet), for generating accurate and efficient surrogate models of CO_2 injection

Table 5
Summary of methods and their characteristics.

Method	Category	Advantages	Disadvantages
Fourier Neural Operators (FNO)	ML Model	Slightly better performance with limited data. Efficient in capturing temporal dynamics.	Potentially higher computational cost than T-UNet.
Transformer UNet (T-UNet)	ML Model	Good at integrating spatial and temporal data. Effective in handling complex patterns.	Can be more sensitive to small training datasets compared to the FNO.
Ensemble Smoother with Multiple Data Assimilation (ESMDA)	DA Method	Does not require gradients. Easy implementation.	Assumes Gaussianity, which can limit its effectiveness in non-Gaussian systems.
Surrogate-based Hybrid ESMDA (SH-ESMDA)	DA Method	Significantly reduces computational time (by over 50%). Maintains physical consistency of the model.	Inherits limitations of the core ESMDA method, such as issues with non-Gaussian distributions.
Surrogate-based Hybrid RML (SH-RML)	DA Method	Provides improved history matching and uncertainty quantification. Avoids the need for adjoint derivations thanks to ML surrogates.	Dependency on the accuracy of ML surrogates for gradient computations.

simulations. Comparative analyses revealed the FNO's slight superiority when training data is limited.

Leveraging these surrogates, we proposed two hybrid methods. Surrogate-based hybrid ESMDA (SH-ESMDA) incorporates the ML models into the Ensemble Smoother with Multiple Data Assimilation (ESMDA), reducing computational time by over 50% while maintaining accuracy. Surrogate-based Hybrid RML (SH-RML) enables variational data assimilation by using automatic differentiation from the neural networks for gradient calculations in the Randomized Maximum Likelihood (RML) optimization. This avoids manual adjoint derivations.

Results showed that SH-RML achieved improved history matching and uncertainty quantification compared to standard ESMDA. The FNO surrogate enabled efficient computation of gradients for the RML optimization. The proposed frameworks thus enhance DA for GCS by integrating machine learning efficiency with the physical reliability of reservoir simulators.

In conclusion, while our study presents significant advancements in integrating AI with DA for GCS applications, it also opens avenues for future research. Particularly, exploring the scalability of our methodologies to larger, more complex reservoir systems, such as real carbonate formations, and real field cases from current GCS projects and investigating the integration of additional neural network architectures, could yield further valuable insights in optimizing GCS operations. The framework shows versatility beyond CO_2 sequestration, presenting opportunities for adaptation to other subsurface modeling applications like geothermal energy and nuclear waste storage. Overall, this study demonstrates the potential of combining machine learning and physics-based models to tackle problems in uncertainty quantification for the energy transition.

CRediT authorship contribution statement

Gabriel Serrão Seabra: Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Nikolaj T. Mücke:** Writing – review & editing, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Vinicius Luiz Santos Silva:** Writing – review & editing, Validation, Methodology, Conceptualization. **Denis Voskov:** Writing – review & editing, Supervision, Software, Methodology, Funding acquisition. **Femke C. Vossepoel:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gabriel Serrao Seabra reports financial support was provided by Petrobras. Vinicius Luiz Santos Silva reports financial support was provided by Petrobras.

Data availability

Data will be made available on request.

Declaration of Generative AI and AI-assisted technologies in the writing process

In the course of preparing this manuscript, the authors utilized ChatGPT from OpenAI to improve readability. Additionally, Grammarly was employed to guarantee grammatical accuracy and to improve overall readability. Subsequent to employing these tools, the authors meticulously reviewed and amended the content as necessary. The authors bear full responsibility for the content of the publication.

Acknowledgments

The authors thank Petróleo Brasileiro S.A. (Petrobras) for sponsoring the doctoral research of Gabriel Serrão Seabra and Vinicius Luiz Santos Silva.

We also thank Guillaume Rongier for providing valuable insights on geological modeling, Alexandre Emerick and Rafael Oliveira for sharing expertise on data assimilation techniques, and Ahmed ElSheikh for fruitful discussions about the machine learning models.

Appendix. Detailed RMSE metrics for FNO and T-UNet models

See Table A.6.

Table A.6
Test RMSE metrics for pressure and CO_2 molar fraction.

Samples	Modes	Width	RMSE (Pressure)	RMSE (CO_2 molar fraction)
FNO				
1000	18	128	4.29081	0.01072
1000	18	64	4.41009	0.01047
1000	6	128	4.28229	0.00895
1000	6	64	4.66067	0.00759
500	18	128	5.80695	0.01089
500	18	64	5.44274	0.01092

(*continued on next page*)

Table A.6 (*continued*).

Samples	Modes	Width	RMSE (Pressure)	RMSE (CO_2 molar fraction)
500	6	128	6.76917	0.01074
500	6	64	6.54552	0.01087
200	18	128	8.22966	0.01156
200	18	64	8.34038	0.01713
200	6	128	8.29742	0.01250
200	6	64	8.20080	0.01262
100	18	128	9.91828	0.01761
100	18	64	10.07389	0.01789
100	6	128	8.75043	0.01727
100	6	64	10.90478	0.01747
T-UNet				
1000	–	–	2.27017	0.00620
500	–	–	6.04151	0.00646
200	–	–	12.78398	0.01721
100	–	–	12.71158	0.03029

References

Agogo, O., Gosselin, O., Aanonsen, S.I., Navdal, J., 2022. Ensemble data space inversion for fast CO_2 injection forecast evaluation. Energy Procedia 159, 111–117.

AlSalmi, H., Elsheikh, A.H., 2023. Automated seismic semantic segmentation using attention U-net. Geophysics 1–75. http://dx.doi.org/10.1190/geo2023-0149.1.

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2021. Combining data assimilation and machine learning to infer unresolved scale parametrization. Phil. Trans. R. Soc. A 379 (2194), 20200086. http://dx.doi.org/10.1098/rsta.2020.0086.

Buizza, C., Casas, C.Q., Nadler, P., Mack, J., Marrone, S., Titus, Z., Cornec, C.L., Heylen, E., Dur, T., Ruiz, L.B., Heaney, C., Lopez, J.A.D., Kumar, K.S., Arcucci, R., 2022. Data learning: Integrating data assimilation and machine learning. J. Comput. Sci. 58, 101525. http://dx.doi.org/10.1016/j.jocs.2021.101525.

Burchette, T.P., 2012. Carbonate rocks and petroleum reservoirs: a geological perspective from the industry. Geol. Soc. Lond. Special Publ. 370 (1), 17–37. http://dx.doi.org/10.1144/sp370.14.

Chen, Y., Voskov, D., 2020. Optimization of CO2 injection using multi-scale reconstruction of composition transport. Comput. Geosci. 24 (2), 819–835.

Cheng, S., Quilodran-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., Xiao, D., Janjic, T., Ding, W., Guo, Y., Carrassi, A., Bocquet, M., Arcucci, R., 2023. Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. http://dx.doi.org/10.48550/ARXIV.2303.10462.

Computer Modelling Group Ltd., 2023. Reservoir Simulation Software. https://www.cmgl.ca/, [Accessed 30 August 2023].

DARTS, 2024. Open-DARTS repository. URL https://gitlab.com/open-darts/open-darts.

de Brito, D.U., Durlofsky, L.J., 2020. Field development optimization using a sequence of surrogate treatments. Comput. Geosci. 25 (1), 35–65. http://dx.doi.org/10.1007/s10596-020-09985-y.

Delottier, H., Doherty, J., Brunner, P., 2023. Data space inversion for efficient uncertainty quantification using an integrated surface and subsurface hydrologic model. Geosci. Model Dev. Discuss. 2023, 1–30.

Dong, Z., Cheng, J., Liu, T., Si, G., Ma, B., 2021. A novel surface parametric method and its application to aerodynamic design optimization of axial compressors. Processes 9 (7), 1230. http://dx.doi.org/10.3390/pr9071230.

DuMux Development Team, 2023. DuMux Simulator. https://dumux.org, [Accessed 30 August 2023].

Emerick, A.A., Reynolds, A.C., 2013a. Ensemble smoother with multiple data assimilation. Comput. Geosci. 55, 3–15. http://dx.doi.org/10.1016/j.cageo.2012.03.011.

Emerick, A.A., Reynolds, A.C., 2013b. Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. Comput. Geosci. 17 (2), 325–350. http://dx.doi.org/10.1007/s10596-012-9333-z.

Evensen, G., Vossepoel, F.C., van Leeuwen, P.J., 2022. Data assimilation fundamentals: a unified formulation of the state and parameter estimation problem. In: Springer Textbooks in Earth Sciences, Geography and Environment, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-030-96709-3.

Fenghour, A., Wakeham, W.A., Vesovic, V., 1998. The viscosity of carbon dioxide. J. Phys. Chem. Ref. Data 27 (1), 31–44.

García-Pintado, J., Paul, A., 2018. Evaluation of iterative Kalman smoother schemes for multi-decadal past climate analysis with comprehensive earth system models. Geosci. Model Dev. 11 (12), 5051–5084. http://dx.doi.org/10.5194/gmd-11-5051-2018.

Geneva, N., Zabaras, N., 2022. Transformers for modeling physical systems. Neural Netw. 146, 272–289.

GEOSX Development Team, 2023. GEOSX Simulator. https://geosx.org, [Accessed 8 March 2023].

Hesthaven, J.S., Rozza, G., Stamm, B., et al., 2016. Certified Reduced Basis Methods for Parametrized Partial Differential Equations. vol. 590, Springer.

IEA, 2022. Carbon capture, utilisation and storage - fuels & technologies. Technical Report, International Energy Agency, URL https://www.iea.org/reports/carbon-capture-utilisation-and-storage-2.

Khait, M., Voskov, D.V., 2017. Operator-based linearization for general purpose reservoir simulation. J. Pet. Sci. Eng. 157, 990–998. http://dx.doi.org/10.1016/j.petrol.2017.08.009.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Korondi, P.Z., Marchi, M., Parussini, L., Poloni, C., 2020. Multi-fidelity design optimisation strategy under uncertainty with limited computational budget. Optim. Eng. 22 (2), 1039–1064. http://dx.doi.org/10.1007/s11081-020-09510-1.

Li, G., Jin, D., Yu, Q., Qi, M., 2023. IB-TransUNet: Combining information bottleneck and transformer for medical image segmentation. J. King Saud Univ.-Comput. Inf. Sci. 35 (3), 249–258. http://dx.doi.org/10.1016/j.jksuci.2023.02.012.

Li, Y., Kokkinaki, A., Darve, E.F., Kitanidis, P., 2017. Smoothing-based compressed state Kalman filter for joint state-parameter estimation: Applications in reservoir characterization and CO_2 storage monitoring. Water Resour. Res. 53 (8), 6844–6867. http://dx.doi.org/10.1002/2016WR020168.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A., 2020. Fourier neural operator for parametric partial differential equations. http://dx.doi.org/10.48550/ARXIV.2010.08895.

Li, Q., Liu, G., 2016. Risk assessment of the geological storage of CO2: A review. In: Geologic Carbon Sequestration: Understanding Reservoir Behavior. Springer, pp. 249–284.

Lyu, X., Voskov, D., 2023. Advanced modeling of enhanced CO2 dissolution trapping in saline aquifers. Int. J. Greenh. Gas Control 127, 103907. http://dx.doi.org/10.1016/j.ijggc.2023.103907.

Machado, M.V.B., Delshad, M., Sepehrnoori, K., 2023. Potential benefits of horizontal wells for CO2 injection to enhance storage security and reduce leakage risks. Appl. Sci. 13 (23), 12830.

March, R., Doster, F., Geiger, S., 2018. Assessment of CO_2 storage potential in naturally fractured reservoirs with dual-porosity models. Water Resour. Res. 54 (3), 1650–1668. http://dx.doi.org/10.1002/2017wr022159.

Mücke, N.T., Bohté, S.M., Oosterlee, C.W., 2021. Reduced order modeling for parameterized time-dependent PDEs using spatially and memory aware deep learning. J. Comput. Sci. 53, 101408.

Nunes, J.P.P., Seabra, G.S., do Amaral, R.F., de Sousa Junior, L.C., 2022. CCS in the Brazilian pre-salt: Current status and geomechanical considerations. In: Rio Oil and Gas Expo and Conference. 22, (2022), Instituto Brasileiro de Petroleo e Gas, pp. 441–442. http://dx.doi.org/10.48072/2525-7579.rog.2022.441.

Ohlberger, M., Rave, S., 2015. Reduced basis methods: Success, limitations and future challenges. arXiv preprint arXiv:1511.02021.

Oliver, D.S., He, N., Reynolds, A.C., 1996. Conditioning permeability fields to pressure data. In: ECMOR V - 5th European Conference on the Mathematics of Oil Recovery. European Association of Geoscientists & Engineers, http://dx.doi.org/10.3997/2214-4609.201406884.

Pintea, S.L., Sharma, S., Vossepoel, F.C., van Gemert, J.C., Loog, M., Verschuur, D.J., 2021. Seismic inversion with deep learning. Comput. Geosci. 26 (2), 351–364. http://dx.doi.org/10.1007/s10596-021-10118-2.

Pour, K.M., Voskov, D., Bruhn, D., 2023. Nonlinear solver based on trust region approximation for CO2 utilization and storage in subsurface reservoir. Geoenergy Sci. Eng. 225, 211698.

Pyrcz, M., Boisvert, J., Deutsch, C., 2009. ALLUVSIM: A program for event-based stochastic modeling of fluvial depositional systems. Comput. Geosci. 35 (8), 1671–1685. http://dx.doi.org/10.1016/j.cageo.2008.09.012.

Quarteroni, A., Manzoni, A., Negri, F., 2015. Reduced Basis Methods for Partial Differential Equations: An Introduction. vol. 92, Springer.

Ringrose, P.S., Meckel, T.A., 2019. Maturing global CO2 storage resources on offshore continental margins to achieve 2DS emissions reductions. Sci. Rep. 9 (1), http://dx.doi.org/10.1038/s41598-019-54363-z.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III. Vol. 18, Springer International Publishing, pp. 234–241.

Rutqvist, J., 2012. The geomechanics of CO2 storage in deep sedimentary formations. Geotech. Geol. Eng. 30, 525–551.

Schlumberger, 2023. ECLIPSE reservoir simulator. https://www.slb.com/reservoir-characterization-and-simulation/software/eclipse, [Accessed 30 August 2023].

Silva, V.L.S., Heaney, C.E., Pain, C.C., 2023. Generative network-based reduced-order model for prediction, data assimilation and uncertainty quantification. In: LatinX in AI Workshop At International Conference on Machine Learning. URL https://openreview.net/forum?id=MGOiK8DCsU.

Sun, A.Y., Durlofsky, L.J., 2019. An integrated framework for optimal monitoring and history matching in CO_2 storage projects. Energy Procedia 159, 103–110.

Taccari, M.L., Nuttall, J., Chen, X., Wang, H., Minnema, B., Jimack, P.K., 2022. Attention U-net as a surrogate model for groundwater prediction. Adv. Water Resour. 163, 104169. http://dx.doi.org/10.1016/j.advwatres.2022.104169.

Tadjer, A., Bratvold, R.B., 2021. Managing uncertainty in geological CO2 storage using Bayesian evidential learning. Energies 14 (6), 1557.

Tang, H., Durlofsky, L.J., 2022. Use of low-fidelity models with machine-learning error correction for well placement optimization. Comput. Geosci. 26 (5), 1189–1206. http://dx.doi.org/10.1007/s10596-022-10153-7.

Tang, G., Wen, J., Ma, X., Feng, Y., Wu, Y., Durlofsky, L.J., 2022. Deep-learning-based coupled flow-geomechanics surrogate model for CO_2 sequestration. Int. J. Greenh. Gas Control 118, 103692.

Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. In: Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, http://dx.doi.org/10.1137/1.9780898717921.

Tarrahi, M., Jafarpour, B., Ghassemi, A., 2015. Integration of microseismic monitoring data into coupled flow and geomechanical models with ensemble Kalman filter. Water Resour. Res. 51 (7), 5245–5260. http://dx.doi.org/10.1002/2014WR016264.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wapperom, M., Tian, X., Novikov, A., Voskov, D., 2023. FluidFlower benchmark: lessons learned from the perspective of subsurface simulation. Transp. Porous Media 1–20.

Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., Benson, S.M., 2022. U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. Adv. Water Resour. 163, 104180. http://dx.doi.org/10.1016/j.advwatres.2022.104180.

Wen, G., Tang, M., Benson, S.M., 2021a. Towards a predictor for CO2 plume migration using deep neural networks. Int. J. Greenh. Gas Control 105, 103223. http://dx.doi.org/10.1016/j.ijggc.2020.103223.

Wen, J., Wu, Y., Durlofsky, L.J., Ma, X., Tang, G., Feng, Y., 2021b. A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. Int. J. Greenh. Gas Control 109, 103488.

White, J.A., Foxall, W., 2016. Assessing induced seismicity risk at CO2 storage projects: Recent progress and remaining challenges. Int. J. Greenh. Gas Control 49, 413–424.

Witte, P.A., Konuk, T., Skjetne, E., Chandra, R., 2023. Fast CO_2 saturation simulations on large-scale geomodels with artificial intelligence-based wavelet neural operators. Int. J. Greenh. Gas Control 126, 103880. http://dx.doi.org/10.1016/j.ijggc.2023.103880.

Xu, P., Zhu, X., Clifton, D.A., 2023. Multimodal learning with transformers: A survey. IEEE Trans. Pattern Anal. Mach. Intell..

Zhang, C., Song, X., Azevedo, L., 2021. U-net generative adversarial network for subsurface facies modeling. Comput. Geosci. 25 (1), 553–573. http://dx.doi.org/10.1007/s10596-020-10027-w.

Zoback, M.D., Gorelick, S.M., 2012. Earthquake triggering and large-scale geologic storage of carbon dioxide. Proc. Natl. Acad. Sci. 109 (26), 10164–10168.