RESEARCH ARTICLE

Open Access

Solving multiphysics-based inverse problems with learned surrogates and constraints

Ziyi Yin[1,*] , Rafael Orozco[1], Mathias Louboutin[1] and Felix J. Herrmann[1]

*Correspondence:
ziyi.yin@gatech.edu

[1] Georgia Institute of Technology
756 West Peachtree Street NW,
Atlanta 30308, GA, USA

Abstract

Solving multiphysics-based inverse problems for geological carbon storage monitoring can be challenging when multimodal time-lapse data are expensive to collect and costly to simulate numerically. We overcome these challenges by combining computationally cheap learned surrogates with learned constraints. Not only does this combination lead to vastly improved inversions for the important fluid-flow property, permeability, it also provides a natural platform for inverting multimodal data including well measurements and active-source time-lapse seismic data. By adding a learned constraint, we arrive at a computationally feasible inversion approach that remains accurate. This is accomplished by including a trained deep neural network, known as a normalizing flow, which forces the model iterates to remain in-distribution, thereby safeguarding the accuracy of trained Fourier neural operators that act as surrogates for the computationally expensive multiphase flow simulations involving partial differential equation solves. By means of carefully selected experiments, centered around the problem of geological carbon storage, we demonstrate the efficacy of the proposed constrained optimization method on two different data modalities, namely time-lapse well and time-lapse seismic data. While permeability inversions from both these two modalities have their pluses and minuses, their joint inversion benefits from either, yielding valuable superior permeability inversions and CO_2 plume predictions near, and far away, from the monitoring wells.

Keywords: Fourier neural operators, Normalizing flows, Multiphysics, Deep learning, Learned surrogates, Learned constraints, Inverse problems

Introduction

In this paper, we introduce a novel learned inversion algorithm designed to address inverse problems based on partial differential equations (PDEs). These problems can be represented using the following general form:

$$\mathbf{d} = \mathcal{H} \circ \mathcal{S}(\mathbf{K}) + \boldsymbol{\epsilon}. \tag{1}$$

Yin *et al. Advanced Modeling and Simulation in Engineering Sciences*(2023)10:14

Page 2 of 27

In this expression, the nonlinear operator $\mathcal{S}$ represents the solution operator of a nonlinear parametric PDE mapping the coefficients $\mathbf{K}$ to the solution. Given numerical solutions of the PDE, partially observed data, collected in the vector $\mathbf{d}$, are modeled by compounding the solution operator with the measurement operator, $\mathcal{H}$, followed by adding the noise term $\boldsymbol{\epsilon}$ with noise level of σ—i.e., $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. This problem is quite general and pertinent to various physical applications, including geophysical exploration [88,89], medical imaging [4], and experimental design [1].

Without loss of generality, we focus on time-lapse seismic monitoring of geological carbon storage (GCS), which involves underground storage of supercritical CO_2 captured from the atmosphere or from industrial smoke stacks [22]. We consider GCS in saline aquifers, which involves multiphase flow physics where CO_2 replaces brine in the porous rocks [64]. In this context, the PDE solution operator, $\mathcal{S}$, serves as the multiphase flow simulator, which takes the gridded spatially varying permeability in the reservoir, $\mathbf{K}$, as input and produces n_t time-varying CO_2 saturation snapshots, $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n_t}]$, as output. The governing equations for the multiphase flow involve Darcy's and the mass conservation law. Detailed information on the governing equations, initial and boundary conditions, and numerical solution schemes can be found in [78] and the references therein. To ensure safety, conformance, and containment of GCS projects, various kinds of time-lapse data are collected to monitor CO_2 plumes. These different data modalities include measurements in wells [21,62], and the collection of gravity [2,63], electromagnetic [14,106], and seismic time-lapse data [5,57,101] that can be used to follow the plume and invert for reservoir properties such as the permeability, $\mathbf{K}$. The latter is the property of interest in this exposition.

Overall, solving for the reservoir model parameter, $\mathbf{K}$, poses significant challenges for two primary reasons:

- the forward modeling operator, $\mathcal{H} \circ \mathcal{S}$, can be ill-posed, resulting in multiple model parameters that fit the observed data equally well. This necessitates the use of regularizers [23,89] in the form of penalties or constraints [73].
- The PDE modeling operator $\mathcal{S}$, and the sensitivity calculations with respect to the model parameters can be computationally expensive for large-scale problems, limiting the efficacy of iterative methods such as gradient-based [49] or Markov chain Monte Carlo [16] methods.

To overcome the second challenge, numerous attempts have been made to replace computationally expensive PDE solves with more affordable approximate alternatives [6,79], which include the use of radial basis functions to learn the complex models from few sample points [75] or reduced-order modeling where the dimension of the model space is reduced [55,81]. More recently, various deep learning techniques have emerged as cheap alternatives to numerical PDE solves [30,35,40,56,71,76,77]. After incurring initial training costs, these neural operators lead to vastly improved computation of PDE solves. Data-driven methods have also been used successfully to learn coarse-to-fine grid mappings of PDEs solves. Because of their advertised performance on approximating solution operators of the multiphase flow in porous media [25,91,92,94,97,98], we will consider Fourier neural operators (FNOs, Li et al. [46], Kovachki et al. [39]) in this work even though alternative choices can be made. Once trained, FNOs produce approximate PDE solutions orders of magnitude faster than traditional solvers [18,25,46,47]. In addition,

Yin et al. [102], Louboutin et al. [52] and Louboutin et al. [54] demonstrated that trained FNOs can replace PDE solution operators during inversion. This latest development is especially beneficial to applications such as GCS where trained FNOs can be used in lieu of numerically costly flow simulators [27,48,78]. However, despite their promising results, unconstrained inversion formulations offer little to no guarantees that the model iterates remain within the statistical distribution on which the FNO was trained initially during inversion. As a consequence, FNOs may no longer produce accurate fluid-flow simulations throughout the iterations, which can lead to erroneous inversion results when the errors become too large, possibly rendering surrogate modeling by FNOs ineffective. To avoid this situation, we propose a constrained formulation where a trained normalizing flow (NF, Rezende and Mohamed [80]) is included as a learned constraint. This added constraint guarantees that the model iterates remain within the desired statistical distribution. Because our approach safeguards the FNO's accuracy, it allows FNOs to act as reliable low-cost neural surrogates replacing costly fluid-flow simulations and gradient calculations that rely on numerically expensive PDE solves during inversion.

The organization of this paper is as follows: First, we introduce FNOs and explore the possibility of replacing the forward modeling operator with a trained FNO surrogate. Next, NFs are introduced. By means of a motivational example, we demonstrate how these learned generative networks control the prediction error of FNOs by ensuring that the model iterates remain in distribution. Based on this motivational example, we propose our novel method for using trained NFs as a learned constraint to guarantee performance of FNO surrogates during inversion. Through four synthetic experiments related to GCS monitoring, the efficacy of our method will be demonstrated.

Fourier neural operators

There is an extensive literature on training deep neural networks to serve as affordable alternatives to computationally expensive numerical simulators [11,35,38,40,56]. Without loss of generality, we limit ourselves in this exposition to the training of a special class of neural operators known as Fourier neural operators (FNOs). These FNOs are designed to approximate numerical solution operators of the PDE solution operator, $\mathcal{S}$, by minimizing the following objective:

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \frac{1}{N} \sum_{j=1}^{N} \|\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{K}^{(j)}) - \mathbf{c}^{(j)}\|_2^2 \quad \text{where} \quad \mathbf{c}^{(j)} = \mathcal{S}(\mathbf{K}^{(j)}). \tag{2}$$

Here, $\mathcal{S}_{\boldsymbol{\theta}}$ denotes the FNO with network weights $\boldsymbol{\theta}$. The optimization aims to minimize the ℓ_2 misfit between numerically simulated PDE solutions, $\mathbf{c}^{(j)}$, and solutions approximated by the FNO, across N training samples (permeability models), $\{\mathbf{K}^{(j)}\}_{j=1}^{N}$ compiled by domain experts. Once trained, FNOs can generate approximate PDE solutions for unseen model parameters orders of magnitude faster than numerical simulations [18,25]. For model parameters that fall within the distribution used to train, approximation by FNOs are reasonably accurate—i.e., $\mathcal{S}_{\boldsymbol{\theta}^*}(\mathbf{K}) \approx \mathcal{S}(\mathbf{K})$, with $\boldsymbol{\theta}^*$ being the minimizer of Eq. 2. We refer to the numerical examples section for details calculating these weights. Before studying the impact of applying these surrogates on samples for the permeability that are out of distribution, let us first consider an example where data is inverted using surrogate modeling.

Inversion with learned surrogates

Replacing PDE solutions by approximate solutions yielded by trained FNO surrogates has two main advantages when solving inverse problems. First, as mentioned earlier, FNOs are orders of magnitude faster than numerical PDE solves, which allows for many simulations at negligible costs [15,42]. Second, existing software for multiphase flow simulations may not always support computationally efficient calculations of sensitivity, e.g. via adjoint-state calculations [13,34,74] of the simulations with respect to their input. In such cases, FNO surrogates are favorable because automatic differentiation on the trained network [26,52,54,99,102] readily provides access to gradients with respect to model parameters. As a result, the PDE solver, $\mathcal{S}$, in Eq. 1 can be replaced by trained surrogate, $\mathcal{S}_{\theta^*}$—i.e., we have

$$\underset{\mathbf{K}}{\text{minimize}} \quad \|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*}(\mathbf{K})\|_2^2 \tag{3}$$

where $\boldsymbol{\theta}^*$ represent the optimized weights minimizing Eq. 2. While the above formulation in terms of trained surrogates has been applied successfully during permeability inversion from time-lapse seismic data [44,54,102], this type of inversion is only valid as long as the (intermediate) permeabilities remain within distribution during the inversion. Practically, this means two things. First, the data need to be in the range of permeability models that are in distribution. This means that there can not be too much noise neither can the observed data be the result of an out-of-distribution permeability. Second, there are no guarantees that the permeability model iterates remain in distribution during inversion even though some bias of the gradients of the surrogate towards in-distribution permeabilities may be expected. To overcome this challenge, we propose to add a learned constraint to Eq. 3 that offers guarantees that the model iterates remain in distribution.

Learned constraints with normalizing flows

As demonstrated by Peters and Herrmann [72], Esser et al. [20], Peters et al. [73] regularization of non-linear inverse problems, such as full-waveform inversion, with constraints, e.g., total-variation [20] or transform-domain sparsity with ℓ_1-norms [45], offers distinct advantages over regularizations based on adding these norms as penalties. Even though constraint and penalty formulations are equivalent for linear inverse problems for the appropriate choice of the Lagrange multiplier, minimizing the constraint formulation leads to completely different solution paths compared to adding a penalty term to the data misfit objective [29]. In the constrained formulation, the model iterates remain at all times within the constraint set while model iterates yielded by the penalty formulation does not offer these guarantees. Peters and Herrmann [72] demonstrated this importance difference for the non-convex problem of full-waveform inversion. For this problem, it proved essential to work with a homotopy where the intersection of multiple handcrafted constraints (intersection of box and size of total-variation-norm ball constraints) are relaxed slowly during the inversion, so the model iterates remain physically feasible and local minima are avoided.

Motivated by these results, we propose a similar approach but now for "data-driven" learned constraints based on normalizing flows (NFs, Rezende and Mohamed [80]). NFs are powerful deep generative neural networks capable of learning to generate samples

from complex distributions [19,54,68,86]. Designed to be invertible, these NFs require the latent and model spaces to share identical dimensions, which confers several advantages:

- unlike variational autoencoders [36] or generative adversarial networks (GANs, Goodfellow et al. [24]), which both have a lower-dimensional latent space, NFs do not impose any intrinsic dimensionality constraints. This flexibility lets NFs capture model space characteristics across high dimensions [37]. Relevantly, concurrent literature has delved into the intrinsic dimensionality of NFs, indicating the potential to using NFs to generate models with inherently lower dimensions [31].
- NFs' inherent invertibility negates the need to store state variables during gradient calculations, enabling memory-efficient training and inversion in large-scale 3D applications, such as in geophysics [43,83,84,86,103–105] and ultrasound imaging [66–70,93].
- because of their invertibility NFs guarantee unique latent codes for all model space samples, including out-of-distribution ones. Therefore, they can still be used to invert for out-of-distribution model parameters, while other methods like GANs may introduce bias [7].

Aside from being invertible, NFs are trained to map samples from a target distribution in the physical space to samples from the standard zero-mean Gaussian distribution noise in the latent space. After training is completed, samples from the target distribution are generated by running the NF in reverse on samples in the latent space from the standard Gaussian distribution. Below, we will demonstrate how NFs can be used to guarantee that the permeability remains in distribution during the inversion.

Training normalizing flows

Given samples from the permeability distribution, $\{\mathbf{K}^{(j)}\}_{j=1}^N$, training NFs entails minimizing the Kullback–Leibler divergence between the base and target distributions [3]. This involves solving the following variational problem:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2} \|\mathcal{G}_{\mathbf{w}}^{-1}(\mathbf{K}_i)\|_2^2 - \log \left| \det \mathrm{J}_{\mathcal{G}_{\mathbf{w}}^{-1}}(\mathbf{K}_i) \right| \right). \tag{4}$$

In this optimization problem, $\mathcal{G}_{\mathbf{w}}^{-1}$ represents the NF, which is parameterized by its network weights $\mathbf{w}$, while $\mathrm{J}_{\mathcal{G}_{\mathbf{w}}^{-1}}$ denotes its Jacobian. By minimizing the ℓ_2-norm, the objective imposes a Gaussian distribution on the network's output and the second log det term prevents trivial solutions, i.e., cases where $\mathcal{G}_{\mathbf{w}}^{-1}$ produces zeros. To ensure alignment between the permeability distributions, Eq. 2 and Eq. 4 are trained on the same dataset consisting of 2000 permeability models examples of which are included in Fig. 1. Each 64×64 permeability model of consists of a randomly generated highly permeable channels (120 mD) in a low-permeable background of 20 mD, where mD denotes millidarcy. Generative examples produced by the trained NF are included in the second row of Fig. 1, which confirm the NF's ability to learn distributions from examples. Aside from generating samples from the learned distribution, trained NFs are also capable of carrying out density calculations, an ability we will exploit below.
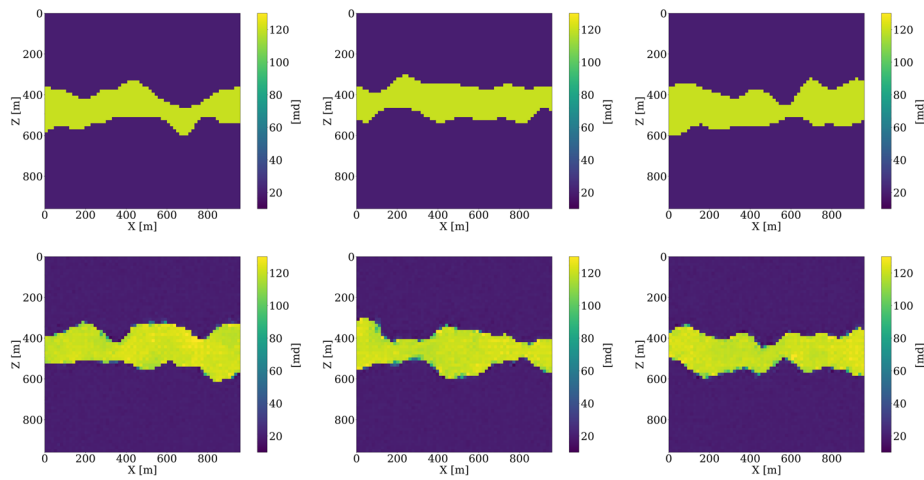
Fig. 1 Permeability models. First row shows the realistic permeability samples for FNO and NF training. Second row shows the generative samples from the trained NF
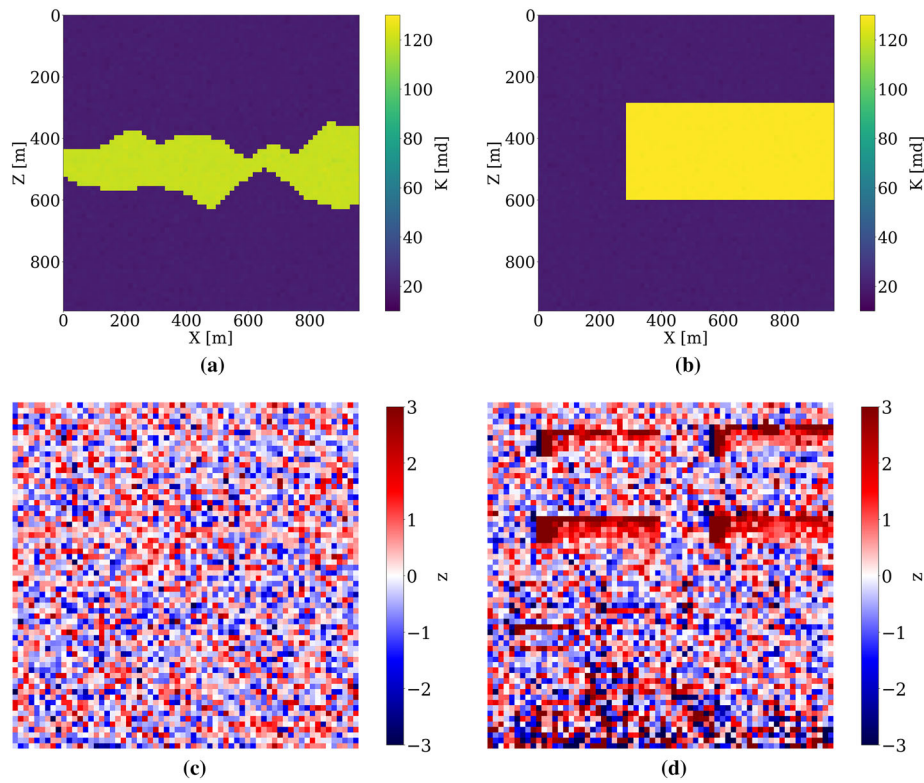

Fig. 2 Sample permeability models in the physical and latent space. **a** An in-distribution permeability model. **b** An out-of-distribution permeability model. **c** An in-distribution permeability model in the latent space. **d** An out-of-distribution permeability model in the latent space
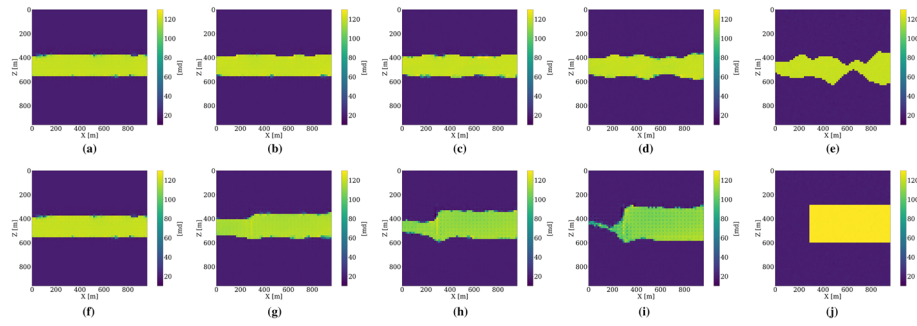
Fig. 3 Projections onto increasing ℓ_2-norm balls for the in- and out-of-distribution examples of Fig. 2. Top row: projections of in-distribution sample. Bottom row: projections of out-of-distribution sample. Each column corresponds to setting $\alpha = 0, 0.1, 0.2, 0.4, 1$ in Eq. 6
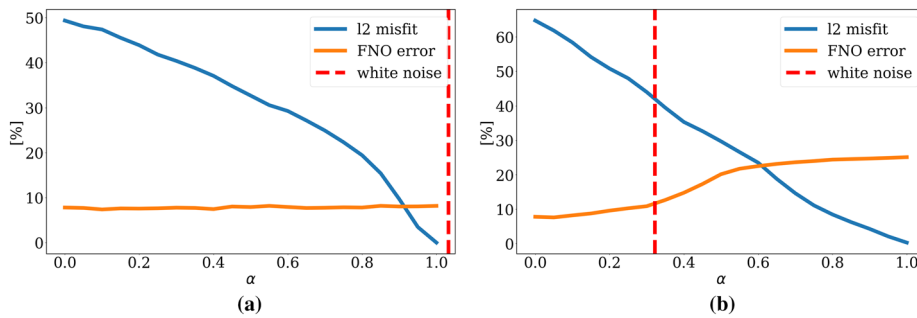

Fig. 4 Latent space projection experiments. **a** Relative ℓ_2 reconstruction error and FNO prediction error for in-distribution sample. **b** The same but for out-of-distribution sample. The blue curve shows the relative ℓ_2 misfit between the permeability models before and after latent space shrinkage. The orange curve shows the FNO prediction error on the permeability model after shrinking the ℓ_2-norm ball. The red dashed line denotes the amplitude of standard Gaussian noise

Trained normalizing flows as constraints

As we mentioned before, adding constraints to the solution of non-convex optimization problems offers guarantees that model iterates remain within constrained sets. When solving inverse problems with learned surrogates, it is important that model iterates remain "in distribution", which can be achieved by recasting the optimization problem in Eq. 3 into the following constrained form:

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\boldsymbol{\theta}^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau. \tag{5}$$

To arrive at this constrained optimization problem, two important changes were made. First, the permeability $\mathbf{K}$ is replaced by the output of a trained NF with trained weights $\mathbf{w}^*$ obtained by minimizing Eq. 4. This reparameterization in terms of the latent variable, $\mathbf{z}$, produces permeabilities that are in distribution as long as $\mathbf{z}$ remains distributed according to the standard normal distribution. Second, we added a constraint on this latent space variable in Eq. 5, which ensures that the latent variable $\mathbf{z}$ remains within an ℓ_2-norm ball of size τ.

To better understand the behavior of a trained normalizing flow in conjunction with the ℓ_2-norm constraint for in- and out-of-distribution examples, we include Figs. 2 and 3. In the latter Figure, nonlinear projections (via latent space shrinkage),

$$\widetilde{\mathbf{K}} = \mathcal{G}_{\mathbf{w}^*}(\alpha\mathbf{z}) \quad \text{where} \quad \mathbf{z} = \mathcal{G}_{\mathbf{w}^*}^{-1}(\mathbf{K}) \tag{6}$$
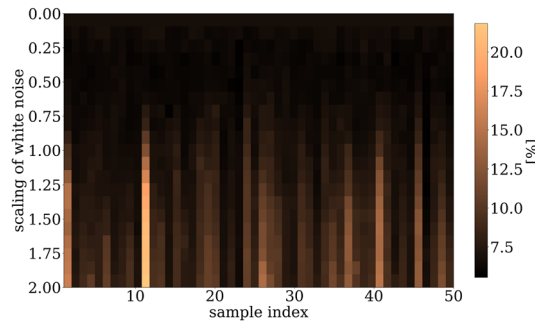
Fig. 5 FNO prediction errors for the latent space shrinkage experiment in Eq. 6 for 50 random realizations of standard Gaussian noise

are plotted as a function of increasing α. We also plot in Fig. 4 the NF's relative nonlinear approximation error, $\|\widetilde{\mathbf{K}} - \mathbf{K}\|_2/\|\mathbf{K}\|_2$, and the corresponding relative FNO prediction error, $\|\mathcal{S}_{\theta^*}(\widetilde{\mathbf{K}}) - \mathcal{S}(\widetilde{\mathbf{K}})\|_2/\|\mathcal{S}(\widetilde{\mathbf{K}})\|_2$ as a function of increasing $0 \leq \alpha \leq 1$. From these plots, we can make the following observations. First, the latent representations (Fig. 2c and d) of the in- and out-of-distribution samples (Fig. 2a and b) clearly show that NF applied to out-of-distribution samples produces a latent variable far from the standard normal distribution, while the latent variable corresponding to the in-distribution example is close to being white Gaussian noise. Quantitatively, the ℓ_2 norm of the latent variables are $0.99\|\mathcal{N}(0, \mathbf{I})\|_2$ and $3.11\|\mathcal{N}(0, \mathbf{I})\|_2$, respectively, where $\|\mathcal{N}(0, \mathbf{I})\|_2$ corresponds to the ℓ_2-norm of the standard normal distribution. Second, we observe from Fig. 3 that for small ℓ_2-norm balls ($\alpha \ll 1$) the projected solutions tend to be close to the most probable sample, which is a flat permeability channel in the middle. This is true for both the in- and out-of-distribution example. As α increases, the in-distribution example is reconstructed accurately when the ℓ_2 norm of the scaled latent variable, $\|\alpha \mathbf{z}\|_2$, is close to the $\|\mathcal{N}(0, \mathbf{I})\|_2$. Clearly, this is not the case for the out-of-distribution example. When $\|\alpha \mathbf{z}\|_2 \approx \|\mathcal{N}(0, \mathbf{I})\|_2$, the reconstruction still looks like an in-distribution permeability sample and is not close to the out-of-distribution sample. However, if $\alpha = 1$, which makes $\|\alpha \mathbf{z}\|_2$ well beyond the norm of the standard normal distribution, the out-of-distribution example is recovered accurately by virtue of the invertibility of NFs, irrespective on their input and what they have been trained on. Third, the relative FNO prediction error for the in-distribution example (Fig. 4a) remains flat while the error of the FNO surrogate increases as soon as $\alpha \approx 0.25$. At that value for α, the projection, $\widetilde{\mathbf{K}}$, is gradually transitioning from being in-distribution to out-of-distribution, which occurs at a non-linear approximation error of about 45%. As expected the plots in Fig. 4 also show a monotonous decay of the nonlinear approximation error as a function of increasing α. To further analyze the effects of the nonlinear projections in Eq. 6, we draw 50 random realizations from the standard normal distribution, scale each of them by $0 \leq \alpha \leq 2$, and calculate the FNO prediction errors on these samples. Figure 5 includes the results of this exercise where each column represents the FNO prediction error calculated for $0 \leq \alpha \leq 2$. From these experiments, we make the following two observations. First, when $\alpha < 0.8$, the FNO consistently makes accurate predictions for all projected samples. Second, as expected, the FNO starts to make less accurate predictions for $\alpha > 1$ with errors that increase as the size of the ℓ_2-norm ball of the latent space expands, demarcating the transition from being in distribution to being out-of-distribution.

In summary, the experiments of Figs. 2, 3, 4 indicate that FNO errors remain small and relatively constant for the in-distribution example. Irrespective of the value of α, the generated samples remain in distribution while moving from the most likely—i.e., a flat high-permeability channel in the middle, to the in-distribution sample as α increases. Conversely, the projection of the out-of-distribution example morphs from being in distribution to being out-of-distribution for $\alpha \geq 0.25$. The FNO prediction errors also increase during this transition from an in-distribution sample to an out-of-distribution sample. Therefore, shrinkage in the latent space by multiplying with a small α can serve as an effective projection that ensures relatively low FNO prediction errors. We will use this unique ability to control the distribution during inversion.

Inversion with progressively relaxed learned constraints

Our main objective is to perform inversions where the multiphase flow equations are replaced with pretrained FNO surrogates. To make sure the learned surrogates remain accurate, we propose working with a continuation scheme where the learned constraint in Eq. 5 is steadily relaxed by increasing the size of the ℓ_2-norm ball constraint. Compared to the more common penalty formulation, where regularization entails adding a Lagrange-multiplier weighted ℓ_2-norm squared, constrained formulations offer guarantees that the model iterates for the latent variable, $\mathbf{z}$, remain within the constraint set—i.e., within the ℓ_2-norm ball of size τ. Using the argument of the previous section, this implies that permeability distributions generated by the trained NF remain in distribution as long as the size of the initial ℓ_2-norm ball, τ_{init}, is small enough (e.g., smaller than $0.6\|\mathcal{N}(0, \mathbf{I})\|_2$, following the observations from Fig. 5). Taking advantage of this trained NF in a homotopy, we propose the following algorithm:

Algorithm 1: Inversion with relaxed learned constraints

---

1   **Input:** initial model parameter $\mathbf{K}_0 \in \mathbb{R}^N$, observed data $\mathbf{d}$, noise level σ
2   **Input:** trained FNO $\mathcal{S}_{\theta^*}$, trained NF $\mathcal{G}_{\mathbf{w}^*}$
3   **Input:** number of inner-loop iterations *maxiter*
4   **Input:** initial ℓ_2 ball size τ_{init}, multiplier $\beta > 1$, final ℓ_2 ball size τ_{final}
5   $\mathbf{z} = \mathcal{G}_{\mathbf{w}^*}^{-1}(\mathbf{K}_0)$
6   $\tau = \tau_{\text{init}}$
7   **while** $\|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2 > \sigma\|\mathcal{N}(0, \mathbf{I})\|_2$ and $\tau \leq \tau_{\text{final}}$ **do**
8   **for** *iter* $= 1 : maxiter$ **do**
9         $\mathbf{g} = \nabla_{\mathbf{z}}\|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2$
10        $\mathbf{z} = \mathcal{P}_{\tau}(\mathbf{z} - \gamma\mathbf{g})$
11   **end**
12      $\tau = \beta\tau$
13   **end**
14   **Output:** inverted model parameter $\mathbf{K} = \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})$

Given observed data, $\mathbf{d}$, trained networks, $\mathcal{S}_{\theta^*}$ and $\mathcal{G}_{\mathbf{w}^*}$, the initial guess for the permeability distribution, $\mathbf{K}_0$, the initial size of the ℓ_2-norm ball, τ_{init}, and the final size of the ℓ_2-norm ball, τ_{final}, Algorithm 1 proceeds by solving a series of constrained optimization problems where the size of the constraint set is increased by a factor of β after each iteration (cf. line 12 in Algorithm 1). The constrained optimization subproblems themselves

(cf. line 8 to 11 of Algorithm 1) are solved with projected gradient descent [10]. Each iteration of the projected gradient descent method first calculates the gradient (cf. line 9 of Algorithm 1), followed by the much cheaper projection of the updated latent variable back onto the ℓ_2-norm ball of size τ via the projection operator $\mathcal{P}_\tau$ (cf. line 10 in Algorithm 1). This projection is a trivial scaling operation if the updated latent variable ℓ_2-norm exceeds the constraint—i.e.,

$$\mathcal{P}_\tau(\mathbf{z}) = \begin{cases} \mathbf{z} & \text{if } \|\mathbf{z}\|_2 \leq \tau \\ \tau\mathbf{z}/\|\mathbf{z}\|_2 & \text{if } \|\mathbf{z}\|_2 > \tau \end{cases} \tag{7}$$

A line search determines the steplength γ [87] for each iteration shown in line 8 to 11. As is common in continuation methods, the relaxed gradient-descent iterations are warm-started with the optimization result from the previous iteration, which at the first iteration is initialized by the latent representation of the initial permeability model, $\mathbf{K}_0$ (cf. line 5 in Algorithm 1). Practically, each subproblem does not need to be fully solved, but only need a few iterations instead. The number of iterations to solve each subproblem is denoted by *maxiter* in line 8 of Algorithm 1. This continuation strategy serves two purposes. First, for small τ's it makes sure the model iterates remain in distribution, so accuracy of the learned surrogate is preserved. Second, by relaxing the constraint slowly, the data residual is gradually allowed to decrease, bringing in more and more features derived from the data. By slowly relaxing the constraint, we find a careful balance between these two purposes as long as progress is made towards the solution when solving the subproblem (cf. line 8 to 11 in Algorithm 1). One notable distinction of the surrogate-assisted inversion, compared to the conventional inversion with relaxed constraints [20], is that the size of the ℓ_2-norm projection ball cannot increase far beyond the ℓ_2-norm of the standard Gaussian white noise on which the NFs are trained. Otherwise, there is no guarantee the learned surrogate is accurate because the NF may generate samples that are out-of-distribution (cf. Fig. 5). This is explicitly incorporated into the stopping criteria, $\tau \leq \tau_{\text{final}}$, in line 7 of Algorithm 1.

Numerical experiments

To showcase the advocacy of the proposed optimization method with relaxed learned constraints, a series of carefully chosen experiments of increasing complexity are conducted. These experiments are designed to be relevant to GCS, which in its ultimate form involves coupling of multiphase flow with the wave equation to perform end-to-end inversion for the permeability given multimodal data. To convince ourselves of the validity of our approach, at all times comparisons will be made between inversion results involving numerical solves of the multiphase equations and inversions yielded by approximations with our learned surrogate.

For all numerical experiments, the "ground-truth" permeability model will be selected from the unseen test set and is shown in Fig. 6a. The inversions will be initiated with the smooth permeability model depicted in Fig. 6b. This initial model, $\mathbf{K}_0$, represents the arithmetic mean of all permeability samples in the training dataset. To ensure that the model iterates remain in distribution, we set the starting ℓ_2-norm ball size to $\tau_{\text{init}} = 0.6\|\mathcal{N}(0, \mathbf{I})\|_2$—i.e., $0.6\times$ the ℓ_2-norm of standard white Gauss noise realizations for the discrete permeability model of 64 by 64 gridpoints. To gradually relax the learned
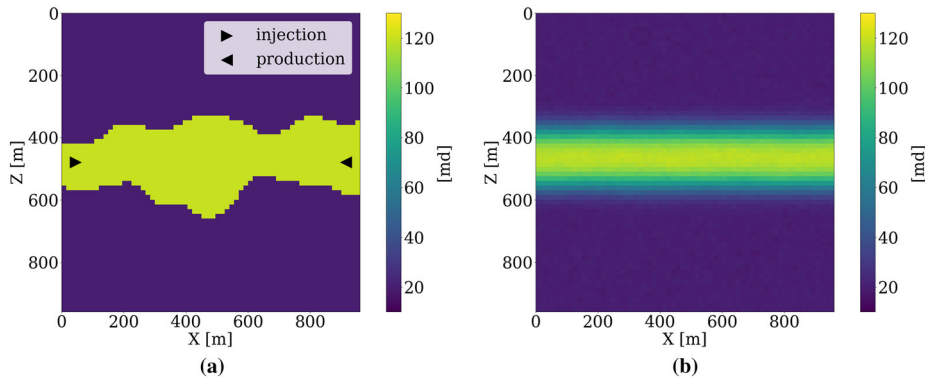
Fig. 6 Permeability models. **a** unknown "ground-truth" permeability model from unseen test set, where the symbols ▶ and ◀ denote the CO_2 injection and brine production location, respectively; **b** initial permeability model, $\mathbf{K}_0$

constraint, the multiplier of the projection ball size is taken to be $\beta = 1.2$, and we set the ultimate projection ball size τ_{final} in Algorithm 1 to be 1.2 times the norm of standard white noise. To limit computational costs of solving the subproblems, we allow each constrained subproblem (cf. line 8 to 11 in Algorithm 1) to perform 8 iterations of projected gradient descent to solve for the latent variable. From practical experience, we found that the proposed inversions are not very sensitive to the choice of these hyperparameters.

To simulate the evolution of injected CO_2 plumes, we make use of the open-source software package Jutul.jl [60,61,100], which for each permeability model, $\mathbf{K}^{(j)}$, solves the immiscible and compressible two-phase flow equations for the CO_2 and brine saturation. As shown in Fig. 6a, an injection well is set up on the left-hand side of the model, which injects supercritical CO_2 with density 700 kg/m^3 at a constant rate of 0.005 m^3/s. To relieve pressure, a production well is included on the right-hand side of the model, which produces brine with density 1000 kg/m^3 with a constant rate of also 0.005 m^3/s. This finally results in approximately a 6% storage capacity after 800 days of CO_2 injection. From these simulations, we collect eight snapshots for the CO_2 concentration, $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n_t}]$ with $n_t = 8$ the number of snapshots that cover a total time period of 800 days. The last five snapshots of these simulations are included in the top row of Fig. 6a. Due to buoyancy effects and well control, the CO_2 plume gradually moves from the left to the right and upwards.

Given these simulated CO_2 concentrations, the optimized weights, $\mathbf{w}^*$, for the FNO surrogate are calculated by minimizing Eq. 2 for $N = 1900$ training pairs, $\{\mathbf{K}^{(j)}, \mathbf{c}^{(j)}\}_{j=1}^N$. Another 100 training pairs are used for validation. After training with 350 epochs, an average of 7% prediction error is achieved for permeability samples from the unseen test set. As observed from Fig. 7, the approximation errors of the FNO are mostly concentrated at the leading edge of the CO_2 plumes. The same permeability models are used to train the NF by minimizing Eq. 4 for 245 epochs using the open-source software package InvertibleNetworks.jl [95]. We use the HINT network structure [41] for the NF. Three generative samples are shown in the second row of Fig. 1. From these examples, we can see that the trained NF is capable of generating random permeability models that resemble the ones in the training samples closely, despite minor noisy artifacts.
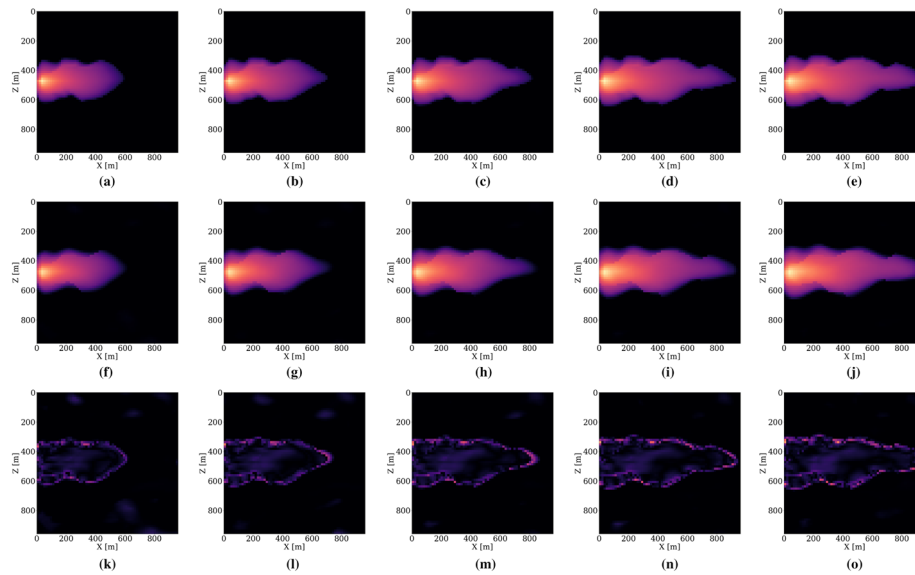
Fig. 7 Five CO$_2$ saturation snapshots after 400, 500, 600, 700, and 800 days. First row shows the CO$_2$ saturation simulated by the PDE. Second row shows the CO$_2$ saturation predicted by the trained FNO. Third row shows the 5× difference between the first row and the second row

Unconstrained/constrained permeability inversion from CO$_2$ saturation data

To demonstrate that permeability inversion with surrogates is indeed feasible, we first consider the idealized, impossible in practice, situation where we assume to have access to the time-lapse CO$_2$ concentration, $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n_t}]$, everywhere, and for all $n_t = 8$ timesteps. In that case, the measurement operator, $\mathcal{H}$ in Eq. 1, corresponds to the identity matrix. Given CO$_2$ concentrations simulated from the "ground-truth" permeability distribution plotted in Fig. 6a, we invert for the permeability by minimizing the unconstrained formulation (cf. Eq. 3) for the correct, yielded by the PDE, and approximate fluid-flow physics, yielded by the trained FNO. The results of these inversions after 100 iterations of gradient descent with back-tracking linesearch [87] are plotted in Fig. 8a and b. From these plots, we observe that the inversion results using PDE solvers delineates most of the upper boundary of the channel accurately. Because there is a null space in the fluid-flow modeling—i.e., this null space mostly corresponds to regions of the permeability model that are barely touched by the CO$_2$ plume (e.g. bottom and right-hand side of the channel)—artifacts are present in the high-permeability channel itself. As expected, the reconstruction of the permeability is also not perfect at the bottom and at the far right of the model. The inversion result with the FNO surrogate is similar but introduces unrealistic artifacts in the high-permeability channel and also outside the channel. These more severe artifacts can be explained by the behavior of the FNO approximation error plotted as the orange curve in Fig. 8e. The error value increases rapidly to 13%, and finally saturates at 10%. This behavior of the error is a manifestation of out-of-distribution model iterates that explain the erroneous behavior of the surrogate and its gradient with respect to the permeability.

Inversions yielded by the relaxed constrained formulation with the trained NF (see Algorithm 1), on the other hand, show virtually artifact free inversion results (see Fig. 8c and d) that compare favorably with the "ground-truth" permeability plotted in Fig. 6. While
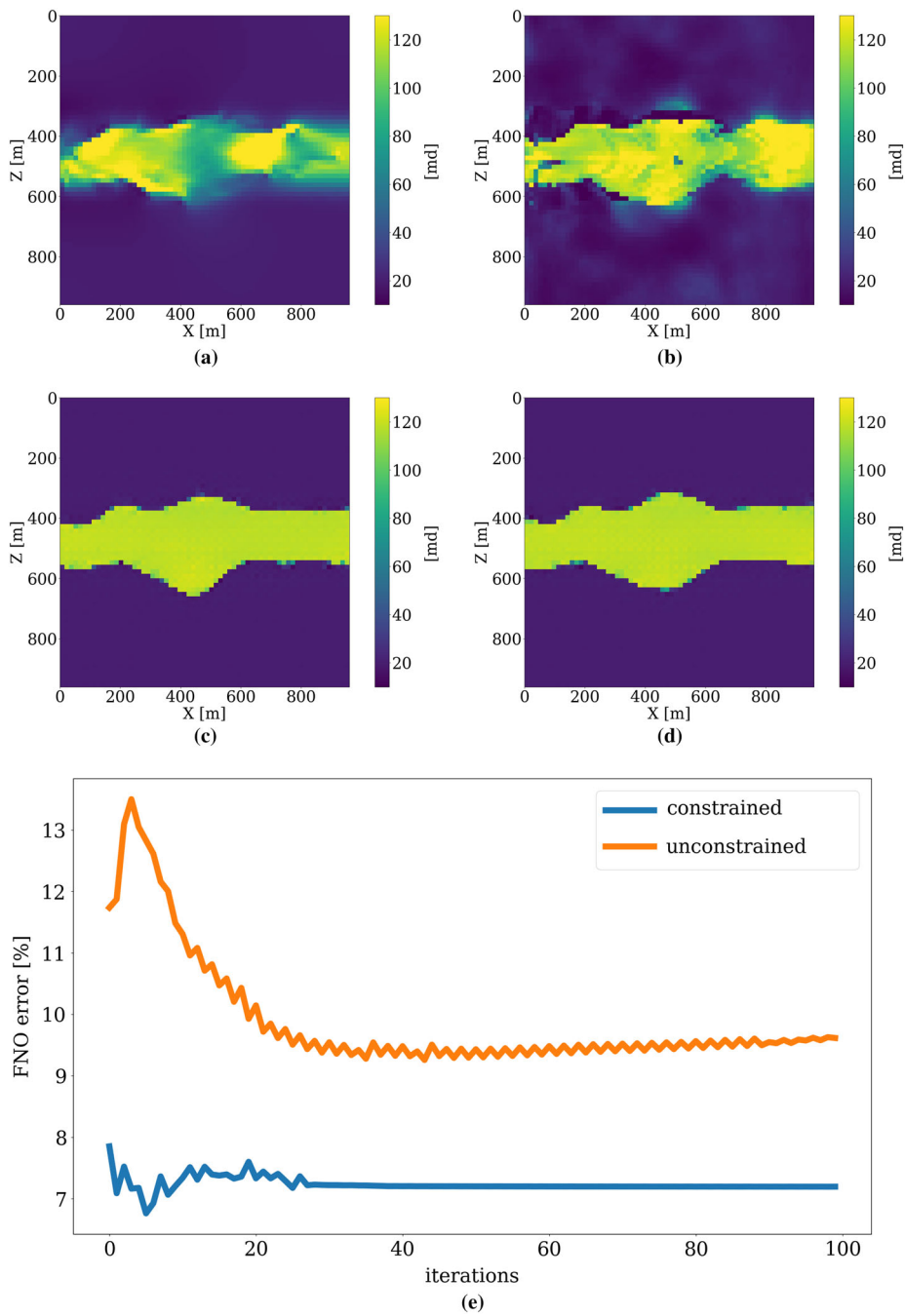
Yin *et al. Advanced Modeling and Simulation in Engineering Sciences* (2023) 10:14

Page 13 of 27

Fig. 8 Permeability inversion from fully observed time-lapse CO_2 saturations. **a** Inversion result with PDE solvers. **b** The same but via the approximate FNO surrogate. **c** Same as **a** but with NF constraint. **d** Same as **b** but with NF constraint. **e** The FNO approximation errors as a function of the number of iterations for the result plotted in **b** and **d**

adding the NF as a constraint obviously adds information, explaining the improved inversion for the accurate physics (Fig. 8c), it also renders the approximate surrogates more accurate, as can be observed from the blue curve in Fig. 8e, where the FNO approximation error is controlled thanks to adding the constraint to the inversion. This behavior underlines the importance of ensuring model iterates to remain within distribution. It also demonstrates the benefits of a solution strategy where we start with a small τ, followed by relaxing the constraint slowly by increasing the size of the constraint set gradually.

Unconstrained/constrained permeability inversion from well observations

While the example of the previous section established feasibility of constrained permeability inversion, it relied on having access to the CO_2 saturation everywhere, which is unrealistic in practice. To address this issue, we first consider permeability inversion from CO_2 saturations, collected at three equally spaced monitoring well locations, for only the first 6 timesteps over the period of 600 days [59]. In this more realistic setting, the measurement operator, $\mathcal{H}$ in Eq. 1, corresponds to a restriction operator that extracts simulated CO_2 saturations at each well location in first six snapshots. The objective function reads

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d}_w - \mathbf{M} \circ \mathcal{S}_{\boldsymbol{\theta}^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau, \tag{8}$$

where $\mathbf{d}_w$ represents the well measurements collected at three well locations through the linear restriction operator $\mathbf{M}$. The goal is to invert for the permeability by minimizing the misfit of the well measurements of the CO_2 saturation without and with constraints on the ℓ_2-norm ball in the latent space. The results of these numerical experiments are included in the first row of Fig. 9, where the differences with respect to the ground truth permeability shown in Fig. 6a are plotted in the second row. Because the part of the permeability that is not touched by the CO_2 plume lives in the null space, we highlight the CO_2 plume in the difference plots by dark color and focus on analyzing errors within the plume region. As expected, the unconstrained inversions based on PDE solves (Fig. 9a) and surrogate approximations (Fig. 9b) are both poorly resolved because of the limited spatial information on the saturation. Contrasting these unconstrained inversions with results for the constrained inversions for the PDE (Fig. 9c) and surrogate (Fig. 9d) again shows the importance of adding constraints to the inversion. Figure 9i clearly demonstrates that the FNO prediction errors remain relatively constant during constrained inversion while the error continues to grow during the unconstrained iterations eventually exceeding 14%. Both constrained results improve significantly, even though they converge to different solutions in the end. This is because history matching is typically an ill-posed problem with many distinctive solutions [12]. This observation further motivates us to consider the experiment below, where time-lapse seismic data are jointly inverted for the subsurface permeability.

Multiphysics end-to-end inversion

Next, we consider the alternative setting for seismic monitoring of geological carbon storage, where the dynamics of the CO_2 plumes are indirectly observed from time-lapse seismic data. In this case, the measurement operator, $\mathcal{H}$, involves the composition of the rock physics modeling operator, $\mathcal{R}$, which converts CO_2 saturations to decreases in the compressional wavespeeds for rocks within the reservoir [9], and the seismic modeling
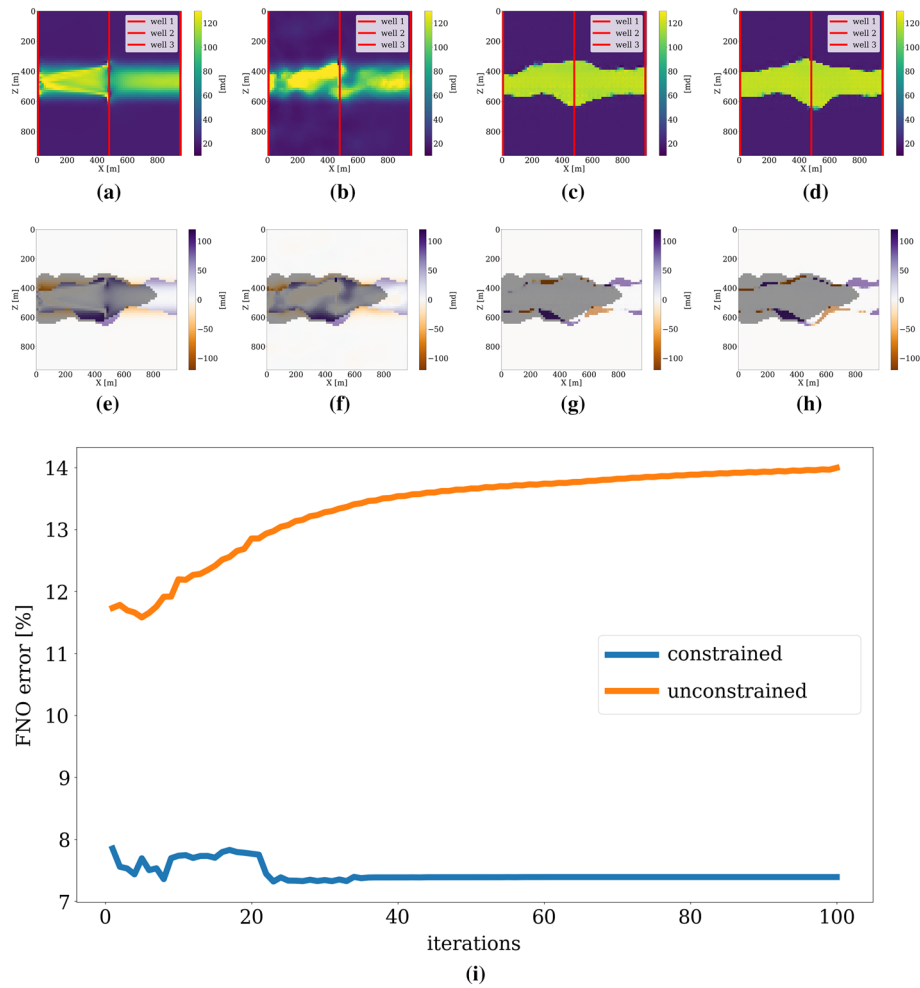
Fig. 9 Permeability inversions from CO_2 saturations sampled at three well locations at 6 early snapshots. The well locations are denoted by the red vertical lines. **a** Unconstrained inversion result based on PDE solves. **b** Same as **a** but now with FNO surrogate approximation. **c** Constrained inversion result based on PDE solves. **d** Same as **c** but now with FNO surrogate approximation. **e**–**h** The error of the permeability inversion results in **a**–**d** compared to the unseen ground truth shown in Fig. 6a. **i** The FNO prediction errors as a function of the number of iterations for **b** and **d**

operator, $\mathcal{F}$, which generates time-lapse seismic data recorded at the receiver locations and based on acoustic wave equation modeling [82]. The multiphysics end-to-end inversion process estimates permeability from time-lapse seismic data via inversion of these nested physics operators for the flow, rock physics, and waves [44]. Following earlier work by Yin et al. [102] and Louboutin, Yin, et al. [54], the fluid-flow PDE modeling is replaced by the trained FNO (cf. Eq. 5), resulting in the following optimization problem:

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d}_{\text{s}} - \mathcal{F} \circ \mathcal{R} \circ \mathcal{S}_{\boldsymbol{\theta}^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau, \tag{9}$$

where $\mathbf{d}_{\text{s}}$ represents the observed time-lapse seismic data. While this end-to-end inversion problem benefits from having remote access to changes in the compressional wavespeed, it may now suffer from null spaces associated with the flow, $\mathcal{S}_{\boldsymbol{\theta}^*}$, and the wave/rock physics, $\mathcal{F} \circ \mathcal{R}$. For instance, the latter suffers from bandwidth limitation of the source function and from limited aperture. Because important components are missing in the observed
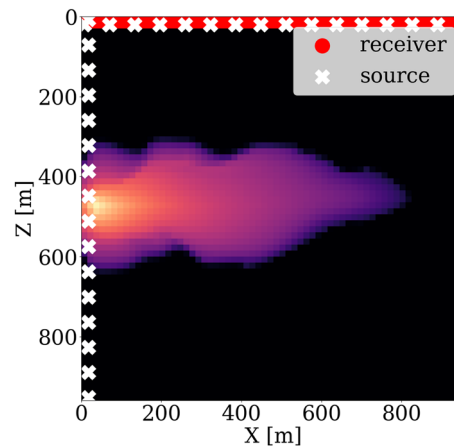
Fig. 10 Seismic acquisition. The white **x** represents the acoustic sources, and the red lines represent the dense receivers. The CO_2 saturation snapshot at day 600 is plotted in the middle, which is the last snapshot that is monitored seismically

data, inversion based on the data objective alone in Eq. 9 are likely to suffer from artifacts that can easily drive the intermediate permeability model iterates out-of-distribution.

To demonstrate capabilities of the proposed relaxed inversion procedure with surrogates for the fluid flow, we assume the baseline to be known—i.e, we assume the brine-filled reservoir with 25% porosity to be acoustically homogeneous prior to CO_2 injection with a compressional wavespeed of 3500 m/s. We use the patchy saturation model [9] to convert the time-dependent CO_2 saturation resulting in < 300 m/s decreases in the wavespeed within the CO_2 plumes. We collect six seismic surveys at the first six snapshots for the CO_2 saturation from day 100 to day 600, which are the same snapshots as the ones used in the previous experiment. For each time-lapse seismic survey, 16 active-seismic sources are located within a well on the left-hand side of the model. We also position 16 sources on the top of the model. Each active source uses a Ricker wavelet with a central frequency of 50 Hz. The transmitted and reflected wavefields are collected by 480 receivers on the top and 480 receivers on the right-hand side of the model. The seismic acquisition is shown in Fig. 10, where the plume at the last seismic vintage (at day 600) is plotted in the middle.

To avoid numerical dispersion, the velocity model is upsampled by a factor of two in both the horizontal and vertical directions, which results in a 7.5 m grid spacing. For the simulations, use is made of the open-source software package JUDI.jl [53, 96] to generate the time-lapse seismic data at the first six snapshots. The fact that this software is based on Devito's wave propagators [51, 58] allows us to do this quickly. For realism, we add 10 dB Gaussian noise to the time-lapse seismic data. Given these six time-lapse vintages, our goal is to invert for the permeability in the reservoir by minimizing the time-lapse seismic data misfit through the nested physics operators shown in Eq. 9.

Inversion results obtained by solving the PDEs for the fluid flow during the inversion are shown in Fig. 11a and c. As before, the inversions benefit majorly from adding the trained NF as a constraint. Remarkably, the end-to-end inversion results shown in Fig. 11a, c, and d are close to the results plotted in Fig. 8a, c, and d, which was obtained with access to the CO_2 saturation everywhere. This reaffirms the notion that time-lapse seismic can indeed provide useful spatial information away from the monitoring wells to estimate the reservoir permeability, which aligns with earlier observations by D. Li et
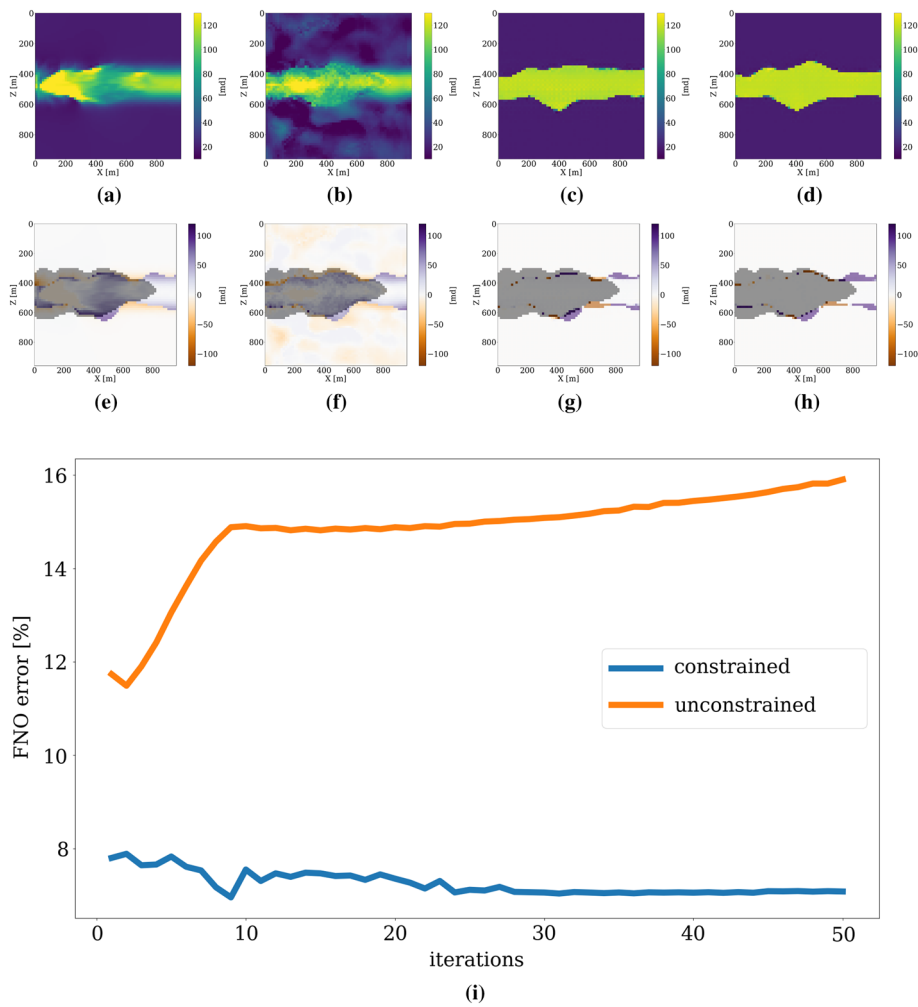
Fig. 11 Permeability inversions from time-lapse seismic data. **a** Inversion result using PDE solvers. **b** The same as **a** but for the FNO surrogate. **c** The same as **a** but with the NF-based constraint. **d** The same as **a** but now for the FNO surrogate with the NF-based constraint. **e–h** The error of the permeability inversion results in **a–d** compared to the unseen ground truth shown in Fig. 6a. **i** The FNO prediction errors as a function of the number of iterations for **b** and **d**

al. [44], Yin et al. [102], and Louboutin, Yin, et al. [54]. Juxtaposing the results for the FNO surrogate without (Fig. 11b) and with the constraint (Fig. 11d) again underlines the importance of adding constraints especially in situations where the forward (wave) operator has a non-trivial nullspace. The presence of this nullspace has a detrimental affect on the unconstrained result obtained by the FNO. Contrary to solutions yielded by the PDE, trained FNOs offer little to no control on the feasibility of the solution, which explains the strong artifacts in Fig. 11b. As we can see from Fig. 11i, these artifacts are mainly due to the FNO-approximation errors that dominate and grow after a few iterations. Conversely, the errors for the constrained case remain more or less flatlined between 7% and 8%. In contrast, using the trained NF as a learned constraint yields better recovery where the errors are minor within the plume region and mostly live on the edges, shown in the second row of Fig. 11.
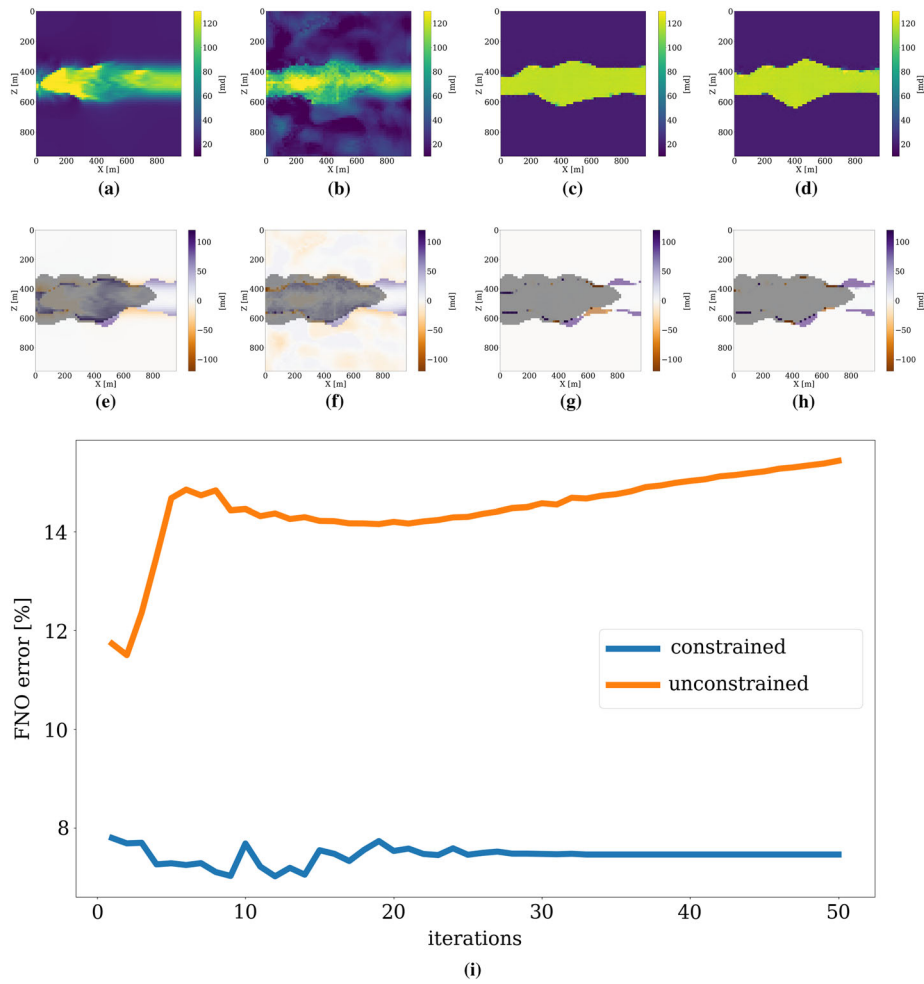
Yin *et al. Advanced Modeling and Simulation in Engineering Sciences*(2023)10:14

Page 18 of 27

Fig. 12 Joint permeability inversions from both time-lapse seismic data and time-lapse well measurements. **a** Inversion result using PDE solvers. **b** The same as **a** but for the FNO surrogate. **c** The same as **a** but with the NF-based constraint. **d** The same as **a** but now for the FNO surrogate with the NF-based constraint. **e**–**h** The error of the permeability inversions in **a**-**d**, compared to the unseen ground truth shown in Fig. 6a. **i** The FNO prediction errors as a function of the number of iterations for **b** and **d**

Table 1 S/N (in dB) and SSIM values of permeability recovery

Inversion method	Well measurement	Time-lapse seismic	Both
Unconstrained inversion with PDE solvers	9.34 dB/0.67	10.50 dB/0.73	10.70 dB / 0.73
Unconstrained inversion with FNO surrogates	9.64 dB/0.68	11.94 dB/0.72	11.98 dB/0.72
Constrained inversion with PDE solvers	12.2 dB / 0.77	14.18 dB / 0.80	15.20 dB/0.85
Constrained inversion with FNO surrogates	11.06 dB/0.74	14.16 dB/0.81	14.92 dB/0.83

Jointly inverting time-lapse seismic data and well measurements

Finally, we consider the most preferred scenario for GCS monitoring, where multiple modalities of data are jointly inverted for the reservoir permeability [32,50]. In our experiment, we consider to jointly invert time-lapse seismic data and well measurements by minimizing the following objective function:

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d}_s - \mathcal{F} \circ \mathcal{R} \circ \mathcal{S}_{\boldsymbol{\theta}^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 + \lambda \|\mathbf{d}_w - \mathbf{M} \circ \mathcal{S}_{\boldsymbol{\theta}^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau.$$

$$(10)$$

This objective function includes both the time-lapse seismic data misfit from Eq. 9 and the time-lapse well measurement misfit from Eq. 8 with a balancing term λ. While better choices can be made, we select this λ in our numerical experiment to be 10, so that the magnitudes of the two terms are relatively the same. The inversion results and differences from the unseen ground truth permeability are shown in Fig. 12, where we again observe large artifacts for the recovery when FNO surrogate is inverted without NF constraints. This behavior is confirmed by the plot for the FNO error curve as a function of the number of iterations. This error finally reaches a value over 15%.

We report quantitative measures for the permeability inversions for all optimization methods and different types of observed data in Table 1 for the signal-to-noise ratios (S/Ns) and the structural similarity index measure (SSIM, Wang et al. [90]). To avoid undue influence of the null space for the permeability, we only calculate the S/N and SSIM values based on the parts of the models that are touched by CO_2 plume. From these values, following observations can be made. First, the NF-constrained permeability inversion are superior in both S/Ns and SSIMs, which demonstrates the efficacy of the learned constraint. Second, by virtue of this NF constraint, the results yielded by either the PDE solver or by the FNO surrogate produce very similar S/Ns and SSIMs. This behavior reaffirms that the trained FNO behavior is similar to the behavior yielded by PDE solver when its inputs remain in-distribution, which is controlled by the NF constraints.

CO_2 plume estimation and forecast

While end-to-end permeability inversion from time-lapse data provides novel access to this important fluid-flow property, the real interest in monitoring GCS lies in determining where CO_2 plumes are and will be in the foreseeable future, say of 100 and 200 days ahead. To demonstrate the value of the proposed surrogates and of the use of time-lapse seismic data, as opposed to time-lapse saturation data measured at the wells only, we in Fig. 7 juxtapose CO_2 predictions obtained from fluid-flow simulations based on the inverted permeabilities in situations where either well data is available (first row), or where time-lapse seismic data is available (second row), or where both data modalities are available (third row). These results are achieved by first inverting for permeabilities using FNO surrogates and NF constraints, followed by running the fluid-flow simulations for additional time steps given the inverted permeabilities yielded by well-only (Fig. 9d), time-lapse data (Fig. 11d), and both (Fig. 12d). From these plots, we draw the following two conclusion. First, the predicted CO_2 plumes estimated from seismic data are significantly more accurate than those obtained by inverting time-lapse saturations measured at the wells only. As expected, there are large errors in the regions away from the wells for the CO_2 plumes estimated from wells shown in the fourth row of Fig. 13. Second, thanks to the NF-constraint, the CO_2 predictions obtained with the computationally beneficial surrogate approximation remain close to the ground truth CO_2 plume plotted in the first row of Fig. 7, with only minor artifacts at the edges. Third, using both seismic data and well measurements produces CO_2 plume predictions with the smallest errors, while the uplift of well measurements on top of seismic observations is modest (comparing the second and the third rows of Fig. 13). Finally, while the CO_2 plume estimates for the past (monitored) vintages (i.e. first three columns of the third row of Fig. 13) are accurate, the near-future forecasts without time-lapse well or seismic data (i.e. last two columns of the
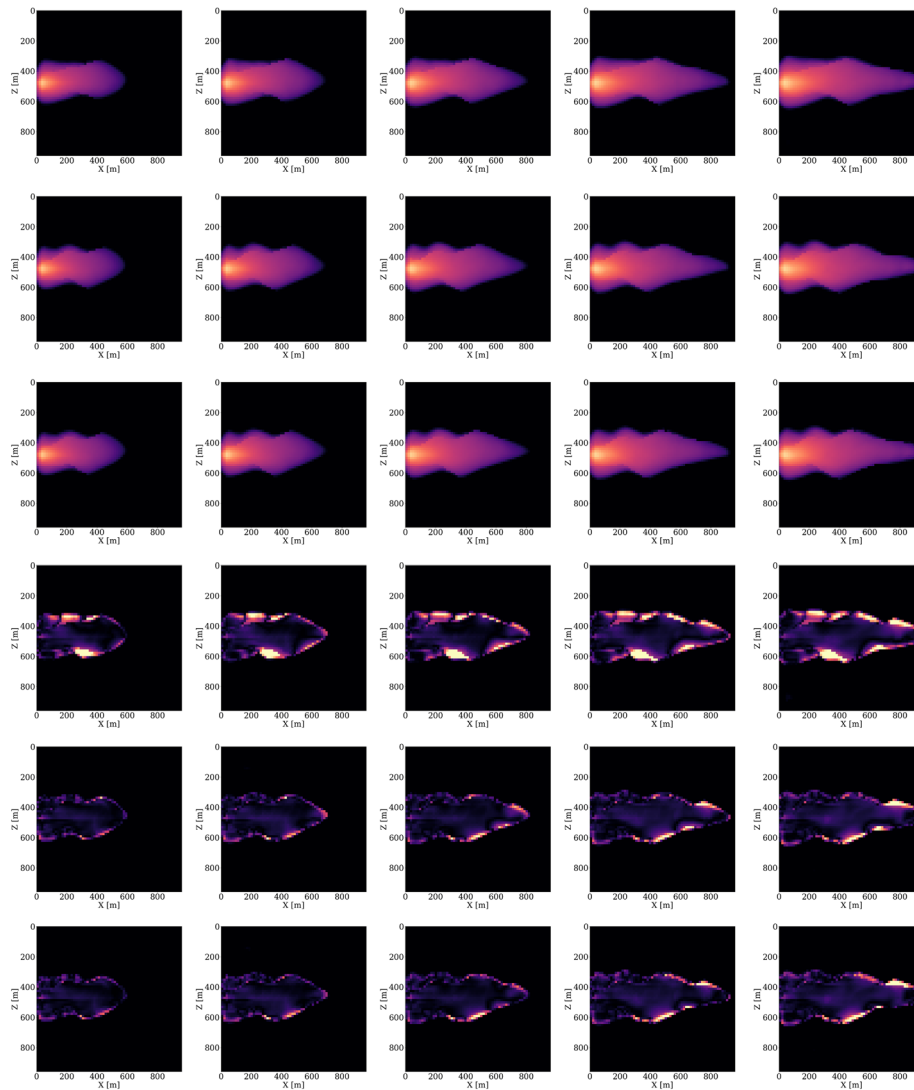
Yin *et al. Advanced Modeling and Simulation in Engineering Sciences* (2023) 10:14

Page 20 of 27

Fig. 13 CO$_2$ plume estimation and forecast using FNO surrogates and NF constraints to invert different modalities of observed data. The first three columns represent past CO$_2$ saturations at day 400, 500, and 600 of the first 600 days of CO$_2$ saturation monitored either through the well measurements or time-lapse data. The last two columns include forecasts for the saturations at future days 700 and 800, where no observed data is available. The first row shows the past and future CO$_2$ estimates yielded by inverting well measurements only. The second row is the same but now inverting time-lapse seismic data. The third row is the same but now jointly inverting well measurements and time-lapse seismic data. The fourth, fifth, and sixth rows show 5× difference between the ground truth CO$_2$ plume (first row of Fig. 7) and the first, second, third row, respectively. The S/Ns for the first, the second, and the third rows are 15.26 dB, 20.14 dB, 20.46 dB, respectively

third row of Fig. 13) could be less accurate. This is because the right-hand side and the bottom of the permeability model are not touched yet by the CO$_2$ plume during the first 600 days. As a result, the error on the permeability recovery on the right-hand side leads to the slightly larger errors on the CO$_2$ plume forecast. Overall, these CO$_2$ forecasts for the future 100 and 200 days match the general trend of the CO$_2$ plume without any observed data despite minor errors. A continuous monitoring system, where multiple modalities of data are being acquired and inverted throughout the GCS project, could allow for updating the reservoir permeability and forecasting the CO$_2$ plume consistently.

Analysis of computational gains

FNOs, and deep neural surrogates in general, have the potential to be orders of magnitude faster than conventional PDE solvers [46], and this speed-up is generally problem-dependent. In our numerical experiments, the PDE solver from Jutul.jl [60,61,100] currently only supports CPUs and we find an average runtime for both the forward and gradient on the 64×64 model to be 10.6 s on average on an 8-core Intel(R) Xeon(R) W-2245 CPU. The trained FNO, implemented using modules from Flux.jl [33], takes 16.4 milliseconds on average for both the forward and gradient. This means that the trained FNO in our case provides $646\times$ speed up compared to conventional PDE solvers. The training of FNO takes about 4 h on an NVIDIA T1000 8GB GPU. Given these numbers, we can calculate the break-even point—i.e., the point where using FNO surrogate becomes cheaper in terms of the overall runtime, by the following formula:

$$\text{Breakeven} = \frac{\text{Generating training set time} + \text{training time}}{\text{PDE solver runtime} - \text{FNO runtime}} \approx 3364. \tag{11}$$

This means that after 3364 calls to the forward simulator and its gradients, the computational savings gained from using the FNO surrogate evaluations during the inversion process balances out the initial upfront costs. These upfront costs include the generation of the training dataset and the training of the FNO. Therefore, after this break-even point of 3364 calls, the use of the FNO surrogate becomes more cost-effective compared to the conventional PDE solver. Because the trained FNO has the potential to generalize to different kinds of inversion problems, and potentially also different GCS sites, 3364 calls is justifiable in practice. However, we acknowledge that a more detailed analysis on a more realistic 4D scale problem will be necessary to understand the potential computational gains and tradeoffs of the proposed methodology. For details on a high-performance computing parallel implementation of FNOs, we refer to Grady et al. [25] who also conducted a realistic performance on large-scale 4D multiphase fluid-flow problems. Even in cases where the computational advances are perhaps challenging to justify, the use of FNOs has the additional benefit by providing access to the gradient with respect to model parameters (i.e. permeability) through automatic differentiation. This feature is important since it is an enabler for inversion problems that involve complex PDE solvers for which gradients are often not readily available, e.g. Gross and Mazuyer [27]. By training FNOs on input–output pairs, "gradient-free" gradient-based inversion is made possible in situations where the simulator does not support gradients.

Discussion and conclusions

Monitoring of geological carbon storage is challenging because of excessive computational needs and demands on data collection by drilling monitor wells or by collecting time-lapse seismic data. To offset the high computational costs of solving multiphase flow equations and to improve permeability inversions from possibly multimodal time-lapse data, we introduce the usage of trained Fourier neural operators (FNOs) that act as surrogates for the fluid-flow simulations. We propose to do this in combination with trained normalizing flows (NFs), which serve as regularizers to keep the inversion and the accuracy of the FNOs in check. Since the computational expense of FNO's online evaluation is negligible compared to numerically expensive partial differential equation solves, FNOs only incur upfront offline training costs. While this obviously presents a major advantage, the

approximation accuracy of FNOs is, unfortunately, only guaranteed when its argument, the permeability, is in distribution—i.e., is drawn from the same distribution as the FNO was trained on. This creates a problem because there is, thanks to the non-trivial null space of permeability inversion, no guarantee the model iterates remain in-distribution. Quite the opposite, our numerical examples show that these iterates typically move out-of-distribution during the (early) iterations. This results in large errors in the FNO and in rather poor inversion results for the permeability.

To overcome this out-of-distribution dilemma for the model iterates during permeability inversion with FNOs, we propose adding learned constraints, which ensure that model iterates remain in-distribution during the inversion. We accomplish this by training a NF on the same training set for the permeability used to train the FNO. After training, the NF is capable of generating in-distribution samples for the permeability from random realizations of standard Gaussian noise in the latent space. We employ this learned ability by parameterizing the unknown permeability in the latent space, which offers additional control on whether the model iterates remain in-distribution during the inversion. After establishing that out-of-distribution permeability models can be mapped to in-distribution models by restricting the ℓ_2-norm of their latent representation, we introduce permeability inversion as a constrained optimization problem where the data misfit is minimized subject to a constraint on the ℓ_2-norm of the latent space. Compared to adding pretrained NFs as priors via additive penalty terms, use of constraints ensures that model iterates remain at all times in-distribution. We show that this holds as long as the size of constraint set does not exceed the size of the ℓ_2-norm ball of the standard normal distribution. As a result, we arrive at a computationally efficient continuation scheme, known as a homotopy, during which the ℓ_2-norm constraint is relaxed slowly, so the data misfit objectives can be minimized while the model iterates remain in distribution.

By means of a series of carefully designed numerical experiments, we were able to establish the advocacy of combining learned surrogates and constraints, yielding solutions to permeability inversion problems that are close to solutions yielded by costly PDE-based methods. The examples also clearly show the advantages of working with gradually relaxed constraints where model iterates remain at all times in distribution with the additional joint benefit of slowly building up the model while bringing down the data misfit, an approach known to mitigate the effects of local minima [20,72,73]. Consequently, the quality of all time-lapse inversions improved significantly without requiring information that goes beyond having access to the training set of permeability models.

While we applied the proposed method to gradient-based iterative inversion, a similar approach can be used for other types of inversion methods, including inference with Markov chain Monte Carlo methods for uncertainty quantification [42]. We also envisage extensions of the proposed method to other physics-based inverse problems [28,99] and simulation-based inference problems [17], where numerical simulations often form the computational bottleneck.

Despite the encouraging results from the numerical experiments, the presented approach leaves room for improvements, which we will leave for future work. For instance, the gradient with respect to the model parameters (permeability) derived from the neural surrogate is not guaranteed to be accurate—e.g. close to the gradient yielded by the adjoint-state method. As recent work by O'Leary-Roseberry et al. [65] has shown, this potential source of error can be addressed by training neural surrogates on the simulator's

Yin *et al. Advanced Modeling and Simulation in Engineering Sciences*(2023)10:14

Page 23 of 27

gradient with respect to the model parameters, provided it is available. Unfortunately, deriving gradients of complex high-performance computing implementations of numerical PDE solvers is often extremely challenging, explaining why this information is often not available. Because our method solely relies on gradients of the surrogate, which are readily available through algorithmic differentiation, we only need access to numerical PDE solvers available in legacy codes. While this approach may go at the expense of some accuracy, this feature offers a distinct practical advantage. However, as with many other machine learning approaches, our learned methods may also suffer from time-lapse observations that are out-of-distribution—i.e., produced by a permeability model that is out-of-distribution. While this is a common problem in data-driven methods, recent developments [85] may remedy this problem by applying latent space corrections, a solution that is amenable to our approach. On the other hand, expanding the latent space's ℓ_2 norm ball during inversion would allow NFs to generate any out-of-distribution model parameter. However, in that case the accuracy of the learned surrogate is not guaranteed. For such cases, transitioning from the learned surrogate to the numerical solver during later iterations may be advantageous and merits further study. The choice for the size of the ℓ_2-norm ball at the beginning and at the end can also be further investigated [8].

While our paper primarily presents a proof of concept through a relatively small 2D experiment, our inversion strategy is designed to scale to large-scale 3D problems. NFs, with their inherent memory efficiency due to invertibility, are already primed for extension to 3D problems. For the learned surrogates, Grady et al. [25] showcases model-parallel FNOs, demonstrating success in simulating 4D multiphase flow physics of over 2.6 billion variables. By combining these strengths, we are optimistic scaling this inversion strategy to 3D.

To end on a positive note and forward looking note, we argue that the presented approach makes a strong case for the inversion of multimodal data, consisting of time-lapse well and seismic data. While inversions from time-lapse saturation data collected from wells are feasible and fall within the realm of reservoir engineering, their performance, as expected, degrades away from the well. We argue that adding active-source seismic provides essential fill-in away from the wells. As such, it did not come to our surprise that joint inversion of multimodal data resulted in the best permeability estimates. From our perspective, our successful combination of these often disjoint data modalities holds future promise when addressing challenges that come with monitoring and control of geological carbon storage and enhanced geothermal systems.

Abbreviations

FNO   Fourier neural operator
GCS   Geological carbon storage
NF   Normalizing flow
PDE   Partial differential equations

Author contributions

 ZY concepted the general outline of the project, implemented the algorithm, designed and conducted the numerical experiments, and wrote the manuscript. RO helped concept the general outline of the project, discussed the approach and helped design the numerical experiments. ML helped concept the general outline of the project and discussed the approach. FH concepted the general outline of the project, discussed the approach, helped design the numerical

Yin *et al. Advanced Modeling and Simulation in Engineering Sciences* (2023) 10:14

Page 24 of 27

experiments, edited the manuscript, and supervised the project. All authors reviewed the results and approved the final version of the manuscript.

Availability of data and materials
The scripts to reproduce the experiments are available on the SLIM GitHub page https://github.com/slimgroup/FNO-NF.jl.

Declarations

Competing interests
The authors declare that they have no competing interests.

References

1. Alexanderian A. Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: a review. Inverse Problems. 2021;37(4): 043001.
2. Alnes H, Eiken O, Nooner S, Sasagawa G, Stenvold T, Zumberge M. Results from sleipner gravity monitoring: updated density and temperature distribution of the CO_2 plume. Energy Procedia. 2011;4:5504–11.
3. Ardizzone L, Kruse J, Wirkert S, Rahner D, Pellegrini EW, Klessen RS, Maier-Hein L, Rother C, Köthe U. Analyzing inverse problems with invertible neural networks; 2018. arXiv Preprint arXiv:1808.04730.
4. Arridge SR. Optical tomography in medical imaging. Inverse problems. 1999;15(2):R41.
5. Arts R, Eiken O, Chadwick A, Peter Z, Van der Meer L, Zinszner B. Monitoring of CO_2 injected at Sleipner using time-lapse seismic data. Energy. 2004;29(9–10):1383–92.
6. Asher MJ, Croke BFW, Jakeman AJ, Peeters LJM. A review of surrogate models and their application to groundwater modeling. Water Resour Res. 2015;51(8):5957–73.
7. Asim M, Daniels M, Leong O, Ahmed A, Hand P. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In: International conference on machine learning; 2020. p. 399–409. PMLR.
8. Aster RC, Borchers B, Thurber CH. Parameter estimation and inverse problems. Amsterdam: Elsevier; 2018.
9. Avseth P, Mukerji T, Mavko G. Quantitative seismic interpretation: applying rock physics tools to reduce interpretation risk. Cambridge: Cambridge University Press; 2010.
10. Beck A. Introduction to nonlinear optimization: theory, algorithms, and applications with MATLAB. Philadelphia: SIAM; 2014.
11. Benitez JAL, Furuya T, Faucher F, Tricoche X, de Hoop MV. Fine-tuning neural-operator architectures for training and generalization; 2023. arXiv Preprint arXiv:2301.11509.
12. Canchumuni SWA, Emerick AA, Pacheco MAC. History matching geological facies models based on ensemble smoother and deep generative models. J Pet Sci Eng. 2019;177:941–58.
13. Cao Y, Li S, Petzold L, Serban R. Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution. SIAM J Sci Comput. 2003;24(3):1076–89.
14. Carcione JM, Gei D, Picotti S, Michelini A. Cross-hole electromagnetic and seismic modeling for CO_2 detection and monitoring in a saline aquifer. J Pet Sci Eng. 2012;100:162–72.
15. Chandra R, Azam D, Kapoor A, Müller DR. Surrogate-assisted Bayesian inversion for landscape and basin evolution models. Geosci Model Dev. 2020;13(7):2959–79.
16. Cowles MK, Carlin BP. Markov Chain Monte Carlo convergence diagnostics: a comparative review. J Am Stat Assoc. 1996;91(434):883–904.
17. Cranmer K, Brehmer J, Louppe G. The frontier of simulation-based inference. Proc Natl Acad Sci. 2020;117(48):30055–62.
18. De Hoop M, Huang DZ, Qian E, Stuart AM. The cost-accuracy trade-off in operator learning with neural networks; 2022. arXiv Preprint arXiv:2203.13181.
19. Dinh L, Sohl-Dickstein J, Bengio S. Density estimation using real nvp; 2016. arXiv Preprint arXiv:1605.08803.
20. Esser E, Guasch L, van Leeuwen T, Aravkin AY, Herrmann FJ. Total variation regularization strategies in full-waveform inversion. SIAM J Imaging Sci. 2018;11(1):376–406.
21. Freifeld BM, Daley TM, Hovorka SD, Henninges J, Underschultz J, Sharma S. Recent advances in well-based monitoring of CO_2 sequestration. Energy Procedia. 2009;1(1):2277–84.
22. Furre AK, Eiken O, Vevatne Alnes H, Vevatne JN, Kiær AF. 20 years of monitoring CO2-injection at sleipner. Energy Procedia. 2017;114:3916–26. https://doi.org/10.1016/j.egypro.2017.03.1523.
23. Golub GH, Hansen PC, O'Leary DP. Tikhonov regularization and total least squares. SIAM J Matrix Anal Appl. 1999;21(1):185–94.
24. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Adv Neural Inf Process Syst. 2014; 27. https://proceedings.neurips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.
25. Grady TJ, Khan R, Louboutin M, Yin Z, Witte PA, Chandra R, Hewett RJ, Herrmann FJ. Model-parallel fourier neural operators as learned surrogates for large-scale parametric PDEs. Comput Geosci. 2023; 105402. https://doi.org/10.1016j.cageo.2023.105402.
26. Griewank A, et al. On automatic differentiation. Math Progr Recent Dev Appl. 1989;6(6):83–107.

27. Gross H, Mazuyer A. GEOSX: a multiphysics, multilevel simulator designed for exascale computing. In: SPE reservoir simulation conference OnePetro; 2021
28. Heidenreich S, Gross H, Henn MA, Elster C, Bär M. A surrogate model enables a Bayesian approach to the inverse problem of scatterometry. J Phys Conf Ser. 2014;490: 012007. (1. IOP publishing)
29. Hennenfent G, van den Berg E, Friedlander MP, Herrmann FJ. New insights into one-norm solvers from the pareto curve. Geophysics. 2008;73(4):A23-26.
30. Hijazi S, Freitag M, Landwehr N. POD-galerkin reduced order models and physics-informed neural networks for solving inverse problems for the Navier–Stokes equations. Adv Model Simul Eng Sci. 2023;10(1):1–38.
31. Horvat C, Pfister JP. Intrinsic dimensionality estimation using normalizing flows. Adv Neural Inf Process Syst. 2022;35:12225–36.
32. Huang L. Geophysical monitoring for geologic carbon storage. Hoboken: Wiley; 2022.
33. Innes M. Flux: elegant machine learning with Julia. J Open Source Softw. 2018;3(25):602.
34. Jansen JD. Adjoint-based optimization of multi-phase flow through porous media—a review. Comput Fluids. 2011;46(1):40–51.
35. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat Rev Phys. 2021;3(6):422–40.
36. Kingma DP, Welling M. Auto-encoding variational Bayes; 2013. arXiv Preprint arXiv:1312.6114.
37. Kobyzev I, Prince SJD, Brubaker MA. Normalizing flows: an introduction and review of current methods. IEEE Trans Pattern Anal Mach Intell. 2020;43(11):3964–79.
38. Kontolati K, Goswami S, Karniadakis GE, Shields MD. Learning in latent spaces improves the predictive accuracy of deep neural operators; 2023. arXiv Preprint arXiv:2304.07599.
39. Kovachki N, Lanthaler S, Mishra S. On universal approximation and error bounds for Fourier neural operators. J Mach Learn Res. 2021;22:13237–312.
40. Kovachki N, Li Z, Liu B, Azizzadenesheli K, Bhattacharya K, Stuart A, Anandkumar A. Neural operator: learning maps between function spaces; 2021. arXiv Preprint arXiv:2108.08481.
41. Kruse J, Detommaso G, Köthe U, Scheichl R. HINT: hierarchical invertible neural transport for density estimation and Bayesian inference. Proc AAAI Conf Artif Intell. 2021;35(8191–99):9.
42. Lan S, Li S, Shahbaba B. Scaling up Bayesian uncertainty quantification for inverse problems using deep neural networks. SIAM/ASA J Uncertainty Quant. 2022;10(4):1684–713.
43. Lensink K, Peters B, Haber E. Fully hyperbolic convolutional neural networks. Res Math Sci. 2022;9(4):60.
44. Li D, Xu K, Harris JM, Darve E. Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation. Water Resour Res. 2020;56(8):e2019WR027032.
45. Li X, Aravkin AY, van Leeuwen T, Herrmann FJ. Fast Randomized full-waveform inversion with compressive sensing. Geophysics. 2012;77(3):A13-17.
46. Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A, Anandkumar A. Fourier neural operator for parametric partial differential equations; 2020. arXiv Preprint arXiv:2010.08895.
47. Li Z, Zheng H, Kovachki N, Jin D, Chen H, Liu B, Azizzadenesheli K, Anandkumar A. Physics-informed neural operator for learning partial differential equations; 2021. arXiv Preprint arXiv:2111.03794.
48. Lie KA. An introduction to reservoir simulation using MATLAB/GNU octave: user guide for the MATLAB reservoir simulation toolbox (MRST). Cambridge: Cambridge University Press; 2019.
49. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Progr. 1989;45(1–3):503–28.
50. Liu M, Vashisth D, Grana D, Mukerji T. Joint inversion of geophysical data for geologic carbon sequestration monitoring: a differentiable physics-informed neural network model. J Geophys Res Solid Earth. 2023;128(3):2022JB025322JB0e2025372.
51. Louboutin M, Luporini F, Lange M, Kukreja N, Witte PA, Herrmann FJ, Velesko P, Gorman GJ. Devito (V3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. Geosci Model Dev. 2019. https://doi.org/10.5194/gmd-12-1165-2019.
52. Louboutin M, Witte PA, Siahkoohi A, Rizzuti G, Yin Z, Orozco R, Herrmann FJ. Accelerating innovation with software abstractions for scalable computational geophysics; 2022. https://doi.org/10.1190/image2022-3750561.1.
53. Louboutin M, Witte P, Yin Z, Modzelewski H, Carlos da Costa K, Nogueira P. Slimgroup/JUDI.jl: V3.2.3 (version v3.2.3). Zenodo; 2023. https://doi.org/10.5281/zenodo.7785440.
54. Louboutin M, Yin Z, Orozco R, Grady II TJ, Siahkoohi A, Rizzuti G, Witte PA, Møyner O, Gorman GJ, Herrmann FJ. Learned multiphysics inversion with differentiable programming and machine learning. Lead Edge. 2023;42(7):452–516. https://doi.org/10.1190/tle42070474.1.
55. Lu K, Jin Y, Chen Y, Yang Y, Hou L, Zhang Z, Li Z, Chao F. Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems. Mech Syst Signal Process. 2019;123:264–97.
56. Lu L, Jin P, Karniadakis GE. Deeponet: learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators; 2019." arXiv Preprint arXiv:1910.03193.
57. Lumley D. 4D seismic monitoring of CO2 sequestration. Lead Edge. 2010;29(2):150–5.
58. Luporini F, Louboutin M, Lange M, Kukreja N, Witte P, Hückelheim J, Yount C, Kelly PHJ, Herrmann FJ, Gorman GJ. Architecture and performance of devito, a system for automated stencil computation. ACM Trans Math Softw(TOMS). 2020;46(1):1–28.
59. Mosser L, Dubrule O, Blunt MJ. Deepflow: history matching in the space of deep generative models; 2019. arXiv Preprint arXiv:1905.05749.
60. Møyner O, Bruer G, Yin Z. Sintefmath/JutulDarcy.jl: V0.2.3 (version v0.2.3). Zenodo; 2023. https://doi.org/10.5281/zenodo.7855628.
61. Møyner O, Johnsrud M, Nilsen HM, Raynaud X, Lye KO, Yin Z. Sintefmath/Jutul.jl: V0.2.6 (version v0.2.6). Zenodo; 2023. https://doi.org/10.5281/zenodo.7855605.
62. Nogues JP, Nordbotten JM, Celia MA. Detecting leakage of brine or CO2 through abandoned wells in a geological sequestration operation using pressure monitoring wells. Energy Procedia. 2011;4:3620–7.

Yin *et al. Advanced Modeling and Simulation in Engineering Sciences*(2023)10:14

Page 26 of 27

63. Nooner SL, Eiken O, Hermanrud C, Sasagawa GS, Stenvold T, Zumberge MA. Constraints on the in situ density of CO2 within the Utsira formation from time-lapse seafloor gravity measurements. Int J Greenhouse Gas Control. 2007;1(2):198–214.

64. Nordbotten JM, Celia MA. Geological storage of CO2: modeling approaches for large-scale simulation. Hoboken: John Wiley Sons; 2011.

65. O'Leary-Roseberry T, Chen P, Villa U, Ghattas O. Derivate informed neural operator: an efficient framework for high-dimensional parametric derivative learning; 2022. arXiv Preprint arXiv:2206.10745.

66. Orozco R, Louboutin M, Herrmann FJ. Memory efficient invertible neural networks for 3D photoacoustic imaging; 2022. arXiv Preprint arXiv:2204.11850.

67. Orozco R, Louboutin M, Siahkoohi A, Rizzuti G, van Leeuwen T, Herrmann F. Amortized normalizing flows for transcranial ultrasound with uncertainty quantification; 2023. https://openreview.net/forum?id=LoJG-IUIlk.

68. Orozco R, Siahkoohi A, Louboutin M, Herrmann FJ. Refining amortized posterior approximations using gradient-based summary statistics; 2023. arXiv Preprint arXiv:2305.08733.

69. Orozco R, Siahkoohi A, Rizzuti G, van Leeuwen T, Herrmann FJ. Photoacoustic imaging with conditional priors from normalizing flows; 2021. https://openreview.net/forum?id=woi1OTvROO1.

70. Orozco R, Siahkoohi A, Rizzuti G, van Leeuwen T, Herrmann FJ. Adjoint operators enable fast and amortized machine learning based Bayesian uncertainty quantification. In: Medical imaging 2023: image processing, edited by Olivier Colliot and Ivana IÅ¡gum, 12464:124641L. International Society for Optics; Photonics; SPIE; 2023. https://doi.org/10.1117/12.2651691.

71. Pestourie R, Mroueh Y, Nguyen TV, Das P, Johnson SG. Active learning of deep surrogates for PDEs: application to metasurface design. NPJ Comput Mater. 2020;6(1):164.

72. Peters B, Herrmann FJ. Constraints versus penalties for edge-preserving full-waveform inversion. Lead Edge. 2017;36(1):94–100.

73. Peters B, Smithyman BR, Herrmann FJ. Projection methods and applications for seismic nonlinear inverse problems with multiple constraints. Geophysics. 2019;84(2):R251-69.

74. Plessix RE. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophys J Int. 2006;167(2):495–503.

75. Powell MJD. Radial basis functions for multivariable interpolation: a review. Algorithms for the approximation of functions and data; 1985. https://doi.org/10.5555/48424.48433.

76. Qian E, Kramer B, Peherstorfer B, Willcox K. Lift & learn: physics-informed machine learning for large-scale nonlinear dynamical systems. Phys D Nonlinear Phenomena. 2020;406: 132401.

77. Rahman MA, Ross ZE, Azizzadenesheli K. U-No: u-shaped neural operators; 2022. arXiv Preprint arXiv:2204.11127.

78. Rasmussen A, Sandve TH, Bao K, Lauser A, Hove J, Skaflestad B, Klofkorn R, Blatt M, Rustad AB, Saevareid O, Lie KA. The open porous media flow reservoir simulator. Comput Math Appl. 2021;81:159–85.

79. Razavi S, Tolson BA, Burn DH. Review of surrogate modeling in water resources. Water Resour Res. 2012. https://doi.org/10.1029/2011WR011527.

80. Rezende D, Mohamed S. Variational inference with normalizing flows. In: International conference on machine learning. PMLR.; 2015. p. 1530–38.

81. Schilders WHA, Van der Vorst HA, Rommes J. Model order reduction: theory, research aspects and applications, vol. 13. Berlin: Springer; 2008.

82. Sheriff RE, Geldart LP. Exploration seismology. Cambridge: Cambridge University Press; 1995.

83. Siahkoohi A, Herrmann FJ. Learning by example: fast reliability-aware seismic imaging with normalizing flows. In: First international meeting for applied geoscience & energy. Society of exploration geophysicists; expanded abstracts; 2021. p. 1580–85 https://doi.org/10.1190/segam2021-3581836.1.

84. Siahkoohi A, Rizzuti G, Louboutin M, Witte PA, Herrmann FJ. Preconditioned training of normalizing flows for variational inference in inverse problems; 2021. https://openreview.net/pdf?id=P9m1sMaNQ8T.

85. Siahkoohi A, Rizzuti G, Orozco R, Herrmann FJ. Reliable amortized variational inference with physics-based latent distribution correction. Geophysics. 2023;88(3):R297-322.

86. Siahkoohi A, Rizzuti G, Orozco R, Herrmann FJ. Reliable amortized variational inference with physics-based latent distribution correction. Geophysics. 2023. https://doi.org/10.1190/geo2022-0472.1.

87. Stanimirović PS, Miladinović MB. Accelerated gradient descent methods with line search. Numer Algorithms. 2010;54:503–20.

88. Tarantola A. Inversion of seismic reflection data in the acoustic approximation. Geophysics. 1984;49(8):1259–66.

89. Tarantola A. Inverse problem theory and methods for model parameter estimation. SIAM; 2005.

90. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12.

91. Wen G, Li Z, Azizzadenesheli K, Anandkumar A, Benson SM. U-FNO—an enhanced fourier neural operator-based deep-learning model for multiphase flow. Adv Water Resour. 2022;163: 104180.

92. Wen G, Li Z, Long Q, Azizzadenesheli K, Anandkumar A, Benson SM. Real-time high-resolution CO2 geological storage prediction using nested fourier neural operators. Energy Environ Sci. 2023;16:1732–41.

93. Wen J, Ahmad R, Schniter P. A conditional normalizing flow for accelerated multi-coil MR imaging; 2023. arXiv Preprint arXiv:2306.01630.

94. Witte PA, Hewett RJ, Saurabh K, Sojoodi A, Chandra R. SciAI4Industry—solving PDEs for industry-scale problems with deep learning; 2022. arXiv Preprint arXiv:2211.12709.

95. Witte PA, Konuk T, Skjetne E, Chandra R. Fast CO2 saturation simulations on large-scale geomodels with artificial intelligence-based wavelet neural operators. Int J Greenhouse Gas Control. 2023;126: 103880.

96. Witte PA, Louboutin M, Kukreja N, Luporini F, Lange M, Gorman GJ, Herrmann FJ. A large-scale framework for symbolic implementations of seismic inversion algorithms in Julia. Geophysics. 2019;84(3):F57-71. https://doi.org/10.1190/geo2018-0174.1.

97. Witte PA, Redmond WA, Hewett RJ, Chandra R. Industry-scale CO2 flow simulations with model-parallel fourier neural operators. In: NeurIPS 2022 workshop tackling climate change with machine learning; 2022.

Yin *et al. Advanced Modeling and Simulation in Engineering Sciences*(2023)10:14

Page 27 of 27

98. Witte P, Louboutin M, Orozco R, Grizzuti AS, Herrmann F, Peters, Páll Haraldsson B, Yin Z. Slim-group/InvertibleNetworks.jl: V2.2.5 (version v2.2.5). Zenodo; 2023. https://doi.org/10.5281/zenodo.7850287.

99. Yang Y, Gao AF, Azizzadenesheli K, Clayton RW, Ross ZE. Rapid seismic waveform modeling and inversion with neural operators. IEEE Trans Geosci Remote Sens. 2023;61:1–12.

100. Yin Z, Grant B, Mathias L. Slimgroup/JutulDarcyRules.jl: V0.2.5 (version v0.2.5). Zenodo; 2023. https://doi.org/10.5281/zenodo.7863970.

101. Yin Z, Erdinc HT, Gahlot AP, Louboutin M, Herrmann FJ. Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection. Lead Edge. 2023;42(1):69–76. https://doi.org/10.1190/tle42010069.1.

102. Yin Z, Siahkoohi A, Louboutin M, Herrmann FJ. Learned coupled inversion for carbon sequestration monitoring and forecasting with fourier neural operators; 2022. https://doi.org/10.1190/image2022-3722848.1.

103. Zhang X, Curtis A. Seismic tomography using variational inference methods. J Geophys Res Solid Earth. 2020;125(4):e2019JB018589.

104. Zhang X, Curtis A. Bayesian geophysical inversion using invertible neural networks. J Geophys Res Solid Earth. 2021;126(7):e2021JB022320.

105. Zhao X, Curtis A, Zhang X. Bayesian seismic tomography using normalizing flows. Geophys J Int. 2021;228(1):213–39. https://doi.org/10.1093/gji/ggab298.

106. Zhdanov MS, Endo M, Black N, Spangler L, Fairweather SA, Hibbs A, Eiskamp GA, Will R. Electromagnetic monitoring of CO2 sequestration in deep reservoirs. First Break. 2013;31(2). https://doi.org/10.3997/1365-2397.31.2.66662.

Publisher's Note