AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR037953

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

A Fluid Flow-Based Deep Learning (FFDL) Architecture for Subsurface Flow Systems With Application to Geologic CO_2 Storage

Zhen Qin[1], Yingxiang Liu[2], Fangning Zheng[1], and Behnam Jafarpour[1]

[1]Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA, USA, [2]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA

Key Points:

- Novel physics-encoded architecture: The Fluid Flow-based Deep Learning (FFDL) architecture features a fluid flow physics-based encoder and a residual-based processor, together with a physics-based operator to capture physical relationships
- Flexibility and performance: FFDL combines the flexibility of deep learning models with the structure of fluid flow equations, resulting in superior performance compared to standard deep learning models such as RUNET
- Application to geologic CO_2 storage: The model is applied to predict the evolution of pressure and saturation fields during geologic CO_2 storage under dynamic injection rates and permeability variations

Correspondence to:

B. Jafarpour,
behnam.jafarpour@usc.edu

Citation:

Qin, Z., Liu, Y., Zheng, F., & Jafarpour, B. (2025). A fluid flow-based deep learning (FFDL) architecture for subsurface flow systems with application to geologic CO_2 storage. *Water Resources Research*, *61*, e2024WR037953. https://doi.org/10.1029/2024WR037953

Received 11 MAY 2024
Accepted 13 DEC 2024

Abstract Prediction of the spatial-temporal dynamics of the fluid flow in complex subsurface systems, such as geologic CO_2 storage, is typically performed using advanced numerical simulation methods that solve the underlying governing physical equations. However, numerical simulation is computationally demanding and can limit the implementation of standard field management workflows, such as model calibration and optimization. Standard deep learning models, such as RUNET, have recently been proposed to alleviate the computational burden of physics-based simulation models. Despite their powerful learning capabilities and computational appeal, deep learning models have important limitations, including lack of interpretability, extensive data needs, weak extrapolation capacity, and physical inconsistency that can affect their adoption in practical applications. We develop a Fluid Flow-based Deep Learning (FFDL) architecture for spatial-temporal prediction of important state variables in subsurface flow systems. The new architecture consists of a physics-based encoder to construct physically meaningful latent variables, and a residual-based processor to predict the evolution of the state variables. It uses physical operators that serve as nonlinear activation functions and imposes the general structure of the fluid flow equations to facilitate its training with data pertaining to the specific subsurface flow application of interest. A comprehensive investigation of FFDL, based on a field-scale geologic CO_2 storage model, is used to demonstrate the superior performance of FFDL compared to RUNET as a standard deep learning model. The results show that FFDL outperforms RUNET in terms of prediction accuracy, extrapolation power, and training data needs.

Plain Language Summary This paper introduces a Fluid Flow-based Deep Learning (FFDL) model for application to subsurface flow prediction. The model offers a computationally efficient prediction tool that combines the flexibility, learning capacity, and efficiency of deep learning models with the structure of fluid flow equations to achieve better training, prediction, and extrapolation performances compared to standard deep learning models. The model provides a computationally efficient surrogate for time-consuming physics-based numerical simulation models that can be used in complex decision-making workflows that require an extensive number of simulation runs. The model improves upon standard deep learning models by reducing training data needs/computation and increasing model fidelity and reliability. Examples from field-scale geologic CO_2 storage problems are used to demonstrate the performance of FFDL relative to a well-established deep learning model (RUNET). The results suggest that, by combining the strengths of deep learning and physics-based models, FFDL can provide an efficient proxy model to facilitate the implementation of complex workflows in subsurface flow systems, including model calibration, uncertainty quantification, and optimization.

1. Introduction

Carbon capture and storage is an important component in reducing the CO_2 emissions by capturing from point sources and injecting it into deep geologic formations. Although geologic CO_2 storage has significant potential, it is still in the early stages. Concerns about CO_2 migration or leakage into shallow aquifers call for robust monitoring and risk management technologies (Celia et al., 2015; Zheng et al., 2021, 2022). While current monitoring methods can track the movement of the injected CO_2 plume, accurately quantifying its volume and migration path remains a challenge for monitoring and verification purposes (Bui et al., 2018). Moreover, predicting the dynamics of pressure buildup and CO_2 migration is essential for guiding decision-making (Zheng et al., 2021) and for assessing real-time risks throughout the life cycle of a project (Ajayi et al., 2019).

Author Contributions:
Conceptualization: Zhen Qin, Yingxiang Liu, Behnam Jafarpour
Data curation: Zhen Qin, Yingxiang Liu, Fangning Zheng
Formal analysis: Zhen Qin
Funding acquisition: Behnam Jafarpour
Investigation: Zhen Qin, Yingxiang Liu
Methodology: Zhen Qin
Project administration: Behnam Jafarpour
Resources: Behnam Jafarpour
Software: Zhen Qin, Fangning Zheng
Supervision: Behnam Jafarpour
Validation: Zhen Qin, Yingxiang Liu
Visualization: Zhen Qin, Fangning Zheng
Writing – original draft: Zhen Qin, Fangning Zheng
Writing – review & editing: Behnam Jafarpour

Reliable prediction of the CO_2 plume migration often requires spatial and temporal analysis, potentially involving detailed simulation models. The injection of CO_2 into subsurface formations triggers a complex multi-component, multiphase flow system. The complex relationships involving miscibility, capillary pressure, and relative permeability, as well as coupled physics lead to nonlinear coupled systems of partial differential equations (PDEs) that are not trivial to solve (Bandilla et al., 2015). Furthermore, field-scale geologic CO_2 storage (GCS) projects span extensive spatial and temporal scales, including both injection and post-injection periods (Ajayi et al., 2019; X. Jiang, 2011). Therefore, numerical simulation for GCS at the field scale can become computationally prohibitive, particularly for complex tasks such as optimization and uncertainty quantification. Another significant challenge is estimating the time-varying storage capacity of the geologic formations, which is influenced by geologic conditions, injectivity and field development plans (Gorecki et al., 2015). Accurate quantification of uncertainty and potential risks typically involves multiple simulation runs, which can impede the implementation of real-time analysis and risk assessments in GCS projects. Consequently, there is a growing need for innovative approaches that can provide accurate and efficient real-time monitoring and forecasting of CO_2 plume migration and pressure buildup during GCS operations.

In recent years, deep learning (DL)-based approaches have emerged as promising alternatives to traditional numerical simulations for predicting the spatial-temporal evolution of fluid dynamics in the subsurface. Specifically, convolutional neural networks (CNNs) that have demonstrated a strong capability for processing image data have found widespread application in predicting the spatial and temporal evolution of subsurface flow systems. Zhu and Zabaras (2018) proposed a fully convolutional encoder-decoder architecture to approximate the mapping from permeability to pressure and velocity maps for a 2-dimensional steady-state Darcy flow problem. Mo, Zhu, et al. (2019) extended the work in Zhu and Zabaras (2018) to predict the responses from a dynamic multiphase flow problem at different time steps. Y. Wang and Lin (2020) designed a custom architecture tailored for single- and two-phase flow systems, incorporating sparsely connected layers to account for the inherent sparse input-output interaction.

Among the family of CNNs, U-Net (Ronneberger et al., 2015) was originally designed for biomedical image segmentation and has emerged as a mainstream model for prediction tasks in the subsurface domain (Z. Jiang et al., 2021; H. Tang et al., 2021; Wen, Tang, & Benson, 2021; Yan, Harp, Chen, & Pawar, 2022). U-Net excels in capturing complex patterns by retaining multi-scale details of input images through skip connections. This feature of U-Net facilitates the integration of high-resolution details, reduces the search space of model parameters, and mitigates the gradient vanishing issue often encountered in deep architectures. Yan, Harp, Chen, and Pawar (2022) proposed a physics-constrained smoother to enhance the pressure prediction generated by U-Net. Their approach incorporates the time step as an additional input, allowing the U-Net model to predict pressure or saturation at different time steps. This method does not explicitly capture the dynamics of the flow system. Instead, it approximates a static mapping conditioned on the time step. Alternatively, spatial-temporal predictions can also be achieved using auto-regressive architecture (Z. Jiang et al., 2021; Mo, Zabaras, et al., 2019), recurrent architecture (M. Tang et al., 2020) or by jointly predicting all time frames (Wen et al., 2022, 2023; Wen, Hay, & Benson, 2021). M. Tang et al. (2020) introduced a combined model called Recurrent R-U-Net that integrates Residual U-Net (R-U-Net) (Wen, Tang, & Benson, 2021) with convolutional long short-term memory (ConvLSTM) network (Shi et al., 2015) to capture the evolution of saturation and pressure in the 2D two-phase waterflooding problem. This combined architecture was later extended to address 3D flow problems in waterflooding (M. Tang et al., 2021) and CO_2 sequestration (M. Tang et al., 2022). Different from the recurrent architecture, Wen et al. (2023) proposed a nested framework of DL models for 4D spatial-temporal prediction of pressure and saturation in a joint way by utilizing Fourier Neural Operator (FNO) (Li et al., 2020). Their methodology involves dividing a large reservoir into multiple scales (levels), with each level employing an FNO model to predict either pressure or saturation. By combining the FNOs of each level sequentially, the states of the entire reservoir can be predicted. Unlike the recurrent nature of ConvLSTM, the FNO approach predicts all time steps simultaneously, enabling parallelized prediction across the time domain. However, the FNO's 4D architecture results in a significantly larger number of trainable parameters, with approximately 80–150 million parameters for each level, making it computationally demanding and potentially resource-intensive when extending it to more complex problems.

Despite the demonstrated effectiveness of DL models, the approaches mentioned above have limitations and are tailored to specific scenarios. Notably, in GCS projects that can span decades for injection and potentially much longer for the post-injection period, the predictive model's capability and flexibility to forecast over arbitrary time

frames are crucial. However, existing DL models are constrained to predicting within fixed periods and encounter challenges in long-term predictions (which require extrapolating power). Deep learning models have limited extrapolation power and do not provide reliable predictions beyond the data distribution defined by the training set (Willard et al., 2022), leading to the out-of-distribution (OOD) generalization problem. As reflected in Mo, Zhu, et al. (2019), the prediction over the OOD time steps becomes an extrapolation task that can challenge deep learning models, as the saturation predictions during the extrapolation may become physically inconsistent. This is particularly important for GCS projects that require long-term prediction horizons. Another important feature is the flexibility of the model in terms of providing predictions over different development scenarios. For instance, the models should be able to capture the response of the storage formations to time-varying controls, which is important in optimizing the performance of the field by dynamically managing the injection strategies based on the in-situ reservoir conditions and monitoring data (Pawar et al., 2015).

To enhance predictive capabilities and overcome limitations, an emerging and active research field focuses on integrating physical principles and domain knowledge into neural networks (NNs) (Faroughi et al., 2023; Willard et al., 2022). A flexible approach is to incorporate governing equations into the loss function to constrain the neural networks during training, known as Physics-informed Neural Networks (PINNs) (Raissi et al., 2019). The PINN framework offers a flexible implementation across various physical systems and has found successful applications in the subsurface domains (Shokouhi et al., 2021; N. Wang et al., 2020; Yan, Harp, Chen, Hoteit, & Pawar, 2022). However, physics-informed approaches often struggle with adhering to boundary conditions, as they implement physical constraints in a "soft" manner. Furthermore, the inherent complexity of multiphase flow in heterogeneous porous media presents challenges in implementing the closed-form residuals of nonlinear PDEs as loss functions (Yan, Harp, Chen, Hoteit, & Pawar, 2022). The training of deep learning models using physics-informed loss functions without labeled data can be affected by the complexity and nonlinearity of the integrated physics. The presence of complex governing equations can lead to highly nonlinear and non-convex loss functions, complicating the training process (Fuks & Tchelepi, 2020). These issues may challenge the use of data-free approaches in more complex problems, including high-dimensional and highly nonlinear systems (Cuomo et al., 2022; Muther et al., 2023).

Another fit-for-purpose alternative involves hard-encoding the underlying physics into the architecture of neural networks, endowing the resulting models with an inductive bias tailored to specific physical problems (Faroughi et al., 2023; Karniadakis et al., 2021). In contrast to physics-informed approaches, physics-encoded architectures impose hard constraints. By capturing the underlying physical dependencies among variables, these architectures demonstrate their connections to Ordinary Differential Equations (ODEs) (E, 2017; E et al., 2017; Lu et al., 2018) and PDEs (Long et al., 2018, 2019; Rao et al., 2021; Ruthotto & Haber, 2020). Long et al. (2018, 2019) proposed PDE-Net to learn PDEs from data. In their work, the CNN is combined with Residual Network (ResNet) (He et al., 2016) to approximate the evolution of PDE with the forward Euler as the temporal discretization. Rao et al. (2021) introduced the product block to emulate the governing terms in PDEs. Their study utilizes convolutional layers to learn spatial dependencies and employs a recurrent form of ResNet for approximating temporal evolution. The resulting DL model has shown good performance in extrapolation tasks. More recently, Dulny et al. (2022) proposed NeuralPDE to combine NeuralODEs (Chen et al., 2018) with the Method of Lines (MOL) using CNNs to approximate the spatial component in PDEs. They state that CNNs can approximate the MOL, a numerical method of solving time-dependent PDEs by representing them as systems of ODEs through spatial discretization. Nonetheless, these studies tend to simplify the governing equations and approximate the dynamics in an explicit form.

In this study, we propose a novel physics-encoded DL model, named Fluid Flow-based Deep Learning (FFDL), for predicting the spatial-temporal evolution of the pressure and saturation in geologic CO_2 storage. To bridge the gaps in the existing models, the FFDL model is designed to handle time-varying well controls and to provide long-term predictions. The architecture of FFDL primarily comprises a physics-based encoder for constructing physically meaningful latent variables, a residual-based processor for the recurrent prediction of latent variables, a control encoder for constructing a latent representation of sink or source term, and a decoder for approximating the mapping from latent variables to outputs, namely pressure and saturation. The physics-based encoder, along with the control encoder, can construct different governing terms in the PDEs for multiphase flow, including accumulation, advection, and sink/source terms. Similar to the previous works (Dulny et al., 2022; Long et al., 2019; Rao et al., 2021), the convolutional layers are employed to capture the spatial dependencies. However, our model extends beyond these by (a) introducing physics-based operators as the activation functions,

(b) constructing latent representations of the governing terms to better approximate the dynamics, and (c) updating the latent governing terms in a coupled and implicit form. We also present a modified Recurrent R-U-Net, based on the work of M. Tang et al. (2022), as a baseline model due to its similar architecture and capability for extension to time-varying control and arbitrary time steps. The predictive performance of FFDL is investigated using a field-scale model of GCS in a saline aquifer. Our results show that FFDL outperforms the Recurrent R-U-Net on test sets featuring unseen permeability, well controls, and time frames. While the developed deep learning framework is capable of handling various inputs, we investigate its prediction and extrapolation performance in specific scenarios by considering variations in permeability and injection rates. We also evaluate the prediction performance of FFDL both within and beyond the training data range, that is, interpolation and extrapolation tasks, respectively.

The remaining sections of this paper are organized as follows. Section 2 presents the problem statement for the spatial-temporal prediction task and introduces the architectures of the proposed physics-based encoder and residual-based processor. In Section 3, we provide a brief overview of the experimental setups and describe the modifications made to Recurrent R-U-Net. The experimental results and discussions are presented in Section 4. Finally, Section 6 offers conclusions on this work and highlights the advantages of the proposed model.

2. Methodology

2.1. Problem Statement

This study addresses a spatial-temporal prediction task governed by a set of coupled PDEs within a 3D reservoir composed of grid blocks $\mathcal{D} \subset \mathbb{R}^3$. The prediction task can be formulated as $\mathbf{X} = \mathcal{F}(\boldsymbol{x}^t, \boldsymbol{m}, \mathbf{U})$, given the inputs $\boldsymbol{x}^t$ as the dynamic states at time step t, $\boldsymbol{m}$ as parameters, and $\mathbf{U}$ as future controls. The output sequence $\mathbf{X} = \{\boldsymbol{x}^{t+1}, \boldsymbol{x}^{t+2}, \dots\}$ represents the dynamic variables over the subsequent time steps, and the operator $\mathcal{F}$ denotes the mapping from inputs to outputs. For each coordinate $\boldsymbol{\omega} \in \mathcal{D}$, the dynamic variable $\boldsymbol{x}^t(\boldsymbol{\omega})$ at time step t consists of pressure $\boldsymbol{p}^t(\boldsymbol{\omega}) \in \mathbb{R}$ and saturation of non-wetting phase $\boldsymbol{S}_n^t(\boldsymbol{\omega}) \in \mathbb{R}$. The input parameters $\boldsymbol{m}(\boldsymbol{\omega}) \in \mathbb{R}^{d_m}$ characterize the model parameters, such as permeability, porosity, grid volume, and elevation, at point $\boldsymbol{\omega}$. The control sequence $\mathbf{U} = \{\boldsymbol{u}^{t+1}, \boldsymbol{u}^{t+2}, \dots\}$ denotes well controls over future time steps. At any given step t, the control variable $\boldsymbol{u}^t$ can be the bottom-hole pressure or flow rate. The value of each grid cell $\boldsymbol{u}^t(\boldsymbol{\omega}) \in \mathbb{R}^{d_u}$ corresponds to the control value (e.g., injection rate) at a well location $\boldsymbol{\omega}$ and is zero if the cell does not contain a well. Superscripts d_x, d_m, and d_u indicate the dimensions of dynamic states, input parameters, and control variables at a point $\boldsymbol{\omega}$, respectively. The goal of this work is to approximate the operator $\mathcal{F}$ with the proposed deep learning model $\mathcal{F}_\theta$, where θ represents the trainable parameters. In this paper, variables denoted in bold represent high-dimensional tensors, while those not in bold refer to scalars. For brevity, we primarily focus on the dimensionality of high-dimensional tensors at individual spatial points $\boldsymbol{\omega}$, rather than the entire spatial domain.

In this approach, we avoid the direct learning of the pressure and saturation dynamics. Instead, we convert these variables into physically meaningful latent variables and learn the dynamics in latent space. The latent representation $\boldsymbol{z}^t = \{\boldsymbol{z}^t_{\text{acc}}, \boldsymbol{z}^t_{\text{adv}}, \boldsymbol{z}^t_{\text{src}}\}$ consists of three components representing the accumulation, advection, and sink/source terms of the governing PDEs in the latent space, respectively. As depicted in Figure 1, the proposed deep learning model mainly consists of the encoder, processor, control encoder, and decoder modules. Initially, the encoder provides an initial estimation of the first two latent variables for the next time step, $\tilde{\boldsymbol{z}}^{t+1}_{\text{acc}}$ and $\tilde{\boldsymbol{z}}^{t+1}_{\text{adv}}$, using $\boldsymbol{x}^t$ and $\boldsymbol{m}$ as inputs. Concurrently, the control encoder provides an initial estimation of the latent variables $\tilde{\boldsymbol{z}}^{t+1}_{\text{src}}$ given the control variables $\boldsymbol{u}^{t+1}$. The processor receives these three initial estimations as input and generates the updated latent variables $\boldsymbol{z}^{t+1}_{\text{acc}}$ and $\boldsymbol{z}^{t+1}_{\text{adv}}$ as outputs. These outputs are then sent to (a) the decoder for the prediction of the dynamic variables $\boldsymbol{x}^{t+1}$, and to (b) the processor itself for the estimation of the new latent variables $\tilde{\boldsymbol{z}}^{t+2}_{\text{acc}}$ and $\tilde{\boldsymbol{z}}^{t+2}_{\text{adv}}$. The decoder takes the updated latent variables $\boldsymbol{z}^{t+1}_{\text{acc}}$ and $\boldsymbol{z}^{t+1}_{\text{adv}}$, as well as the static variable $\boldsymbol{m}$, as inputs. By including $\boldsymbol{m}$, the decoder is designed to decouple the effects of parameters from the latent variables, functioning as the inverse of the encoder. In a recurrent manner, the processor utilizes the previous latent variables as initial estimations and updates them for the next time step by integrating $\boldsymbol{z}^{t+1}_{\text{src}}$ as external inputs. To articulate the model's function, the process $\mathbf{X} = \mathcal{F}(\boldsymbol{x}^t, \boldsymbol{m}, \mathbf{U})$ can be modularized as follows:

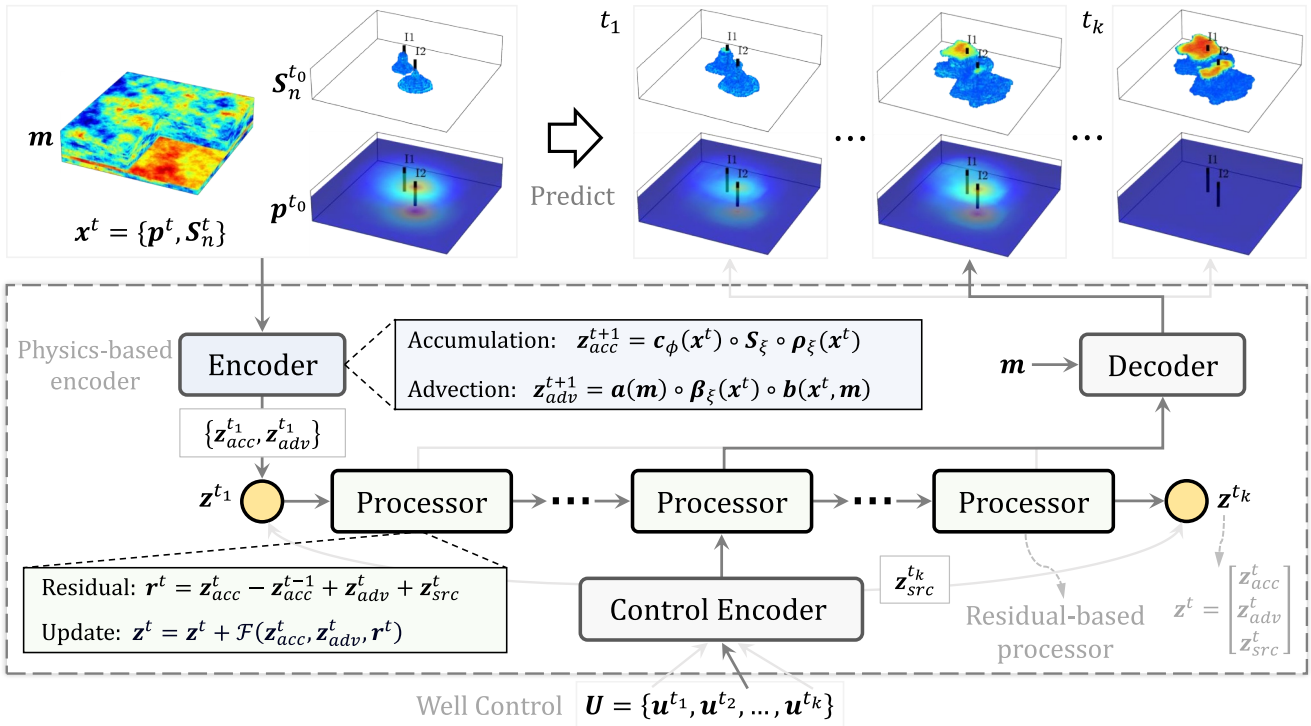

Figure 1. Overview of the proposed Fluid Flow-based Deep Learning (FFDL) model for predicting spatial-temporal pressure and saturation in geologic CO_2 storage. The process is applied to the entire reservoir with four modules: the physics-based encoder and control encoder to project input into physically meaningful latent variables; the residual-based processor to evolve the latent variables given well controls; and the decoder to project latent variables back to pressure and saturation.

$$
\begin{aligned}
\{\tilde{z}_{\text{acc}}^{t+1}, \tilde{z}_{\text{adv}}^{t+1}\} &= \mathcal{F}_{\theta_{\text{InputToLatent}}}(x^t, m), \\
\{\tilde{z}_{\text{src}}^{t+n}\} &= \{\mathcal{F}_{\theta_{\text{ControlToLatent}}}(u^{t+n})\}, n = 1, 2, \ldots, \\
\{z_{\text{acc}}^{t+n}, z_{\text{adv}}^{t+n}\} &= \{\mathcal{F}_{\theta_{\text{LatentToLatent}}}(\tilde{z}_{\text{acc}}^{t+1}, \tilde{z}_{\text{adv}}^{t+1}, \tilde{z}_{\text{src}}^{t+1})\}, n = 1, 2, \ldots, \\
\{x^{t+n}\} &= \{\mathcal{F}_{\theta_{\text{LatentToOutput}}}(z_{\text{acc}}^{t+n}, z_{\text{adv}}^{t+n}, m)\}, n = 1, 2, \ldots,
\end{aligned}
\tag{1}
$$

where, the operators $\mathcal{F}_{\theta_{\text{InputToLatent}}}$, $\mathcal{F}_{\theta_{\text{ControlToLatent}}}$, $\mathcal{F}_{\theta_{\text{LatentToLatent}}}$, and $\mathcal{F}_{\theta_{\text{LatentToOutput}}}$ represent the encoder, control encoder, processor, and decoder modules, respectively.

2.2. Physics-Encoded Deep Learning Model

One of the main contributions of this work is the design of the encoder and processor modules, as depicted in Figure 2. In contrast, the control encoder and decoder are implemented using standard convolutional layers with simpler configurations. This subsection delineates (a) the physical operators used in the encoder, and (b) the detailed architectures of the encoder and processor. A comprehensive description of the governing equations for geologic CO_2 storage is provided in Appendix A. The explanation of the control encoder and decoder is provided in Appendix B.

In our model, the encoder is designed to capture the relationship between the latent representations of the governing terms (i.e., accumulation and advection) and various input variables (e.g., pressure, saturation, permeability, etc.) using physical operators. Recent works have designed neural operators to either (a) construct the latent variables in the spectral domain (Li et al., 2020; Wen et al., 2022; Xiong et al., 2023) or (b) introduce high-frequency features to the DL model (H. Wu et al., 2023). Although these advanced architectures effectively address the issue of spectral bias (Rahaman et al., 2019; Tancik et al., 2020), the constructed latent variables lack interpretability, resulting in a black-box architecture. In contrast, we adopt the concept of physical operators to map the input variables to physically meaningful latent variables that represent accumulation and advection

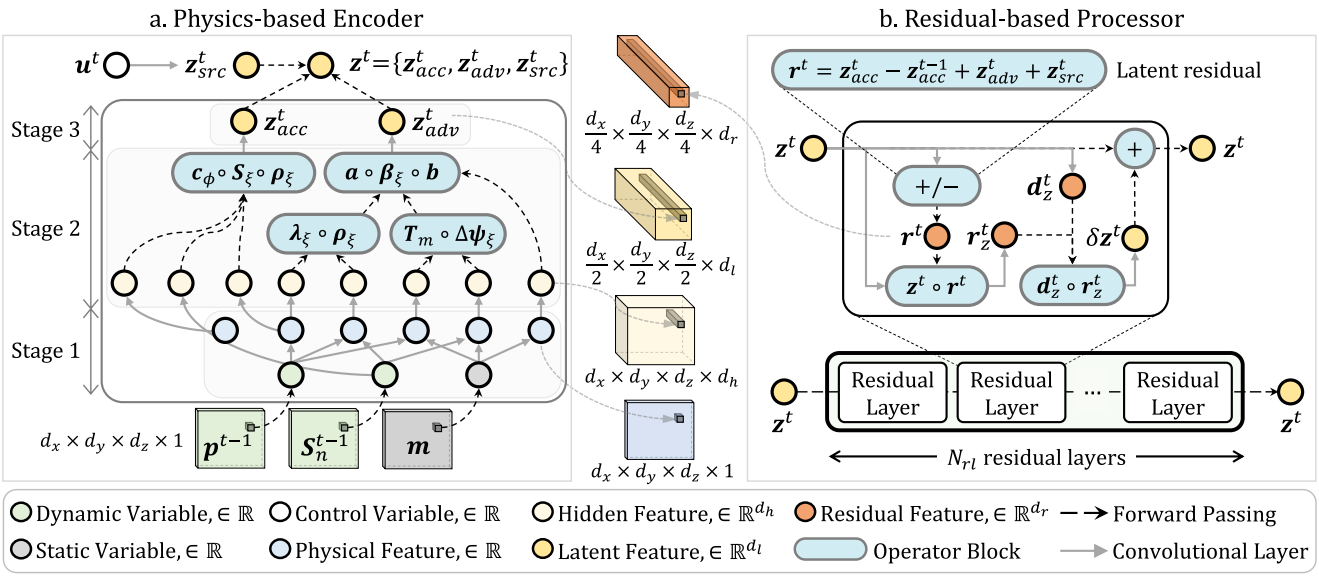

Figure 2. Overview of the physics-based encoder (left) and the residual-based processor (right). Each circle represents a variable corresponding to a single voxel in the 3D representation of the reservoir. The encoder consists of three stages: (1) mapping the inputs (dynamic and static) to physical features, (2) mapping previous features to hidden features, and (3) calculating physical latent variables. For each time step, the processor utilizes a series of residual layers to iteratively update the latent variables. The operator block incorporates non-parametric operators, including element-wise product, addition, and subtraction. The forward passing operation involves directly passing the variable to the next layer without any additional transformations or modifications.

terms. This design is inspired by the Operator-Based Linearization approach (Voskov, 2017), which represents the governing terms as combinations of operators and linearizes them within the parameter space (pressure, saturation, and component) to facilitate the application of numerical solvers.

Two primary motivations underpin the transition from input variables to physics-based latent representations. First, the governing equations of the system involve implicit and nonlinear dependencies on pressure and saturation, challenging the derivation of explicit update forms. Previous studies have utilized ResNet (or skip connections) for direct updates of variables in explicit form. These methods often rely on simplifications of PDEs (J. Nagoor Kani & Elsheikh, 2019; Y. Wang & Lin, 2020) or are tailored to specific problem characteristics (Rao et al., 2021). In contrast to direct updates of dynamic variables, our method focuses on the dynamics within the latent space by mapping them onto physics-based latent variables. This design circumvents the assumptions required by previous works, offering a more thorough treatment of the underlying physics. Second, predicting the dynamics within the latent space offers our model flexibility and generalizability. The physics-based encoder can flexibly adapt to various physical terms, such as capillary and gravity terms, through the design of physics-based operators. Moreover, the physical operators enable our model to effectively generalize across diverse input variables and learn the physical relationship between diverse inputs and outputs. Consequently, this design potentially extends the applicability of our model to various tasks and scenarios.

2.2.1. Physical Operators

For geologic CO_2 sequestration sites, such as saline aquifers, the system consists of two primary phases: a water-rich phase and a CO_2-rich phase. In this work, we simplify the CO_2-brine system to an immiscible two-fluid-phase system with no internal component gradient. As a result, the mass balance equation can be written in terms of phase-based balance equations:

$$\frac{\partial}{\partial t}\left(\phi S_\xi \rho_\xi\right) + \nabla \cdot \left(\rho_\xi \mathbf{v}_\xi\right) = \rho_\xi \tilde{q}_\xi, \quad \xi \in \{w, n\}, \tag{2}$$

where ϕ is the porosity; S_ξ and ρ_ξ are the saturation and density of phase ξ, respectively; $\mathbf{v}_\xi$ is the volumetric flux vector for phase ξ; $\tilde{q}_\xi$ is external sources or sinks of volumetric rate per unit volume of phase ξ. The phase ξ is n for the non-wetting phase (supercritical CO_2) and w for the wetting phase (brine).

We rewrite the Equation 2 as a combination of operators in an algebraic form for the whole reservoir in three-dimensional space $\mathcal{D}$:

$$\boldsymbol{r}_{\xi}(\boldsymbol{x},\boldsymbol{m},\boldsymbol{u}) = \left(z_{\mathrm{acc},\xi}(\boldsymbol{x}) - z_{\mathrm{acc},\xi}\left(\boldsymbol{x}^{t-1}\right)\right) - z_{\mathrm{adv},\xi}(\boldsymbol{x},\boldsymbol{m}) + z_{\mathrm{src},\xi}(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{0}, \tag{3}$$

where $\boldsymbol{r}_{\xi}$ is the residual of the governing equation for phase ξ over the entire reservoir; $z_{\mathrm{acc},\xi}$, $z_{\mathrm{adv},\xi}$, and $z_{\mathrm{src},\xi}$ are the accumulation, advection, and sink/source terms for phase ξ, respectively.

The governing terms are calculated through the physical operators. Detailed derivation of the governing equations and operators are provided in Appendix A. Here, we define the operators as follows

$$z_{\mathrm{acc},\xi}(x) = c_{\phi} \circ S_{\xi} \circ \boldsymbol{\rho}_{\xi} = \left(1 + c_r \circ \left(p - p_{ref}\right)\right) \circ S_{\xi} \circ \boldsymbol{\rho}_{\xi}, \tag{4}$$

$$z_{\mathrm{adv},\xi}(x,m) = a(m) \circ \sum_{l} \boldsymbol{\beta}_{\xi,l}(x) \circ b_l(x,m), \tag{5}$$

$$z_{\mathrm{src},\xi}(x,u) = a(m) \circ \boldsymbol{\rho}_{\xi} \circ q_{\xi} = a(m) \circ \boldsymbol{\rho}_{\xi} \circ V \circ \tilde{q}_{\xi}, \tag{6}$$

$$a(m) = \Delta t \; I_{PV}, \tag{7}$$

$$\boldsymbol{\beta}_{\xi,l}(x) = \boldsymbol{\lambda}_{\xi} \circ \boldsymbol{\rho}_{\xi}, \tag{8}$$

$$b_l(x,m) = T_m^l \circ \Delta\boldsymbol{\psi}_{\xi}^l = T_m^l \circ \left(\boldsymbol{\psi}_{\xi}^v - \boldsymbol{\psi}_{\xi}^u\right), \tag{9}$$

where the symbol $\circ$ represents the element-wise product; $\boldsymbol{c}_{\phi}$ is defined as an update multiplier for the initial porosity; $\boldsymbol{c}_r$, $\boldsymbol{p}_{ref}$, and $\boldsymbol{V}$ are rock compressibility, reference pressure, and grid volume, respectively; $\boldsymbol{q}_{\xi}$ is the volumetric flow rate for phase ξ; Δt is a scalar and represents the time interval; $\boldsymbol{I}_{PV}$ is the inverse of the initial pore volume of a grid cell $\phi_0 V$; $\boldsymbol{\lambda}_{\xi}$ is the mobility of phase ξ; $\boldsymbol{T}_m^l$ is the geometric part of the transmissibility of interface l between two grids u and v; $\Delta\boldsymbol{\psi}_{\xi}^l$ is the phase potential difference between the two grid cells u and v, of which the phase potentials are $\boldsymbol{\psi}_{\xi}^u$ and $\boldsymbol{\psi}_{\xi}^v$, respectively. For the design of the encoder, the phase potential is simplified by neglecting the capillary pressure and represented as $\boldsymbol{\psi}_{\xi} = \boldsymbol{p} + \boldsymbol{\rho}_{\xi} \circ \boldsymbol{D}$. The variable $\boldsymbol{D} = g\boldsymbol{d}$ refers to the gravity term defined as the product of the gravitational acceleration (g) and the elevation (or depth) of grid cells ($\boldsymbol{d}$). The parameter $\boldsymbol{m}$ is defined as a set of three components $\{\boldsymbol{T}_m, \boldsymbol{I}_{PV}, \boldsymbol{D}\}$.

2.2.2. Physics-Based Encoder

The encoder consists of three stages in a sequence (See Figure 2a). The first stage of the encoder takes the dynamic variables $\boldsymbol{x}$ and parameter $\boldsymbol{m}$ as inputs. The output from the first stage is a set of physical features $\boldsymbol{f}$ used in the operators and defined as follows:

$$\boldsymbol{f} = \left\{\boldsymbol{c}_{\phi}, \boldsymbol{\rho}_{\xi}, \boldsymbol{\lambda}, \Delta\boldsymbol{p}, \Delta\left(\boldsymbol{\rho}_{\xi} \circ \boldsymbol{D}\right)\right\}. \tag{10}$$

The dependencies of $\boldsymbol{c}_{\phi}$ and $\boldsymbol{\rho}_{\xi}$ on pressure and the dependency of $\boldsymbol{\lambda}_{\xi}$ on pressure and saturation are parameterized using fully-connected (FC) NN, which is named state-related operator (Figure 3a). In this work, we parameterize the state-related operator using a two-layer fully connected NN with the hidden unit d_o. The Gaussian Error Linear Unit (Hendrycks & Gimpel, 2023) serves as the nonlinear activation function after each layer. Features $\Delta\boldsymbol{p}$ and $\Delta\left(\boldsymbol{\rho}_{\xi} \circ \boldsymbol{D}\right)$ are the differential pressure and gravity terms used in the potential difference $\Delta\boldsymbol{\psi}$, which are calculated through the spatial-related operator (Figure 3b). In a structured grid system, $\boldsymbol{p}$ and $\boldsymbol{D}$ are first padded with reflection padding around the edges of tensors, which serves as a closed boundary. Given the padded terms, $\Delta\boldsymbol{p}$ and $\Delta\left(\boldsymbol{\rho}_{\xi} \circ \boldsymbol{D}\right)$ are computed along the x-, y-, and z-directions. The differential operator is implemented by applying the discretization stencil $\frac{1}{2} \times [-1, 0, 1]$ along each of the three directions.

In the second stage, convolutional layers are used to project all the components of dynamic variable $\boldsymbol{x}$, input parameters $\boldsymbol{m}$, and physical feature $\boldsymbol{f}$ into a high-dimensional hidden feature $\boldsymbol{h}$. Each component of the hidden feature $\boldsymbol{h}$ has a dimensionality of d_h at the point $\boldsymbol{\omega}$. Hidden features are then passed to the operator blocks defined

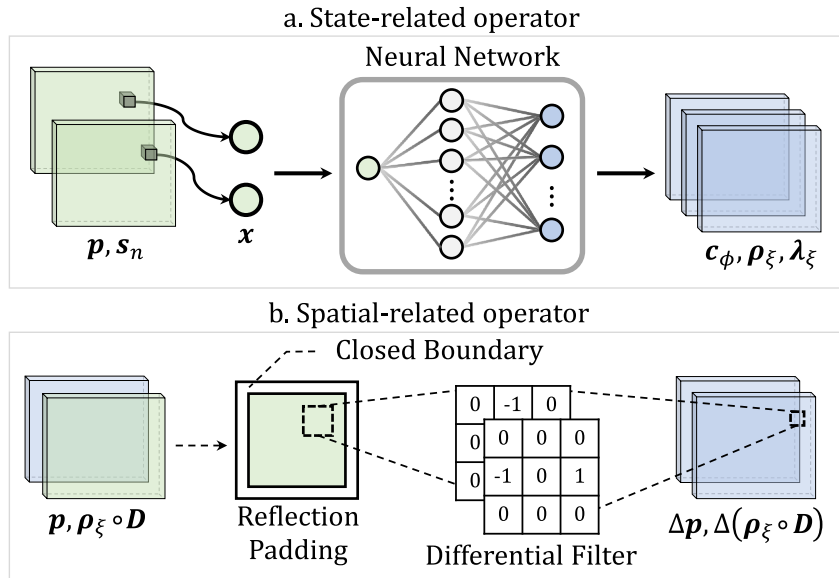

Figure 3. Illustration of (a) state-related operator and (b) spatial-related operator in the first stage of the physics-based encoder. The state-related operator is parameterized by a fully connected neural network. The differential pressure (or fluid potential) is calculated under a closed boundary, which is approximated by reflection padding.

in Equations 4–9. The operator blocks return the high-dimensional representations of the accumulation and advection terms. The goal of extending the feature dimensionality before the operator blocks is to project the physical relationships in a high-dimensional space. Instead of applying nonlinear functions, the element-wise product is used to introduce nonlinearity in the encoder.

In the third stage, we further map the accumulation and advection terms into a high-dimensional latent space with the dimension of d_l ($d_l > d_h$). Simultaneously, the latent variables are downsampled through the convolutional layers, reducing the spatial dimension by a factor of 2. The outputs of this stage are the latent variables z^t_{acc} and z^t_{adv}. The downsampling is employed to increase the receptive field of the convolutional layers while reducing the GPU memory demand.

While the notation of latent variables does not specify the fluid phase, the latent variables are computed separately for the wetting and non-wetting phases and then concatenated. As a result, each latent variable sent to the next layer encompasses the latent representations for the two phases. As shown in Figure 2a, the dynamic input $\boldsymbol{x}$ is defined for the previous time step $t-1$, while the latent variable $\boldsymbol{z}$ corresponds to the current time step t. Therefore, the purpose of the encoder is to generate an initial estimate of the latent variables for time step t. These latent variables will be further updated by the processor, taking into account the external effect of the control variable. The encoder $\mathcal{F}_{\theta_{\text{InputToLatent}}}$ can then be decomposed into three stages as follows:

$$\begin{aligned} \boldsymbol{f} &= \mathcal{F}_{\theta_{\text{InputToFeature}}}(\boldsymbol{x}, \boldsymbol{m}), \\ \boldsymbol{h} &= \mathcal{F}_{\theta_{\text{FeatureToHidden}}}(\boldsymbol{f}, \boldsymbol{x}, \boldsymbol{m}), \\ \{\boldsymbol{z}_{\text{acc}}, \boldsymbol{z}_{\text{adv}}\} &= \mathcal{F}_{\theta_{\text{HiddenToLatent}}}(\boldsymbol{h}), \end{aligned} \tag{11}$$

where $\mathcal{F}_{\theta_{\text{InputToFeature}}}$, $\mathcal{F}_{\theta_{\text{FeatureToHidden}}}$, and $\mathcal{F}_{\theta_{\text{HiddenToLatent}}}$ refer to the three stages of the physics-based encoder, respectively. The dimensions d_o, d_h, and d_l in these three stages are the hyperparameters defined by users. The selection of these hyperparameters and their feasible ranges are provided in Appendix B.

Neural networks parameterize the state-related operator in the first stage with d_o hidden units and learn the nonlinear dependencies with complicated formulas, such as Equation of State (EoS). In contrast to the parameterization of the state-related operator, the spatial and physical operator blocks are non-parametric and serve as hard constraints to enforce adherence to the underlying physics.

2.2.3. Residual-Based Processor

The processor is implemented as a variant of recurrent neural network with a customized recurrent unit, as illustrated in Figure 2b. The inputs to the processor are the governing terms in the latent space z^t to represent the accumulation z^t_{acc}, advection z^t_{adv}, and sink/source z^t_{src} for wetting and non-wetting phases. The latent representation of the sink/source term comes from the control encoder $\mathcal{F}_{\theta_{\text{ControlToLatent}}}$. Each customized recurrent unit, also referred to as a processor block, consists of N_{rl} residual layers that iteratively update the latent variables $z^t_{\text{acc}}, z^t_{\text{adv}}$, and z^t_{src}, where N_{rl}, denotes the number of residual layers. To make the proposed model more memory-efficient and to increase the receptive field of the convolutional kernel, we further reduce the spatial dimension of the latent variables $z^t_{\text{acc}}, z^t_{\text{adv}}$, and $z^t_{\text{src}} \in \mathbb{R}^{D\times d_l}$ by a factor of 2 and simultaneously project them onto the feature space with a higher dimensionality of d_r ($d_r > d_l$). The resulting latent features are used to calculate the residual term δz^t, which is then projected back to the original dimension $D \times d_l$ and added to the latent variable z^t.

The residual layer updates the latent variables based on the concept of the residual for governing equations. For a numerical solver, the Newton-Raphson method is commonly applied, which is written as:

$$\boldsymbol{J}(\boldsymbol{x}^k)\left(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\right) = -\boldsymbol{r}(\boldsymbol{x}^k), \tag{12}$$

where $\boldsymbol{J}$ is the Jacobian matrix at the inner iteration step k of the nonlinear solver; $\boldsymbol{r}(\boldsymbol{x}^k)$ denotes the residual of the governing equations for iteration k.

In this study, the design of the processor is inspired by the Newton-Raphson method. Instead of iteratively updating the dynamic variable $\boldsymbol{x}^k$, we update the latent variables and parameterize the update procedure using neural networks. First, we rewrite the Equation 12 as follows:

$$\begin{aligned}\delta\boldsymbol{x}^k &= \boldsymbol{x}^{k+1}-\boldsymbol{x}^k = -\left(\boldsymbol{J}^k\right)^{-1}\boldsymbol{r}^k \\ &\approx \mathcal{F}\left(\boldsymbol{r}^k,\boldsymbol{x}^k\right) \approx \mathcal{F}_{\theta_{\text{ResidualToDiff}}}\left(\boldsymbol{r}^k,z^k\right).\end{aligned} \tag{13}$$

The dependency of input difference $\delta\boldsymbol{x}^k$ on terms $\boldsymbol{x}^k$ and $\boldsymbol{r}^k$ is parameterized by the operator $\mathcal{F}_{\theta_{\text{ResidualToDiff}}}$. To further facilitate the update, we introduce another operator $\mathcal{F}_{\theta_{\text{DiffToDiff}}}$ to convert the difference $\delta\boldsymbol{x}^k$ to the latent space, which is shown as follows:

$$\delta z^k = z^{k+1}-z^k \approx \mathcal{F}_{\theta_{\text{DiffToDiff}}}\left(\delta\boldsymbol{x}^k\right), \tag{14}$$

where $\mathcal{F}_{\theta_{\text{DiffToDiff}}}$ approximates the mapping between the input difference $\delta\boldsymbol{x}^k$ and the latent difference δz^k. Finally, δz^k can be written as:

$$\delta z^k = \mathcal{F}_{\theta_{\text{DiffToDiff}}}\left(\mathcal{F}_{\theta_{\text{ResidualToDiff}}}\left(\boldsymbol{r}^k,z^k\right)\right). \tag{15}$$

For brevity, we neglect the superscript k in the following derivation of the processor unit. For each residual layer within the processor unit, we first calculate the residual term $\boldsymbol{r}$ as follows:

$$\boldsymbol{r}^t = z^t_{\text{acc}} - z^{t-1}_{\text{acc}} + z^t_{\text{adv}} + z^t_{\text{src}}. \tag{16}$$

Here, the negative sign is integrated into the advection term without the loss of correctness. The above equation is similar to Equation 3, with the distinction that each term in Equation 3 corresponds to a specific phase. The latent variables in the above equation encompass representations for both phases. In the first residual layer of each processor block, the terms z^t_{acc} and z^{t-1}_{acc} in Equation 16 are identical. In other words, the accumulation term z^{t-1}_{acc} sent from the previous processor will serve as both the previous state z^{t-1}_{acc} and the current state z^t_{acc}. Consequently, the previous state z^{t-1}_{acc} remains fixed, while the current state z^t_{acc} is updated in subsequent residual layers.

After calculating the latent residual $\boldsymbol{r}^t$, each of the latent variables (i.e., $z^t_{\text{acc}}, z^t_{\text{adv}}$ and z^t_{src}) is multiplied by $\boldsymbol{r}^t$, and their products are sent to the operator $\mathcal{F}_{\theta_{\text{ResidualToDiff}}}$ as inputs, expressed as follows:

$$\boldsymbol{r}_z^t = \mathcal{F}_{\theta_{\text{ResidualToDiff}}}\left(\begin{bmatrix} \boldsymbol{z}_{\text{acc}}^t \\ \boldsymbol{z}_{\text{adv}}^t \\ \boldsymbol{z}_{\text{src}}^t \end{bmatrix} \circ \begin{bmatrix} \boldsymbol{r}^t \\ \boldsymbol{r}^t \\ \boldsymbol{r}^t \end{bmatrix}\right). \tag{17}$$

In this step, the output $\boldsymbol{r}_z^t$ refers to the term $-\left(\boldsymbol{J}^k\right)^{-1}\boldsymbol{r}^k$ in the latent space. Then, the latent residual term $\delta \boldsymbol{z}^t$ can be derived as follows:

$$\boldsymbol{d}_z^t = \mathcal{F}_{\theta_{\text{LatentToDiff}}}\left(\boldsymbol{z}^t\right), \tag{18}$$

$$\delta \boldsymbol{z}^t = \mathcal{F}_{\theta_{\text{DiffToDiff}}}\left(\boldsymbol{d}_z^t \circ \boldsymbol{r}_z^t\right), \tag{19}$$

where the operator $\mathcal{F}_{\theta_{\text{LatentToDiff}}}$ is to transform the latent variable into a new term to introduce nonlinearity. Then, the operator $\mathcal{F}_{\theta_{\text{DiffToDiff}}}$ takes the product $\boldsymbol{d}_z^t \circ \boldsymbol{r}_z^t$ as input and returns the latent residual term $\delta \boldsymbol{z}^t$. Therefore, in each residual layer, the latent variable z^t is updated as

$$z^t = z^t + \delta z^t. \tag{20}$$

In this study, the design of the residual layer is heuristic, as the updating procedure is performed in a high-dimensional latent space and is parameterized by neural networks. The high-dimensional representation enables the latent variable to capture comprehensive information about the input variables. The processor unit in this study is referred to as a residual-based processor, as it is composed of residual layers and draws inspiration from the concept of residuals in the governing equations.

3. Experimental Setup

In this study, we conduct a comprehensive evaluation of the proposed physics-encoded DL model through three distinct experiments. First, we assess the predictive capability of the model using unseen permeability inputs in Section 4.1 and compare our model with the Recurrent R-U-Net over the public data set proposed by M. Tang et al. (2021). Second, we investigate the model's performance in the presence of unseen pairs of control variables and permeability inputs (Section 4.2). This assessment requires the model to generalize and handle complex and diverse scenarios. Lastly, we explore the model's extrapolation capability by applying it to the prediction over the post-injection period while the training set only covers the injection period (Section 4.3). This extrapolation task poses a significant challenge due to the distinctive dynamics involved. In addition to these three experiments, we investigate the effect of the physics-based encoder on the latent representations of the governing terms by replacing the physics-based encoder with a traditional convolutional encoder (Section 4.4).

In addition to the comparison based on publicly available data sets (M. Tang et al., 2021), we also applied our model to a set of simulated data sets generated by a synthetic 3D simulation model of deep saline aquifer (Appendix C). The training data sets include simulation models for both single-well and two-well scenarios, introducing sufficient diversity in the data set. However, we test our model on the two-well scenario only, as it is more complex than the one-well scenario. The simulation model consists of CO_2 injection over 15 years, followed by a post-injection period of 85 years. For the third experiment, we specifically utilize the initial 15 years of the post-injection as our test set. This selection is based on the observation that the pressure and plume dynamics reach a quasi-steady state beyond this time frame and do not show observable changes over time. The simulated data set consists of 30 steps, where each step represents 1 year. During the injection, the well controls are randomly perturbed for each year while being constrained to have a total injection amount over 15 years. For detailed information regarding the simulation model and data generation, please refer to Appendix C.

In the examples with the public data set, the models are trained to predict the whole simulation time with 10 steps by taking the initial state and permeability as inputs, following the setup in M. Tang et al. (2021). In the other experiments conducted in this study, however, the training process involves feeding the models with dynamic states from any time step and predicting only the subsequent 8 years rather than the entire 15 years. Exposing the model to dynamic inputs of different time frames can enable models to prevent overfitting and accurately learn the

dynamics. Consequently, each training sample, comprising 16 time frames (15 years plus an initial state), is divided into eight data samples, each spanning 8 time frames. During testing, the DL models predict the entire 15 years over injection (or 30 years over both injection and post-injection periods) without taking any ground truth dynamic state as input.

In the experiment on the public data set, we propose a modified version of Recurrent R-U-Net by making three key modifications to the architecture of the original model proposed by M. Tang et al. (2021). These modifications were aimed at improving the model's performance and aligning it with the setup of our model to ensure fair and meaningful comparisons. First, we replaced Batch Normalization (BatchNorm) (Ioffe & Szegedy, 2015) in the original model with Group Normalization (GroupNorm) (Y. Wu & He, 2018). This change was motivated by the observation that GroupNorm offers better generalizability and transferability compared to BatchNorm (Kolesnikov et al., 2020). Second, we introduced a dummy control input to the modified Recurrent R-U-Net and our model for the public data set. The control variable in the public data set is fixed over time and not included as input in the original Recurrent R-U-Net. This change is to make the use of modified Recurrent R-U-Net and our model consistent throughout this work. The dummy control variable for each time step is a 3D tensor, with values of 0 for grid blocks with no well, 1 for producers, and -1 for injectors. Lastly, we redefined the static variable $\boldsymbol{m}$ by (a) adding $\boldsymbol{I}_{PV}$ and $\boldsymbol{D}$ as part of the input, and (b) replacing permeability $\boldsymbol{K}$ with the term $\boldsymbol{T}_m$.

The pressure and injection rate are normalized using Min-Max normalization. Saturation is not normalized since it typically ranges between 0 and 1. To normalize the inverse grid volume $\boldsymbol{I}_{PV}$, we scale it by multiplying a reference grid volume of $50 \times 50 \times 2.5$ m^3. Using a uniform layer thickness, the depth $\boldsymbol{D}$ varies from 0 to 19, representing the depths of grid blocks from the first to the 20th layer. The term $\boldsymbol{T}_m$ contains zero values to represent the closed boundary and exhibits a right-skewed distribution with a mean significantly higher than the median. Consequently, we apply the inverse hyperbolic sine transformation to transform $\boldsymbol{T}_m$ before feeding it into a deep learning model.

The training in this study employs the Adam optimizer. In our proposed model and the modified Recurrent R-U-Net, we apply the relative ℓ_2-loss (Wen et al., 2022) as the loss function, which is defined as follows:

$$L\left(\{\boldsymbol{S},\boldsymbol{p}\}_{t_1:t_2},\{\hat{\boldsymbol{S}},\hat{\boldsymbol{p}}\}_{t_1:t_2}\right)=\frac{\|\boldsymbol{S}_{t_1:t_2}-\hat{\boldsymbol{S}}_{t_1:t_2}\|_2}{\|\boldsymbol{S}_{t_1:t_2}\|_2}+\frac{\|\boldsymbol{p}_{t_1:t_2}-\hat{\boldsymbol{p}}_{t_1:t_2}\|_2}{\|\boldsymbol{p}_{t_1:t_2}\|_2}, \tag{21}$$

where $\{\hat{\boldsymbol{S}},\hat{\boldsymbol{p}}\}_{t_1:t_2}$ represents the predicted saturation and pressure over time steps from t_1 to t_2. In this work, the variance of the norms $\|\boldsymbol{S}_{t_1:t_2}\|_2$ or $\|\boldsymbol{p}_{t_1:t_2}\|_2$ are considerably high due to the dynamic range of time frames, which can vary widely. Using the mean squared error (MSE) loss function encourages inaccurate predictions at the start of injection. This is mainly because the saturation of the non-wetting phase in the initial time frames is inconsequential. Deep learning models that fail to predict the saturation of these time frames will still generate low values of MSE loss. The choice of relative error as the loss function will mitigate this issue. The details of the training process and configuration are provided in Appendix B.

4. Results

4.1. Testing on Unseen Permeability

In this experiment, we first compare three models on the public data set (M. Tang et al., 2021): the original Recurrent R-U-Net (the baseline model), the modified Recurrent R-U-Net, and our proposed model. The data set consists of 2,923 samples spanning 1,000 days and is generated by a numerical simulation model of a two-phase oil reservoir for a waterflooding problem. In this experiment, models are trained to predict oil saturation using different amounts of training samples: 200, 500, 1,000, 1,500, 2,000, and 2,500. A separate set of 50 samples is used for model validation, while the remaining 373 samples form the test set.

Figure 4a shows that our model consistently achieves lower test losses across all sizes of training samples compared to the other two models. Notably, our model trained on 1,000 samples demonstrates a comparable root mean square error (RMSE) of saturation prediction to the baseline model trained on 2,500 samples. Figure 4b demonstrates the distribution of saturation RMSE for each time step, where the baseline model exhibits higher mean and variance in test losses across time steps compared to our model. Both modified Recurrent R-U-Net and our model accurately capture the saturation front for the 200 and 2,500 training sample cases with lower RMSE

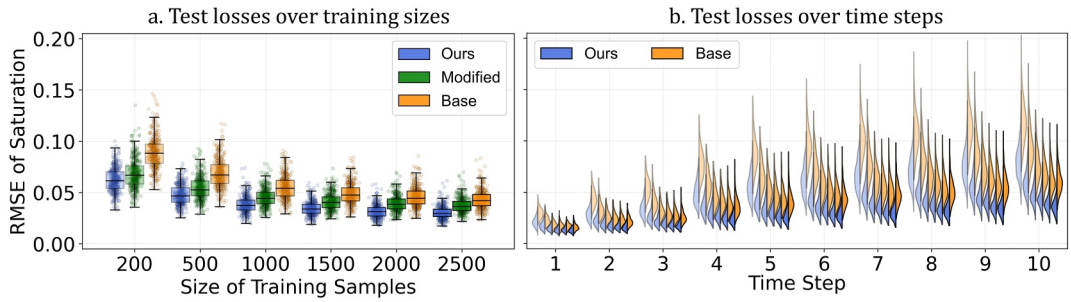

Figure 4. Distributions of test losses of saturation for different models. (a) Test losses for different sizes of training samples. (b) Test losses of saturation for different time steps. The different transparencies in panel (a) represent the various sizes of training samples ranging from 200 to 2,500 from left to right. Each time step in panel (b) contains the six bars to represent distributions of root mean square error values for the cases of training samples ranging from 200 to 2,500 from left to right. "Modified" refers to the modified Recurrent R-U-Net. "Base" refers to the baseline R-U-Net. "Ours" refers to our model.

values. In the subsequent experiments, we will continue using the modified Recurrent R-U-Net for comparison and exclude the original version. The original Recurrent R-U-Net has a different setup of input and does not support time-varying control variables.

Next, we proceed to evaluate our model in the simulation model of subsurface CO_2 storage reservoir. Our proposed model and modified Recurrent R-U-Net are trained using 700 simulated samples to predict pressure and CO_2 saturation over a 15-year injection. The 700 simulated samples are then divided into 5,600 training data samples, each of which spans eight years. The validation and test sets consist of 100 and 200 simulated samples, respectively. The training and test sets have the same injection controls but different permeability. For brevity, we will refer to the modified Recurrent R-U-Net as RUNET in the following experiments. As shown in Figure 5 (left), our proposed model consistently exhibits lower mean and variance of RMSE compared to RUNET. It is worth noting that the saturation error exhibits increasing trends over time due to the spreading of the saturation front. In contrast, the pressure error is more stationary, attributed to the presence of an aquifer region surrounding the storage reservoir, which facilitates pressure dissipation and maintains it within a certain range.

Figures 6 and 7 provide 3D visualizations of saturation and pressure distributions, respectively. The errors in both saturation and pressure tend to occur in areas where there are significant changes (or gradients). Specifically, for

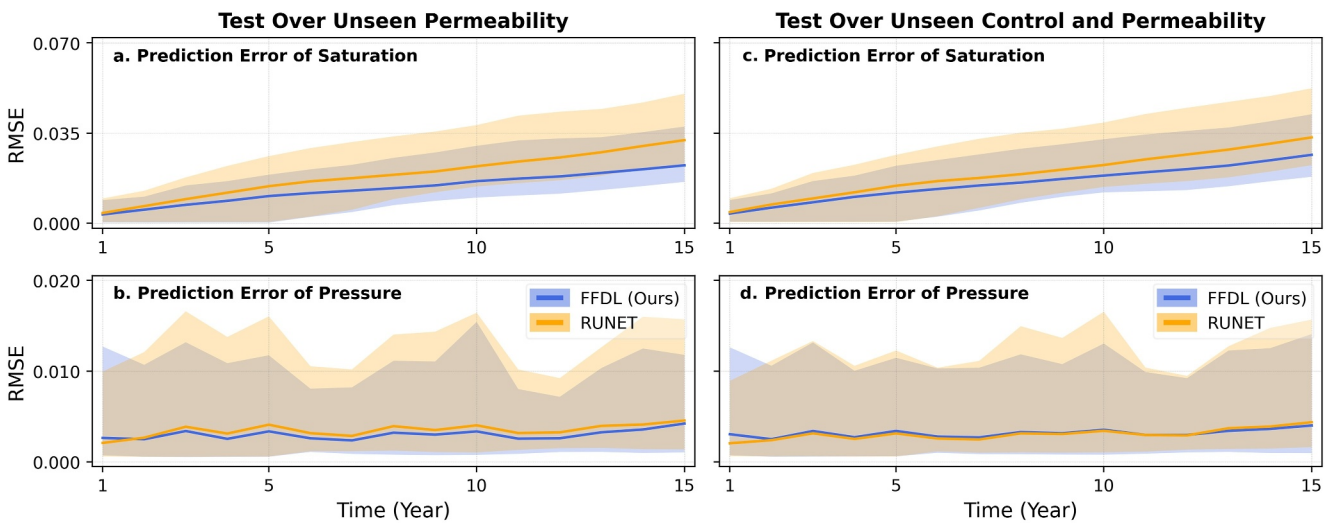

Figure 5. Prediction errors of normalized saturation and normalized pressure over the test set for two models: our proposed model (Ours) and modified Recurrent R-U-Net (RUNET). Left: Test errors with unseen permeability; Right: Test errors with unseen control and permeability. The error band and solid line represent a 95% confidence interval and the median of root mean square error, respectively. Subfigures (a) and (b) show the normalized saturation and pressure errors, respectively, for test cases with unseen permeability maps. Subfigures (c) and (d) present the normalized saturation and pressure errors for test cases with both unseen controls and permeability maps.

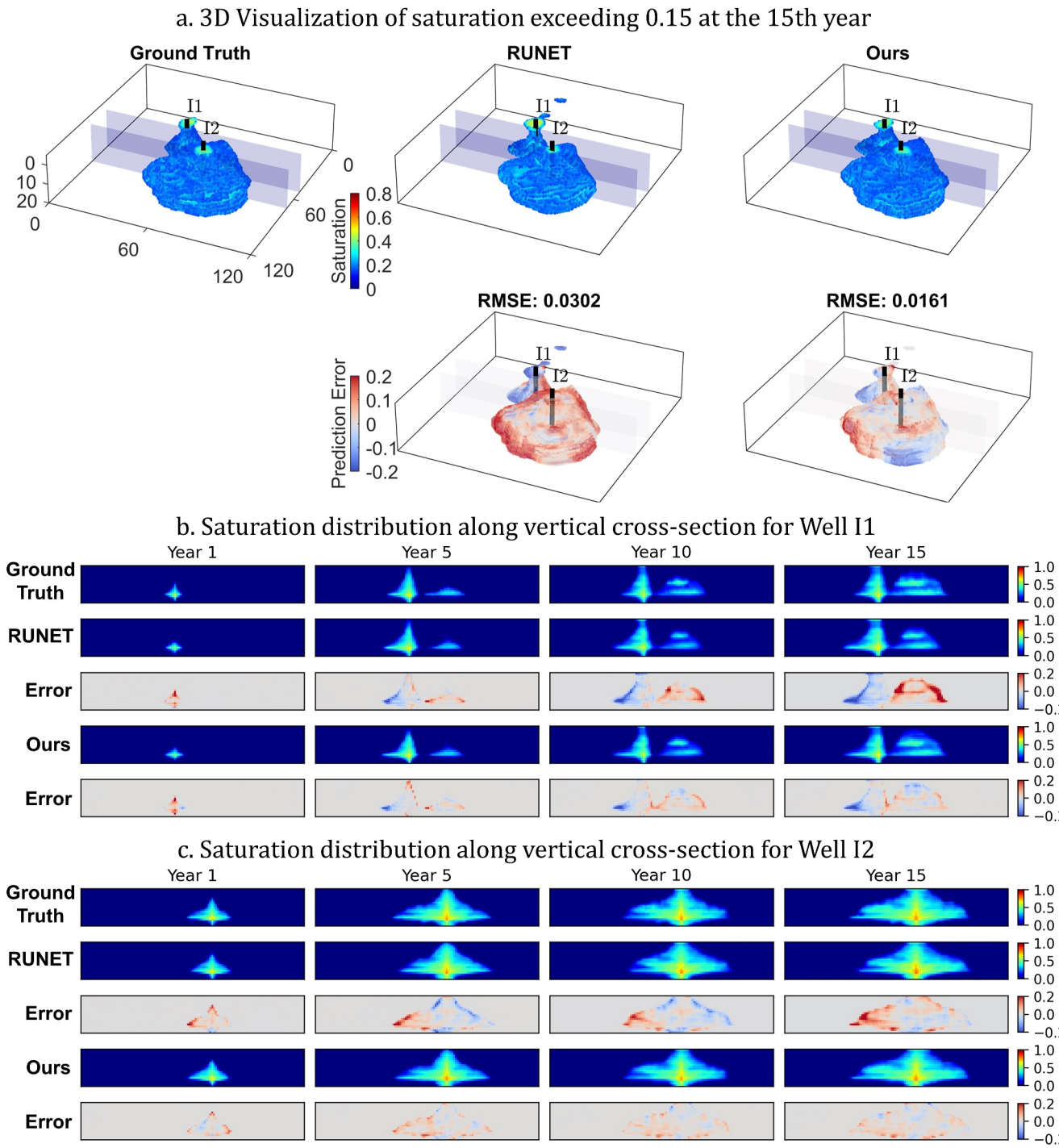

Figure 6. Visualization of saturation prediction over unseen permeability. The 3D visualization in panel (a) displays the saturation distribution in the 15th year. In panels (b, c), the 2D saturation distributions represent the *x*-*z* planes intersecting Wells I1 and I2, respectively. These planes are depicted as transparent cross-sections in panel (a).

saturation, the errors are prominent near the front of the CO_2 plume, where the saturation values experience rapid transitions. On the other hand, for pressure, the errors are more noticeable in the vicinity of the wells. The 3D pressure map changes significantly over time due to variations in injection control. As reflected in Figures 6 and 7, our model outperforms the RUNET in capturing these two distinct patterns.

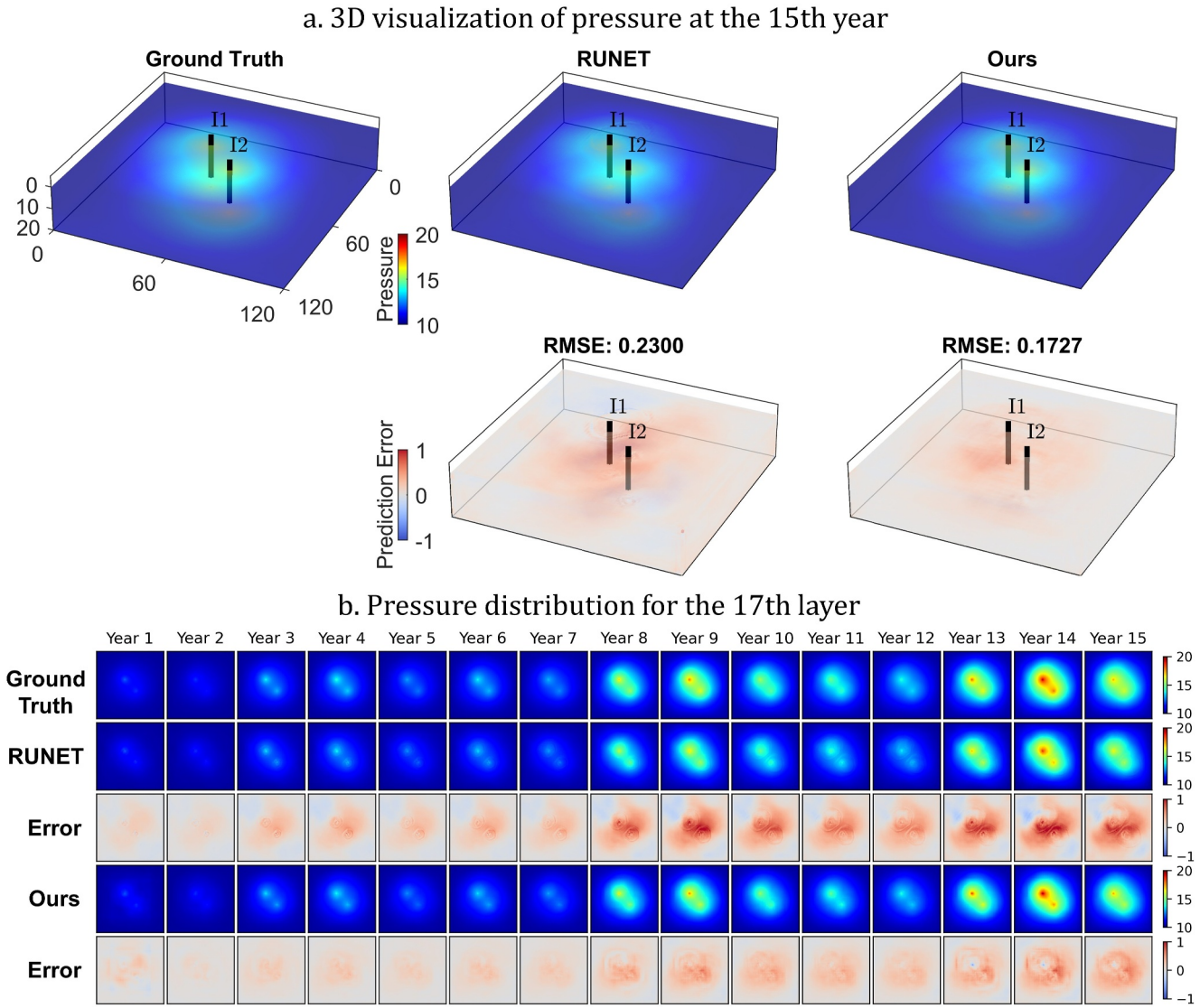

Figure 7. Visualization of pressure prediction over unseen permeability. In panel (b), the 2D pressure distributions represent the 17th layer of the reservoir, the layer where injection wells are located. The pressure values are measured in MPa, and the root mean square error is also reported in MPa.

4.2. Testing on Unseen Control and Permeability

In this experiment, we further evaluate the performances of two models (our model and RUNET) on the test set where both control and permeability are unseen during training. The training, validation, and test sets still consist of 700, 100, and 200 simulated samples, respectively. As depicted in Figure 8, the distribution of injection rates differs between the training and test sets, even though both data sets have the same constraint of total injection rate. Notably, the allocation of injection rates in the test set exhibits more significant variations compared to the training set, demonstrating a less balanced allocation of injection rates. The allocation of the injection rate for the test set falls outside the distribution covered by the training set. For control optimization problems, the control variable can evolve toward its extremity during the process of optimization (Qin et al., 2023). The out-of-distribution injection rate resembles the challenges caused by the extremity of the control variable in the optimization task and challenges the model's robustness in handling diverse control scenarios (Qin et al., 2024).

Figure 5 (right) illustrates that the prediction performances of the two models are comparable in predicting pore pressure, while our model is more accurate in predicting saturation than RUNET. Given the variations in control variables in the test set, predicting saturation and pressure becomes more challenging for both models compared

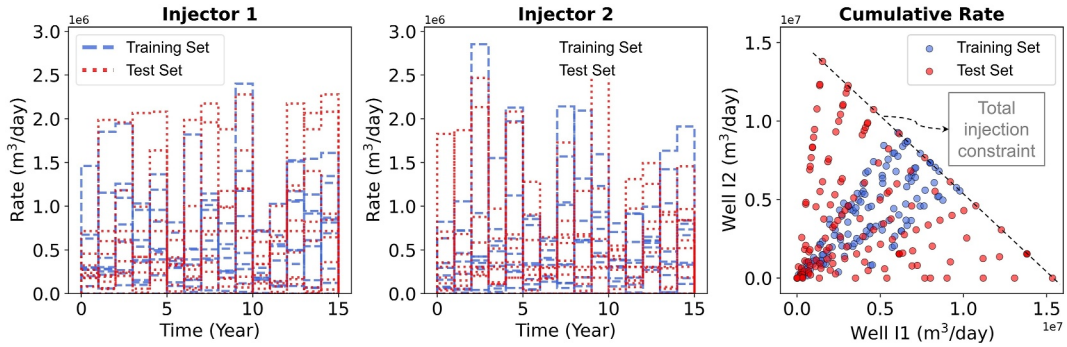

Figure 8. Visualization of injection control in the training and test sets. The left and middle figures depict the time-series injection rates for wells I1 and I2, respectively. The right figure shows the QQ-plot of cumulative injection rates for the two wells, illustrating the differences in the distribution between the two data sets.

to the previous experiment. Figures 9 and 10 provide visualizations of saturation and pore pressure predictions for both models. In Figure 9, it can be observed that the saturation prediction from RUNET shows significant errors around the top layer, leading to physically inconsistent results. In contrast, our model exhibits only slight discrepancies near the saturation front in the top layer, indicating its improved accuracy and ability to maintain consistency with the underlying physics. As shown in Figure 10, the pressure prediction from RUNET exhibits errors that are evenly distributed around the wells, while the prediction error from our model is more concentrated. This distinct behavior between the two models may be attributed to differences in their architectures and will be further investigated in our future work.

4.3. Extrapolation Over Post-Injection

In the previous experiments, the deep learning models were trained and tested over a 15-year injection. During testing, DL models predicted the future states over the same period based on different sets of inputs. In this experiment, we extend the application of these trained models to predict 30 years, including both injection and post-injection. Specifically, models trained in previous experiments (Sections 4.1 and 4.2) are directly applied to predict the last 15 years, from the 16th to the 30th year, with the initial state to be predicted saturation and pressure in the 15th year. During post-injection, injection rates are set to zero.

In Figure 11, the prediction performances of two models are evaluated for both injection and post-injection periods. It can be observed that our model consistently exhibits significantly lower prediction errors for saturation than RUNET. Furthermore, our model's prediction errors remain relatively stable during post-injection, whereas errors from RUNET keep increasing for the saturation prediction. On the other hand, the pore pressure of the entire reservoir becomes more uniform and converges to the boundary pressure during post-injection. Therefore, pressure prediction becomes less challenging for the two models, as both exhibit decreasing prediction errors after the 15th year when the post-injection starts. Despite the comparable performance of pressure prediction for both models during the injection period (Figure 11d), our model consistently outperforms RUNET in terms of lower mean and variance of prediction errors during post-injection. This indicates the superior robustness and accuracy of our model in handling the distinct dynamics of the post-injection period.

Figure 12 visualizes three examples of the saturation predictions for two models in the 15th and 30th years. It can be observed from all three examples that, during the post-injection period, CO_2 is dominantly driven by the buoyant force and gets accumulated at the top layer, which is overburdened by cap rock. A common failure of RUNET is its physically inconsistent prediction during extrapolation. As shown in the first two examples, the predicted CO_2 plume from RUNET in the 30th year shows disconnected patterns and deviates significantly from the ground truth. In the last two examples, the prediction from RUNET exhibits a relatively low RMSE in the 15th year, which is close to the error of our model. However, RUNET still fails to capture the CO_2 migration during post-injection and even shrinks the CO_2 plume toward the end of the prediction. On the other hand, our model can accurately extrapolate the CO_2 plume during post-injection. Among the three models, our model accurately predicts the saturation front compared with a slight increase in RMSE.

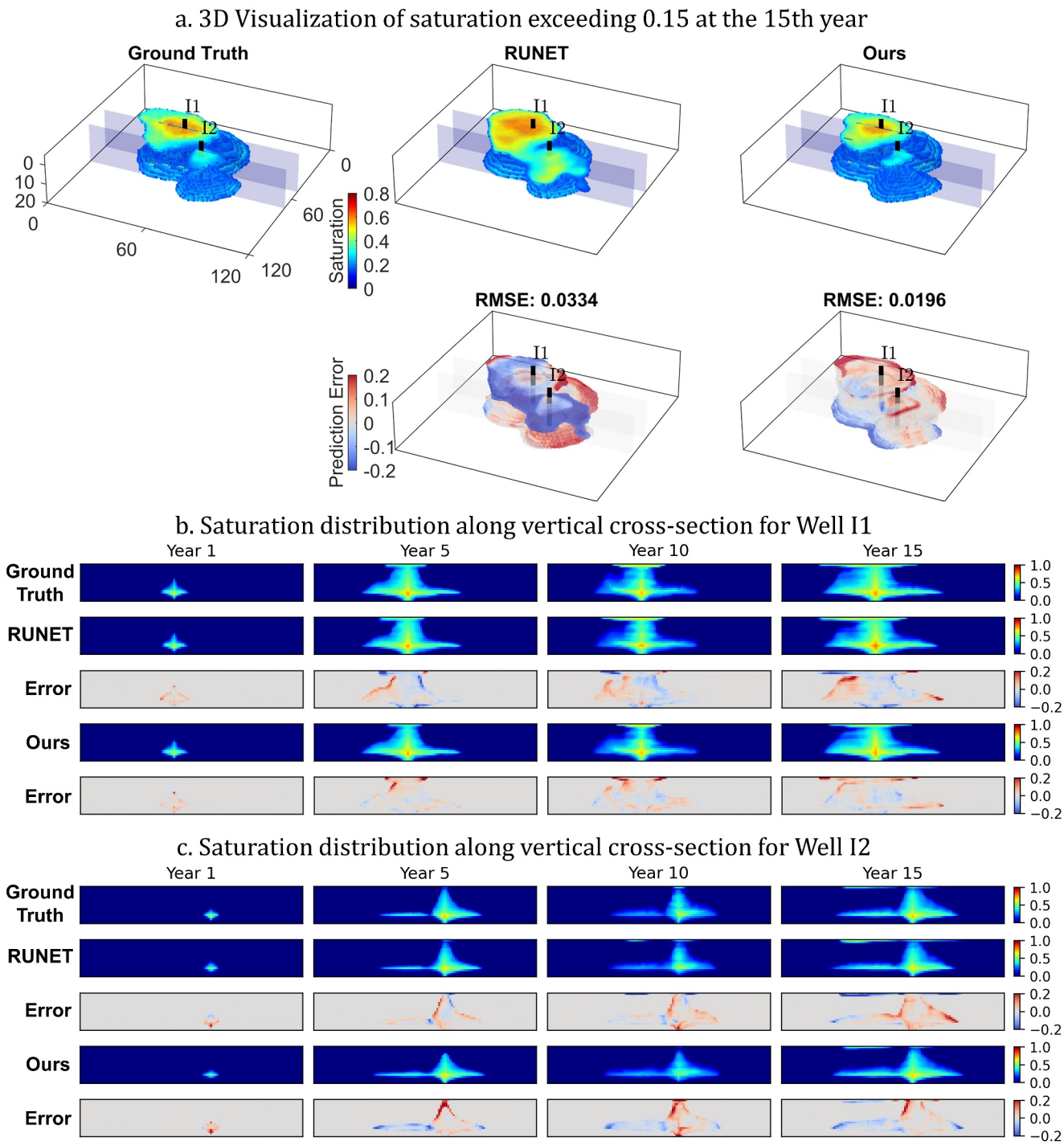

Figure 9. Visualization of saturation prediction over unseen control and permeability. Subfigure (a) compares the 3D saturation predictions from RUNET and our model with the ground truth in the 15th year. Subfigures (b) and (c) compare the 2D saturation predictions from two models along the vertical cross-section passing through injector Well 1 and Well 2, respectively.

Figure 13 visualizes the pressure prediction of the 17th layer of the reservoir for two models during post-injection. It can be observed that the pressure dissipates during post-injection and reaches the aquifer boundary, which serves as a constant boundary condition due to its extensive volume. In contrast to RUNET, our model demonstrates an accurate prediction of the pressure dissipation and eventual convergence to the aquifer pressure.

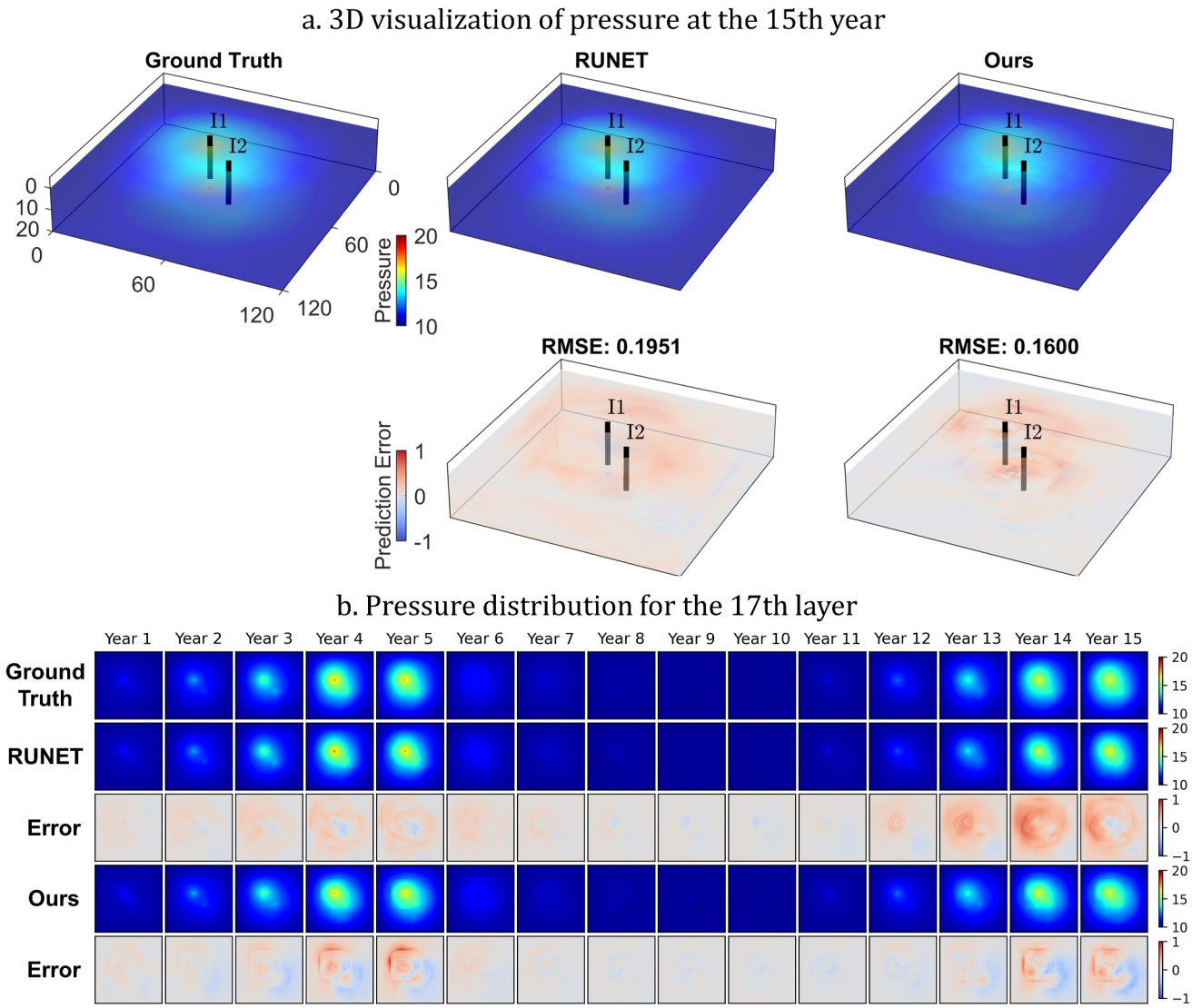

Figure 10. Visualization of pressure prediction over unseen control and permeability. The pressure values are measured in MPa, and the root mean square error is also reported in MPa. Subfigure (a) compares the 3D pressure predictions from RUNET and our model with the ground truth in the 15th year. Subfigure (b) presents the pressure profile of the injection layer from RUNET and our model compared to the ground truth from Year 1 to Year 15.

These findings underscore the superior performance of our model in capturing long-term behavior and accurately predicting the saturation evolution beyond the injection period.

It is important to note that predicting the post-injection period is an extrapolation task, as it falls outside the time range covered by the training set. Extrapolation tasks can pose challenges for deep learning models, as the statistical input-output relationship during extrapolation may deviate from what was learned from the training set. In the case of CO_2 migration, the dominant driving force changes between the injection and post-injection periods, which can result in a different statistical input-output relationship. This change in driving force can undermine the prediction performance of deep learning models. Therefore, accurate prediction during post-injection requires the models to capture the underlying physics of the system, rather than solely relying on the learned statistical relationship from the training set.

4.4. Investigation of Latent Representation

To further investigate the effect of introducing the physics-based encoder on the prediction performance, we introduce an additional DL model. This model shares the identical modules as the proposed physics-encoded DL

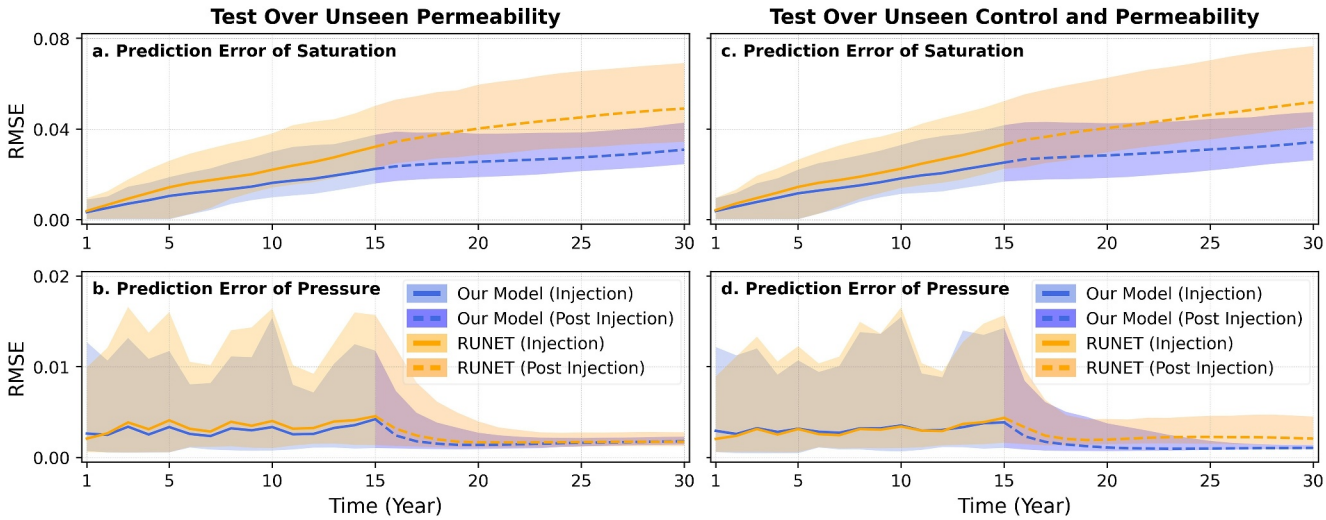

Figure 11. Prediction errors of normalized saturation and normalized pressure over injection and post-injection for the examples of (left) unseen permeability and (right) unseen control and permeability. The first 15 years denote the injection, while the last 15 years denote the post-injection where injection rates are zeros for two wells. Subfigures (a) and (b) show the normalized saturation and pressure errors, respectively, for test cases with unseen permeability. Subfigures (c) and (d) present the normalized saturation and pressure errors for test cases with both unseen control and permeability.

model, with the exception that the physics-based encoder is replaced with convolutional encoders. This alteration facilitates a more focused investigation of the effect. The convolutional encoder comprises two separate encoders to produce the latent representations of accumulation and advection terms, respectively. Each encoder shares a similar architecture as the encoder in the RUNET. In the case of the accumulation term, the input exclusively encompasses the initial states of dynamic variables (i.e., pressure and saturation). As for the advection term, dynamic and static variables are concatenated together and fed into the encoder as input. Subsequently, the latent variables representing the accumulation and advection terms are then sent to the processor blocks, the same as the pathway of the physics-based encoder.

Figure 14 illustrates the comparison between two models: one equipped with the physics-based encoder (referred to as "Ours") and the other with convolutional encoders (referred to as "ConvEnc"). Both models are trained and tested using the same setup as the second experiment (Section 4.2). The results show that both models demonstrate similar RMSE values for pressure and saturation during injection. However, our model consistently outperforms ConvEnc during the post-injection period. On the other hand, as shown in Figure 15, ConvEnc significantly outperforms RUNET in predicting pressure and saturation, encompassing both injection and post-injection periods. Given that all three models possess a similar number of trainable parameters (over 3 million), these results exhibit the superior performances of the physics-based encoder and the residual-based processor in contrast to the convolutional encoder and ConvLSTM, respectively.

Since the governing terms follow certain PDE constraints, it is essential to construct a compact latent space. The latent representations of accumulation and advection terms are expected to exhibit similar patterns with the true governing terms to contain physically consistent features. Therefore, we further investigate the latent representations of two governing terms derived from our model and ConvEnc and discuss their physical meanings. Figure 16 illustrates an example of the latent representations of accumulation (Figure 16a) and advection terms (Figure 16b) for both our model and ConvEnc. The latent representations are the outputs of the last processor layer. To reduce the redundancy in latent representations, we first employ Principal Component Analysis on d_l latent variables ($d_l = 32$) for each time step separately. Here, only the first three principal components (PCs) are visualized for brevity. Moreover, Figures 16a and 16b present the actual values of the wetting-phase accumulation and advection in the vertical direction in the 13th year as references.

As shown in Figure 16a, the true accumulation term is dominantly affected by saturation (Figure 16d), with only slight variations outside the CO_2 plume as affected by pressure (Figure 16c). This slight variation is because the density of liquid-phase water is insensitive to pressure changes and that reservoir porosity is set to be homogeneous in this work. The components from our model exhibit similar patterns with pressure contour and

10.1029/2024WR037953

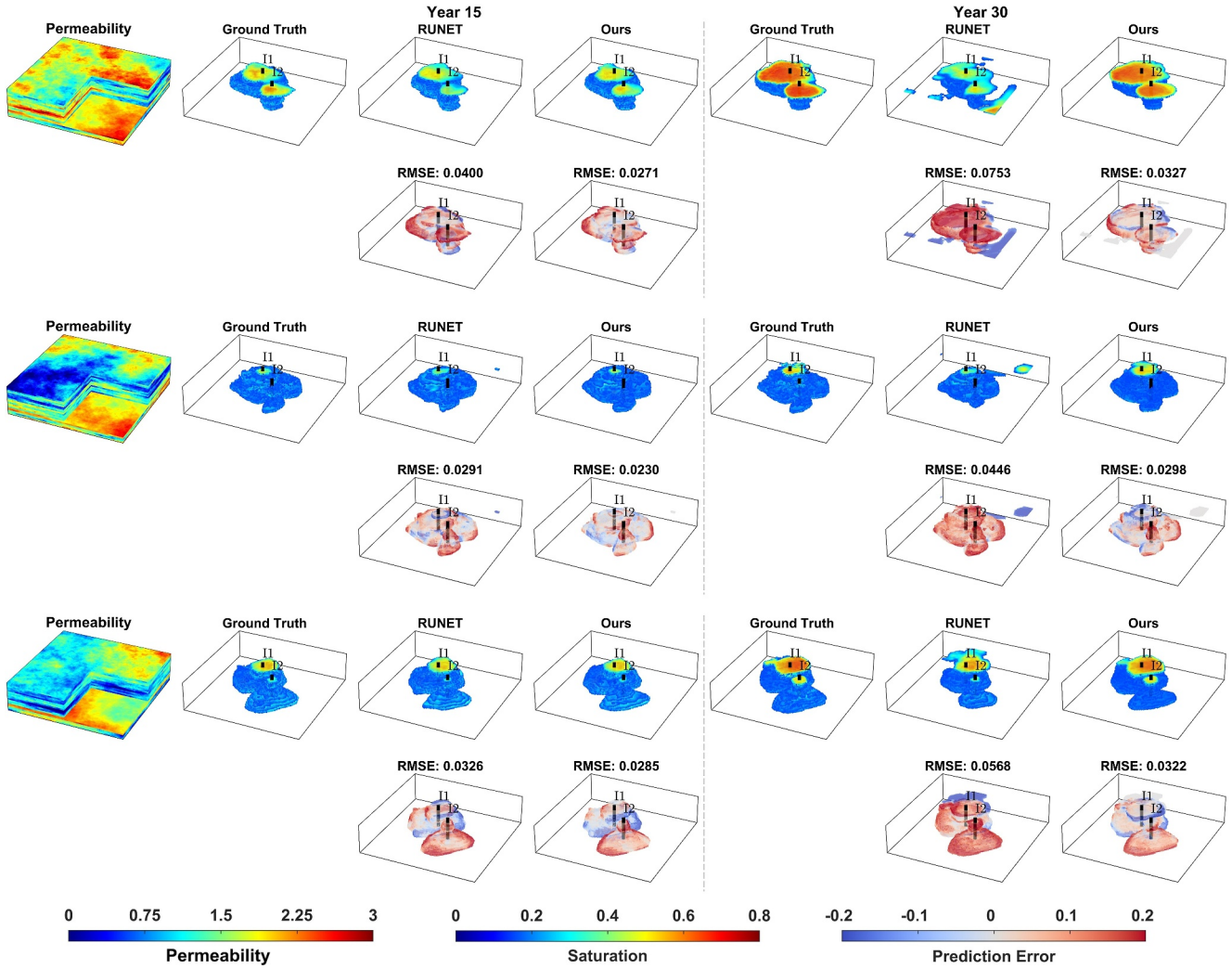

Figure 12. Visualization of saturation prediction of three different cases in the 15th and 30th years. Each case has different permeability and injection control as inputs. The permeability map is in the unit of $\log_{10}(mD)$.

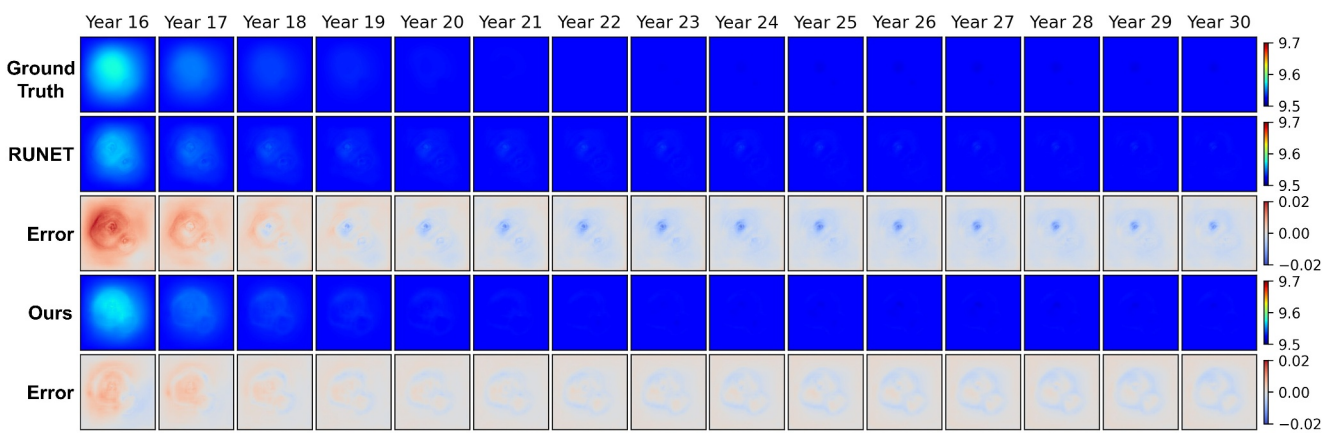

Figure 13. Visualization of pressure prediction over the post-injection period.

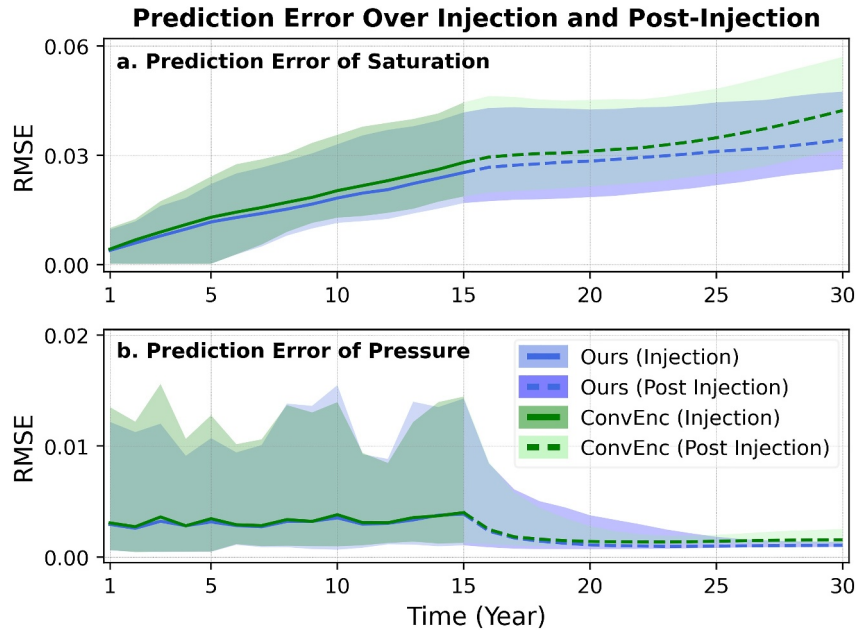

Figure 14. Prediction errors of normalized saturation (a) and normalized pressure (b) over injection and post-injection for two predictive models with different encoders: convolutional and physics-based.

saturation front, as reflected in the first and third components. The components from ConvEnc also reflect these two features, as shown in the first PC. However, as depicted in the second PCs, the latent accumulation terms in ConvEnc exhibit high-frequency patterns similar to permeability. This arises from the updating process within the processor blocks, where the latent accumulation and advection mutually influence each other. These high-frequency patterns are distinct from the accumulation terms depicted in Figure 16a. The residual-based processor model passes the latent accumulation to the next time step to approximate the dynamics of the mass balance governing equation. Latent representations that are not aligned with the true governing terms could hinder the learning process and lead to inefficient learning of the dynamics.

Figure 16b illustrates the PCs of latent representations of the advection term for both models. Notably, the latent advection terms generated by our model closely align with the true values and exhibit high-fidelity patterns. In contrast, ConvEnc produces overly smooth latent advection patterns. This distinction between the two models may stem from the inclusion of the physics-based encoder. Specifically, in our model, the overall receptive field for permeability spans only three grid blocks. This design choice prevents over-smoothness during the calculation of the latent advection term, thus preserving intricate details. On the other hand, ConvEnc's receptive field follows the default setup of RUNET and spans 12 blocks, resulting in an inconsistent inductive bias with the advection term. While ConvEnc can learn high-fidelity patterns from the data, it requires more data to constrain its behavior, hindering efficient learning. Despite our model's superiority over ConvEnc, we do not dive into introducing any

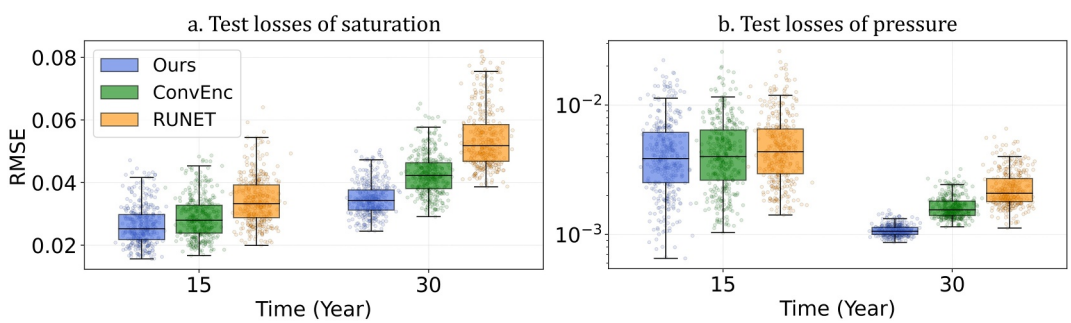

Figure 15. Distributions of test losses of (a) saturation and (b) pressure for different models.

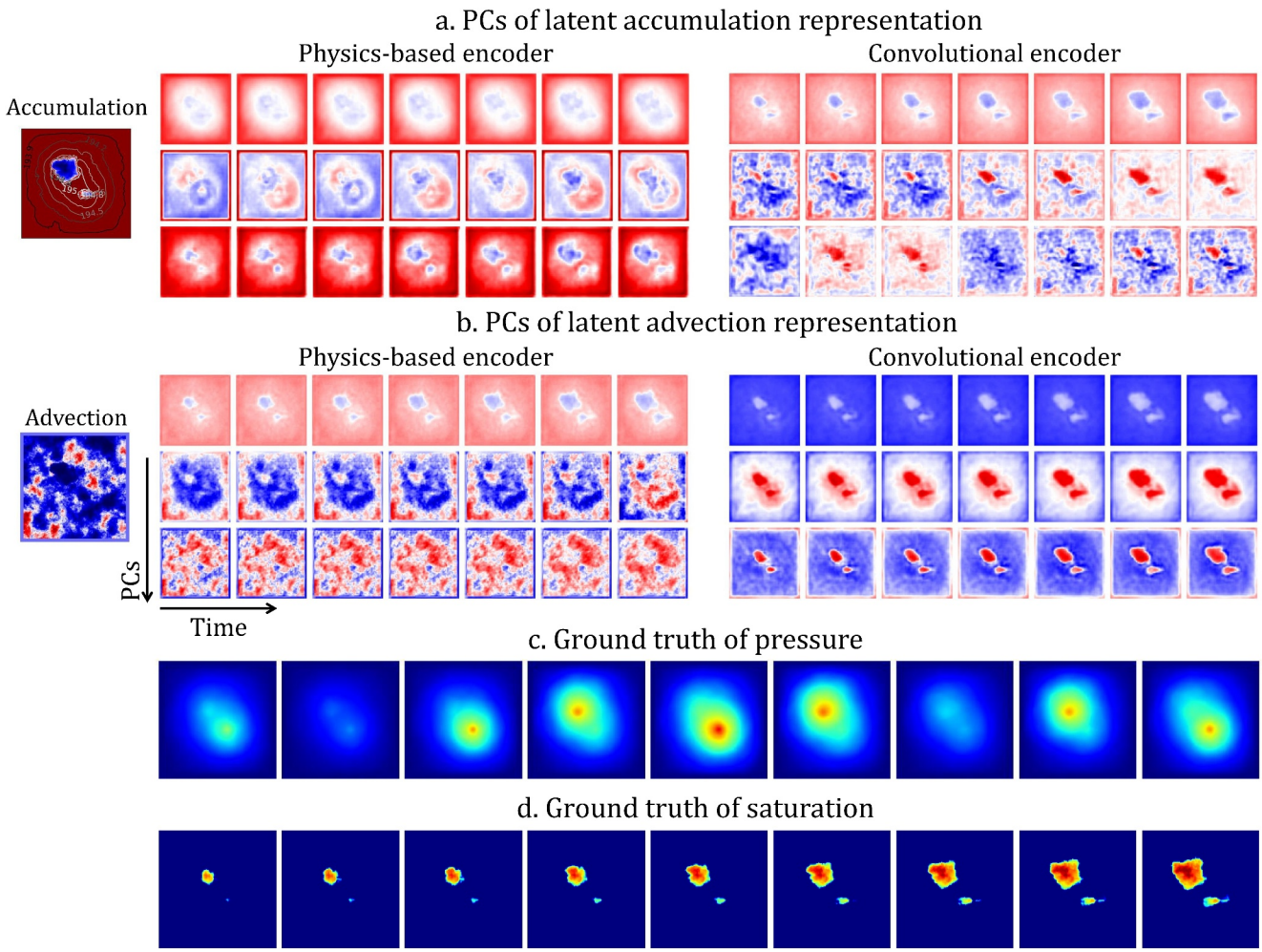

Figure 16. Visualization of the first three principal components of latent representations of (a) accumulation and (b) advection terms from two deep learning models: one with a physics-based encoder (left column) and one with a convolutional encoder (right column). Each row in panels (a, b) represents the evolution of a principal component from the 9th to the 15th year. (c, d) visualize the ground truths of pressure and saturation of the first layer over the last 9 years of injection. The true accumulation and z-direction advection of the wetting phase are provided in panels (a, b) for reference.

metrics to evaluate the physical consistency of latent variables. This aspect extends beyond the scope of this work and could potentially be explored in future endeavors.

5. Discussion

The experimental results validate the effectiveness of our proposed model in addressing the aforementioned limitations faced by traditional deep learning models. Our model exhibits data efficiency, as demonstrated using a public data set (Section 4.1), and consistently outperforms traditional deep learning models when handling various types of inputs (Sections 4.1 and 4.2). The physical operators enable our model to approximate the underlying physics and learn the complex input-output relationship efficiently. Furthermore, our model exhibits superior performance compared to RUNET in the extrapolation to the post-injection period (Section 4.3), where RUNET fails to extrapolate the saturation front and generates physically inconsistent predictions. The recurrent architecture, along with regular time intervals for each recurrent unit, allows our model and RUNET to effectively predict over arbitrary lengths of time steps. Unlike RUNET, which lacks interpretability and functions as a black box, the proposed model employs modular components, each with a specific role in capturing the dynamics of the system.

5.1. Advantages of Physics-Encoded Framework

Integration of Physical Priors with Learning Capacity: The physics-encoded framework is constrained by physical priors and balances interpretability with adaptability. The physical priors are in the form of the structure of the subsurface flow equations, without the detailed complexity of the exact processes involved in specific systems/applications. The architecture resembles the general form of the flow equations that is applicable to a wide range of problems. The training data related to each specific application is then used to learn the flow behavior in the application of interest. For instance, rather than incorporating the full complexity of physics involved in geologic CO_2 storage, which involves multi-component multiphase flow, the model incorporates the structure of two-phase immiscible flow dynamics. The integrated physical relation serves as an inductive bias, tailoring the DL model's structure to flow problems, facilitating its learning of the flow behavior from training data. During training, the parameters of the DL model are tuned to capture the specific dynamics of the systems, which is not explicitly encoded in the prior. It is important to note that the model is not expected to capture the physical processes that are not manifested in the structure of model nor the training data. In the cases where the dynamics used in the training stage changes significantly during the prediction phase, additional considerations (such as retraining with data pertaining to the new physics) may be needed to enhance the prediction power of the model.

Flexibility and Broader Applicability: The FFDL's applicability to different subsurface reservoir settings and time-varying well controls makes it a valuable tool for efficient implementation of decision-making workflows, including inverse modeling, uncertainty quantification, and optimization. The main objective of this paper is to introduce FFDL as a flexible, efficient, and tailored deep learning architecture for subsurface flow problems. While the specific inputs and outputs of the model can be designed according to the objective of each study, the examples in this paper included the variability in permeability heterogeneity and well injection controls. The applicability of FFDL extends beyond geologic CO_2 storage and the specific scenarios studied in this work. Notably, the public data set used in Section 4.1 addresses a two-phase immiscible fluid flow problem.

Computational Efficiency: A primary motivation for developing DL models is to accelerate predictions, as traditional numerical simulators are computationally demanding and often limit their utility in field management workflows. In this study, FFDL is introduced to address some of the limitations of existing DL proxy models by incorporating physical priors. The simulated data needed to train standard DL proxy models can be computationally demanding to generate. The incorporated prior in FFDL reduces the offline training data needs (and computational demand), as shown in our sensitivity analysis. This efficiency makes FFDL a viable option for tasks requiring iterative and/or ensemble evaluations, such as robust optimization and closed-loop control. Once trained, the inference cost of FFDL is extremely low, similar to other DL models. Predictions can be produced in near-real-time, typically within seconds, even for large-scale reservoirs. In contrast, classical numerical simulators may require hours or days for similar tasks. This computational advantage is especially pronounced in workflows involving thousands of simulations, such as ensemble-based inverse modeling or robust optimization, where FFDL can leverage GPU-enabled parallelization to handle multiple realizations simultaneously.

5.2. Data Support and Inductive Bias

While encoding physical mechanisms into the model architecture is expected to improve its extrapolation power, the prediction performance also depends heavily on the information provided by the training data. As highlighted by Wilson and Izmailov (2020), the convergence of a model toward a true solution often hinges on two key factors: sufficient data support and appropriate inductive bias. Data support refers to the diversity and completeness of the training data in capturing the essential dynamics of the system. Without adequate representations of mechanisms such as dissolution or diffusion, the model lacks the necessary information to learn and predict their effects. On the other hand, the inductive bias that is introduced through the architecture embeds prior knowledge about the system's general behavior and guides the model toward physically plausible solutions. These two factors are complementary: data support provides the foundation for learning, while inductive bias ensures efficient and accurate generalization. However, caution must be exercised in using FFDL in extrapolation tasks, especially when the dominant physics in the extrapolation problem differs significantly from the one used during the training. For example, the subsurface conditions and physical processes that are involved during CO_2 injection are typically different from those that are present many years after stopping the injection. These factors must be taken into account in effective design and application of FFDL (or any DL model).

Some of the future research in this area includes improving the inductive bias used in FFDL and enhancing the data support for it. The integration of additional physical mechanisms, such as dissolution and diffusion, can improve the model's predictive accuracy and efficiency. Extending the physics encoded in the FFDL architecture can be used to enhance its learning efficiency and application to more complex tasks. Increasing the diversity and representativeness of the training data sets can also facilitate the inclusion of mechanisms that are not adequately represented in the training data sets. This will help the model generalize better across scenarios with varying input conditions and physical processes.

In addition to these advancements, incorporation of physics-informed loss functions can further improve the fidelity of the model in representing the behavior of the physical models. While PINNs have shown proficiency in learning simpler tasks, they face challenges when applied to complex systems such as geologic CO_2 storage, where the governing PDEs exhibit high nonlinearity and complexity (Krishnapriyan et al., 2021).

Additional extensions and methodological innovations are particularly crucial for the practical deployment of large, data-intensive models in real-world scenarios. The ability to generalize across diverse input conditions and adapt to varying physical processes will significantly enhance the framework's robustness and applicability across a wide range of subsurface engineering problems.

6. Conclusion

In this work, we propose a novel DL architecture that incorporates the general structure of the fluid flow equations, as a form of inductive bias, to enhance its training and prediction performance. The model is used to predict the spatial-temporal behavior of subsurface flow systems. We applied the model to predict the evolution of pressure and saturation during geologic CO_2 storage. Traditional deep learning models, which solely rely on learning statistical input-output relationships, face challenges in extrapolation tasks, require large amounts of training data, and can produce physically inconsistent predictions. To address these limitations, we introduce FFDL as a new deep learning framework that incorporates the general form of the fluid flow physics equations into its encoder and processor units. The new architecture improves the ability of the model to learn the general behavior of the underlying physics. The examples in this paper considered variability in permeability heterogeneity and dynamic well control settings, to make predictions over both the injection and post-injection periods.

The flexibility and applicability of this framework are demonstrated with examples from geologic CO_2 storage, which involves complex physics than is included in the model architecture. However, integration of the incorporated fluid flow structure and training data from the relevant physics resulted in superior predictive capabilities to RUNET as a well-established DL model. The proposed FFDL model is general and can be applied to other subsurface flow applications involving spatial-temporal predictions, such as groundwater flow, geothermal energy, and hydrocarbon reservoirs. While this study focuses on a specific scope, future research can explore ways to adapt this framework to broader applications by incorporating additional physical processes and employing physics-informed loss functions. Such adaptations may help further improve the model's accuracy and data efficiency. In summary, this work presents a novel deep learning framework that combines the benefits of data-driven models (including efficiency, learning capacity, and flexibility) with physics-based inductive bias (including interpretability and extrapolation power) to develop an efficient and proxy model to enable computationally complex subsurface flow tasks. By embedding physical knowledge directly into the architecture, the model offers a promising framework for developing interpretable and data-efficient models for adoption in various subsurface flow applications.

Appendix A: Governing Equations

We adopt the Operator-Based Linearization approach (Voskov, 2017) to derive governing equations and implement the finite-volume discretization for the mass balance equation. For a system with n_c components and n_p phases, the transport equations can be written as follows:

$$\frac{\partial}{\partial t}\left(\phi \sum_{\xi=1}^{n_p} x_{i,\xi}\rho_\xi S_\xi\right) + \nabla \cdot \sum_{\xi=1}^{n_p}\left(x_{i,\xi}\rho_\xi \mathbf{v}_\xi\right) = \sum_{\xi=1}^{n_p} x_{i,\xi}\rho_\xi \tilde{q}_\xi, \quad i = 1, \ldots, n_c, \tag{A1}$$

where ϕ is porosity; $x_{i,\xi}$ is the mass fraction of component i in phase ξ; ρ_ξ is the density of phase ξ; S_ξ is the saturation of phase ξ; $\mathbf{v}_\xi$ is the volumetric flux vector (or Darcy flux) for phase ξ; $\tilde{q}_\xi$ is the external sources or sinks of volumetric rate per unit volume in phase ξ.

For a geologic CO_2 storage site where CO_2 is injected into the deep saline aquifer, the phases $\xi \in \{w,n\}$ could be w for the wetting phase (brine) and n for the non-wetting phase (supercritical CO_2). The multiphase extension of Darcy's equation is applied to describe the flow of each phase and is written as:

$$\mathbf{v}_\xi = -\mathbf{K}\frac{k_{r,\xi}}{\mu_\xi}\left(\nabla p_\xi - \rho_\xi g \nabla d\right) = -\lambda_\xi \mathbf{K}\left(\nabla p_\xi - \rho_\xi g \nabla d\right), \quad \xi \in \{w,n\}, \tag{A2}$$

where $\mathbf{K}$ is the permeability tensor, $k_{r,\xi}$ is the relative permeability, μ_ξ is the phase viscosity, $\lambda_\xi = k_{r,\xi}/\mu_\xi$ is the phase mobility, p_ξ is the phase pressure, g is the gravitational acceleration, and d represents the depth.

The capillary pressure is defined as the difference between the pressure of two phases and is a function of saturation:

$$p_c(s_w) = p_n - p_w, \tag{A3}$$

where p_n and p_w are the pressures of non-wetting (n) and wetting phases (w).

By applying the finite-volume discretization and backward Euler approximation in time, the mass balance equation can be discretized as:

$$\begin{aligned} V\left(\left(\phi\sum_{\xi=1}^{n_p} x_{i,\xi}\rho_\xi S_\xi\right)^{(n+1)} - \left(\phi\sum_{\xi=1}^{n_p} x_{i,\xi}\rho_\xi S_\xi\right)^{(n)}\right) \\ - \Delta t\sum_{l\in L}\left(\sum_{\xi=1}^{n_p} x_{i,\xi}^{l}\rho_\xi^{l} T_\xi^{l}\Delta\psi_\xi^{l}\right) + \Delta t\sum_{\xi=1}^{n_p} x_{i,\xi}\rho_\xi q_\xi = 0, \end{aligned} \tag{A4}$$

where V is the volume of the grid cell; T_ξ^l is the phase transmissibility between two grid cells connected by the interface l; $\Delta\psi_\xi^l$ is the phase potential difference across the interface l considering pressure, capillary pressure, and gravity; q_ξ is the volumetric rate of phase ξ.

In this work, we simplify the CO_2-brine system to an immiscible two-fluid-phase system with no internal component gradient. Then the mass balance equation reduces to the phase-base balance equation, written as:

$$\frac{\partial}{\partial t}\left(\phi S_\xi \rho_\xi\right) + \nabla\cdot\left(\rho_\xi \mathbf{v}_\xi\right) = \rho_\xi \tilde{q}_\xi, \quad \xi \in \{w,n\}. \tag{A5}$$

The discretized mass balance equation in Equation A4 is then transformed into:

$$V\left(\left(\phi\rho_\xi S_\xi\right) - \left(\phi\rho_\xi S_\xi\right)^t\right) - \Delta t\sum_{l\in L}\rho_\xi^l T_\xi^l \Delta\psi_\xi^l = \Delta t\rho_\xi q_\xi. \tag{A6}$$

For the two grid cells u and v with the interface of l, T_ξ^l is defined as follows:

$$T_\xi^l = \frac{\bar{k}_{u,v}A_l}{L_{u,v}}\lambda_\xi = T_m\lambda_\xi, \tag{A7}$$

where $\bar{k}_{u,v}$ is the harmonic average of permeability between the grid cells u and v; A_l is the area of interface l; $L_{u,v}$ is the distance between the grid cells u and v.

The porosity ϕ in can be represented as $\phi_0\left(1 + c_r\left(p - p_{ref}\right)\right)$. By denoting $1 + c_r\left(p - p_{ref}\right)$ as c_ϕ, Equation A6 can be written as:

$$\left(\left(c_\phi \rho_\xi S_\xi\right) - \left(c_\phi \rho_\xi S_\xi\right)^t\right) - \frac{\Delta t}{V\phi_0}\sum_{l\in L}\rho_\xi^l \lambda_\xi^l T_m^l \Delta\psi_\xi^l = \frac{\Delta t}{V\phi_0}\rho_\xi q_\xi. \tag{A8}$$

By presenting the Equation A8 in algebraic form and the definition in Equations 4–9, Equation A8 can be rewritten as follows:

$$\begin{aligned} &\boldsymbol{r}_\xi\left(\boldsymbol{x},\boldsymbol{m},\boldsymbol{u}^t\right) = \\ &\left(z_{\mathrm{acc},\xi}(\boldsymbol{x}) - z_{\mathrm{acc},\xi}\left(\boldsymbol{x}^t\right)\right) - z_{\mathrm{adv},\xi}(\boldsymbol{x},\boldsymbol{m}) + z_{\mathrm{src},\xi}(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{0}. \end{aligned} \tag{A9}$$

Appendix B: Implementation Details

B1. Description of Decoder and Control Encoder

At each time step t, the control encoder takes the control variable $\boldsymbol{u}^t$ as input and generates the latent variable z_{src}^t as output. The control encoder consists of two 3D convolutional (Conv3D) layers with linear activation functions and a kernel size of 3 on x-, y-, and z-directions. The output from Conv3D layers is downsampled by reducing the spatial dimension by a factor of 2 and projected to the latent space with the feature dimension of d_l.

At each time step t, the decoder takes the latent variables z_{acc}^t and z_{adv}^t as well as the static variable $\boldsymbol{m}$ as inputs and generates the dynamic states $\boldsymbol{S}_n^t$ and $\boldsymbol{p}^t$ as outputs (Figure B1). First, the latent variables from the processor layer are projected back to the original dimension D through the upsampling layers. Second, the inputs to the decoder are projected to the space with the dimension of d_{d_1} for each grid block. The projected features are then fused together to be the variable $\boldsymbol{z}^t$ through addition and concatenation. The role of the input $\boldsymbol{m}$ here is to remove the effect of static variables, $\boldsymbol{T}_m$ and $\boldsymbol{I}_{PV}$, that are coupled in the latent variable z_{adv}^t. Then, the dynamic variable $\boldsymbol{x}^t = \left\{\boldsymbol{p}^t, \boldsymbol{S}_n^t\right\}$ is learned from the latent variable $\boldsymbol{z}^t$ through two Conv3D layers with the dimension of output channels to be d_{d_2} and d_{d_3}, respectively.

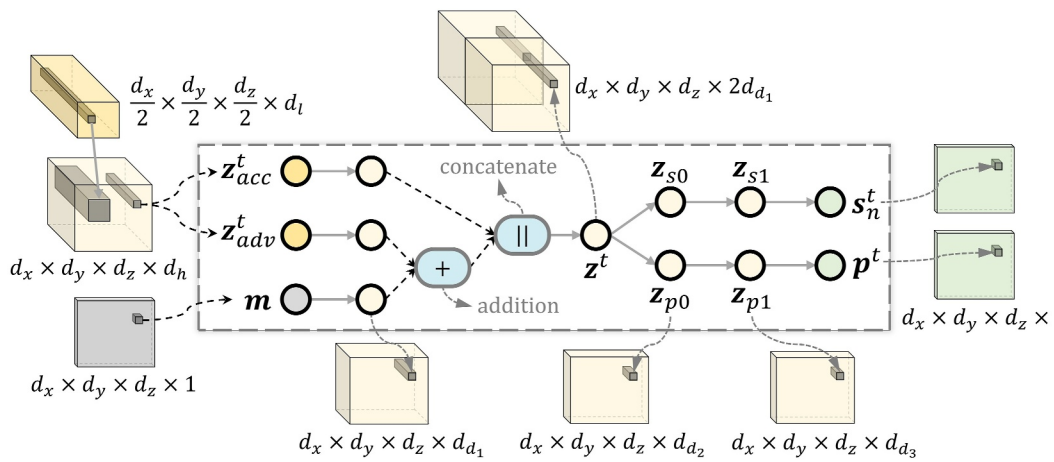

Figure B1. Illustration of the architecture of the decoder.

B2. Hyperparameters of Deep Learning Models

The hyperparameters of deep learning models, including our proposed model and the modified Recurrent R-U-Net are summarized in Tables B1 and B2. Both two models consist of approximately 3 million trainable parameters.

Table B1
Hyperparameters of Each Module of the Proposed Model: Encoder, Processor, and Decoder

Module	Hyperparameters	Value	Range
Encoder	Physical Operator Dimension, d_o	40	{10, 20, 30, 40}
	Hidden Dimension, d_h	16	{8, 12, 16, 20}
	Latent Dimension, d_l	32	{24, 32, 48}
Processor	Dimension of Residual Feature, d_r	48	{32, 48, 64}
	Number of Residual Layers, N_{rl}	3	{2, 3, 4}
Decoder	Dimension of the 1^{st} Layer, d_{d_1}	16	{8, 16, 32}
	Dimension of the 2^{nd} Layer, d_{d_2}	8	{4, 8, 12, 16}
	Dimension of the 3^{rd} Layer, d_{d_3}	4	{4, 6, 8}

Note. The third column denotes the selected values of hyperparameters. The last column denotes the feasible range of the hyperparameters.

Table B2
Hyperparameters of Modified Recurrent R-U-Net

Hyperparameter	Selected size
Channels	[16, 32, 64]
# of Groups for GroupNorm Layer	4

B3. Configuration of Training

The training in this study employs the Adam optimizer as the chosen optimization algorithm, with the decay of the learning rate implemented gradually. Specifically, the learning rate decreases progressively over every step size of training samples by being multiplied by the decay factor. Further details about the optimization hyperparameters can be found in Table B3. The number of samples used for training, validation, and test sets for all models and experiments are tabulated in Table B4. All experiments are implemented using PyTorch (Paszke et al., 2019) and conducted on a single NVIDIA A100 40 GB GPU. Summaries of model efficiency comparison and the input configuration are presented in Tables B5 and B6, respectively. As shown in the input configuration, the first dimension of three variables denotes the batch size of 1. The second and third dimensions of the control variable are the number of time steps of the prediction horizon during training and the number of components (water and CO_2), respectively. The second dimension of the dynamic variable denotes the number of dynamic variables including pressure and saturation. The second dimension of static variable represents the number of static variables, including the inverse of grid volume, depth, and transmissibility of six faces over x-, y-, and z-directions for a structured grid system. The dimension $128 \times 128 \times 20$ represents the grid sizes of the reservoir over three directions.

Table B3
Hyperparameters for the Training of Our Model and RUNET in All Experiments

Hyperparameter	Value
Batch Size	2
Learning Rate	1e−4
Step Size	4,500
Decay Factor	0.9
Gradient Clipping	40

Table B4
The Numbers of Samples Used for Training, Validating, and Testing all Models in Three Experiments

Experiment	Training	Validate	Test
Exp 1 (Public Data set)	200, 500, 1,000, 1,500, 2,000, 2,500	50	373
Exp 1 (Our Data set)	700	100	200
Exp 2	700	100	200

Table B5
Comparison of Model Efficiency

Model	Parameter (million)	Runtime (s/iter)
Ours	3.3951	0.2238
RUNET	3.0371	0.4201
ConvEnc	3.5225	0.4012

Note. ConvEnc refers to the model sharing the same architecture as the proposed DL model except the encoder is the convolutional encoder.

Table B6
Input Size Used for the Comparison of Model Efficiency

Input variable	Input size
Control Variable	$1 \times 8 \times 2 \times 128 \times 128 \times 20$
Dynamic Variable	$1 \times 2 \times 128 \times 128 \times 20$
Static Variable	$1 \times 8 \times 128 \times 128 \times 20$

Appendix C: Numerical Simulation Model and Data Generation

In this section, we describe the numerical simulation model used to generate the training and test data for the described physics-encoded DL model for predicting the spatial-temporal evolution of the pressure and saturation for subsurface CO_2 storage. Figure C1 shows a synthetic 3D two-phase flow simulation model constructed using

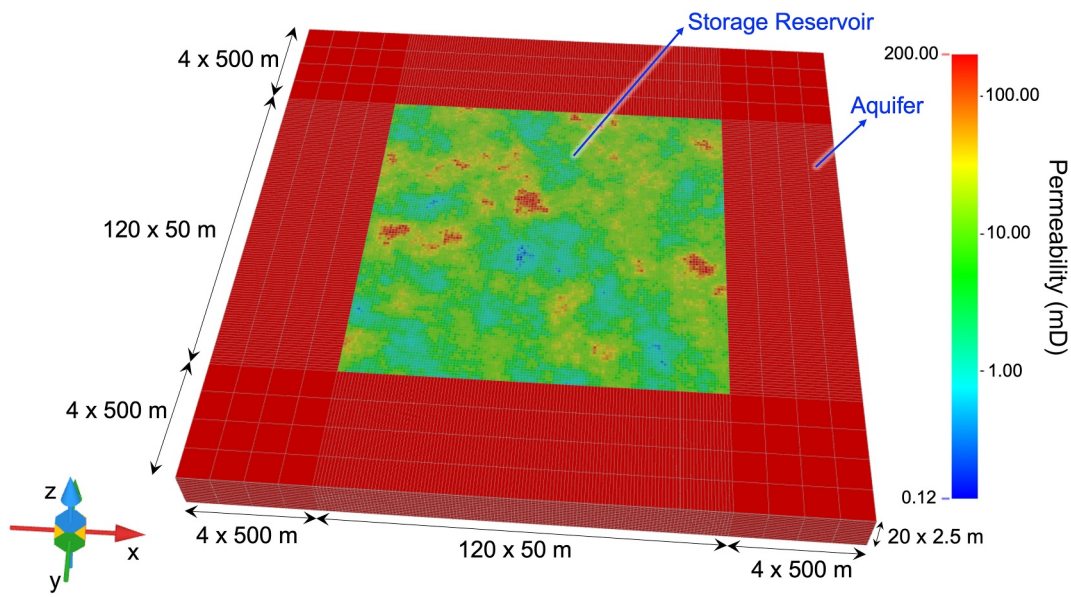

Figure C1. Numerical simulation model of 3D deep saline aquifer reservoir.

Table C1
Numerical Simulation Model Settings

	Reservoir region	Aquifer region
Grid numbers × Grid size (meter) in x and y	120 × 50	8 × 500
Grid numbers × Grid size (meter) in z	20 × 2.5	20 × 2.5
Permeability (mD)	0.1 − 2000	200
Porosity	0.18	0.18
Pore pressure gradient (kPa/m)	9.8	9.8
k_v/k_h	0.1	0.1
Temperature ($°C$) (isothermal condition)	40	40

CMG-GEM (Computer Modeling Group (CMG), 2019). The storage reservoir region is encompassed by an aquifer region used to allow pressure dissipation and minimize the effects of no-flow boundaries. The zero-flux boundary conditions are imposed on the reservoir boundaries. The storage reservoir is at a depth of 1,000 m with a vertical grid dimension of 20 grid blocks × 2.5 m and a horizontal grid dimension of 120 grid blocks × 50 m, resulting in a total grid number of 288,000. The aquifer region is situated at the same depth and exhibits identical vertical and horizontal resolutions for the middle section as the storage reservoir. The two sides of the horizontal resolutions of the aquifer region are coarsened to 4 grid blocks × 500 m to reduce computational cost. The total grid number of the entire simulation model is 327,680. The reservoir region has heterogeneous permeability values ranging from 0.1 to 2,000 mD. Table C1 summarizes the simulation model settings. The injection spans 15 years, followed by a monitoring period of 85 years without CO_2 injection.

In this paper, we assume an isothermal environment for the CO_2 injection and storage processes. Various trapping mechanisms of CO_2 storage, including geologic, residual, and solubility trapping (Blunt, 2010), are considered in the simulation model. We use the Brooks-Corey relative permeability model (Brooks, 1965) with $n = 2$ to calculate the CO_2-water relative permeability. The maximum residual gas saturation (S_{grmax}) is set to be 0.4. The capillary pressure for the sand-CO_2-brine system is referred to Plug and Bruining (2007) of the drainage capillary pressure curve for supercritical CO_2 at 40°C. The gas density is calculated with the Peng-Robinson EoS (Peng & Robinson, 1976). The gas viscosity is calculated from the Jossi, Stiel and Thodos correlation (Poling et al., 2001). The density and viscosity of the aqueous phase are calculated from the Rowe and Chou correlation (Rowe & Chou, 1970) and the Kestin correlation (Kestin et al., 1981), respectively.

Uncertainties in the geologic properties of storage reservoirs, particularly in CO_2 sequestration sites such as saline aquifers, arise due to limited understanding and measurement of geologic representation at various scales (Ma, 2011). These uncertainties stem from factors such as reservoir heterogeneity, anisotropy, and lateral variation resulting from uncertain rock composition, texture, pore structure, and rock type, which can influence the stress state and strength differences within the rocks (Middleton et al., 2012). Consequently, geomechanical properties become uncertain, affecting the capacity and costs of CO_2 storage in carbon capture and storage systems during pre-injection, injection, and post-injection periods (Anderson, 2017). To enhance the robustness of the developed deep learning model and effectively address geologic uncertainty, we employ sampling techniques and ensemble representations. These techniques allow us to incorporate varying levels of heterogeneity range and generate multiple realizations of the permeability map. We utilize sequential Gaussian simulation with the Stanford Geostatistical Modeling Software (Remy et al., 2009) for this purpose. The geometric anisotropy is characterized by a spherical variogram model where the maximum (x-direction) and medium (y-direction) ellipsoid ranges are set isotropically with three distinct scenarios of 20, 60, and 100 units. The minimum ellipsoid range is represented by two distinct scenarios, namely 2 and 5 units. Consequently, a comprehensive set of six scenarios, each featuring different ellipsoid ranges, is utilized to generate a total of 100 realizations per scenario, thereby resulting in a total of 600 realizations of the 3D permeability maps. Figure C2 shows the permeability maps of the first four realizations from each of the six distinct scenarios.

Given each realization of the permeability map, the simulation data is generated based on various cases of well injection schedules. The well location is fixed at 64,64,17 for the one-well scenario and the well indices of the two-well scenario are fixed at {55,55,17} and {75,75,17}. The total CO_2 injection volume, $V^{inj}_{T_{CO_2}}$, over 15 years

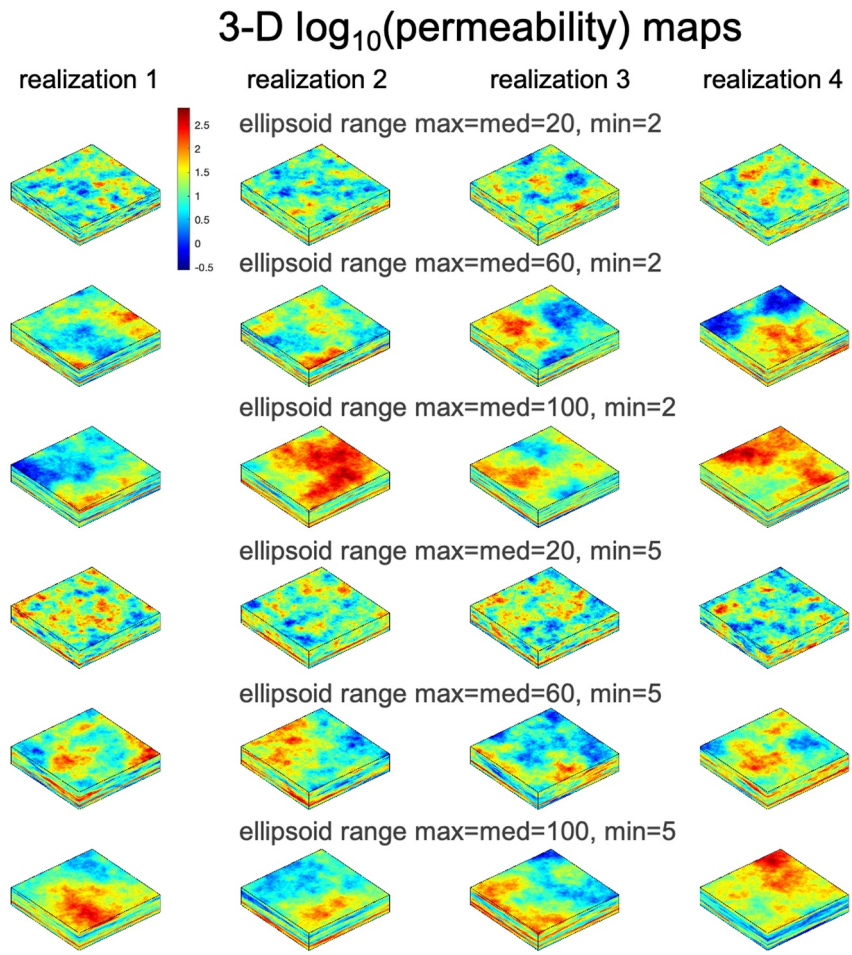

Figure C2. The 3D Permeability maps of the first four realizations from each of the six scenarios with distinct ellipsoid ranges.

of injection is fixed to be $3.5e + 09\ m^3$ ($\sim 10\ mtCO_2$) for both one-well and two-well scenarios. To ensure comprehensive coverage of the realistic response space, we divide the injection duration into three intervals, each spanning 5 years. We assign each interval with an injection volume, $V_{interval_x}$, with one of V_{high}, V_{med}, or V_{low}, where $V_{high} = 0.5V^{inj}_{T_{CO_2}}$, $V_{med} = 0.35V^{inj}_{T_{CO_2}}$, and $V_{low} = 0.15V^{inj}_{T_{CO_2}}$. We then create 10 different combinations of those injection volumes as shown in Table C2 and Figure C3a. The injection rate, $u\ (m^3/day)$, changes every year, and $U = \{u^{t_1}, u^{t_2}, \ldots, u^{t_k}\}$ where $k = 15$. We randomly assign a portion of the injection volume to each year based on its corresponding interval. Let

$$p = rand(1,5) = [p_1, p_2, p_3, p_4, p_5], \tag{C1}$$

and injection rate u at time t_i is calculated by

$$u^{t_i} = \frac{p_i}{sum(p)} \frac{V_{interval_x}}{365}, \tag{C2}$$

where p_i corresponds to the same injection interval x where u^{t_i} belongs to.

Figure C3b shows an example of random injection rate allocation based on total injection allocation in Figure C3a. In this way, a total of 6,000 simulation runs are performed and the spatial and temporal results of

10.1029/2024WR037953

Table C2
Ten Combinations of the Percentage Injection Volume for Each Interval

	$V_{interval_1}$	$V_{interval_2}$	$V_{interval_3}$
Combination 1	V_{high}	V_{med}	V_{low}
Combination 2	V_{high}	V_{low}	V_{med}
Combination 3	V_{med}	V_{high}	V_{low}
Combination 4	V_{med}	V_{low}	V_{high}
Combination 5	V_{low}	V_{high}	V_{med}
Combination 6	V_{low}	V_{med}	V_{high}
Combination 7	V_{avg}	V_{avg}	V_{avg}
Combination 8	V_{high}	0	V_{high}
Combination 9	V_{high}	V_{high}	0
Combination 10	0	V_{high}	V_{high}

Note. V_{high}, V_{med}, V_{low}, and V_{avg} correspond to 50%, 35%, 15%, and 33.33% of the total injection volume, respectively.

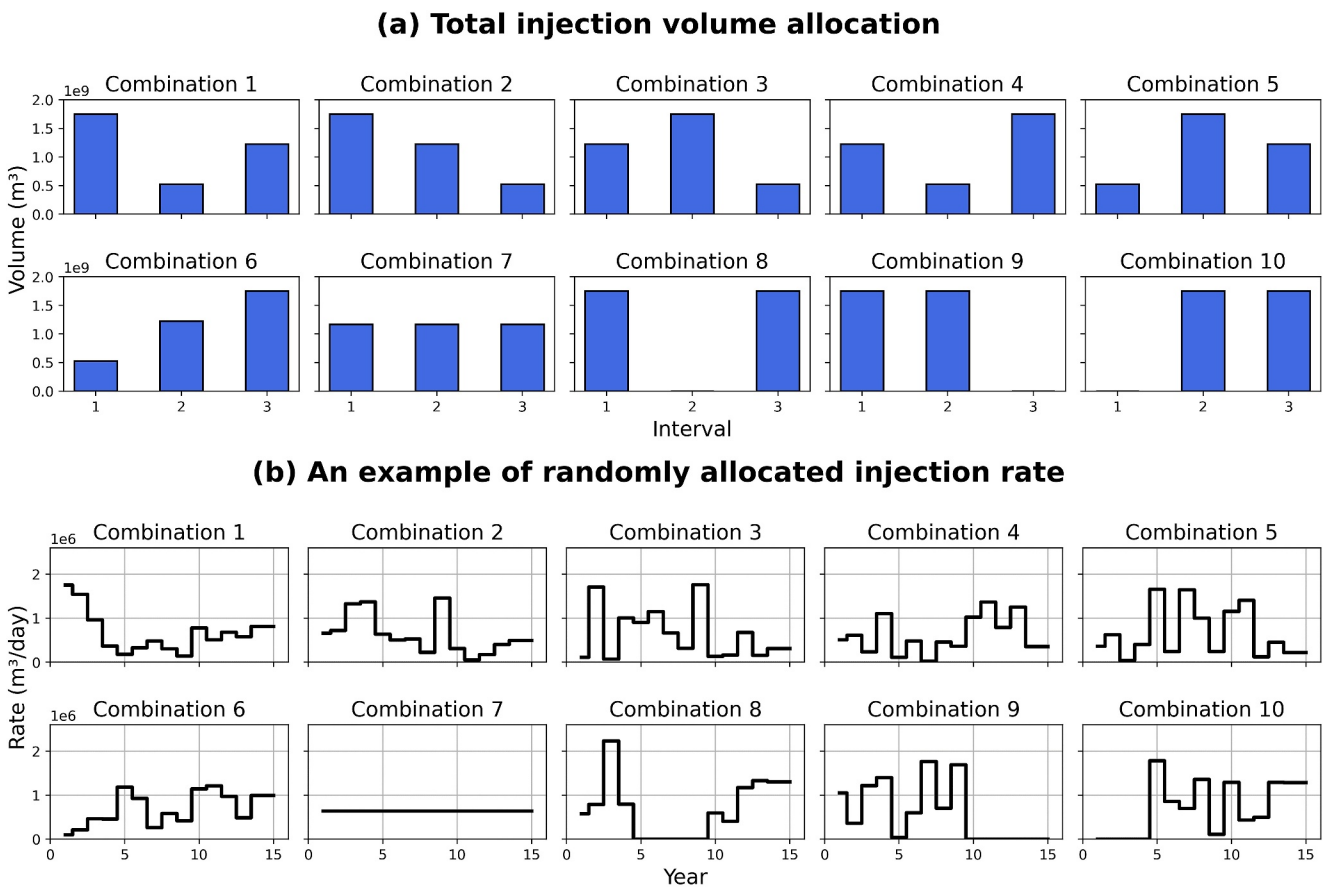

Figure C3. Ten combinations of the percentage injection volume for each interval. (a) Total injection volume allocation for each interval. (b) an example of random injection rate allocation based on total injection allocation on (a).

pressure and gas saturation are generated. The described data generation approach maximizes our ability to explore a wide range of injection rate combinations.

For the two-well scenario, the injection rate is distributed into two wells with specific portions:

$$\left[u_{w_1}^{t_i}, u_{w_2}^{t_i}\right] = u^{t_i}\left[q_i, 1 - q_i\right], \tag{C3}$$

where q_i is the percentage of injection amount assigned to well 1, $(1 - q_i)$ is the percentage of injection amount assigned to well 2, and $q = [0.1 : 0.1 : 1]$ for a total of 10 injection schedules for each permeability realization.

Data Availability Statement

The data, codes, and examples used in this study have been uploaded to the FAIR-compliant Zenodo online repository (link: https://zenodo.org/records/14594842) as well as our research website at http://sees.usc.edu.

Acknowledgments
The authors acknowledge the support provided by the Energi Simulation Foundation through an Industrial Research Chair on Subsurface Energy Data Science at the University of Southern California.

References

Ajayi, T., Gomes, J. S., & Bera, A. (2019). A review of CO_2 storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches. *Petroleum Science*, *16*(5), 1028–1063. https://doi.org/10.1007/s12182-019-0340-8

Anderson, S. T. (2017). Cost implications of uncertainty in CO_2 storage resource estimates: A review. *Natural Resources Research*, *26*(2), 137–159. https://doi.org/10.1007/s11053-016-9310-7

Bandilla, K. W., Celia, M. A., Birkholzer, J. T., Cihan, A., & Leister, E. C. (2015). Multiphase modeling of geologic carbon sequestration in saline aquifers. *Ground Water*, *53*(3), 362–377. https://doi.org/10.1111/gwat.12315

Blunt, M. (2010). Carbon dioxide storage. Grantham Institute Briefing Paper, 4.

Brooks, R. H. (1965). *Hydraulic properties of porous media*. Colorado State University.

Bui, M., Adjiman, C. S., Bardow, A., Anthony, E. J., Boston, A., Brown, S., et al. (2018). Carbon capture and storage (CCS): The way forward. *Energy & Environmental Science*, *11*(5), 1062–1176. https://doi.org/10.1039/C7EE02342A

Celia, M. A., Bachu, S., Nordbotten, J. M., & Bandilla, K. W. (2015). Status of CO_2 storage in deep saline aquifers with emphasis on modeling approaches and practical simulations. *Water Resources Research*, *51*(9), 6846–6892. https://doi.org/10.1002/2015WR017609

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. In *Advances in neural information processing systems* (Vol. 31, pp. 6571–6583). Curran Associates, Inc.

Computer Modeling Group (CMG). (2019). Gem user manual.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, *92*(3), 88. https://doi.org/10.1007/s10915-022-01939-z

Dulny, A., Hotho, A., & Krause, A. (2022). NeuralPDE: Modelling dynamical systems from data. In *KI 2022: Advances in artificial intelligence* (pp. 75–89). Springer International Publishing.

E, W. (2017). A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, *1*(5), 1–11. https://doi.org/10.1007/s40304-017-0103-z

E, W., Han, J., & Jentzen, A. (2017). Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, *5*(4), 349–380. https://doi.org/10.1007/s40304-017-0117-6

Faroughi, S. A., Pawar, N., Fernandes, C., Raissi, M., Das, S., Kalantari, N. K., & Mahjour, S. K. (2023). Physics-Guided, physics-informed, and physics-encoded neural networks in scientific computing. *arXiv:2211.07377*. https://doi.org/10.48550/arXiv.2211.07377

Fuks, O., & Tchelepi, H. A. (2020). Limitations of physics informed machine learning for nonlinear two-phase transport in porous media. *Journal of Machine Learning for Modeling and Computing*, *1*(1), 19–37. https://doi.org/10.1615/JMachLearnModelComput.2020033905

Gorecki, C. D., Ayash, S. C., Liu, G., Braunberger, J. R., & Dotzenrod, N. W. (2015). A comparison of volumetric and dynamic CO_2 storage resource and efficiency in deep saline formations. *International Journal of Greenhouse Gas Control*, *42*, 213–225. https://doi.org/10.1016/j.ijggc.2015.07.018

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 770–778).

Hendrycks, D., & Gimpel, K. (2023). Gaussian error linear units (gelus).

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd international conference on machine learning* (Vol. 37, pp. 448–456). PMLR.

Jiang, X. (2011). A review of physical modelling and numerical simulation of long-term geological storage of CO_2. *Applied Energy*, *88*(11), 3557–3566. https://doi.org/10.1016/j.apenergy.2011.05.004

Jiang, Z., Tahmasebi, P., & Mao, Z. (2021). Deep residual u-net convolution neural networks with autoregressive strategy for fluid flow predictions in large-scale geosystems. *Advances in Water Resources*, *150*, 103878. https://doi.org/10.1016/j.advwatres.2021.103878

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Kestin, J., Khalifa, H. E., & Correia, R. J. (1981). Tables of the dynamic and kinematic viscosity of aqueous nacl solutions in the temperature range 20–150 c and the pressure range 0.1–35 mpa. *Journal of Physical and Chemical Reference Data*, *10*(1), 71–88. https://doi.org/10.1063/1.555641

Kolesnikov, A., Beyer, L., Zhai, X., Puigcerver, J., Yung, J., Gelly, S., & Houlsby, N. (2020). Big transfer (bit): General visual representation learning. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16* (pp. 491–507). Springer International Publishing.

Krishnapriyan, A., Gholami, A., Zhe, S., Kirby, R., & Mahoney, M. W. (2021). Characterizing possible failure modes in physics-informed neural networks. In *Advances in neural information processing systems* (Vol. 34, pp. 26548–26560). Curran Associates, Inc.

Li, Z., Kovachki, N. B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A. M., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *CoRR*. abs/2010.08895. https://arxiv.org/abs/2010.08895

Long, Z., Lu, Y., & Dong, B. (2019). PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, *399*, 108925. https://doi.org/10.1016/j.jcp.2019.108925

Long, Z., Lu, Y., Ma, X., & Dong, B. (2018). PDE-Net: Learning PDEs from data. In *Proceedings of the 35th international conference on machine learning* (pp. 3208–3216). PMLR.

Lu, Y., Zhong, A., Li, Q., & Dong, B. (2018). Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *Proceedings of the 35th international conference on machine learning* (pp. 3276–3285). PMLR.

Ma, Y. Z. (2011). Uncertainty analysis in reservoir characterization and management: How much should we know about what we don't know? In *Uncertainty analysis and reservoir modeling* (pp. 1–15). American Association of Petroleum Geologists. https://doi.org/10.1306/13301404M963458

Middleton, R. S., Keating, G. N., Viswanathan, H. S., Stauffer, P. H., & Pawar, R. J. (2012). Effects of geologic reservoir uncertainty on CO_2 transport and storage infrastructure. *International Journal of Greenhouse Gas Control*, *8*, 132–142. https://doi.org/10.1016/j.ijggc.2012.02.005

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019). Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. *Water Resources Research*, *55*(5), 3856–3881. https://doi.org/10.1029/2018WR024638

Mo, S., Zhu, Y., Zabaras, N., Shi, X., & Wu, J. (2019). Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resources Research*, *55*(1), 703–728. https://doi.org/10.1029/2018WR023528

Muther, T., Dahaghi, A. K., Syed, F. I., & Van Pham, V. (2023). Physical laws meet machine intelligence: Current developments and future directions. *Artificial Intelligence Review*, *56*(7), 6947–7013. https://doi.org/10.1007/s10462-022-10329-8

Nagoor Kani, J., & Elsheikh, A. H. (2019). Reduced-order modeling of subsurface multi-phase flow models using deep residual recurrent neural networks. *Transport in Porous Media*, *126*(3), 713–741. https://doi.org/10.1007/s11242-018-1170-7

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems* (Vol. 32, pp. 8026–8037). Curran Associates, Inc.

Pawar, R. J., Bromhal, G. S., Carey, J. W., Foxall, W., Korre, A., Ringrose, P. S., et al. (2015). Recent advances in risk assessment and risk management of geologic CO_2 storage. *International Journal of Greenhouse Gas Control*, *40*, 292–311. https://doi.org/10.1016/j.ijggc.2015.06.014

Peng, D. Y., & Robinson, D. B. (1976). A new two-constant equation of state. *Industrial & Engineering Chemistry Fundamentals*, *15*(1), 59–64. https://doi.org/10.1021/i160057a011

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

Poling, B. E., Prausnitz, J. M., & O'connell, J. P. (2001). *Properties of gases and liquids*. McGraw-Hill Education.

Qin, Z., Jiang, A., Faulder, D., Cladouhos, T. T., & Jafarpour, B. (2023). Efficient optimization of energy recovery from geothermal reservoirs with recurrent neural network predictive models. *Water Resources Research*, *59*(3), e2022WR032653. https://doi.org/10.1029/2022WR032653

Qin, Z., Jiang, A., Faulder, D., Cladouhos, T. T., & Jafarpour, B. (2024). Physics-Guided deep learning for prediction of energy production from geothermal reservoirs. *Geothermics*, *116*, 102824. https://doi.org/10.1016/j.geothermics.2023.102824

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., et al. (2019). On the spectral bias of neural networks. In *Proceedings of the 36th international conference on machine learning* (Vol. 97, pp. 5301–5310). PMLR.

Raissi, M., Perdikaris, P., & Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Rao, C., Sun, H., & Liu, Y. (2021). Hard encoding of physics for learning spatiotemporal dynamics.

Remy, N., Boucher, A., & Wu, J. (2009). *Applied geostatistics with SGEMS: A user's guide*. Cambridge University Press. Retrieved from https://books.google.com/books?id=AFlegl0OLacC

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing.

Rowe, A. M., Jr., & Chou, J. C. (1970). Pressure-volume-temperature-concentration relation of aqueous sodium chloride solutions. *Journal of Chemical and Engineering Data*, *15*(1), 61–66. https://doi.org/10.1021/je60044a016

Ruthotto, L., & Haber, E. (2020). Deep neural networks motivated by partial differential equations. *Journal of Mathematical Imaging and Vision*, *62*(3), 352–364. https://doi.org/10.1007/s10851-019-00903-1

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 28, pp. 802–810). Curran Associates, Inc.

Shokouhi, P., Kumar, V., Prathipati, S., Hosseini, S. A., Giles, C. L., & Kifer, D. (2021). Physics-informed deep learning for prediction of CO_2 storage site response. *Journal of Contaminant Hydrology*, *241*, 103835. https://doi.org/10.1016/j.jconhyd.2021.103835

Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., et al. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in neural information processing systems* (Vol. 33, pp. 7537–7547). Curran Associates, Inc.

Tang, H., Fu, P., Sherman, C. S., Zhang, J., Ju, X., Hamon, F., et al. (2021). A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. *International Journal of Greenhouse Gas Control*, *112*, 103488. https://doi.org/10.1016/j.ijggc.2021.103488

Tang, M., Ju, X., & Durlofsky, L. J. (2022). Deep-learning-based coupled flow-geomechanics surrogate model for CO_2 sequestration. *International Journal of Greenhouse Gas Control*, *118*, 103692. https://doi.org/10.1016/j.ijggc.2022.103692

Tang, M., Liu, Y., & Durlofsky, L. J. (2020). A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *Journal of Computational Physics*, *413*, 109456. https://doi.org/10.1016/j.jcp.2020.109456

Tang, M., Liu, Y., & Durlofsky, L. J. (2021). Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3D subsurface flow. *Computer Methods in Applied Mechanics and Engineering*, *376*, 113636. https://doi.org/10.1016/j.cma.2020.113636

Voskov, D. V. (2017). Operator-based linearization approach for modeling of multiphase multi-component flow in porous media. *Journal of Computational Physics*, *337*, 275–288. https://doi.org/10.1016/j.jcp.2017.02.041

Wang, N., Zhang, D., Chang, H., & Li, H. (2020). Deep learning of subsurface flow via theory-guided neural network. *Journal of Hydrology*, *584*, 124700. https://doi.org/10.1016/j.jhydrol.2020.124700

Wang, Y., & Lin, G. (2020). Efficient deep learning techniques for multiphase flow simulation in heterogeneous porous media. *Journal of Computational Physics*, *401*, 108968. https://doi.org/10.1016/j.jcp.2019.108968

Wen, G., Hay, C., & Benson, S. M. (2021). CCSNet: A deep learning modeling suite for CO_2 storage. *Advances in Water Resources*, *155*, 104009. https://doi.org/10.1016/j.advwatres.2021.104009

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. *Advances in Water Resources*, *163*, 104180. https://doi.org/10.1016/j.advwatres.2022.104180

Wen, G., Li, Z., Long, Q., Azizzadenesheli, K., Anandkumar, A., & Benson, S. M. (2023). Real-time high-resolution CO_2 geological storage prediction using nested fourier neural operators. *Energy & Environmental Science*, *16*(4), 1732–1741. https://doi.org/10.1039/D2EE04204E

Wen, G., Tang, M., & Benson, S. M. (2021). Towards a predictor for CO_2 plume migration using deep neural networks. *International Journal of Greenhouse Gas Control*, *105*, 103223. https://doi.org/10.1016/j.ijggc.2020.103223

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, *55*(4), 1–37. https://doi.org/10.1145/3514228

Wilson, A. G., & Izmailov, P. (2020). Bayesian deep learning and a probabilistic perspective of generalization. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 4697–4708). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2020/file/322f62469c5e3c7dc3e58f5a4d1ea399-Paper.pdf

Wu, H., Hu, T., Luo, H., Wang, J., & Long, M. (2023). Solving high-dimensional PDEs with latent spectral models.

Wu, Y., & He, K. (2018). Group normalization. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3–19).

Xiong, W., Huang, X., Zhang, Z., Deng, R., Sun, P., & Tian, Y. (2023). Koopman neural operator as a mesh-free solver of non-linear partial differential equations.

Yan, B., Harp, D. R., Chen, B., Hoteit, H., & Pawar, R. J. (2022). A gradient-based deep neural network model for simulating multiphase flow in porous media. *Journal of Computational Physics*, *463*, 111277. https://doi.org/10.1016/j.jcp.2022.111277

Yan, B., Harp, D. R., Chen, B., & Pawar, R. (2022). A physics-constrained deep learning model for simulating multiphase flow in 3D heterogeneous porous media. *Fuel*, *313*, 122693. https://doi.org/10.1016/j.fuel.2021.122693

Zheng, F., Jahandideh, A., Jha, B., & Jafarpour, B. (2021). Geologic CO_2 storage optimization under geomechanical risk using coupled-physics models. *International Journal of Greenhouse Gas Control*, *110*, 103385. https://doi.org/10.1016/j.ijggc.2021.103385

Zheng, F., Jha, B., & Jafarpour, B. (2022). Optimization of CO_2 storage and leakage through caprock fracturing using coupled flow-geomechanics-fracturing simulation. *ECMOR 2022*, *2022*(1), 1–16. https://doi.org/10.3997/2214-4609.202244084

Zhu, Y., & Zabaras, N. (2018). Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, *366*, 415–447. https://doi.org/10.1016/j.jcp.2018.04.018