Electronic Supplementary Information (ESI) for "Real-time High-resolution CO_2 Geological Storage Prediction using Nested Fourier Neural Operators"

Gege Wen,[a*] Zongyi Li,[b] Qirui Long,[a] Kamyar Azizzadenesheli,[c]
Anima Anandkumar,[b,c] Sally M. Benson[a]

[a]Energy Sciences and Engineering,
Stanford University, Stanford, 94305, CA, USA

[b]Computing and Mathematical Sciences,
California Institute of Technology, Pasadena, 91125, CA, USA

[c]NVIDIA Corporation, Santa Clara, 95051, CA, USA

[*]Corresponding author: gegewen@stanford.edu.

This PDF file includes:

Supplementary Text

1 Governing equation

We consider the immiscible multiphase flow problem with CO_2 and water in the context of CO_2 geological storage. The multiphase flow is governed by the basic mass balance equations, and the general forms of mass accumulations are written as (*1*):

$$\frac{\partial M^{\eta}}{\partial t} = -\nabla \cdot \mathbf{F}^{\eta} + q^{\eta}, \tag{1}$$

where η denotes component of CO_2 or $water$, $\mathbf{F}$ denotes the flux, and q denotes the source term. For each component, the mass accumulation term is summed over phase $p = w$ (wetting) or n (non-wetting)

$$M^{\eta} = \varphi \sum_{p} S_p \rho_p X_p^{\eta}. \tag{2}$$

Here φ is the porosity, S_p is the saturation of phase p, and X_p^{η} is the mass fraction of component η in phase p. Component $water$ is the wetting phase in most storage formations with siliciclastic rocks (*2*). Meanwhile, CO_2 and $water$ have mutual solubility; therefore, a small amount of CO_2 dissolves into the wetting phase, and a small amount of $water$ vaporizes into the non-wetting phase.

For both components, the advective mass flux $\mathbf{F}^{\eta}|_{adv}$ is obtained by summing over phases p,

$$\mathbf{F}^{\eta}|_{adv} = \sum_{p} X^{\eta} \mathbf{F}_p = \sum_{p} X^{\eta} \big(- k \frac{k_{r,p} \rho_p}{\mu_p} (\nabla P_p - \rho_p \mathbf{g}) \big) \tag{3}$$

where each individual phase flux $\mathbf{F}_p$ is governed by the multiphase flow extension of Darcy's law derived from the Navier-Stokes equation (*3*). k denotes the absolute permeability, $k_{r,p}$ is the relative permeability of phase p that non-linearly depends on S_p, μ_p is the viscosity of phase p that depends on P_p, and $\mathbf{g}$ is the gravitational acceleration.

Due to the effect of capillarity, the fluid pressure P_p of each phase is

$$P_n = P_w + P_c \tag{4}$$

$$P_w = P_w \tag{5}$$

where the capillary pressure P_c is a non-linear function of S_p. Additionally, porosity φ, density ρ_p, and the flow composition in Equation 2 and Equation 3 are also non-linear functions that depend on P_p.

2 Local Grid Refinement

We inject CO_2 into a 3D saline formation with multiple vertical injection wells. The reservoir is 160,000 m $\times$ 160,000 m $\times$ 100 m with a dip angle along the x-axis. The large spatial domain is chosen to mimic saline formations with infinite-acting boundaries; no flow boundary conditions are used for the top and bottom of the formation. On the global level (level 0), we used a grid resolution of 1,600 m $\times$ 1,600 m $\times$ 20 m to simulate the entire reservoir domain.

Each reservoir has 1 to 4 injection wells. Around each injection well, we apply high-resolution local refinements to capture the multi-phase flow processes. Specifically, since CO_2 is lighter than the formation fluid, the CO_2 plume will migrate upward due to gravity forces. As shown in the sensitivity analysis presented in (*4*), large grid thicknesses, such as the 20 m at level 0, will suppress gravity override and leads to underestimation of the plume footprint. In addition, the injection of CO_2 leads to rapid near-well pressure buildup. Using large lateral grids will greatly underestimate the maximum pressure buildup. Large lateral grids will also ignore the dry-out zone near the injection wells (*5*). Therefore, we gradually refine the grid cell size from 1,600 m $\times$ 1,600 m $\times$ 20 m to 20 m $\times$ 20 m $\times$ 2 m, and named each refinement from level 1 being the coarsest to level 4 being the finest. We refined the entire reservoir depth along the z-axis in levels 1-4 in order to ensure that the gravity override effect is always modeled by high-resolution grids. We chose the finest level grid thickness of 2 m as it provides decent estimates of the plume migration process. For example, as shown in *(4)*, the plume radius error simulated by 2 m-thick grids is within 3% compared to the converged grid thickness of 0.5 m. The injection wells are placed in the middle of level 4. The level 1 refinements never overlap with each other. Table S1 summarizes detailed information for each level, and Figure S1 provides visualizations for levels 0 to 4 grid in a 4-well reservoir.

As for the x and y-axis, we designed the LGRs so that the CO_2 plume always stays within levels 1 to 4. Specifically, levels 2 to 4 are placed on the right-hand side of level 1 (Figure S1) because the simulated reservoir dips downwards along the x-axis. CO_2 is more likely to migrate up-dip, especially for high permeable reservoirs. This design ensures the CO_2 gas saturation plume never enters the level 0 grids. Overall, by using this four-level grid refinement on the x, y, and z-axis, we accurately captured the CO_2 plume migration, near well pressure response, as well as far-field pressure inference in multi-well reservoirs.

3 Data set generation

To demonstrate the potential to make the Nested FNO a general-purpose numerical simulator alternative, we sampled a wide range of inputs to cover most realistic scenarios for CO_2 injection into saline formations. This differs from most 'surrogate' modeling approaches with fixed reservoir conditions and permeability maps constrained by a prior. The following sections introduce the sampling ranges for reservoir conditions, injection schemes, and permeability fields. Sampling ranges are summarized in Table S2.

3.1 Reservoir conditions

We sample the reservoir depth from 800 m to 4,500 m. For the upper bound, injecting deeper than 800 m ensures that the CO_2 plume stays in the supercritical state. For the lower bound, our sampling range is deeper than commonly used ranges for CCS ($\sim$ 3,000 m) to include deep offshore formations such as in the Gulf of Mexico. Reservoir temperatures are calculated according to the reservoir depth and a geothermal gradient variable, which we sampled from 15 to 35 C$^\circ$ per km. Within each formation, we assume isothermal conditions. Deeper reservoirs lead to higher initial pressures and temperatures, which in turn introduce a more diverse range of CO_2 properties (e.g., density, solubility, and viscosity).

Our data set also considers a range of reservoir dip angles. Since CO_2 is lighter than formation fluid, it would naturally migrate up-dip. The influence on plume migration is particularly significant for large dip angles and high permeability formations. We sampled dip angles from 0 to 2° to consider this effect and then discarded the reservoirs that are shallower than 800m or deeper than 4,500 m. The dip angle is applied to the x-axis, and the initial hydrostatic pressure in each cell is assigned according to the cell depth. Figure 1 in the main manuscript document demonstrates a reservoir with four injection wells, where the deepest and shallowest injection depths are more than 1,000 meters apart vertically due to the dip angle. As a result, the plume migration and pressure buildup around these two wells are influenced by significantly different reservoir conditions.

3.2 Injection scenarios

We randomly place one to four wells in each reservoir. Each well is at least 5,000 m away from the reservoir boundaries to avoid pressure influence. Each well has a constant injection rate

sampled from 0.5 to 2 MT/year over 30 years. We also sampled perforation lengths from 20 m to 100 m for each injection well and randomly placed the perforation interval along the well.

3.3 Permeability fields

To generate the 3D heterogeneous permeability fields, we used a modified version of Stanford Geostatistical Modeling Software (SGeMS) (*6*) to create random media according to 3D spatial correlations. For each permeability field, we sampled random correlation lengths on the x-axis from 800 m to 4,000 m, y-axis from 800 m to 4,000 m, and z-axis from 4 m to 20 m. These spatial correlation ranges are chosen to cover typical depositional environments (i.e., shoreface, deltaic, marginal marine) found in existing CO_2 storage and hydrocarbon reservoirs (*7–11*). The random field is generated at the resolution of level 2 and subsequently up-sampled for levels 0-1 level and down-sampled for levels 3-4 using tri-linear interpolation. Note that the up-sampling and down-sampling of the permeability map is *not* performed to capture the flow characteristics. Instead, it simply re-grids the random field to another resolution while maintaining the spatial correlations for training. This is different from an 'up-scaling' context, where the coarser permeability field is expected to capture the flow of the finer permeability field.

Then, we scaled the distribution of each permeability field according to a pair of the randomly sampled ln permeability mean and ln permeability standard deviation (Table S2). The maximum cell permeability in the data set is 14 Darcy, and the minimum is kept at 0.1 mD. Figure 1 (main text) shows an example of the permeability field.

4 Numerical simulation setting

We use a full physics numerical simulator, ECLIPSE (e300), to solve the governing PDEs. ECLIPSE uses the finite difference method with upstream weighting for spatial discretization and the fully implicit method for temporal discretization (*12*). The benchmark study by Class et al. (*13*) compared a wide variety of numerical simulators for CO_2 storage problems (i.e., ECLIPSE, TOUGH2 (*14*), GEM-CMG (*15*)). They found that ECLIPSE has leading computational performance for compositional multiphase flow among the studied simulators. Therefore, we chose ECLIPSE to generate the training dataset in this study. ECLIPSE uses the non-conforming grid technique to simulate the domain with local grid refinement. We do not explicitly simulate molecular diffusion and hydrodynamic dispersion. However, numerical simulation results include some unavoidable numerical diffusion and dispersion resulting

from finite difference gradient approximation. Since we applied LGRs around each injection well, the number of cells in each reservoir ranges from $\sim$0.3 million to $\sim$1 million, depending on the number of injection wells. We run each ECLIPSE numerical simulation on 20 parallel Intel®Xeon®E5-2640 CPU using Stanford University's HPC cluster. The average ECLIPSE run times are summarized in Table S6 (b).

Corey's curve and Leverette-J function scaling is used for relative permeability and capillary pressure. Relative permeability curves control the mobility of the liquid and the gas phase. Lower irreducible water saturation generally leads to more compact CO_2 plumes compared to higher irreducible water saturation (*16, 17*). Capillary pressure curves describe the capillary entry pressure and the rate of capillary pressure rising as CO_2 invades rock pores. It is also an important measurement of the wettability of the reservoir (*18–20*). In our previous work, we demonstrated that these curves could also be variables in deep learning models by sampling the irreducible water saturation and the scaling factor for capillary pressure curves (*21, 22*). Since the methodology of this paper focuses on predicting temporal-3D outputs with locally refined grids and multiple injection wells, we chose to prioritize incorporating injection scenarios, reservoir conditions, and permeability fields instead of rock properties in our training dataset; however, rock properties can be incorporated using the same approach described in (*21, 22*) in future work.

After running the numerical simulation, we extract 24-time snapshots for gas saturation and pressure buildup during 30 years of injection. We sampled the time snapshots with higher frequencies at the beginning of the injection and lower frequencies near the end of the injection. This technique is commonly used in numerical simulations to guide engineering decisions.

The data set includes 3011 cases for level 0 and 7374 for levels 1-4 because each reservoir can have one to four injection wells. We then split the data into training/validation/testing sets at an 8/1/1 ratio, which gives us 2408/301/302 for level 0 and 5916/731/727 for LGR models.

5 Fourier Neural Operator

The computational domain of the Nested Fourier Neural Operator (FNO) is a 3D space with time, $D = \Omega \times T$, where T is the time interval of 30 years and Ω is the 3D spatial domain. The Nested FNO consists of nine FNOs for five levels of sub-domains where each subdomain corresponds to a level of grid refinement. The inputs and outputs are functions defined on the 4D domain from the corresponding function spaces $\mathcal{A} = \mathcal{U} = L^{\infty}(D)$.

Each FNO model learns an infinite-dimensional-space mapping from a finite collection of input-output observations (i.e. data set $\{a_j, u_j\}_{j=1}^N$). We use n-point space-time discretization $D_j = \{\xi_1, \ldots, \xi_n\} \subset D$, where $\xi = (\omega, t)$, to numerically represent $a_j|_{D_j} \in \mathbb{R}^{n \times d_a}$ and $u_j|_{D_j} \in \mathbb{R}^{n \times d_u}$ so that a_j consists of the input coefficient functions and u_j consists of output functions of gas saturation S and pressure buildup P. FNO is a type of neural operator (Definition 1) where we use the Fourier integral kernel operator (Definition 2) as the linear integral operator to achieve efficient and accurate training.

Definition 1 (Neural operator)**.** *A L-layered neural operator $\mathcal{G}_\theta$ is defined as:*

$$\mathcal{G}_\theta := \mathcal{Q} \circ (W_L + \mathcal{K}_L) \circ \ldots \circ \sigma(W_1 + \mathcal{K}_1) \circ \mathcal{P}, \tag{6}$$

where $\mathcal{P}$ is a pointwise operator that expands the input function to a higher co-dimension space, parameterized with $\mathcal{P} : \mathbb{R}^{d_a} \to \mathbb{R}^{d_{l=1}}$. In each layer for $l \in 1...L$, $\mathcal{K}$ is a linear integral operator, W is a linear matrix operator, and σ is a non-linear activation. Lastly, $\mathcal{Q}$ is a pointwise operator that projects the function to the output space, parameterized with $\mathcal{Q} : \mathbb{R}^{d_{l=L}} \to \mathbb{R}^{d_u}$.

All parameters in $\mathcal{P}$, $\mathcal{K}_l$, W_l, and $\mathcal{Q}$ are learned through training. By stacking multiple neural operator layers, $\mathcal{G}_\theta$ can approximate the mapping of high dimensional functions with strong non-linearity.

Definition 2 (Fourier integral kernel operator)**.** *An integral kernel operator $\mathcal{K}$ is defined by*

$$\big(\mathcal{K}(v_l)\big)(\xi) = \int_D \kappa(\xi, \xi) v_l(\xi') \mathrm{d}\xi', \quad \forall \xi' \in D. \tag{7}$$

To efficiently parameterize kernel $\mathcal{K}$, the FNO method (23) represents v_l in the Fourier space to directly parameterize κ by its Fourier coefficients:

$$\big(\mathcal{K}(v_l)\big)(\xi) = \mathcal{F}^{-1}\big(R \cdot \mathcal{F}(v_l)\big)(\xi), \quad \forall \xi \in D. \tag{8}$$

where R is the Fourier transform of a periodic function κ, $\mathcal{F}$ denotes a Fourier transform of a function $f : D \to \mathbb{R}^c$ and $\mathcal{F}^{-1}$ is its inverse. To further improve computational efficiency and reduce memory consumption, we truncate the Fourier series at a maximum number of modes (k_{max}), and then parameterize R with the truncated Fourier coefficients. Therefore, we can implement R using a linear parameterization as

$$\big(R \cdot \mathcal{F}(v_l)\big)_{k,i} = \sum_{j=1}^{c} R_{k,i,j} (\mathcal{F}(v_l))_{k,j}, \quad \forall k = 1, ..., k_{max},\ i = 1, ..., c. \tag{9}$$

Using the Fourier integral kernel operator as $\mathcal{K}_{1...L}$ gives us the FNO architecture. Since the data is locally provided on uniform grids, we utilize Fast Fourier Transform (FFT) to implement and approximate the Fourier transform. The speed and low memory requirement of FFT make our proposed method computationally fast with a significantly low memory footprint. We denote FNO with $\mathcal{G}$ (in short of $\mathcal{G}_\theta$) throughout this paper.

We summarized the parameters in the following tables: level 0 (Table S3), level 1 (Table S4), and levels 2 to 4 (Table S5). We used 4 Fourier layers for each sub-model with a width of 28. The modes (t, x, y, z) are $(4, 20, 20, 2)$ for the level 0 and $(6, 10, 10, 10)$ for levels 1 to 4.

6 Training procedure

We use Nvidia A100-SXM GPUs for the training. Each model fits into the GPU with a batch size of one. We use the relative $l2$-loss function for the training because it provides normalized loss across cases, leading to excellent gradient propagation compared to a vanilla $l2$-loss or RMSE error. Each model was trained for 30 and 40 epochs and fine-tuned for around 10 epochs until the validation set error plateaus. Instead of relative $l2$-loss, we monitored the model level plume gas saturation error (main text, Equation 4) and relative pressure buildup error (main text, Equation 3) on the validation set as an indicator of when to stop training. We do not monitor the relative $l2$-loss because it is normalized over the sample's norm, making the magnitudes carry different meanings among different refinement levels.

Note that the numerical simulator domains at levels 0 to 3 are annular because they do not include information at the finer refinement levels. To help with the FNO learning (which prefers continuous functions), we construct the continuous training domains by down-sampling the information from the finer LGRs to populate the annular domains. Our experiments show that the continuous training domains provide more efficient learning than the annular simulator domains. The down-sampled information is only used during the training and is not reported in the final prediction. Figure S2 shows the simulator domain and training domain at each level.

7 Speedup analysis

Once the Nested FNO model is trained, we can directly use it as a numerical simulator alternative for CO_2 gas saturation and pressure buildup prediction in 3D saline formations. Therefore, we compared the ML model prediction time with the numerical simulation time to compute

the computational efficiency speed-up. We summarized each model's training and prediction times in the Nested FNO in Table S6 (a). The machine learning models are evaluated on Nvidia A100-SXM GPUs the prediction time is calculated by taking the average over 1,000 cases. All models take $\sim$0.005 s to evaluate with negligible variances. The prediction time depends on the number of injection wells: $t_{total} = t_{global} + n_{well} \times \sum_{i=1}^{4} t_i$. We run each ECLIPSE numerical simulation on 20 parallel Intel®Xeon®E5-2640 CPU using Stanford HPC cluster. The run time also depends on the number of injection wells since more wells lead to more cells.

Table S6 (b) shows that the Nested FNO provides 4×10^5 to 7 $\times10^5$ times speed-up compared to traditional numerical simulation. Interestingly, four-well cases with larger cell counts have the most significant speed-up. The advantage of using Nested FNO is more prominent with higher dimensional cases, e.g., more injection wells or levels of refinements. In addition, Nested FNO can predict pressure buildup or gas saturation separately, whereas numerical simulation always calculates all variables regardless of whether they are needed. The speed-up is calculated considering separate predictions for gas saturation and pressure buildup.

8 Accuracy summary

Table S7 shows an accuracy summary for train, validation, and test sets in the simulator domain, the training domain with separate prediction, and the training domain with sequential prediction.

9 Probabilistic assessment

We use Nested FNO to conduct probabilistic assessments on CO_2 plume footprint and maximum pressure buildup. We consider CO_2 injection through 4 injection wells where each well injects at a 1MT/year rate. All wells inject through a perforation located at the bottom 20 meters of the reservoir. We generated 1,000 permeability realizations where the permeability average and standard deviation for each map are 85mD and 100mD, respectively. The correlation lengths on x, y, and z directions are 2000, 4000, and 20 meters, respectively. The 10 realizations shown in Figure 4 (main text) were taken from the same injection well.

CO_2 plume footprint is defined as the area of land under the separate phase CO_2 plume, which has a gas saturation larger than 0.01. We calculate the plume footprint using the Nested FNO's gas saturation prediction and the xy-cross section area of each grid cell. The plume footprint reported in Figure 4 (main text) is the reservoir's total plume footprint; the maximum

pressure buildup reported in Figure 4 (main text) is the reservoir's maximum.

Fig. S1.

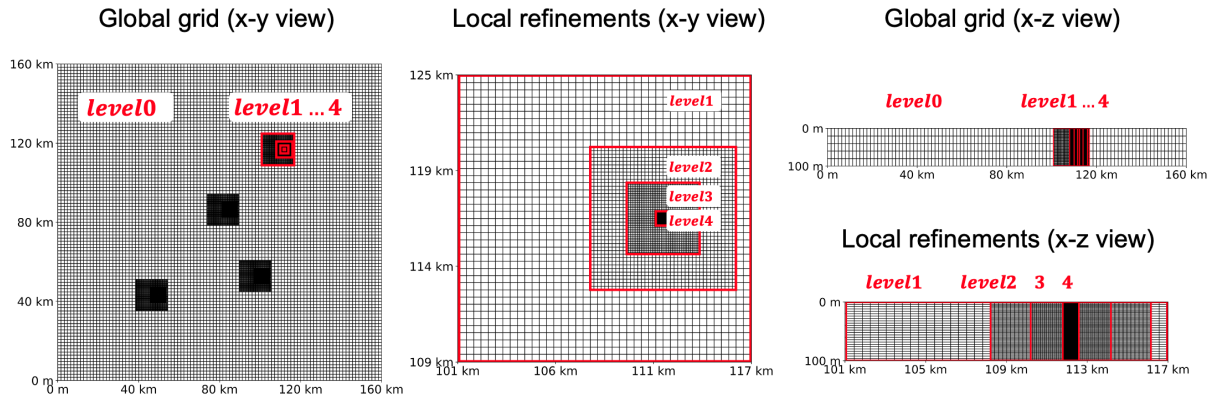

Figure S1: Grid visualizations for the entire reservoir (x-y view), local refinements around the injection well (x-y view), entire reservoir (x-z view) and local refinements around the injection well (x-z view).

Fig. S2.

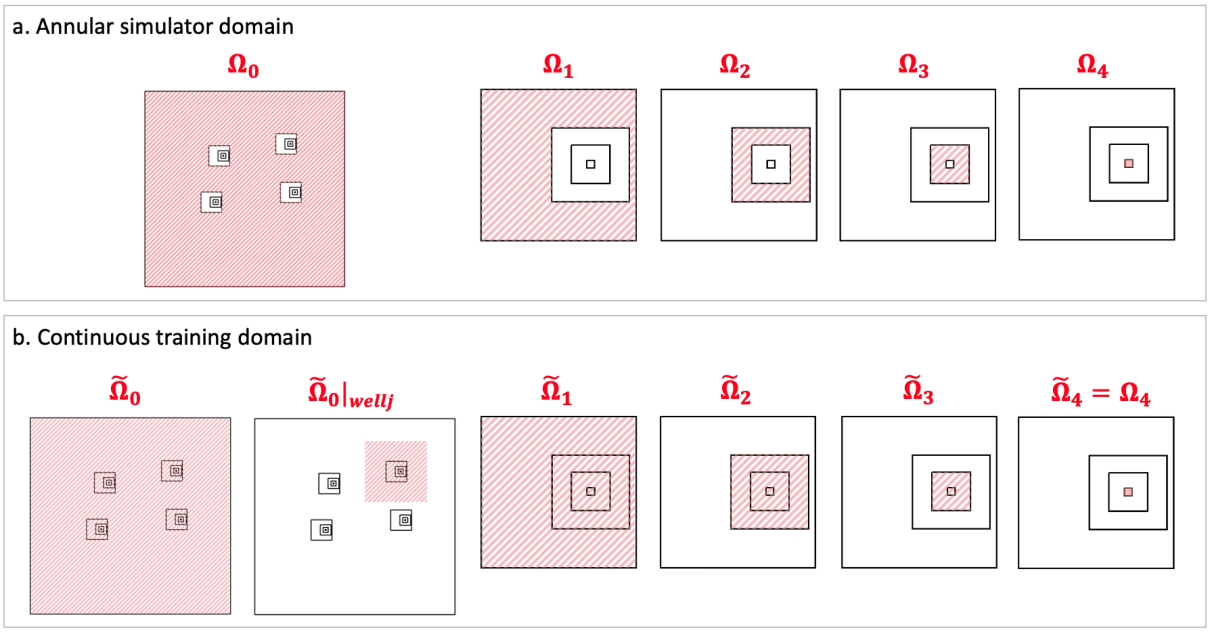

Figure S2: Simulator domain and training domain for each level

Table S1.

Table S1: Spatial dimension, cell size, matrix dimension, and the number of cells for level 0-4 grids.

	Spatial dimension (m)	Cell size (m)	Matrix dim	# of cells
level 0	160,000×160,000×100	1,600×1,600×20	100×100×5	50,000
level 1	16,000×16,000×100	400×400×4	40×40×25	40,000
level 2	8,000×8,000×100	200×200×2	40×40×50	80,000
level 3	4,000×4,000×100	100×100×2	40×40×50	80,000
level 4	800×800×100	20×20×2	40×40×50	80,000

Table S2.

Table S2: Variables parameters and sample ranges used for generating the input data set.

Variable type	Sampling parameter	Distribution	Unit
Permeability map	x-axis correlation	$X \sim \mathcal{U}[800, 4000]$	m
	y-axis correlation	$X \sim \mathcal{U}[800, 4000]$	m
	z-axis correlation	$X \sim \mathcal{U}[4, 20]$	m
	reservoir permeability mean	$X \sim \mathcal{U}[4.09, 5.01]$	$\ln mD$
	reservoir permeability std	$X \sim \mathcal{U}[0.25, 1]$	$\ln mD$
Reservoir cond.	Reservoir center depth	$X \sim \mathcal{U}[800, 4500]$	m
	Geothermal gradient	$X \sim \mathcal{U}[15, 35]$	°C/km
	Dip angle	$X \sim \mathcal{U}[0, 2]$	°C/km
Injection design	# of wells	$n \in \{1, 2, 3, 4\}$	-
	Injection rate	$X \sim \mathcal{U}[0.5, 2]$	MT/y
	Perforation thickness	$X \sim \mathcal{U}[20, 100]$	m
	Perforation location	Randomly placed	-

Table S3.

Table S3: FNO model parameters for the level 0 level. The `Padding` denotes a padding operator that accommodates the non-periodic boundaries; `Linear` denotes the linear transformation to lift the input to the high dimensional space, and the projection back to original space; `Fourier4d` denotes the 4D Fourier operator; `Conv1d` denotes the bias term; `Add` operation adds the outputs together; `GELU` denotes a Gaussian Error Linear Units layer.

Layer	Operation	Output Shape
Input	-	(24, 100, 100, 5, 8)
Padding	`Padding (8)`	(40, 116, 116, 21, 8)
Lifting	`Linear`	(40, 116, 116, 21, 28)
Fourier 1	`Fourier4d/Conv1d/Add/GELU`	(40, 116, 116, 21, 28)
Fourier 2	`Fourier4d/Conv1d/Add/GELU`	(40, 116, 116, 21, 28)
Fourier 3	`Fourier4d/Conv1d/Add/GELU`	(40, 116, 116, 21, 28)
Fourier 4	`Fourier4d/Conv1d/Add`	(40, 116, 116, 21, 28)
De-padding	`Depadding (8)`	(24, 100, 100, 5, 28)
Projection 1	`Linear`	(24, 100, 100, 5, 112)
Projection 2	`Linear`	(24, 100, 100, 5, 1)

Table S4.

Table S4: FNO model parameters for the level 1 level. The `Padding` denotes a padding operator that accommodates the non-periodic boundaries; `Linear` denotes the linear transformation to lift the input to the high dimensional space, and the projection back to original space; `Fourier4d` denotes the 4D Fourier operator; `Conv1d` denotes the bias term; `Add` operation adds the outputs together; `GELU` denotes a Gaussian Error Linear Units layer.

Layer	Operation	Output Shape
Input	-	(24, 40, 40, 25, 9)
Padding	`Padding (8)`	(40, 56, 56, 41, 9)
Lifting	`Linear`	(40, 56, 56, 41, 28)
Fourier 1	`Fourier4d/Conv1d/Add/GELU`	(40, 56, 56, 41, 28)
Fourier 2	`Fourier4d/Conv1d/Add/GELU`	(40, 56, 56, 41, 28)
Fourier 3	`Fourier4d/Conv1d/Add/GELU`	(40, 56, 56, 41, 28)
Fourier 4	`Fourier4d/Conv1d/Add`	(40, 56, 56, 41, 28)
De-padding	`Depadding (8)`	(24, 40, 40, 25, 28)
Projection 1	`Linear`	(24, 40, 40, 25, 112)
Projection 2	`Linear`	(24, 40, 40, 25, 1)

Table S5.

Table S5: FNO model parameters for levels 2-4. The `Padding` denotes a padding operator that accommodates the non-periodic boundaries; `Linear` denotes the linear transformation to lift the input to the high dimensional space, and the projection back to original space; `Fourier4d` denotes the 4D Fourier operator; `Conv1d` denotes the bias term; `Add` operation adds the outputs together; `GELU` denotes a Gaussian Error Linear Units layer.

Layer	Operation	Output Shape
Input	-	(24, 40, 40, 50, 9)
Padding	`Padding (8)`	(40, 56, 56, 66, 9)
Lifting	`Linear`	(40, 56, 56, 66, 28)
Fourier 1	`Fourier4d/Conv1d/Add/GELU`	(40, 56, 56, 66, 28)
Fourier 2	`Fourier4d/Conv1d/Add/GELU`	(40, 56, 56, 66, 28)
Fourier 3	`Fourier4d/Conv1d/Add/GELU`	(40, 56, 56, 66, 28)
Fourier 4	`Fourier4d/Conv1d/Add`	(40, 56, 56, 66, 28)
De-padding	`Depadding (8)`	(24, 40, 40, 50, 28)
Projection 1	`Linear`	(24, 40, 40, 50, 112)
Projection 2	`Linear`	(24, 40, 40, 50, 1)

Table S6.

Table S6: (a). Prediction time, training time, and the number of parameters for levels 0-4 models. The prediction time is calculated based on an average of 250 runs. (b) Speed up calculation for ECLIPSE vs. Nested FNO. We show the number of cells and average run time of the ECLIPSE simulation and the Nested FNO prediction time, both averaging over 250 cases. The speed-up is calculated for predicting gas saturation and pressure buildup separately.

(a)

Model	Prediction (s)	Training (hr/epoch)	# of parameter
level 0	0.005	0.96	80,288,461
level 1	0.005	1.43	150,534,889
level 2	0.005	2.20	150,534,889
level 3	0.005	2.20	150,534,889
level 4	0.005	2.20	150,534,889

(b)

# of Wells	# of Cells	ECLIPSE (hr)	Nested FNO (s)	Speed-up
1	296,300	2.75	0.025	3.96×10^5
2	542,600	6.43	0.045	5.14×10^5
3	788,900	11.00	0.065	6.09×10^5
4	1,035,200	15.93	0.085	6.75×10^5

Table S7.

Table S7: Accuracy summary for train, validation, and test sets in the simulator domain, the training domain with separate prediction, and the training domain with sequential prediction.

(a) Pressure buildup

Simulator domain	Train (%)	Validation (%)	Test (%)
δ^P	0.34	0.48	0.47
Training domain, separate	Train (%)	Validation (%)	Test (%)
δ_0^P	0.02	0.02	0.02
δ_1^P	0.09	0.11	0.10
δ_2^P	0.15	0.16	0.16
δ_3^P	0.14	0.14	0.14
δ_4^P	0.42	0.47	0.45
Training domain, sequential	Train (%)	Validation (%)	Test (%)
δ_0^P	0.02	0.02	0.02
δ_1^P	0.08	0.15	0.16
δ_2^P	0.19	0.31	0.30
δ_3^P	0.38	0.50	0.51
δ_4^P	0.59	0.81	0.82

(b) Gas saturation

Simulator domain	Train (%)	Validation (%)	Test (%)
δ^S	1.16	1.81	1.79
Training domain, separate	Train (%)	Validation (%)	Test (%)
δ_1^S	0.78	1.32	1.27
δ_2^S	0.84	1.02	1.00
δ_3^S	0.48	0.61	0.61
δ_4^S	0.67	0.79	0.74
Training domain, sequential	Train (%)	Validation (%)	Test (%)
δ_1^S	0.73	1.36	1.39
δ_2^S	1.17	1.88	1.91
δ_3^S	1.00	1.74	1.77
δ_4^S	1.16	1.85	1.82

References

1. K. Pruess, C. M. Oldenburg, G. Moridis, Tough2 user's guide version 2, *Tech. rep.*, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States) (1999).

2. R. Pini, S. C. Krevor, S. M. Benson, Capillary pressure and heterogeneity for the CO_2/water system in sandstone rocks at reservoir conditions, *Advances in Water Resources* **38**, 48 (2012).

3. M. J. Blunt, *Multiphase flow in permeable media: A pore-scale perspective* (Cambridge university press, 2017).

4. G. Wen, S. M. Benson, CO_2 plume migration and dissolution in layered reservoirs, *International Journal of Greenhouse Gas Control* **87**, 66 (2019).

5. L. André, Y. Peysson, M. Azaroual, Well injectivity during co2 storage operations in deep saline aquifers–part 2: Numerical simulations of drying, salt deposit mechanisms and role of capillary forces, *International Journal of Greenhouse Gas Control* **22**, 301 (2014).

6. N. Remy, A. Boucher, J. Wu, *Applied Geostatistics with SGeMS: A User's Guide* (Cambridge University Press, 2009).

7. F. P. Shepard, D. G. Moore, Central texas coast sedimentation: Characteristics of sedimentary environment, recent history, and diagenesis: Part 1, *AAPG Bulletin* **39**, 1463 (1955).

8. T. Ito, T. Nakajima, Z. Xue, Geological reservoir characterization and modelling of a co2 storage aquifer: A case study of the nagaoka site, japan, *Energy Procedia* **114**, 2792 (2017).

9. Y. Z. Ma, X. Zhang, *Quantitative geosciences: Data analytics, geostatistics, reservoir characterization and modeling* (Springer, 2019).

10. R. S. Jayne, H. Wu, R. M. Pollyea, Geologic co2 sequestration and permeability uncertainty in a highly heterogeneous reservoir, *International Journal of Greenhouse Gas Control* **83**, 128 (2019).

11. M. Fawad, M. J. Rahman, N. H. Mondol, Seismic reservoir characterization of potential co2 storage reservoir sandstones in smeaheia area, northern north sea, *Journal of Petroleum Science and Engineering* **205**, 108812 (2021).

12. Schlumberger, Eclipse reservoir simulation software reference manual (2014).

13. H. Class, *et al.*, A benchmark study on problems related to co 2 storage in geologic formations: summary and discussion of the results, *Computational geosciences* **13**, 409 (2009).

14. K. Pruess, The tough codes—a family of simulation tools for multiphase flow and transport processes in permeable media, *Vadose zone journal* **3**, 738 (2004).

15. C. M. Group, Gem user guide. (2006).

16. C. Doughty, K. Pruess, Modeling supercritical carbon dioxide injection in heterogeneous porous media, *Vadose Zone Journal* **3**, 837 (2004).

17. M. Flett, R. Gurton, I. Taggart, *SPE Asia Pacific Oil and Gas Conference and Exhibition* (OnePetro, 2004).

18. W.-J. Plug, J. Bruining, Capillary pressure for the sand–co2–water system under various pressure conditions. application to co2 sequestration, *Advances in Water Resources* **30**, 2339 (2007).

19. A. Al-Menhali, B. Niu, S. Krevor, Capillarity and wetting of carbon dioxide and brine during drainage in berea sandstone at reservoir conditions, *Water Resources Research* **51**, 7895 (2015).

20. H. Wu, R. S. Jayne, R. M. Pollyea, A parametric analysis of capillary pressure effects during geologic carbon sequestration in a sandstone reservoir, *Greenhouse Gases: Science and Technology* **8**, 1039 (2018).

21. G. Wen, C. Hay, S. M. Benson, CCSNet: a deep learning modeling suite for CO_2 storage, *Advances in Water Resources* p. 104009 (2021).

22. G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, S. M. Benson, U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow, *Advances in Water Resources* **163**, 104180 (2022).

23. Z. Li, *et al.*, Fourier neural operator for parametric partial differential equations, *arXiv preprint arXiv:2010.08895* (2020).