Supplementary Notes

Contents

1 Neural network architectures

The CoNFiLD-geo framework is constructed through a synergistic integration of two complementary neural networks: the Conditional Neural Field (CNF) and the Latent Diffusion Model (LDM). The architectural and training details of both components are elaborated in the following subsections

1.1 Conditional neural field architecture

The schematic of the neural field with full-projection conditioning is illustrated in Fig. S1. The sinusoidal representation network (SIREN; lower box) [1] is modulated in a layer-wise manner by a dedicated modulation network. The SIREN receives spatial coordinates $\boldsymbol{X}$ as input and predicts the physical field $\boldsymbol{\Phi}$ at the corresponding spatial locations. The modulation network (upper box) takes as input a learnable latent vector $\boldsymbol{L}$, which serves as a concise encoding of the underlying field. It consists of a series of linear layers that project the latent vector to each hidden layer of the SIREN, enabling consistent and expressive modulation throughout the network.

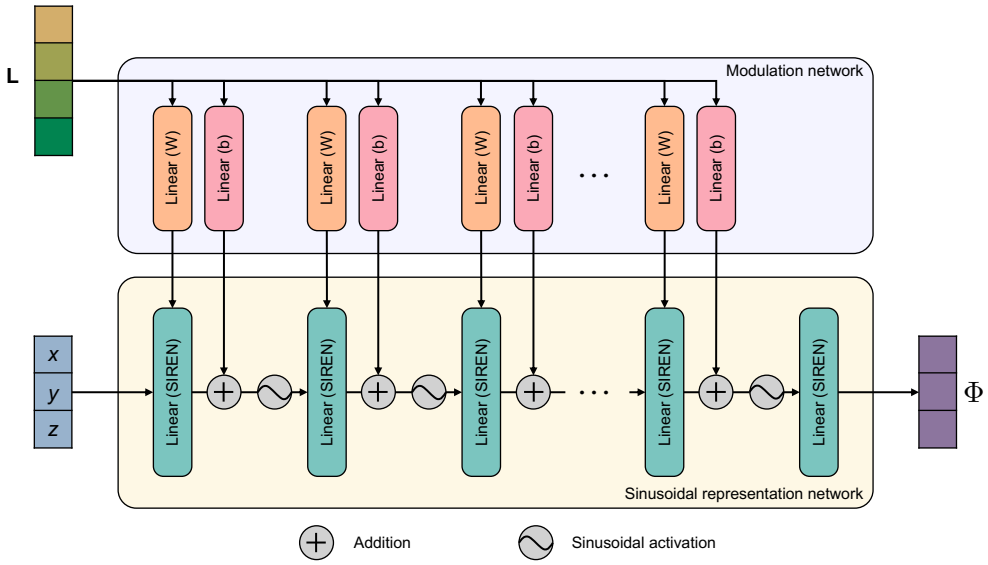

Fig. S1 Conditional neural field architecture. The upper box is the modulation network. The lower box shows the SIREN base network.

1.2 Latent diffusion model architecture

We implement the improved U-Net architecture [2] as the denoising backbone in our diffusion model. It predicts the corresponding noise component $\boldsymbol{\epsilon}_\tau$ given a noisy latent image $\boldsymbol{z}_\tau$ at diffusion timestep τ. As illustrated in Fig. S2, the U-Net is composed of hierarchical residual blocks interconnected via skip concatenations. The residual block contains two convolutional layers with group normalization and SiLU activation. Within each block, the intermediate tensor is modulated by the embedded diffusion timestep, enabling noise level conditioning throughout the network. Multi-head spatial attention blocks are incorporated at the lower-resolution stages to enhance the learning of fine-scale representations.

1.3 CoNFiLD-geo hyperparameters

The neural network and training hyperparameters of CoNFiLD-geo for the three cases are summerized in Table S1. The CNF is trained in a distributed manner across five NVIDIA RTX 4090 GPUs, while the LDM is trained on a single NVIDIA RTX 4090. Note that the hyperparameter ω_0, which controls the frequency bandwidth of the SIREN, varies across the three cases. Its value is selected via empirical tuning to accommodate the spatiotemporal complexity of each scenario. We adopt an alternating training strategy for the CNF, in which the latent vectors are updated on a per-batch basis with the CNF parameters held fixed, followed by updating the CNF parameters while keeping the latent vectors temporarily frozen. Moreover, we assign different learning rates to the CNF and the latent vectors, with the latter set to be ten times larger. This setting helps mitigate overfitting in the CNF and promotes stable convergence during training [3].

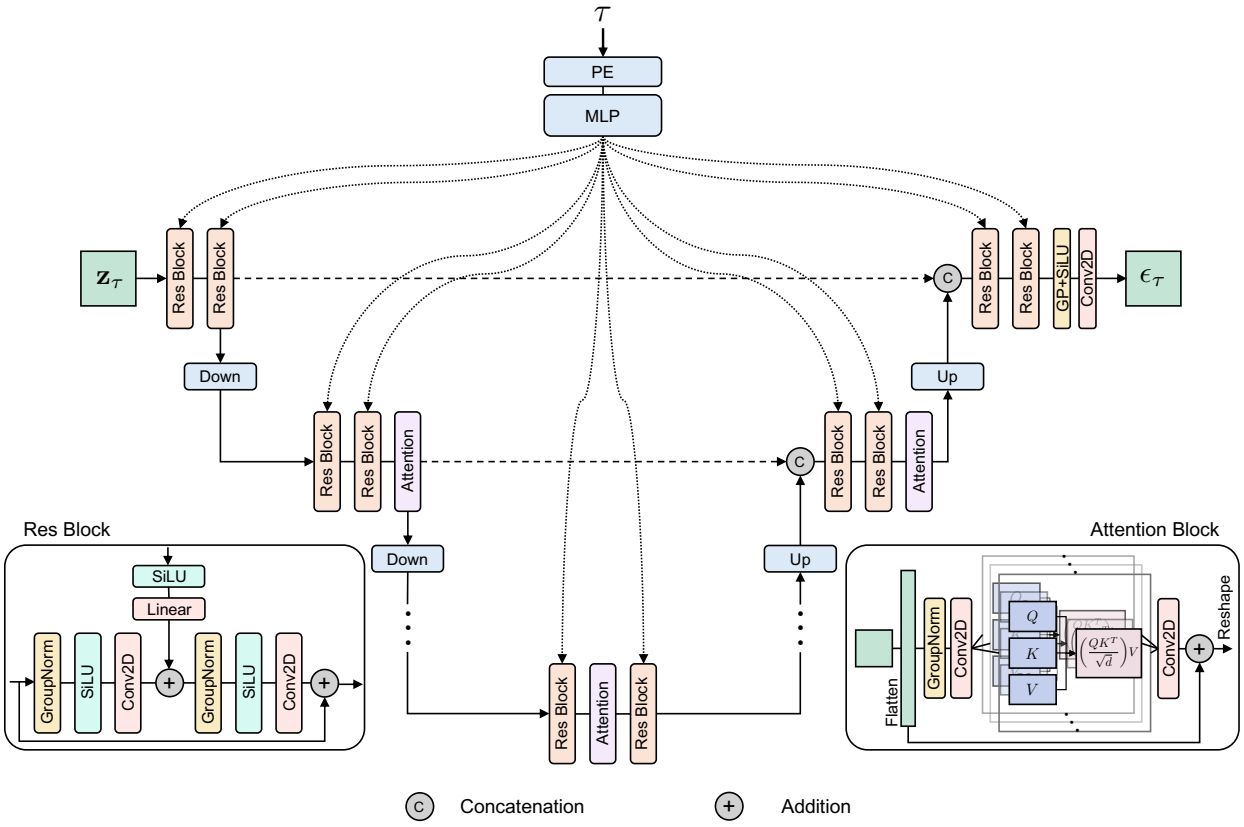

Fig. S2 U-Net architecture of the LDM. PE denotes positional embedding. MLP stands for multi-layer perceptron. Down and Up refer to downsampling and upsampling operations, respectively. The residual block and the multi-head spatial attention block are depicted in the lower left and lower right panels, respectively.

Table S1 Hyperparameters of CoNFiLD-geo. The U-Net channel configuration specifies the number of feature channels in the residual blocks across six stages of the network.

Case		2D synthetic reservoir	3D Sleipner reservoir	3D stratigraphically complex reservoir
CNF	Latent size (N_l)	256	256	384
	Hidden layer size	128		
	Number of hidden layers	5		
	SIREN ω_0	5	15	20
	Batch size	600	300	250
	Training epochs	5000		
	Latent learning rate	1×10^{-3}		
	Network learning rate	1×10^{-4}		
LDM	U-Net channels	[64,64,128,128,256,256]		
	Latent image size ($N_t \times N_l$)	64×256	128×256	128×384
	Batch size	8	32	16
	Noise schedule	Cosine		
	Diffusion steps	1000		
	Learning rate	5×10^{-5}		

2 Numerical model setup

In this section, we present the auxiliary constitutive relationships that complete the closed form of the governing equations. Detailed configurations of the numerical experiments are also provided. All simulations are performed using the ECO2N module of the finite-volume simulator TOUGH3 [4] on an Intel i9-13900K CPU workstation.

2.1 Constitutive relationships of subsurface multiphase flow

In a multiphase flow system, the capillary pressure $P_c(\leq 0)$ measures the pressure difference between two fluid phases that arises due to the interfacial tension between the fluid phases,

$$P_c = P_w - P_{nw}, \tag{1}$$

where P_w is the pressure of the wetting phase (e.g., the aqueous brine phase) and P_{nw} is the pressure of the non-wetting phase (e.g., the gaseous CO_2 phase). The van Genuchten [5] capillary pressure relationship is used in this work,

$$P_c = -P_0\left((S^*)^{-\frac{1}{\lambda}} - 1\right)^{1-\lambda}, \tag{2}$$

subject to the restriction,

$$-P_{max} \leq P_c \leq 0, \tag{3}$$

where $S^* = (S_l - S_{lr})/(1 - S_{lr})$ is the effective saturation, S_{lr} is the irreducible water saturation, P_{max} is the maximum capillary pressure, P_0 is the entry capillary pressure, and λ is the exponential constant.

Relative permeability quantifies the effective permeability of a given fluid phase in a multiphase porous medium, normalized by its permeability under single-phase flow conditions. The Corey's [6] relationship is chosen to characterize the CO_2-brine system,

$$k_{r,l} = \hat{S}^4, \tag{4}$$

$$k_{r,g} = (1 - \hat{S})^2(1 - \hat{S}^2), \tag{5}$$

where $\hat{S} = (S_l - S_{lr})/(1 - S_{lr} - S_{gr})$ is the effective saturation and S_{gr} is the residual gaseous saturation.

The above constitutive relationships, together with the governing mass conservation law, form a closed system of partial differential equations (PDEs) that can be solved numerically using methods such as the finite volume method. The specific values of the constitutive parameters for the three cases are detailed in Table S2.

Table S2 Constitutive parameter values.

Capillary pressure parameters	Value
λ	0.254
S_{lr}	0.1
P_0	19,600 Pa
P_{max}	12,500 Pa
Relative permeability parameters	**Value**
S_{lr}	0.1
S_{gr}	0.01

2.2 Case 1: CO_2 drainage in heterogeneous reservoirs

Case 1 simulates the migration of CO_2 in a synthetic heterogeneous 2D reservoir. A horizontal saline aquifer spanning 640×640 m^2 is discretized using a 64×64 uniform grid. The initial in-situ pore pressure is 12 MPa with a constant temperature of 45 °C. The salinity of the aquifer is assumed to be 15% by weight. The porosity has a constant value of 0.2. No-flux boundary conditions are imposed at the top and bottom boundaries. A constant CO_2 injection rate of 0.45 kg/s is prescribed at the left boundary, while the right boundary is set to maintain constant hydraulic pressure and saturation. The simulation spans 500 days and consists of 64 time steps. Detailed model setup is also list in Table S3.

Table S3 Numerical setup for Case 1.

Domain size	Number of mesh	Simulation time	Number of time steps	CO_2 injection rate
640 m × 640 m	64 × 64	500 days	64	0.45 kg/s

The reservoir is characterized by random log-normal permeability fields to represent a wide range of geological heterogeneity. The distribution has a mean of ln(100) (corresponding to a geometric mean of 100 mD) and a standard deviation of 1.0. An isotropic correlation length of 80 m is assumed along the horizontal directions. A total of 2,000 permeability field realizations are obtained using the Gaussian covariance model of the open-source geostatistical Python package GSTools [7].

2.3 Case 2: field-scale CO_2 sequestration at the Sleipner site

In Case 2, we present a realistic GCS scenario based on an actual field-scale project to demonstrate the capability of CoNFiLD-geo in addressing practical geo-engineering problems. This case study serves as a representative testbed for evaluating the framework's performance under realistic geological complexity and operational constraints. This subsection provides an overview of the Sleipner GCS project in Norway and details the setup of the numerical model.

Since October 1996, Statoil and its Sleipner partners have been injecting CO_2 — separated from natural gas extracted at the Sleipner field — into the Utsira Formation, a saline aquifer located at a depth of 1012 m below sea level [8]. The Utsira Formation primarily consists of highly porous, very permeable, and weakly consolidated sandstones, located at depths ranging from approximately 800 m to 1100 m. Around the injection site, the stratigraphy comprises nine sandstone units, designated Utsira L1 (the lowermost) to Utsira L9 (the uppermost), each separated by low-permeability shale layers (Fig. S3). Shale layers separating L1 through L8 are relatively thin ($\sim$ 1 m), while the shale between L8 and L9 is significantly thicker ($\sim$ 7 m). Overlying L9 is a thick shale formation with very low permeability, serving effectively as a caprock. Therefore, L9 forms a relatively isolated hydrological system from surrounding sandstone units, making it a favorable candidate for studying CO_2 migration within the reservoir.

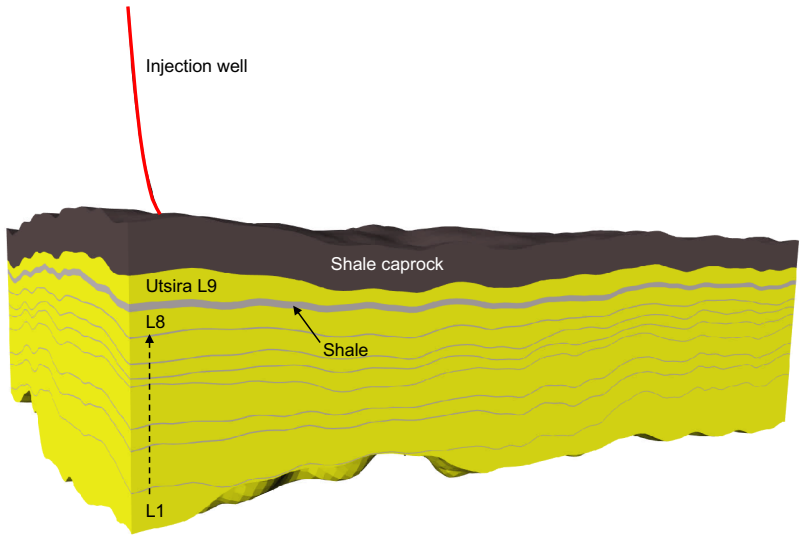

Fig. S3 The Sleipner simulation model comprising all nine sandstone layers. The red line indicates the CO_2 injection well. The target layer used in this study, Utsira L9, is vertically isolated by low-permeability shale layers. Modified from [9].

We select Utsira L9 from the Sleipner 2019 Benchmark Model [9] as the target reservoir for simulation (Fig. S3). In the horizontal plane, the reservoir extends 3.2 km in the x-direction and 5.9 km in the y-direction. The reservoir topography (i.e., thickness) is interpreted from geophysical survey data, with an average thickness of approximately 50 m. The reservoir depth ranges from -854 m to -770 m. The lateral resolution of the model grid is set to 50 × 50 m, balancing the need to resolve fine-scale characteristics of

the CO_2 plume against the computational tractability. As the horizontal extent of the domain is significantly larger than its vertical thickness, vertical CO_2 plume dynamics are expected to have minimal impact. Accordingly, the model employs a single-layer grid in the vertical direction to reduce computational complexity while retaining the essential features of lateral migration. Despite employing only a single grid layer in the vertical direction, the injected CO_2 exhibits complex migration behavior as gravity is accounted for in the simulation. This complexity results from the interplay between buoyancy and lateral driving forces (e.g., viscous and capillary forces), compounded by the structural undulations of the reservoir. The perforated injection well is consistent with the actual project configuration. The injection rate is set to a constant value of 32 kg/s, equivalent to a storage capacity of 1 Mt/year, closely matching the target capacity of the reference site. Impermeable boundary conditions are applied at the top and bottom, reflecting the sealing effect of overlying and underlying shale formations. Constant hydraulic boundary conditions are imposed on the lateral sides to emulate connectivity with an infinite-acting aquifer. An injection period of 10 years is considered, which is discretized into 128 time steps. Detailed model setup is also summarized in Table S4.

Table S4 Numerical setup for Case 2.

Lateral domain size	Number of mesh	Simulation time	Number of time steps	CO_2 injection rate
3.2 km × 5.9 km	64 × 118 × 1	10 years	128	32 kg/s

An in-situ pressure of 9 MPa is prescribed based on the reservoir depth, and pressure equilibrium is established through a preliminary simulation phase prior to injection. The reservoir temperature is constant at 32 °C. The salinity of the pore water is 33500 ppm [10]. The Utsira sandstone has a high porosity of 0.36 [9]. Permeability is modeled as a spatial random field based on the Gaussian covariance model. Permeability values are sampled from a log-normal distribution with a logarithmic mean of $\ln 1000$ and a standard deviation of 0.4. The correlation lengths are set to 500 m and 1000 m in the x and y directions, respectively

2.4 Case 3: CO_2 injection and brine production in stratigraphically complex reservoirs

Case 3 represents an extension of Case 2 toward increased geological and operational complexity, with stochastically generated reservoir structures and the simultaneous modeling of CO_2 injection and brine production. While the reservoir shares the same lateral extent as Case 2, its vertical depth and thickness are stochastically defined using geostatistical tools to account for structural uncertainty. Specifically, the reservoir depth is modeled as a Gaussian random field with a mean of -800 m and a standard deviation of 200 m. Similarly, the thickness is modeled as a Gaussian random field with a mean of 30 m and a standard deviation of 60 m. The isotropic correlation lengths for depth and thickness are set to 400 m and 500 m, respectively. The relatively large correlation length for thickness ensures that no negative values occur, even though the standard deviation exceeds the mean. To better characterize the structural complexity of stratigraphically heterogeneous reservoirs, the model domain is discretized using an unstructured triangular mesh, allowing enhanced fidelity in representing complex reservoir geometry. Note that the unstructured mesh can be naturally handled by CNF, whereas traditional convolution-based methods are limited to structured grids. The domain is discretized into 7,720 triangular elements, which offer adequate spatial resolution to resolve the dynamic behavior of CO_2 in the reservoir. In Case 3, a closed hydrological system is assumed. To maintain safe reservoir pressure, an additional brine production well is introduced. The bottomhole pressure (BHP) of the production well is controlled at 9.5 MPa for effective pressure management. The model setup is also listed in Table S5. Porosity and permeability are assumed to be constant at 0.36 and 2000 mD, respectively. All other in-situ conditions are identical to those in Case 2.

Table S5 Numerical setup for Case 3.

Lateral domain size	Number of mesh	Simulation time	Number of time steps	CO_2 injection rate	Production BHP
3.2 km × 5.9 km	7720	10 years	128	32 kg/s	9.5 MPa

3 Data restoration from incomplete observations

In this section, we demonstrate the capability of CoNFiLD-geo to perform data restoration, that is, to recover the complete field of interest from incomplete or missing observations. We consider two scenarios: (i) spatially occluded CO_2 plume monitoring data, and (ii) permeability fields with missing regions, as detailed in the following subsections. For simplicity, we adopt Case 1 as a representative test case. More complex scenarios can be seamlessly accommodated within the CoNFiLD-geo framework without requiring any structural modifications.

3.1 Spatially occluded CO_2 plume monitoring data

CO_2 plume monitoring data obtained from time-lapse seismic surveys may suffer from spatial incompleteness due to subsurface occlusions or acquisition-related issues such as equipment malfunction [11]. Such data loss can hinder accurate interpretation and prediction, necessitating reliable data restoration techniques. We showcase the novel application of CoNFiLD-geo to simultaneously infer the unknown permeability field and restore the complete CO_2 plume dynamics from spatially occluded plume monitoring data.

With the reference data shown in Fig.S4a, the damaged monitoring data is defined as removing a central subregion of the CO_2 saturation field (Fig.S4b). This degradation is mathematically formulated as a spatiotemporal masking operation. Fig. S4c presents three representative examples of the generated permeability fields along with their corresponding CO_2 saturation snapshots. CoNFiLD-geo can consistently restore the complex spatiotemporal CO_2 migration patterns within the occluded region, integrating smoothly with the surrounding observed data without noticeable discrepancies at the interfaces (Fig.S4c). The overall reconstructed CO_2 saturation fields appear physically plausible and closely match the reference, as further corroborated by the Structural Similarity Index Measure (SSIM) evaluation metric (Fig.S4e). Along the profiling line (dash-dot lines indicated in Fig.S4b), the variation in reconstructed saturation generally follows the reference trend, with slight discrepancies and elevated uncertainties within the occluded region (Fig.S4f,g). Moreover, CoNFiLD-geo is capable of inversely inferring the underlying permeability fields from the incomplete CO_2 saturation monitoring data. The high-permeability pathways are reasonably well captured by the model (Fig.S4c), showing a general agreement with the reference subsurface structure (Fig.S4d).

3.2 Permeability fields with missing region

For field-scale GCS projects, petrophysical properties, such as permeability, may be entirely unknown in certain regions due to the absence of geological surveys or the loss of historical subsurface data. This lack of information is common in areas with limited exploration, inaccessible terrains, or incomplete data archives, posing significant challenges for accurate reservoir characterization and predictive modeling. In such contexts, CoNFiLD-geo offers a promising solution for inferring missing information based on incomplete observations.

Take the same parameter-solution pair from testing set as reference data (Fig.S5a), we deliberately remove a central subregion of the permeability field and treat it as completely unknown (Fig.S5b). This spatially masked region serves as the conditioning input for CoNFiLD-geo, which aims to infer the missing permeability values and generate the corresponding spatiotemporal evolution of the CO_2 saturation field. The restored full permeability fields closely resemble the reference (Fig.S5c), with the mean probabilistic density function (PDF) of the generated samples aligning well with that of the reference (Fig.S5d). In addition, the permeability values along the profiling lines (dash-dot lines indicated in Fig. S5b) exhibit a generally consistent trend with the reference, suggesting that the model reasonably captures the spatial variation within the unknown region (Fig.S5f,g). The increase in uncertainty from the periphery to the center of the masked region (Fig.S5f,g) arises from stronger spatial covariance with the surrounding known geological information. Consequently, the uncertainty is lower near the boundaries and becomes more pronounced toward the center, where the influence of known data diminishes. The simultaneously predicted CO_2 plumes share similar overall spatiotemporal patterns with the reference, particularly in capturing the major fingering structures along the highly permeable pathway (Fig.S5a,c). The SSIM metric is initially high and gradually decreases over time as the CO_2 plume evolves and becomes more complex. Nevertheless, it remains consistently above 0.6 throughout the simulation, indicating reliable reconstruction performance (Fig.S5e).

In summary, the proposed CoNFiLD-geo framework enables flexible data restoration from either parameter or solution space via zero-shot conditional generation, with the ability to quantify associated uncertainties. This capability paves the way for broader applications in real-world subsurface modeling and monitoring tasks, particularly in scenarios involving damaged or missing data.

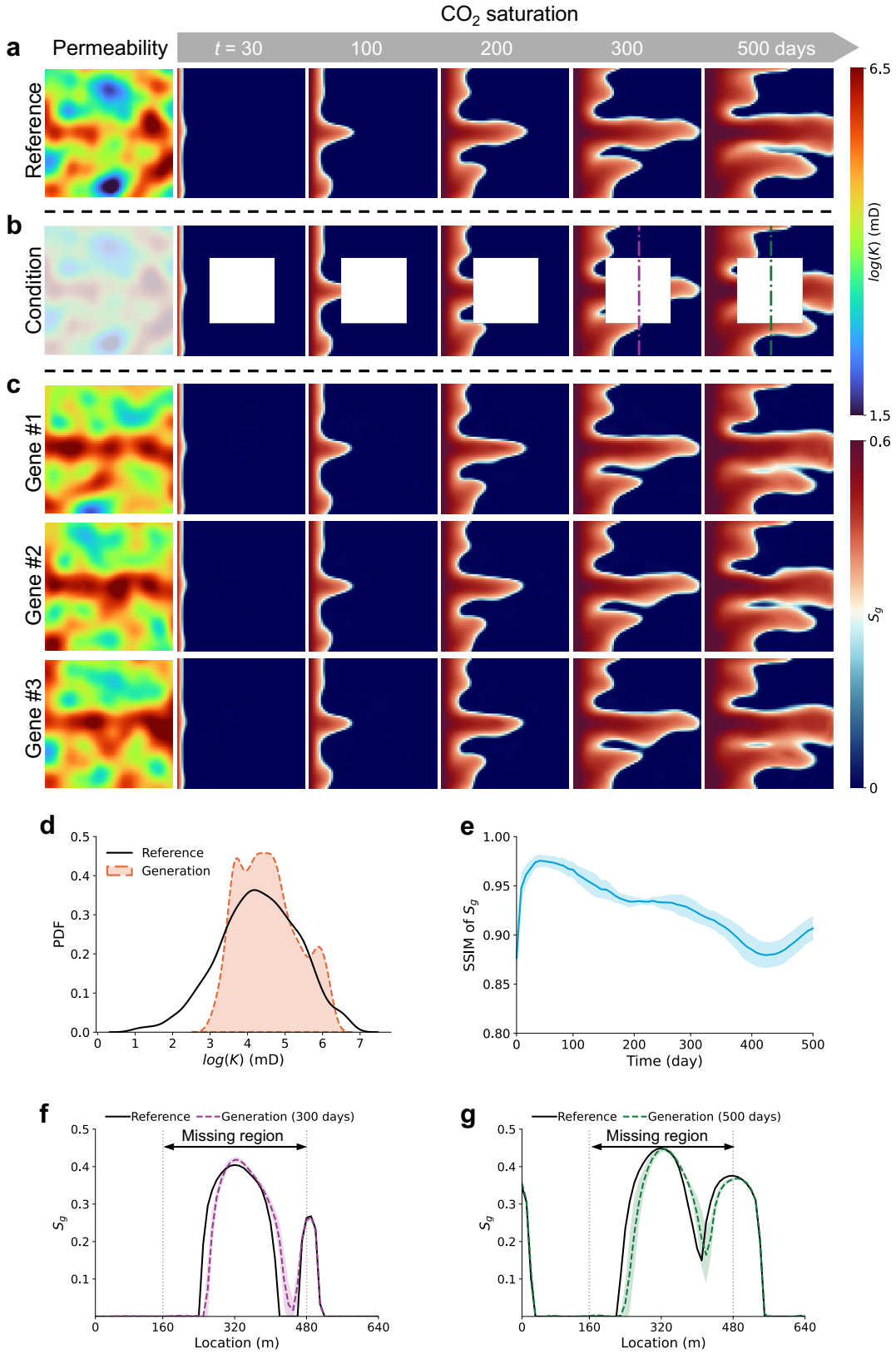

Fig. S4 Data restoration under spatially occluded CO_2 plume monitoring data. (a) Reference permeability field and corresponding CO_2 saturation trajectory with snapshots at 30, 100, 200, 300, and 500 days. (b) Spatially occluded CO_2 plume observations used as conditional inputs for CoNFiLD-geo. (c) Three representative generated permeability fields and their corresponding CO_2 saturation dynamics. (d) Probabilistic density function (PDF) plot of the reference permeability field the sample-mean generated permeability fields. (e) Temporal variation of SSIM for CO_2 saturation, with shaded regions indicating standard deviation. (f) CO_2 saturation profile along the purple dash-dot line at 300 days (see panel b), with shaded regions denoting uncertainty. (g) Same as (f) but at 500 days.

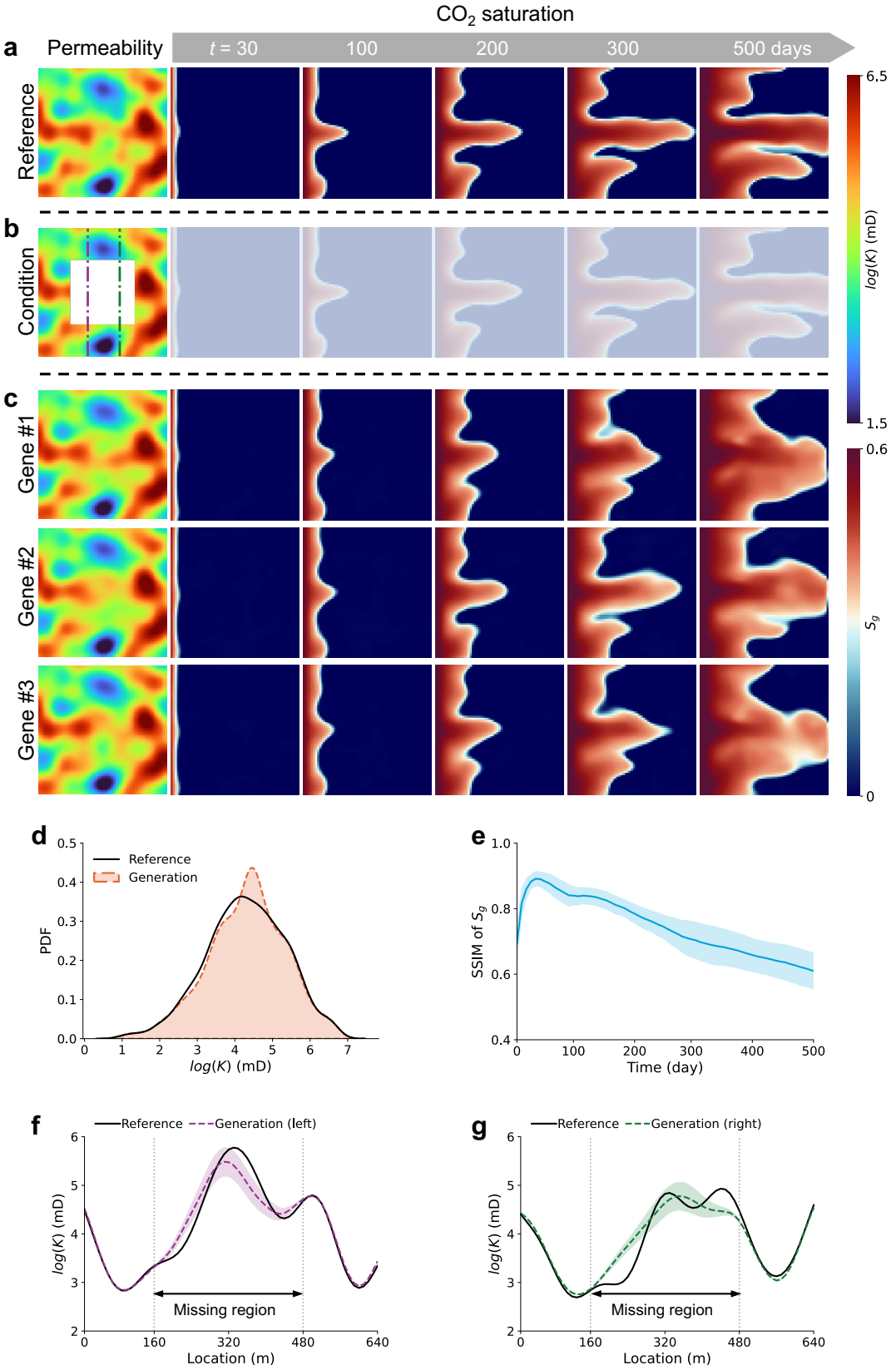

Fig. S5 Data restoration under permeability with missing region. (a) Reference permeability field and corresponding CO_2 saturation trajectory with snapshots at 30, 100, 200, 300, and 500 days. (b) Permeability field with a missing central region used as conditional inputs for CoNFiLD-geo. (c) Three representative generated permeability fields and their corresponding CO_2 saturation dynamics. (d) PDF plot of the reference permeability field the sample-mean generated permeability fields. (e) Temporal variation of SSIM for CO_2 saturation, with shaded regions indicating standard deviation. (f) Permeability profile along the purple dash-dot line on the left-hand side (see panel b), with shaded regions denoting uncertainty. (g) Same as (f) but on the right-hand side.

4 Comparison with deterministic forward models

CoNFiLD-geo, by virtue of its probabilistic nature, serves as a generative modeling framework that approximates the conditional distribution of the target field given partial observations or inputs. In contrast, deterministic models yield only single-point predictions and inherently lack the capacity to characterize uncertainty. In the context of GCS, numerous studies have explored deep learning-based surrogate models as deterministic forward solvers to replace computationally expensive numerical simulations. In this section, we compare CoNFiLD-geo with the well-established U-FNO [12] surrogate model to demonstrate CoNFiLD-geo's capability for forward modeling with quantified uncertainty. This feature is particularly important in practical scenarios, where full knowledge of input parameters is often unavailable or subject to significant uncertainty.

4.1 Baseline U-FNO model

We adopt U-FNO [12], a U-Net enhanced Fourier Neural Operator (FNO) network, as our baseline deterministic model. U-FNO integrates the global modeling capabilities of neural operators with the local feature extraction strengths of Convolutional Neural Networks (CNNs), and has demonstrated superior performance over vanilla FNO [13] and CNNs in geological carbon sequestration tasks [12, 14].

The deterministic model learns a functional mapping from the parameter space to the solution space, i.e., $\mathscr{G}_\theta : \boldsymbol{M} \in \mathbb{R}^{N_d \times N_m} \mapsto \boldsymbol{U} \in \mathbb{R}^{N_d \times N_t \times N_u}$. A U-FNO network $\mathscr{G}_\theta$ with I Fourier layers and J U-Fourier layers is formulated as,

$$\mathscr{G}_\theta = \mathscr{Q} \circ \underbrace{\sigma(\mathscr{K}_J + \mathscr{U}_J + \mathscr{W}_J) \circ \ldots \circ \sigma(\mathscr{K}_1 + \mathscr{U}_1 + \mathscr{W}_1)}_{\text{U-Fourier layers}} \circ \underbrace{\sigma(\mathscr{K}_I + \mathscr{W}_I) \circ \ldots \circ \sigma(\mathscr{K}_1 + \mathscr{W}_1)}_{\text{Fourier layers}} \circ \mathscr{P}, \tag{6}$$

where $\mathscr{P}$ and $\mathscr{Q}$ denote the linear lifting and projecting operator, respectively. In each intermediate layer, $\mathscr{K}$ denotes the kernel integral operator, $\mathscr{U}$ represents a U-Net CNN operator, $\mathscr{W}$ is a linear operator, and σ is a nonlinear activation function. The kernel integral operator applied on layer input function v_l is defined by,

$$\big(\mathscr{K}(v_l)\big)(x) \coloneqq \int \kappa(x, y) v_l(y) dy, \tag{7}$$

and is parameterized in the Fourier space [13],

$$\big(\mathscr{K}(v_l)\big)(x) = \mathscr{F}^{-1}\big(\mathscr{R} \cdot \mathscr{F}(v_l)\big)(x), \tag{8}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ represents the Fourier transformation and its inverse, respectively, and $\mathscr{R}$ is the Fourier transform of a periodic function κ.

Table S6 U-FNO model architecture. The Padding layer is used to accommodate the non-periodic boundaries. The first two channels of the output tensor denote the spatial grids along the x and y directions, respectively, the third channel denotes time dimension, the last channel denotes the feature dimension. The input features include the permeability field and the time step information.

Name	Layer	Output shape
Input	-	(64, 64, 64, 2)
Padding	Padding	(72, 72, 72, 2)
Lifting	Linear	(72, 72, 72, 36)
Fourier 1	Fourier3d/Conv1d/Add/ReLu	(72, 72, 72, 36)
Fourier 2	Fourier3d/Conv1d/Add/ReLu	(72, 72, 72, 36)
Fourier 3	Fourier3d/Conv1d/Add/ReLu	(72, 72, 72, 36)
U-Fourier 1	Fourier3d/Conv1d/UNet3d/Add/ReLu	(72, 72, 72, 36)
U-Fourier 2	Fourier3d/Conv1d/UNet3d/Add/ReLu	(72, 72, 72, 36)
U-Fourier 3	Fourier3d/Conv1d/UNet3d/Add/ReLu	(72, 72, 72, 36)
Projection 1	Linear	(72, 72, 72, 128)
Projection 2	Linear	(72, 72, 72, 1)
De-padding	-	(64, 64, 64, 1)

In this section, we utilize U-FNO as a data-driven surrogate model to approximate the spatiotemporal reservoir responses using the corresponding geological parameters. The input to U-FNO consists of the geological parameters concatenated with broadcasted time step scalars. The target output comprises N_u state variables at N_t time steps. In practice, we train N_u identical neural networks from scratch for each state variable, as proposed in Wen et al [12]. Since U-FNO is constrained to structured data formats due to its reliance on convolution operations, we train it on the first 2D case. The detailed architecture of U-FNO is illustrated in Table S6.

U-FNO is trained as a forward deterministic mapping using full observations of $\boldsymbol{M}$. Following the approach of Huang et al [15], during the testing stage, U-FNO is tasked with predicting $\boldsymbol{U}$ given either a low-resolution $\boldsymbol{M}$ or sparse measurements of $\boldsymbol{M}$, in a manner consistent with CoNFiLD-geo. We present the comparison results in the following two subsections.

4.2 Forward modeling with low-resolution permeability

We begin with predicting the CO_2 saturation based on low-resolution permeability. Results for pressure are expected to exhibit analogous trends and are therefore omitted for brevity. This synthetic test case reflects practical scenarios where permeability information is only coarsely available — for example, as interpreted from sparse monitoring wells or derived through spatial averaging. As illustrated in Fig. S6a and b, CoNFiLD-geo consistently produces results that closely match the reference solutions, whereas U-FNO exhibits significant performance degradation as input resolution decreases, occasionally generating artifacts even at early time steps. It is noteworthy that, beyond predicting CO_2 saturation, CoNFiLD-geo is also capable of generating high-resolution permeability fields inferred from low-resolution inputs. This enhanced flexibility stems from the unified treatment of inverse and forward modeling within a single generative framework. The generated high-resolution permeability fields more closely resemble the reference fields as the input resolution increases, as expected (Fig S6c). CoNFiLD-geo can fully recover the reference field using only a 16×16 resolution input, highlighting its robustness to severely under-resolved inputs (Fig S6c). The SSIM and root mean square error (RMSE) metrics for CO_2 saturation over time are presented in Fig. S6d and e. While U-FNO yields only single-point predictions, CoNFiLD-geo produces ensembles that not only align more closely with reference data but also enable explicit uncertainty quantification. CoNFiLD-geo outperforms U-FNO at most time steps, and notably achieves superior performance across the entire time horizon when the input resolution is as low as 16×16. Overall, CoNFiLD-geo demonstrates clear advantages over U-FNO in terms of robustness to coarse inputs, unified forward and inverse modeling capability, and the ability to quantify predictive uncertainty.

4.3 Forward modeling with sparse measurements of permeability

We further evaluate the forward prediction performance of both models under sparse observations of the permeability field. In Fig. S7, we present the results for cases where only a small fraction of the permeability values are known, specifically 0.05%, 1%, and 3% of the total number of grid cells. CoNFiLD-geo can simultaneously recover the full permeability field as well as the spatiotemporal dynamics of the corresponding CO_2 saturation (Fig. S7a,b), offering a flexible forward and inverse modeling capability in a unified framework. Nevertheless, U-FNO fails to generate accurate predictions under sparse observations, yielding overly smoothed and physically implausible plume migration patterns (Fig. S7a,b). The PDF plots demonstrate that CoNFiLD-geo can produce decent estimation of the full permeability field using only 3% probed data, highlighting its robustness under extremely sparse observation conditions (Fig. S7c). The error metrics further showcase the superior forward prediction accuracy of CoNFiLD-geo compared to U-FNO (Fig. S7d,e). These results underscore the advantage of probabilistic modeling over deterministic approaches in handling realistic GCS scenarios, as it enables uncertainty quantification and robust inference under sparse or low-resolution observations, which are inherent capabilities of the proposed CoNFiLD-geo framework.

4.4 Forward modeling with fully observed permeability

We also include results under the condition of full permeability field availability, although such a setting is rarely attainable in realistic GCS scenarios. It is observed that CoNFiLD-geo is capable of generating decent permeability fields and consistent accurate CO_2 saturation dynamics compared to the reference (Fig. S8a,b). In this fully observed setting, the performance of CoNFiLD-geo and U-FNO is largely comparable (Fig. S8b,c). Based on the error metrics (Fig. S8d,e), CoNFiLD-geo shows slightly lower accuracy than U-FNO, particularly at later time steps. We argue that this trade-off is acceptable, as CoNFiLD-geo offers additional capabilities beyond deterministic prediction — most notably, the ability to handle partial

observations and to quantify uncertainties. These advantages are critical in realistic GCS scenarios where full knowledge of input parameters is rarely available.

It should be noted that the CoNFiLD-geo framework could be further enhanced by incorporating physics-informed sampling strategies [15–17]. Importantly, the incorporation of PDE constraints is confined to the sampling stage, preserving the zero-shot conditional generation capability of CoNFiLD-geo without requiring additional training. In such a case, the generated parameter–solution pairs would inherently satisfy the constraints imposed by the governing PDEs, thereby enabling the model to surpass deterministic approaches in terms of physical consistency and realism. This direction will be the focus of our future research.

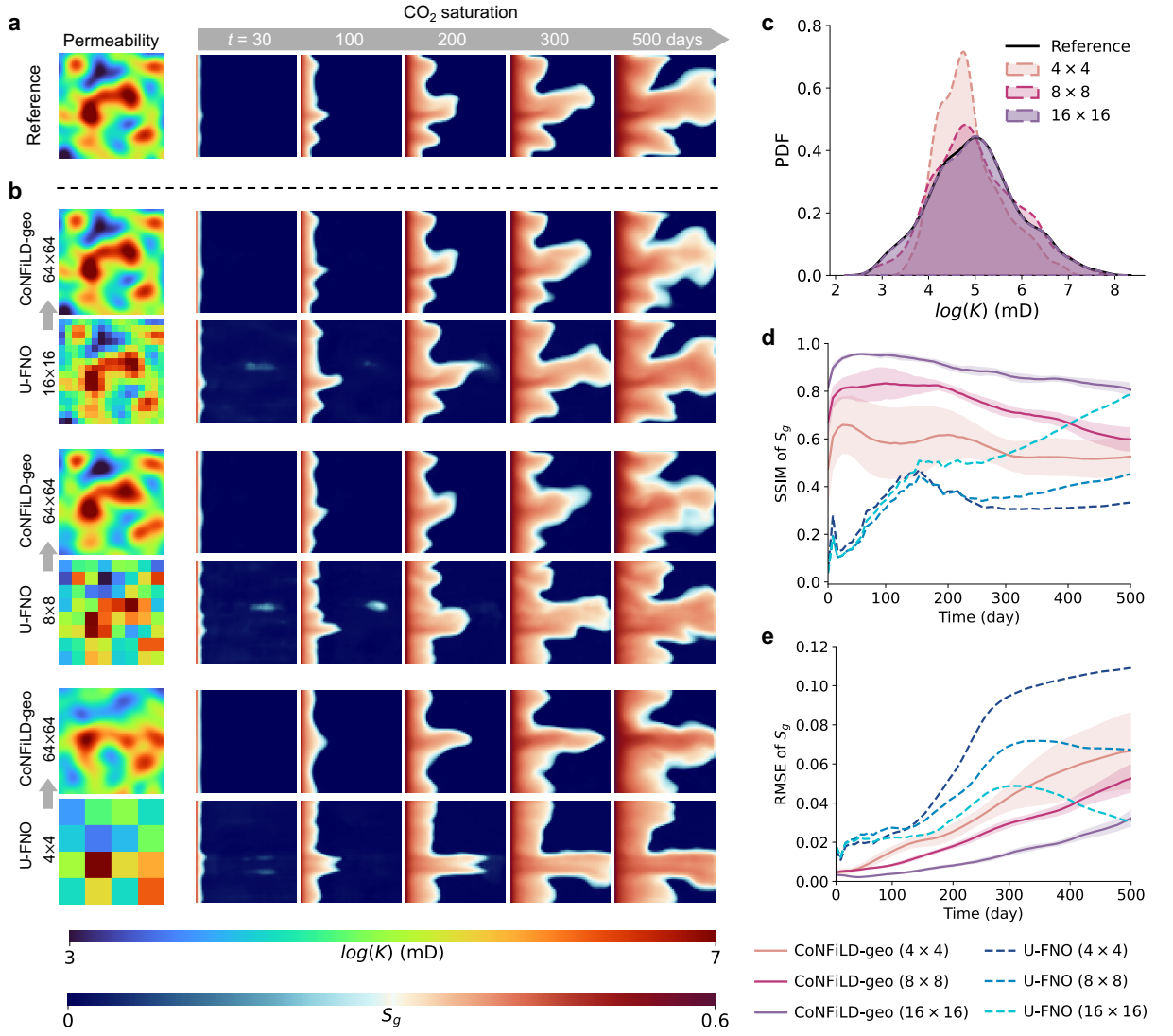

Fig. S6 Comparison of forward modeling performance between CoNFiLD-geo and U-FNO under coarsely estimated permeability fields. (a) The reference high-resolution permeability field and CO_2 saturation trajectory, with snapshots taken at 30, 100, 200, 300, and 500 days. (b) From bottom to top, the resolution of the conditioning low-resolution permeability increases from 4×4 to 16×16. For each resolution, the first row shows the CoNFiLD-geo inferred permeability and the corresponding CO_2 saturation snapshots; the second row displays the low-resolution permeability input (used as both the input to U-FNO and the conditioning guidance for CoNFiLD-geo), along with the CO_2 saturation predicted by U-FNO. (c) PDF plots comparing the reference permeability and the CoNFiLD-geo inferred permeability under varying conditioning resolutions. (d) Temporal variation of SSIM for CO_2 saturation predicted by CoNFiLD-geo and U-FNO at varying input resolutions. The shaded regions denote the standard deviation. (e) Temporal variation of RMSE for CO_2 saturation predicted by CoNFiLD-geo and U-FNO at varying input resolutions. The shaded regions denote the standard deviation.

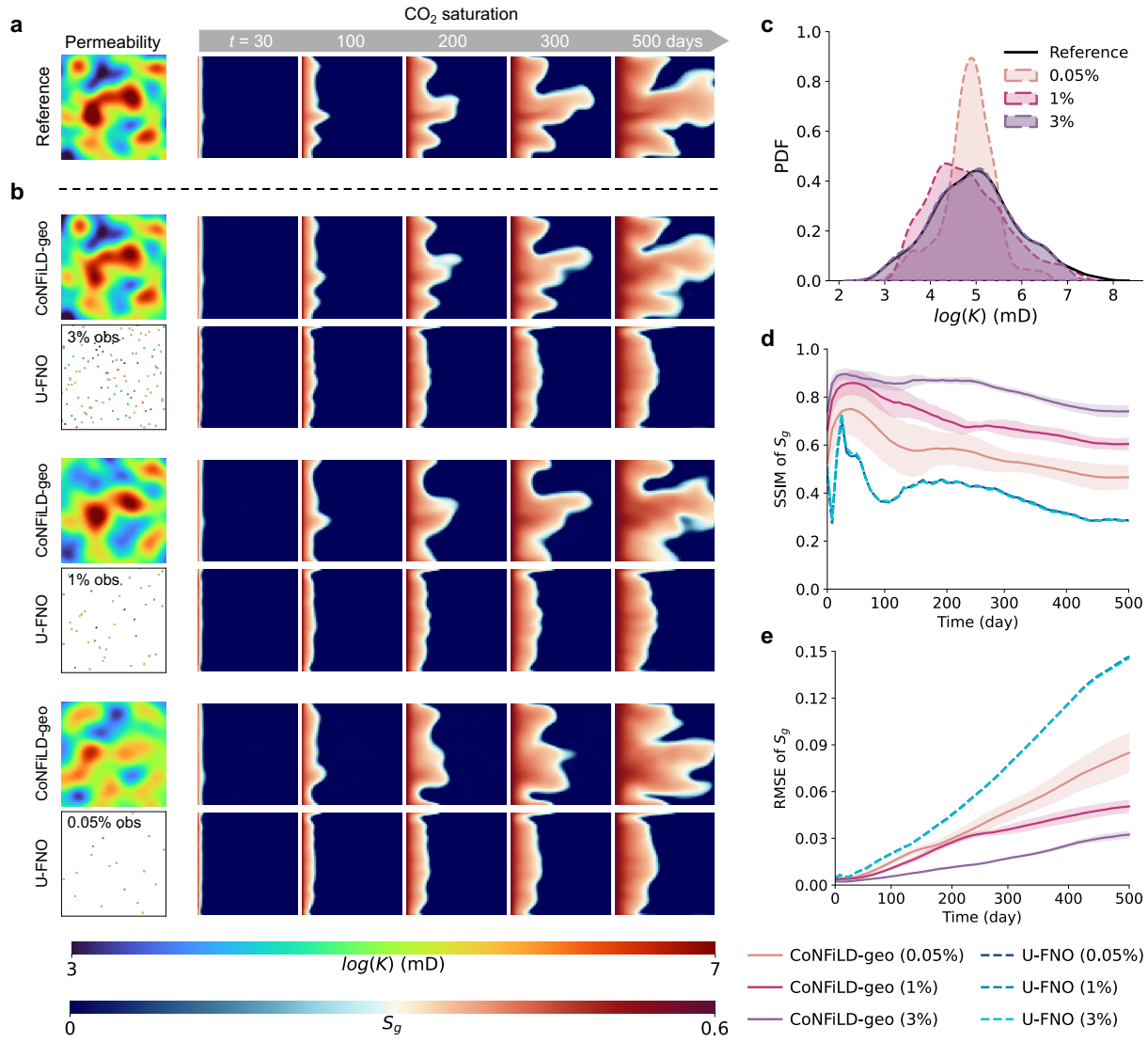

Fig. S7 Comparison of forward modeling performance between CoNFiLD-geo and U-FNO under sparsely observed permeability fields. (a) The reference full permeability field and CO_2 saturation trajectory, with snapshots taken at 30, 100, 200, 300, and 500 days. (b) From bottom to top, the known fraction of the full permeability increases from 0.05% to 3%. For each level of sparsity, the first row shows the CoNFiLD-geo reconstructed permeability and the corresponding CO_2 saturation snapshots; the second row displays the sparsely observed permeability input (used as both the input to U-FNO and the conditioning guidance for CoNFiLD-geo), along with the CO_2 saturation predicted by U-FNO. (c) PDF plots comparing the reference permeability and the CoNFiLD-geo inferred permeability under different levels of observation sparsity. (d) Temporal variation of SSIM for CO_2 saturation predicted by CoNFiLD-geo and U-FNO across varying observation sparsity. The shaded regions denote the standard deviation. (e) Temporal variation of RMSE for CO_2 saturation predicted by CoNFiLD-geo and U-FNO across varying observation sparsity. The shaded regions denote the standard deviation.

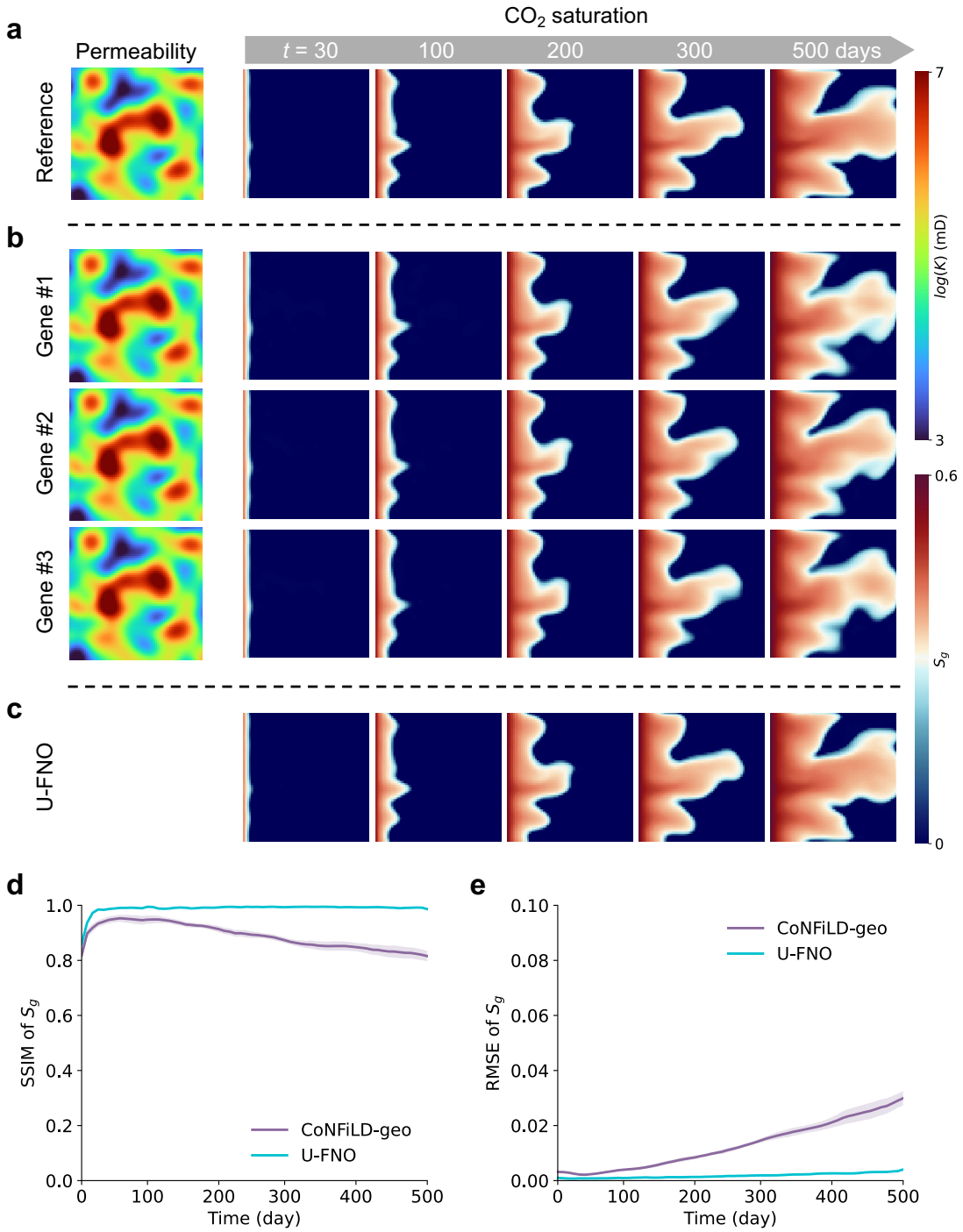

Fig. S8 Comparison of forward modeling performance between CoNFiLD-geo and U-FNO under fully observed permeability fields. (a) The reference full permeability field and CO_2 saturation trajectory, with snapshots taken at 30, 100, 200, 300, and 500 days. (b) Three demonstrating generated permeability fields and their CO_2 saturation dynamics of CoNFiLD-geo. (c) U-FNO's predicted CO_2 saturation snapshots. (d) Temporal variation of SSIM for CO_2 saturation predicted by CoNFiLD-geo and U-FNO. The shaded regions denote the standard deviation. (e) Temporal variation of RMSE for CO_2 saturation predicted by CoNFiLD-geo and U-FNO. The shaded regions denote the standard deviation.

5 Dimension reduction using Proper Orthogonal Decomposition (POD)

The proposed CoNFiLD-geo framework consists of two primary components: a dimensionality reduction module (CNF) and a probabilistic modeling module (LDM). Owing to its modular design, each component can be flexibly replaced with alternative techniques depending on the specific application or domain requirements. In our previous work, we have evaluated the performance of LDM against other generative models such as variational autoencoders (VAEs) and generative adversarial networks (GANs) [18].

For the dimensionality reduction module, a variety of alternatives exist. Classical methods such as Proper Orthogonal Decomposition (POD) [19] approximate system states using a fixed set of orthogonal basis functions weighted by projection coefficients. Learning-based approaches, including convolutional neural networks (CNNs) [20] and graph neural networks (GNNs) [21], have also been widely adopted for surrogate modeling. CNNs are effective on structured grids but are limited to regular spatial domains, while GNNs can operate on unstructured meshes but typically produce node-specific outputs and lack the ability to query continuous spatial locations.

Among these methods, POD is particularly relevant as a comparator due to its structural similarity to CNF — both reconstruct fields as combinations of basis functions and instance-dependent coefficients. In the subsequent analysis, we therefore adopt POD as the primary baseline to assess CNF's performance in representing complex spatiotemporal variability.

5.1 Formulation of POD

In CNF, the SIREN can be interpreted as a set of basis functions, while the latent vector $\boldsymbol{L}$ encodes the instance-specific coefficients. The modulation process effectively resembles a functional multiplication between the basis and the coefficients, resulting in the reconstructed field of interest. This formulation is conceptually analogous to Proper Orthogonal Decomposition (POD) [22], where a system state is approximated by a weighted sum of orthogonal spatial modes. Unlike POD, however, the basis functions in CNF are not fixed or orthogonal but are implicitly parameterized by the neural network, allowing for enhanced representational flexibility.

To obtain the basis functions (modes), we first stack the snapshots of N_s samples as,

$$\boldsymbol{\Phi}' = [\boldsymbol{\Phi}_1(\boldsymbol{x}, t_1), \boldsymbol{\Phi}_1(\boldsymbol{x}, t_2), ..., \boldsymbol{\Phi}_{N_s}(\boldsymbol{x}, t_{N_t})] \in \mathbb{R}^{N_h \times N_a}, \tag{9}$$

where $N_h = N_d \times N_\Phi$ is the full dimension size, $N_a = N_s \times N_t$ is the total number of spatiotemporal snapshots, and $N_h < N_a$ in our case. To reduce the model, a low-rank approximation with rank $N_l \ll N_h$ needs to be sought for the full-order space, $\mathcal{S} = \text{span}\{\boldsymbol{\Phi}_1(\boldsymbol{x}, t_1), \boldsymbol{\Phi}_1(\boldsymbol{x}, t_2), ..., \boldsymbol{\Phi}_{N_s}(\boldsymbol{x}, t_{N_t})\} \subset \mathbb{R}^{N_h}$.

The POD leverages the singular value decomposition (SVD) of $\boldsymbol{\Phi}'$ for dimension reduction,

$$\boldsymbol{\Phi}' = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{Q}^T, \tag{10}$$

with $\boldsymbol{P} \in \mathbb{R}^{N_h \times N_h}$ and $\boldsymbol{Q} \in \mathbb{R}^{N_a \times N_a}$ being orthogonal matrices, and $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1, \sigma_2, ..., \sigma_{N_h}\}$ containing the singular values. The column space of $\boldsymbol{P}$ spans the full-order space $\mathcal{S}$. Let $\tilde{\boldsymbol{P}} \in \mathbb{R}^{N_h \times N_l}$ denote the first N_l columns of $\boldsymbol{P}$, representing the reduced set of orthogonal bases in the the lower-dimensional subspace. Then, for a new testing snapshot $\boldsymbol{\Phi}_i^{\text{test}}$, the corresponding coefficients used to reconstruct the field are given by $\tilde{\boldsymbol{P}}^T\boldsymbol{\Phi}_i^{\text{test}}$, with the reconstruction error quantified by $\|\boldsymbol{\Phi}_i^{\text{test}} - \tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^T\boldsymbol{\Phi}_i^{\text{test}}\|^2$.

5.2 Comparison between CNF and POD

We take Case 1 as a benchmark case due to its simplicity and ease of implementation. It is worth noting that the performance gap between CNF and POD is expected to widen in Cases 2 and 3, which involve more complex spatiotemporal patterns. For a fair comparison, both CNF and POD are trained on the same training dataset and evaluated on the same testing dataset.

Fig. S9 presents the results of reconstructing a testing trajectory using CNF and POD. Panels (a) and (b) illustrate the reconstruction outcomes when the reduced dimensionality is set to be the same for both methods, i.e., $N_l^{\text{POD}} = N_l^{\text{CNF}} = 256$. The contour maps indicate that the CNF-reconstructed fields exhibit a closer agreement with the reference solution. In contrast, POD shows noticeable residual errors, particularly in the saturation variable (Fig. S9a). This is further validated in Fig. S9b, where the RMSE of POD is several orders of magnitude higher than that of CNF. In particular, the saturation RMSE increases monotonically over time, indicating cumulative reconstruction error. The poor performance of POD in this case is primarily

attributed to its limited ability to capture transient dynamics. Since POD constructs a fixed set of global basis functions via linear decomposition, it inherently lacks the flexibility to adapt to time-varying features, particularly in systems exhibiting nonlinear or nonstationary behavior. As a result, the method tends to underrepresent evolving patterns, especially during the plume development stages, where strong transients dominate the solution structure. In contrast, CNF leverages learnable, instance-specific representations, enabling it to better accommodate localized and time-dependent variations. The variation of reconstruction error (summed over all three channels) shows that POD requires more than 2048 dimensions to achieve an accuracy comparable to that of CNF (Fig. S9c). However, such high-dimensional latent representations impose a substantial computational burden, rendering them impractical for downstream modeling with LDM. The envelope of POD RMSE variation further indicates that POD remains a suboptimal choice for model reduction when considering both reconstruction fidelity and computational efficiency (Fig. S9c).

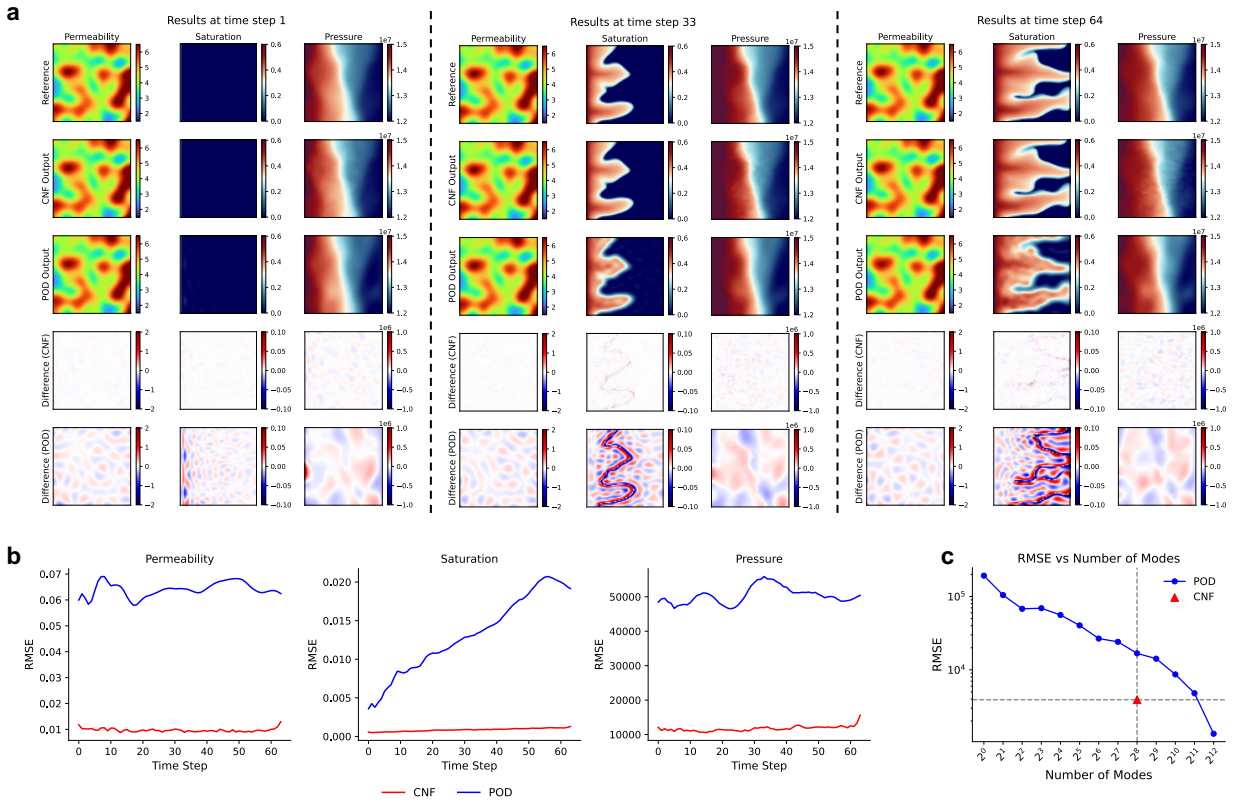

Fig. S9 Testing trajectory performance for CNF and POD. (a) Contour plots of permeability, saturation, and pressure at three selected time steps. The first row shows the reference solution, followed by reconstructions from CNF and POD in the second and third rows, respectively. The fourth and fifth rows present the corresponding residual errors for CNF and POD. (b) Temporal evolution of RMSE for CNF and POD across the three variables. (c) POD reconstruction RMSE as a function of the number of retained modes.

6 Pressure generation results

We have presented the results for pressure in this section. The contours of the generated pressure fields, along with the corresponding evaluating metrics, are elaborated in the following subsections for the three cases. These results serve as a complementary visualization to the main text.

6.1 Case 1: CO_2 drainage in heterogeneous reservoirs

For the 2D CO_2 drainage case, the pressure field exhibits relatively minor temporal variation compared to the transient dynamics of the saturation field. The pressure is elevated near the left-side injection boundary and decreases toward the right-side outflow boundary, forming a transitional gradient that aligns with the dominant direction of CO_2 migration (Fig. S10a,b). Notably, the spatial distribution of pressure is influenced by the underlying heterogeneous permeability field. CoNFiLD-geo can basically infer the pressure fields from low-resolution CO_2 plume monitoring data, with the 16×16 resolution scenario yielding the most accurate results (Fig. S10a). As the observation resolution increases, the similarity between the generated and the reference pressure field improves, and the associate uncertainty correspondingly decreases (Fig. S10c). These findings are in agreement with the saturation results discussed in the main text.

The generated pressure field under sparse well observations are illustrated in Fig. S10b and d. As the number of observation well increases, the generated fields manifest improved agreement with the reference (Fig. S10b). Correspondingly, the standard deviation of SSIM decreases (Fig. S10d), reflecting reduced uncertainty in the predictions, which also aligns with the conclusions drawn in the main text.

These results demonstrate that CoNFiLD-geo is capable of accurately generating state variables such as pressure, enabling future extensions of the framework to more complex multiphysics coupling scenarios, including thermal, mechanical and chemical fields. From the contour visualizations, minor ripple-like artifacts can be observed, which stem from the fixed frequency spectrum of the SIREN. Although these artifacts have limited influence on the overall pressure field patterns, they could potentially be mitigated through careful tuning of the activation constant ω_0.

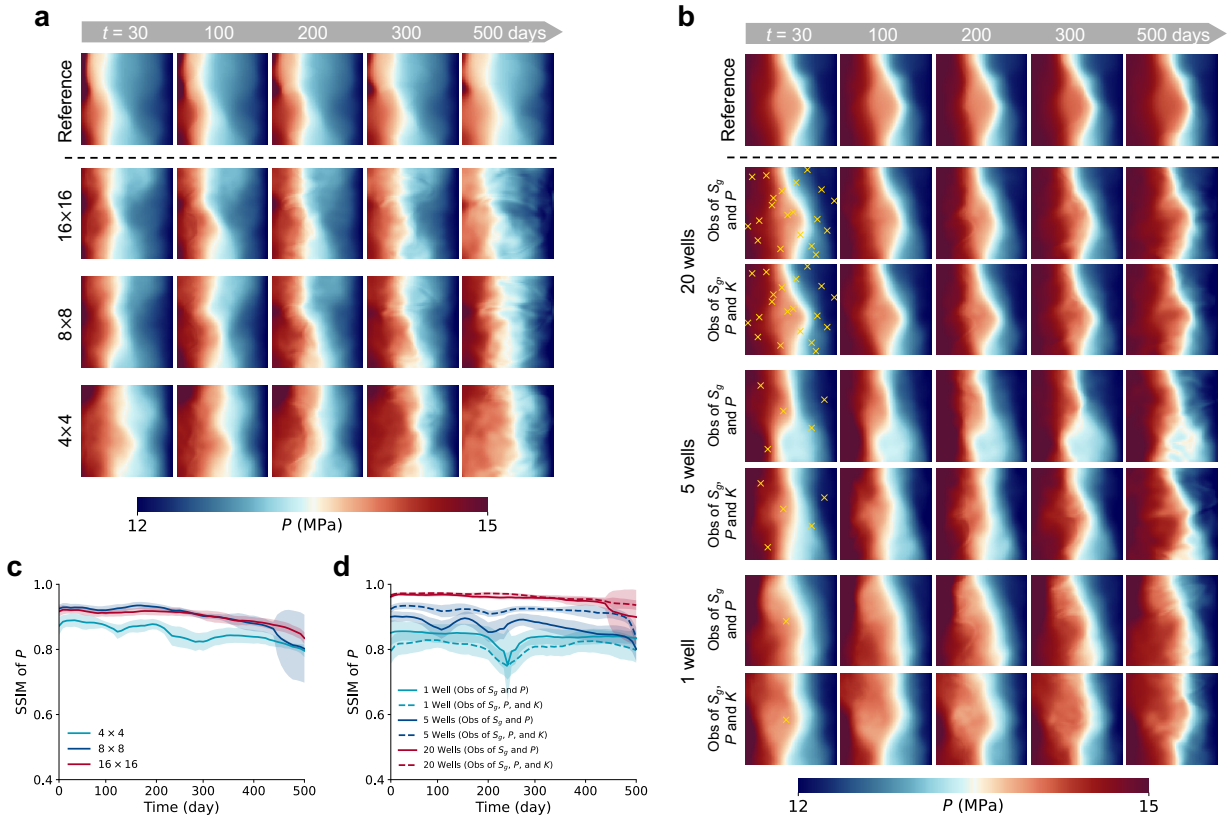

Fig. S10 Pressure generation results for Case 1. (a) Reference and generated pressure fields at 30, 100, 200, 300, and 500 days under different CO_2 resolution conditions. (b) Reference and generated pressure fields with varying numbers of monitoring wells. (c) SSIM of pressure predictions under different CO_2 resolution conditions; shaded regions indicate standard deviation. (d) SSIM of pressure predictions under varying numbers of monitoring wells; shaded regions indicate standard deviation.

6.2 Case 2: field-scale CO_2 sequestration at the Sleipner site

For the field-scale Sleipner case, elevated pressure accumulates around the injection well due to injection-induced over-pressurization (see reference contours in Fig. S11a). Over time, this overpressure gradually decreases as the effect spreads through the open boundaries, and the system transitions into a relatively stable diffusive regime. The spatial distribution of pressure is significantly affected by the realistic stratigraphic complexity and the heterogeneity of the permeability field. Among the three monitoring strategies, CO_2 plume monitoring yields the most accurate inference results, followed by the scenario with 18 monitoring wells. The case with only 6 monitoring wells exhibits the lowest accuracy and the highest uncertainty, as indicated by the broader shaded region in the RMSE plot (Fig. S11a,b). Despite these differences, CoNFiLD-geo demonstrates reliable capability in generating spatiotemporal pressure dynamics that align well with the reference, highlighting its robustness in modeling state variable evolution under various monitoring conditions. This is further substantiated by the RMSE metric, which remains consistently below 2×10^4 Pa (Fig. S11b), representing a relative error of less than 0.25% with respect to the reference pressure magnitude of approximately 9×10^6 Pa.

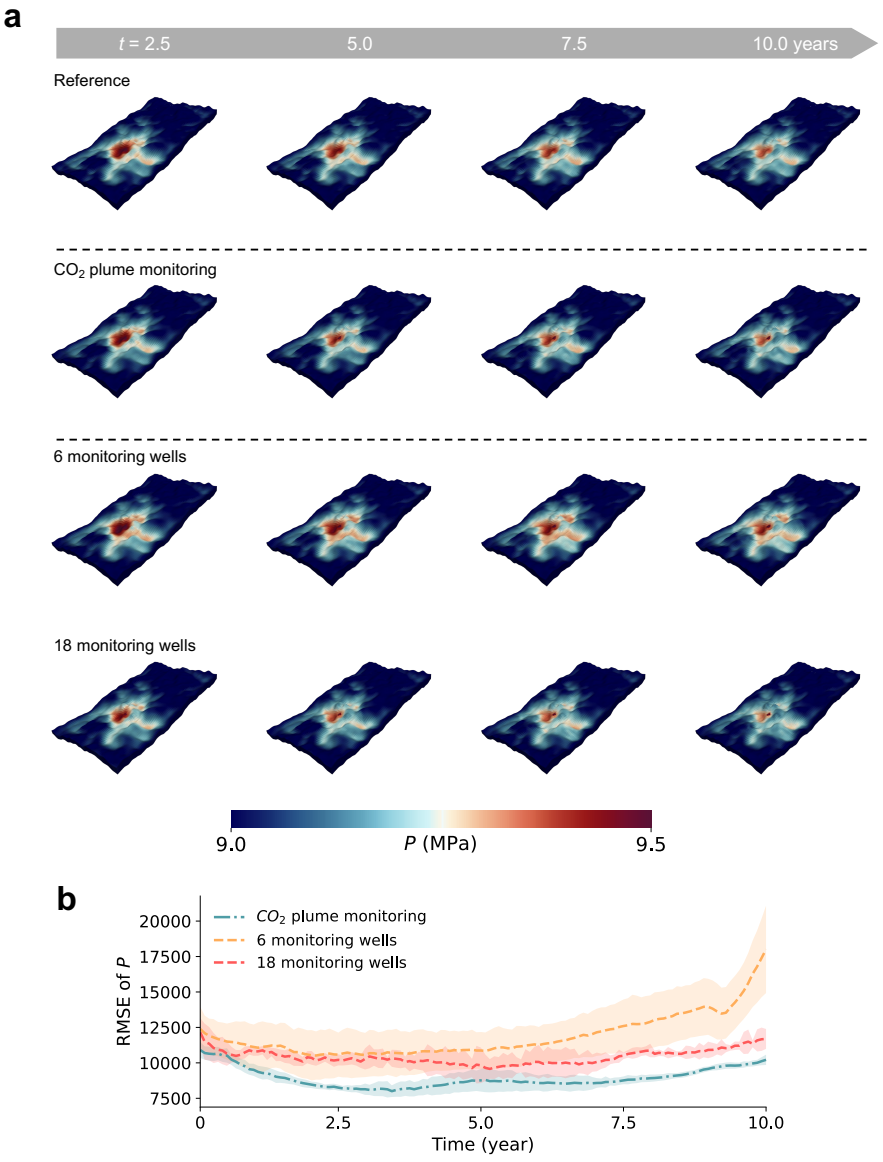

Fig. S11 Pressure generation results for Case 2. (a) Reference and generated pressure fields at 2.5, 5, 7.5 and 10 years under three monitoring strategies (conditions): CO_2 plume monitoring, 6 monitoring wells, and 18 monitoring wells. (b) Temporal variation of RMSE for the three conditions, with shaded areas indicating the standard deviation across generated samples.

6.3 Case 3: CO_2 injection and brine production in stratigraphically complex reservoirs

The third case simulates the co-existence of CO_2 injection and brine production within a closed hydrological system featuring stratigraphically complex reservoir geometry. The pressure is high near the injection well and relatively low near the production well (Fig. S12a). Owing to the closed boundary conditions, the overall level of over-pressurization is higher than that observed in Case 2. The spatial distribution of pressure is primarily governed by the reservoir geometry (i.e., depth and thickness). Again, the spatiotemporal dynamics of pressure is stable compared to that of saturation, as the system has reached a quasi-steady seepage state. Fig. S12b presents the pressure contours generated by CoNFiLD-geo under different monitoring strategies. All monitoring strategies yield results that align well with the reference, except for the CO_2 plume monitoring strategy, which shows noticeable deviations. The RMSE evaluation metric plotted in Fig. S12c also shows that this strategy produces the highest error and the largest standard deviation, indicating both lower similarity to the reference and increased uncertainty in the predicted fields. These findings suggest that relying solely on CO_2 plume monitoring may be suboptimal for pressure inference in stratigraphically complex reservoirs. Overall, CoNFoLD-geo is able to conditionally generate decent pressure fields, especially when geoseismic survey data and sparse well measurements are available. For all monitoring strategies except CO_2 plume monitoring, the RMSE remains consistently below 1×10^5 Pa (Fig. S12c), which is significantly lower than the reference pressure magnitude of around 11×10^6 Pa and corresponds to a relative error of less than 1%.

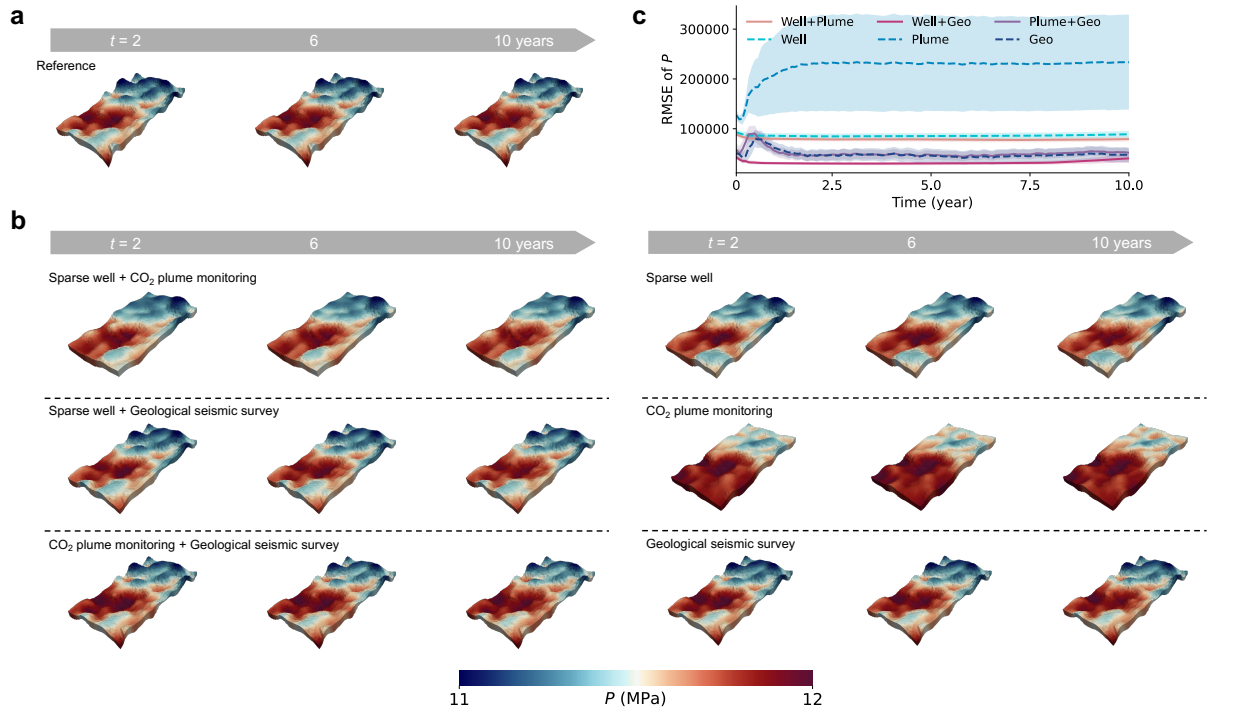

Fig. S12 Pressure generation results for Case 3. (a) Reference pressure fields at 2, 6 and 10 years. (b) Generated pressure fields under various monitoring strategies. (c) Temporal variation of RMSE for the six monitoring strategies, with shaded areas indicating the standard deviation across generated samples.

7 Additional conditional generation results

In the main text, we presented only one randomly selected sample from the generated ensemble of 10 for brevity. In this section, we provide additional visualizations from different generated samples at more time steps to facilitate further evaluation by the reader. We have also included the results of regularly placed monitoring wells for Case 1. Moreover, in the main text, only the multi-source monitoring results are presented for the scenario involving CO_2 injection and brine production in stratigraphically complex reservoirs. For completeness, the results under single-source monitoring conditions are provided in the third subsection, while the corresponding pressure results have already been presented in the previous section.

7.1 Case 1: CO_2 drainage in heterogeneous reservoirs

When monitoring wells are arranged in a regular grid-like pattern, the corresponding conditional generation results are presented in Fig. S13 and S14. The generated permeability and saturation fields show improved agreement with the reference as the number of well measurements increases, accompanied by a reduction in uncertainty. As expected, the inferred permeability fields are more accurate when direct observations of permeability values are available.

Additional conditional generation results for 16×16-resolution CO_2 plume observations and 20 sparse well measurements (with probed permeability) are shown in Fig. S15. When low-resolution CO_2 plume data are available, the inferred saturation fields closely match the reference, as the direct observations in the associated functional space effectively guide the generation. The uncertainty is minimal, evidenced by the high similarity among the three generated samples. Although the predicted pressure fields capture the global distribution well, the local details are comparatively less accurate than those of the saturation results. The generated permeability fields basically capture the structural patterns of the reference, effectively distinguishing between high- and low-permeability regions. The variations among the results reflect the uncertainty associated with the inferred permeability fields (Fig. S15a). When sparse well measurements are available, the reconstructed saturation fields generally resemble the global pattern of the reference, with minor discrepancies near the plume front. The pressure fields appear more consistent with the reference, as they exhibit smoother variations that are easier to characterize. The inferred permeability fields align well with the reference, as the direct permeability probes at sparse well locations provide critical information that enables the model to approximate the underlying heterogeneity (Fig. S15b).

7.2 Case 2: field-scale CO_2 sequestration at the Sleipner site

Fig. S16 presents the additional generation results at finer time steps conditioned on 18 monitoring wells. Both the saturation and pressure fields successfully reproduce the reference data, showcasing the model's excellent capability to reconstruct high-dimensional state variables from extremely sparse observations. The strong similarity among the generated samples indicates a low level of uncertainty in the inferred solutions. The inferred permeability fields also show close match with the reference, with both highly and lowly permeable regions accurately captured by the model. Regions lacking observation wells, such as the peripheral zones of the permeability field, exhibit increased variation/uncertainty, as expected.

7.3 Case 3: CO_2 injection and brine production in stratigraphically complex reservoirs

Fig. S17 complements Fig. 5 in the main text by providing results under single-source monitoring conditions. From a visual perspective, direct observation of the CO_2 plume yields the most accurate saturation results. In contrast, the saturation fields conditioned on "sparse well" and "geological seismic survey" data exhibit some deviations near the plume edge. The inferred reservoir geometry (depth and thickness) closely matches the reference when geological seismic survey data are available, as they serve as informative conditioning data, albeit with inherent noise. The remaining two types of conditional information deliver inferior performance. In particular, the CO_2 plume monitoring tends to produce an over-smoothed representation of the reservoir stratigraphy.

Fig. S18 provides additional generation results conditioned on both CO_2 plume monitoring and geological seismic survey data, as this combined monitoring strategy yields the most balanced and reliable outputs across different state variables. The generated geomodel and reservoir responses under this conditioning scenario are in good accordance with the reference data, with minor inter-sample variations suggesting low uncertainty. This highlights CoNFiLD-geo's effectiveness in inverse modeling of geological heterogeneity and flow behavior.

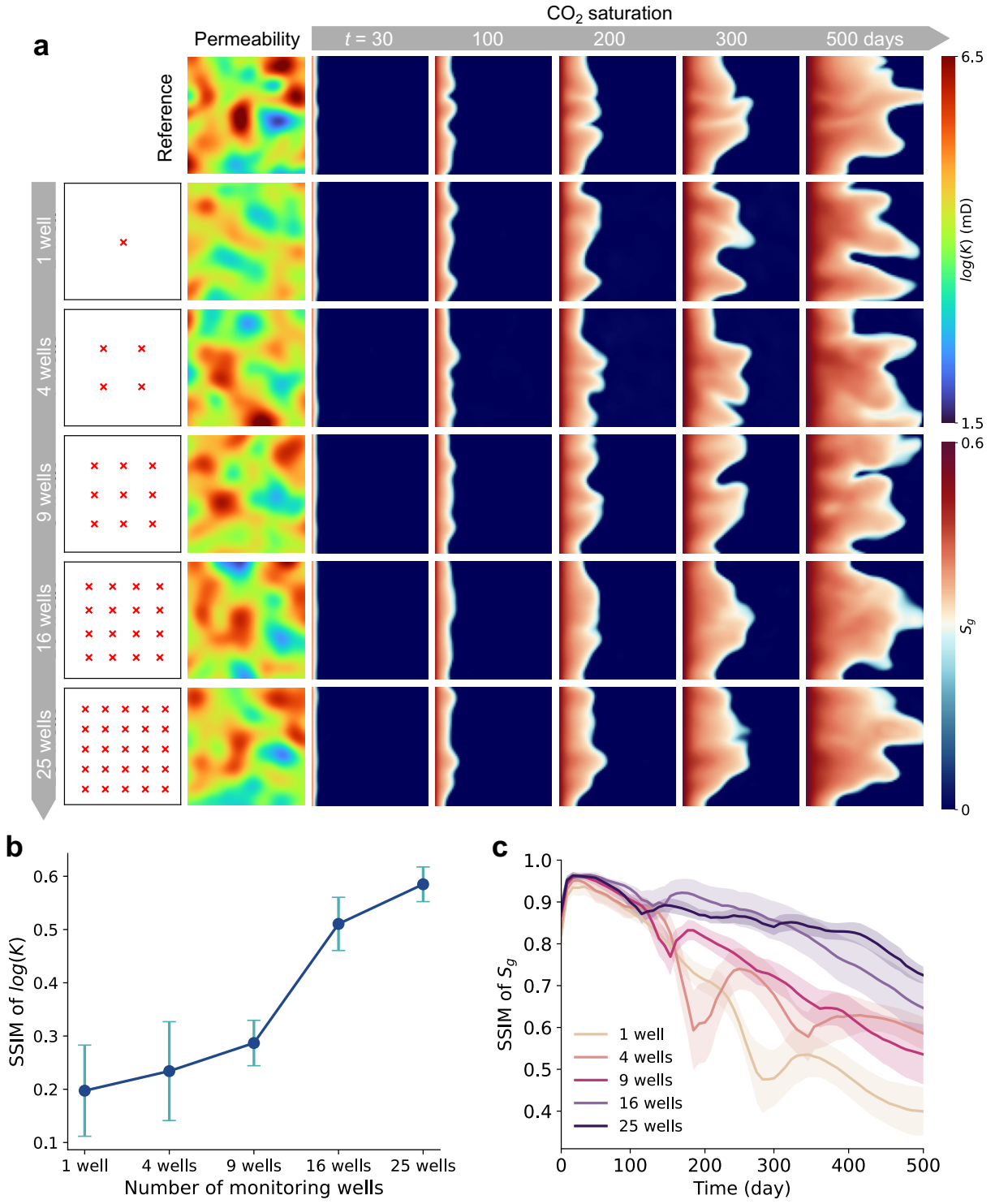

Fig. S13 Conditional generation results under regularly placed monitoring wells for Case 1. Monitoring wells provide observational data on pressure and saturation. (a) Reference and generated permeability fields, along with saturation dynamics at 30, 100, 200, 300, and 500 days. The monitoring well configurations (indicated by red crosses) are displayed in the first column. (b) SSIM of permeability as a function of the number of monitoring wells. Error bars indicate the standard deviation. (c) Temporal variation of SSIM for saturation under different numbers of monitoring wells, with shaded areas denoting the standard deviation.

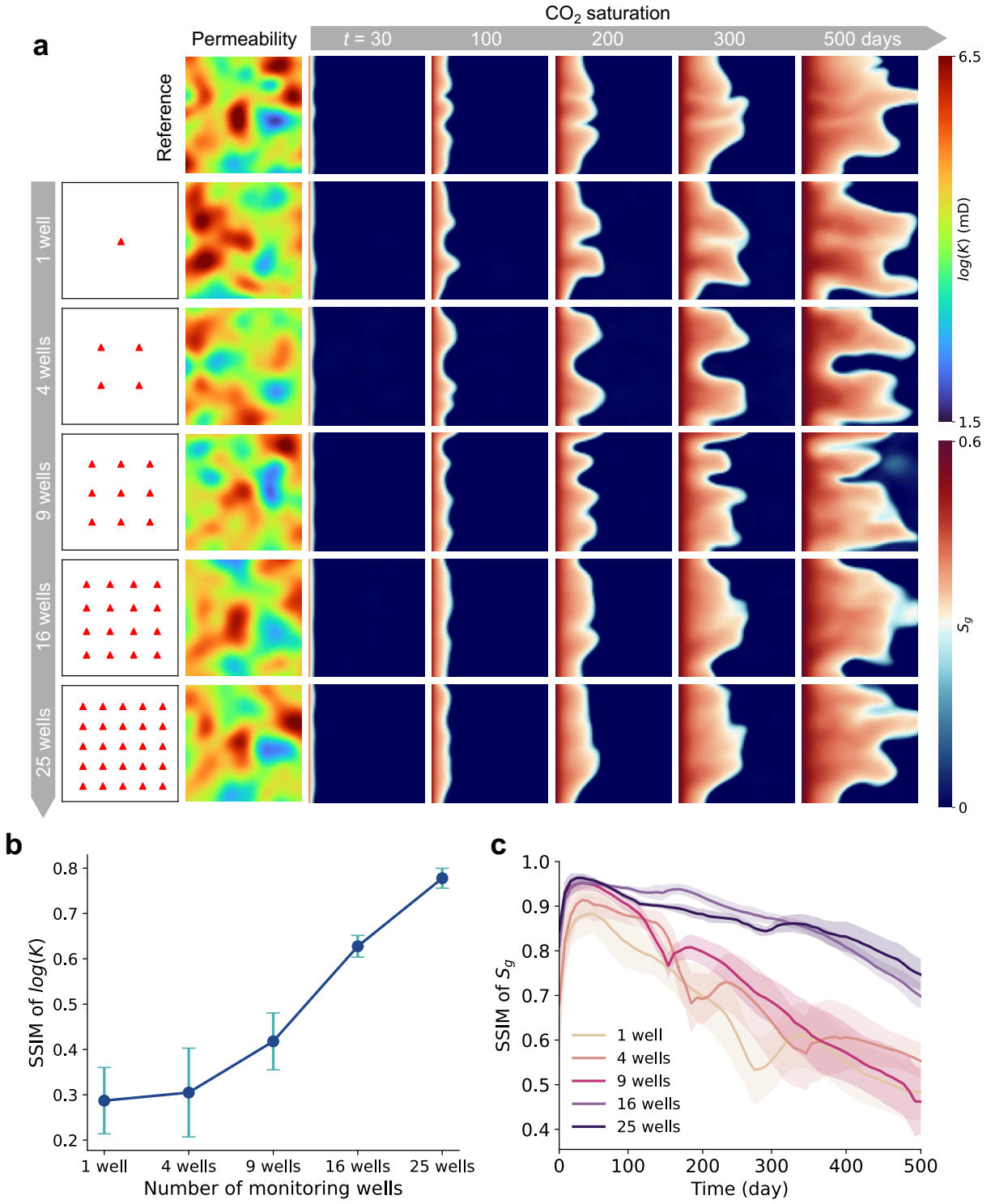

Fig. S14 Conditional generation results under regularly placed monitoring wells for Case 1. Monitoring wells provide observational data on pressure, saturation, and permeability. (a) Reference and generated permeability fields, along with saturation dynamics at 30, 100, 200, 300, and 500 days. The monitoring well configurations (indicated by red triangles) are displayed in the first column. (b) SSIM of permeability as a function of the number of monitoring wells. Error bars indicate the standard deviation. (c) Temporal variation of SSIM for saturation under different numbers of monitoring wells, with shaded areas denoting the standard deviation.

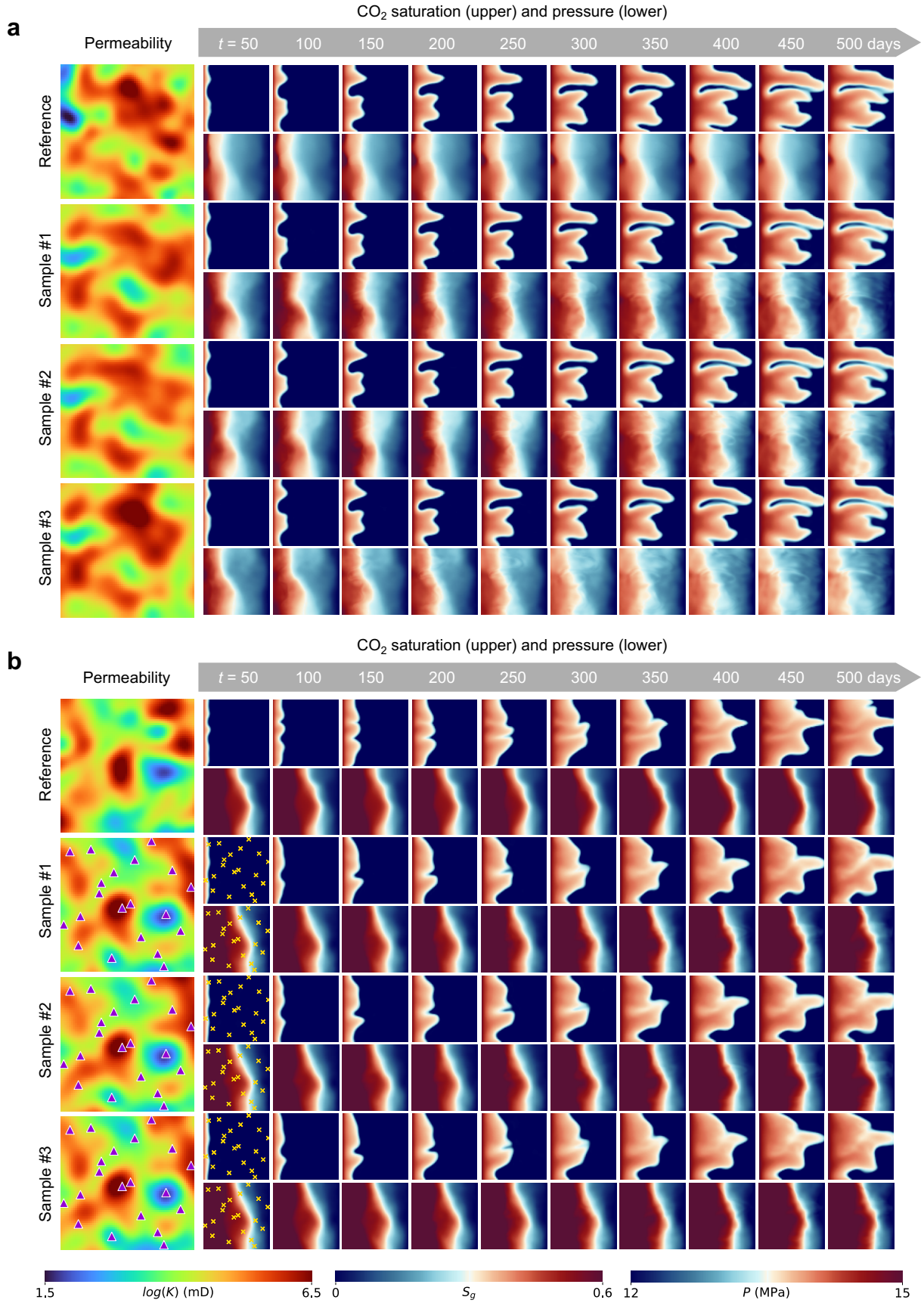

Fig. S15 Additional generation results at finer time steps for Case 1. The reference and the three randomly selected samples are shown. (a) Condition: 16×16 CO_2 plume monitoring data. (b) Condition: 20 sparse monitoring wells with permeability, pressure and saturation probed.

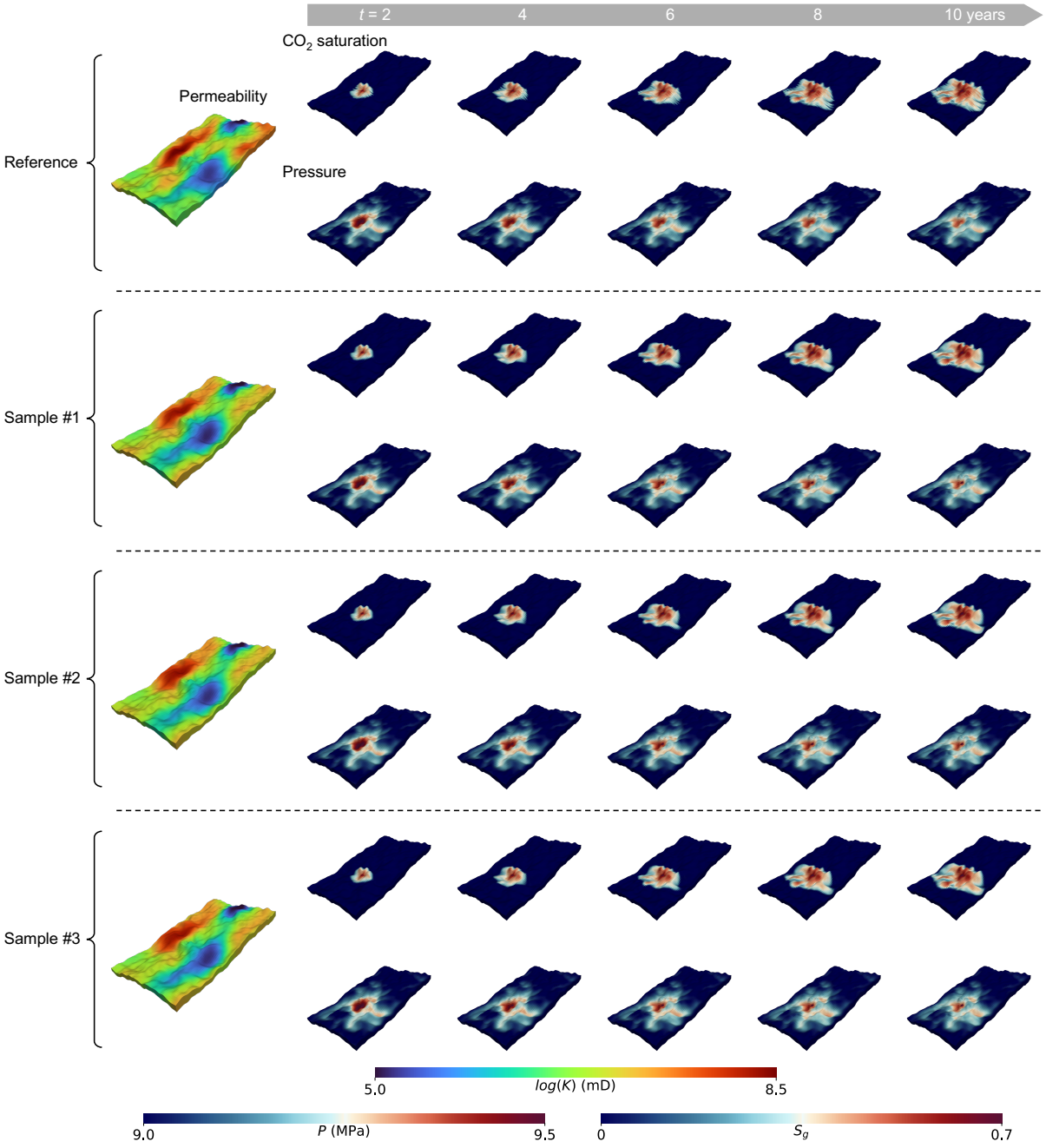

Fig. S16 Additional generation results at finer time steps for Case 2 under the condition of 18 sparse monitoring wells. From top to bottom, the reference and the three randomly selected samples from the generated ensemble are displayed.

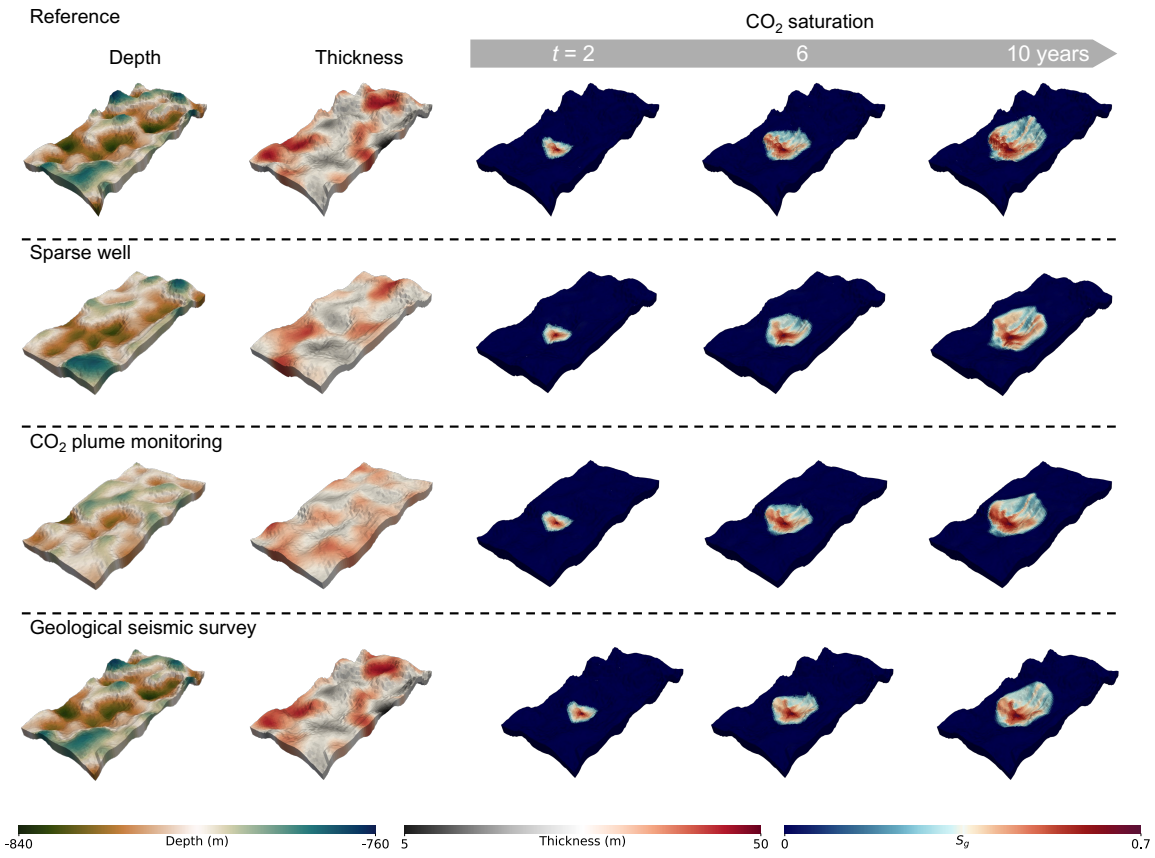

Fig. S17 Conditional generation results under single-source monitoring data. From top to bottom: the reference, results conditioned on sparse well data, results from CO_2 plume monitoring, and results from geological seismic survey.

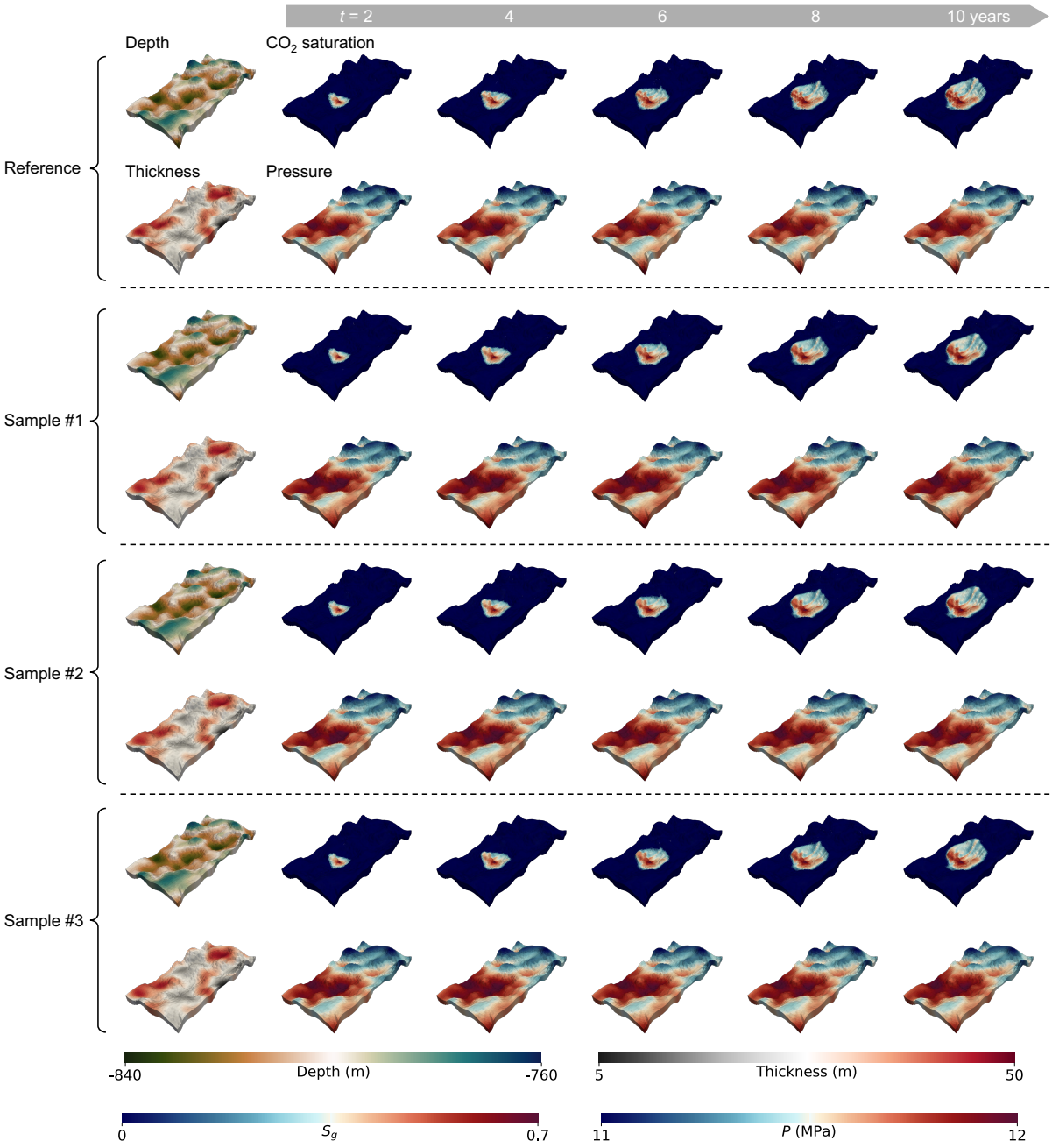

Fig. S18 Additional generation results at finer time steps for Case 3 conditioned on multi-source monitoring data consisting of 5% noisy CO_2 plume observations and geological seismic survey data. From top to bottom, the reference and the three randomly selected samples from the generated ensemble are displayed.

8 Unconditional generation results

A remarkable feature of CoNFiLD-geo is its ability to perform unconditional generation, in addition to the conditional generation explored so far. Unconditional generation requires no external observations to guide the process, allowing the model to rapidly synthesize physically plausible geomodels and corresponding reservoir responses in a fully data-driven manner. This capability effectively functions as a fast numerical emulator, enabling efficient preliminary uncertainty analysis, prior ensemble construction, and reservoir screening. These applications are also instrumental for pre-operational planning, risk assessment, and strategy development in GCS projects.

8.1 Case 1: CO_2 drainage in heterogeneous reservoirs

Fig. S19 shows the generated 4 samples starting for different Gaussian white noise. The unconditionally generated samples demonstrate a wide range of geological heterogeneity in the permeability fields, reflecting the model's ability to capture diverse subsurface scenarios. Meanwhile, the resulting saturation and pressure fields consistently exhibit physically coherent patterns, indicating the model's strong generalization capability and adherence to physical processes. Notably, the CO_2 plume migration predominantly follows the highly permeable pathways, in accordance with the transport mechanisms characteristic of subsurface multiphase flow systems.

8.2 Case 2: field-scale CO_2 sequestration at the Sleipner site

As shown in Fig. S20, CoNFiLD-geo explores a wide spectrum of permeability realizations. The generated saturation and pressure fields, while exhibiting some overall similarity, are locally influenced by the spatial heterogeneity of permeability. Their global patterns, however, are primarily governed by the complex stratigraphic structure of the reservoir. The generated parameter-solution pairs reveal physically plausible behaviors. For example, in Sample #2, the presence of a low-permeability region (blue) near the injection well limits the plume's expansion, resulting in significant overpressure buildup.

8.3 Case 3: CO_2 injection and brine production in stratigraphically complex reservoirs

Unconditional generated reservoirs and the corresponding reservoir responses are presented in Fig. S21. CoNFiLD-geo is capable of synthesizing multifarious reservoirs characterized by stratigraphic complexity and distinct geological features. The generated reservoir pairs also conform to the governing physical principles of subsurface multiphase flow. For instance, supercritical CO_2 preferentially accumulates at shallower depths (shown in green), as its buoyancy drives it to rise above denser formation fluids. In addition, thicker reservoirs offer greater storage capacity, which helps confine the plume and results in a more compact spatial distribution. These physical behaviors are effectively captured by CoNFiLD-geo.

In summary, the above unconditional generation results demonstrate that CoNFiLD-geo can efficiently synthesize diverse and physically consistent geomodels along with their corresponding dynamic responses, without relying on external observations. The generated results capture key subsurface flow behaviors, highlighting the model's strong generalization and physical awareness. Future research will explore integrating physics-informed sampling to further enhance the physical consistency of generated outputs.

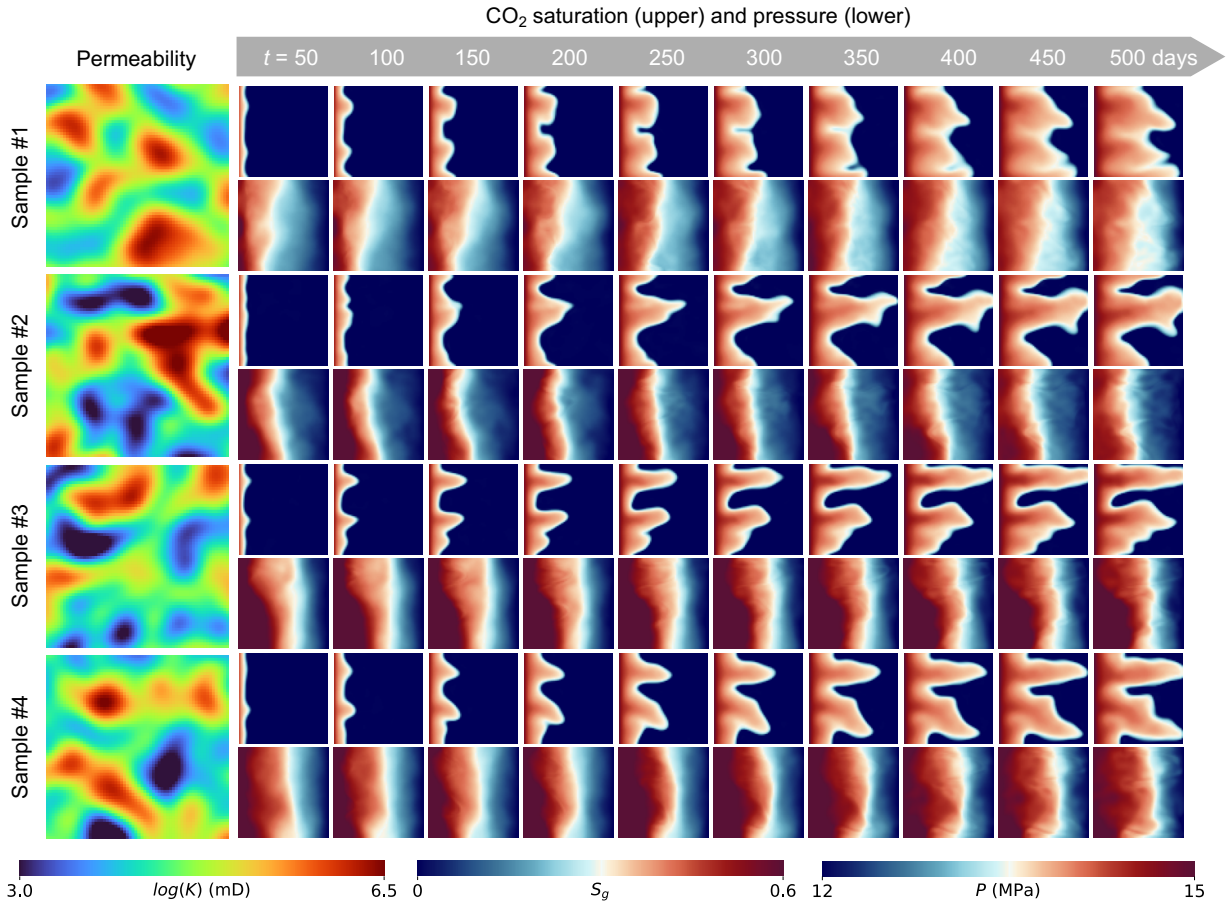

Fig. S19 Unconditional generation results for Case 1 over a 500-day period. From top to bottom, four generated samples are shown, each initiated from a different Gaussian white noise input.

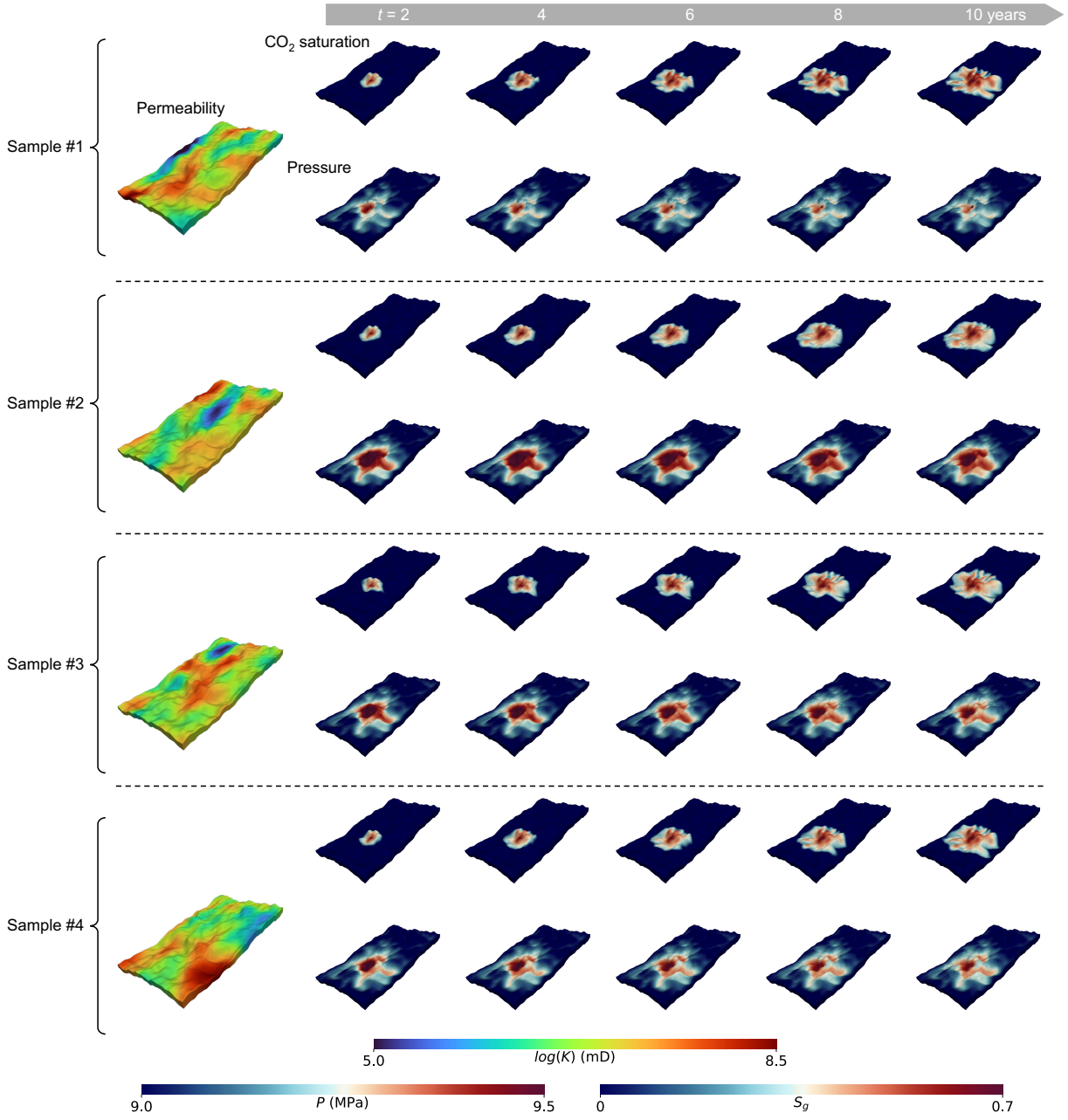

Fig. S20 Unconditional generation results for Case 2 over a 10-year period. From top to bottom, four generated samples are shown, each initiated from a different Gaussian white noise input.

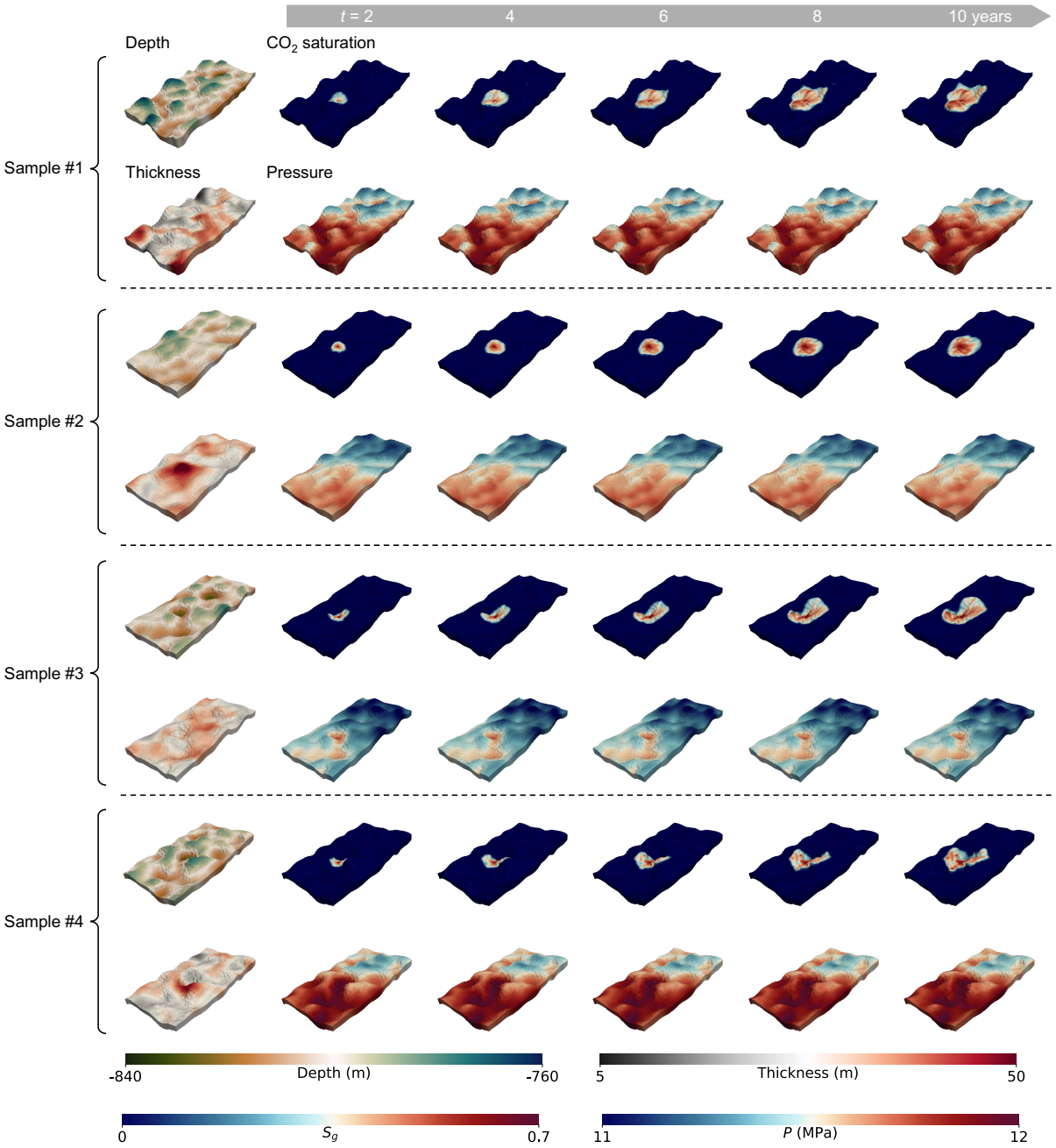

Fig. S21 Unconditional generation results for Case 3 over a 10-year period. From top to bottom, four generated samples are shown, each initiated from a different Gaussian white noise input.

9 Computational time

Table S7 presents the computational cost of CoNFiLD-geo under conditional and unconditional settings, alongside the corresponding cost of traditional numerical simulation. As a fast numerical emulator, CoNFiLD-geo is able to unconditionally generate a geomodel and its corresponding reservoir responses in approximately 20 seconds, which is substantially less than the time required by conventional numerical simulation. For conditional generation, the time cost varies depending on observational information and model complexity, ranging from 48 to 192 seconds for the three representative cases, which remains within an acceptable and efficient computational range. It should be noted that CoNFiLD-geo directly generates posterior samples via zero-shot generation, in contrast to traditional gradient-based or gradient-free methods that progressively approximate the posterior through iterative calls to a surrogate model or a numerical simulator. In addition, CoNFiLD-geo can generate samples in batch mode, facilitating the fast evaluation of uncertainties. The computational cost of numerical simulation is significantly higher, ranging from 5 to 30 minutes depending on the complexity of the case. Hence, CoNFiLD-geo opens a new pathway toward the real-time uncertainty quantification of GCS using generative AI.

Table S7 Computational time of CoNFiLD-geo for unconditional generation, conditional generation and numerical simulation. The reported conditional generation time corresponds to the following scenarios: (i) 20 sparse well measurements for Case 1; (ii) 18 sparse well measurements for Case 2; (iii) 18 sparse well measurements for Case 3.

	Case 1	Case 2	Case 3
Unconditional generation (s)	22	19	20
Conditional generation (s)	62	48	192
Numerical simulation (min)	5	18	30

References

[1] Sitzmann, V., Martel, J. N. P., Bergman, A. W., Lindell, D. B. & Wetzstein, G. Implicit Neural Representations with Periodic Activation Functions (2020). URL http://arxiv.org/abs/2006.09661. ArXiv:2006.09661 [cs].

[2] Nichol, A. & Dhariwal, P. Improved Denoising Diffusion Probabilistic Models (2021). URL http://arxiv.org/abs/2102.09672. ArXiv:2102.09672 [cs].

[3] Serrano, L. *et al.* Operator Learning with Neural Fields: Tackling PDEs on General Geometries .

[4] Jung, Y., Pau, G. S. H., Finsterle, S. & Pollyea, R. M. TOUGH3: A new efficient version of the TOUGH suite of multiphase flow and transport simulators. *Computers & Geosciences* **108**, 2–7 (2017). URL https://linkinghub.elsevier.com/retrieve/pii/S0098300416304319.

[5] van Genuchten, M. T. A closed for equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc.* **44**, 892–898 (1980).

[6] Corey, A. T. The interrelation between gas and oil relative permeabilities. *Prod. Month.* **19**, 38–41 (1954).

[7] Müller, S., Schüler, L., Zech, A. & Heße, F. GSTools v1.3: a toolbox for geostatistical modelling in Python. *Geoscientific Model Development* **15**, 3161–3182 (2022). URL https://gmd.copernicus.org/articles/15/3161/2022/. Publisher: Copernicus GmbH.

[8] Arts, R., Chadwick, A., Eiken, O., Thibeau, S. & Nooner, S. Ten years' experience of monitoring CO2 injection in the Utsira Sand at Sleipner, offshore Norway. *First Break* **26** (2008). URL https://www.earthdoc.org/content/journals/0.3997/1365-2397.26.1115.27807.

[9] Equinor. Sleipner 2019 benchmark model. https://co2datashare.org/dataset/sleipner-2019-benchmark-model (2020). DOI: 10.11582/2020.00004.

[10] Singh, V. *et al.* Reservoir Modeling of CO2 Plume Behavior Calibrated Against Monitoring Data From Sleipner, Norway (2010). URL https://onepetro.org/SPEATCE/proceedings/10ATCE/10ATCE/SPE-134891-MS/102052.

[11] Reynolds, J. M. An Introduction to Applied and Environmental Geophysics .

[12] Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A. & Benson, S. M. U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources* **163**, 104180 (2022). URL https://linkinghub.elsevier.com/retrieve/pii/S0309170822000562.

[13] Li, Z. *et al.* Fourier Neural Operator for Parametric Partial Differential Equations (2021). URL http://arxiv.org/abs/2010.08895. ArXiv:2010.08895 [cs, math].

[14] Chu, A. K., Benson, S. M. & Wen, G. Deep-Learning-Based Flow Prediction for CO2 Storage in Shale–Sandstone Formations. *Energies* **16**, 246 (2022). URL https://www.mdpi.com/1996-1073/16/1/246.

[15] Huang, J., Yang, G., Wang, Z. & Park, J. J. DiffusionPDE: Generative PDE-Solving Under Partial Observation (2024). URL http://arxiv.org/abs/2406.17763. ArXiv:2406.17763 [cs].

[16] Raissi, M., Perdikaris, P. & Karniadakis, G. 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 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S0021999118307125.

[17] Bastek, J.-H., Sun, W. & Kochmann, D. M. Physics-Informed Diffusion Models (2025). URL http://arxiv.org/abs/2403.14404. ArXiv:2403.14404 [cs].

[18] Du, P., Parikh, M. H., Fan, X., Liu, X.-Y. & Wang, J.-X. Conditional neural field latent diffusion model for generating spatiotemporal turbulence. *Nature Communications* **15**, 10416 (2024). URL https://www.nature.com/articles/s41467-024-54712-1.

[19] Guo, M. & Hesthaven, J. S. Data-driven reduced order modeling for time-dependent problems. *Computer Methods in Applied Mechanics and Engineering* **345**, 75–99 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S0045782518305334.

[20] Vlachas, P. R., Arampatzis, G., Uhler, C. & Koumoutsakos, P. Multiscale simulations of complex systems by learning their effective dynamics. *Nature Machine Intelligence* **4**, 359–366 (2022).

[21] Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).

[22] Berkooz, G., Holmes, P. & Lumley, J. The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows. *Annual Review of Fluid Mechanics* **25**, 539–575 (2003).