Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physics-Driven Hierarchical Cascade Surrogate for Compositional Simulation of Geological CO2 and H2S Storage

Stepan Zainulin

zainulin.s@gubkin.ru

National University of Oil and Gas "Gubkin University"

Anna Storozheva

National University of Oil and Gas "Gubkin University"

Research Article

Keywords: Geological CO2 storage, hierarchical cascade neural network, CO2 plume prediction, long-term blind forecast

Posted Date: July 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10080033/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Physics-Driven Hierarchical Cascade Surrogate for Compositional Simulation of Geological CO_2 and H_2S Storage

Stepan Zainulin[1,2*] and Anna Storozheva[1]

[1]Department of Offshore Oil and Gas Field Development, National University of Oil and Gas "Gubkin University", 65/1 Leninsky Prospekt, Moscow, 119991, Russia.

[2]Global Climate Stabilisation Research Department, Yu. A. Izrael Institute of Global Climate and Ecology, 20B Glebovskaya St., Moscow, 107258, Russia.

*Corresponding author(s). E-mail(s): zainulin.s@gubkin.ru;
Contributing authors: storozheva.a@gubkin.ru;

Abstract

Accurate long-term prediction of CO_2 plume migration and inter-phase compositional redistribution is a central challenge in carbon capture and storage (CCS). Existing machine-learning surrogates predict bulk phase saturations and pressure, but none resolves the full compositional state of reservoir fluids or has been evaluated on a blind centennial forecast. This study presents the Hierarchical Cascade Surrogate for CO_2 Storage (HCS-CO_2S), a block-based multi-layer perceptron surrogate trained on a compositional *tNavigator* (Rock Flow Dynamics, Moscow, Russia) simulation of CO_2/H_2S co-injection ($\mathbf{31\times31\times22}$ corner-point grid, **21,142** active cells, injection stream 0.8/0.2 mol/mol, BHP = 450 bar). The surrogate predicts, for each active cell, eight mandatory targets: reservoir pressure, three phase saturations, overall CO_2 mole fraction, dissolved-in-water CO_2, and gas- and oil-phase CO_2 mole fractions. The mapping is decomposed into five physically-ordered cascade levels mirroring the causal structure of compositional flow: a shared residual encoder, a thermobaric head, a PVT head, a gas-gated saturation head enforcing the unit-sum constraint exactly, and a composition head with hard phase masks; well-influence descriptors derived from the injection schedule provide an implicit spatial embedding for the point-wise MLP. Model performance is evaluated through a blind forecast in which the terminal step pair is excluded from training entirely, demonstrating that the hierarchical

cascade suppresses inter-level error accumulation over the centennial horizon and reproduces the spatial distribution of all compositional targets at the terminal step.

Keywords: Geological CO_2 storage, hierarchical cascade neural network, CO_2 plume prediction, long-term blind forecast

1 Introduction

Limiting global temperature rise to 2 °C requires large-scale deployment of carbon capture and storage (CCS) technology (IPCC 2022). The Intergovernmental Panel on Climate Change (IPCC) and the International Energy Agency (IEA) project that achieving net-zero emissions will require storing hundreds of billions of tonnes of CO_2 in deep geological formations (IPCC 2023; IEA et al. 2024). Deep saline aquifers and depleted hydrocarbon reservoirs offer the largest storage capacity through a combination of structural, capillary, solubility and mineral trapping (Eyitayo et al. 2025; Li et al. 2023). Reliable prediction of CO_2 plume migration, reservoir pressure evolution and storage complex integrity over planning horizons of tens to hundreds of thousands of years is therefore a primary requirement for CCS site design and regulatory certification (Miocic et al. 2019; Xiao et al. 2025).

Three-dimensional compositional simulation is the standard tool for CCS reservoir modelling. CO_2 migration in a porous medium is governed by the interplay of viscous, capillary and gravitational forces together with dissolution and mineralisation reactions (Zhong et al. 2019). Compositional simulators solve coupled nonlinear partial differential equations for mass conservation and the generalised Darcy law using finite-volume methods (Lie 2019; Voskov et al. 2017; Moortgat et al. 2012), and must resolve phase equilibrium (flash) at every grid cell and time step (Barros and Pires 2021; Zheng et al. 2023). Probabilistic risk assessment requires tens of thousands of such simulations on grids comprising millions of cells (Pawar et al. 2026), making flash calculations the dominant computational bottleneck in CCS modelling (Jing et al. 2026; Hua et al. 2024; Samnioti and Gaganis 2023).

Machine-learning (ML) surrogates have been developed to replace full-physics simulations at speedups of 10^3–10^6 (Yan et al. 2022; Wen et al. 2021a, 2023). Existing surrogates fall broadly into three architectural families (Wang et al. 2024b; Samnioti and Gaganis 2023; Du et al. 2024). The first family comprises convolutional spatial surrogates. Residual U-Net (R-U-Net) (Tang et al. 2020, 2021) demonstrated image-to-image regression for dynamic CO_2 plume fields and was extended to coupled flow-geomechanics settings (Tang et al. 2022), CO_2-EOR regimes (Meng et al. 2024), and constrained saturation forecasting (Zhang et al. 2025). Graph network surrogates extend the CNN family to arbitrary well configurations (Tang and Durlofsky 2025), and deep-learning workflows based on discrete fracture networks have been demonstrated for heterogeneous fractured reservoirs (Tariq et al. 2023b,a). Convolutional history-matching parameterisations have also achieved substantial uncertainty reduction in 3D storage systems (Han et al. 2024; Jiang and Durlofsky 2024). The second

family comprises operator-learning architectures: U-FNO (Wen et al. 2022) overcame the resolution dependency of CNNs; Nested FNO (Wen et al. 2023) enabled basin-scale 3D modelling with speedups approaching 7×10^5; and physics-informed FNO variants incorporated mass-conservation constraints for site-screening applications (Pawar et al. 2026). Multifidelity FNO extensions address near-well grid sensitivity (Ma et al. 2024). The third family encompasses encoder-based methods, including early deep neural network predictors for CO_2 plume migration (Wen et al. 2021b), convolutional autoencoder forecasting (Fan et al. 2023), spatiotemporal ED-ConvLSTM networks with transfer learning (Fan et al. 2024), VAE-based plume reconstruction from sensor data (Nagao et al. 2024), few-shot RNN surrogates (Du et al. 2026), and ConvLSTM-VQ-VAE models for CO_2 flooding (Dong et al. 2023). Physics-aware proxy models (Coutinho et al. 2021) and reinforcement learning frameworks (Chen et al. 2024) extend this family to operational optimisation. At the MLP end, point-wise surrogates for CO_2 site response (Shokouhi et al. 2021; Omosebi et al. 2022) and deep learning models for reservoir simulation with time-varying well controls (Jin et al. 2020) demonstrated that block-wise cell prediction achieves competitive accuracy at low computational cost. TransUNet-based architectures (Zhang and Li 2025) and dimensionless Transformer models (Li et al. 2026) have further extended the MLP family to spatial plume forecasting.

A parallel line of work has targeted the acceleration of thermodynamic flash calculations. Neural networks replacing stability testing and phase-split computations yield speedups of 40–800× while preserving mass-conservation (Barros and Pires 2021; Zheng et al. 2023; Hua et al. 2024; Jing et al. 2026). Physics-informed deep learning has been applied to CO_2 compositional simulation in EOR settings to improve solver convergence in near-miscible regimes (Sun et al. 2023), and ML-based compositional PVT modelling — including solution gas-oil ratio prediction — has been demonstrated with neural network and tree-based approaches (Larestani et al. 2025).

Despite this progress, four limitations constrain existing spatial surrogates:

- CNN, U-Net and GAN architectures require regular Cartesian grids (Zhong et al. 2019; Meng et al. 2024; Stepien et al. 2023). The industrial standard for reservoir modelling is stratigraphic corner-point grids, characterised by cell wedging and geometric discontinuities (Lie 2019). Applying these architectures to corner-point grids requires upscaling that reduces structural resolution and physical consistency (Wang et al. 2024a; Du et al. 2024).
- Physics-informed neural networks (PINNs) embed governing PDEs in the loss function and are mesh-free (Pour and Voskov 2023). In heterogeneous 3D formations, however, automatic differentiation at sharp saturation fronts produces singular gradients, and pointwise residual minimisation does not guarantee integral mass conservation, which leads to error accumulation over long forecast horizons (Pawar et al. 2026).
- Operator-learning architectures (FNO, U-FNO, Nested FNO) require full spatial field inputs and are therefore incompatible with corner-point grids without explicit remeshing (Wen et al. 2022; Ma et al. 2024). More fundamentally, all existing spatial surrogates predict only bulk gas saturation S_g and pressure P; none resolves the molar fractions of CO_2 in the gas phase (y_{CO_2}), the aqueous phase (x_{CO_2}), or the

overall dissolved inventory (Wen et al. 2021a, 2023; Tang et al. 2022). This prevents direct quantification of the solubility trapping fraction, which grows over decades to centuries and constitutes a substantial portion of long-term storage security (Voskov et al. 2017; Li et al. 2023).

- Point-wise MLP surrogates operate on individual cell properties and are compatible with unstructured grids (Omosebi et al. 2022; Shokouhi et al. 2021), but without geometric context the network minimises MSE by averaging over cells with similar petrophysical properties, producing spatially blurred saturation fields that fail to resolve displacement fronts (Zhang et al. 2025; Omosebi et al. 2022).

A further limitation common to all architectures is temporal extrapolation. Most published surrogates are validated only within the time window of their training data; blind forecasting at centennial scales — the horizon required for regulatory certification of storage sites (Miocic et al. 2019) — has not been demonstrated. This is particularly important because dominant trapping mechanisms operate on different timescales: structural and residual trapping dominate during injection, while solubility and mineral trapping increase progressively over decades to centuries (Eyitayo et al. 2025; Li et al. 2023). A surrogate that cannot resolve inter-phase compositional redistribution will accumulate systematic inventory errors over this horizon. The importance of compositional forecast accuracy extends to CO_2 utilisation contexts, where phase behaviour predictions directly support EOR scheme design and long-term reservoir management (Zainulin and Storozheva 2025; Du et al. 2024).

This study presents the Hierarchical Cascade Surrogate for CO_2 Storage (HCS-CO_2S), a block-based MLP surrogate for compositional CO_2/H_2S co-injection CCS modelling on corner-point grids. The model advances existing surrogates in three respects. First, it adopts a five-level physically-ordered cascade — thermobaric, PVT, saturation and composition heads conditioned sequentially on upstream outputs — that makes inter-phase CO_2 mole fractions first-class prediction targets, a capability absent from all spatial surrogates reviewed above (Wen et al. 2021a; Tang et al. 2022; Wen et al. 2022, 2023). The cascade design is informed by ML-based flash acceleration work (Barros and Pires 2021; Zheng et al. 2023; Hua et al. 2024; Jing et al. 2026), but applies the same causal logic at the level of a complete spatial surrogate. Second, the feature space is extended with dimensionless well-influence descriptors derived from the parsed injection schedule, providing an implicit spatial embedding that allows the point-wise MLP to learn large-scale plume dynamics without full-field image inputs. Third, the model is evaluated through a blind 100-year forecast in which the terminal step pair is excluded from training entirely, directly addressing the regulatory requirement for centennial storage integrity assessment (Miocic et al. 2019; IEA et al. 2024).

2 Methods

The prediction of the long-term compositional fate of co-injected CO_2 and H_2S is cast as learning a cell-wise nonlinear operator that maps a known reservoir state and its static context to the target compositional state at a future report step. The surrogate is constructed as a physically ordered *hierarchical cascade* in which each level is

conditioned on the predictions of all upstream levels, so that the data-driven model follows the causal structure of compositional flow. This section defines the simulation data and learning task (Section 2.1), the temporal decomposition and blind validation protocol (Section 2.2), the feature construction and well-influence representation (Section 2.3), the cascade architecture (Section 2.4), the training objective and physical constraints (Section 2.5), the evaluation metrics for plume migration and phase partitioning (Section 2.6), and the implementation details (Section 2.7). The end-to-end workflow is summarised in Fig. 1.

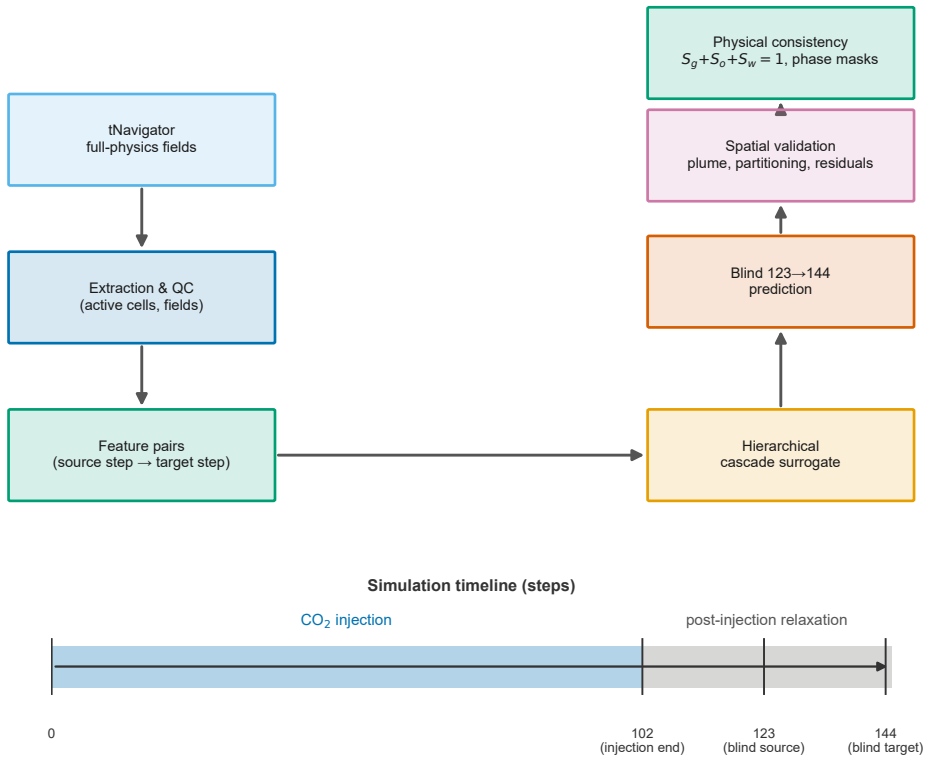


Figure 1 End-to-end data and modelling workflow. Full-physics fields produced by the compositional simulator are extracted and quality-controlled into active-cell arrays, assembled into source-to-target feature pairs, and used to train the hierarchical cascade surrogate. The trained model produces a blind prediction for the held-out step pair ($123 \rightarrow 144$), which is then assessed by spatial validation of plume extent and CO_2 partitioning and by physical-consistency checks (saturation closure $S_g + S_o + S_w = 1$ and hard phase-presence masks). The bottom timeline shows the CO_2 injection and post-injection relaxation phases and the blind validation steps.

2.1 Full-physics simulation data and learning task

The training corpus is generated by a single high-fidelity, fully implicit compositional simulation of CO_2 and H_2S co-injection into a deep saline formation performed in *tNavigator* (Rock Flow Dynamics, Moscow, Russia). The simulator resolves the coupled mass-conservation equations and the generalised Darcy law with a finite-volume discretisation, together with a cubic equation-of-state flash that partitions each component between the supercritical CO_2-rich phase and the aqueous brine at every grid cell and every time step (Lie 2019; Moortgat et al. 2012; Voskov et al. 2017). This recurrent flash calculation constitutes the dominant computational bottleneck in compositional simulation (Jing et al. 2026; Hua et al. 2024).

The reservoir is represented on a structured corner-point grid of size $31 \times 31 \times 22$ containing $N_a = 21{,}142$ active cells. The dynamic solution is recorded at $N_t = 145$ report steps. Injection of the CO_2-rich stream (molar composition 0.8 CO_2 / 0.2 H_2S) proceeds through nine injectors under bottom-hole-pressure control (BHP = 450 bar) until the injection-end step (step 102), after which the wells are shut in and the plume enters a gravitational and dissolution-driven relaxation phase. The post-injection period governs long-term containment security over horizons of 10^3 to 10^5 years (IEA et al. 2024).

The learning task is defined at the granularity of individual active cells. For a cell c and an ordered source/target step pair (t_s, t_t), the surrogate $\mathcal{F}_\theta$ with parameters θ maps a feature vector $\mathbf{x}_c(t_s, t_t)$ to the target compositional state $\hat{\mathbf{y}}_c$,

$$\hat{\mathbf{y}}_c = \mathcal{F}_\theta(\mathbf{x}_c(t_s,\, t_t))\,, \tag{1}$$

where the target vector collects the mandatory base variables and, when the corresponding fields are present in the simulator output, an optional set of thermal and PVT variables,

$$\mathbf{y}_c = \Big(\underbrace{P,\ S_g, S_o, S_w,\ z_{\mathrm{CO_2}}, R_{sw}, x_{\mathrm{CO_2}}, y_{\mathrm{CO_2}}}_{\text{eight base targets}},\ \underbrace{T, \rho_g, \rho_o, \rho_w, \mu_g, \mu_o, \mu_w}_{\text{optional PVT/thermal}}\Big). \tag{2}$$

The eight base targets are the reservoir pressure P; the gas, oil and water saturations S_g, S_o, S_w; the overall CO_2 mole fraction $z_{\mathrm{CO_2}}$; and the phase-resident CO_2 fractions in the aqueous (R_{sw}), oleic ($x_{\mathrm{CO_2}}$) and gaseous ($y_{\mathrm{CO_2}}$) phases. The optional variables — phase densities ρ_g, ρ_o, ρ_w and viscosities μ_g, μ_o, μ_w and the reservoir temperature field T — are appended to $\mathbf{y}_c$ only when a genuine simulator field is detected. The reservoir temperature field $T(\mathbf{r}, t)$ is treated as a learnable target solely in this case; it is never synthesised from depth or any other proxy variable, and it must be distinguished from the scalar critical temperatures of the components (e.g. $T_c^{\mathrm{CO_2}}$, $T_c^{\mathrm{H_2S}}$), which are thermodynamic constants internal to the equation of state and are never predicted.

2.2 Temporal decomposition and blind validation protocol

The surrogate learns a direct source-to-target operator between report steps. The available report steps are decomposed into ordered pairs (t_s, t_t) with $t_s < t_t$; each pair

defines one supervised example per active cell, in which the dynamic state at t_s is the input and the compositional state at t_t is the regression target. This formulation allows a single model to represent the strongly non-stationary behaviour of the system across both the injection and post-injection regimes, consistent with deep-learning surrogates that learn temporal evolution operators for CO_2 storage (Wen et al. 2021a, 2023; Zhang et al. 2025).

A *blind validation protocol* is adopted to obtain an unbiased estimate of long-horizon predictive skill. The terminal step pair ($t_s = 123$, $t_t = 144$) is removed in its entirety from the training set, so that neither the source nor the target step of this transition appears during optimisation. The surrogate must therefore extrapolate to the deep future of the storage operation from intermediate dynamics only. This held-out transition probes the long-term integrity assessment requirement for geological storage sites (Miocic et al. 2019; IEA et al. 2024) and provides the basis for the spatial validation reported in Section 2.6.

2.3 Feature construction and well-influence representation

The input feature vector combines three blocks: the static reservoir context, the dynamic source-step state, and a set of well-influence descriptors,

$$\mathbf{x}_c = \Big[\underbrace{\mathbf{s}_c}_{\text{static}}, \underbrace{\mathbf{d}_c(t_s)}_{\text{dynamic}}, \underbrace{\mathbf{w}_c}_{\text{wells}} \Big]. \tag{3}$$

The static block $\mathbf{s}_c$ encodes petrophysical flow-and-storage properties together with cell geometry: porosity ϕ, directional permeabilities as $\log_{10} k$, pore volume, depth, cell dimensions and normalised cell coordinates. These properties control storativity and preferential migration pathways of the buoyant plume. The dynamic block $\mathbf{d}_c(t_s)$ carries the reservoir state at the source step and, optionally, short temporal lags.

Because a point-wise multilayer perceptron has no intrinsic awareness of the spatial arrangement of the injectors, the feature vector is extended with a compact set of dimensionless well-influence descriptors $\mathbf{w}_c$ that introduce near-wellbore spatial context explicitly. The injectors are recovered from the field schedule (trajectories, completed intervals and control keywords) by a self-contained parser, and the resulting geometry is converted into features normalised by the reservoir diagonal L, independent of grid resolution. Writing $d_{cw} = \|\mathbf{r}_c - \mathbf{r}_w\|/L$ for the normalised distance from cell c to injector w, the well-influence vector contains

$$\begin{aligned}
d_c^{\min} &= \min_w d_{cw}, && \text{(distance to nearest injector)} \\
I_c^{\mathrm{G}} &= \sum_w \exp\big[-(d_{cw}/\sigma)^2\big], && \text{(Gaussian influence)} \\
I_c^{\mathrm{inv}} &= \sum_w \frac{1}{d_{cw} + \varepsilon}, && \text{(inverse-distance influence)} \\
n_c^{\mathrm{near}} &= \#\{w : d_{cw} < r_{\mathrm{near}}\}, && \text{(injector density)}
\end{aligned} \tag{4}$$

together with the vertical overlap between the cell and the perforated interval, a dimensionless depth-to-completion, the first few explicit per-well distances, and two operational flags encoding the injection schedule: an indicator of active injection $\mathbb{1}[t_s \leq 102]$ and a normalised post-shut-in relaxation timer $\max(0,\, t_s - 102)/N_t$. When cell coordinates are unavailable the builder falls back to normalised grid indices $(i, j, k)/(n_x, n_y, n_z)$, preserving resolution-independence.

2.4 Hierarchical cascade surrogate architecture

The operator $\mathcal{F}_\theta$ in Eq. (1) is decomposed into five physically ordered levels (Fig. 2) in which each downstream level is conditioned on the outputs of all upstream levels. This ordering follows the causal structure of sequential compositional flow solvers: the thermobaric state sets fluid properties, fluid properties and thermobaric state drive phase saturations, and the phase split constrains component mole fractions. Physics-guided architectural constraints of this kind have been advocated for subsurface ML surrogates to improve physical consistency and data efficiency (Shokouhi et al. 2021; Omosebi et al. 2022).

Level 1 — Static and state encoder. A shared residual multilayer-perceptron encoder E_θ compresses the concatenated feature vector of Eq. (3) into a latent state $\mathbf{h}_c \in \mathbb{R}^d$,

$$\mathbf{h}_c = E_\theta(\mathbf{x}_c). \tag{5}$$

This latent state summarises the static petrophysical context, the well-influence geometry and the source-step dynamics, and is supplied to every downstream head.

Level 2 — Thermobaric head. Pressure is predicted first, conditioned on the latent state alone (Lie 2019),

$$\hat{P}_c = g_\theta^P(\mathbf{h}_c), \qquad \hat{T}_c = g_\theta^T(\mathbf{h}_c) \;\; \text{(thermal mode only)}. \tag{6}$$

The temperature head g_θ^T is instantiated only when a genuine spatially distributed temperature field is detected; otherwise the model operates in isothermal mode and predicts pressure alone. The thermobaric vector $\boldsymbol{\tau}_c = [\hat{P}_c]$ or $[\hat{P}_c, \hat{T}_c]$ is passed downstream.

Level 3 — PVT head. Conditioned on the latent state and the thermobaric vector, the model predicts the phase densities and viscosities,

$$\hat{\boldsymbol{\pi}}_c = g_\theta^{\mathrm{PVT}}\big([\,\mathbf{h}_c,\, \boldsymbol{\tau}_c\,]\big), \qquad \hat{\boldsymbol{\pi}}_c = (\hat{\rho}_g, \hat{\rho}_o, \hat{\rho}_w, \hat{\mu}_g, \hat{\mu}_o, \hat{\mu}_w). \tag{7}$$

This head is present only when PVT fields are available in the simulator output.

Level 4 — Saturation head. The three phase saturations are produced conditional on $[\mathbf{h}_c, \boldsymbol{\tau}_c, \hat{\boldsymbol{\pi}}_c]$ through a gas-gated construction that enforces the unit-sum constraint exactly. The network emits a gate logit and an amplitude logit for the non-wetting gas phase, and the residual liquid volume is split between the oleic and aqueous phases by a softmax,

$$\hat{S}_g = \sigma(\ell_c^g)\,\sigma(\ell_c^a), \qquad (\hat{f}_o, \hat{f}_w) = \mathrm{softmax}(\ell_c^o, \ell_c^w), \tag{8}$$

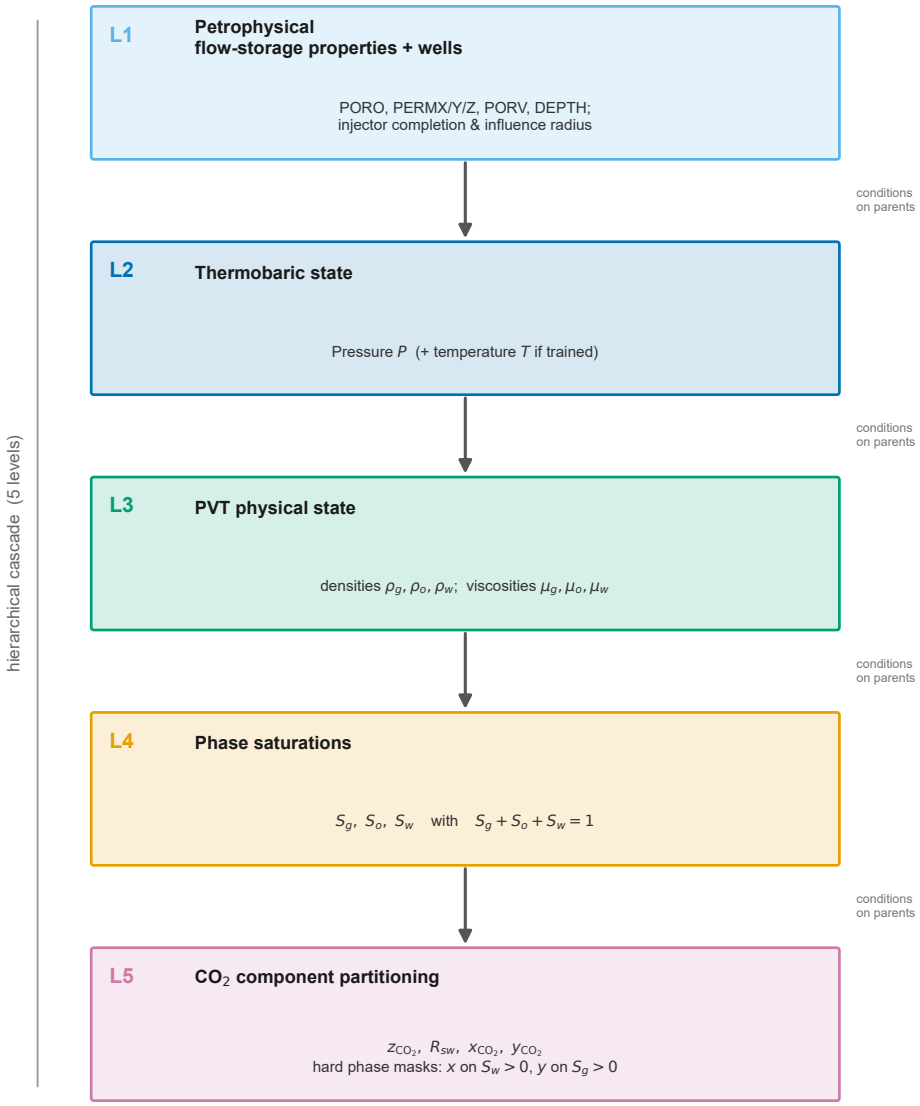


Figure 2 Hierarchical five-level cascade architecture of the storage surrogate. Each level is conditioned on the physical state inferred at the levels above it: (L1) static reservoir petrophysical flow-and-storage properties and well completions, (L2) thermobaric pressure (and temperature when a genuine field is available), (L3) PVT densities and viscosities, (L4) phase saturations constrained to sum to unity, and (L5) CO_2 component partitioning with hard phase-presence masks.

$$\hat{S}_o = (1 - \hat{S}_g)\,\hat{f}_o, \qquad \hat{S}_w = (1 - \hat{S}_g)\,\hat{f}_w \quad \Longrightarrow \quad \hat{S}_g + \hat{S}_o + \hat{S}_w = 1, \tag{9}$$

where $\sigma(\cdot)$ is the logistic function. This decomposition is motivated by the sparsity of the gas saturation field: the gate $\sigma(\ell_c^g)$ suppresses the gas prediction in background aquifer cells where $S_g \approx 0$, while the amplitude term controls the gas content where the plume is present.

Level 5 — Composition head and hard phase masks. Conditional on all upstream outputs and the predicted saturations, the composition head returns the CO_2 distribution across phases,

$$\big(\hat{z}_{\mathrm{CO_2}}, \hat{R}_{sw}, \hat{x}_{\mathrm{CO_2}}, \hat{y}_{\mathrm{CO_2}}\big) = g_\theta^{\mathrm{comp}}\big([\,\mathbf{h}_c, \boldsymbol{\tau}_c, \hat{\boldsymbol{\pi}}_c, \hat{S}_g, \hat{S}_o, \hat{S}_w\,]\big). \tag{10}$$

The overall mole fraction $\hat{z}_{\mathrm{CO_2}}$ is always produced. Because a component cannot reside in a phase that is absent from a cell, the phase-resident fractions are subjected to *hard phase masks*,

$$\begin{aligned} \hat{y}_{\mathrm{CO_2}} &\leftarrow \hat{y}_{\mathrm{CO_2}} \cdot \mathbb{1}\Big[\hat{S}_g \geq \theta_g\Big], \\ \hat{x}_{\mathrm{CO_2}} &\leftarrow \hat{x}_{\mathrm{CO_2}} \cdot \mathbb{1}\Big[\hat{S}_o \geq \theta_o\Big], \\ \hat{R}_{sw} &\leftarrow \hat{R}_{sw} \cdot \mathbb{1}\Big[\hat{S}_w \geq \theta_w\Big], \end{aligned} \tag{11}$$

with thresholds $\theta_g = \theta_o = 0.01$ and $\theta_w = 0.05$, mirroring the phase masks used when the supervised targets are assembled.

2.5 Training objective and physical constraints

The parameters θ are obtained by minimising a composite objective that sums mean-squared-error terms over the explicit cascade groups, weighted to balance the heterogeneous magnitudes and physical importance of the variables,

$$\mathcal{L}(\theta) = w_P\,\mathcal{L}_P + w_T\,\mathcal{L}_T + w_{\mathrm{PVT}}\,\mathcal{L}_{\mathrm{PVT}} + w_S\,\mathcal{L}_S + w_z\,\mathcal{L}_z + w_{\mathrm{comp}}\,\mathcal{L}_{\mathrm{comp}} + w_m\,\mathcal{L}_{\mathrm{mask}}. \tag{12}$$

Here $\mathcal{L}_P$ and $\mathcal{L}_T$ penalise the thermobaric predictions, $\mathcal{L}_{\mathrm{PVT}}$ the densities and viscosities, $\mathcal{L}_S$ the saturations, and $\mathcal{L}_z$ and $\mathcal{L}_{\mathrm{comp}}$ the overall and phase-resident CO_2 fractions, respectively. A representative cascade term has the masked, normalised form

$$\mathcal{L}_\bullet = \frac{1}{\sum_c m_c} \sum_c m_c \left\| \frac{\hat{\mathbf{y}}_c^\bullet - \mathbf{y}_c^\bullet}{s^\bullet} \right\|_2^2, \tag{13}$$

where $m_c \in \{0,1\}$ is an availability mask that excludes cells or fields absent from the data, and $s^\bullet$ is a per-variable scale. Pressure, temperature and PVT variables are standardised by z-scoring; bounded saturations and mole fractions are optimised in their natural $[0,1]$ units. The consistency term

$$\mathcal{L}_{\mathrm{mask}} = \frac{1}{N_a} \sum_c \Big(\hat{y}_{\mathrm{CO_2},c}^2\, \mathbb{1}[\hat{S}_{g,c} < \theta_g] + \hat{x}_{\mathrm{CO_2},c}^2\, \mathbb{1}[\hat{S}_{o,c} < \theta_o] + \hat{R}_{sw,c}^2\, \mathbb{1}[\hat{S}_{w,c} < \theta_w] \Big) \tag{14}$$

penalises residual component mass predicted in an absent phase, reinforcing the hard masks of Eq. (11) during optimisation. Saturation closure (Eq. 9) is guaranteed by construction and requires no penalty term. The network is trained with the Adam optimiser and a plateau-based learning-rate scheduler; convergence behaviour, per-component losses and the generalisation gap are reported in Fig. 3.

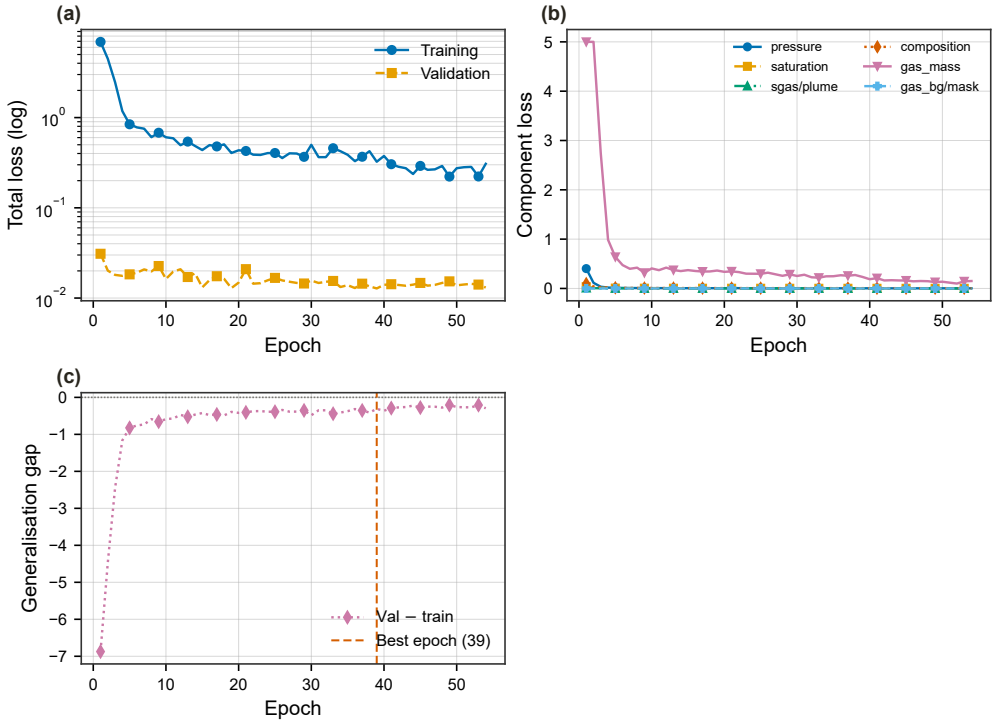


Figure 3 Training convergence dynamics of the hierarchical cascade surrogate: (a) total training and validation loss versus epoch on a logarithmic ordinate; (b) the learning-rate schedule; (c) the per-component cascade losses (pressure/thermobaric, saturations, gas-plume, composition and mass/mask terms); and (d) the generalisation gap (validation minus training loss) with the best, minimum-validation-loss, epoch marked. Curves are distinguished by line style and marker so that they remain legible in grayscale.

2.6 Evaluation metrics for plume migration and phase partitioning

Predictive skill is quantified on the blind transition of Section 2.2 with complementary regression, spatial and mass-balance metrics, following established practice for CO_2 storage surrogates (Wen et al. 2021a; Zhang et al. 2025; Du et al. 2024). For each target field we report the coefficient of determination, the mean absolute error and

the root-mean-square error,

$$R^2 = 1 - \frac{\sum_c (y_c - \hat{y}_c)^2}{\sum_c (y_c - \bar{y})^2}, \quad \mathrm{MAE} = \frac{1}{N_a} \sum_c |y_c - \hat{y}_c|, \quad \mathrm{RMSE} = \sqrt{\frac{1}{N_a} \sum_c (y_c - \hat{y}_c)^2}. \tag{15}$$

For fields that are near-constant over the active domain (for example the water viscosity μ_w), the denominator of R^2 collapses and the statistic becomes uninformative; in these cases R^2 is reported as not-available while the absolute-error metrics are retained.

The accuracy of plume localisation is assessed by thresholding the gas saturation to obtain predicted and reference plume masks $\hat{\Omega}$ and Ω, and computing precision, recall, F_1 score and intersection-over-union,

$$\mathrm{Precision} = \frac{|\hat{\Omega} \cap \Omega|}{|\hat{\Omega}|}, \quad \mathrm{Recall} = \frac{|\hat{\Omega} \cap \Omega|}{|\Omega|}, \quad \mathrm{F}_1 = \frac{2\,\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \quad \mathrm{IoU} = \frac{|\hat{\Omega} \cap \Omega|}{|\hat{\Omega} \cup \Omega|}. \tag{16}$$

Bulk containment is verified through a pore-volume-weighted gas-inventory error and the displacement of the plume centroid. With cell pore volumes V_c,

$$\Delta_{\mathrm{inv}} = \frac{\sum_c V_c \hat{S}_{g,c} - \sum_c V_c S_{g,c}}{\sum_c V_c S_{g,c}}, \qquad \Delta_{\mathrm{cen}} = \left\| \hat{\mathbf{r}}_{\mathrm{cen}} - \mathbf{r}_{\mathrm{cen}} \right\|, \quad \mathbf{r}_{\mathrm{cen}} = \frac{\sum_c V_c S_{g,c} \mathbf{r}_c}{\sum_c V_c S_{g,c}}. \tag{17}$$

The fidelity of the phase split is summarised by a phase-partitioning error that aggregates the discrepancy of the phase-resident CO_2 fractions over cells in which the carrier phase is present,

$$\mathcal{E}_{\mathrm{part}} = \frac{1}{|\mathcal{P}|} \sum_{j \in \{g,o,w\}} \sum_{c \in \mathcal{P}_j} \left| \hat{c}_{j,c} - c_{j,c} \right|, \qquad c_j \in \{\, y_{\mathrm{CO_2}}, x_{\mathrm{CO_2}}, R_{sw} \,\}, \tag{18}$$

where $\mathcal{P}_j$ is the set of cells with phase j present and $\mathcal{P} = \bigcup_j \mathcal{P}_j$. Together these metrics characterise the geometry of the migrating plume and the thermodynamic partitioning of CO_2, the two quantities of primary interest for storage-security assessment (IEA et al. 2024; Samnioti and Gaganis 2023).

2.7 Implementation and reproducibility

The surrogate and its training and prediction stages are implemented in Python with PyTorch; the figure-generation, schedule-parsing and feature-construction utilities depend only on NumPy and Matplotlib and use no proprietary libraries, so the entire post-processing pipeline runs headless and cross-platform (Windows, macOS and Linux). The cascade is backward compatible: in the absence of PVT and temperature fields the model reduces to predicting the eight base targets, and all cascade-specific metadata are optional and defaulted when absent. All extracted fields, feature definitions, well-schedule summaries and per-epoch training logs are serialised to disk, and

every figure reported here is regenerated deterministically from these on-disk products, supporting reproducibility of the reported results.

3 Results

The HCS-CO_2S surrogate is evaluated on a single held-out long-horizon prediction task. Results are reported for reservoir pressure, phase saturations, the overall CO_2 mole fraction, the liquid and vapour phase CO_2 fractions, and the PVT context fields, with emphasis on the quantities relevant to storage-security assessment: plume localisation, gas inventory, phase partitioning and satisfaction of physical constraints.

3.1 Blind long-horizon prediction setup

The trained surrogate is assessed in a strictly blind configuration. Using the hierarchical cascade over levels L1–L5 with hard phase-presence masks, the compositional reservoir state at report step 144 is predicted from the state at step 123, a step pair excluded from training. The active domain contains $N_a = 21{,}142$ cells on a $31 \times 31 \times 22$ grid. The model operates in isothermal / PVT-only mode: reservoir temperature is not a trained target, and the PVT fields (densities and viscosities) are predicted as physical-context variables that condition fluid mobility and pressure response.

The phase-saturation closure $S_g + S_o + S_w = 1$ holds with a mean absolute residual of 2.04×10^{-8} and a maximum of 1.19×10^{-7} across all active cells, confirming that the gas-gated saturation head produces a physically admissible phase state.

Figure 3 shows the convergence of the training run: the total training and validation losses decrease together and stabilise without divergence, the per-component panel confirms that no single cascade term dominates or collapses, and the generalisation gap remains small at the selected epoch, indicating that the retained model is neither under- nor over-fitted on the training pairs. All results below correspond to the minimum-validation-loss checkpoint.

3.2 Overall field reconstruction accuracy

Table 1 summarises the per-variable blind-validation accuracy for the primary state and compositional targets. Reservoir pressure is reconstructed with a mean absolute error (MAE) of 0.798 bar and a root mean square error (RMSE) of 1.16 bar against a field with a dynamic range of approximately 14 bar (true range 406.8–420.8 bar), giving $R^2 = 0.827$. The two liquid-phase saturations are recovered with closely matched accuracy (S_o: $R^2 = 0.914$, MAE $= 0.042$; S_w: $R^2 = 0.915$, MAE $= 0.042$).

The gas saturation S_g is a strongly zero-inflated field in which the CO_2 plume occupies only a small fraction of the domain, so the full-field MAE (2.77×10^{-4}) is dominated by the large background of cells at $S_g \approx 0$, while $R^2 = 0.837$ reflects discrimination between gas-bearing and background cells. The plume itself is assessed separately in Section 3.3. The overall CO_2 mole fraction ZMF_5, the most directly storage-relevant compositional field, is predicted with $R^2 = 0.913$ (MAE $= 0.033$).

Spatial cross-sections of the true and predicted fields are compared in Fig. 4.

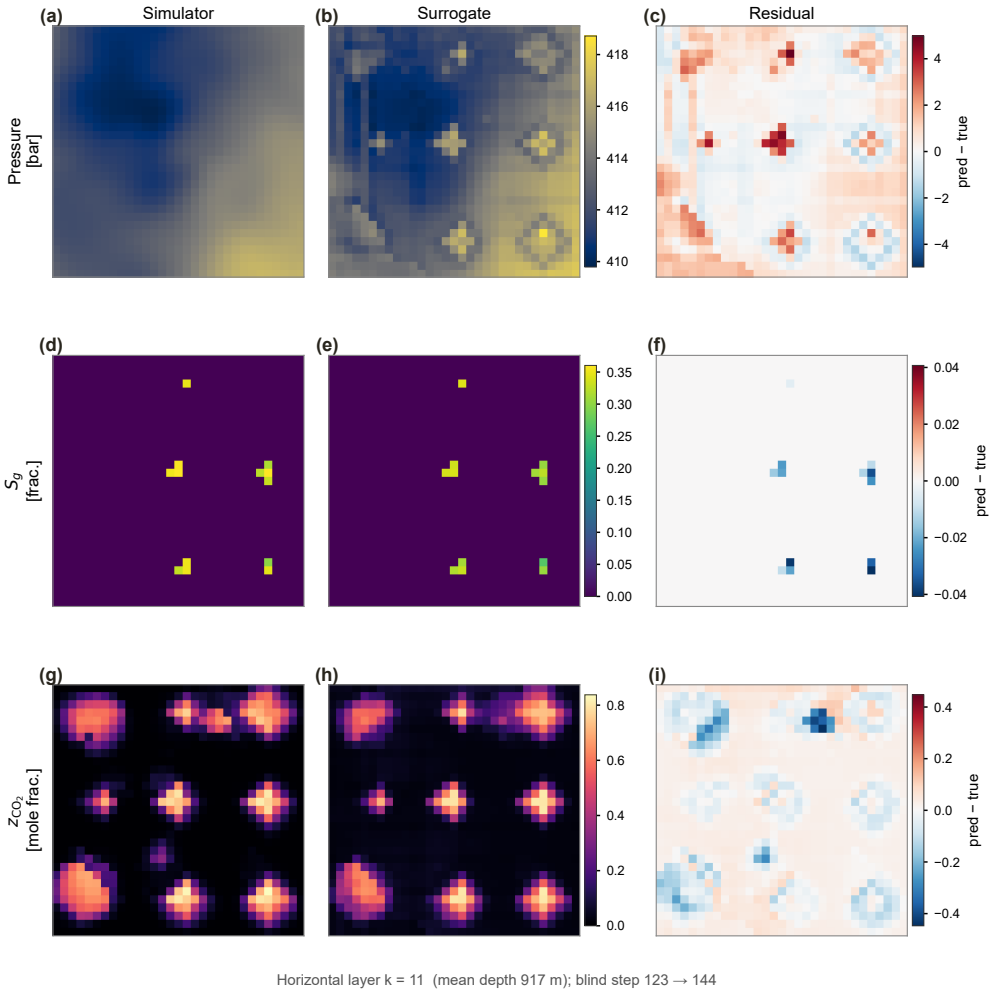


Figure 4 Spatial comparison between the reservoir simulator (left), the surrogate prediction (centre) and their residual (right) for the blind step 123 → 144 on the plume-core horizontal layer. Rows show pressure, gas saturation S_g, and the CO_2 overall mole fraction ZMF_5. Simulator and surrogate panels share a common per-variable colour scale; residual panels use a symmetric diverging scale.

The pressure row of Fig. 4 shows that the surrogate reproduces the large-scale pressure pattern, with residuals confined to a small number of cells associated with post-injection relaxation (Section 3.6). The gas-saturation row shows that the predicted plume occupies the same compact region as the simulator field, with non-trivial residuals only along the plume body and its immediate margin. The ZMF_5 row confirms that the spatial extent of the compositionally significant CO_2 region is reproduced, with residuals that are small and spatially unstructured. Taken together, the residual columns show that surrogate error is concentrated where the physics is most dynamic rather than spread diffusely across the field.

3.3 Carbon dioxide plume localisation and migration

Treating plume detection as a binary spatial classification at the gas-saturation threshold $S_g > 0.05$, the model recovers every truly gas-bearing cell: precision = 0.857, recall = 1.000, $F_1 = 0.923$, and intersection-over-union IoU = 0.857, with 72 true positives, 12 false positives and no false negatives (Table 2). The plume is captured in full, with a modest over-detection of 12 peripheral cells; this asymmetry is acceptable for conservative containment screening.

Within the true plume ($S_g > 0.05$, $n = 72$), gas saturation is slightly underpredicted: the in-plume true mean is 0.329 (max 0.362) versus a predicted mean of 0.304 (max 0.353), a bias of -0.025, with 71 of 72 plume cells underpredicted. The within-plume saturation correlation is correspondingly weaker ($R^2 = 0.339$), indicating that the model localises the plume more accurately than it reproduces the internal saturation gradient. This behaviour is consistent with the gas gate, which activates on 69 of the 72 true gas cells (gate recall = 0.806) at a false-positive rate of 5.2×10^{-4}: the gate is highly specific but slightly conservative, which suppresses spurious background gas while contributing to the mild in-plume underprediction.

Plume migration is captured to near grid resolution. The predicted gas-mass centroid is displaced from the true centroid by $(\Delta i, \Delta j, \Delta k) = (-0.99, 0.53, -0.31)$ cells, a total shift of 1.16 cell-widths. The gas inventory restricted to detected plume cells is underestimated by 7.8%, mirroring the in-plume saturation bias, whereas the inventory integrated over the full field is overestimated by 10.2% due to the 12 over-detected peripheral cells.

Figure 5 makes these diagnostics concrete. The mask-overlay panel shows the predicted plume enclosing the simulator plume with a thin rim of over-detected peripheral cells, the spatial signature of perfect recall combined with precision of 0.857. The centroid markers nearly coincide, illustrating the 1.16-cell-width migration error, and the vertical-extent panel confirms that the plume is captured at the correct depths.

3.4 Carbon dioxide phase partitioning and compositional consistency

In addition to plume localisation, the surrogate predicts how the CO_2 component is partitioned between the aqueous and non-aqueous phases, which controls the dissolution-trapping fraction. The dissolved gas/water ratio RSW_5 is predicted with $R^2 = 0.905$ (MAE = 0.830 sm^3/m^3, RMSE = 1.87 sm^3/m^3), and the liquid-phase CO_2 mole fraction XMF_5 with $R^2 = 0.926$ (MAE = 0.024) over the $n = 20{,}745$ cells where the aqueous phase is present. The vapour-phase CO_2 mole fraction YMF_5 is meaningful only inside the free-gas plume ($n = 72$); over this restricted support it is recovered with low absolute error (MAE = 0.014, RMSE = 0.019) and a near-zero residual mean (+0.006), although the narrow value range yields a pessimistic $R^2 = 0.627$.

Figure 6 clarifies why the partitioning fields behave differently. The RSW_5 and XMF_5 scatter panels lie tightly along the identity line across the full ∼20,700-cell aqueous support, reflecting R^2 values of 0.905 and 0.926. The YMF_5 scatter is restricted to 72 free-gas cells and spans a narrow value interval, so although the absolute errors are small, the explained-variance statistic is intrinsically weak over such a restricted

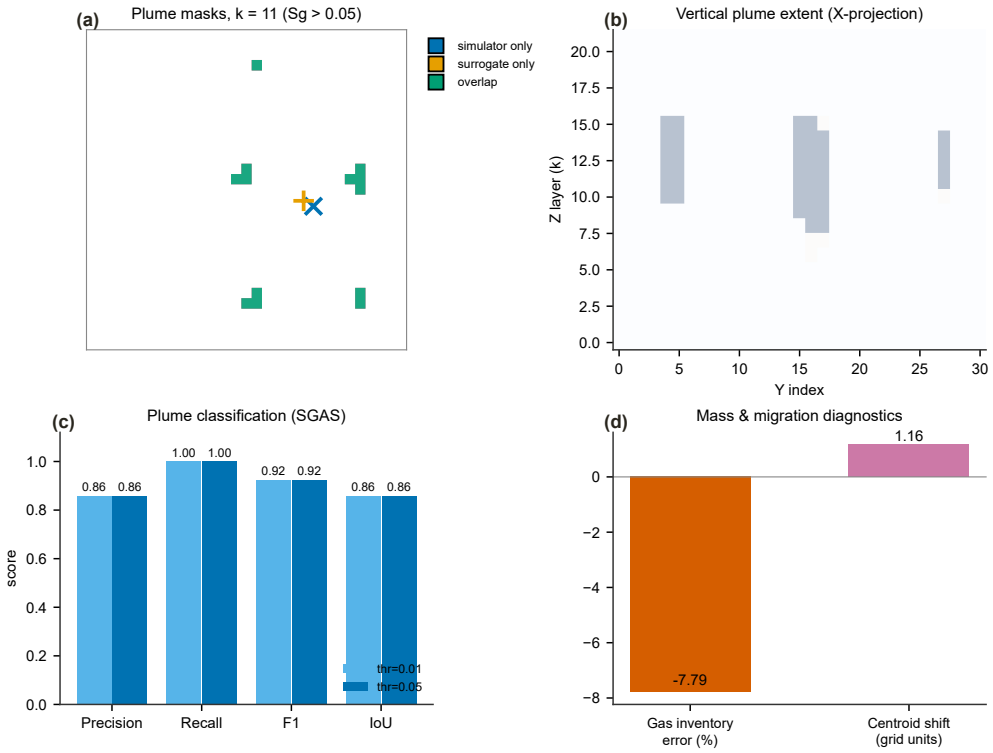


Figure 5 Gas-plume localisation and migration diagnostics for the blind step 123 → 144: simulator/surrogate plume-mask overlay with mass-weighted centroids, the vertical plume extent, the classification scores at thresholds 0.01 and 0.05, and the gas-inventory error and centroid shift.

support. The limited YMF_5 R^2 therefore reflects gas-cell sparsity and value range, not a systematic compositional bias.

Two consistency checks support the compositional predictions. First, saturation closure holds to $\sim 10^{-8}$ (Section 3.1). Second, the binary structure of ZMF_5 is well reproduced: at threshold 0.05, precision = 0.813, recall = 0.903, F_1 = 0.856, IoU = 0.748 (n_{true} = 5,115, n_{pred} = 5,679); at the stricter threshold 0.10, precision = recall = F_1 = 0.882 and IoU = 0.789 (n_{true} = 4,196, n_{pred} = 4,198), indicating that the surrogate reproduces the spatial extent of the compositionally significant CO_2 region without systematic inflation.

3.5 PVT-context prediction

The PVT fields condition fluid mobility and pressure response and are not primary storage targets. Gas and oil densities are reconstructed with R^2 = 0.917 (DENG, MAE = 2.92) and R^2 = 0.937 (DENO, MAE = 3.55), and the corresponding viscosities with R^2 = 0.845 (VISG) and R^2 = 0.949 (VISO). Water density (DENW) is predicted with small absolute error (MAE = 0.245, RMSE = 0.486) but a more modest R^2 =

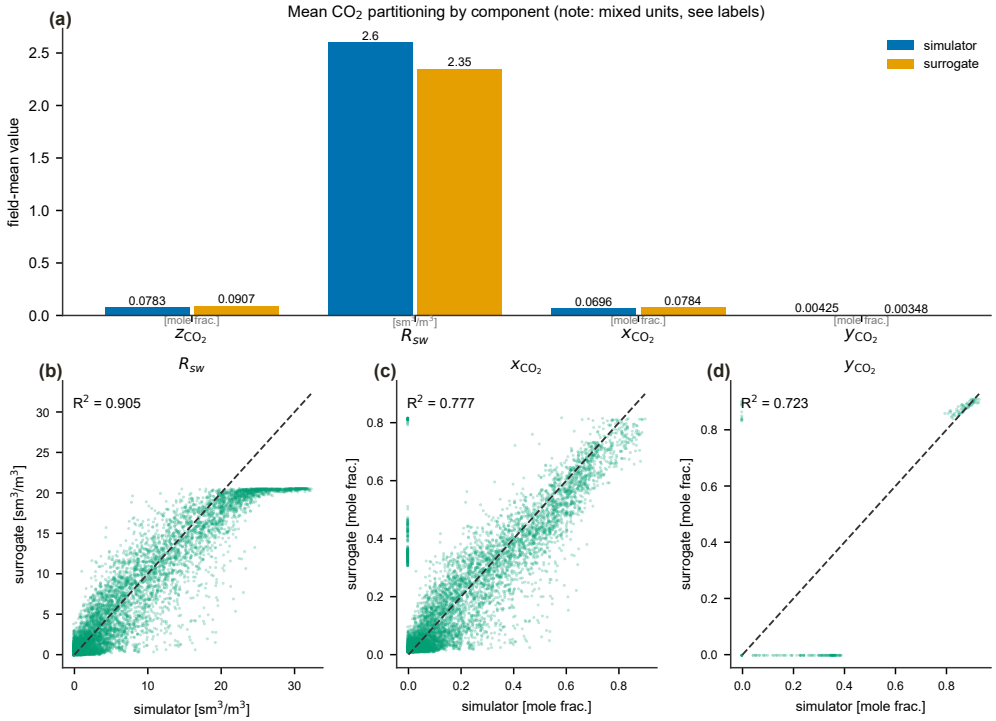


Figure 6 CO_2 component partitioning for the blind step 123 → 144: field-mean composition (overall mole fraction z, dissolved gas/water ratio R_{sw}, liquid-phase mole fraction x and vapour-phase mole fraction y; mixed units as labelled) together with cell-wise simulator-versus-surrogate scatter and identity lines for R_{sw}, x and y.

0.526, reflecting the narrow dynamic range of brine density under the isothermal, near-incompressible conditions of this case.

Water viscosity (VISW) requires a note on metric interpretation. Its MAE (5.57×10^{-3}) and RMSE (6.14×10^{-3}) are small in absolute terms, but the field is essentially constant across the domain. For such a near-constant target the variance of the truth is negligible, so the R^2 statistic is numerically ill-conditioned and assumes a large negative value. This behaviour reflects the properties of R^2 for near-constant fields, not a deterioration in prediction accuracy. VISW is therefore characterised by its absolute error, and R^2 is reported as not meaningful in Table 3. The supplementary PVT scatter (Fig. 7) illustrates this directly: the density and gas/oil viscosity panels follow the identity line, whereas the VISW panel collapses to a near-constant value.

3.6 Residual structure and model limitations

The residual fields (Fig. 8) characterise the surrogate's remaining weaknesses.

The violin panels show symmetric, sharply peaked residual distributions centred near zero for the compositional and saturation fields, indicating the absence of systematic bias. The most material limitation is in the pressure field. Although the bulk

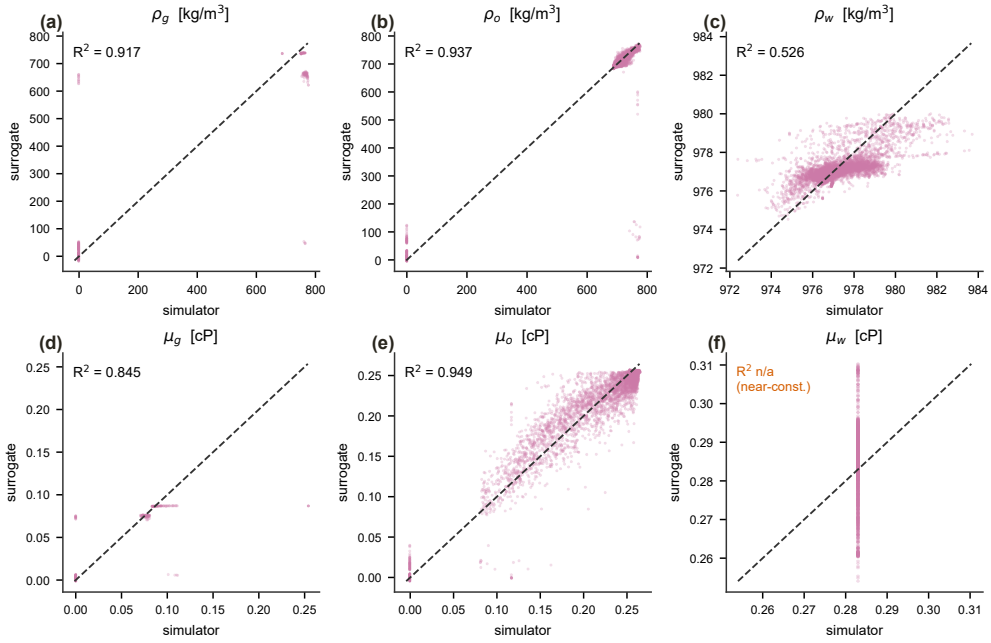


Figure 7 Supplementary PVT context for the blind step 123 → 144: simulator-versus-surrogate scatter for fluid densities and viscosities. Near-constant fields such as water viscosity (VISW) report R^2 as not available to avoid a misleading statistic and are instead characterised by their absolute error.

of the pressure residuals is tightly distributed (median +0.49 bar, 5th–95th percentile interval [−0.50, 2.13] bar), the distribution is positively skewed (mean +0.659 bar) and carries a heavy tail reaching a maximum absolute error of 11.3 bar. These outliers are associated with post-injection pressure relaxation: the point-wise MLP does not impose spatial diffusion or smoothness constraints, and therefore cannot fully reproduce the smooth relaxation of localised pressure anomalies after injection ceases. The depth-profile panel localises these anomalies to specific layers, consistent with a small number of relaxation-affected cells. A Laplacian smoothness penalty on the pressure field, or a spatial graph encoder propagating information between neighbouring cells, represent possible extensions for future work.

The second limitation is the mild in-plume underprediction discussed in Section 3.3: gas saturation bias of −0.025 with 71/72 cells underpredicted, and over-detection of 12 peripheral cells. The gas gate is tuned for high specificity (FPR $= 5.2 \times 10^{-4}$) at the cost of a sub-unity recall (0.806), making the surrogate conservative inside the plume while over-detecting peripheral cells. For containment screening this is acceptable, but it introduces the threshold-dependent inventory errors noted above (−7.8% within the detected plume, +10.2% over the full field).

A compact overview of the per-target accuracy and plume classification scores is provided in Fig. 9.

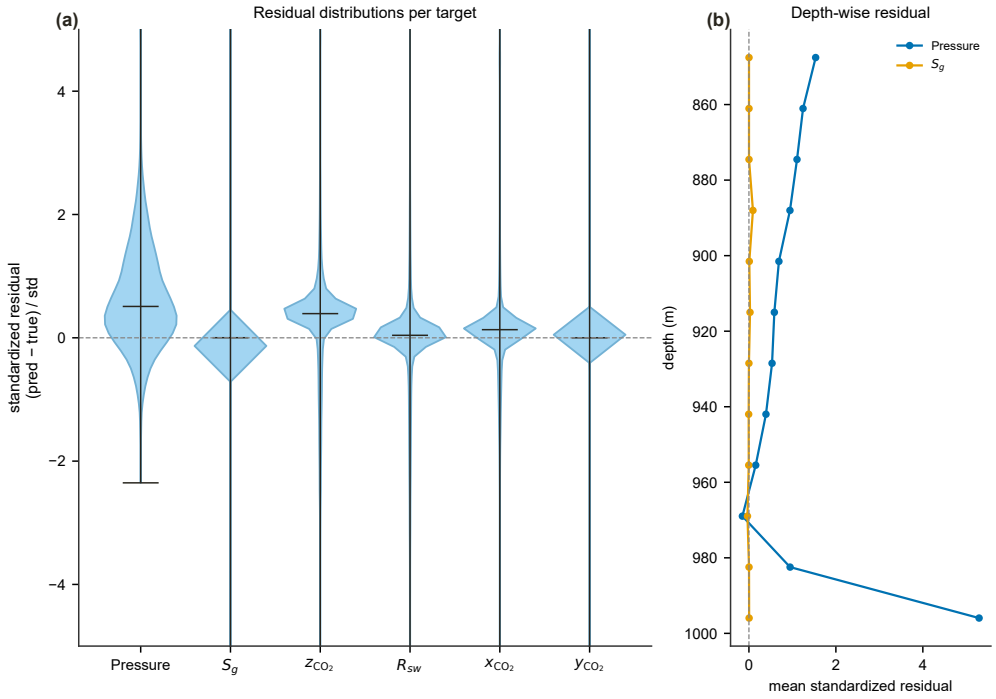


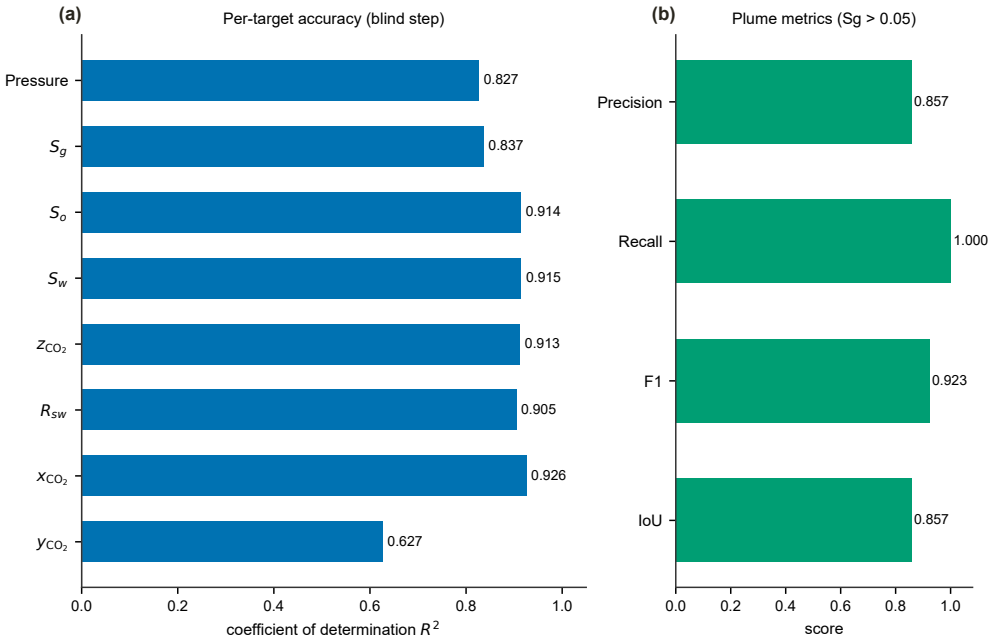
Figure 8 Residual diagnostics for the blind step 123 → 144: standardised residual distributions per target (violin plots) and the depth-binned mean standardised residuals for pressure and gas saturation. The residual is defined as prediction minus simulator truth.

Figure 9 consolidates the headline numbers: the R^2 panel groups the high-accuracy compositional and saturation targets ($R^2 \gtrsim 0.90$) together with the pressure field ($R^2 = 0.827$), while the plume-classification panel highlights the perfect recall and $F_1 = 0.923$ that anchor the localisation result.

These results demonstrate blind generalisation in time — a long-horizon 123 → 144 prediction on a held-out step pair for a single geological model. Generalisation across geological realisations is not claimed; validating transfer to different reservoir grids remains an explicit direction for future work.

3.7 Computational performance

The blind prediction over all 21,142 active cells, including model loading, normalisation, full hierarchical cascade inference and metric computation, completes in 22.4 s on CPU without requiring GPU resources at inference time, making the surrogate suitable for many-realisation uncertainty-quantification workflows in which the bottleneck is ensemble size rather than single-run fidelity.



Note: VISW R^2 is excluded from the main accuracy summary because the field is near-constant, making R^2 unreliable; see supplementary figure S1 for PVT context.

Figure 9 Summary of blind-validation accuracy for the step 123 → 144: per-target coefficient of determination R^2 and the gas-plume classification scores at $S_g > 0.05$. Water viscosity is excluded from the R^2 panel as a near-constant field.

Table 1 Blind long-horizon validation accuracy (123 → 144) for the primary state and compositional targets. MAE and RMSE are reported in the native unit of each variable; n is the number of active cells over which the metric is evaluated. ZMF_5, RSW_5, XMF_5 and YMF_5 refer to the overall CO_2 mole fraction, dissolved gas/water ratio, liquid-phase mole fraction and vapour-phase mole fraction, respectively.

Variable	MAE	RMSE	R^2	n	Unit
PRESSURE	0.798	1.163	0.827	21 142	bar
SGAS	2.77×10^{-4}	7.79×10^{-3}	0.837	21 142	frac
SOIL	0.042	0.082	0.914	21 142	frac
SWAT	0.042	0.081	0.915	21 142	frac
ZMF_5	0.033	0.049	0.913	21 142	frac
RSW_5	0.830	1.875	0.905	21 142	sm^3/m^3
XMF_5	0.024	0.045	0.926	20 745	frac
YMF_5	0.014	0.019	0.627	72	frac

Table 2 Plume, compositional and physical-consistency diagnostics for the blind 123 → 144 prediction. Classification metrics (P, R, F_1, IoU) treat field reconstruction as binary spatial detection at the indicated threshold. Inventory and centroid diagnostics quantify integrated CO_2 mass and the displacement of the gas-mass centroid in cell-widths.

Diagnostic	Value
SGAS detection ($S_g > 0.05$): P / R / F_1 / IoU	0.857 / 1.000 / 0.923 / 0.857
SGAS detection: TP / FP / FN	72 / 12 / 0
In-plume S_g mean (true / pred)	0.329 / 0.304
In-plume S_g bias; underpredicted cells	−0.025; 71/72
Plume-only SGAS R^2	0.339
ZMF_5 detection (> 0.05): P / R / F_1 / IoU	0.813 / 0.903 / 0.856 / 0.748
ZMF_5 detection (> 0.10): P / R / F_1 / IoU	0.882 / 0.882 / 0.882 / 0.789
Gas inventory error (plume, $S_g > 0.05$)	−7.8%
Gas inventory error (full field)	+10.2%
Gas-mass centroid shift ($\Delta i, \Delta j, \Delta k$; dist.)	(−0.99, 0.53, −0.31); 1.16 cell-widths
Gas gate recall / FPR ($S_g > 0.01$)	0.806 / 5.2×10^{-4}
Saturation closure $\lvert S_g + S_o + S_w - 1\rvert$ (mean / max)	2.04×10^{-8} / 1.19×10^{-7}

Table 3 PVT-context prediction accuracy for the blind 123 → 144 case. For water viscosity (VISW) the field is near-constant and R^2 is numerically ill-conditioned; this variable is characterised by its absolute error only.

Variable	MAE	RMSE	R^2
DENG (gas density)	2.920	19.72	0.917
DENO (oil density)	3.547	22.19	0.937
DENW (water density)	0.245	0.486	0.526
VISG (gas viscosity)	3.56×10^{-4}	3.29×10^{-3}	0.845
VISO (oil viscosity)	4.04×10^{-3}	8.91×10^{-3}	0.949
VISW (water viscosity)	5.57×10^{-3}	6.14×10^{-3}	not meaningful

4 Discussion

4.1 Implications for compositional carbon dioxide storage surrogate modelling

The results show that a feature-based, cell-wise hierarchical cascade can reproduce the coupled pressure, saturation and compositional state of a compositional CO_2-storage simulation over a long blind horizon while honouring saturation closure $S_g + S_o + S_w = 1$ to $\sim 10^{-7}$. The surrogate belongs to the same family of deep-learning proxies developed to accelerate geological CO_2 storage forecasting (Wen et al. 2021a, 2022, 2023; Pawar et al. 2026; Zhang et al. 2025), but differs in two respects that are important for compositional storage. First, it predicts not only the bulk state (pressure and phase saturations) but the explicit partitioning of the CO_2 component between the aqueous and non-aqueous phases (ZMF_5, RSW_5, XMF_5 and YMF_5), which governs dissolution trapping in compositional models (Moortgat et al. 2012; Voskov et al. 2017). Second, it operates on the native simulation grid and completes the full blind prediction in 22.4 s on CPU without requiring a specialised accelerator,

which is directly relevant to many-realisation screening and uncertainty-quantification workflows (Pawar et al. 2026; Li et al. 2023; Du et al. 2024).

4.2 Importance of plume-aware and phase-consistent evaluation

An important methodological consideration is that a low full-field error can conceal poor storage-relevant behaviour, and conversely a modest full-field statistic does not necessarily indicate a weak surrogate. The gas saturation S_g is strongly zero-inflated: the free-gas plume occupies only 72 of 21,142 active cells, so the full-field MAE (2.77×10^{-4}) is dominated by the inactive background and says little about the plume itself. The quantities relevant to storage security — where the CO_2 is, how far it has migrated and how much is present — are captured by plume-aware diagnostics: the surrogate detects every gas-bearing cell (recall 1.000, $F_1 = 0.923$), places the gas-mass centroid within 1.16 cell-widths of the truth, and reproduces the compositional CO_2 extent (ZMF_5 detection $F_1 = 0.882$ at threshold 0.10). These results motivate reporting classification, centroid and inventory metrics alongside regression error, and justify restricting YMF_5 evaluation to its physical support. Evaluating storage surrogates on plume-aware and phase-consistent metrics is consistent with the emphasis on plume migration and containment in the broader storage literature (Wen et al. 2023; Zhang et al. 2025). The near-perfect recall with a thin rim of over-detected peripheral cells is a favourable error structure for conservative containment screening, since it overestimates rather than underestimates the plume footprint.

4.3 Role of the hierarchical cascade and well-influence features

The accuracy of the phase partitioning follows directly from the hierarchical cascade and the physical conditioning it imposes. By predicting the thermobaric pressure and PVT context first and conditioning the saturation and composition levels on that state, the cascade respects the physical dependency chain of a compositional model rather than treating all targets as independent regression outputs. The hard phase-presence masks and the saturation-closure head then guarantee that the emitted state is admissible — $S_g + S_o + S_w = 1$ to $\sim 10^{-8}$ — which is not automatic for a multi-output regressor. This level-wise design is conceptually aligned with physics-aware and physics-framed proxy strategies that embed governing relations or staged physical reasoning into the network architecture (Omosebi et al. 2022; Shokouhi et al. 2021; Coutinho et al. 2021). The cell-wise, feature-based formulation together with explicit well-influence features also allows the surrogate to operate directly on the native corner-point reservoir grid and on the irregular active-cell geometry, without resampling onto a regular tensor as image-based convolution proxies require.

4.4 Limitations and sources of residual error

Three limitations warrant discussion. First, the pressure field carries a heavy-tailed residual (mean +0.659 bar, maximum absolute error 11.3 bar) concentrated in a small number of cells affected by post-injection pressure relaxation. Because pressure is predicted cell-wise with no spatial diffusion or smoothness constraint, the model cannot

fully reproduce the smooth spatial relaxation of localised pressure anomalies after injection ceases; this is the principal candidate for architectural improvement. Second, gas saturation is slightly underpredicted inside the plume (bias -0.025, 71/72 cells underpredicted) and the plume footprint is mildly over-detected, a consequence of a gas gate tuned for high specificity (FPR $= 5.2 \times 10^{-4}$) at a sub-unity recall (0.806); the resulting inventory error is threshold-dependent (-7.8% inside the detected plume, $+10.2\%$ over the full field). Third, two metric-level observations require careful interpretation: the limited YMF_5 $R^2 = 0.627$ reflects the sparsity and narrow value range of the 72 free-gas cells rather than a compositional bias (MAE $= 0.014$), and the negative R^2 for water viscosity is a numerical artefact of a near-constant field with negligible truth variance, so VISW is correctly characterised by its absolute error (5.57×10^{-3}). The PVT fields are predicted as physical context that conditions fluid mobility and pressure response, not as primary storage endpoints; their high R^2 values ($\geq$0.92 for DENG, DENO and VISO) support the thermobaric state but should not be read as the headline storage result.

4.5 Transferability to other geological grids and future work

The present study demonstrates blind temporal generalisation on a single geological model; transfer to different geometries, boundary conditions and rock properties remains to be tested. The feature-based formulation is structurally suited to this extension, because the model consumes per-cell physical and well-influence features rather than a fixed-size image tensor, allowing it to be fine-tuned on, or trained jointly over, multiple realisations (Wen et al. 2022; Zhang et al. 2025; Li et al. 2023).

Several concrete directions follow from the limitations above. The pressure relaxation tail motivates a spatially aware extension — either a Laplacian smoothness penalty on the predicted pressure or a graph/spatial encoder propagating information between neighbouring cells — to capture the diffusion-like relaxation that a pointwise model omits; such spatial inductive biases are characteristic of convolution-, Fourier-neural-operator- and graph-based proxies (Wen et al. 2021a, 2022), whereas physics-informed losses offer a complementary route to enforce governing constraints directly (Shokouhi et al. 2021; Coutinho et al. 2021). The gas-gate conservatism and in-plume bias motivate plume-amplitude-aware training and a recalibrated gate to reduce inventory error without sacrificing recall. Finally, multi-realisation training and temporal rollout validation across additional geological scenarios would assess transfer and error accumulation over chained predictions, which is essential before such surrogates are applied to the screening and uncertainty-quantification workflows that motivate compositional storage proxies (Pawar et al. 2026; Samnioti and Gaganis 2023; Du et al. 2024; Li et al. 2023).

Statements and Declarations

- **Funding:** The authors did not receive support from any organization for the submitted work.
- **Competing interests:** The authors have no competing interests to declare that are relevant to the content of this article.

- **Availability of data and materials:** The datasets generated and analysed during the hydrodynamic modelling of geological CO_2 sequestration are available from the corresponding author on reasonable request.
- **Code availability:** The source code of the hierarchical cascade surrogate model developed in this study is available from the corresponding author on reasonable request.
- **Authors' contributions:** Stepan Zainulin: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. Anna Storozheva: Supervision, Writing – review and editing.
- **Use of AI and large language models:** During the preparation of this manuscript, the authors used large language model (LLM)-based AI-assistants to improve the language, grammar, and clarity of the text. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the content of the published article.

References

Barros WQ, Pires AP (2021) Application of neural network to speed-up equilibrium calculations in compositional reservoir simulation. Artificial Intelligence in Geosciences 2:202–214. https://doi.org/10.1016/j.aiig.2022.03.004, URL https://linkinghub.elsevier.com/retrieve/pii/S2666544122000119

Chen J, Gildin E, Kompantsev G (2024) Optimization of pressure management strategies for geological CO2 storage using surrogate model-based reinforcement learning. International Journal of Greenhouse Gas Control 138:104262. https://doi.org/10.1016/j.ijggc.2024.104262, URL https://linkinghub.elsevier.com/retrieve/pii/S1750583624002056

Coutinho EJR, Dall'Aqua M, Gildin E (2021) Physics-Aware Deep-Learning-Based Proxy Reservoir Simulation Model Equipped With State and Well Output Prediction. Front Appl Math Stat 7:651178. https://doi.org/10.3389/fams.2021.651178, URL https://www.frontiersin.org/articles/10.3389/fams.2021.651178/full

Dong P, Liao X, Zhang L, et al (2023) A Surrogate Model for Numerical Reservoir Simulation of CO2 Flooding and Storage Based on Deep Learning. In: SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition. SPE, Jakarta, Indonesia, p D022S003R006, https://doi.org/10.2118/215253-MS, URL https://onepetro.org/SPEAPOG/proceedings/23APOG/23APOG/D022S003R006/535147

Du L, Du J, Fang Z, et al (2026) A Recurrent Neural Network Surrogate Model with Few-Shot Strategy for CO2 Storage in Deep Subsurface Saline Aquifer with Limited Direct Numerical Simulation Samples. Lithosphere 2026(2):lithosphere_2025_139. https://doi.org/10.2113/lithosphere_2025_139, URL https://pubs.geoscienceworld.org/lithosphere/article/2026/2/lithosphere_2025_139/728376/A-Recurrent-Neural-Network-Surrogate-Model-with

Du X, Salasakar S, Thakur G (2024) A Comprehensive Summary of the Application of Machine Learning Techniques for CO2-Enhanced Oil Recovery Projects. MAKE 6(2):917–943. https://doi.org/10.3390/make6020043, URL https://www.mdpi.com/2504-4990/6/2/43

Eyitayo S, Arbad N, Okere C, et al (2025) Advancing geological storage of carbon dioxide (CO2) with emerging technologies for climate change mitigation. Int J Environ Sci Technol 22(6):5023–5056. https://doi.org/10.1007/s13762-024-06074-w, URL https://link.springer.com/10.1007/s13762-024-06074-w

Fan M, Lu D, Liu S (2023) A deep learning-based direct forecasting of CO2 plume migration. Geoenergy Science and Engineering 221:211363. https://doi.org/10.1016/j.geoen.2022.211363, URL https://www.sciencedirect.com/science/article/pii/S2949891022000513

Fan M, Wang H, Zhang J, et al (2024) Advancing spatiotemporal forecasts of CO2 plume migration using deep learning networks with transfer learning and interpretation analysis. International Journal of Greenhouse Gas Control 132:104061. https://doi.org/10.1016/j.ijggc.2024.104061, URL https://www.sciencedirect.com/science/article/pii/S1750583624000045

Han Y, Hamon FP, Jiang S, et al (2024) Surrogate model for geological CO 2 storage and its use in hierarchical MCMC history matching. Advances in Water Resources 187:104678. https://doi.org/10.1016/j.advwatres.2024.104678, URL https://linkinghub.elsevier.com/retrieve/pii/S0309170824000654

Hua J, Du X, Yang F, et al (2024) Data-driven guided physics-informed segmented neural network for liquid–vapor flash calculation. Physics of Fluids 36(10):106624. https://doi.org/10.1063/5.0237206, URL https://pubs.aip.org/pof/article/36/10/106624/3316808/Data-driven-guided-physics-informed-segmented

IEA, Bason D, Nourollah H, et al (2024) Geological Storage of CO2: Seal Integrity Review. Tech. rep., IEAGHG, https://doi.org/10.62849/2024-06, URL https://ieaghg.org/publications/geological-storage-of-co2-seal-integrity-review/

IPCC (2022) Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty, 1st edn. Cambridge University Press, https://doi.org/10.1017/9781009157940, URL https://www.cambridge.org/core/product/identifier/9781009157940/type/book

IPCC (2023) Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st edn. Cambridge University Press, https://doi.org/10.1017/9781009157896, URL https://www.cambridge.org/core/product/identifier/9781009157896/type/book

Jiang S, Durlofsky LJ (2024) History matching for geological carbon storage using data-space inversion with spatio-temporal data parameterization. International Journal of Greenhouse Gas Control 134:104124. https://doi.org/10.1016/j.ijggc.2024.104124, URL https://linkinghub.elsevier.com/retrieve/pii/S1750583624000677

Jin ZL, Liu Y, Durlofsky LJ (2020) Deep-learning-based surrogate model for reservoir simulation with time-varying well controls. Journal of Petroleum Science and Engineering 192:107273. https://doi.org/10.1016/j.petrol.2020.107273, URL https://linkinghub.elsevier.com/retrieve/pii/S0920410520303533

Jing H, Liu J, Pan H, et al (2026) Efficient Phase Equilibrium Calculations of Shale Reservoir Fluids—Part III: Deep Learning–Based Acceleration and Integration into Compositional Simulation. SPE Journal 31(02):1179–1198. https://doi.org/10.2118/231421-PA, URL https://onepetro.org/SJ/article/31/02/1179/794214/Efficient-Phase-Equilibrium-Calculations-of-Shale

Larestani A, Sahebalzamani S, Hemmati-Sarapardeh A, et al (2025) Compositional modeling of solution gas–oil ratio (Rs): a comparative study of tree-based models, neural networks, and equations of state. Sci Rep 15(1):8428. https://doi.org/10.1038/s41598-025-91132-7, URL https://www.nature.com/articles/s41598-025-91132-7

Li C, Wang X, Yu W, et al (2026) A General Surrogate Model for CO2 Flooding Dynamic Prediction Based on Dimensionless Features and Implicit Time. ACS Omega 11(11):18486–18505. https://doi.org/10.1021/acsomega.6c01704, URL https://doi.org/10.1021/acsomega.6c01704

Li N, Feng W, Yu J, et al (2023) Recent Advances in Geological Storage: Trapping Mechanisms, Storage Sites, Projects, and Application of Machine Learning. Energy Fuels 37(14):10087–10111. https://doi.org/10.1021/acs.energyfuels.3c01433, URL https://doi.org/10.1021/acs.energyfuels.3c01433

Lie KA (2019) An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST), 1st edn. Cambridge University Press, https://doi.org/10.1017/9781108591416, URL https://www.cambridge.org/core/product/identifier/9781108591416/type/book

Ma X, Zhong R, Zhan J, et al (2024) Enhancing subsurface multiphase flow simulation with Fourier neural operator. Heliyon 10(18):e38103. https://doi.org/10.1016/j.heliyon.2024.e38103, URL https://linkinghub.elsevier.com/retrieve/pii/S2405844024141345

Meng S, Fu Q, Tao J, et al (2024) Predicting CO2-EOR and storage in low-permeability reservoirs with deep learning-based surrogate flow models. Geoenergy Science and Engineering 233:212467. https://doi.org/10.1016/j.geoen.2023.212467, URL https://linkinghub.elsevier.com/retrieve/pii/S2949891023010540

Miocic JM, Gilfillan SMV, Frank N, et al (2019) 420,000 year assessment of fault leakage rates shows geological carbon storage is secure. Sci Rep 9(1):769. https://doi.org/10.1038/s41598-018-36974-0, URL https://www.nature.com/articles/s41598-018-36974-0

Moortgat J, Li Z, Firoozabadi A (2012) Three-phase compositional modeling of CO_2 injection by higher-order finite element methods with CPA equation of state for aqueous phase. Water Resources Research 48(12):2011WR011736. https://doi.org/10.1029/2011WR011736, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR011736

Nagao M, Yao C, Onishi T, et al (2024) An efficient deep learning-based workflow for CO2 plume imaging considering model uncertainties with distributed pressure and temperature measurements. International Journal of Greenhouse Gas Control 132:104066. https://doi.org/10.1016/j.ijggc.2024.104066, URL https://linkinghub.elsevier.com/retrieve/pii/S1750583624000094

Omosebi OA, Oldenburg CM, Reagan M (2022) Development of lean, efficient, and fast physics-framed deep-learning-based proxy models for subsurface carbon storage. International Journal of Greenhouse Gas Control 114:103562. https://doi.org/10.1016/j.ijggc.2021.103562, URL https://linkinghub.elsevier.com/retrieve/pii/S1750583621003133

Pawar S, Chandra A, Panda A, et al (2026) Quantifying the Uncertainty of the 3D Evolution of CO2-Mass and Pressure Plumes for Site Screening Applications with Physics-Informed Fourier Neural Operators. SPE Journal pp 1–19. https://doi.org/10.2118/228266-PA, URL https://onepetro.org/SJ/article/doi/10.2118/228266-PA/795667/Quantifying-the-Uncertainty-of-the-3D-Evolution-of

Pour KM, Voskov D (2023) PHYSICS-INFORMED NEURAL NETWORKS BASED ON SEQUENTIAL TRAINING FOR CO2 UTILIZATION AND STORAGE IN SUBSURFACE RESERVOIR. J Mach Learn Model Comput 4(4):27–40. https://doi.org/10.1615/JMachLearnModelComput.2023048866, URL https://www.dl.begellhouse.com/journals/558048804a15188a,24f8785e4156a0df,157258503daf8310.html

Samnioti A, Gaganis V (2023) Applications of Machine Learning in Subsurface Reservoir Simulation—A Review—Part I. Energies 16(16):6079. https://doi.org/10.3390/en16166079, URL https://www.mdpi.com/1996-1073/16/16/6079

Shokouhi P, Kumar V, Prathipati S, et al (2021) Physics-informed deep learning for prediction of CO2 storage site response. Journal of Contaminant Hydrology 241:103835. https://doi.org/10.1016/j.jconhyd.2021.103835, URL https://linkinghub.elsevier.com/retrieve/pii/S0169772221000747

Stepien M, Ferreira CA, Hosseinzadehsadati S, et al (2023) Continuous conditional generative adversarial networks for data-driven modelling of geologic

CO 2 storage and plume evolution. Gas Science and Engineering 115:204982. https://doi.org/10.1016/j.jgsce.2023.204982, URL https://linkinghub.elsevier.com/retrieve/pii/S2949908923001103

Sun R, Pan H, Xiong H, et al (2023) Physical-informed deep learning framework for CO 2 -injected EOR compositional simulation. Engineering Applications of Artificial Intelligence 126:106742. https://doi.org/10.1016/j.engappai.2023.106742, URL https://linkinghub.elsevier.com/retrieve/pii/S0952197623009260

Tang H, Durlofsky LJ (2025) Graph network surrogate model for optimizing the placement of horizontal injection wells for CO2 storage. International Journal of Greenhouse Gas Control 145:104404. https://doi.org/10.1016/j.ijggc.2025.104404, URL https://linkinghub.elsevier.com/retrieve/pii/S1750583625001021

Tang M, Liu Y, Durlofsky LJ (2020) A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. Journal of Computational Physics 413:109456. https://doi.org/10.1016/j.jcp.2020.109456, URL https://linkinghub.elsevier.com/retrieve/pii/S0021999120302308

Tang M, Liu Y, Durlofsky LJ (2021) Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3D subsurface flow. Computer Methods in Applied Mechanics and Engineering 376:113636. https://doi.org/10.1016/j.cma.2020.113636, URL https://linkinghub.elsevier.com/retrieve/pii/S0045782520308215

Tang M, Ju X, Durlofsky LJ (2022) Deep-learning-based coupled flow-geomechanics surrogate model for CO2 sequestration. International Journal of Greenhouse Gas Control 118:103692. https://doi.org/10.1016/j.ijggc.2022.103692, URL https://linkinghub.elsevier.com/retrieve/pii/S1750583622001104

Tariq Z, Xu Z, Gudala M, et al (2023a) Deep-Learning-Based Surrogate Model to Predict CO2 Saturation Front in Highly Heterogeneous Naturally Fractured Reservoirs: A Discrete Fracture Network Approach. In: SPE Reservoir Characterisation and Simulation Conference and Exhibition. SPE, Abu Dhabi, UAE, p D021S007R006, https://doi.org/10.2118/212658-MS, URL https://onepetro.org/SPERCSC/proceedings/22RCSC/22RCSC/D021S007R006/515792

Tariq Z, Yan B, Sun S (2023b) Application of Image Processing Techniques in Deep-Learning Workflow to Predict CO2 Storage in Highly Heterogeneous Naturally Fractured Reservoirs: A Discrete Fracture Network Approach. In: Middle East Oil, Gas and Geosciences Show. SPE, Manama, Bahrain, p D011S033R006, https://doi.org/10.2118/213359-MS, URL https://onepetro.org/SPEMEOS/proceedings/23MEOS/23MEOS/D011S033R006/517261

Voskov DV, Henley H, Lucia A (2017) Fully compositional multi-scale reservoir simulation of various CO 2 sequestration mechanisms. Computers & Chemical Engineering 96:183–195. https://doi.org/10.1016/j.compchemeng.2016.09.021, URL

https://linkinghub.elsevier.com/retrieve/pii/S0098135416303076

Wang H, Guo R, Dalton LE, et al (2024a) Comparative Assessment of U-Net-Based Deep Learning Models for Segmenting Microfractures and Pore Spaces in Digital Rocks. SPE Journal 29(11):5779–5791. https://doi.org/10.2118/215117-PA, URL https://onepetro.org/SJ/article/29/11/5779/555191/Comparative-Assessment-of-U-Net-Based-Deep

Wang Y, Chu H, Lyu X (2024b) Deep learning in CO2 geological utilization and storage: Recent advances and perspectives. Adv Geo-Energy Res 13(3):161–165. https://doi.org/10.46690/ager.2024.09.01, URL https://www.yandy-ager.com/index.php/ager/article/view/1350

Wen G, Hay C, Benson SM (2021a) CCSNet: A deep learning modeling suite for CO 2 storage. Advances in Water Resources 155:104009. https://doi.org/10.1016/j.advwatres.2021.104009, URL https://linkinghub.elsevier.com/retrieve/pii/S0309170821001640

Wen G, Tang M, Benson SM (2021b) Towards a predictor for CO2 plume migration using deep neural networks. International Journal of Greenhouse Gas Control 105:103223. https://doi.org/10.1016/j.ijggc.2020.103223, URL https://linkinghub.elsevier.com/retrieve/pii/S1750583620306484

Wen G, Li Z, Azizzadenesheli K, et al (2022) U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. Advances in Water Resources 163:104180. https://doi.org/10.1016/j.advwatres.2022.104180, URL https://linkinghub.elsevier.com/retrieve/pii/S0309170822000562

Wen G, Li Z, Long Q, et al (2023) Real-time high-resolution CO_2 geological storage prediction using nested Fourier neural operators. Energy Environ Sci 16(4):1732–1741. https://doi.org/10.1039/D2EE04204E, URL https://xlink.rsc.org/?DOI=D2EE04204E

Xiao J, Wei C, Lin D, et al (2025) Dynamic CO2 Leakage Risk Assessment of the First Chinese CCUS-EGR Pilot Project in the Maokou Carbonate Gas Reservoir in the Wolonghe Gas Field. Energies 18(17):4478. https://doi.org/10.3390/en18174478, URL https://www.mdpi.com/1996-1073/18/17/4478

Yan B, Chen B, Robert Harp D, et al (2022) A robust deep learning workflow to predict multiphase flow behavior during geological C O 2 sequestration injection and Post-Injection periods. Journal of Hydrology 607:127542. https://doi.org/10.1016/j.jhydrol.2022.127542, URL https://linkinghub.elsevier.com/retrieve/pii/S0022169422001172

Zainulin SM, Storozheva AE (2025) Evaluation of the effectiveness of the CO2-EOR method for enhanced oil recovery in the offshore Sakhalin Island field. IZVESTIYA 336(3):51–62. https://doi.org/10.18799/24131830/2025/3/4656, URL

https://izvestiya.tpu.ru/archive/article/view/4656

Zhang D, Li H (2025) Efficient prediction of subsurface CO2 plume migration in heterogeneous reservoirs using the TransUNet deep learning network. Geoenergy Science and Engineering 255:214106. https://doi.org/10.1016/j.geoen.2025.214106, URL https://linkinghub.elsevier.com/retrieve/pii/S2949891025004646

Zhang X, Xiao T, Moodie N, et al (2025) A Dynamic and Constrained Machine Learning Framework for CO_2 Saturation Prediction in Geological Carbon Storage Reservoirs. Journal of Geophysical Research: Machine Learning and Computation 2(4):e2025JH000772. https://doi.org/10.1029/2025JH000772, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000772

Zheng Z, Di Y, Yu E (2023) DL-KF modeling for acceleration of flash calculations in phase equilibrium using deep learning methods. Front Earth Sci 10:1041589. https://doi.org/10.3389/feart.2022.1041589, URL https://www.frontiersin.org/articles/10.3389/feart.2022.1041589/full

Zhong Z, Sun AY, Jeong H (2019) Predicting CO_2 Plume Migration in Heterogeneous Formations Using Conditional Deep Convolutional Generative Adversarial Network. Water Resources Research 55(7):5830–5851. https://doi.org/10.1029/2018WR024592, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018WR024592