

Beyond Physics Residuals: Local Conservation, Symmetry and Physics-Distance Reliability in Two-Phase Reservoir Surrogates under Distribution Shift

Franck-Hilaire Essiagne*, Moussa Camara, Kouassi Louis Kra

Laboratoire des Sciences Géographiques du Génie Civil et des Géosciences, Ecole Supérieure de Chimie du Pétrole et de l'Energie (ESCPE), Institut National Polytechnique Félix HOUPOUET-BOIGNY (INP-HB), Yamoussoukro, Côte d'Ivoire
Email: *franck.essiagne@inphb.ci

How to cite this paper: Essiagne, F.-H., Camara, M. and Kra, K.L. (2026) Beyond Physics Residuals: Local Conservation, Symmetry and Physics-Distance Reliability in Two-Phase Reservoir Surrogates under Distribution Shift. *Journal of Analytical Sciences, Methods and Instrumentation*, 16, 21-41.

<https://doi.org/10.4236/jasmi.2026.162002>

Received: June 2, 2026

Accepted: June 26, 2026

Published: June 29, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Physics-informed reservoir surrogates are often evaluated by state error or equation residual alone, although neither establishes local conservation, symmetry consistency, or reliability under distribution shift. We separate these properties in a controlled two-dimensional incompressible oil-water benchmark. A flux-predictive U-Net outputs pressure, saturation increment, and oriented intercell total fluxes. Soft finite-volume residual regularization, exact minimum-norm cellwise flux projection, and horizontal/vertical reflection canonicalization are varied in a three-seed $2 \times 2 \times 2$ factorial design across seven IID/OOD regimes. Exact local projection reduces saturation error in every regime and enforces cellwise balance to numerical precision but does not guarantee accurate pressure or long-horizon dynamics. Canonicalization gives large additional gains only for symmetry-equivalent well shifts: under combined geology plus symmetry OOD, saturation nRMSE decreases from 0.0216 without canonicalization to 0.00994 with the full method, whereas its incremental benefit is not significant for non-symmetry well or hard compound OOD after Holm correction. Converged FNO and PINO-style baselines remain less accurate in saturation/front metrics. Projector-forensics controls show learned fluxes reduce saturation error by 25% - 56% relative to a zero-flux projector-only baseline, while reference simulator fluxes require only approximately 10^{-6} relative correction. On a separately generated 56-case hold-out, pre-projection correction magnitude predicts failure (Spearman rho = 0.648; AUROC = 0.894 for saturation nRMSE > 0.025; severe-OOD AUROC = 0.971). The results establish that conservation, symmetry, and residual consistency address distinct failure modes, and that conservation correction

provides an interpretable physics-distance reliability signal.

Keywords

Reservoir Surrogate, Two-Phase Flow, Finite-Volume Conservation, Symmetry Canonicalization, Physics-Informed Machine Learning, Out-of-Distribution, Reliability

1. Introduction

Repeated simulation of multiphase flow through heterogeneous porous media is central to reservoir management, uncertainty quantification, optimization, history matching and subsurface energy applications. Classical reservoir simulators remain the numerical reference for resolving mass balance and nonlinear constitutive coupling [1], while physics-informed machine learning and operator-learning methods provide increasingly capable reduced-order and many-query alternatives [2]-[6].

Recent operator-learning work has expanded from transfer and out-of-distribution analysis to boundary-informed operators and reservoir-specific multiphase surrogates [7]-[18]. In porous-media flow, finite-volume physics, theory-guided convolutional models, adaptive physics-informed networks, DeepONet variants and reservoir-specific neural surrogates have been used to represent pressure, saturation, wells, geological heterogeneity and control variation [19]-[35]. These developments establish strong accuracy baselines, but they do not by themselves make soft residual reduction, exact local conservation and admissible symmetry equivalent notions.

Three distinctions are therefore central to this study. First, a finite physics penalty can reduce a residual without making conservation exact; discretization-aware and conservative learning methods illustrate both the value and the limits of residual-based enforcement [36]-[43]. Second, exact conservation can be imposed on an inaccurate flux, so a conservative prediction may still place a saturation front incorrectly. Third, equivariance applies only to transformations that are genuine symmetries of the complete physical problem; geometric and canonicalization perspectives make this distinction explicit [44]-[47]. Reliability and uncertainty under distribution shift must consequently be evaluated separately from physical feasibility [48]-[50]. The present study tests these distinctions directly in transient two-phase reservoir rollouts.

The present study is motivated by a failure observed in a preliminary surrogate: a global water-inventory correction reduced material-balance drift to near zero, and soft finite-volume residuals were also improved, yet a reversed injector-producer geometry still produced a severely misplaced saturation front until the complete physical problem was reflection-canonicalized. That observation suggested that state accuracy, equation consistency, conservation and symmetry act on

different error components rather than forming a single scalar notion of “physics”. The study was therefore redesigned around falsifiable constraint-sufficiency tests instead of incremental network tuning.

The contribution is fourfold. i) We use a flux-predictive surrogate followed by an exact cellwise finite-volume projection and verify with oracle and zero-flux controls that the projection is a correction layer rather than a hidden substitute simulator. ii) We perform a matched three-seed factorial experiment that separates soft physics, hard local conservation and reflection canonicalization across seven distribution-shift regimes. iii) We extend autoregressive evaluation from the 80-step training horizon to 240 steps to distinguish conserved rollouts from dynamically accurate rollouts. iv) We test the pre-projection correction norm as a physics-distance reliability indicator on a separately generated, threshold-frozen holdout. The resulting message is deliberately narrower than a universal robustness claim: structural constraints are useful because they remove specific failure modes, and their value depends on which structure the shift preserves.

2. Governing Problem and Structural Constraints

2.1. Two-Phase Finite-Volume Benchmark

We consider incompressible, immiscible oil-water flow on a two-dimensional Cartesian grid with no-flow external boundaries. Let p denote pressure, S_w water saturation, K absolute permeability, ϕ porosity, q the signed well source/sink field and $\lambda_t = \lambda_w + \lambda_o$ total mobility. Neglecting gravity and capillarity, the pressure equation is the total-volume balance

$$\nabla \cdot (K \lambda_t (S_w) \nabla p) + q = 0$$

and the water equation can be written in conservative form

$$\phi \frac{\partial S_w}{\partial t} + \nabla \cdot [f_w(S_w) \mathbf{u}_t] = q_w, \quad f_w = \frac{\lambda_w}{\lambda_t}$$

Quadratic Corey-type normalized relative permeabilities are used with connate water and residual oil saturations of 0.15, water viscosity 1, oil viscosity 5 and porosity 0.20. The numerical reference uses a cell-centered finite-volume pressure solve and an upwind explicit saturation update. The grid has 24×24 cells, $\Delta t = 0.005$ and 80 steps for the primary data set. Injector and producer completions each occupy 3×3 cell blocks. Heterogeneous log permeability is generated from Gaussian random fields with regime-dependent correlation and anisotropy parameters.

2.2. Exact Local Conservation as an Affine Projection

The learned model predicts oriented internal total fluxes $\hat{\mathbf{q}}$ together with pressure and saturation increment. Let $\mathbf{A}$ be the cell-face incidence/divergence matrix and $\mathbf{b}$ the balanced cell source vector induced by the wells. The physically admissible total-flux set is the affine manifold $\mathbf{A}\mathbf{q} = \mathbf{b}$. We use the minimum-norm Euclidean projection

$$\mathbf{q}^* = \arg \min_{\mathbf{q}} \|\mathbf{q} - \hat{\mathbf{q}}\|_2^2 \quad \text{subject to } A\mathbf{q} = \mathbf{b}$$

For the regular Cartesian graph, the solution is obtained from the sparse normal system $AA^\top \lambda = A\hat{\mathbf{q}} - \mathbf{b}$ with one gauge degree of freedom removed, followed by $\mathbf{q}^* = \hat{\mathbf{q}} - A^\top \lambda$. Water flux is reconstructed with upwind fractional flow and the next saturation is obtained from the projected water balance. Because $S_o = 1 - S_w$ and the fluids are incompressible, exact total balance combined with the water update also determines the oil inventory balance. The projection is therefore stronger than the preliminary global-inventory correction: it removes cellwise divergence inconsistency rather than only reservoir-integrated drift.

Source-sink feasibility is enforced when each case is generated. The total injection rate is distributed uniformly over the 3×3 injector block and the same total magnitude is withdrawn uniformly over the 3×3 producer block, so the discrete source vector sums to zero up to floating-point roundoff. Horizontal internal flux is positive from cell (i, j) to $(i, j+1)$, and vertical internal flux is positive from (i, j) to $(i+1, j)$; the incidence matrix therefore contributes $+1$ to the oriented upstream cell and -1 to its neighbor. The reference pressure system is solved by preconditioned conjugate gradients with relative tolerance 10^{-9} , absolute tolerance 10^{-11} and 2000 maximum iterations, with a sparse direct fallback. The conservation projector removes one gauge degree of freedom and solves the reduced graph-Laplacian system by sparse factorization; consequently, the reported post-projection local-balance residual (about 10^{-15} in the frozen experiments) reflects the direct sparse solve and floating-point arithmetic rather than the pressure-solver stopping tolerance.

The hard projection changes face fluxes but does not change the predicted pressure. Pressure nRMSE is therefore exactly unchanged by applying the flux projector to a fixed network output. Because the corrected flux is not recomputed from the unchanged pressure field, the projection does not guarantee Darcy pressure-flux consistency and may change the Darcy residual. We therefore interpret the hard layer only as an exact divergence/conservation constraint; Darcy consistency remains a separate soft-training and diagnostic quantity. This distinction is important in the results, where exact conservation can coexist with appreciable pressure error.

We record the relative correction

$$E_{\text{proj}} = \frac{\|\mathbf{q}^* - \hat{\mathbf{q}}\|_2}{\|\hat{\mathbf{q}}\|_2 + \varepsilon}$$

which later serves as a candidate reliability score. Importantly, E_{proj} is not assumed to be uncertainty a priori; its relation to true state error is evaluated on an untouched holdout.

2.3. Exact Reflection Canonicalization

The controlled benchmark admits exact horizontal and vertical reflections when the complete problem is transformed: permeability, current saturation, well masks and

signed source field are reflected together, the network is evaluated in a canonical injector-producer orientation, and predictions are mapped back to physical coordinates. For an admissible discrete reflection g with transform T_g , the ideal solution operator G satisfies $G(T_g a) = T_g G(a)$. Canonicalization implements this equivariance without changing the backbone architecture. We do not interpret this as invariance to arbitrary well geometry, continuous rotation, gravity changes or other transformations that are not exact symmetries of the stated benchmark.

3. Surrogate Models and Training

3.1. Flux-Predictive Convolutional Backbone

The principal surrogate is a compact U-Net with 272,908 trainable parameters. Seven input channels are used after coordinate augmentation: normalized log permeability, current water saturation, normalized signed well source, injector mask, producer mask and two Cartesian coordinate channels. Four output channels are decoded as normalized pressure, saturation increment, horizontal face flux and vertical face flux. Flux targets are computed from the reference finite-volume pressure and mobility fields, so the learned flux has a direct numerical meaning rather than being an unconstrained latent variable. **Figure 1** summarizes the complete canonicalize-predict-project-update inference pathway.

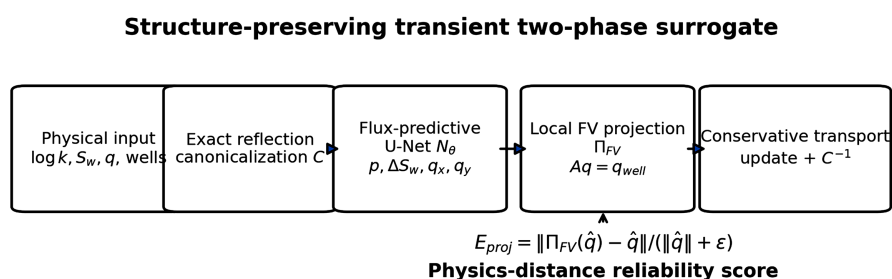


Figure 1. Structure-preserving surrogate. The network predicts pressure, saturation increment and oriented face fluxes in canonical coordinates. The hard finite-volume projector enforces cellwise total balance, and its correction magnitude is retained as a physics-distance reliability score.

Inference uses two distinct saturation pathways. For unprojected variants, the network-predicted saturation increment is added to the current saturation and clipped to the admissible interval. For every hard-projection variant, that predicted increment is not used in the recursive state update after projection and is never added a second time. Instead, the projected total face flux is partitioned with upwind fractional flow and the next saturation is recomputed directly from the discrete water balance. The pressure head remains the network prediction and is not recomputed by the projector. Thus the hard projector replaces the learned saturation-increment update while retaining the learned pressure and corrected face-flux information.

Training uses 60 independent reservoir realizations and validation uses 10. Only every eighth simulator state is labeled, producing 600 supervised training

transitions and 100 validation transitions. The data objective assigns equal weight to normalized pressure and saturation-increment losses and higher weight to the two flux channels. Three independently initialized data models use seeds 20260903 - 20260905.

3.2. Soft Finite-Volume Physics

The soft-physics variant adds differentiable finite-volume penalties for total balance, water transport and Darcy consistency of the predicted flux. With $\mathcal{L}_D$ denoting the supervised pressure/saturation/flux loss, the training objective is

$$\mathcal{L} = \mathcal{L}_D + 0.02 \mathcal{L}_{\text{balance}} + 0.05 \mathcal{L}_{\text{transport}} + 0.005 \mathcal{L}_{\text{Darcy}}$$

These weights were fixed before the final factorial analysis. The purpose of the soft terms is not to guarantee equality but to test whether residual regularization adds predictive value once flux supervision and a hard inference-time projection are already present.

3.3. FNO and PINO-Style External Baselines

External operator baselines use the same five physical input channels plus coordinates and predict pressure and saturation increment. The FNO has four spectral layers, width 16 and eight retained Fourier modes per direction (133,506 parameters). The PINO-style baseline uses the identical FNO backbone and adds the same discrete pressure-balance and water-transport residuals used for evaluation. Because this implementation uses finite-volume residuals at the training resolution, consistent with discretization-guided physics-informed approaches [18] [19] [36] [37], we use the term PINO-style descriptively rather than claiming reproduction of a particular published PINO formulation.

All three FNO and three PINO-style runs were first trained for 30 epochs, then warm-started under a fixed validation-only convergence protocol using ReduceLROnPlateau. Best checkpoints occurred at epochs 48, 58 and 42 for FNO and 64, 65 and 54 for PINO-style. No IID or OOD test result was used for checkpoint selection. Incomplete parameter-matched and from-scratch pilots were excluded from the frozen evidence rather than selectively reported.

4. Experimental Design

4.1. Distribution-Shift Hierarchy

The primary evaluation contains seven mutually defined regimes, each with 12 independent reservoirs. The distinction between symmetry-equivalent and genuinely non-equivalent shifts is central to interpretation. **Table 1** summarizes the seven regimes and their scientific roles.

The permeability generator is parameterized by a Gaussian-filter correlation scale, an anisotropy ratio and an orientation. Training, validation and IID cases use correlation scale 1.5 - 4.0 cells, anisotropy 0.7 - 1.6 and orientation 0°. Geology-OOD cases use correlation scale 5.0 - 7.5 cells, anisotropy 2.0 - 3.0 and

orientation 0° or 90° . In all regimes the standardized Gaussian field is rescaled by an independently sampled log-permeability amplitude in 0.7 - 1.3 and centered at $\log(100)$ in the simulator scaling.

Training-support injector centers are $\{(3,3), (3,5), (5,3), (5,5)\}$ and producer centers are $\{(18,18), (18,20), (20,18), (20,20)\}$. Reflection-equivalent well OOD uses the pairs $((3,20), (20,3)), ((5,20), (18,3)), ((3,18), (20,5))$ and $((5,18), (18,5))$. Genuine non-symmetry well OOD cycles through $((3,12), (20,12)), ((12,3), (12,20)), ((7,3), (16,20)), ((3,9), (20,15)), ((8,5), (15,18))$ and $((5,12), (18,8))$. Training-support, IID, geology-OOD and well-OOD rates are sampled uniformly from 0.08 - 0.16 injected pore volumes per unit simulation time; control OOD uses 0.20 - 0.28. Hard compound OOD combines the geology-OOD range, the non-symmetry well layouts and the 0.20 - 0.28 rate range. Each evaluation regime contains 12 independently generated reservoirs.

Table 1. Distribution-shift hierarchy used for the primary 84-reservoir evaluation.

Regime	Shift	Scientific role
IID	corr 1.5 - 4.0; anis. 0.7 - 1.6; 0° ; training-support wells; rate 0.08 - 0.16 PV/time	In-distribution accuracy and baseline drift
Geology OOD	corr 5.0 - 7.5; anis. 2.0 - 3.0; $0^\circ/90^\circ$; training-support wells; rate 0.08 - 0.16	Material-property extrapolation without geometry change
Symmetry-well OOD	corr 1.5 - 4.0; anis. 0.7 - 1.6; exact reflected well pairs; rate 0.08 - 0.16	Tests exact equivariant transfer
Combined symmetry OOD	geology-OOD statistics + reflected well pairs; rate 0.08 - 0.16	Tests symmetry under simultaneous geology shift
Non-symmetry well OOD	training-support geology + six interior/edge well layouts; rate 0.08 - 0.16	Tests genuine geometric extrapolation
Control OOD	training-support geology/wells; rate 0.20 - 0.28	Tests operating-condition extrapolation
Hard compound OOD	geology-OOD statistics + non-symmetry layouts + rate 0.20 - 0.28	Combined genuine shift

4.2. Factorial Ablation and Metrics

The matched three-seed factorial contains eight methods: Data, PI, Data + local projection, PI + local projection, symmetry-canonicalized Data, symmetry-canonicalized PI, SC + local projection and full SC-CPI. Every seed-method combination is evaluated on the same physical reservoirs, giving 2016 seed-case-method observations. Saturation error is normalized by the mobile saturation interval $1 - S_{or} - S_{wc}$, pressure error by the reference pressure standard deviation, and water-front fidelity is measured with intersection over union at $S_w > 0.20$. Projection correction is averaged across the autoregressive rollout.

The primary inferential unit is the physical reservoir, not an individual seed prediction. For each prespecified contrast, the metric is first averaged over the three independently trained seeds within each of the 12 physical reservoirs, after

which a two-sided paired Wilcoxon signed-rank test is applied at $N = 12$. Holm family-wise correction is then applied across the same 42 prespecified contrasts. As a sensitivity analysis, the full seed-level factorial data are also analyzed with physical reservoir as the clustering unit, so the three predictions belonging to one reservoir are not treated as independent replicates. Effect sizes and corrected physical-case-level tests are emphasized throughout.

4.3. Long-Horizon, Projector and Reliability Stress Tests

Three additional experiments address alternative explanations. First, a projector-forensics study applies the hard projection to reference simulator fluxes and to a zero-flux baseline. If the projector were behaving as a hidden approximate simulator, zero-flux projection would approach the learned model and reference fluxes would require a large change; the opposite behavior is expected for a genuine correction layer. Second, independent numerical references are generated to 240 steps, three times the training rollout, for IID, combined symmetry OOD and hard compound OOD. Third, reliability is evaluated using one scalar score per physical reservoir: the arithmetic mean of the relative pre-to-post projection correction over the 80-step autoregressive rollout. A saturation nRMSE threshold of 0.025 was pre-specified on the development set as a low-single-digit mobile-saturation error criterion for separating nominal from elevated-error trajectories. The correction threshold was then frozen before a separately generated 56-case holdout (eight cases per regime) was evaluated. On that holdout, 15 of 56 reservoirs (26.8%) exceed the 0.025 failure threshold. Bootstrap confidence intervals resample physical reservoirs with replacement, not rollout steps or seed predictions. The holdout is not used for training, model selection, projector tuning or threshold choice.

The final reproducibility audit hashes geology arrays and their horizontal/vertical reflection orbits. It found no exact or reflection-equivalent training geology in evaluation sets and no duplicated geologies across the audited scopes. The complete data freeze contains 60 training, 10 validation, 84 primary evaluation and 56 reliability-holdout reservoirs.

4.4. Reproducibility Environment and Frozen Model Artifact

The revised reproducibility archive pins the computational environment used for the frozen analysis (Python 3.13.5, NumPy 2.3.5, SciPy 1.17.0, PyTorch 2.10.0 CPU, scikit-learn 1.8.0, pandas 2.2.3 and statsmodels 0.14.6), documents all random seeds, architecture dimensions, training schedules, normalization constants and validation-selected operator epochs, and includes the simulator/data-generation code, evaluation and statistical-analysis scripts, the frozen per-case prediction/metric workbook, figure-generation assets and SHA-256 manifests. Because the final strong-study checkpoint binaries were not retained in the original submission archive, the revised archive identifies the frozen per-case outputs plus complete architecture/training/checkpoint-selection metadata as the equivalent reproducible model artifact rather than claiming that unavailable binaries are present.

5. Results

5.1. Exact Local Projection Is the Dominant Transport Stabilizer

Figure 2 and **Table 2** summarize the three-seed factorial results. Hard local conservation reduces saturation error in every distribution-shift regime. On IID cases, mean saturation nRMSE decreases from 0.0240 for Data to 0.0153 for Data + projection. On geology OOD the reduction is 0.0305 to 0.0161, and on hard compound OOD it is 0.0949 to 0.0397. The corresponding hard-conservation contrasts remain significant after Holm correction in all seven regimes. The result is therefore not confined to the deliberately symmetry-aligned test.

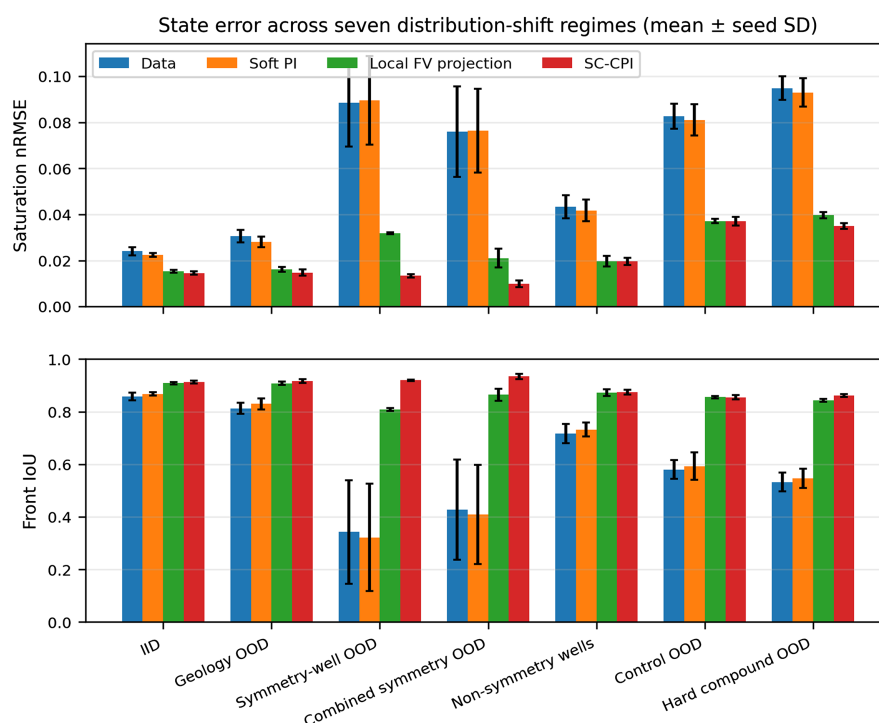


Figure 2. Three-seed factorial performance across seven regimes. Bars show seed means and error bars show seed standard deviation. Local FV projection consistently reduces saturation error; symmetry adds its largest gains where the physical shift is an exact reflection.

Table 2. Mean saturation and front performance over three training seeds and 12 physical cases per regime.

Regime	Data S nRMSE	Local CP	SC-CP	SC-CPI	SC-CPI front IoU
IID	0.0240	0.0153	0.0153	0.0145	0.9129
geology OOD	0.0305	0.0161	0.0161	0.0148	0.9170
symmetry-well OOD	0.0885	0.0319	0.0142	0.0133	0.9193
combined symmetry OOD	0.0760	0.0210	0.0104	0.0099	0.9342
non-symmetry well OOD	0.0433	0.0197	0.0201	0.0196	0.8745
control OOD	0.0826	0.0371	0.0371	0.0371	0.8550
hard compound OOD	0.0949	0.0397	0.0356	0.0351	0.8619

The improvement is especially large for displaced wells. In symmetry-well OOD, Data gives 0.0885 saturation nRMSE and 0.342 front IoU, while Data + local projection gives 0.0319 and 0.808. This demonstrates that local balance is not merely an accounting correction: by constraining the learned intercell flux network to a source-consistent divergence field, it materially changes transport trajectories.

At the same time, conservation is not sufficient for full state correctness. Pressure nRMSE remains high for noncanonicalized reversed-well models even after projection because the projection adjusts fluxes to satisfy divergence but does not reconstruct a correct pressure solution. This separation between conservative transport and constitutively accurate pressure is central: an exact equality removes one infeasible error component but cannot eliminate all tangent-to-manifold model error. Because pressure is held fixed by design, hard projection cannot improve pressure error for a given raw network prediction; any pressure improvement must come from the learned model or symmetry transformation rather than the conservation solve.

5.2. Equivariance Is Not General OOD Robustness

Figure 3 isolates the incremental symmetry effect after exact local conservation. Canonicalization produces its strongest improvement when the shift is actually generated by an admitted reflection. After local projection, adding symmetry to the PI model reduces saturation nRMSE from 0.0317 to 0.0133 in symmetry-well OOD and from 0.0216 to 0.0099 in combined symmetry OOD. With physical reservoir as the paired inferential unit and Holm correction across the 42 prespecified contrasts, the adjusted p-value is 0.0205 for both comparisons.

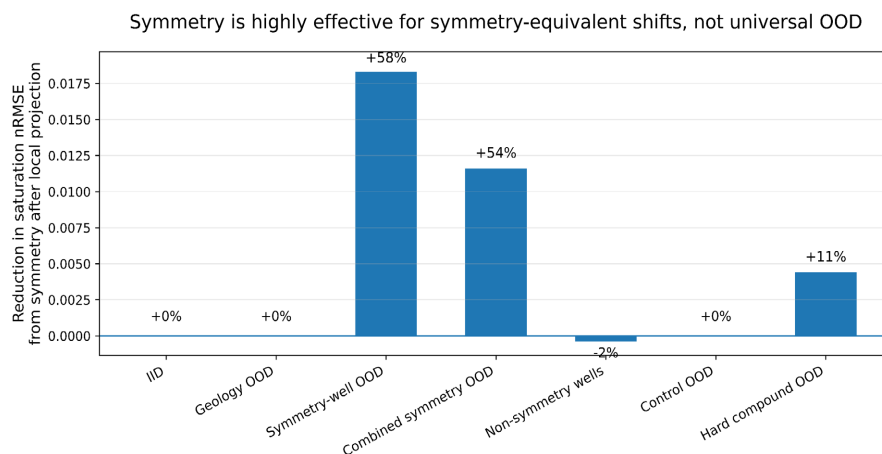


Figure 3. Incremental effect of reflection canonicalization after hard local conservation. The benefit is large for symmetry-equivalent well shifts, approximately zero for IID/control cases, and small or absent for genuinely non-equivalent geometries.

The same statement does not extend to genuinely new geometry. For non-symmetry well OOD the full canonicalized model has 0.0196 saturation nRMSE versus 0.0193 without canonicalization, and the corrected comparison is not significant.

For the hard compound shift the numerical reduction is modest (0.0394 to 0.0351) and again does not survive Holm correction. The proper interpretation is therefore exact zero-label transfer within a physical equivalence class, not geometry-agnostic extrapolation.

5.3. Soft Residual Physics Is Not Automatically Additive

Once the model is supervised on fluxes and exact conservation is imposed structurally, the additional soft finite-volume loss contributes much less to state accuracy than the hard projector. Across the prespecified paired tests, most soft-physics-after-conservation contrasts are not significant after Holm adjustment. For example, hard compound OOD saturation nRMSE is 0.0397 for Data + projection and 0.0394 for PI + projection, while the canonicalized counterparts are 0.0356 and 0.0351. This does not imply that residual information is useless; rather, a lower residual is not equivalent to a more accurate OOD state once a stronger equality constraint is already enforced.

Figure 4 compares the validation-selected operator baselines with SC-CPI. The external neural-operator comparison reinforces this distinction. The converged FNO reaches mean saturation nRMSE 0.0343 on IID and 0.1111 on hard OOD; the PINO-style model reaches 0.0384 and 0.1168. In contrast, SC-CPI gives 0.0145 and 0.0351. PINO-style substantially lowers finite-volume residuals relative to FNO in several regimes, but it does not attain the best front or saturation accuracy. Physics residual and predictive fidelity should therefore be reported together rather than used as proxies for one another.

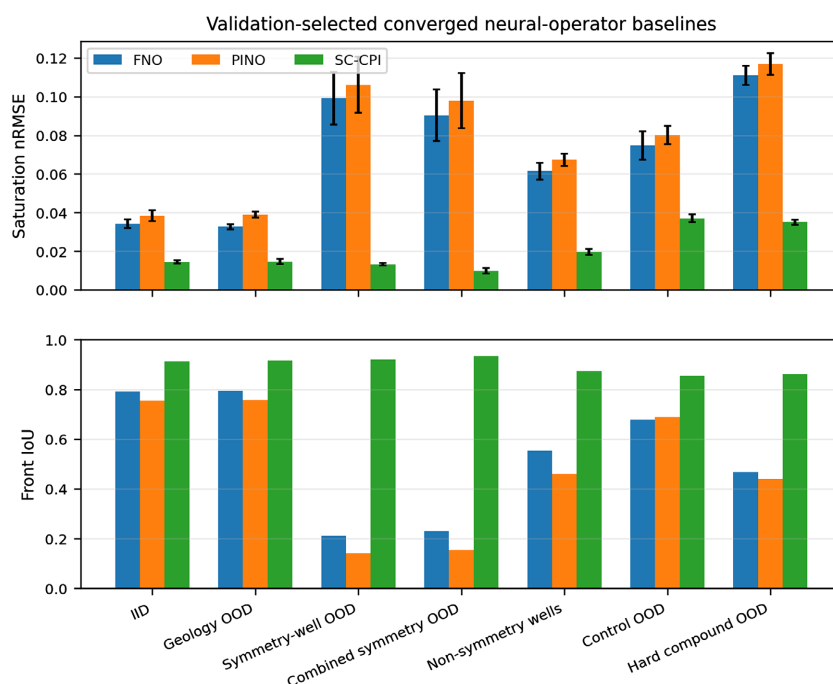


Figure 4. Validation-selected converged FNO and PINO-style baselines compared with SC-CPI. Operator checkpoints were selected using validation data only and then evaluated once on the seven regimes.

5.4. Projector Forensics: The Network Still Carries the Dynamics

Figure 5 reports the projector-forensics controls. The oracle-flux test validates the numerical projection. When reference simulator fluxes are projected, the relative correction is only 1.4×10^{-6} - 3.1×10^{-6} across regimes and the post-projection local residual is approximately 4×10^{-17} . By contrast, learned-flux corrections range from 0.177 on IID to 0.709 on hard OOD, showing that the network is close to but not generally on the conservation manifold.

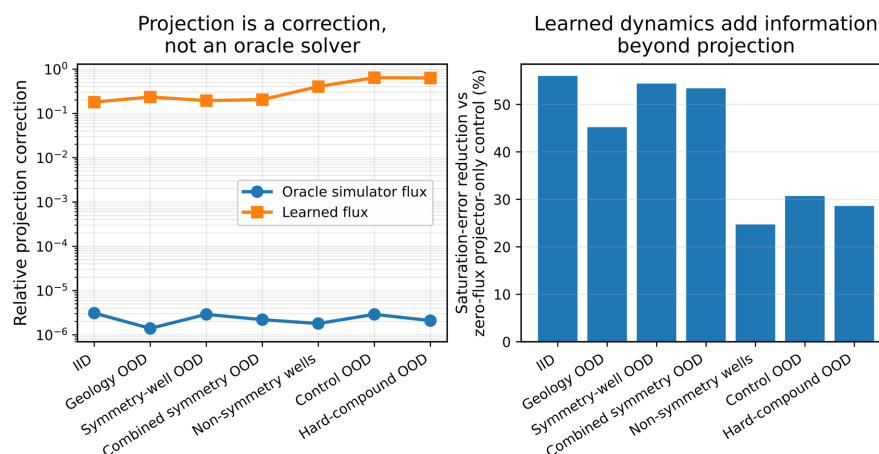


Figure 5. Projector forensics. Reference simulator fluxes are already on the conservation manifold up to numerical tolerance, whereas learned fluxes require finite correction. The learned + projected model materially outperforms a zero-flux/projector-only control.

A more demanding control starts the projector from zero intercell flux. Because a divergence constraint with wells contains a strong source-sink topology prior, this control is not trivial and can produce plausible displacement patterns. Nevertheless, learned + projected flux reduces saturation error relative to the zero-flux projector-only control by 56% on IID, 53% on combined symmetry OOD and 29% on hard OOD; reductions span approximately 25% - 56% across all seven regimes. Thus, the projector contributes meaningful physics, but the learned flux still contributes substantial state information.

5.5. Exact Conservation Slows Long-Horizon Drift but Does Not Eliminate Model Error

Figure 6 shows the beyond-training-horizon stress test. At 240 autoregressive steps, three times the original training rollout, the raw learned-flux model accumulates substantial saturation error and bound violations. On IID cases its saturation nRMSE reaches 0.147, compared with 0.071 after local projection. On combined symmetry OOD, the raw value is 0.273, local projection gives 0.101, and SC + projection gives 0.063. On hard OOD the corresponding values are 0.247, 0.127 and 0.110.

Projected trajectories preserve cellwise total balance at approximately 10^{-15} and exhibit zero post-projection saturation-bound violation in these stress tests, yet

their state error continues to grow. This is a useful negative result: exact conservation removes inventory drift but cannot repair accumulated constitutive or learned-flux error. The combined symmetry test also illustrates the diagnostic role of the correction norm. At step 240 the noncanonicalized projected model requires a correction of 1.072, whereas the canonicalized trajectory requires 0.279; the latter remains both closer to the manifold and more accurate.

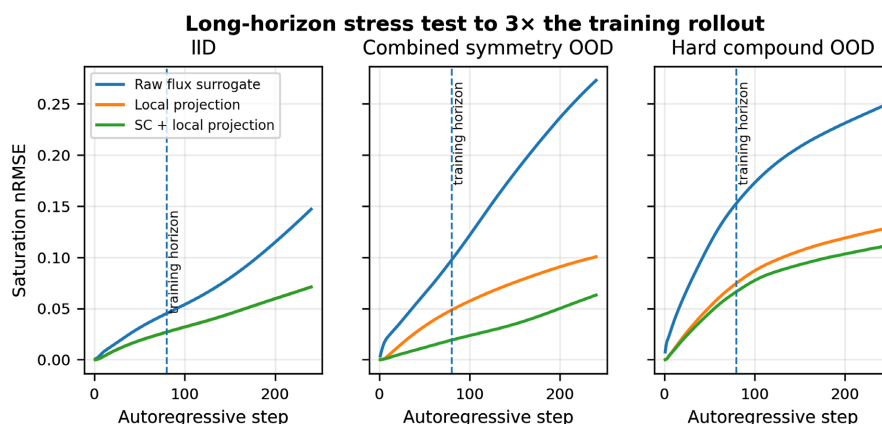


Figure 6. Long-horizon saturation error to 240 autoregressive steps. The dashed line marks the 80-step training horizon. Hard local conservation substantially suppresses drift but does not make long-time dynamics exact.

5.6. Projection Correction Predicts Failure on an Independent Holdout

Figure 7 summarizes the independent reliability holdout. The most practically useful result is that the pre-projection correction magnitude contains information about prediction risk. On the 84-case development evaluation, has Spearman correlation 0.700 with saturation error and AUROC 0.943 for detecting cases with saturation nRMSE > 0.025. A threshold of 0.4518 was then frozen before generating the independent holdout.



Figure 7. Independent post-threshold reliability holdout. Left: projection correction versus saturation error; dotted lines show the error criterion and frozen correction threshold. Middle: ROC for detecting saturation nRMSE > 0.025. Right: selective-prediction risk decreases as high-correction cases are rejected.

On the separately generated 56-case holdout, the correlation remains 0.648 (95% bootstrap CI 0.467 - 0.768; $p = 6.60 \times 10^{-8}$). Error-detection AUROC is

0.894 (95% CI 0.786 - 0.974) with AUPRC 0.819; severe genuine-OOD discrimination reaches AUROC 0.971. At the frozen threshold, sensitivity is 0.800, specificity 0.805 and negative predictive value 0.917. Thus, a low correction is not a formal guarantee of accuracy, but it is an empirically useful indicator that the raw learned flux lies close to the physically feasible manifold in the tested system.

The selective-prediction interpretation is operational: accepting lower-correction cases monotonically reduces mean saturation error in the development risk-coverage curve. A deployment system could therefore use E_{proj} as one signal for escalating difficult cases to a high-fidelity simulator rather than forcing the surrogate to answer every query. Because this study uses a controlled benchmark, the threshold itself is not transferable to another simulator or field model without recalibration.

5.7. Qualitative Displacement and Computational Cost

Figure 8 compares representative final-step saturation fields. The qualitative fields reflect the aggregate statistics: the exact reflection shift can be canonicalized into the learned orientation, whereas the hard case still requires genuine extrapolation. This difference is intentionally visible rather than hidden behind a single averaged OOD score.

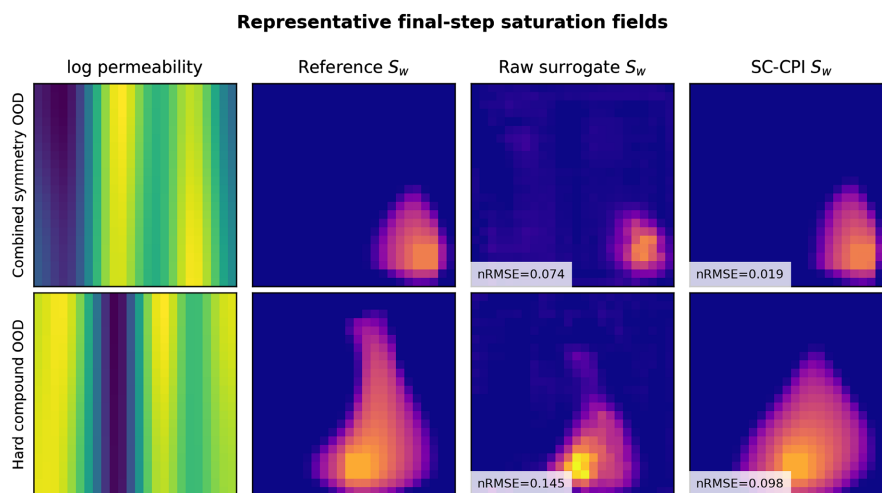


Figure 8. Representative final-step saturation fields for a symmetry-equivalent combined OOD case and a genuinely hard compound OOD case. SC-CPI markedly improves plume/front geometry but does not remove all error in the non-equivalent hard case.

The stronger model does not accelerate a single 24×24 CPU case: the SC flux + projection step requires 0.00773 s versus 0.00716 s for the compact finite-volume step, a speedup of only $0.93 \times$. Under batch-64 inference, however, per-case time is 0.000461 s, corresponding to $15.5 \times$ throughput relative to sequential finite-volume execution. We therefore make no single-query acceleration claim. The potential computational benefit is a many-query throughput benefit, which is the setting relevant to ensembles and optimization.

6. Discussion

6.1. A Hierarchy of Constraint Sufficiency

The experiments support a hierarchy rather than a scalar notion of physical correctness. Soft residual training makes certain equation violations less costly during learning but does not impose an equality. Hard projection guarantees a specific discrete conservation law but leaves pressure, constitutive consistency and front position free to remain wrong. Symmetry canonicalization guarantees consistency only along admitted group orbits. None of these operations can replace missing information about a genuinely new geological or control regime. This separation explains why combining physically motivated modules can improve robustness without implying that every module should improve every metric.

The zero-flux control is particularly instructive. A conservation projector with specified sources and sinks contains meaningful topology and can create a plausible flux field even without neural information. Without this control, a post-processing method could appear to validate a learned surrogate while actually doing much of the physical work itself. The oracle and zero-flux tests should therefore be considered useful audit patterns for future conservation-projected machine-learning studies.

6.2. Equivariance versus Extrapolation

The well-reversal experiment could easily be overinterpreted. Its dramatic improvement is not evidence that reflection canonicalization learns arbitrary well-layout changes. It shows that known physical redundancy can be removed exactly. When the well pattern is not related by that redundancy, the benefit becomes small. This distinction matters for reservoir surrogate benchmarking because an OOD split can be geometrically unfamiliar to the network while still being physically equivalent under a known group action. Such a test is valuable, but it should be labeled as symmetry transfer rather than general OOD.

6.3. Residual Minimization and Predictive Error Are Different Objectives

The PINO-style comparison and the soft-physics ablation both caution against using a PDE residual as a surrogate for state accuracy. A residual is an important diagnostic, especially when labels are scarce, but several conservative or low-residual states can still differ materially in transport geometry. In multiphase flow, front position and well response are particularly sensitive because errors in fractional-flow direction accumulate autoregressively. Reporting pressure/saturation error, front fidelity, local conservation and residual metrics together is therefore more informative than optimizing one quantity in isolation.

6.4. Physics-Distance as a Reliability Signal

E_{proj} has a direct physical interpretation: it is the magnitude of the correction required to return a raw learned flux to the discrete conservation manifold. Unlike

a generic latent-space distance, it is measured in the variables on which the hard physical equality acts. The independent-holdout result suggests that this quantity can identify confident-looking but physically distant extrapolations. It should not be treated as a calibrated probability of error. Rather, it can complement ensemble uncertainty or other epistemic indicators in a selective surrogate-simulator policy. A natural next study would compare active-learning acquisition based on projection distance with ensemble variance under a larger public simulator benchmark.

6.5. Limitations and Scope

The evidence is intentionally controlled. The simulator is two-dimensional, incompressible and immiscible; gravity and capillarity are omitted; the grid is regular; well controls are simplified; and the main benchmark uses one injector and one producer. The exact reflection canonicalizer is appropriate only for the admitted discrete symmetry and would need stabilizer-aware, weighted or learned treatments for broader groups [44]-[47]. Generalization to corner-point or irregular geometries would require geometry-aware representations and grid-consistent conservation operators [34] [36] [37]. Finally, the reliability threshold is an empirical decision rule validated on an independent holdout, not a universal uncertainty guarantee; broader calibration should be assessed with established scientific-machine-learning UQ methodologies [48]-[50].

These limitations also define a disciplined boundary for the current contribution. Recent work already addresses multiphase neural operators, finite-volume physics-informed learning, reservoir-specific surrogates, exact conservation, equivariance and scientific-machine-learning uncertainty [10]-[35] [42]-[50]. The novelty claimed here is therefore not the first use of FNOs, finite-volume residuals, projection, symmetry or uncertainty concepts in isolation. It is the controlled transient reservoir study of what soft physics, hard local conservation and admissible symmetry individually guarantee and fail to guarantee under distinct distribution shifts, together with independent validation of the conservation-correction magnitude as a physics-distance risk signal.

7. Conclusion

A flux-predictive, locally projected and reflection-canonicalized surrogate was used to test the sufficiency of three common forms of physical structure in transient two-phase reservoir learning. The final evidence supports five conclusions. Exact cellwise finite-volume projection is the most consistent transport stabilizer across the tested distribution shifts, but exact conservation does not guarantee an accurate pressure field or exact long-term dynamics. Reflection canonicalization provides strong zero-label transfer when the new configuration is genuinely symmetry-equivalent, but not when wells or controls define a different physical problem. Soft residual physics is useful diagnostically but is not automatically additive once flux supervision and a hard conservation equality are present. Projector forensics show that the learned flux carries substantial information beyond the

physical correction layer. Finally, the magnitude of the pre-projection correction predicts state failure on an independently generated holdout, providing an interpretable physics-distance signal for selective use of the surrogate. These results argue for evaluating reservoir surrogates by a vector of structural properties rather than by RMSE or residual alone.

Data and Code Availability

The revised submission package includes a frozen reproducibility archive containing the simulator and data-generation code, synthetic benchmark cases from the underlying study, exact architecture and training configurations, dependency versions, evaluation/statistical scripts, frozen per-case outputs, the final result workbook used for reported tables, publication figure assets, validation-selected operator checkpoint metadata and SHA-256 manifests. The original final strong-study checkpoint binaries were not retained; accordingly, the archive provides the frozen per-case outputs together with the complete configuration and checkpoint-selection metadata as the equivalent reproducible model artifact and states this limitation explicitly.

CRedit Authorship Contribution Statement

Franck-Hilaire Essiagne: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing original draft. Moussa Camara: Conceptualization, Methodology, Software, Validation. Kouassi Louis Kra: Supervision, Resources, Writing, review and editing, Conceptualization, Methodology, Software, Validation.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions

Franck-Hilaire Essiagne: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing—original draft. Kouassi Louis Kra: Supervision, Resources, Writing—review and editing, Conceptualization, Methodology, Software, Validation. Moussa Camara: Conceptualization, Methodology, Software, Validation.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work.

References

- [1] Rasmussen, A.F., Sandve, T.H., Bao, K., Lauser, A., Hove, J., Skaflestad, B., *et al.* (2021) The Open Porous Media Flow Reservoir Simulator. *Computers & Mathematics with*

- Applications*, **81**, 159-185. <https://doi.org/10.1016/j.camwa.2020.05.014>
- [2] Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *Journal of Computational Physics*, **378**, 686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>
 - [3] Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021) Physics-Informed Machine Learning. *Nature Reviews Physics*, **3**, 422-440. <https://doi.org/10.1038/s42254-021-00314-5>
 - [4] Lu, L., Jin, P., Pang, G., Zhang, Z. and Karniadakis, G.E. (2021) Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators. *Nature Machine Intelligence*, **3**, 218-229. <https://doi.org/10.1038/s42256-021-00302-5>
 - [5] Azizzadenesheli, K., Kovachki, N., Li, Z., Liu-Schiaffini, M., Kossaifi, J. and Anandkumar, A. (2024) Neural Operators for Accelerating Scientific Simulations and Design. *Nature Reviews Physics*, **6**, 320-328. <https://doi.org/10.1038/s42254-024-00712-5>
 - [6] Subedi, U. and Tewari, A. (2026) Operator Learning: A Statistical Perspective. *Annual Review of Statistics and Its Application*, **13**, 123-148. <https://doi.org/10.1146/annurev-statistics-042424-070908>
 - [7] Goswami, S., Kontolati, K., Shields, M.D. and Karniadakis, G.E. (2022) Deep Transfer Operator Learning for Partial Differential Equations under Conditional Shift. *Nature Machine Intelligence*, **4**, 1155-1164. <https://doi.org/10.1038/s42256-022-00569-2>
 - [8] Lara Benitez, J.A., Furuya, T., Faucher, F., Kratsios, A., Tricoche, X. and de Hoop, M.V. (2024) Out-of-Distributional Risk Bounds for Neural Operators with Applications to the Helmholtz Equation. *Journal of Computational Physics*, **513**, Article 113168. <https://doi.org/10.1016/j.jcp.2024.113168>
 - [9] Fang, Z., Wang, S. and Perdikaris, P. (2024) Learning Only on Boundaries: A Physics-Informed Neural Operator for Solving Parametric Partial Differential Equations in Complex Geometries. *Neural Computation*, **36**, 475-498. https://doi.org/10.1162/neco_a_01647
 - [10] Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A. and Benson, S.M. (2022) U-FNO—An Enhanced Fourier Neural Operator-Based Deep-Learning Model for Multiphase Flow. *Advances in Water Resources*, **163**, Article 104180. <https://doi.org/10.1016/j.advwatres.2022.104180>
 - [11] Jiang, Z., Zhu, M. and Lu, L. (2024) Fourier-MIONet: Fourier-Enhanced Multiple-Input Neural Operators for Multiphase Modeling of Geological Carbon Sequestration. *Reliability Engineering & System Safety*, **251**, Article 110392. <https://doi.org/10.1016/j.res.2024.110392>
 - [12] Badawi, D. and Gildin, E. (2025) Neural Operator-Based Proxy for Reservoir Simulations Considering Varying Well Settings, Locations, and Permeability Fields. *Computers & Geosciences*, **196**, Article 105826. <https://doi.org/10.1016/j.cageo.2024.105826>
 - [13] Ma, X., Zhong, R., Zhan, J. and Zhou, D. (2024) Enhancing Subsurface Multiphase Flow Simulation with Fourier Neural Operator. *Heliyon*, **10**, e38103. <https://doi.org/10.1016/j.heliyon.2024.e38103>
 - [14] Kadeethum, T., Verzi, S.J. and Yoon, H. (2024) An Improved Neural Operator Framework for Large-Scale CO₂ Storage Operations. *Geoenery Science and Engineering*, **240**, Article 213007. <https://doi.org/10.1016/j.geoen.2024.213007>
 - [15] Santos, E.S., Barros, G.F., Oliveira, A.C.N., Silva, R.M., Freitas, R.S.M., Valiveti, D.M.,

- et al.* (2026) Hybrid DeepONet Surrogates for Multiphase Flow in Porous Media. *Journal of Computational Physics*, **561**, Article 114981. <https://doi.org/10.1016/j.jcp.2026.114981>
- [16] Huang, P., Leng, Y., Lian, C. and Liu, H. (2024) Porous-DeepONet: Learning the Solution Operators of Parametric Reactive Transport Equations in Porous Media. *Engineering*, **39**, 94-103. <https://doi.org/10.1016/j.eng.2024.07.002>
- [17] Diab, W., Chaabi, O., Alkobaisi, S., Awotunde, A. and Al Kobaisi, M. (2024) Learning Generic Solutions for Multiphase Transport in Porous Media via the Flux Functions Operator. *Advances in Water Resources*, **183**, Article 104609. <https://doi.org/10.1016/j.advwatres.2023.104609>
- [18] Yan, X., Lin, J., Ju, Y., Zhang, Q., Zhang, Z., Zhang, L., *et al.* (2025) A Finite-Volume Based Physics-Informed Fourier Neural Operator Network for Parametric Learning of Subsurface Flow. *Advances in Water Resources*, **205**, Article 105087. <https://doi.org/10.1016/j.advwatres.2025.105087>
- [19] Zhang, Z., Yan, X., Liu, P., Zhang, K., Han, R. and Wang, S. (2023) A Physics-Informed Convolutional Neural Network for the Simulation and Prediction of Two-Phase Darcy Flows in Heterogeneous Porous Media. *Journal of Computational Physics*, **477**, Article 111919. <https://doi.org/10.1016/j.jcp.2023.111919>
- [20] Yan, X., Lin, J., Wang, S., Zhang, Z., Liu, P., Sun, S., *et al.* (2024) Physics-Informed Neural Network Simulation of Two-Phase Flow in Heterogeneous and Fractured Porous Media. *Advances in Water Resources*, **189**, Article 104731. <https://doi.org/10.1016/j.advwatres.2024.104731>
- [21] Hanna, J.M., Aguado, J.V., Comas-Cardona, S., Askri, R. and Borzacchiello, D. (2022) Residual-Based Adaptivity for Two-Phase Flow Simulation in Porous Media Using Physics-Informed Neural Networks. *Computer Methods in Applied Mechanics and Engineering*, **396**, Article 115100. <https://doi.org/10.1016/j.cma.2022.115100>
- [22] Almajid, M.M. and Abu-Al-Saud, M.O. (2022) Prediction of Porous Media Fluid Flow Using Physics Informed Neural Networks. *Journal of Petroleum Science and Engineering*, **208**, Article 109205. <https://doi.org/10.1016/j.petrol.2021.109205>
- [23] Wang, N., Chang, H. and Zhang, D. (2022) Surrogate and Inverse Modeling for Two-Phase Flow in Porous Media via Theory-Guided Convolutional Neural Network. *Journal of Computational Physics*, **466**, Article 111419. <https://doi.org/10.1016/j.jcp.2022.111419>
- [24] Yan, B., Harp, D.R., Chen, B., Hoteit, H. and Pawar, R.J. (2022) A Gradient-Based Deep Neural Network Model for Simulating Multiphase Flow in Porous Media. *Journal of Computational Physics*, **463**, Article 111277. <https://doi.org/10.1016/j.jcp.2022.111277>
- [25] Yan, B., Harp, D.R., Chen, B. and Pawar, R. (2022) A Physics-Constrained Deep Learning Model for Simulating Multiphase Flow in 3D Heterogeneous Porous Media. *Fuel*, **313**, Article 122693. <https://doi.org/10.1016/j.fuel.2021.122693>
- [26] Yan, B., Chen, B., Robert Harp, D., Jia, W. and Pawar, R.J. (2022) A Robust Deep Learning Workflow to Predict Multiphase Flow Behavior during Geological CO₂ Sequestration Injection and Post-Injection Periods. *Journal of Hydrology*, **607**, Article 127542. <https://doi.org/10.1016/j.jhydrol.2022.127542>
- [27] Li, J., Zhang, D., He, T. and Zheng, Q. (2023) Uncertainty Quantification of Two-Phase Flow in Porous Media via the Coupled-TgNN Surrogate Model. *Geoenergy Science and Engineering*, **221**, Article 211368. <https://doi.org/10.1016/j.geoen.2022.211368>
- [28] Chakraborty, A., Rabinovich, A. and Moreno, Z. (2024) Physics-Informed Neural

- Networks for Modeling Two-Phase Steady State Flow with Capillary Heterogeneity at Varying Flow Conditions. *Advances in Water Resources*, **185**, Article 104639. <https://doi.org/10.1016/j.advwatres.2024.104639>
- [29] Alhubail, A., Fahs, M., Lehmann, F. and Hoteit, H. (2024) Modeling Fluid Flow in Heterogeneous Porous Media with Physics-Informed Neural Networks: Weighting Strategies for the Mixed Pressure Head-Velocity Formulation. *Advances in Water Resources*, **193**, Article 104797. <https://doi.org/10.1016/j.advwatres.2024.104797>
- [30] Basha, N., Arcucci, R., Angeli, P., Anastasiou, C., Abadie, T., Casas, C.Q., *et al.* (2024) Machine Learning and Physics-Driven Modelling and Simulation of Multiphase Systems. *International Journal of Multiphase Flow*, **179**, Article 104936. <https://doi.org/10.1016/j.ijmultiphaseflow.2024.104936>
- [31] Zhu, H., Chen, Z., Gao, X., Chen, X., Yu, W. and Sepehrnoori, K. (2025) A Physics-Informed Adaptive-Wavelet Neural Network (PIAWNN) for Reservoir Simulation with Two-Phase Flow. *Advances in Water Resources*, **207**, Article 105174. <https://doi.org/10.1016/j.advwatres.2025.105174>
- [32] Wang, Q., Zha, W., Li, D., Li, X., Shen, L. and Shi, Z. (2026) Parameterized Analytical Solution of Seepage Equation for Reservoir Simulation Using Physics-Informed Kolmogorov-Arnold Network without Labels. *Journal of Computational Physics*, **548**, Article 114599. <https://doi.org/10.1016/j.jcp.2025.114599>
- [33] Wang, Q., Li, X., Li, D., Zha, W. and Shen, L. (2026) Asymptotic Solution Neural Network for Solving Multi-Well Seepage Equation with Deep Integration of Superposition Principle and Laplacian Distance Decay Weighting in Reservoir Simulation. *Journal of Computational Physics*, **558**, Article 114845. <https://doi.org/10.1016/j.jcp.2026.114845>
- [34] Lin, J., Yan, X., Zhang, K., Zhang, Q., Zhang, Z., Zhang, L., *et al.* (2026) Deep Learning-Based Upscaling for Reservoir Models on Corner-Point Grids via Finite-Volume Physics-Informed Fourier Neural Operator. *Petroleum Science*, **23**, 5662-5692. <https://doi.org/10.1016/j.petsci.2026.04.037>
- [35] Donnelly, J., Daneshkhah, A. and Abolfathi, S. (2024) Physics-Informed Neural Networks as Surrogate Models of Hydrodynamic Simulators. *Science of The Total Environment*, **912**, Article 168814. <https://doi.org/10.1016/j.scitotenv.2023.168814>
- [36] Gao, H., Sun, L. and Wang, J. (2021) Phygeonet: Physics-Informed Geometry-Adaptive Convolutional Neural Networks for Solving Parameterized Steady-State PDEs on Irregular Domain. *Journal of Computational Physics*, **428**, Article 110079. <https://doi.org/10.1016/j.jcp.2020.110079>
- [37] Ranade, R., Hill, C. and Pathak, J. (2021) Discretizationnet: A Machine-Learning Based Solver for Navier-Stokes Equations Using Finite Volume Discretization. *Computer Methods in Applied Mechanics and Engineering*, **378**, Article 113722. <https://doi.org/10.1016/j.cma.2021.113722>
- [38] Sun, L., Gao, H., Pan, S. and Wang, J. (2020) Surrogate Modeling for Fluid Flows Based on Physics-Constrained Deep Learning without Simulation Data. *Computer Methods in Applied Mechanics and Engineering*, **361**, Article 112732. <https://doi.org/10.1016/j.cma.2019.112732>
- [39] Wang, S., Teng, Y. and Perdikaris, P. (2021) Understanding and Mitigating Gradient Flow Pathologies in Physics-Informed Neural Networks. *SIAM Journal on Scientific Computing*, **43**, A3055-A3081. <https://doi.org/10.1137/20m1318043>
- [40] Yu, J., Lu, L., Meng, X. and Karniadakis, G.E. (2022) Gradient-Enhanced Physics-Informed Neural Networks for Forward and Inverse PDE Problems. *Computer Methods in Applied Mechanics and Engineering*, **393**, Article 114823.

- <https://doi.org/10.1016/j.cma.2022.114823>
- [41] Wu, C., Zhu, M., Tan, Q., Kartha, Y. and Lu, L. (2023) A Comprehensive Study of Non-Adaptive and Residual-Based Adaptive Sampling for Physics-Informed Neural Networks. *Computer Methods in Applied Mechanics and Engineering*, **403**, Article 115671. <https://doi.org/10.1016/j.cma.2022.115671>
 - [42] Jagtap, A.D., Kharazmi, E. and Karniadakis, G.E. (2020) Conservative Physics-Informed Neural Networks on Discrete Domains for Conservation Laws: Applications to Forward and Inverse Problems. *Computer Methods in Applied Mechanics and Engineering*, **365**, Article 113028. <https://doi.org/10.1016/j.cma.2020.113028>
 - [43] Cardoso-Bihlo, E. and Bihlo, A. (2025) Exactly Conservative Physics-Informed Neural Networks and Deep Operator Networks for Dynamical Systems. *Neural Networks*, **181**, Article 106826. <https://doi.org/10.1016/j.neunet.2024.106826>
 - [44] Gerken, J.E., Aronsson, J., Carlsson, O., Linander, H., Ohlsson, F., Petersson, C., *et al.* (2023) Geometric Deep Learning and Equivariant Neural Networks. *Artificial Intelligence Review*, **56**, 14605-14662. <https://doi.org/10.1007/s10462-023-10502-7>
 - [45] Celledoni, E., Ehrhardt, M.J., Etmann, C., Owren, B., Schönlieb, C. and Sherry, F. (2021) Equivariant Neural Networks for Inverse Problems. *Inverse Problems*, **37**, Article 085006. <https://doi.org/10.1088/1361-6420/ac104f>
 - [46] Ma, G., Wang, Y., Lim, D., Jegelka, S. and Wang, Y. (2024) A Canonicalization Perspective on Invariant and Equivariant Learning. *Advances in Neural Information Processing Systems* 37, Vancouver, 10-15 December 2024, 60936-60979. <https://doi.org/10.52202/079017-1948>
 - [47] Bulusu, S., Favoni, M., Ipp, A., Müller, D.I. and Schuh, D. (2021) Generalization Capabilities of Translationally Equivariant Neural Networks. *Physical Review D*, **104**, Article 074504. <https://doi.org/10.1103/physrevd.104.074504>
 - [48] Xu, Y., Kohtz, S., Boakye, J., Gardoni, P. and Wang, P. (2023) Physics-Informed Machine Learning for Reliability and Systems Safety Applications: State of the Art and Challenges. *Reliability Engineering & System Safety*, **230**, Article 108900. <https://doi.org/10.1016/j.res.2022.108900>
 - [49] Psaros, A.F., Meng, X., Zou, Z., Guo, L. and Karniadakis, G.E. (2023) Uncertainty Quantification in Scientific Machine Learning: Methods, Metrics, and Comparisons. *Journal of Computational Physics*, **477**, Article 111902. <https://doi.org/10.1016/j.jcp.2022.111902>
 - [50] Guo, L., Wu, H., Wang, Y., Zhou, W. and Zhou, T. (2024) IB-UQ: Information Bottleneck Based Uncertainty Quantification for Neural Function Regression and Neural Operator Learning. *Journal of Computational Physics*, **510**, Article 113089. <https://doi.org/10.1016/j.jcp.2024.113089>