International Journal of Geosciences, 2026, 17(9), 671-702
https://www.scirp.org/journal/ijg
ISSN Online: 2156-8367
ISSN Print: 2156-8359

Graph Neural Networks for Unstructured Reservoir Flow Prediction: Conservation, Multi-Resolution Transfer, and Dual Physical Connectivity

Franck-Hilaire Essiagne*, Koffi Eugene Kouadio, Kouassi Louis Kra, Moussa Camara

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 Houphouet-Boigny (INP-HB), Yamoussoukro, Côte d'Ivoire
Email: *franck.essiagne@inphb.ci

How to cite this paper: Essiagne, F.-H., Kouadio, K.E., Kra, K.L. and Camara, M. (2026) Graph Neural Networks for Unstructured Reservoir Flow Prediction: Conservation, Multi-Resolution Transfer, and Dual Physical Connectivity. *International Journal of Geosciences*, **17**, 671-702.
https://doi.org/10.4236/ijg.2026.179032

Received: August 22, 2026
Accepted: September 20, 2026
Published: September 23, 2026

Abstract

Graph neural networks (GNNs) operate naturally on irregular reservoir grids, but low state error does not establish conservation, constitutive consistency, mesh transfer, physical interpretability, or numerical value. We audit a physics-certified framework for incompressible two-phase flow on unstructured finite-volume graphs. Pressure is exactly scalar-rate equivariant; a graph head predicts antisymmetric edge flux; and a discrete divergence projection enforces local continuity before conservative upwind transport. Projection reduces flux error by 47.1% - 57.1% and long-rollout saturation error by 73.6% - 86.1%, while driving non-well continuity defects from order unity to 10^{-14}. A post-projection diagnostic nevertheless finds inverse-transmissibility-weighted Darcy defects of 0.27 - 0.59, proving that continuity does not imply constitutive compatibility. Under a strict matched budget (eight training reservoirs, 140 updates, ≈481.6 × 10^3 node-time presentations), multi-resolution exposure reduces projected-flux error by 42.0% - 73.2% at all five graph sizes; pressure gains are significant at three sizes and unresolved at two. In 24 multiwell reservoirs, hydraulic and tracer connectivity have mean injector-row rank correlation 0.632 at 0.5 PVI and top-producer agreement 73.6%. A tolerance-conditioned warm-start selector closely tracks the lower-iteration initializer, but reported savings are PCG iterations, not end-to-end runtime. The results support GNNs as structure-aware proposals coupled to discrete conservation and solver diagnostics rather than unconstrained simulator replacements.

Keywords

Graph Neural Networks, Reservoir Simulation, Unstructured Mesh, Finite-Volume Conservation, Multiscale Learning, Interwell Connectivity

1. Introduction

Reservoir simulation is a conservation problem on irregular geological discretizations. Heterogeneous transmissibility, saturation-dependent mobility, wells, faults, and non-neighbour connections couple globally communicating pressure to locally advective transport. Graph neural networks are attractive because this discretization is already relational: cells are nodes and physical connections are edges [1]-[7].

The novelty threshold has moved beyond showing that a GNN can predict pressure and saturation on an irregular mesh. Recent work already covers unstructured reservoir optimization, porous-media transport, faulted CO_2 systems, multiscale graph processing, and physics-informed irregular-grid learning [3]-[13]. A stronger question is whether learned states and fluxes obey discrete physics and remain useful under mesh, geology, control, and well-topology shift.

Three methodological problems are especially consequential. A network can violate an exact control symmetry despite modest field error; an antisymmetric edge field can cancel globally while violating cellwise continuity; and a graph can transfer topologically yet fail when graph size leaves the training range. These failures require tests beyond state RMSE: exact equivariance, local balance, constitutive consistency, long-horizon transport, and explicit OOD evaluation.

Connectivity introduces a fourth issue. Dynamic interwell connectivity is commonly inferred from production-injection fluctuations, material-balance models, attention mechanisms, or streamline allocation [14]-[16]. These quantities are useful, but they are not physically identical. A pressure perturbation can communicate rapidly across a permeable path without the injected material reaching the same producer on the same timescale. We therefore distinguish hydraulic connectivity, defined by control-to-response sensitivity, from transport connectivity, defined by tagged-material attribution. Their agreement or disagreement becomes a falsifiable physical result rather than an interpretation of an attention coefficient.

The study has four linked blocks. A rate-equivariant GNN predicts pressure per unit total rate and one antisymmetric flux per edge, followed by finite-volume divergence projection. A matched-budget experiment tests multi-resolution scale transfer and coarse aggregation. A multiwell benchmark independently measures hydraulic response and tagged-tracer transport connectivity. Finally, learned warm starts are compared with the previous converged pressure and selected by a tolerance-conditioned classifier.

The findings are deliberately mixed. Projection greatly improves continuity and transport but leaves a measurable non-potential Darcy-cycle component. Matched-budget multi-resolution exposure robustly improves projected-flux transfer, whereas pressure gains vary by graph size and transmissibility aggregation is not universally superior to Euclidean aggregation. Hydraulic and transport connectivity are correlated but non-equivalent within individual injectors. These results identify both useful structure and its limits.

2. Governing Equations and Discrete Reference Model

We consider incompressible, immiscible oil-water flow in a two-dimensional porous layer. Gravity and capillarity are omitted so that the experiments isolate graph topology, permeability heterogeneity, mobility, wells, and transport. This is intentionally narrower than a compositional field simulator; conclusions are restricted to the tested regime. The domain is discretized by an unstructured Delaunay/Voronoi-like finite-volume graph. Cell volumes, contact geometry, harmonic permeability, and phase mobility determine the transmissibility used by the reference solver.

At a fixed saturation state, the pressure equation for a free control volume i is written as a conservative transmissibility balance

$$\sum_{j\in N(i)} T_{ij}\lambda_{t,ij}\left(p_i - p_j\right) = q_i \tag{1}$$

where total mobility is evaluated from the local saturation-dependent phase mobilities. Producer cells are held at the reference pressure and injection is specified as a balanced source distribution, with well-pressure treatment following the conventional well-block interpretation used in reservoir simulation [12]-[17]. The sparse linear system is solved to numerical precision to generate reference pressure and total face fluxes.

Water saturation is advanced with a first-order upwind finite-volume update. For cell i,

$$\phi V_i \frac{S_{w,i}^{n+1} - S_{w,i}^{n}}{\Delta t} + \sum_{j\in N(i)} F_{ij}^{w} = q_i^{w} \tag{2}$$

with water flux obtained by upwinding fractional flow according to the sign of the total face flux. The reference time step is selected so that the benchmark trajectories show substantial front movement without saturation clipping. All principal reference runs were audited for clipping and global water balance; the latter remains at floating-point scale.

The reference discretization is also the definition of physical consistency used throughout the paper. This choice avoids comparing a neural prediction against a continuous PDE residual that is different from the numerical task that produced the labels. On each interior connection, the same transmissibility and mobility convention is used to assemble pressure, compute reference total flux, evaluate learned-flux error, and form the projection constraint. Source terms are assembled at the control-volume level, and producer pressure constraints are removed from the free-cell continuity equations before normalization. Consequently, a projected flux satisfying the free-cell divergence constraint is conservative with respect to the exact discrete operator used by the benchmark, rather than merely small under a surrogate residual. This distinction becomes important in Section

5.1, where antisymmetry alone preserves global cancellation but leaves large cell-wise divergence errors.

3. Physics-Certified Graph Formulation

3.1. Rate-Equivariant Graph State Representation

Each cell node contains normalized coordinates, log permeability, porosity/pore-volume information, saturation, source indicators, producer/injector flags, and graph-based hydraulic coordinates. Physical edges contain relative displacement, distance, transmissibility-related information, and local geometric features. The pressure head does not predict pressure directly. Because the balanced incompressible pressure equation is linear in a common source-rate scale Q at fixed saturation, the model predicts pressure per unit rate and reconstructs pressure as

$$p_\theta(Q) = Q\tilde{p}_\theta \tag{3}$$

This makes scalar-rate equivariance exact rather than statistical. In Benchmark A, a direct-pressure GNN violates this transformation by 14.8% - 22.4% across ID and OOD suites, whereas the normalized formulation has zero violation by construction. Hydraulic-resistance shortest-path coordinates are also supplied to expose long-range connectivity without requiring the message-passing depth itself to span the reservoir.

3.2. Antisymmetric Edge-Flux Prediction

A second decoder predicts one scalar total flux for each unique physical edge. The same scalar is assigned with opposite signs to the two directed incidences. Thus, the model enforces

$$\hat{F}_{ij} = -\hat{F}_{ji} \tag{4}$$

exactly. This architectural antisymmetry is useful but insufficient: it guarantees that interior fluxes cancel when summed globally, yet it does not constrain the divergence at an individual non-well cell. The experiments explicitly test this distinction rather than equating global cancellation with local finite-volume conservation.

3.3. Discrete Divergence Projection

Let B denote the oriented cell-edge incidence operator on free cells and let q be the assembled source vector. The raw edge prediction is projected onto the affine finite-volume constraint set by solving

$$F_{proj} = \arg\min_{F:BF=q} \left\| F - \hat{F} \right\|_{T^{-1}}^2 \tag{5}$$

where the transmissibility-weighted norm penalizes corrections in hydraulic resistance units. In implementation, the Lagrange multiplier system uses the same sparse elliptic operator class as the pressure balance. The projection is therefore treated here as a physics layer and diagnostic experiment, not as a claimed free acceleration mechanism. A separate limited-iteration study confirms this distinction: 8 - 16 Jacobi-preconditioned CG projection iterations reduce but do not eliminate continuity defect, whereas the exact projection is required to reach machine-precision balance in the present implementation.

The projection can be interpreted as the minimum hydraulic-energy correction required to make the learned edge field admissible. Because the constraint is linear, the correction has the form of a graph potential difference and therefore modifies neighbouring edge fluxes coherently rather than clipping individual edges independently. This is materially different from adding a continuity penalty during training: the penalty can encourage low average divergence but cannot guarantee an admissible state at inference, whereas projection supplies an explicit certificate. We evaluate the cost of that certificate separately. In particular, the exact solve and the finite-iteration approximations use identical raw neural proposals, so improvements in transport can be attributed to enforcement of the discrete constraint rather than to retraining or additional labels.

Divergence admissibility is weaker than constitutive admissibility. We therefore add a post-projection Darcy diagnostic on each held-out state. Using the same saturation-dependent transmissibility, mobility convention, and zero-gauge producer datum as the reference discretization, we compute the unique Darcy-consistent gradient flux and compare it with the projected learned flux in an inverse-transmissibility-weighted energy norm. The residual difference is a divergence-free cycle component on free cells. We report both its relative energy and the edge-sign agreement between the projected flux and the Darcy flux associated with the neural pressure head plus the projection potential. This diagnostic test whether a locally conservative edge field is also compatible with a pressure-gradient constitutive relation; projection is not assumed to guarantee that property.

3.4. Residual Certification and Conservative Rollout

The pressure proposal is independently evaluated with the assembled finite-volume matrix A. We use the normalized discrete residual

$$r = \frac{\|Ap - q\|^2}{\|q\|^2 + \varepsilon} \tag{6}$$

as a label-free admissibility indicator. Where high pressure accuracy is required, preconditioned conjugate gradients (PCG) correct the proposal to a specified residual tolerance. Saturation is always advanced through finite-volume upwind transport, using either pressure-derived certified fluxes or, in the new edge-flux experiments, the projected learned flux. This separation is deliberate: a learned field proposes, while a known discrete operator supplies conservation and numerical certification.

3.5. Multi-Resolution Training and Coarse Communication

The single-resolution flux model is trained on eight heterogeneous reservoirs with approximately 100 - 145 interior seed points. To test scale transfer, a second curriculum uses eight training reservoirs spanning 100, 140, 200, 270, 350, 450, 600, and 750 interior seed points, with held-out validation meshes at 180 and 380. The message-passing architecture and parameter count remain fixed at 66,242 trainable parameters. Coarse communication is evaluated with random groups, Euclidean k-means groups, and a transmissibility-weighted heavy-edge aggregation. Matched multi-resolution Euclidean and transmissibility models are used to distinguish the effect of training-scale diversity from the effect of the coarse-graph rule.

3.6. Dual Physical Connectivity

For each of 24 independent three-injector/three-producer reservoirs, hydraulic connectivity is obtained from a high-fidelity unit perturbation of each injector while producers remain at fixed bottom-hole pressure. The response matrix is normalized by the total induced production response for that injector,

$$C_{ab}^{H} = \frac{\Delta q_b / \Delta u_a}{\sum_{k \in P} \Delta q_k / \Delta u_a} \tag{7}$$

where a indexes injectors and b indexes producers. Transport connectivity is obtained from conservative passive tagged tracers injected separately from each injector and accumulated to a specified pore-volume-injected horizon. The allocation is

$$C_{ab}^{T}(\tau) = \frac{m_{a \to b}(\tau)}{\sum_{k \in P} m_{a \to k}(\tau)} \tag{8}$$

and breakthrough time is recorded when producer tracer cut first exceeds 10^{-3}. This produces independent labels for pressure communication and material communication. The tracer integrator is audited for bounded concentration and mass error; the representative case has maximum tracer mass error 1.13×10^{-10} and concentration within numerical tolerance of [0, 1].

3.7. Realistic Warm-Start Audit and Tolerance-Conditioned Selector

A sequential reservoir solver rarely starts every pressure solve from zero. Krylov iteration, preconditioning, and inexact-solve theory provide the numerical context for this comparison [18]-[21]. Benchmark B therefore compares the learned correction with a rate-scaled previous converged pressure. A residual-only gate is intentionally included as a negative control: it almost never selects the graph correction because smaller initial residual does not necessarily imply fewer iterations to a finite PCG target. A lightweight tolerance-conditioned selector instead predicts which candidate will require fewer iterations from pre-solve features such as support change, active-well count, rate ratio, candidate residuals, correction magnitude, graph size, and requested tolerance. Conceptually,

$$z_{sel} = \arg\min_{z \in \{prev, GNN\}} N(z; \tau, x) \tag{9}$$

where the selector is trained only on training/validation transitions and frozen before OOD evaluation. The purpose is not to replace PCG but to decide when learned information is numerically worth using. The complete physics-certified workflow is summarized in Figure 1.

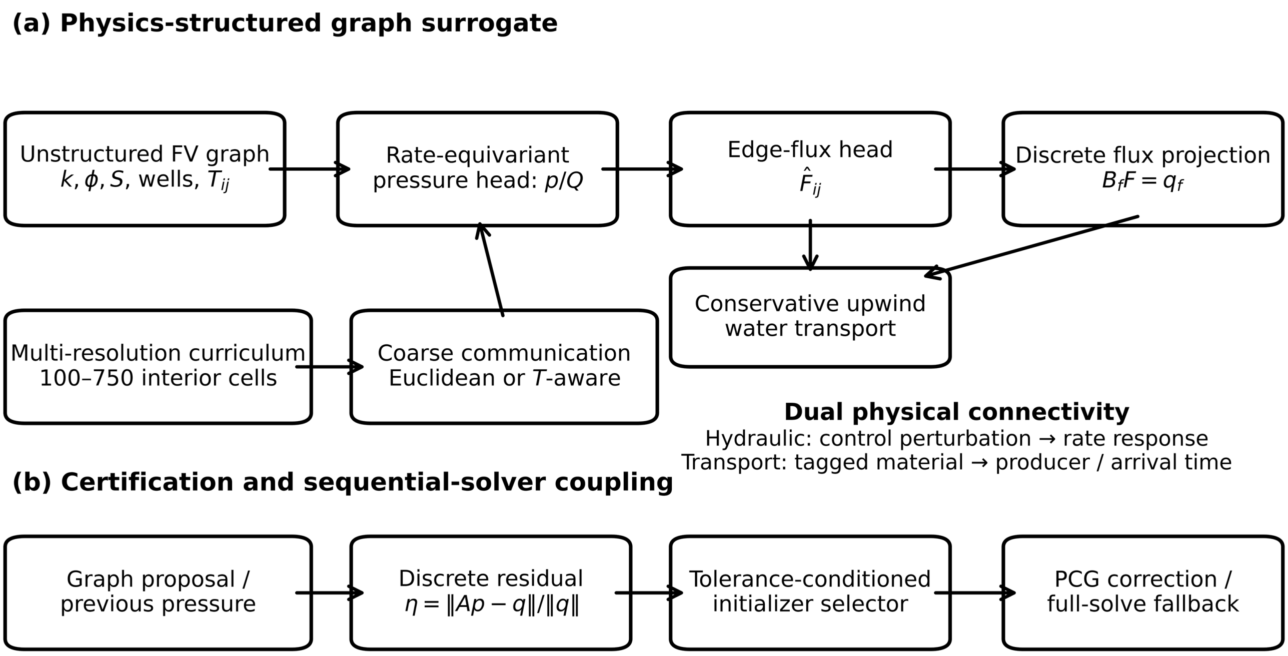

Figure 1. Expanded physics-certified workflow. The model predicts rate-normalized pressure and antisymmetric edge fluxes on the native unstructured graph. Discrete residual evaluation certifies pressure proposals; a finite-volume divergence projection enforces local continuity of learned edge fluxes before conservative transport. Multi-resolution training/coarse communication and the dual hydraulic-transport connectivity benchmark test scale transfer and physical interpretability, while the numerical-solver branch retains the realistic warm-start audit and tolerance-conditioned selector.

Broader subsurface-surrogate literature supplies additional comparators rather than directly comparable leaderboards. Reservoir neural operators, uncertainty-aware surrogates, and physics-informed models motivate the state/flux and robustness diagnostics [22]-[40]; general neural-operator studies and recent large-scale porous-media applications motivate the scale-transfer controls [41]-[47]; and physics-constrained dynamics plus classical multiscale finite-volume work motivate structure-preserving and coarse-grid comparisons [48]-[50]. Because these studies use different grids, physics, datasets, and compute budgets, their published errors are used only for methodological context.

4. Experimental Design and Statistical Analysis

All reported measurements are computational experiments generated with fixed random seeds. No field-measured data are claimed. The reference model uses correlated lognormal permeability fields on independently generated unstructured meshes, heterogeneous porosity, randomly perturbed well locations, and piecewise-varying controls. Entire geological realizations are held out; cells and time snapshots from a test reservoir never appear in training. Five principal suites are used: held-out ID geology, finer-mesh OOD, stronger/shorter-correlation heterogeneity OOD, control-amplitude OOD, and a combined mesh-geology-control shift. The multiwell solver benchmark separately adds topology changes and rapid switching.

4.1. Reproducibility Specification

Single-pair domains are 500 m × 500 m with 15 - 25 m thickness. Producer-completion cells impose the pressure datum $p = 0$ Pa; only pressure differences are used. The common rate scale Q is the total injector-patch source rate before SI conversion. Base Q is 700 - 1300 m^3/day and is allocated to injector cells in proportion to pore volume. Training controls use three piecewise-constant multipliers in 0.8 - 1.2; control-OOD segments use 0.40 - 0.60, 1.25 - 1.40, and 0.65 - 0.80. Producer rates emerge from fixed-pressure sinks. Initial water saturation is 0.20. Corey curves use $S_{wc} = S_{or} = 0.15$, quadratic exponents, $\mu_w = 1$ mPa·s and $\mu_o = 5$ mPa·s; gravity and capillarity are omitted.

Porosity is clipped to 0.13 - 0.30 around mean 0.21 and standard deviation 0.025. Permeability is a 64-mode correlated lognormal field with 100 - 220 mD geometric mean, training log-standard deviation about 0.55 - 1.05, and correlation parameter 70 - 140 m; stronger OOD uses 1.25 and 65 m. Pressure is implicit and saturation uses first-order upwind finite volume with $\Delta t = 1.25$ days. The edge-flux GNN has 66,242 parameters, 40 latent channels, two fine and two coarse message-passing blocks, and separate pressure-per-Q and unique-edge-flux heads. AdamW uses learning rate of 1.2×10^{-3}, weight decay of 2×10^{-6} and gradient cap of 1.0; the loss is Smooth-L1 pressure plus 1.25× Smooth-L1 flux, with no conservation penalty. The single-resolution model uses 320 training and 80 validation transitions. Exact seeds and scripts are archived. Table A1 summarizes the reference-solver and tracer quality-control checks, while Table A2 lists the complete physical, numerical, graph-model, and training settings used for reproducibility.

4.2. Evaluation and Statistical Inference

Benchmark A uses 40-step two-phase rollouts and reports pressure relative L2 error, normalized saturation error, production-rate error, and normalized discrete residual. The new edge-flux benchmark uses the same family of reference equations and additionally reports relative unique-edge flux error, free-cell continuity defect, saturation clipping mass, and long-horizon transport error. Raw and projected predictions are paired on identical reservoirs and schedules. Bootstrap confidence intervals are computed by resampling whole reservoir realizations rather than individual cells.

The scale-transfer experiment deliberately extrapolates beyond the multi-resolution training range. One-step frozen-model tests contain approximately 314, 588, 1128, 2180, and 4252 graph nodes. This experiment is a stress test, not a claim of industrial million-cell scalability. The dual-connectivity benchmark comprises 12 moderate-heterogeneity reservoirs with log-permeability standard deviation 0.7 and 12 strong-heterogeneity reservoirs with standard deviation 1.35. For each realization, all nine injector-producer pairs are evaluated through independent perturbation and tagged-tracer calculations at 0.25, 0.5, and 1.0 pore volumes injected (PVI).

Inference is at the reservoir-realization level. Paired changes are formed per realization before bootstrap resampling; cells, edges, time steps, and injector-producer pairs are not treated as independent reservoirs. The revised scale ablation uses eight training reservoirs in each regime, 140 optimizer updates, the same architecture/optimizer/validation cases, and 481,632 versus 481,600 node-time presentations (0.0066% difference), followed by eight unseen reservoirs at each graph size. Connectivity inference is performed per injector because row-normalized producer fractions are dependent: Spearman uses average ranks for ties, top-link agreement is set-valued, zero-recovery rows would be excluded, and breakthrough absent by 1.0 PVI is right-censored. Reservoir-cluster bootstrap resamples whole reservoirs while retaining all injector rows. The four experimental blocks and their falsifiable questions are summarized in Table 1.

Table 1. Experimental blocks and their falsifiable questions.

Study block	Training/cases	Held-out evaluation	Primary question
Residual-certified pressure	8 training reservoirs	6 ID + 4 per principal OOD suite	Can residual correction stabilize two-phase rollouts?
Edge-flux/projection	8 single-resolution training reservoirs	5 ID + 4 per OOD suite	Is antisymmetry sufficient, or is local divergence projection needed?
Multi-resolution transfer	8 meshes spanning 100 - 750 interior seeds	standard suites + 314 - 4252-node scaling	What controls mesh-scale transfer?
Dual connectivity	No learned labels; 24 high-fidelity multiwell reservoirs	216 injector-producer relations per horizon	Are hydraulic and material connectivity equivalent?
Solver selector	Multiwell transition training/validation	ID + mesh, heterogeneity, topology, rapid-switch, combined OOD	Does learning add value beyond a previous-pressure warm start?

Models are trained with AdamW and selected exclusively from held-out validation loss. The edge-flux architecture is fixed at 66,242 trainable parameters across hierarchy and scale ablations. The budget-matched scale experiment samples all eight reservoir cases in every optimizer update and equalizes total node-time exposure to within 0.007%; the mean edge loss remains one scalar term per update with the same weight in both regimes. Primary paired percentage reductions use 5000 reservoir-level bootstrap replicates. The tolerance-conditioned solver selector is trained on training/validation transitions only and frozen before OOD evaluation.

5. Results

5.1. Antisymmetry Is Not Local Conservation; Divergence Projection Enforces Local Continuity

The raw edge decoder predicts a single oriented scalar per physical connection, so its flux is antisymmetric by construction. Nevertheless, its mean normalized continuity defect on non-producer cells is 1.23 - 3.97 across the five single-resolution evaluation suites. This result is important because global internal cancellation could otherwise be misreported as "conservation." The raw antisymmetric field redistributes equal and opposite flux between neighbouring cells, but its divergence generally fails to match the local source vector.

The finite-volume projection changes the result qualitatively. Mean continuity defect falls to 6.15×10^{-15} - 7.57×10^{-14}. At the same time, projection reduces relative flux error by 55.9% in ID, 57.1% on finer meshes, 47.1% under stronger heterogeneity, 52.5% under control OOD, and 50.5% under the combined shift. All paired bootstrap intervals remain positive; the weakest interval, for heterogeneity, is 32.4% - 57.2%. Thus, the physics projection is not merely repairing the divergence statistic: it moves the learned flux closer to the high-fidelity flux field.

The constitutive audit qualifies the conservation result. Relative to the Darcy-consistent flux on the same saturation state, the inverse-transmissibility-weighted defect of the projected edge field is 0.271 [0.221, 0.313] ID, 0.369 [0.272, 0.448] on mesh OOD, 0.359 [0.315, 0.437] under stronger heterogeneity, 0.281 [0.222, 0.340] for control OOD, and 0.588 [0.331, 0.925] for combined OOD. The divergence-free cycle component accounts for about 26% - 47% of projected-flux energy, while Darcy sign agreement is 0.78 - 0.84. Thus, projection enforces continuity but not full pressure-gradient compatibility. Table A7 reports the full suite-wise constitutive-consistency statistics, and Figure A5 shows the corresponding Darcy-cycle defect and edge-sign agreement across the ID and OOD suites.

Transport consequences are even larger. Relative long-rollout saturation error is reduced by 73.6% - 86.1% across the same suites. Raw fluxes produce substantial accumulated saturation clipping in ID, mesh, and combined OOD tests; the exact projection eliminates clipping in all principal projected rollouts. In the representative combined-shift trajectory, final saturation error drops from 29.0% of the dynamic saturation range to 7.0%. The result supports a more precise claim than "hard conservation helps": antisymmetry alone is too weak, whereas projection onto the local discrete divergence constraint changes both physical admissibility and predictive transport accuracy. Table 2 and Figure 2 summarize the paired conservation, flux, and transport effects of projection.

The projection is an elliptic constrained correction, and the present exact implementation should not be confused with a cheap algebraic post-process. A limited-iteration projection study provides an additional negative result: on ID reservoirs, 8 and 16 preconditioned CG correction iterations leave mean continuity defects of 0.320 and 0.259 and saturation errors of 16.3% and 14.7%, respectively, compared with 3.99% saturation error after the exact projection. On OOD mesh cases, 16 iterations reduce saturation error to 8.58% but still leave continuity 0.295. We therefore make no acceleration claim for exact flux projection; its present contribution is structural certification and a benchmark for future approximate/differentiable projection schemes. Table A3 reports the full limited-iteration projection audit and shows how finite PCG budgets improve the raw field without reproducing the exact conservation result.

Table 2. Effect of exact finite-volume edge-flux projection. Reductions compare paired raw and projected rollouts from the same trained transmissibility model.

Suite	Raw continuity	Projected continuity	Flux-error reduction, % [95% CI]	Saturation-error reduction, % [95% CI]
ID	1.476	1.05×10^{-14}	55.9 [44.8, 67.0]	77.7 [72.0, 84.4]
OOD mesh	2.466	1.34×10^{-14}	57.1 [47.9, 69.2]	86.1 [79.8, 92.8]
OOD heterogeneity	1.751	7.57×10^{-14}	47.1 [32.4, 57.2]	77.3 [75.5, 79.1]
OOD controls	1.229	6.15×10^{-15}	52.5 [45.0, 60.0]	73.6 [61.2, 81.8]
OOD combined	3.969	3.37×10^{-14}	50.5 [38.2, 59.3]	83.3 [80.6, 86.5]

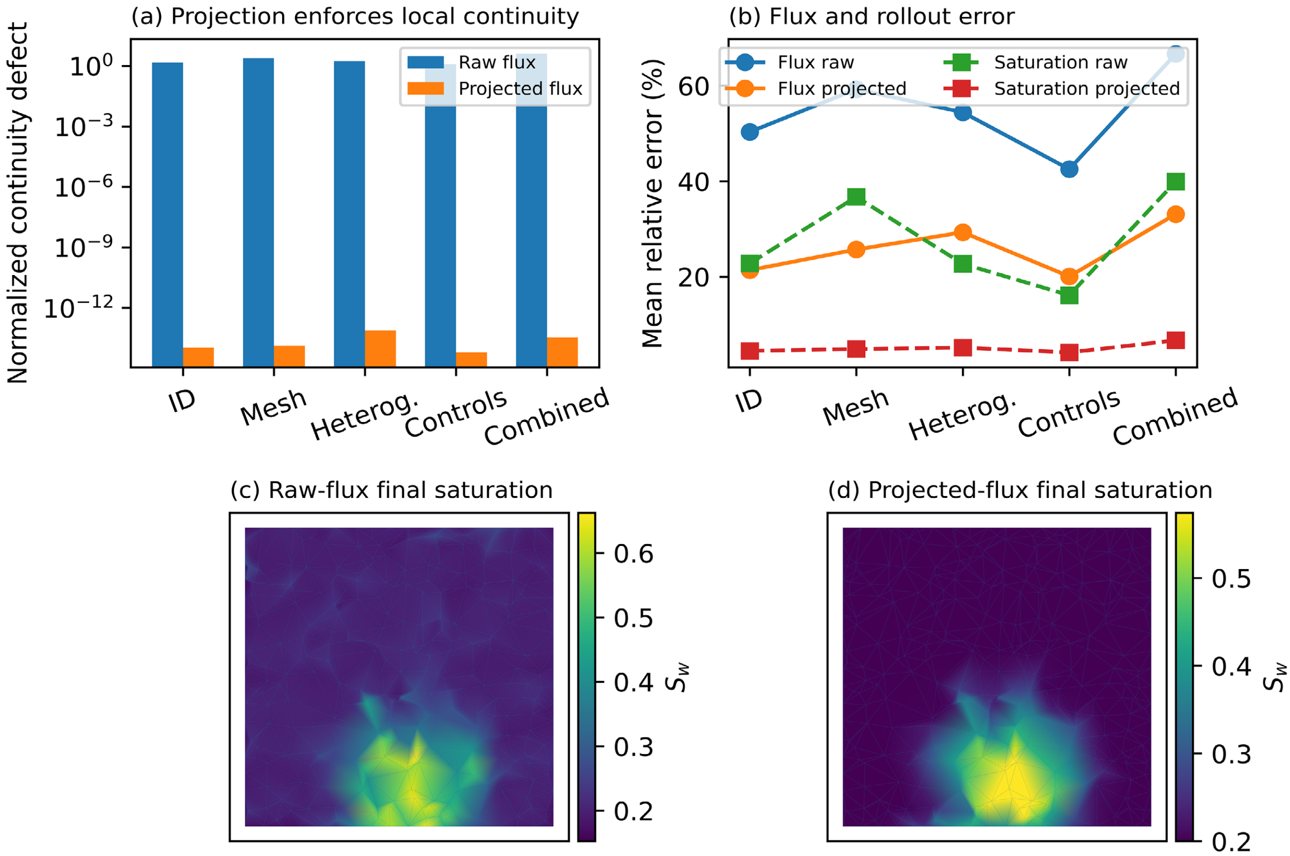

Figure 2. Local conservation and transport effects of edge-flux projection. Raw antisymmetric graph fluxes retain order-one free-cell continuity defects. Projection drives the discrete defect to machine precision, reduces relative face-flux error across every ID/OOD suite, and materially improves long-horizon saturation. The spatial example shows a combined-shift case in which the projected flux prevents the large front distortion produced by the raw learned flux.

5.2. Multi-Resolution Exposure, Not a Universally Superior Hierarchy, Controls Scale Transfer

The original scale sweep used unequal label exposure and is therefore not used for the revised primary inference. We reran the comparison under a matched budget: eight single-resolution and eight multi-resolution reservoir cases, the identical 66,242-parameter transmissibility model, the same AdamW settings and validation cases, 140 optimizer updates in each regime, and 481,632 versus 481,600 node-time presentations. The single-resolution updates draw 19 - 20 state graphs per update from eight 173-node training meshes; each multi-resolution update draws one state from each of eight meshes spanning 148 - 858 graph nodes. The resulting node-time mismatch is only 0.0066%.

Each graph size is evaluated on eight unseen reservoirs. Multi-resolution training reduces projected-flux error by 42.0% [34.6, 50.2], 49.5% [42.1, 56.6], 61.3% [59.3, 63.5], 65.6% [63.1, 68.2], and 73.2% [69.1, 77.5] from 314 to 4252 nodes. Pressure reduction is 35.7% [9.3, 57.9], 25.3% [−6.2, 52.4], 58.7% [39.6, 74.0], 23.7% [−6.6, 51.8], and 62.6% [55.9, 69.0]. Hence projected-flux improvement is resolved at every size, whereas pressure improvement is unresolved at 588 and 2180 nodes.

The coarse-graph hypothesis remains more nuanced. In the separate equal-parameter hierarchy experiment, Euclidean aggregation often yields slightly lower pressure error on large graphs, whereas transmissibility aggregation can yield lower projected-flux error. We therefore continue to reject a universal claim that transmissibility-aware pooling is intrinsically superior. The revised budget-matched experiment sharpens the conclusion: exposure to multiple resolutions is a reproducible driver of flux transfer, but its pressure benefit is scale dependent and should not be summarized by a single 65% - 79% range. **Table 3** and **Figure 3** report the matched-budget scale-transfer results and their reservoir-bootstrap intervals.

5.3. Hydraulic Response and Material Transport Are Related but Not Interchangeable

The dual-connectivity experiment independently tests what "interwell connectivity" means. Flattened 3 × 3 matrix correlations are 0.744, 0.769, and 0.773 at 0.25, 0.5, and 1.0 PVI, but the primary analysis is injector based. Across 72 injector rows, mean hydraulic-transport Spearman agreement is 0.590 [0.465, 0.701], 0.632 [0.514, 0.736], and 0.583 [0.465, 0.694]. Tie-aware top-producer agreement is 69.4%, 73.6%, and 75.0%, respectively.

Heterogeneity amplifies the distinction. At 0.5 PVI, mean injector-row correlation is 0.681 [0.556, 0.792] for moderate heterogeneity and 0.583 [0.403, 0.750] for the strong group. All 72 injector rows have nonzero tracer recovery. Nine of 216 producer pairs (4.17%) have no breakthrough above 10^{-3} by 1.0 PVI and are right-censored. Among observed producers, hydraulic-versus-breakthrough rank agreement is 0.556 [0.417, 0.681], and the hydraulically strongest producer is also the earliest observed breakthrough producer for 69.4% [58.3, 80.6] of injectors. **Table A5** summarizes the matrix-level and injector-level connectivity measures, while **Figure A4** shows the reservoir-level distributions of rank correlation and top-link mismatch at 0.5 PVI.

Tracer horizon also changes what "connected" means. Average ranks handle ties and top-link agreement uses tied-max sets. Mean recovered injected tracer is only 1.1% at 0.25 PVI, then 8.7% at 0.5 and 30.5% at 1.0 PVI. Row normalization therefore describes the allocation of recovered material among producers; the unrecovered fraction is reported separately. **Figure A6** and **Table A8** give row-level distributions and censoring counts.

Table 3. Training-budget-matched zero-shot graph-size scaling. Both regimes use eight training reservoir cases, 140 optimizer updates, the same architecture and optimizer, and approximately 481.6 × 10^3 node-time presentations; n = 8 unseen reservoirs per graph size. Errors are one-step relative L2 quantities after exact flux projection where applicable.

Graph nodes (n = 8)	Single p err.	Multi p err.	p reduction [95% CI]	Single flux err.	Multi flux err.	Flux reduction [95% CI]
314	36.1%	19.4%	35.7% [9.3, 57.9]	38.0%	21.6%	42.0% [34.6, 50.2]
588	54.3%	31.6%	25.3% [−6.2, 52.4]	53.7%	27.0%	49.5% [42.1, 56.6]
1128	51.5%	17.7%	58.7% [39.6, 74.0]	55.7%	21.5%	61.3% [59.3, 63.5]
2180	56.1%	40.1%	23.7% [−6.6, 51.8]	79.7%	27.6%	65.6% [63.1, 68.2]
4252	79.9%	27.1%	62.6% [55.9, 69.0]	111.7%	28.8%	73.2% [69.1, 77.5]

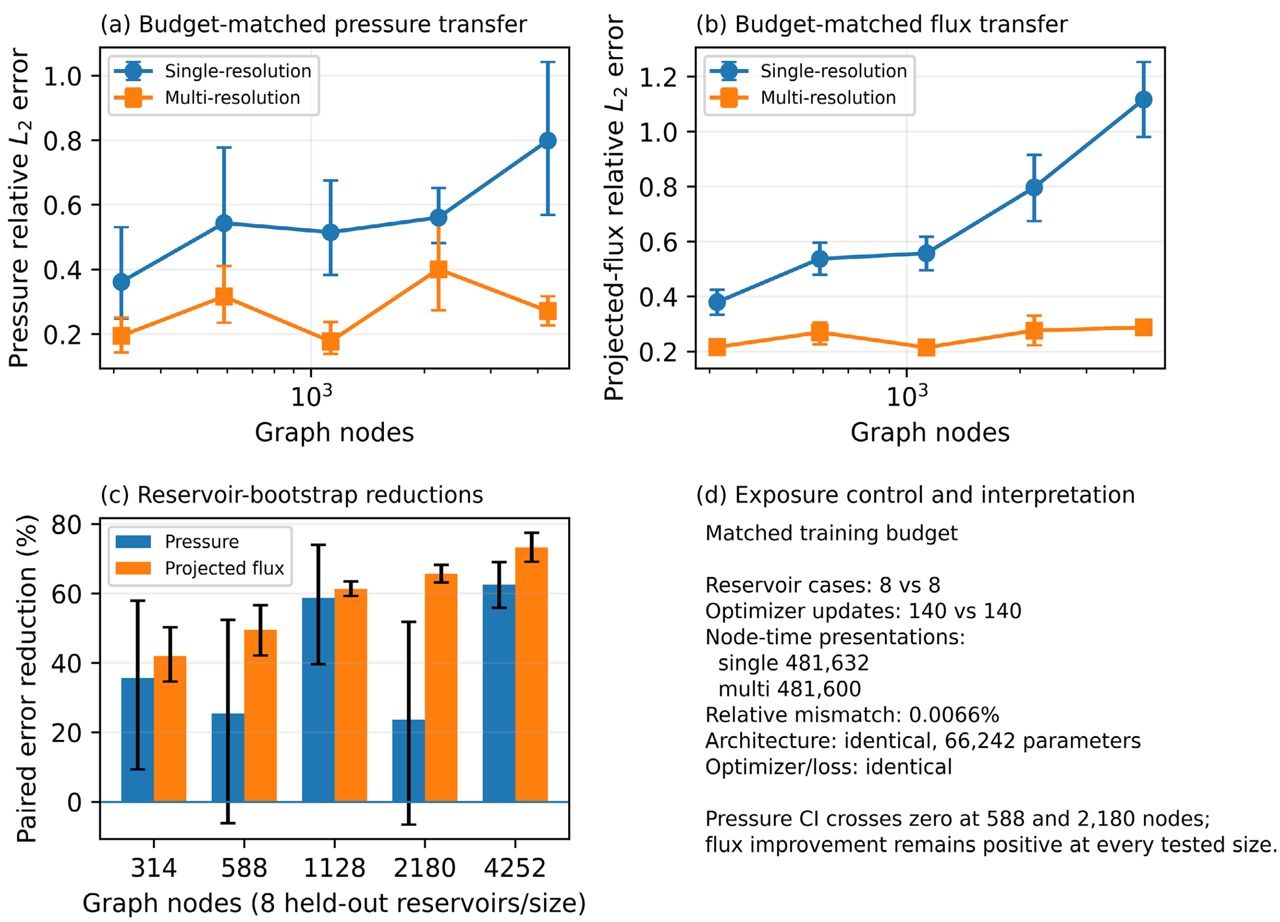

Figure 3. Training-budget-matched multi-resolution transfer. Error bars are reservoir-bootstrap 95% intervals from eight unseen reservoirs at each graph size. The two training regimes use equal reservoir counts, optimizer updates, architecture, optimizer, and node-time exposure. Multi-resolution training produces consistently lower projected-flux error; pressure improvement is substantial at several scales but its interval crosses zero at 588 and 2180 nodes.

The representative strong-heterogeneity case makes the mechanism visible. Hydraulic response is relatively diffuse, with each injector distributing approximately 30% - 40% of its induced production among the three producers. At 0.5 PVI, tracer allocation is much more selective: one injector sends 58.5% of recovered tracer to the middle producer, while another sends 55.2% to the third producer. For two of the three injectors the highest-ranked hydraulic connection differs from the highest tracer allocation. The same case has hydraulic-transport rank correlation 0.283 and hydraulic-breakthrough correlation 0.367. Table 4 and Figure 4 summarize the matrix-level and injector-level connectivity comparisons.

Table 4. Hydraulic versus tracer-derived transport connectivity. Matrix-level Spearman correlation is retained as a descriptive statistic; injector-row correlation and tie-aware top-link agreement are the primary within-injector measures. Confidence intervals use reservoir-cluster bootstrap resampling.

Group	PVI	Matrix ρ (H, T)	Injector-row ρ [95% CI]	Top-link agreement [95% CI]	Mean tracer recovery
Moderate	0.50	0.818	0.681 [0.556, 0.792]	72.2% [61.1, 83.3]	6.7%
Strong	0.50	0.719	0.583 [0.403, 0.750]	75.0% [61.1, 88.9]	10.6%
All	0.25	0.744	0.590 [0.465, 0.701]	69.4% [58.3, 79.2]	1.1%
All	0.50	0.769	0.632 [0.514, 0.736]	73.6% [65.3, 81.9]	8.7%
All	1.00	0.773	0.583 [0.465, 0.694]	75.0% [65.3, 83.3]	30.5%

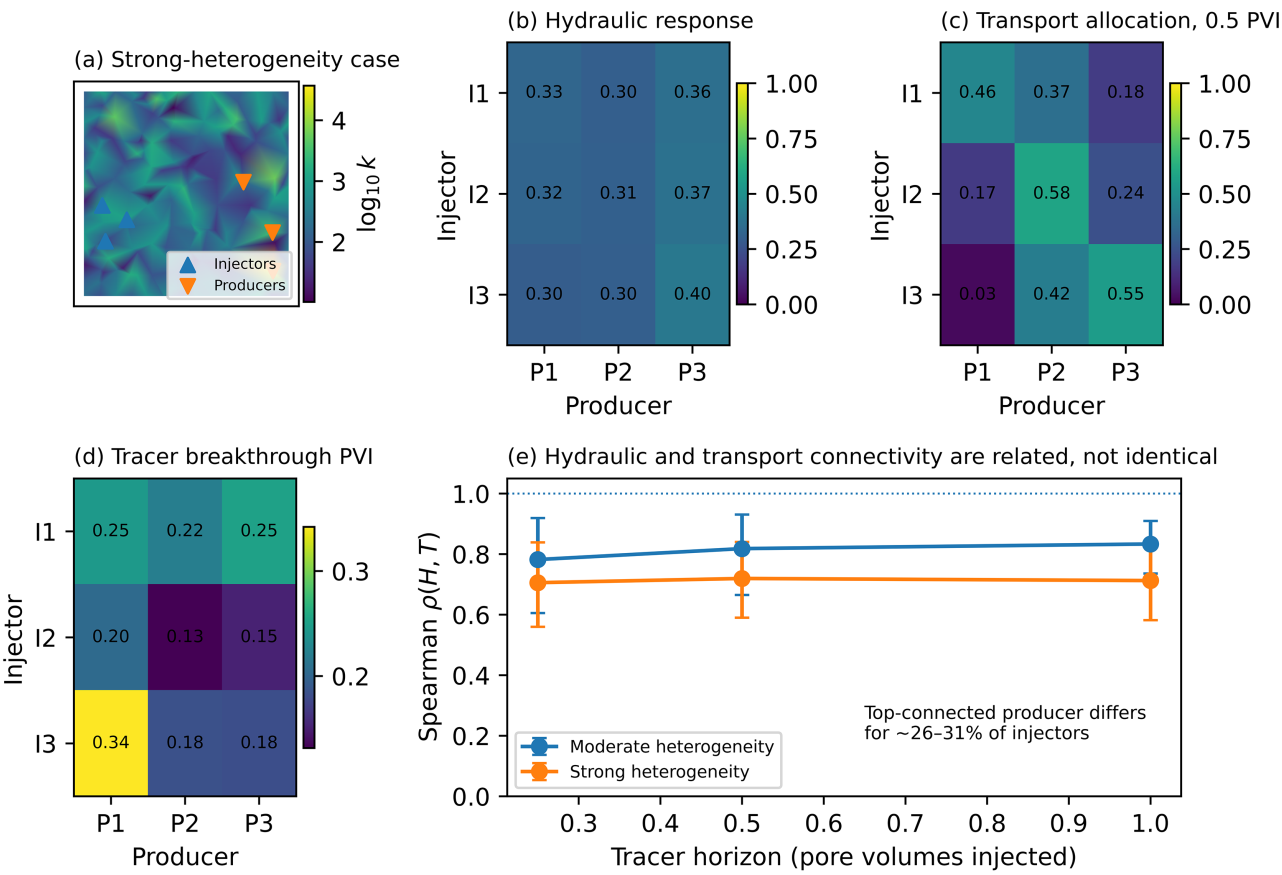

Figure 4. Dual physical connectivity benchmark. Hydraulic connectivity is obtained from high-fidelity injector perturbations; transport connectivity and breakthrough are obtained from conservative tagged tracers. The representative strong-heterogeneity reservoir shows that diffuse pressure response can coexist with selective material pathways. Across 24 reservoirs the two matrices remain positively correlated but disagree on the strongest injector-producer relation in approximately one quarter to one third of injectors.

5.4. Residual Certification Remains Necessary for High-Accuracy Pressure Prediction

The new flux experiments do not replace the earlier pressure-certification result. In Benchmark A, the raw rate-equivariant pressure proposal has a mean relative error of 27.9% - 43.8% across the five suites. PCG correction to normalized residual tolerance 0.1 reduces pressure error to 4.67% ID, 3.91% mesh OOD, 3.18% heterogeneity OOD, and 3.46% control OOD. The deliberately combined mesh-geology-control shift remains harder at 12.05%, establishing a failure envelope rather than a universal-generalization claim. Corresponding normalized saturation errors are 2.31% - 5.94%, and production-rate errors are 1.77% - 13.95%.

The discrete residual also functions as a label-free reliability indicator. Under fixed-budget correction, residual and true pressure error achieve Spearman correlation 0.976 in the heterogeneity shift and 0.914 in the combined shift. This matters because the exact reference pressure is unavailable in deployment, whereas the assembled residual is available wherever the finite-volume operator and current state are known. Residual certification therefore complements, rather than competes with, predictive uncertainty: it directly tests compatibility with the simulator discretization used to define the task.

The rate-equivariant ablation likewise survives the broader study. Direct pressure prediction violates the known common rate scaling by 15.61% ID, 15.22% on finer meshes, 14.77% under stronger heterogeneity, 18.88% for unseen controls, and 22.39% in the combined shift. Predicting pressure per unit total rate eliminates this failure structurally. The result illustrates a general design principle: where an exact symmetry of the discrete physics is known, architectural equivariance is preferable to hoping that finite training data recover it.

5.5. Realistic Solver Baselines Change the Acceleration Claim

A zero-start solver comparison initially made the graph proposal appear highly attractive: at moderate residual targets it substantially reduced PCG iterations. That comparison proved misleading once the rate-scaled previous converged pressure was introduced. During smooth sequential evolution the classical warm start is extremely strong, often requiring only one to four iterations. Used indiscriminately, the graph correction can therefore increase work. This negative result is retained because it materially changes the interpretation of "GNN acceleration."

The tolerance-conditioned selector recovers algebraic value where the state changes abruptly. Here "PCG work" means iterations after initializer selection; each iteration contains one sparse pressure-matrix product, one Jacobi application, and vector operations. It excludes graph construction, feature assembly, GNN inference, both candidate residual evaluations, and selector inference, so we claim PCG-iteration reduction rather than wall-time speedup. At tolerance 0.1, the selector reduces mean iterations by 14.1% [6.3, 21.9] on mesh OOD, 14.3% [4.8, 27.7] under stronger heterogeneity, 19.0% [15.7, 22.2] under rapid switching, and 4.4% [0.5, 8.4] combined; ID crosses zero. On transitions with a unique better candidate, selector agreement is 97.3% - 99.0%, with only 0.03 - 0.18 mean iteration regret relative to an oracle. Table A6 reports the selector results across all requested residual tolerances, and Table A9 gives strict lower-candidate agreement and iteration regret relative to the oracle at tolerance 0.1. Table 5 and Figure 5 report the selector's PCG-iteration behaviour and agreement with the lower-iteration candidate.

Table 5. Tolerance-conditioned initializer selection at residual tolerance 0.1. "PCG iterations" excludes graph/GNN/residual/selector overhead. Strict agreement is computed only on transitions for which the two candidate iteration counts differ. Intervals crossing zero are not interpreted as established iteration reduction.

Suite	Previous PCG	Selector PCG	Reduction [95% CI]	Strict lower-candidate agreement	Switch-step PCG previous → selector
ID	2.122	2.020	5.2% [−4.5, 14.4]	97.3%	27.03 → 25.67
OOD mesh	2.750	2.347	14.1% [6.3, 21.9]	99.0%	42.63 → 35.81
OOD heterogeneity	2.526	2.153	14.3% [4.8, 27.7]	99.0%	32.48 → 27.85
OOD topology	1.224	1.000	14.6% [−5.1, 40.2]	98.5%	15.00 → 12.25
OOD rapid switch	4.255	3.485	19.0% [15.7, 22.2]	97.9%	23.17 → 18.97
OOD combined	3.643	3.510	4.4% [0.5, 8.4]	97.9%	44.63 → 43.00

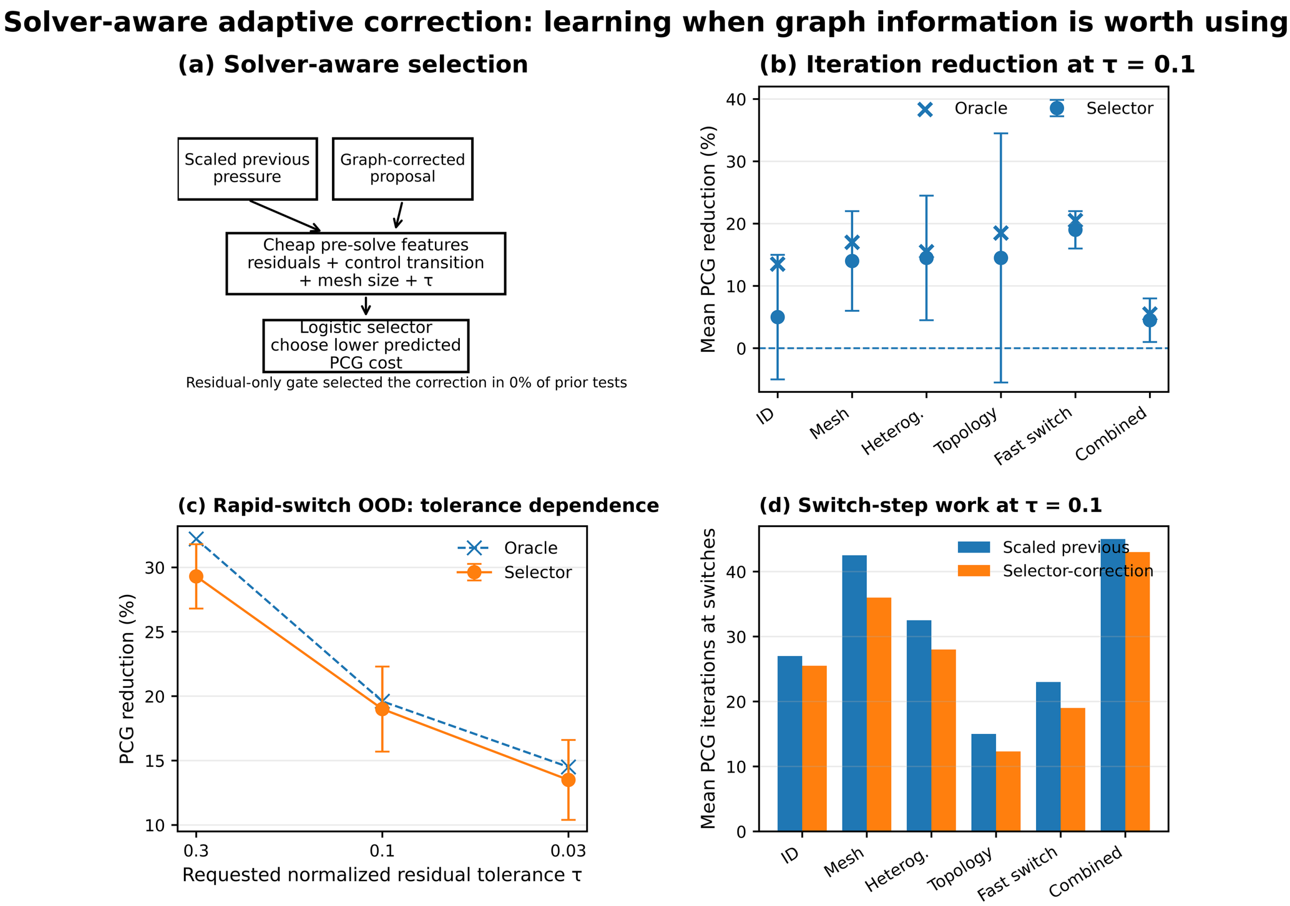

Figure 5. Realistic solver audit and tolerance-conditioned selection. The previous converged pressure is a strong sequential warm start, so graph initialization is not universally beneficial. The learned selector uses pre-solve transition information and requested tolerance to invoke the graph correction selectively, with the clearest benefit under rapid control switching and isolated OOD shifts.

6. Discussion

First, conservation must be given a precise discrete definition. Antisymmetry guarantees pairwise global cancellation, whereas the finite-volume divergence constraint enforces local continuity. The reviewer-requested audit adds a second distinction: continuity is still weaker than the Darcy pressure-flux constitutive relation. Projection removes the divergence error but can retain a divergence-free cycle component of the neural edge field. Thus, state error, local balance, constitutive defect, and transport error should be reported jointly rather than allowing any single diagnostic to stand for physical correctness.

Second, mesh-independence cannot be inferred merely from using a graph representation. The budget-matched rerun shows that multi-resolution exposure robustly improves projected-flux transfer under equal node-time and optimizer budgets, but pressure gains are not statistically resolved at every graph size. The hierarchy result remains similarly qualified: transmissibility aggregation can preserve flux structure, yet it is not universally superior to Euclidean aggregation. For the tested architecture, scale diversity is the more reproducible intervention, especially for flux prediction.

Third, dual connectivity gives interpretability an external physical test. Dynamic graph connectivity, capacitance-resistance models, and streamline allocation [14]-[16] are useful but need not represent the same physics. The within-injector analysis shows that pressure interference and material sweep can rank producers differently, especially under stronger heterogeneity. Connectivity should therefore be named by its measurement process rather than inferred from one generic edge weight.

Fourth, physical structure does not remove the need for numerical baselines. Exact flux projection is an elliptic correction and is not advertised as faster than a pressure solve. Likewise, the pressure GNN is not a universal Krylov accelerator when compared with the previous converged pressure. The selector is useful because it preserves the strong classical warm start on easy transitions and invokes the graph correction selectively.

For deployment, the evidence suggests a layered trust policy. Known symmetries are enforced structurally; graph resolution and controls are checked against training support; learned proposals are evaluated by discrete residual, continuity, and constitutive diagnostics; and difficult states are corrected by finite-volume projection or Krylov iteration. The discrete simulator operator therefore remains part of inference rather than merely a source of training labels.

Taken together, the strongest contribution is not a claim that one architecture replaces a reservoir simulator. It is an evaluation framework in which known symmetry, local conservation, mesh-scale exposure, physical connectivity labels, OOD tests, and realistic solver baselines are treated as first-class evidence. This makes several negative results publishable rather than embarrassing: transmissibility pooling is not universally best; exact flux projection is not free; the graph initializer does not beat a strong warm start on routine steps; and the largest combined distribution shift remains difficult. Those boundaries make the positive findings more credible.

7. Limitations and Future Work

The experiments are two-dimensional, with incompressible immiscible oil-water physics and no gravity, capillarity, compositional flash, geomechanics, or thermal effects. The graph-size stress test reaches 4252 nodes, far below industrial million-cell models. It establishes a scale-transfer trend, not production-scale performance. Future work should repeat the multi-resolution experiment on public 3-D benchmarks such as SPE10 sectors and report memory, preprocessing, partitioning, and wall time in addition to accuracy.

The exact divergence projection uses a sparse elliptic solve of the same operator class as the pressure problem. Its present value is therefore conservation and diagnosis, not guaranteed acceleration. Moreover, exact divergence does not remove the measured non-potential cycle component, so a future constitutive projection would need to constrain both continuity and compatibility with a pressure potential. Approximate differentiable projection, mixed flux-pressure architectures, multigrid projection, or learned preconditioning could reduce cost or inconsistency, but such claims must be benchmarked against strong numerical warm starts rather than zero initialization.

The dual-connectivity benchmark uses passive tracers and controlled perturbations under the same simulator. Field data could introduce model discrepancy, uncertain well completions, rate measurement noise, and unmodeled crossflow. The next validation step should combine public reservoir models with tracer or pulse-test observations where available. Calibrated predictive uncertainty also remains incomplete: bootstrap intervals quantify population uncertainty in reported metrics, not per-case epistemic uncertainty. Deep ensembles and risk coverage analysis are logical extensions, but they should support rather than obscure the present conservation/connectivity thesis.

8. Conclusions

This study asks what evidence is required before an unstructured reservoir GNN can be regarded as physically and numerically credible. Five conclusions emerge. Exact rate normalization removes a 14.8% - 22.4% control-scaling failure of direct pressure regression. Antisymmetric fluxes are not locally conservative: divergence projection drives continuity defects to machine precision and improves flux and saturation, but a measurable Darcy-cycle component remains. Under a strict matched budget, multi-resolution training improves projected-flux transfer at all five graph sizes, while pressure gains are significant at three. Hydraulic and tracer connectivity are distinct: at 0.5 PVI, mean within-injector rank correlation is 0.632 and top-producer agreement is 73.6%. Finally, the learned initializer closely tracks the lower-iteration candidate under OOD transitions, but its demonstrated benefit is PCG iterations rather than end-to-end runtime because graph and selector overheads are excluded.

The resulting perspective is a hybrid one. GNNs are most useful as structure-aware proposals that exploit irregular connectivity, known equivariances, learned flux patterns, and control-transition information. Finite-volume operators remain valuable for conservation, residual certification, and correction. This division of labour produces more defensible claims than either an unconstrained end-to-end surrogate or a physics penalty evaluated only through state RMSE. The accompanying code, trained weights, fixed seeds, numerical outputs, and figure-generation scripts are packaged for independent rerun and extension.

Author Contributions

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

Data and Code Availability

All numerical evidence in this manuscript is computational evidence. The submission package contains scenario generation, model training, evaluation, bootstrap analysis, dual-connectivity tracer calculations, multi-resolution scaling tests, figure-generation scripts, retained model weights, fixed seeds, and machine-readable result files.

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] Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G. (2009) The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, **20**, 61-80. https://doi.org/10.1109/tnn.2008.2005605

[2] Wu, T., Wang, Q., Zhang, Y., Ying, R., Cao, K., Sosic, R., *et al*. (2022) Learning Large-Scale Subsurface Simulations with a Hybrid Graph Network Simulator. *Proceedings of the* 28*th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington DC, 14-18 August 2022, 4184-4194. https://doi.org/10.1145/3534678.3539045

[3] Jiang, J. (2024) Simulating Multiphase Flow in Fractured Media with Graph Neural Networks. *Physics of Fluids*, **36**, Article 023115. https://doi.org/10.1063/5.0189174

[4] Tang, H. and Durlofsky, L.J. (2024) Graph Network Surrogate Model for Subsurface Flow Optimization. *Journal of Computational Physics*, **512**, Article 113132. https://doi.org/10.1016/j.jcp.2024.113132

[5] Ju, X., Hamon, F.P., Wen, G., Kanfar, R., Araya-Polo, M. and Tchelepi, H.A. (2024) Learning CO_2 Plume Migration in Faulted Reservoirs with Graph Neural Networks. *Computers & Geosciences*, **193**, Article 105711.

https://doi.org/10.1016/j.cageo.2024.105711

[6] Chen, H.-Y., Xue, L., Liu, L., Zou, G., Han, J., Dong, Y., *et al*. (2025) Physics-Informed Graph Neural Network for Predicting Fluid Flow in Porous Media. *Petroleum Science*, **22**, 4240-4253. https://doi.org/10.1016/j.petsci.2025.06.007

[7] Gudala, M. and Yan, B. (2025) Graph Neural Network for Multi-Physics Geothermal Simulation with Discrete Fracture Network. *Advances in Water Resources*, **205**, Article 105062. https://doi.org/10.1016/j.advwatres.2025.105062

[8] Aavatsmark, I. (2002) An Introduction to Multipoint Flux Approximations for Quadrilateral Grids. *Computational Geosciences*, **6**, 405-432. https://doi.org/10.1023/a:1021291114475

[9] Aarnes, J.E., Krogstad, S. and Lie, K. (2006) A Hierarchical Multiscale Method for Two-Phase Flow Based Upon Mixed Finite Elements and Nonuniform Coarse Grids. *Multiscale Modeling & Simulation*, **5**, 337-363. https://doi.org/10.1137/050634566

[10] Aarnes, J.E., Kippe, V. and Lie, K. (2005) Mixed Multiscale Finite Elements and Streamline Methods for Reservoir Simulation of Large Geomodels. *Advances in Water Resources*, **28**, 257-271. https://doi.org/10.1016/j.advwatres.2004.10.007

[11] Aarnes, J.E., Krogstad, S. and Lie, K. (2008) Multiscale Mixed/Mimetic Methods on Corner-Point Grids. *Computational Geosciences*, **12**, 297-315. https://doi.org/10.1007/s10596-007-9072-8

[12] Gerritsen, M.G. and Durlofsky, L.J. (2005) Modeling Fluid Flow in Oil Reservoirs. *Annual Review of Fluid Mechanics*, **37**, 211-238. https://doi.org/10.1146/annurev.fluid.37.061903.175748

[13] Gao, H., Zahr, M.J. and Wang, J.-X. (2022) Physics-Informed Graph Neural Galerkin Networks: A Unified Framework for Solving PDE-Governed Forward and Inverse Problems. *Computer Methods in Applied Mechanics and Engineering*, **390**, Article 114502. https://doi.org/10.1016/j.cma.2021.114502

[14] Huang, Z.-Q., Wang, Z.-X., Hu, H.-F., Zhang, S.-M., Liang, Y.-X., Guo, Q. and Yao, J. (2024) Dynamic Interwell Connectivity Analysis of Multi-Layer Waterflooding Reservoirs Based on an Improved Graph Neural Network. *Petroleum Science*, **21**, 1062-1080. https://doi.org/10.1016/j.petsci.2023.11.008

[15] Yousef, A.A., Gentil, P.H., Jensen, J.L. and Lake, L.W. (2006) A Capacitance Model to Infer Interwell Connectivity from Production and Injection Rate Fluctuations. *SPE Reservoir Evaluation & Engineering*, **9**, 630-646. https://doi.org/10.2118/95322-pa

[16] Thiele, M.R. and Batycky, R.P. (2006) Using Streamline-Derived Injection Efficiencies for Improved Waterflood Management. *SPE Reservoir Evaluation & Engineering*, **9**, 187-196. https://doi.org/10.2118/84080-pa

[17] Peaceman, D.W. (1978) Interpretation of Well-Block Pressures in Numerical Reservoir Simulation. *Society of Petroleum Engineers Journal*, **18**, 183-194. https://doi.org/10.2118/6893-pa

[18] Hestenes, M.R. and Stiefel, E. (1952) Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, **49**, 409-436. https://doi.org/10.6028/jres.049.044

[19] Benzi, M. (2002) Preconditioning Techniques for Large Linear Systems: A Survey. *Journal of Computational Physics*, **182**, 418-477. https://doi.org/10.1006/jcph.2002.7176

[20] Saad, Y. and Schultz, M.H. (1986) GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM Journal on Scientific and Statistical Computing*, **7**, 856-869. https://doi.org/10.1137/0907058

[21] Eisenstat, S.C. and Walker, H.F. (1996) Choosing the Forcing Terms in an Inexact Newton Method. *SIAM Journal on Scientific Computing*, **17**, 16-32. https://doi.org/10.1137/0917003

[22] 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

[23] Kadeethum, T., Verzi, S.J. and Yoon, H. (2024) An Improved Neural Operator Framework for Large-Scale CO_2 Storage Operations. *Geoenergy Science and Engineering*, **240**, Article 213007. https://doi.org/10.1016/j.geoen.2024.213007

[24] 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

[25] Wen, G., Hay, C. and Benson, S.M. (2021) CCSNet: A Deep Learning Modeling Suite for CO_2 Storage. *Advances in Water Resources*, **155**, Article 104009. https://doi.org/10.1016/j.advwatres.2021.104009

[26] Mo, S., Zhu, Y., Zabaras, N., Shi, X. and Wu, J. (2019) Deep Convolutional Encoder-decoder Networks for Uncertainty Quantification of Dynamic Multiphase Flow in Heterogeneous Media. *Water Resources Research*, **55**, 703-728. https://doi.org/10.1029/2018wr023528

[27] Zhu, Y. and Zabaras, N. (2018) Bayesian Deep Convolutional Encoder-Decoder Networks for Surrogate Modeling and Uncertainty Quantification. *Journal of Computational Physics*, **366**, 415-447. https://doi.org/10.1016/j.jcp.2018.04.018

[28] Zhu, Y., Zabaras, N., Koutsourelakis, P. and Perdikaris, P. (2019) Physics-Constrained Deep Learning for High-Dimensional Surrogate Modeling and Uncertainty Quantification without Labeled Data. *Journal of Computational Physics*, **394**, 56-81. https://doi.org/10.1016/j.jcp.2019.05.024

[29] 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

[30] 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

[31] Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M. and Piccialli, F. (2022) Scientific Machine Learning through Physics-Informed Neural Networks: Where We Are and What's Next. *Journal of Scientific Computing*, **92**, Article No. 88. https://doi.org/10.1007/s10915-022-01939-z

[32] Tartakovsky, A.M., Marrero, C.O., Perdikaris, P., Tartakovsky, G.D. and Barajas-Solano, D. (2020) Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems. *Water Resources Research*, **56**, e2019WR026731. https://doi.org/10.1029/2019wr026731

[33] 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

[34] Kenzhebek, Y., Imankulov, T., Bekele, S.D., Panfilova, I. and Akhmed-Zaki, D. (2025) Coupled Pressure and Saturation Prediction for Two-Phase Flow in Porous Media

Using Physics-Informed Neural Networks (PINNs). *Computer Methods in Applied Mechanics and Engineering*, **446**, Article 118229. https://doi.org/10.1016/j.cma.2025.118229

[35] Abbasi, J., Moseley, B., Kurotori, T., Jagtap, A.D., Kovscek, A.R., Hiorth, A., *et al.* (2025) History-Matching of Imbibition Flow in Fractured Porous Media Using Physics-Informed Neural Networks (PINNs). *Computer Methods in Applied Mechanics and Engineering*, **437**, Article 117784. https://doi.org/10.1016/j.cma.2025.117784

[36] Walter, L., Parisio, F., Kong, Q., Hanson-Hedgecock, S. and Vilarrasa, V. (2025) Bridging the Gap for Sparse Observations: Data-Driven versus Physics-Informed Neural Networks for a Transient Flow Problem in a Heterogeneous Porous Medium. *Journal of Geophysical Research*: *Machine Learning and Computation*, **2**, e2025JH000733. https://doi.org/10.1029/2025jh000733

[37] Kashefi, A. and Mukerji, T. (2023) Prediction of Fluid Flow in Porous Media by Sparse Observations and Physics-Informed PointNet. *Neural Networks*, **167**, 80-91. https://doi.org/10.1016/j.neunet.2023.08.006

[38] Sorokin, A.G., Pachalieva, A., O'Malley, D., Hyman, J.M., Hickernell, F.J. and Hengartner, N.W. (2024) Computationally Efficient and Error Aware Surrogate Construction for Numerical Solutions of Subsurface Flow through Porous Media. *Advances in Water Resources*, **193**, Article 104836. https://doi.org/10.1016/j.advwatres.2024.104836

[39] Zhao, M., Wang, Y., Gerritsma, M. and Hajibeygi, H. (2024) A Physics-Constraint Neural Network for CO_2 Storage in Deep Saline Aquifers during Injection and Post-Injection Periods. *Advances in Water Resources*, **193**, Article 104837. https://doi.org/10.1016/j.advwatres.2024.104837

[40] Latrach, A., Malki, M.L., Morales, M., Mehana, M. and Rabiei, M. (2024) A Critical Review of Physics-Informed Machine Learning Applications in Subsurface Energy Systems. *Geoenergy Science and Engineering*, **239**, Article 212938. https://doi.org/10.1016/j.geoen.2024.212938

[41] 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

[42] 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

[43] Jain, N., Roy, S., Kodamana, H. and Nair, P. (2025) Scaling the Predictions of Multiphase Flow through Porous Media Using Operator Learning. *Chemical Engineering Journal*, **503**, Article 157671. https://doi.org/10.1016/j.cej.2024.157671

[44] 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

[45] Liu, J., Pan, H., Sun, W., Jing, H. and Gong, B. (2025) Extension of Fourier Neural Operator from Three-Dimensional (X, Y, T) to Four-Dimensional (X, Y, Z, T) Subsurface Flow Simulation. *Mathematical Geosciences*, **57**, 359-391. https://doi.org/10.1007/s11004-024-10152-7

[46] Mao, S., Carbonero Gonzales, A.R. and Mehana, M.Z.S. (2025) Deep Learning for Subsurface Flow: A Comparative Study of U-Net, Fourier Neural Operators, and Transformers in Underground Hydrogen Storage. *Journal of Geophysical Research*: *Machine Learning and Computation*, **2**, e2024JH000401.

https://doi.org/10.1029/2024jh000401

[47] Chandra, A., Koch, M., Pawar, S., Panda, A., Azizzadenesheli, K., Snippe, J., *et al*. (2026) Accelerating Porous Media Flow Simulations with Fourier Neural Operators: An Application to Geologic Storage of CO_2. *Advanced Theory and Simulations*, **9**, e00747. https://doi.org/10.1002/adts.202500747

[48] Geneva, N. and Zabaras, N. (2020) Modeling the Dynamics of PDE Systems with Physics-Constrained Deep Auto-Regressive Networks. *Journal of Computational Physics*, **403**, Article 109056. https://doi.org/10.1016/j.jcp.2019.109056

[49] Jenny, P., Lee, S.H. and Tchelepi, H.A. (2003) Multi-Scale Finite-Volume Method for Elliptic Problems in Subsurface Flow Simulation. *Journal of Computational Physics*, **187**, 47-67. https://doi.org/10.1016/s0021-9991(03)00075-5

[50] Lie, K.-A., Krogstad, S., Ligaarden, I.S., Natvig, J.R., Nilsen, H.M. and Skaflestad, B. (2012) Open-Source MATLAB Implementation of Consistent Discretisations on Complex Grids. *Computational Geosciences*, **16**, 297-322. https://doi.org/10.1007/s10596-011-9244-4

Appendix

A1. Scope and Evidence Hierarchy

This supplement documents diagnostics and ablations supporting the five main figures and the reviewer-requested reproducibility, constitutive-consistency, budget-matching, injector-level connectivity, and solver-work audits. All numerical evidence is synthetic computational evidence. The reference solver and learned models use fixed random seeds and explicit train/validation/test separation.

A2. Reference-Solver and Tracer Quality Control

Principal two-phase reference trajectories were rejected or regenerated if the explicit transport safety guard clipped saturation. The retained suites show non-trivial saturation motion and global water-balance defects at floating-point scale. The independent tracer benchmark was also audited for mass conservation and bounded concentration.

Table A1. Reference-solver quality-control statistics for Benchmark A.

Suite	Mean nodes	Mean ΔS	Min ΔS	Total clipping	Max |balance defect|
ID	173	0.396	0.322	0	3.47e−17
OOD-mesh	280	0.387	0.328	0	2.26e−17
OOD-heterogeneity	173	0.395	0.388	0	2.95e−17
OOD-controls	173	0.365	0.355	0	3.47e−17
OOD-combined	300	0.337	0.317	0	2.08e−17

Across the 24 dual-connectivity reservoirs, the passive-tracer integrator used between 225 and 2097 explicit steps depending on case stability. The maximum recorded tracer mass error was 1.43×10^{-10}; concentration remained within numerical tolerance of [0, 1], with an observed maximum of 1.0000000000000167. These audits are important because the hydraulic-versus-transport comparison would be uninterpretable if tracer attribution itself violated material conservation.

A3. Model Architecture and Optimization

Table A2. Reproducibility specification for the principal physical, numerical, graph-model, and training settings. Exact scripts, seeds, sampled transitions, and retained weights are included in the revision archive.

Component	Setting
Domain/unstructured grid	500 m × 500 m 2-D layer; independently generated Delaunay/Voronoi-like cell graph; thickness 15 - 25 m
Pressure datum	Producer-completion cells impose $p = 0$ Pa gauge; only pressure differences enter evaluation
Common rate scale Q	Total injector-patch source rate before SI conversion; base $Q = 700$ - 1300 m³/day

Continued

Source/sink allocation	Injector source distributed to completion cells in proportion to pore volume; producer rate emerges from fixed-pressure sink
Control schedules	Training: 3 piecewise-constant multipliers 0.8 - 1.2; control OOD: 0.40 - 0.60, 1.25 - 1.40, 0.65 - 0.80
Fluids	Incompressible oil-water; $\mu_w = 1$ mPa·s; $\mu_o = 5$ mPa·s; gravity and capillarity omitted
Relative permeability	Corey-type quadratic curves; $S_{wc} = 0.15$, $S_{or} = 0.15$; initial $S_w = 0.20$
Porosity	Normal mean 0.21, standard deviation 0.025, clipped to 0.13 - 0.30
Permeability field	64 random Fourier/cosine modes; lognormal; geometric mean 100 - 220 mD; train log-$\sigma \approx$ 0.55 - 1.05; correlation parameter 70 - 140 m; strong OOD log-$\sigma = 1.25$, correlation 65 m
Well geometry/constraints	Separated injector and producer patches; producer completion pressure fixed at datum; no prescribed producer rate
Time integration	Pressure implicit each step; water first-order upwind finite volume; $\Delta t = 1.25$ days; retained reference trajectories have zero saturation clipping
Benchmark A pressure GNN	3 message-passing layers; 32 latent channels; rate-equivariant pressure-per-Q representation
Edge-flux GNN	66,242 parameters; 40 latent channels; 2 fine + 2 coarse message-passing blocks; pressure-per-Q and unique-edge-flux-per-Q heads
Edge-flux loss	Smooth-L1 pressure + 1.25 × Smooth-L1 unique-edge flux; no conservation penalty
Edge-flux optimizer	AdamW; initial lr 1.2×10^{-3} (budget rerun 1.1×10^{-3}); weight decay 2×10^{-6}; gradient-norm cap 1.0
Single-resolution sampled states	8 training reservoirs × 40 transitions = 320; 2 validation reservoirs × 40 = 80
Budget-matched scale rerun	8 reservoirs/regime; 140 optimizer updates/regime; 481,632 vs 481,600 node-time presentations; same architecture, optimizer family and validation reservoirs
Dual connectivity	24 reservoirs; 3 injectors × 3 producers; perturbation hydraulic labels and tagged tracers at 0.25, 0.5 and 1.0 PVI; breakthrough threshold 10^{-3}
Tolerance-conditioned selector	Logistic classifier trained on training/validation transitions only; features include candidate residuals, support/rate changes, correction magnitude, graph size and requested tolerance

A4. Antisymmetry, Local Conservation, and Projection Budget

The raw flux head predicts one value per unique edge and assigns the opposite sign to the reverse incidence. This guarantees pairwise cancellation in the global internal sum but does not force the divergence of the edge field to equal the free-cell source vector. Figure A1 shows the hierarchy ablation after exact projection; Figure A2 shows the case-level reduction in long-horizon saturation error when the same learned flux proposal is projected before transport.

The limited-iteration experiment constrains the computational claim. For ID reservoirs, 16 Jacobi-preconditioned CG iterations reduce continuity defect from 1.37 to 0.259 and saturation error from 26.7% to 14.7%, but exact projection reaches approximately 1.64×10^{-14} continuity defect and 3.99% saturation error. The same pattern persists under mesh, heterogeneity, control, and combined shifts. Exact projection is therefore used as a physics-certification layer in the paper, not advertised as a free acceleration mechanism.

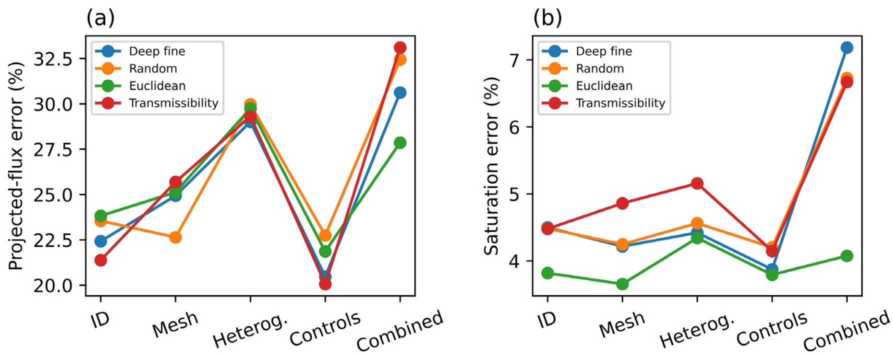

Figure A1. Single-resolution hierarchy ablation after exact divergence projection. No coarse rule dominates every pressure and transport metric; the Euclidean model is often competitive or superior in saturation error, while other rules can be preferable in flux error.

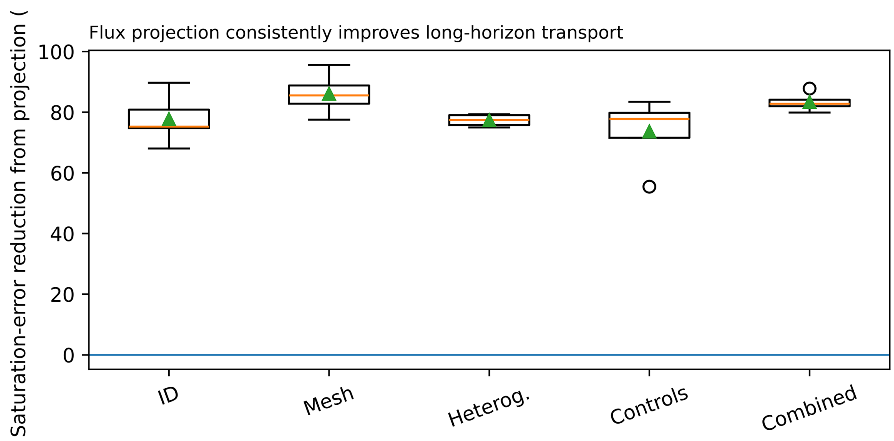

Figure A2. Reservoir-level reduction in long-horizon saturation error obtained by projecting the raw antisymmetric neural flux before conservative transport. Boxes summarize held-out cases and triangles denote means.

Table A3. Limited-iteration projection audit. "Exact" denotes convergence of the sparse correction solve. Finite budgets improve the raw field but do not reproduce the exact conservation result.

Suite	Projection budget	Continuity defect	Flux error	Saturation error	Clip mass
ID	0	1.37	54.00%	26.69%	2258.2
ID	8	0.32	39.26%	16.29%	1138.6
ID	16	0.259	36.51%	14.72%	1836.9
ID	exact	1.64e−14	21.08%	3.99%	0.0

Continued

OOD-mesh	0	1.63	52.30%	28.38%	1753.4
OOD-mesh	8	0.409	41.76%	14.55%	672.3
OOD-mesh	16	0.295	37.53%	8.58%	0.0
OOD-mesh	exact	1.9e−14	24.32%	3.82%	0.0
OOD-heterogeneity	0	1.5	50.19%	18.12%	0.0
OOD-heterogeneity	8	0.428	42.87%	9.10%	66.1
OOD-heterogeneity	16	0.401	42.05%	7.58%	0.0
OOD-heterogeneity	exact	7.54e−14	27.05%	3.90%	0.0
OOD-controls	0	1.29	49.47%	17.59%	0.0
OOD-controls	8	0.322	35.40%	8.58%	218.8
OOD-controls	16	0.202	32.21%	4.80%	0.0
OOD-controls	exact	8.6e−15	22.26%	3.54%	0.0
OOD-combined	0	2.34	57.18%	28.47%	6968.3
OOD-combined	8	0.689	46.83%	21.44%	2175.8
OOD-combined	16	0.51	44.50%	16.36%	1403.6
OOD-combined	exact	5.97e−14	25.51%	3.96%	0.0

A5. Multi-Resolution Transfer and Hierarchy Falsification

The original unequal-exposure sweep is retained only as exploratory development evidence and is superseded for primary inference by Table A4. Under the strict matched budget, projected-flux improvement remains positive at all five sizes, whereas the pressure confidence interval crosses zero at 588 and 2180 nodes. The separate transmissibility-versus-Euclidean hierarchy comparison in Figure A3 is unchanged and remains a falsification result: no coarse rule dominates all field and flux metrics.

Table A4. Training-budget-matched frozen-model scale transfer. Both regimes use eight training reservoir cases, 140 optimizer updates, identical 66,242-parameter models, and 481,632 versus 481,600 node-time presentations. Each graph size contains eight unseen reservoirs; confidence intervals are reservoir-level paired bootstraps.

Nodes (n = 8)	p single	p multi	p reduction [95% CI]	Flux single	Flux multi	Flux reduction [95% CI]
314	36.1%	19.4%	35.7% [9.3, 57.9]	38.0%	21.6%	42.0% [34.6, 50.2]
588	54.3%	31.6%	25.3% [−6.2, 52.4]	53.7%	27.0%	49.5% [42.1, 56.6]
1128	51.5%	17.7%	58.7% [39.6, 74.0]	55.7%	21.5%	61.3% [59.3, 63.5]
2180	56.1%	40.1%	23.7% [−6.6, 51.8]	79.7%	27.6%	65.6% [63.1, 68.2]
4252	79.9%	27.1%	62.6% [55.9, 69.0]	111.7%	28.8%	73.2% [69.1, 77.5]

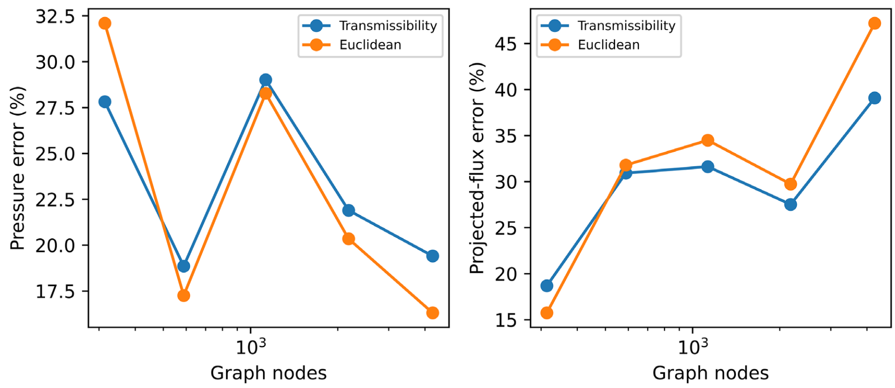

Figure A3. Matched multi-resolution transmissibility versus Euclidean hierarchy on progressively larger graphs. Euclidean aggregation gives lower pressure error at the four largest sizes, whereas transmissibility aggregation gives lower projected-flux error at those sizes.

A6. Dual Physical Connectivity

Across all 24 reservoirs, every injector row has positive tracer recovery at each reported horizon, so no row is discarded. Spearman ranks use average ranks for ties. Top-link agreement treats tied maxima as sets and counts any overlap as agreement. Mean injected tracer recovered is 1.1%, 8.7%, and 30.5% at 0.25, 0.5, and 1.0 PVI. Nine of 216 producer pairs (4.17%) have no breakthrough above 10^{-3} by 1.0 PVI; these pairs are right-censored rather than assigned an artificial rank. Among observed producers, mean within-injector hydraulic-versus-breakthrough Spearman agreement is 0.556 [0.417, 0.681] and earliest-breakthrough top agreement is 69.4% [58.3, 80.6].

Table A5. Hydraulic-transport agreement with matrix-level and injector-level summaries. Matrix ρ is descriptive; within-injector rank correlation and tie-aware top-link agreement are the primary measures.

Group	PVI	Matrix ρ (H, T)	Injector-row ρ [95% CI]	Top agreement [95% CI]	Tracer recovery
Moderate	0.50	0.818	0.681 [0.556, 0.792]	72.2% [61.1, 83.3]	6.7%
Strong	0.50	0.719	0.583 [0.403, 0.750]	75.0% [61.1, 88.9]	10.6%
All	0.25	0.744	0.590 [0.465, 0.701]	69.4% [58.3, 79.2]	1.1%
All	0.50	0.769	0.632 [0.514, 0.736]	73.6% [65.3, 81.9]	8.7%
All	1.00	0.773	0.583 [0.465, 0.694]	75.0% [65.3, 83.3]	30.5%

A7. Sequential Solver Selector at All Tolerances

The selector uses only pre-solve transition features and the requested tolerance. "PCG work" throughout this study is algebraic iteration count after initializer selection: one iteration comprises one sparse pressure-matrix product, one Jacobi preconditioner application, and associated vector/dot operations. The metric excludes graph construction, feature assembly, GNN correction inference, both candidate residual evaluations, and selector inference; accordingly, the revision claims iteration reduction rather than end-to-end wall-time speedup. The residual-only negative control almost never selects the graph correction, showing that smaller initial residual is not identical to lower finite-tolerance Krylov cost.

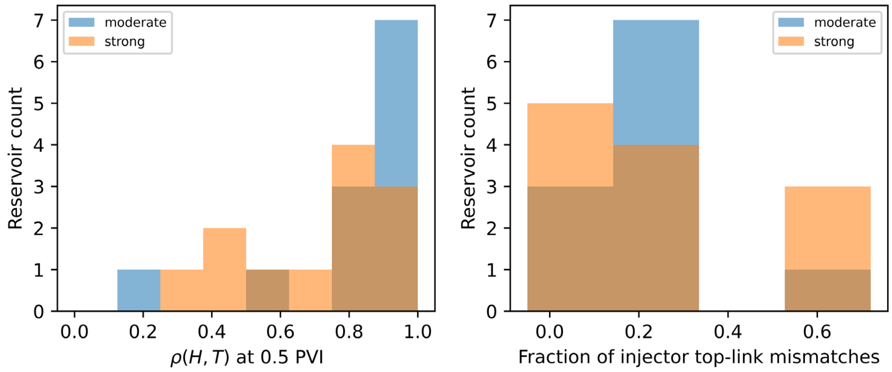

Figure A4. Reservoir-level distributions of hydraulic-transport rank correlation and injector top-link mismatch at 0.5 PVI. Strong heterogeneity broadens the distribution and increases the frequency of weak agreement.

Table A6. Tolerance-conditioned selector results across requested residual tolerances. "Previous", "Selector", and "Oracle" are mean PCG iteration counts after initialization; graph/GNN/residual/selector overhead is excluded. The graph-selection-frequency column is not selector accuracy. Intervals crossing zero are not interpreted as established iteration reduction.

Suite	Tol.	Previous	Selector	Oracle	Graph-selection frequency	Reduction [95% CI]
ID	0.3	1.173	0.935	0.901	7.8%	21.1% [13.6, 28.4]
ID	0.1	2.122	2.020	1.837	7.8%	5.2% [−4.5, 14.4]
ID	0.03	3.568	3.401	3.163	7.8%	4.4% [−9.7, 14.4]
Mesh OOD	0.3	1.189	1.061	0.974	7.1%	9.4% [−8.2, 23.8]
Mesh OOD	0.1	2.750	2.347	2.276	7.1%	14.1% [6.3, 21.9]
Mesh OOD	0.03	4.245	4.097	3.959	7.1%	3.5% [−1.4, 8.4]
Heterogeneity OOD	0.3	0.944	0.781	0.750	7.7%	18.3% [10.5, 26.1]
Heterogeneity OOD	0.1	2.526	2.153	2.112	7.7%	14.3% [4.8, 27.7]
Heterogeneity OOD	0.03	3.714	3.480	3.372	7.7%	5.3% [−3.0, 13.6]
Topology OOD	0.3	0.546	0.520	0.474	6.6%	3.6% [−19.0, 26.2]
Topology OOD	0.1	1.224	1.000	0.969	6.6%	14.6% [−5.1, 40.2]
Topology OOD	0.03	2.143	1.755	1.735	6.6%	17.3% [11.4, 25.0]
Rapid-switch OOD	0.3	1.740	1.224	1.173	17.9%	29.3% [26.9, 31.8]
Rapid-switch OOD	0.1	4.255	3.485	3.454	17.9%	19.0% [15.7, 22.2]
Rapid-switch OOD	0.03	6.903	6.000	5.939	17.9%	13.4% [10.4, 17.5]
Combined OOD	0.3	1.837	1.648	1.587	8.2%	8.8% [0.4, 17.0]
Combined OOD	0.1	3.643	3.510	3.474	8.2%	4.4% [0.5, 8.4]
Combined OOD	0.03	5.444	5.281	5.138	8.2%	3.7% [−3.3, 10.7]

A8. Reproducibility and Claim Boundaries

The original submission reproducibility archive remains the source for the unchanged pressure/residual benchmark, multiwell warm-start and selector experiments, edge-flux/projection experiments, dual-connectivity tracer calculations, model weights, fixed seeds, and figure-generation code. For this revision, the response package adds machine-readable CSV exports of the matched-budget scale-transfer results, post-projection Darcy-consistency audit, injector-row connectivity analysis, and selector-agreement audit, together with the revised high-resolution figures. These files document the numerical values used in the response and revised tables without implying that the journal-facing resubmission is a replacement for a permanent public repository.

The present evidence does not establish industrial million-cell runtime, compositional or thermal physics, capillarity, geomechanics, or field-data validity. The 4252-node scale test is deliberately described as a mesh-transfer stress test. Exact projection is treated as a conservation/certification solve rather than a speedup. Bootstrap intervals quantify uncertainty across the sampled synthetic reservoirs but do not constitute calibrated per-case epistemic uncertainty.

A9. Post-Projection Constitutive-Consistency Audit

Exact local divergence is necessary but not sufficient for Darcy consistency. For each held-out state, the projected flux is compared with the Darcy-consistent gradient flux computed from the same saturation-dependent transmissibility/mobility field and the same producer pressure datum. The primary diagnostic is the inverse-transmissibility-weighted norm of the divergence-free cycle component relative to the projected-flux energy. A second metric reports edge-sign agreement with the pressure-gradient flux. Confidence intervals resample whole reservoir realizations.

Table A7. Post-projection constitutive-consistency audit. A zero Darcy-cycle defect would indicate that the locally conservative edge field is also compatible with a pressure potential in the tested discrete metric.

Suite	n	Darcy-cycle defect [95% CI]	Cycle energy fraction [95% CI]	Darcy sign agreement [95% CI]
ID	5	0.271 [0.221, 0.313]	0.261 [0.215, 0.299]	0.838 [0.816, 0.852]
OOD mesh	4	0.369 [0.272, 0.447]	0.343 [0.259, 0.408]	0.810 [0.796, 0.827]
OOD heterogeneity	4	0.359 [0.315, 0.437]	0.336 [0.301, 0.398]	0.787 [0.755, 0.819]
OOD controls	4	0.281 [0.222, 0.340]	0.269 [0.217, 0.321]	0.815 [0.774, 0.852]
OOD combined	4	0.588 [0.331, 0.924]	0.468 [0.314, 0.643]	0.784 [0.766, 0.797]

A10. Injector-Level Connectivity Agreement and Censoring

Row-normalized transport values are dependent within each injector, so flattened matrix correlation is not used as the primary inferential statistic. The revised analysis calculates Spearman agreement separately within each injector row and applies reservoir-cluster bootstrap resampling. Average ranks are used for ties; top-link agreement is set-valued; zero-recovery rows would be excluded; and producer pairs without breakthrough above 10^{-3} by 1.0 PVI are right-censored. In the present 24-reservoir dataset no injector row is unrecovered, while 9/216 producer pairs are censored for breakthrough.

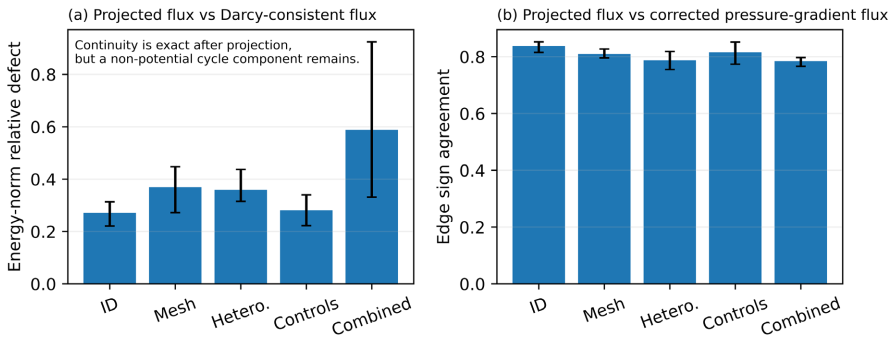

Figure A5. Post-projection constitutive consistency. The divergence projection enforces local continuity to numerical precision but leaves a nonzero divergence-free, non-potential component; the combined OOD suite has the largest defect.

Table A8. Per-injector hydraulic-transport agreement with explicit tie and censoring rules. Confidence intervals use reservoir-cluster bootstrap resampling.

Group/horizon	Reservoirs	Injectors	Row ρ [95% CI]	Top-link agreement [95% CI]	Unrecovered rows
all, 0.25 PVI	24	72	0.590 [0.465, 0.701]	69.4% [58.3, 79.2]	0
all, 0.5 PVI	24	72	0.632 [0.514, 0.736]	73.6% [65.3, 81.9]	0
all, 1.0 PVI	24	72	0.583 [0.465, 0.694]	75.0% [65.3, 83.3]	0
moderate, 0.5 PVI	12	36	0.681 [0.556, 0.792]	72.2% [61.1, 83.3]	0
strong, 0.5 PVI	12	36	0.583 [0.403, 0.750]	75.0% [61.1, 88.9]	0
Breakthrough, observed	24	72	0.556 [0.417, 0.681]	69.4% [58.3, 80.6]	9/216 pairs censored

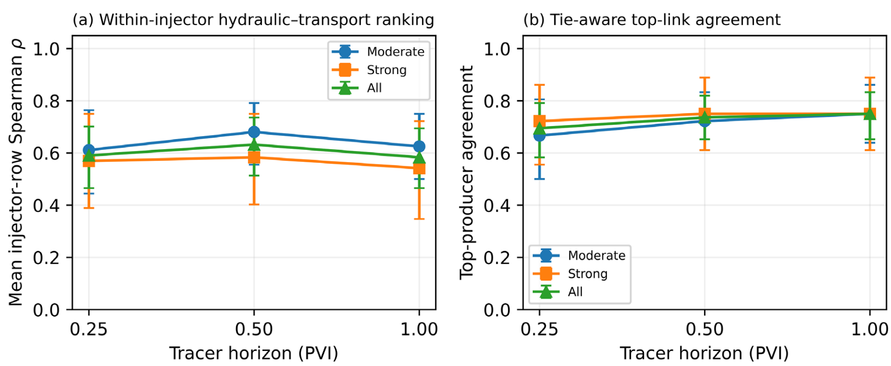

Figure A6. Injector-level connectivity agreement. Row-wise analysis confirms positive but incomplete agreement and avoids treating dependent row-normalized producer fractions as independent matrix entries.

A11. PCG-Work Definition and Selector Agreement Audit

The selector comparison is an algebraic-work audit, not an end-to-end runtime benchmark. Candidate residuals and classifier features are deliberately evaluated before the PCG solve, but their construction time is outside the reported iteration metric. To test whether the selector actually chooses the lower-iteration candidate, we recomputed every transition at tolerance 0.1 and excluded ties from the strict-agreement denominator. Oracle-equivalent means the selected candidate attains the minimum of the two-candidate iteration counts, including ties.

Table A9. Selector agreement with the lower-PCG-iteration candidate at requested normalized residual tolerance 0.1. Strict agreement excludes ties; regret is mean iterations per transition relative to an oracle chooser.

Suite	Transitions	Tie fraction	Strict lower-candidate agreement	Oracle-equivalent	Iteration regret vs oracle
ID	294	1.0%	97.3%	97.3%	0.184
OOD mesh	196	1.0%	99.0%	99.0%	0.071
OOD heterogeneity	196	1.0%	99.0%	99.0%	0.041
OOD topology	196	1.0%	98.5%	98.5%	0.026
OOD rapid switch	196	1.0%	97.9%	98.0%	0.046
OOD combined	196	1.0%	97.9%	98.0%	0.036