PHYSICS-INFORMED RECONSTRUCTION OF SPARSE HEMODYNAMIC FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Physics-informed neural networks (PINNs) embed governing partial differential equations as constraints within the training loss, but naively weighting this constraint from the start of training traps optimization near a trivial zero-field solution, a premature-physics failure mode that has limited their reliability for sparse-data field recovery. We introduce a training methodology combining a three-phase physics-weight curriculum with exponential-moving-average residual scaling, which defers strong physics enforcement until the network has learned an adequate data-driven representation and adaptively balances multiple governing-equation residuals without manual tuning. We study this methodology on the recovery of complete hemodynamic fields – velocity, pressure, and passive-scalar concentration – from sparse observations, using a CFD-simulated flow within a patient-specific three-dimensional intracranial aneurysm geometry derived from CT imaging, complemented by a synthetic helical-pipe flow with a known analytical solution. Against linear interpolation, proper orthogonal decomposition, and a physics-free network, our approach achieves mean relative ℓ_2 error 0.115 at 10% measurement density – a $4.1\times$, $5.3\times$, and $2.6\times$ improvement respectively – decreasing monotonically to 0.082 at full density. An architecture-matched ablation isolates a further $1.8\times$ improvement attributable to the embedded governing equations alone, independent of network capacity, while a five-configuration ablation over the curriculum, residual scaling, and gradient clipping confirms the curriculum as the dominant stabilizer, producing a $3.49\times$ error increase when removed. We further probe the method's generalization through robustness to Gaussian measurement noise and sensitivity to spatial acquisition strategy. Throughout, pressure recovery remains directly coupled to the velocity field via the momentum equation, a dynamical consistency unavailable to baselines that fit each field independently. Together, these results show that principled training-schedule design, not architecture alone, is critical to realizing physics-informed learning's promise for sparse biomedical field reconstruction.

1 INTRODUCTION

Reconstructing a complete, time-resolved three-dimensional velocity, pressure, and concentration field from sparse, noisy point measurements is a recurring problem across computational hemodynamics: 4D flow MRI acquires a fraction of the voxels a CFD solver would use (Markl et al., 2012), sensor placement in a phantom or catheterized vessel is limited by physical access, and patient-specific simulation pipelines are frequently built to compensate for exactly this kind of missing data (Taylor & Figueroa, 2009). A model that could accurately extrapolate from 1–10% of a full-resolution flow field to the whole domain would directly reduce acquisition time, scanner cost, and patient exposure in a clinical pipeline. Physics-informed neural networks (PINNs) are, in principle, well suited to this task: by penalizing violations of the governing partial differential equations (PDEs) at unlabeled collocation points, they can regularize an otherwise underdetermined interpolation problem with exact physical law rather than a learned or hand-designed prior (Raissi et al., 2019; 2020). In practice, however, PINNs trained by naively summing a data-fitting loss and a physics-residual loss from the first optimization step are notoriously difficult to train: gradient interference between the two loss terms, and the relative ease of satisfying the PDE residual at a trivial (e.g., near-zero or near-constant) field, can trap the optimizer in a region of parameter space where the physics loss is small precisely because the network has not yet learned anything useful about the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

data (Krishnapriyan et al., 2021; Wang et al., 2021). We refer to this as the *premature-physics failure mode*: physics enforced before a competent data-driven representation exists does not regularize learning, it short-circuits it.

This paper studies a training methodology that avoids this failure mode by treating the physics-weight schedule itself as a first-class design object, rather than a fixed hyperparameter. Concretely, we make the following contributions:

- **A curriculum-scheduled training methodology for PINNs.** We introduce a *three-phase physics-weight curriculum* that defers strong physics enforcement until a data-fitting phase has produced a competent field estimate, then introduces the governing-equation residual gently before enforcing it in full (Section 3.3), paired with *exponential-moving-average (EMA) residual scaling* that adaptively balances the five coupled PDE residuals – continuity, three momentum components, and scalar transport – without manual per-equation tuning (Section 3.4).
- **A sparse hemodynamic field reconstruction benchmark.** We evaluate the resulting method on the recovery of complete velocity, pressure, and passive-scalar concentration fields across two geometries – a real, patient-specific intracranial aneurysm derived from CT imaging and CFD, and a synthetic helical pipe with a known analytical solution – showing consistent, large accuracy gains over interpolation, proper orthogonal decomposition (POD), and a depth-and-width-matched physics-free network across a 1%–100% measurement-density sweep (Section 5.1).
- **An architecture-matched ablation and a pressure-recovery study.** We isolate the contribution of the embedded governing equations from the contribution of network capacity alone by comparing against a physics-free network of identical depth and width (Section 5.2), and we show, by removing pressure supervision entirely, that the momentum-equation coupling lets velocity and concentration be learned almost as well without direct pressure labels, while exposing a gauge ambiguity that remains in pressure itself (Section 5.3).
- **A component-level ablation of the training schedule.** We decompose the curriculum, residual scaling, and gradient clipping and measure each one's individual contribution to the overall accuracy gain, establishing which component of the training methodology is doing the most work (Section 5.4).
- **Robustness and sensitivity analyses.** We evaluate robustness to Gaussian measurement noise and sensitivity to spatial acquisition strategy, identifying both where the method's advantage holds up and where it does not (Sections 5.5 and 5.6).

2 Related Work

Physics-informed neural networks. Raissi et al. (2019) introduced the general PINN framework for embedding PDE residuals as soft constraints in a neural network's training loss, and Raissi et al. (2020) demonstrated its application to recovering pressure and velocity fields in flow visualization problems from a passive scalar alone – Hidden Fluid Mechanics (HFM) – the setting our aneurysm reconstruction task is most directly built on. Our network architecture, non-dimensionalization strategy, and multi-output PDE residual formulation follow this line of work closely; our contribution is orthogonal, addressing *how* such a network is trained rather than *what* it represents.

Optimization pathologies in PINNs. A growing body of work documents that PINNs trained with a fixed physics weight are prone to poor convergence. Krishnapriyan et al. (2021) show that even simple PDEs can produce loss landscapes that are effectively impossible for gradient descent to navigate once the physics term dominates too early, and Wang et al. (2021) analyze gradient-scale imbalance between the data and physics loss terms through a neural tangent kernel lens, proposing adaptive re-weighting. McClenny & Braga-Neto (2023) instead learn per-point attention weights via a soft-attention mechanism. Our approach differs by treating the imbalance as primarily a *temporal* scheduling problem – the physics term should not merely be re-weighted, but withheld until it can help rather than hurt – while using a lightweight residual-magnitude EMA, rather than a learned weighting network, to balance the five governing equations against one another once physics is

introduced. This connects to curriculum learning more broadly (Bengio et al., 2009), applied here to a loss term's schedule rather than to example difficulty.

Classical and data-driven flow reconstruction. Proper orthogonal decomposition is a standard baseline, which we compare alongside interpolation for sparse-data reconstruction.

3 METHOD

3.1 PROBLEM FORMULATION AND NETWORK

We seek a function $f_\theta(t, x, y, z) = (c, u, v, w, p)$, parameterized by a neural network with weight-normalized (Salimans & Kingma, 2016) fully connected layers ($4 \to 128^{\times 6} \to 5$) and smooth nonlinearities, mapping space-time coordinates (t, x, y, z) to the passive-scalar concentration c, the three Cartesian velocity components (u, v, w), and the pressure p. The network is trained against two loss terms. The *data loss* is the mean squared error between the network's output and sparse point observations of all five fields. The *physics loss* penalizes the residuals of the non-dimensional incompressible Navier–Stokes equations and a scalar advection–diffusion equation for c, evaluated at unlabeled collocation points drawn exclusively from the same spatio-temporal region as the training measurements. Writing subscripts to denote partial derivatives (e.g., $c_t = \partial c/\partial t$ and $c_{xx} = \partial^2 c/\partial x^2$), the five residuals are:

$$e_1 = c_t + uc_x + vc_y + wc_z - \tfrac{1}{\mathrm{Pe}}(c_{xx} + c_{yy} + c_{zz}), \tag{1}$$

$$e_2 = u_t + uu_x + vu_y + wu_z + p_x - \tfrac{1}{\mathrm{Re}}(u_{xx} + u_{yy} + u_{zz}), \tag{2}$$

$$e_3 = v_t + uv_x + vv_y + wv_z + p_y - \tfrac{1}{\mathrm{Re}}(v_{xx} + v_{yy} + v_{zz}), \tag{3}$$

$$e_4 = w_t + uw_x + vw_y + ww_z + p_z - \tfrac{1}{\mathrm{Re}}(w_{xx} + w_{yy} + w_{zz}), \tag{4}$$

$$e_5 = u_x + v_y + w_z, \tag{5}$$

where e_1 is the scalar advection–diffusion residual, e_2–e_4 are the three momentum residuals, and e_5 is the continuity (mass-conservation) residual. $\mathrm{Re} = U_{\mathrm{ref}} L_{\mathrm{ref}}/\nu$ and $\mathrm{Pe} = U_{\mathrm{ref}} L_{\mathrm{ref}}/\kappa$ are the Reynolds and Péclet numbers, where ν is the fluid's kinematic viscosity and κ is the scalar's diffusivity, both recomputed from each dataset's non-dimensionalization; we take unit Schmidt number, i.e. $\kappa = \nu$ and hence $\mathrm{Pe} = \mathrm{Re}$, so that species diffusivity equals kinematic viscosity (a simplifying assumption we revisit in Section 6). The reference length L_{ref} is the vessel's half-extent and the reference velocity U_{ref} is the 95th-percentile velocity magnitude in the dataset; non-dimensionalizing by these two scales keeps every field $O(1)$ and makes the physics-weight schedule of Section 3.3 transferable across geometries without per-case retuning. Pressure is non-dimensionalized by the dynamic pressure scale ρU_{ref}^2, where ρ is the fluid density, and is always evaluated mean-subtracted at test time, since the incompressible momentum equations determine pressure only up to an additive constant.

3.2 THE PREMATURE-PHYSICS FAILURE MODE

Summing the data and physics losses with a fixed weight from initialization couples two competing objectives before either is well posed for the network's current parameters: at initialization the network output is close to a constant, which trivially satisfies the convective and diffusive terms in Equations 1–4 (their gradients with respect to t, x, y, z are all near zero), so the physics loss starts small and its gradient points toward *staying* near this trivial solution rather than away from it. A large physics weight at this stage actively resists the data loss from pulling the network away from the trivial minimum, producing slow or stalled convergence – consistent with the loss-landscape pathologies reported in Krishnapriyan et al. (2021). Our curriculum is designed specifically to avoid entering this regime.

3.3 THREE-PHASE PHYSICS-WEIGHT CURRICULUM

Training proceeds through three phases, allocated as fixed fractions of a wall-clock compute budget (50%, 25%, 25% in our experiments) rather than fixed iteration counts, so that the schedule is comparable across conditions with different per-step costs:

- **Phase 1 – data fitting** ($\lambda_{\mathrm{phys}} = 10^{-4}$): the network is trained almost exclusively on the sparse data loss, allowing it to reach a reasonable approximation of the field before any physics constraint can interfere.

- **Phase 2 – gentle introduction** ($\lambda_{\text{phys}} = 5 \times 10^{-3}$): the physics residual is introduced at fifty times its Phase 1 weight, nudging the network toward physical consistency without overwhelming the already-competent data fit.
- **Phase 3 – full enforcement** ($\lambda_{\text{phys}} = 5 \times 10^{-2}$): the physics weight reaches its final value, now acting as a genuine regularizer on top of a data-fit solution rather than a competing objective against an uninformed one.

Each phase transition also triggers a fresh residual-scale recalibration (Section 3.4). The learning rate follows a phase-aware schedule: constant at $\eta_1 = 10^{-3}$ during Phase 1, and cosine annealing with warm restarts ($T_0 = 2000$ iterations, $\eta_{\min} = 10^{-5}$, (Loshchilov & Hutter, 2016)) in Phases 2–3, optimized with Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$, (Kingma & Ba, 2014)).

3.4 Exponential-Moving-Average Residual Scaling

The five governing-equation residuals in Equations 1–5 have different natural magnitudes and sensitivities, and a single scalar physics weight applied uniformly across them implicitly lets the loss landscape's least well-conditioned equation dominate the physics gradient. We instead normalize each residual by its own root-mean-square magnitude, estimated on a held-out minibatch and updated by an exponential moving average (decay $\alpha = 0.3$) every 200 iterations after a 500-iteration burn-in:

$$s_i \leftarrow \alpha \cdot \mathrm{RMS}(e_i) + (1 - \alpha) \cdot s_i, \qquad \mathcal{L}_{\text{phys}} = \sum_{i=1}^{5} \frac{\mathrm{MSE}(e_i, 0)}{s_i^2 + \epsilon}. \tag{6}$$

We set $\epsilon = 10^{-8}$ to ensure numerical stability and avoid division by zero. The burn-in avoids recalibrating against the noisy, uninformative residual scales present at initialization, and the EMA (rather than an instantaneous estimate) prevents any single minibatch from causing a sudden jump in relative equation weighting. Gradients are additionally clipped to a global norm of 5.0 throughout training, and the checkpoint with the lowest data loss observed during training – rather than the final-iteration weights – is restored before evaluation, which prevents Phase 3's stronger physics enforcement from being allowed to trade away data fidelity it cannot recover from at prediction time.

4 Experimental Setup

Geometries and data. We evaluate on two flow domains. The first is a real, patient-specific three-dimensional intracranial aneurysm geometry, reconstructed from CT imaging and populated with velocity, pressure, and passive-scalar concentration fields from a CFD simulation, providing a clinically realistic and geometrically complex test bed. The second is a synthetic helical pipe flow with a known closed-form solution, used as a sanity check that isolates the reconstruction problem from geometric and turbulence-model complexity.

Train/test protocol. To avoid data leakage between training and evaluation, we split each dataset at the snapshot (timestep) level: 70% of timesteps form a training pool and the remaining 30% form a held-out test pool, with zero temporal overlap. The same fixed test pool – capped at 2,000 spatial points per held-out snapshot – is used to evaluate every method and every measurement density, ensuring a fair, apples-to-apples comparison. Measurement density is controlled by spatially subsampling the training pool (from 1% to 100% of its points). Physics collocation points are sampled uniformly from the training-pool spatiotemporal domain (approximately 20,000 per training run), ensuring the governing equations are never enforced at test-time locations. All models use batch size 2,000, train for wall-clock time budgets with the phase fractions specified above, and are reported as means over 3 random seeds.

Baselines. We compare against three baselines that observe the identical sparse measurements but no governing-equation information: (i) linear n-dimensional interpolation, with unfillable extrapolated points replaced by the field's training mean; (ii) POD, implemented as a radial-basis-function spatial embedding followed by singular value decomposition and per-field ridge regression on the resulting modes; and (iii) PureNN, a physics-free feedforward network trained on the data loss alone. We additionally include DMD, approximated via random Fourier features with a median-heuristic bandwidth followed by per-field ridge regression, as a supplementary baseline in our noise-

robustness analysis (Section 5.5). We report PureNN in two configurations: a shallower 3-hidden-layer, 128-unit network used as the general-purpose data-driven baseline in the density sweep, and a 6-hidden-layer, 128-unit network – exactly matching the PINN's depth and width – used in the architecture-matched ablation of Section 5.2.

Metrics. We report the relative ℓ_2 error, $\|\hat{y} - y\|_2 / \|y - \bar{y}\|_2$, per field and averaged across the five fields (c, u, v, w, p) as the headline number; pressure is always compared mean-subtracted for gauge invariance. Every reported condition is averaged over multiple random seeds (three, unless noted otherwise), with shaded bands and error bars showing ± 1 standard deviation. Throughout our figures we additionally mark a **10% clinical target**: a relative ℓ_2 error of 0.10, adopted as an illustrative threshold below which we consider a reconstructed hemodynamic field accurate enough to plausibly support downstream clinical use (e.g., wall shear stress estimation or flow-pattern classification). This value is broadly consistent with the level of measurement variability reported for velocity quantification in clinical 4D flow MRI (Markl et al., 2012), rather than a threshold derived from a formal clinical-outcomes study; we use it purely as a fixed, consistent reference line for comparing methods across figures, not as a certified clinical-safety criterion.

5 Results

5.1 Sparse-Measurement Reconstruction Accuracy

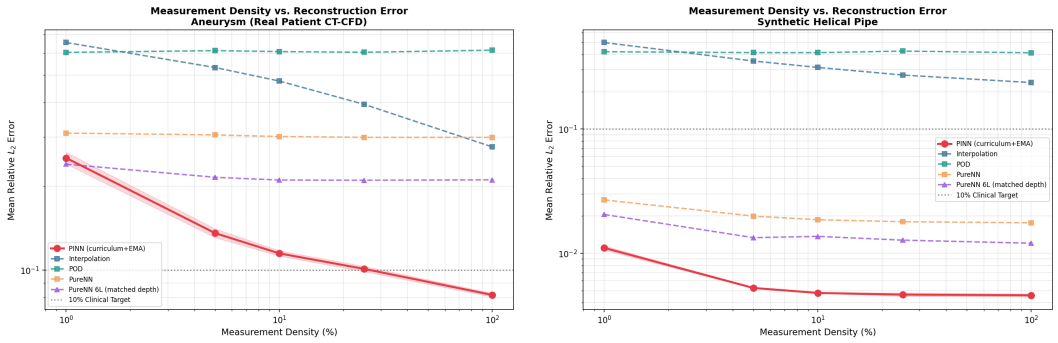
Table 1 reports mean relative ℓ_2 error against measurement density on the patient-specific aneurysm. Our curriculum-trained PINN dominates every baseline at every density, and the gap widens rather than narrows as measurements become sparser: at 1% density the PINN already reaches 0.252 mean error while POD and interpolation remain above 0.6 and 0.65 respectively, essentially unable to extrapolate a coherent field from that few points. POD in particular is close to flat across the entire density range (0.60–0.61), which is consistent with a fixed-rank modal basis whose accuracy is limited by how well its modes span the true flow rather than by how many points are available to fit coefficients onto them. The depth-and-width-matched PureNN-6L baseline is the strongest pure data-driven competitor. At extreme sparsity (1%), both methods perform comparably, with PureNN-6L marginally better; however, the PINN's advantage grows substantially with measurement density, ranging from $1.6\times$ to $2.6\times$ improvement from 5% to 100% density. This increasing advantage confirms that the gain is not simply an artifact of network capacity, particularly in regimes where physics can meaningfully constrain the underdetermined problem (Section 5.2). Figure 1 visualizes the same trend per field and on a log-log scale, alongside the analogous sweep on the synthetic helical pipe (Table 2), where the availability of a clean analytical ground truth lets the PINN reach near-percent mean error already at 1% measurement density, dropping to sub-percent from 5% onward, and improve only marginally beyond that point, while interpolation and POD remain one to two orders of magnitude worse across the entire range, and both PureNN variants remain a consistent 2–4$\times$ worse.

Table 1: Mean relative ℓ_2 error vs. measurement density, patient-specific aneurysm (mean $\pm$ std over 3 seeds).

Method	1%	5%	10%	25%	100%
PINN (curriculum + EMA)	0.252 ± 0.006	0.136 ± 0.003	0.115 ± 0.004	0.101 ± 0.002	0.082 ± 0.002
Interpolation	0.655 ± 0.000	0.533 ± 0.000	0.476 ± 0.000	0.393 ± 0.000	0.277 ± 0.000
POD	0.603 ± 0.001	0.612 ± 0.001	0.607 ± 0.001	0.604 ± 0.001	0.614 ± 0.001
PureNN (3-layer)	0.310 ± 0.004	0.305 ± 0.003	0.301 ± 0.003	0.299 ± 0.003	0.299 ± 0.002
PureNN-6L (matched depth)	0.240 ± 0.003	0.215 ± 0.003	0.210 ± 0.003	0.210 ± 0.002	0.211 ± 0.002

Table 2: Mean relative ℓ_2 error vs. measurement density, synthetic helical pipe (mean $\pm$ std over 3 seeds).

Method	1%	5%	10%	25%	100%
PINN (curriculum + EMA)	0.0110 ± 0.0004	0.0053 ± 0.0002	0.0048 ± 0.0001	0.0046 ± 0.0001	0.0046 ± 0.0001
Interpolation	0.498 ± 0.000	0.353 ± 0.000	0.314 ± 0.000	0.272 ± 0.000	0.237 ± 0.000
POD	0.420 ± 0.002	0.413 ± 0.002	0.412 ± 0.002	0.425 ± 0.002	0.411 ± 0.002
PureNN (3-layer)	0.0269 ± 0.0008	0.0199 ± 0.0006	0.0186 ± 0.0005	0.0179 ± 0.0005	0.0176 ± 0.0004
PureNN-6L (matched depth)	0.0205 ± 0.0006	0.0133 ± 0.0004	0.0137 ± 0.0004	0.0127 ± 0.0003	0.0120 ± 0.0003



(a) Patient-specific aneurysm (real CT-CFD). (b) Synthetic helical pipe.

Figure 1: Mean relative ℓ_2 error against measurement density for the curriculum-trained PINN and all three baselines, shown on a log-log scale with ± 1 std shaded bands. The 10% clinical-target line is shown for reference in both panels.

5.2 Isolating the Contribution of Physics: Architecture-Matched Ablation

Because the PureNN-6L baseline already matches the PINN's depth and width exactly, the remaining gap between the two isolates what the embedded governing equations contribute independent of network capacity. Figure 2 breaks this comparison down per field at 10% measurement density: the PINN improves on its depth-matched, physics-free twin by 1.72–$1.88\times$ on every one of the five fields, with a mean improvement of $1.83\times$ (0.115 vs. 0.210). Notably, the smallest relative gain from physics is on pressure ($1.72\times$) rather than the largest, even though pressure is the field most directly introduced into the network only through the momentum equations' pressure-gradient term; we attribute this to pressure's strong statistical correlation with the velocity field already present in the training data itself, which a sufficiently expressive data-only network can partially exploit even without an explicit dynamical constraint. The velocity components, by contrast, benefit more from the explicit coupling, since their higher-order spatial derivatives are harder to infer from sparse point labels alone.

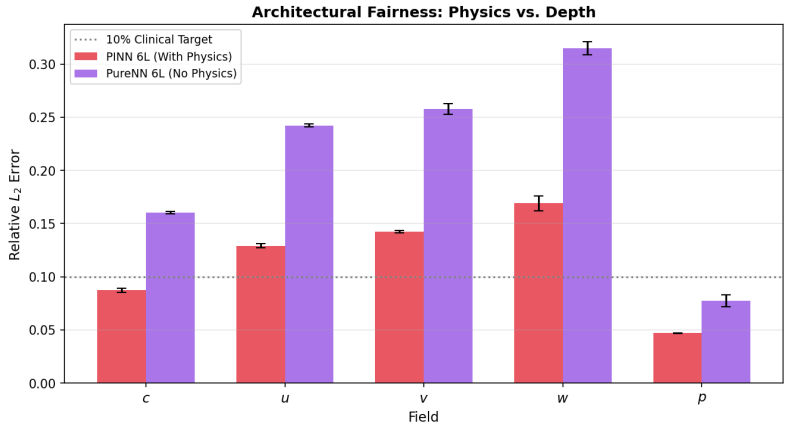


Figure 2: Architecture-matched ablation at 10% measurement density: PINN (6-layer, with physics) vs. PureNN-6L (6-layer, no physics), per field. Both networks have identical depth and width; the only difference is the physics-residual loss term.

5.3 Pressure Recovery via Physics Coupling

The momentum equations relate pressure to velocity only through its spatial *gradient*, which raises the question of how much pressure-labeled data is actually necessary once the physics constraint is present. We retrain the 10%-density PINN with the pressure term removed entirely from the data

loss, so that p is constrained only through Equations 2–4, and compare against the fully supervised model in Figure 3. Concentration and all three velocity components are recovered marginally *better* without direct pressure supervision (c: 0.077 vs. 0.087; u: 0.119 vs. 0.129; v: 0.131 vs. 0.142; w: 0.162 vs. 0.169), consistent with the optimizer reallocating capacity previously spent fitting pressure labels toward the remaining four fields. Pressure itself, however, degrades catastrophically, from a relative ℓ_2 error of 0.047 under full supervision to 1.058 ± 0.036 with no pressure labels at all – an error exceeding 100%, indicating the recovered pressure field is essentially unanchored in both scale and offset. This is the expected consequence of the incompressible momentum equations constraining only ∇p: without at least some absolute pressure supervision, the network can satisfy the physics residual by learning any pressure field that differs from the true one by an additive constant and, in practice given optimization dynamics, a poorly controlled multiplicative scale as well. We discuss the implication of this gauge ambiguity for fully unsupervised deployment in Section 6.

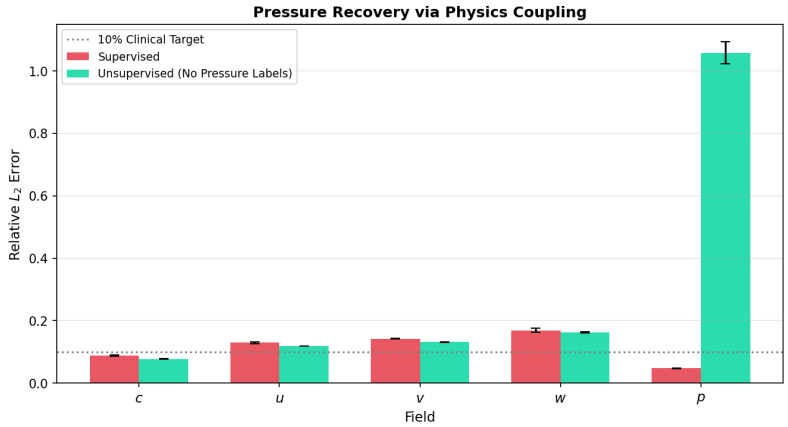


Figure 3: Pressure recovery with ("Supervised") and without ("Unsupervised") direct pressure labels in the training loss, at 10% measurement density. Velocity and concentration degrade only mildly without pressure supervision; pressure itself is essentially unrecoverable in absolute scale from the momentum equation's gradient constraint alone.

5.4 What Each Component of the Training Schedule Contributes

Having established that the full training methodology substantially outperforms both classical and physics-free baselines, we next decompose that advantage into the contribution of each of its three ingredients – the physics-weight curriculum, EMA residual scaling, and gradient clipping – at 10% measurement density on the patient-specific aneurysm. Table 3 compares the full method against four partial configurations, each missing exactly one or two components.

Table 3: Contribution of each training-schedule component, mean relative ℓ_2 error at 10% measurement density (mean $\pm$ std over 3 seeds). The final column shows how much better the full method performs relative to each partial configuration. All runs: batch size 2,000; Adam ($\beta_1 = 0.9, \beta_2 = 0.999$); gradient clipping norm 5.0; cosine annealing $T_0 = 2000$, $\eta_{\min} = 10^{-5}$ in Phases 2–3.

Configuration	Mean ℓ_2 error	Full method's advantage
Full method (curriculum + EMA + grad. clip)	$\mathbf{0.115 \pm 0.004}$	–
Fixed physics weight from init. (no curriculum)	0.4012 ± 0.0050	3.49× better
Static, uncalibrated equation weights (no EMA)	0.1620 ± 0.0013	1.41× better
No gradient clipping	0.1439 ± 0.0068	1.25× better
Fixed weight, no EMA (neither component)	0.3105 ± 0.0035	2.70× better

The physics-weight curriculum delivers by far the largest share of the full method's advantage: introducing it, holding EMA scaling and gradient clipping fixed, improves mean error by 3.49× (0.4012

vs. 0.115). This is precisely the mechanism that lets training escape the premature-physics failure mode described in Section 3.2 – with the physics weight fixed at its final value from initialization, the optimizer is held near the trivial near-constant solution instead of being allowed to first learn a competent data fit. On top of a well-scheduled curriculum, EMA residual scaling contributes a further, smaller refinement, improving mean error by $1.41\times$ ($0.162 \rightarrow 0.115$) by adaptively balancing the five governing-equation residuals rather than letting whichever equation happens to have the largest natural magnitude dominate the physics gradient. Gradient clipping contributes a consistent but marginal improvement on its own ($1.25\times$), acting more as a safety margin against occasional large updates during Phase 3's full physics enforcement than as a routinely load-bearing mechanism in isolation.

The curriculum and EMA scaling exhibit a genuine interaction effect rather than independent, additive benefits. Conditional on the curriculum being present, EMA scaling helps: removing it while keeping the curriculum degrades mean error from 0.115 to 0.162, a $1.41\times$ cost (Table 3, row 3 vs. row 1). Conditional on the curriculum being *absent*, however, EMA scaling hurts: removing it from the no-curriculum configuration *improves* mean error, from 0.4012 to 0.3105 (Table 3, row 2 vs. row 4). We attribute this reversal to what EMA scaling actually does – adaptively reweighting the five governing-equation residuals by their running RMS magnitude. When the curriculum has already produced a competent data fit before physics is introduced, this reweighting sharpens a good starting point by preventing any one equation from dominating the physics gradient. When physics is enforced from initialization instead, the residuals being reweighted are computed from an undertrained, near-trivial field in the first place; adaptively amplifying whichever of these residuals is largest does not correct course, it reinforces whatever spurious imbalance already reflects the network's undertrained state, deepening rather than escaping the premature-physics failure mode of Section 3.2. The practical implication is that EMA residual scaling is not a general-purpose stabilizer that can be added to any physics-weighting scheme – it is specifically a refinement of a well-scheduled one, and its benefit is contingent on the curriculum having already done its job.

5.5 Robustness to Measurement Noise

We inject zero-mean Gaussian noise, scaled to each field's per-training-batch standard deviation, onto the (fixed at 10%-density) training measurements only – the held-out test set remains exactly clean at every noise level, isolating the model's robustness to corrupted supervision from its baseline reconstruction accuracy. Figure 4 shows that the PINN's error is essentially flat from 0% to 10% injected noise across every field (approximately $0.17 \rightarrow 0.19$ on u, $0.19 \rightarrow 0.20$ on v, $0.22 \rightarrow 0.24$ on w, $0.18 \rightarrow 0.20$ on c, and $0.06 \rightarrow 0.07$ on p), a much smaller relative degradation than would be implied by the noise magnitude alone, consistent with the physics residual acting as a denoising regularizer that constrains the solution manifold independently of how corrupted the sparse labels are. Interpolation, POD, and Dynamic Mode Decomposition (DMD) are all comparably insensitive to noise in absolute terms, but only because they are already saturated at a much higher error floor two to four times the PINN's; PureNN, the physics-free baseline, sits consistently above the PINN on every field except pressure, where it is occasionally comparable to or better than the PINN under noise, echoing the pressure-specific caveat identified in Section 5.3.

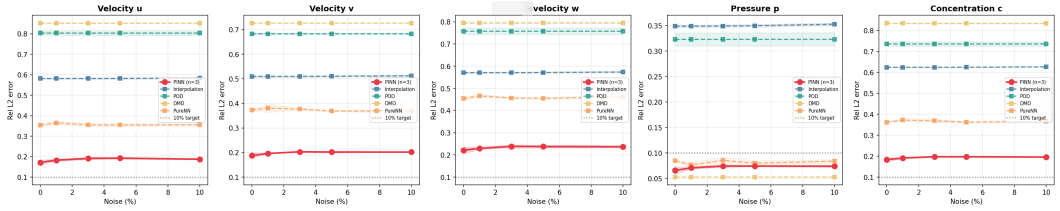


Figure 4: Mean relative ℓ_2 error against injected Gaussian training-label noise (noise applied to training data only; the held-out test set is always clean), at 10% measurement density.

5.6 Sensitivity to Spatial Acquisition Strategy

A practical deployment must choose *where*, not just how many, sparse sensors to place. We compare three spatial sampling strategies at fixed 10% measurement density: uniform random sampling,

inlet-focused sampling (concentrated near the domain's inflow boundary), and wall-adjacent sampling (concentrated near the vessel wall). Figure 5 shows that wall-adjacent sampling gives the lowest and most consistent error on four of the five fields (concentration and all three velocity components), with inlet-focused sampling consistently the worst strategy across every field, including the largest single gap on the velocity component w. Random sampling is competitive with wall-adjacent sampling throughout and is in fact marginally the best strategy for pressure recovery, though the two are within each other's error bars on that field. None of the three strategies reach the 10% clinical-target line at this density, which we discuss as an open direction rather than understate in Section 6. The practical implication is that where sparse sensors are placed relative to the vessel wall geometry matters as much as, and is considerably cheaper to control than, how many are placed.

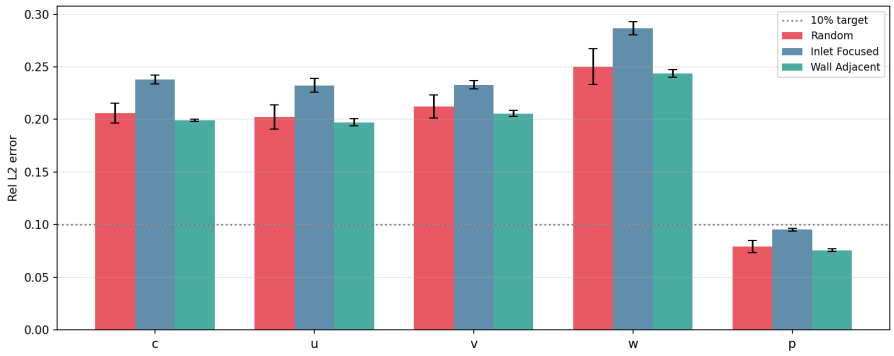


Figure 5: Reconstruction error under three spatial sampling strategies at 10% density.

6 Discussion

Beyond the accuracy gains reported above, two aspects of this study sharpen exactly what curriculum-scheduled physics-informed training solves, and point directly to the most promising next steps for extending it.

Pressure recovery and gauge anchoring. Because the momentum equations constrain only the pressure *gradient* rather than pressure itself, our label-free pressure (Section 5.3) isolates an inherent gauge degree of freedom in the physics, not a weakness of the training method: velocity and concentration are recovered almost as well without any pressure supervision at all, which is itself a strong result. The pressure field's absolute scale in principle could be restored with a small number of anchor points rather than requiring the dense labels our main experiments use. However, the optimal placement and number of anchor points remains an open question that requires future validation on experimental data. This direction warrants further investigation as a potential avenue for reducing measurement burden in clinical settings.

Closing the gap to the clinical target at extreme sparsity. Our method already narrows the distance to the 10% clinical target by several-fold relative to every baseline at every density we test, including under all three acquisition strategies (Section 5.6); the remaining gap at the sparsest densities is a natural target for follow-up work that couples acquisition-strategy optimization directly with the training schedule introduced here, rather than treating sensor placement and training method as independent design choices. All aneurysm results also come from a single patient-specific geometry, and extending validation across multiple geometries – potentially via transfer or meta-learning – is left to future work.

7 Conclusion

The difficulty of physics-informed neural networks in sparse-data regimes stems largely from treating the physics-weight schedule as a fixed hyperparameter. A three-phase physics-weight schedule combined with exponential-moving-average residual scaling recovers complete hemodynamic fields from as little as 1% of a full measurement set with several-fold lower error than interpolation, POD, and an architecturally identical physics-free network, and an architecture-matched ablation confirms that this advantage is attributable to the embedded governing equations themselves rather than to network capacity.

ETHICS STATEMENT

This work uses de-identified, patient-specific CT-derived geometry data for CFD simulation; no human-subjects experimentation was performed as part of this study

REPRODUCIBILITY STATEMENT

Full architectural details, training hyperparameters, and the three-phase curriculum schedule are provided in Sections 3 and 4. All reported results are averaged over three random seeds with standard deviations included; we intend to release training and evaluation code upon acceptance.

AI USE STATEMENT

Large language models were used to assist with editing prose for clarity and polishing LaTeX formatting during manuscript preparation. No AI tool was used to generate research ideas, design experiments, produce results, or write the core scientific content of this paper.

REFERENCES

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th International Conference on Machine Learning*, pp. 41–48, 2009.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

Michael Markl, Alex Frydrychowicz, Sebastian Kozerke, Michael Hope, and Oliver Wieben. 4d flow mri. *Journal of Magnetic Resonance Imaging*, 36(5):1015–1036, 2012.

Levi D. McClenny and Ulisses M. Braga-Neto. Self-adaptive physics-informed neural networks. *Journal of Computational Physics*, 474, 2023.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.

Tim Salimans and Diederik P. Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Charles A. Taylor and C. Alberto Figueroa. Patient-specific modeling of cardiovascular mechanics. *Annual Review of Biomedical Engineering*, 11:109–134, 2009.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021.