PINNeAPPle

TECHNICAL REPORT · PUBLIC BENCHMARK

# FluidsBench / AirfRANS nine results, one measured idea

The evaluator never reads a submitted wall shear: it re-integrates it from the submitted field. Nine entries built on that fact, from 29.4 to 75.4, with the component that is still unsolved.

| 75.442 | 36.95% | 2,760,708 |
|---|---|---|
| overall score, full split, of 100 | wall shear rel L2 — above the 20% cap | trainable parameters, one model |

Document BENCH-01 · version 1.0 · 2026
Dataset: AirfRANS (airfrans-native-v1), splits full · scarce · aoa_extrapolation
Method: point-wise coordinate network, exact no-slip wall layer, shape descriptor
PINNeAPPle · LAIS & YAN LTDA · pinneapple-labs.github.io

AirfRANS is the airfoil dataset of the public FluidsBench benchmark: RANS solutions around NACA 4- and 5-digit profiles, 1,000 cases, three official splits. The benchmark scores a submitted field with a composite of four caps and two force correlations, and the whole interest of the exercise is one property of how it does that.

### What the evaluator actually measures

It never reads a submitted wall shear or a submitted force coefficient. It takes the submitted velocity and pressure at the native internal points of each case, reconstructs the tangential gradient at the surface nodes, and integrates that gradient into a wall shear and into the drag and lift coefficients through its own polyline quadrature over the airfoil curve. Both the near-wall layer and the field around it therefore enter the score, and neither alone is measured by it: what is measured is the agreement between them.

| COMPOSITE COMPONENT | CAP | WHAT HAPPENS AT THE CAP |
|---|---|---|
| Surface pressure, relative L2 | 15% | the component scores zero |
| Surface wall shear, relative L2 | 20% | the component scores zero |
| Flow-domain velocity, relative L2 | 12% | the component scores zero |
| Flow-domain pressure, relative L2 | 15% | the component scores zero |
| Cd $R^2$, Cl $R^2$ | — | correlation, no cap |
| Velocity profile $R^2$, diagnostics | — | weight 0.25 in the composite |

### The nine entries

Six on the full split (200 test cases), two on aoa_extrapolation (196), one on scarce. They are one architecture and one set of ideas measured in sequence, not nine unrelated attempts: each entry changes one thing and reports what the change was worth.

> **Why six entries on one split.** A benchmark entry is a fixed artefact, so the only way to publish a measured sequence is one submission per step. The five full-split steps are 63.200 → 65.259 → 70.070 → 72.503 → 75.442, and every step names the one change that produced it.

The previous architecture of this campaign — a coordinate network predicting the perturbation of the field — had plateaued at 29.438 on the full split, 29.697 on aoa_extrapolation and 24.591 on scarce: three splits, one number.

### What was measured on the best previous entry

| QUANTITY | MEASURED | WHY IT MATTERS |
|---|---|---|
| Wall velocity, where it must be exactly zero | 3.7% of U∞ | the no-slip condition is not held |
| First cell distance from the surface | y+ ≈ 0.2 | the gradient is read inside the sublayer |
| Viscous drag against the reference value | 5.6× | the integrated wall shear is off by a factor |
| Pressure-drag error ahead of 10% chord | 73% | the leading edge carries the force |
| Flow-domain velocity, relative L2 | 12.15% | against a 12% cap — the field sits at the cap |
| Cd $R^2$ | −104.13 | the drag correlation is not usable at all |

A point-wise network has no reason to place the wall velocity at zero, and no reason to make the derivative it produces at the surface agree with the field around it. Both are properties the physics asks for and the loss did not: the training target was the field, while the score is the derivative at the boundary and its integral along the chord.

> **The ceiling is representational, and it was measured.** Blending the *true* field into a linear wall layer with $\partial u_t / \partial d = +\tau_w/\nu$ makes the evaluator reconstruct the wall shear to 0.85% and 1.68% on two cases, at every $\lambda$ between $2{\cdot}10^{-5}$ and $3{\cdot}10^{-3}$ chords. The architecture can represent the wall shear almost exactly; what was missing was a parametrisation that makes it the thing being predicted.

### What this report claims

- That the score of this benchmark is dominated by the coherence between the near-wall layer and the field around it, and that the coherence can be built in rather than hoped for.
- That each of the four changes in section 3 carries a measured part of the 63.2 → 75.4 gain, and that the two that do not carry it are reported as not carrying it.
- That the wall shear is **not** solved: it is 36.95% on the full split against a 20% cap, and the entry that scores 75.4 still takes zero on that component.

Four changes, each one measured against the same 20-case held-out set before it entered a submission.

### 1. An exact no-slip wall layer

The velocity is written as $U = A(d)\cdot U_{outer} + (1 - A(d))\cdot s\cdot d$ with $A(d) = 1 - \exp(-(d/\lambda)^2)$ and $\lambda = 5\cdot10^{-5}$ chords, the viscous sublayer. No-slip holds by construction, and the first derivative of that expression at the wall *is* the slope head s — the quantity the evaluator reconstructs. The wall shear stops being an accident of the field and becomes an output of the model.

### 2. A per-case shape descriptor

A point-wise network cannot tell two airfoils apart at the same location: the same coordinate on the upper surface of a NACA 0012 and of a NACA 4412 is two different physical points. The network also reads the case's upper and lower surfaces on 32 chord bins, 64 channels. A nearest-neighbour probe on the wall slope puts the error at 75.7% without the descriptor and 42.1% with it, which is what makes it worth the channels.

### 3. Force consistency in the loss

Drag and lift are reconstructed from the predicted curve pressure and wall slope through the evaluator's own polyline quadrature, reproduced here to five decimals against `Simulation.force_coefficient` and asserted in tests. The squared error then enters the objective, so the model is trained on the same integral it is scored on. The ablation in section 4 says this term does not move the wall shear on its own; it is in the loss because it prices the quantity the force score reads.

### 4. The readout held at the flat-plate law

Instead of letting the wall readout co-adapt with the trunk, the readout is held at the flat-plate skin-friction law while the field trains, and refitted afterwards on the frozen field. Section 9 gives the measurement that motivated it, and it is the single largest step of the series.

One ablation run per change, on a deliberately small budget: 5 epochs, one seed, the same 20 held-out cases. This is where the four changes are separated from each other.

| CONFIGURATION | VELOCITY | PRESSURE | WALL SHEAR | CD $R^2$ |
|---|---|---|---|---|
| base, 5 epochs | 9.28% | 6.27% | 97.2% | −3.85 |
| without the shape descriptor | 14.69% | 46.20% | 101.1% | −28.87 |
| wall layer too thin ($\lambda = 10^{-7}$) | 11.07% | 3.98% | 948.5% | −4292 |
| without slope supervision | 8.55% | 4.73% | 97.3% | −4.32 |
| without the force term | 7.66% | 4.35% | 97.2% | −2.18 |

### Reading it

Two of the four changes are not what carries the score, and the ablation is explicit about it. The shape descriptor and the wall-layer thickness are the additions that carry it: removing either makes the field collapse — 6.27% of pressure becomes 46.20% without the descriptor, and the wall shear goes to 948.5% when the layer is thinner than the first cell.

Slope supervision and the force term have no measurable effect on the wall shear at this budget (97.3% and 97.2% against 97.2% of the base), and they are reported that way rather than credited. What decides the wall shear turned out to be the stage that produces the readout — section 9 — not the terms added to the field loss.

> **The number that made the next four entries possible.** With the descriptor and the wall layer in place, the held-out embedding of the entry that scored 63.200 reaches 43.5% of wall shear and −5.03 of Cd $R^2$. Everything after it is measured against that pair, not against the base row.

Five steps on the 200-case test set. Each row is a submitted package; the narrative below names the one change that separates it from the row above.

| SUBMISSION | SCORE | PRESSURE | WALL SHEAR | VELOCITY | PRESSURE (DOMAIN) | CD $R^2$ |
|---|---|---|---|---|---|---|
| wall-aware-full-v1 | 63.200 | 2.93% | 78.72% | 4.49% | 3.66% | −1.433 |
| wall-frozen-full-v1 | 65.259 | 2.28% | 58.59% | 4.07% | 2.59% | −0.214 |
| wall-frozen-readout-full-v1 | 70.070 | 2.28% | 43.83% | 3.95% | 2.59% | +0.311 |
| wall-leading-edge-full-v1 | 72.503 | 2.28% | 39.17% | 3.93% | 2.59% | +0.472 |
| wall-chord-weight-full-v1 | 75.442 | 2.28% | 36.95% | 3.94% | 2.59% | +0.668 |

### 63.200 — the wall layer, the descriptor and the force term

The first entry of the series. All four field components leave their cap for the first time in the campaign (field_score 0 → 58.03), the velocity profile $R^2$ goes from 0.782 to 0.968, and the wall shear falls from 758.6% to 78.72%. It is still far from the 20% cap and the drag correlation is negative.

### 65.259 — the field gets the budget, the readout stays fixed

Same recipe, 28 epochs instead of 14, readout held at the flat-plate law. The field recovers past the previous entry on every component, the wall shear drops to 58.59%, and Cd $R^2$ reaches −0.214 — one step from crossing zero.

### 70.070 — only the readout moves

Eight epochs with the trunk frozen and only the readout refitted on that field. Nothing else changes, so the field numbers do not move, and the wall shear falls to 43.83% while Cd $R^2$ goes positive for the first time in the full split (+0.311). force_score 39.98 → 58.67.

The last two steps of the series, and the two variants that were measured and rejected between them.

### 72.503 — the loss tilted towards the leading edge

The force coefficients are integrals of the slope along the chord, and 73% of the pressure-drag error of the previous architecture sat ahead of 10% chord. Weighting the readout loss by $1 + 2(1 - x/c)$ puts capacity where the force is: wall shear 39.17%, Cd $R^2$ +0.472, force_score 68.30.

### 75.442 — the square of the distance to the chord

The same recipe with the weight squared, 16 epochs. This concentrates the readout on the first per cent of the chord, where the slope varies fastest: wall shear 36.95%, Cd $R^2$ +0.668, force_score 80.08, and the field untouched at 3.94% and 2.59%.

### Two variants that lost, and what they close

A cubic weight reached 29.9% of wall shear with Cd +0.14, and the quadratic weight trained for 24 epochs reached 26.8% with +0.14 — both measured on the same held-out set and both worse than the 22.1% and +0.45 of the chosen configuration. Concentrating further, or training the readout longer, undoes what it got right. The readout tuning is closed: what remains is not in it.

> **The accumulated gain, and its shape.** 63.200 → 75.442 is +12.24 points, and 80% of it comes from the wall shear falling from 78.72% to 36.95% and the drag correlation crossing zero. Both came from the readout stage, not from the field, which barely moved from the second entry on (4.49% → 3.93% of velocity, 3.66% → 2.59% of pressure).

The same recipe, refitted from scratch in the split's own training regime. This is the split where the previous entry of the campaign had the worst result of all: 1107% of wall shear.

Fitted on aoa_train (804 cases, a shuffled 764/20/20 partition, so no full_train case enters the fit): the readout held at the law while the field receives 28 epochs, then refitted over the frozen trunk with the quadratic weight on the distance to the chord. No test case is used for fitting, selection, tuning or normalisation.

| COMPONENT | WALL-AWARE-AOA (55.009) | WALL-CHORD-WEIGHT-AOA (67.239) |
|---|---|---|
| Surface pressure rel L2 (cap 15%) | 6.29% | 5.48% |
| Surface wall shear rel L2 (cap 20%) | 120.59% | 48.71% |
| Flow-domain velocity (cap 12%) | 7.82% | 5.65% |
| Flow-domain pressure (cap 15%) | 8.63% | 7.43% |
| Cd $R^2$ | +0.222 | +0.702 |
| Cl $R^2$ | 0.9981 | 0.9994 |
| Velocity profile $R^2$ | 0.941 | 0.969 |
| field_score · force_score · diagnostic_score | 36.36 · 53.24 · 94.07 | 45.02 · 82.07 · 96.85 |

The +0.702 of drag correlation is the best of this campaign on any split. The entry that leads this split reaches 0.832 and 85.355 overall; this one is second, and the gap is the wall shear (48.71% against 4.94%) — the same gap that separates the entries on the full split.

The scarcest split was the campaign's worst line and the cheapest to close, and it is the one place where the recipe does not transfer.

Refitting the readout with 200 cases **diverges**: on the held-out set the wall shear is 38.15% at the end of the field stage and 127.25% after the refit. The submitted entry therefore stops the recipe one stage earlier and uses the field stage alone — readout held at the law, 28 epochs on 200 cases, no second stage.

| COMPONENT | PP-COORDINATE-NET-SCARCE-V1 (24.591) | WALL-FIELD-SCARCE-V1 (52.533) |
|---|---|---|
| Surface pressure rel L2 (cap 15%) | 20.61% | 5.01% |
| Surface wall shear rel L2 (cap 20%) | 2704.3% | 48.42% |
| Flow-domain velocity (cap 12%) | 22.35% | 8.73% |
| Flow-domain pressure (cap 15%) | 41.58% | 6.29% |
| Cd $R^2$ | −1995.1 | −0.678 |
| Cl $R^2$ | 0.970 | 0.9981 |
| Velocity profile $R^2$ | 0.596 | 0.907 |
| overall_score | 24.591 | 52.533 |

+27.9 points with a 200-case training run and no new test case. What still separates it from the top of that split is exactly what separates the entries on the full split: wall shear 48.42% against 3.63%, and velocity 8.73% against 1.40%.

> **The negative result that belongs in the report.** On the scarce split the refit of the readout is not neutral, it is destructive: the same second stage that is worth +4.81 points on the full split costs the wall shear a factor of three when only 200 cases are available. It is reported as measured rather than left out.

This is the finding the whole series rests on: the wall shear is not a property of the readout, it is a property of the **pair** (field, readout).

| CONFIGURATION (20 HELD-OUT CASES) | VELOCITY | PRESSURE | WALL SHEAR | CD $R^2$ |
|---|---|---|---|---|
| embedded entry (63.200) | 4.31% | 2.41% | 43.5% | −5.03 |
| flat-plate prior, co-trained readout | 4.64% | 2.62% | 38.5% | −1.88 |
| readout frozen in the law, field trained | 5.80% | 3.01% | 30.9% | −0.97 |
| trunk frozen, readout refitted | 5.64% | 3.01% | 29.6% | −1.41 |
| readout refitted over the co-adapted trunk | 4.60% | 2.62% | 75.4% | −1.34 |

Five budgets, the same 20 held-out cases, only the part of the model allowed to move changes. Freezing the readout at the flat-plate law while the field trains gives the best wall — 30.9% against 43.5% for the co-adapted entry. Refitting the readout afterwards over a trunk that co-adapted with it makes it *worse*: 75.4%.

The mechanism is the one the evaluator imposes. A readout that moves together with the trunk produces a wall gradient consistent with the field it was trained against, but the evaluator integrates the gradient of the field *as submitted*, and a gradient that co-adapted to a moving field does not survive that step. Holding the readout at an external law, and refitting it only on a field that has stopped moving, keeps the two consistent by construction.

### Two families of ideas closed by measurement

The wall-shear error is not one of magnitude: giving every case its own least-squares scale takes the mean per-case error from 32.7% to 28.1% on the full split, and a single constant is worse; the error is also uniformly spread, since the ten worst cases carry 8.3% of the sum. Nor does the field dictate it: raising the wall sampling share from 0.25 to 0.50 sends the velocity from 3.13% to 7.96% and Cd $R^2$ from +0.449 to −0.437, while the wall shear moves from 22.05% to 23.38%. What decides the wall shear is the stage of the readout.

The 20-case held-out set chose every configuration in this report, and it was systematically optimistic. It is shown here because a score predicted from it would have been wrong by a factor of two.

| WHERE | HELD-OUT (20 CASES) | OFFICIAL TEST |
|---|---|---|
| wall shear, aoa_extrapolation (196 cases) | 22.47% | 48.71% |
| velocity, scarce (180 cases) | 8.02% | 8.73% |
| wall shear, scarce | 38.15% | 48.42% |

The gap is not noise from a small sample. Twenty cases do not cover the tails of incidence and Reynolds that carry the error. The replacement set takes 61 cases from the final half of the same shuffled pool — so that every training cut up to 400 cases stays disjoint from it — stratified in three inlet-velocity and three angle levels, with a composition that matches the release (34% of cases above 8° against 31%, 33% above 70 m/s against 37%).

### And it reversed a conclusion

On the 20-case set the arm with a panel-method prior looked better on the wall (46.1% against 127.3%). On the 61 cases the plain field wins clearly: 44.9% against 62.1% of wall shear, and 8.57% against 9.93% of velocity. Every comparison in this report was therefore made on the 61.

> **The methodological point.** A held-out set chooses a configuration; it does not predict a score. The campaign states the held-out number and the official number side by side for that reason, and never quotes the first as a result.

Five statements that the submitted packages do not support, written before anyone asks.

- **The wall shear is still above its cap in every entry**: 36.95% on full, 48.71% on aoa_extrapolation, 48.42% on scarce, against a 20% cap. The 0.10-weight wall component scores zero in all nine, and 75.442 is a score with that component at zero.
- **The drag component needs Cd $R^2$ above zero to score at all**: the full split gets there from the third entry on, the AoA split with the last one, the scarce split does not (−0.678).
- **The readout that supplies the wall gradient has a relative error of 370%** on its own training cases, with a correlation of 0.24 against the truth, while a nearest-neighbour model over the same channels reaches 42%. The information is present in the inputs; the open item is the parametrisation of the readout, not the features.
- **The panel-method prior is not in any submitted entry**: the arm was measured (44.9% against 62.1% of wall shear on the 61-case set) and lost, so the lever the leading entry of this split uses is one this campaign measured and did not adopt.
- **No external validation**: every number is the benchmark's own reconstruction on the benchmark's own test cases. There is no experiment, no second solver and no independent implementation behind them.

> **What comes next, and why it is not in this report.** The remaining 25 points of the composite sit in two components: the wall shear (needs ≤ 20%) and the drag correlation. Both require changing the readout itself — a stream-function parametrisation or a per-case boundary-layer model — not the loss, the sampling or the budget. Neither is measured yet, so neither is claimed.

Everything in this document is re-derivable from the nine submitted packages with the benchmark's own tooling and no private data.

| WHAT | WHERE |
|---|---|
| The nine packages | FluidsBench/fluidsbench-submission → submissions/airfrans/pp-coordinate-net-wall-*-v1 |
| Per-case metrics | metrics/cases.json in each package |
| Submitted field | discretization/ and profiles/ in each package |
| Evaluation evidence | evaluation-evidence.json — command, metric values, SHA-256 of every input |
| Method record | submission.json → methodology — architecture, training stages, compute, checkpoints |

### The gate

Each package is validated with the benchmark's own contributor gate, which needs no private data: `python scripts/validate_submission.py --contributor-stage submissions/airfrans/<package>`. It re-reads the submitted field, re-applies the scoring contract and prints the case and profile counts — 200 cases and 1,600 profile series on the full split, 196 and 1,568 on aoa_extrapolation. All nine pass it.

### Compute and licence

One process on one device: an Apple M-series integrated GPU through Metal (MPS). The reported wall-clock is the training script measured end to end — 289.1 s for the 16-epoch readout refit, 1,972.1 s for the 28-epoch field stage. The model is a residual point-wise network, 4 blocks of width 512, 16 Fourier bands, three heads, 2,760,708 trainable parameters — one model, one seed, one run. The submitted material is Apache-2.0; the AirfRANS release keeps its own licence.

The packages as submitted, in the order the series was produced. The epochs are those of the final stage recorded in each package's method block.

| PACKAGE | SPLIT | SCORE | WALL SHEAR | CD $R^2$ | EPOCHS |
|---|---|---|---|---|---|
| wall-prior-full-v1 | full | 62.136 | 76.74% | −0.691 | 14 |
| wall-aware-full-v1 | full | 63.200 | 78.72% | −1.433 | 14 |
| wall-frozen-full-v1 | full | 65.259 | 58.59% | −0.214 | 28 |
| wall-frozen-readout-full-v1 | full | 70.070 | 43.83% | +0.311 | 8 |
| wall-leading-edge-full-v1 | full | 72.503 | 39.17% | +0.472 | 16 |
| wall-chord-weight-full-v1 | full | 75.442 | 36.95% | +0.668 | 16 |
| wall-aware-aoa-v1 | aoa_extrapolation | 55.009 | 120.59% | +0.222 | 14 |
| wall-chord-weight-aoa-v1 | aoa_extrapolation | 67.239 | 48.71% | +0.702 | 16 |
| wall-field-scarce-v1 | scarce | 52.533 | 48.42% | −0.678 | 28 |

All nine are prefixed `pp-coordinate-net-` in the submission directory and share one architecture: residual point-wise network, 4 blocks of width 512, 16 Fourier bands, a per-case shape descriptor and an exact no-slip wall layer.