Diagnosing Ocean Vertical Velocity from Surface Fields with ML

Hanfu Gong, supervised by Nicolas Grisouard

Oceananigans reimplementation of the He & Mahadevan coastal-upwelling experiments; 25-member ensemble; U-Net, plain CNN, and HM24-style CNN training

25	97 × 385 × 33 × 121 = 149,118,585	$R^2 = 0.605$	$R^2 = 0.73$
Oceananigans ensemble members	grid points per ensemble member	our HM24-style surface-only CNN	HM24 surface-only CNN benchmark

1. Motivation and HM24 context

Scientific target. Infer subsurface vertical velocity from remotely observable surface fields:

$$\mathcal{F}_\theta : (u_{\text{surf}}, v_{\text{surf}}, b_{\text{surf}}) \to w(x, y, z).$$

Why this matters. Submesoscale w controls vertical exchange of heat, nutrients, and carbon, but it is difficult to observe directly.

He & Mahadevan (GRL 2024) reference problem.

- Idealized coastal-upwelling simulations in a 500 m deep, 96×384 km re-entrant channel.
- HM24 compared MLR, RF, and CNN approaches for diagnosing w from surface data.
- Their CNN used a 32×32 km surface patch (with unit stride) and predicted a vertical profile $w(z)$ below the patch center.

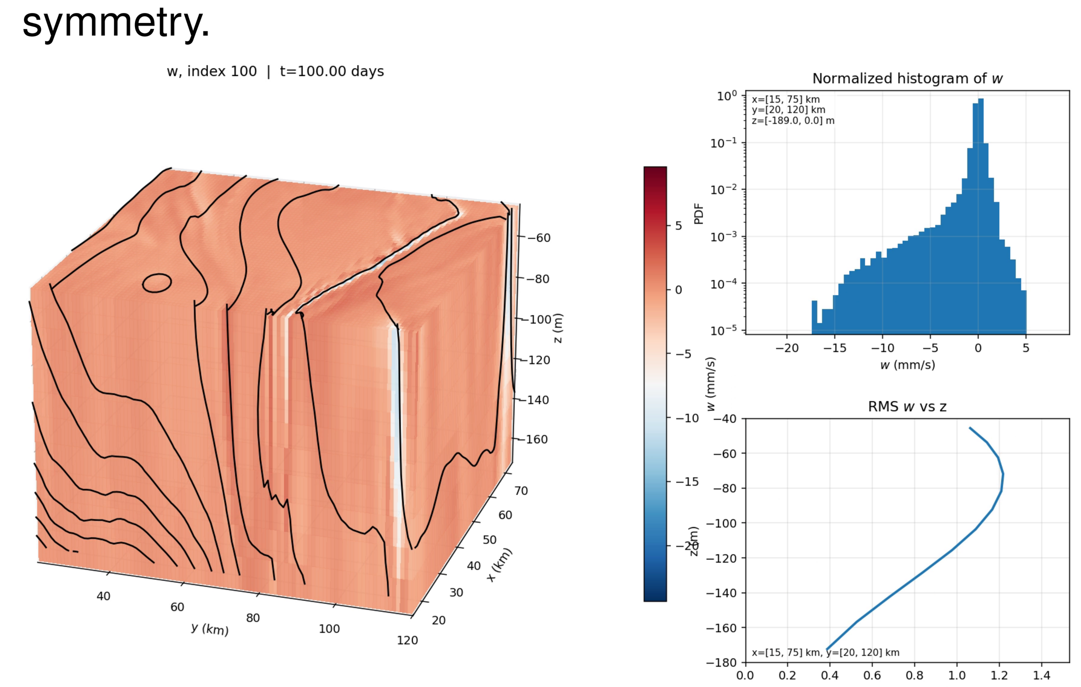
2. Oceananigans simulation design

HM24 implementation

- The model follows from a previous work (He & Mahadevan (JGR Oceans 2021)).
- They use a PSOM governed by 3-D Nonhydrostatic Boussinesq equations.

Reimplementation choices.

- Nonhydrostatic Boussinesq model, WENO-5 advection, RK3 time stepping in Oceananigans.
- f-plane at $15°$N, alongshore wind stress ramped over 10 days.
- Horizontal spacing: 1 km; 32 stretched vertical levels over 500 m with nonuniform heights.
- b-diffusivity uses $K_h = 1$ m^2 s^{-1} plus wind-dependent $K_z(z)$.
- Small deterministic v perturbation breaks exact alongshore symmetry.



HM24-like model w diagnostic diagrams from ensemble member 19.

Expanded parameters. We replace HM24's 3×3 sweep with a 5×5 sweep:

$$\tau = [0.01, 0.0325, 0.055, 0.0775, 0.1]\ \mathrm{N\,m^{-2}},$$

$$N^2 = [1, 3.25, 5.5, 7.75, 10] \times 10^{-5}\ \mathrm{s^{-2}}.$$

3. C-grid to ML grid

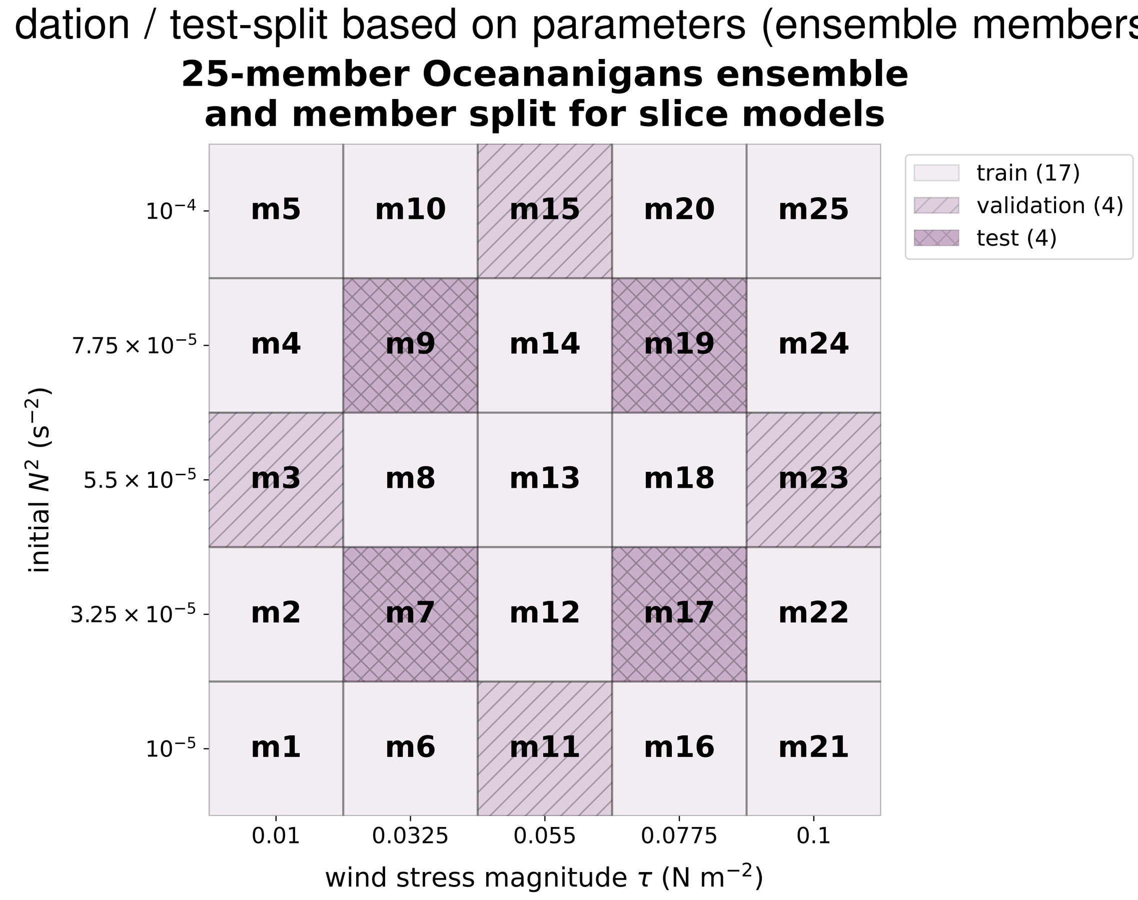
The training scripts collocate predictors and targets on tracer-cell centers:

b	already center/center/center; use the uppermost center level.
u	average adjacent x-faces to x centers with periodic wrapping.
v	average adjacent y-faces to y centers.
w	linearly interpolate z-faces to tracer z centers.

For slice models, the target $z = -200$ m maps to the nearest tracer-center level, $z \approx -197.85$ m.

4. 25-member ensemble split

For our own U-Net and CNN runs, we choose a train / validation / test-split based on parameters (ensemble members).



Each cell is one Oceananigans run. Slice models use a member split: train on 17 members, validate on 4, test on 4.

For the HM24-style CNN, we mimic their train / test-split by choosing a 20-day period from each ensemble member, using 4 days for training and 1 day for testing with adjacent dates separated by 5 days.

5. Training methods

Current archive: neural models on Oceananigans data.

- **U-Net slice model:** predicts a full horizontal w map at one target depth from surface fields.
- **Plain CNN slice model:** same target and inputs, but a simpler residual fully convolutional baseline.
- **HM24-style patch CNN:** predicts 21 depths from 0 to -200 m at the center of a local surface patch.

Training emphasis

- Surface-only inputs for current results: $(u_{\text{surf}}, v_{\text{surf}}, b_{\text{surf}})$.
- Member-split evaluation tests generalization to unseen wind/stratification combinations.
- Staggered C-grid variables are averaged/interpolated before training.

6. Data splits and comparability

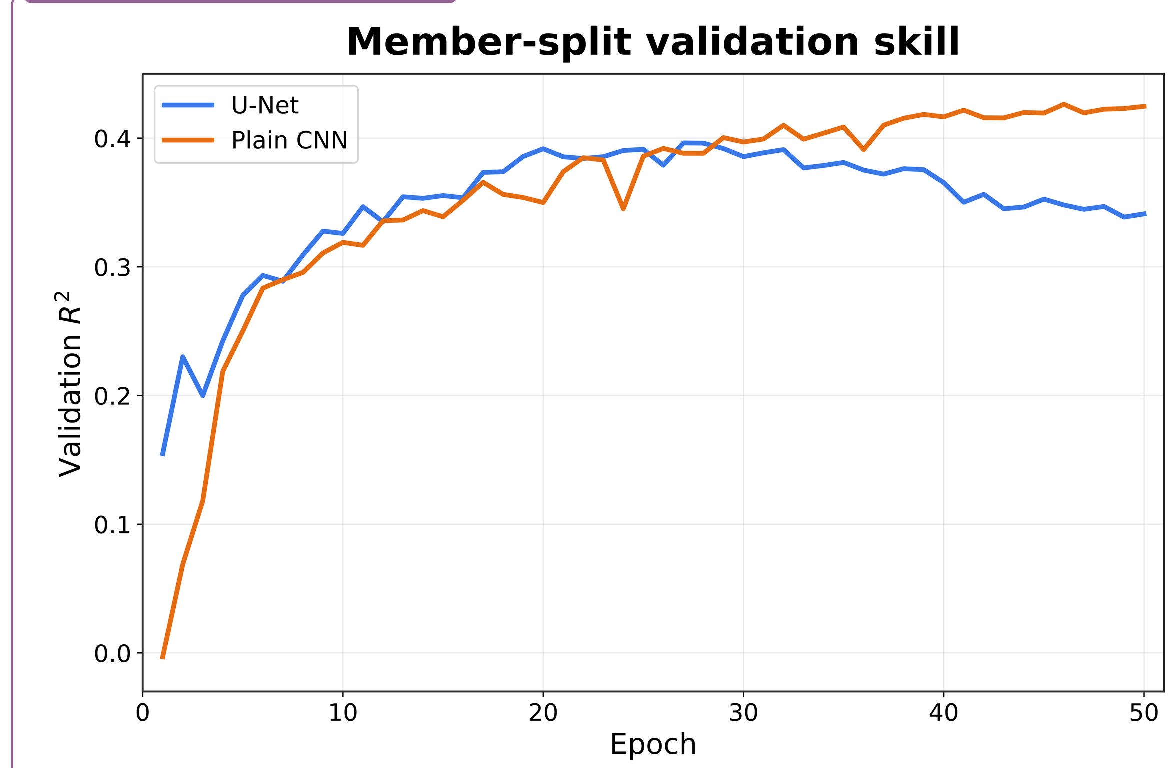
Slice models: harder member split.

- Train/validation/test members are disjoint.
- The task is to generalize to unseen parameter combinations.
- The predicted target is one deep upper-ocean slice over the full domain.

Patch CNN: closer to HM24.

- Uses all 25 members, but different output times for train and test.
- Predicts a center-column profile from local 32×32 surface patches.
- Evaluation is upper-200 m aggregate skill, matching HM24 more closely.

7. Training progress



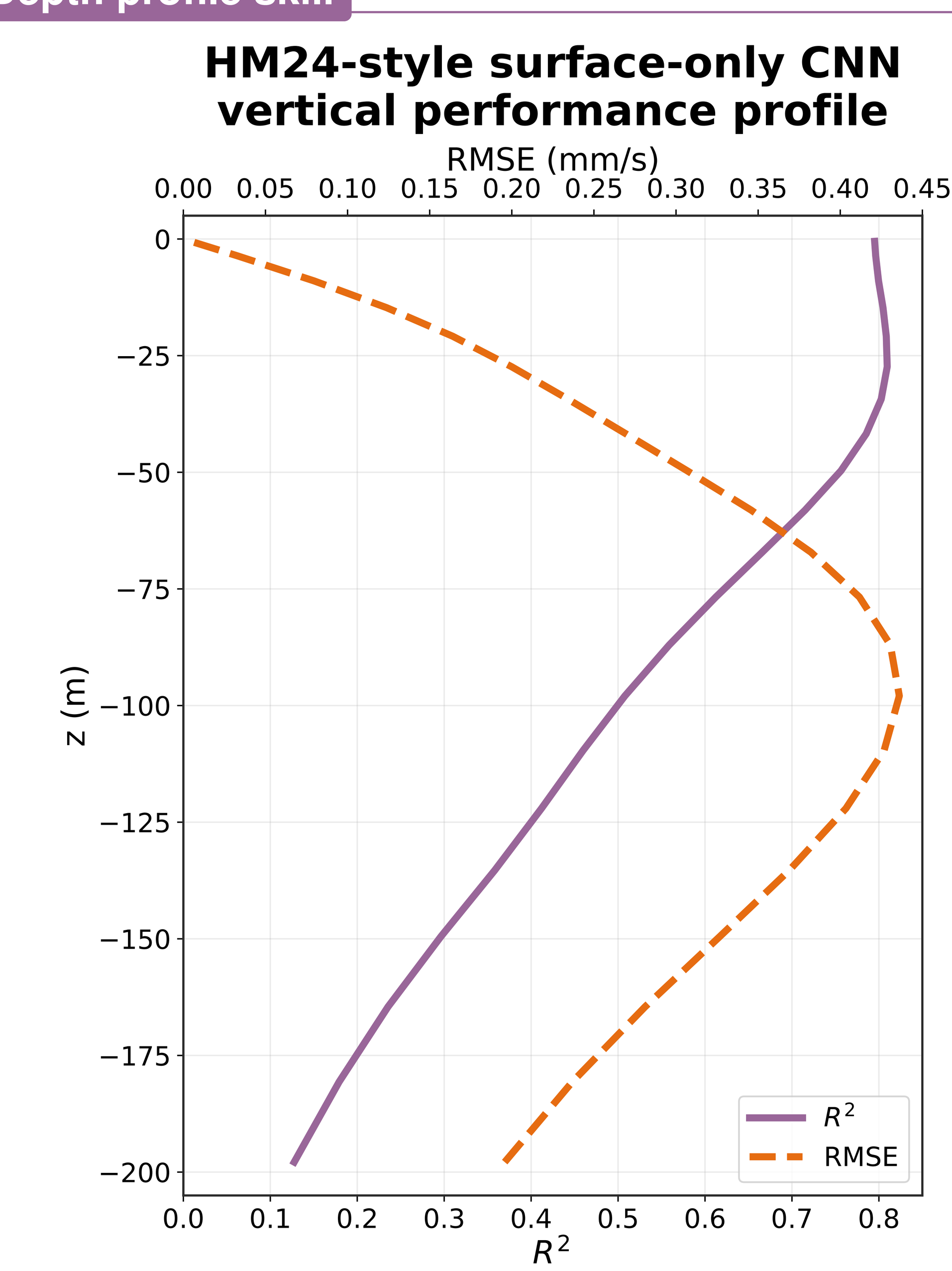
Plain CNN slightly outperforms U-Net on the member-split slice task; improvement slows after about 20–30 epochs.

8. Metric definitions

$$R^2 = 1 - \frac{\sum_i (w_i - \hat{w}_i)^2}{\sum_i (w_i - \bar{w})^2}, \qquad \mathrm{RMSE} = \sqrt{\langle (\hat{w} - w)^2 \rangle}.$$

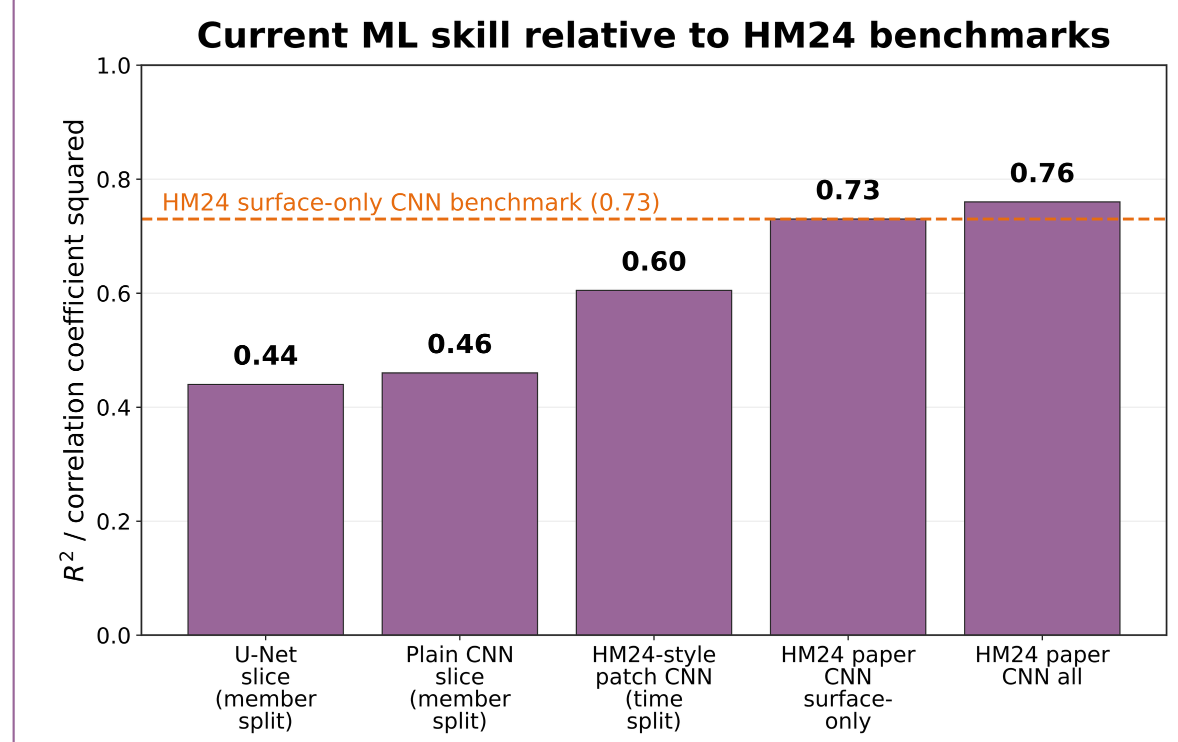
HM24's table reports an upper-200 m aggregate R^2 over the test set, not a simple average of depthwise R^2 values. We therefore compare both aggregate and depth-resolved skill when possible.

9. Depth profile skill



The HM24-style CNN is strongest near the surface and loses skill with depth, consistent with surface fields being more directly connected to mixed-layer w.

10. Current results



Model/test setting	Skill
U-Net slice, member split	$R^2 \approx 0.44$
Plain CNN slice, member split	$R^2 \approx 0.46$
HM24-style patch CNN, time split	$R^2 \approx 0.60$
HM24 paper CNN surface-only	$R^2 \approx 0.73$
HM24 paper CNN all inputs	$R^2 \approx 0.76$

11. Interpretation

- Current member-split results are lower than HM24's time-split CNN benchmarks.
- Part of the gap is expected: member-split testing is stricter, target design differs, and our data come from an Oceananigans reimplementation rather than HM24's PSOM output.
- The patch CNN is closest to HM24 because it predicts a center-column vertical profile instead of a full 2D map.
- Skill is strongest near the surface and decreases with depth.

12. Why are we below HM24?

1. **Stricter generalization:** member split versus time split.
2. **Different target:** full 2D deep slice versus center-column profile.
3. **Expanded ensemble:** 25 members increase variability and challenge.
4. **Input sensitivity:** surface divergence, stratification effects, and coarsening choices can change skill.

13. Next steps

- Reproduce HM24's train / test time split and patch-CNN setup.
- Move to PINNs for better performance.
- Try iterative and GNN-like architectures.

14. Acknowledgements

We wish to thank Shixun Huang, Han Wang, and Maksym Veremchuk for the helpful discussions and suggestions.

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

Nous remercions le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) de son soutien.

This research was enabled in part by support provided by the Digital Research Alliance of Canada (alliancecan.ca).

Physics UNIVERSITY OF TORONTO

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

NSERC CRSNG