FUTURE-QUERY WIND FORECASTING FOR QUADROTOR PREVIEW CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Quadrotor control can exploit wind preview to anticipate aerodynamic disturbances, yet onboard sensing is sparse and confined to the traversed trajectory, leaving future control queries outside direct sensing support. We introduce EgoCast, a framework for turning sparse onboard observations into short-horizon wind preview at future control queries. Although these queries are unobserved, recent sensing and vehicle motion still provide structure that can be carried toward them. EgoCast therefore decomposes forecasting into a sensing-supported transported reference and a learned residual for unresolved correction and short-horizon evolution; label-anchored physics further constrains this residual where direct supervision is sparse. Under matched training, transport–residual forecasting reduces future-query RMSE by 18.5–23.6% relative to direct full-field prediction. Physical supervision further improves forecasts and benefits 21 of 25 sensing–supervision configurations across shifted observation layouts, with generally larger gains farther from historical observations. These forecast gains reduce aerodynamic force-preview RMSE by 29.1–39.3% relative to the data-only counterpart and improve closed-loop tracking. Adding EgoCast preview to DOB-MPC reduces tracking error on all four tested trajectories relative to DOB-MPC, showing that anticipatory wind preview complements reactive disturbance estimation. Together, these results establish future-query wind forecasting as a practical interface between sparse onboard sensing and preview control.

1 INTRODUCTION

Quadrotor control can use wind preview to anticipate aerodynamic disturbances (Mendez et al., 2023). However, preview control requires wind information ahead of the vehicle, while onboard observations accumulate only along the traversed trajectory (Achermann et al., 2024; Tasim & Sun, 2026). This mismatch creates a *sensing–query support gap* between where wind is observed and where future control information is required (Figure 1). We formulate *future-query wind forecasting* as the problem of predicting short-horizon local wind at these unvisited control queries from sparse onboard observations.

Prior work has advanced both sparse/mobile field inference and future flow forecasting, including prediction from limited observations (Achermann et al., 2024; Ebers et al., 2024; Yin et al., 2023; Lu & Papadakis, 2026). Onboard preview combines these challenges in a moving-query setting: observations are sparse and confined to the traversed trajectory, while the required targets lie both in the future and at spatially displaced, unvisited control queries. As the vehicle moves, the geometry between past sensing and future queries changes continuously, requiring forecasts to extrapolate evolving local wind while accounting for this moving spatial support.

Future control queries lie outside direct sensing support, yet they are not devoid of usable structure: recent observations and vehicle motion define wind structure that can be transported toward them. Rather than forecasting the future field from scratch, we therefore decompose prediction into a sensing-supported transported reference and a learned residual that captures unresolved correction and short-horizon evolution. **EgoCast** operationalizes this decomposition through Ego-Transport and residual forecasting. Label-anchored physics further constrains the learned forecast where direct training supervision is sparse. The resulting continuous wind field is queried along the control

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.

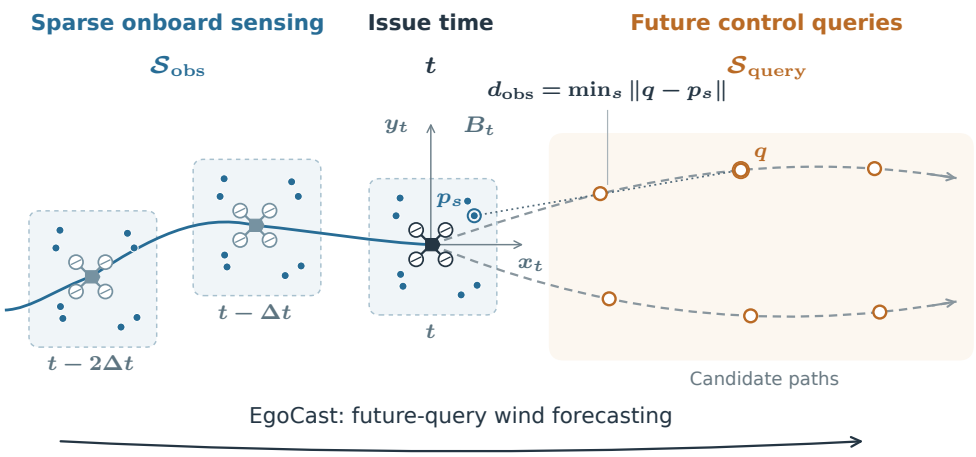

Figure 1: **Forecasting across the sensing–query support gap.** Past onboard observations are aligned in the issue-time frame, and the continuous wind forecast supports queries along different future paths. The closed-loop experiments use future control queries extrapolated from current motion. The dotted segment connects a future control query to its nearest historical observation. Boxes group discrete samples; the two-dimensional geometry is schematic.

horizon and mapped to aerodynamic force preview for model predictive control (MPC) (Mayne et al., 2000).

Matched comparisons show that transport–residual forecasting outperforms direct full-field prediction at future control queries, reducing RMSE by 18.5–23.6% under both data-only and physics-supervised training. Under changed observation layouts, physical supervision improves prediction in 21 of 25 sensing–supervision configurations, with generally larger gains farther from historical observations. These forecasting gains translate downstream, reducing aerodynamic force-preview RMSE by 29.1–39.3% relative to the data-only counterpart and improving closed-loop tracking. With disturbance-observer control (Xu et al., 2025), adding EgoCast preview to our DOB-MPC baseline reduces tracking error on all four tested trajectories, showing that anticipatory wind preview complements reactive disturbance estimation.

Our contributions are:

- **Future-query wind forecasting for preview control.** We formulate short-horizon wind prediction at future, unvisited control queries from sparse onboard observations, making explicit the sensing–query support gap between onboard sensing and the future wind information required for preview control.
- **Transport–residual forecasting with EgoCast.** We introduce a structured forecasting framework that preserves sensing-supported structure while focusing learning on unresolved short-horizon dynamics.
- **Label-anchored physics under sparse supervision.** We embed sparse training velocity labels into local physical constraints, coupling measured values to neighboring predictions beyond directly supervised locations.
- **Control-grounded evaluation of future wind preview.** We connect future-query wind forecasting to aerodynamic force preview and closed-loop control, including disturbance-observer control to assess the complementary roles of anticipatory preview and reactive estimation.

2 Related Work

Sparse wind-field reconstruction and forecasting. WindSeer (Achermann et al., 2024) reconstructs volumetric wind from sparse measurements and terrain, while related work reconstructs

fields from sparse or mobile sensors (Zhou et al., 2026; Ebers et al., 2024; Santos et al., 2023; Fukami et al., 2021). SHRED (Williams et al., 2024) and sparse-flow forecasting further show that limited histories can support subsequent field prediction (Yin et al., 2023; Lu & Papadakis, 2026). These methods do not explicitly use vehicle motion to transport sensing-supported wind structure to spatially displaced, unvisited future control queries.

Physical supervision under sparse data. Physics-informed learning supplements sparse labels with governing equations (Raissi et al., 2019; Karniadakis et al., 2021), supporting wind and mean-flow reconstruction (Tasim & Sun, 2026; Patel et al., 2024). PINP incorporates coupled physical quantities into flow forecasting (Chen et al., 2025). Conventional objectives evaluate label and physics losses separately, so measured values do not directly anchor neighboring physical relations. Measurement-embedded local constraints partly close this gap (Mo & Magri, 2025), but do not address future fields conditioned on moving observation histories.

Wind preview and disturbance compensation. Wind-aware controllers use remote sensing, local flow measurements, or learned dynamics (Mendez et al., 2023; Simon et al., 2023; O'Connell et al., 2022; Wang et al., 2025). Online wind mapping already provides spatial feedforward along predictive-control horizons (Torok et al., 2026). Disturbance observers provide reactive compensation from vehicle responses (Xu et al., 2025). EgoCast focuses on short-horizon wind evolution at unvisited future control queries, conditioned on sparse moving observation histories.

3 METHOD

EgoCast combines Ego-Transport and residual forecasting into a continuous wind forecast, trained with label-anchored physics and independent collocation constraints. The forecast is converted into force preview for MPC. Figure 2 illustrates the field construction.

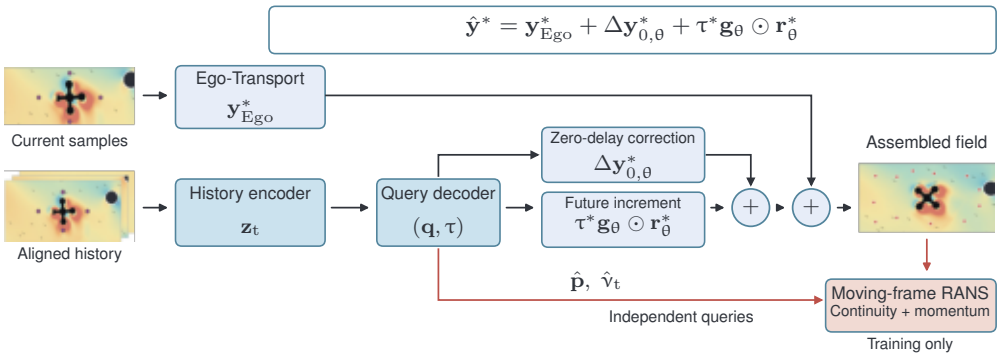

Figure 2: **EgoCast field construction.** An Ego-Transport reference, a zero-delay correction, and a gated future evolution increment form the forecast. Moving-frame continuity and momentum constraints provide training supervision.

3.1 EGO-TRANSPORT AND RESIDUAL FORECASTING

At issue time t, $\mathcal{H}_t^{(M)}$ contains the available velocity–pressure observations, channel-validity masks, and vehicle-motion measurements, with at most M observation sites per frame (the sensing budget). Known poses align historical observations to the issue-time body frame B_t, and a history encoder summarizes them into a shared context $\boldsymbol{z}_t$. Future queries use a local frame extrapolated from the current pose and motion state. At local position $\boldsymbol{q} \in \mathbb{R}^3$ and future delay $\tau > 0$, the model predicts

$$\hat{\boldsymbol{u}}_{t+\tau}(\boldsymbol{q}) = F_\theta^u(\mathcal{H}_t^{(M)}, \boldsymbol{q}, \tau). \tag{1}$$

Ego-Transport constructs a deterministic wind reference from recent sensing and vehicle motion. It interpolates the latest velocity observations using inverse-square-distance weights (Shepard, 1968) and queries this field at locations given by extrapolated motion, preserving the spatial structure supported by these observations. Conditioned on $\boldsymbol{z}_t$, residual forecasting supplies a current-field

correction and unresolved future evolution. The assembled forecast is

$$\hat{\boldsymbol{y}}^*_\theta(\boldsymbol{q}^*, \tau^* \mid \boldsymbol{z}_t) = \boldsymbol{y}^*_{\mathrm{Ego}}(\boldsymbol{q}^*, \tau^*) + \Delta\boldsymbol{y}^*_{0,\theta}(\boldsymbol{q}^* \mid \boldsymbol{z}_t) + \tau^* \boldsymbol{g}_\theta(\boldsymbol{q}^*, \tau^* \mid \boldsymbol{z}_t) \odot \boldsymbol{r}^*_\theta(\boldsymbol{q}^*, \tau^* \mid \boldsymbol{z}_t). \tag{2}$$

Here $\boldsymbol{y} = [u, v, w, p]^\top$, p is auxiliary pressure used by the physical constraints, $\boldsymbol{y}^*_{\mathrm{Ego}} = [(\boldsymbol{u}^*_{\mathrm{Ego}})^\top, 0]^\top$, and stars denote normalized quantities. The two learned terms constitute residual forecasting: $\Delta\boldsymbol{y}^*_{0,\theta}$ corrects the reference at zero delay using the observation history, and $\tau^*\boldsymbol{g}_\theta \odot \boldsymbol{r}^*_\theta$ models subsequent evolution. The sigmoid gate $\boldsymbol{g}_\theta$ scales the change rate, while τ^* makes the evolution increment vanish at zero delay. History alignment, normalization, and architecture details are given in Appendices A.1–A.3.

3.2 Label-anchored physics

During training, sparse velocity labels directly supervise future wind predictions. Label-anchored physics embeds these labels in local continuity and momentum relations within the same history-conditioned forecast, constraining neighboring predictions at unlabeled locations. Following the label-insertion approach of Mo & Magri (2025), we construct a $3 \times 3 \times 3$ finite-difference stencil near each selected training label. The residual center lies one stencil spacing from the labelled site, placing the known velocity at an axial neighbor. The stencil values are

$$\boldsymbol{u}_h(\boldsymbol{q}_j) = \begin{cases} \mathrm{stopgrad}\left(\boldsymbol{u}_{\mathrm{CFD}}(\boldsymbol{q}_s)\right), & \boldsymbol{q}_j = \boldsymbol{q}_s, \\ \hat{\boldsymbol{u}}_\theta(\boldsymbol{q}_j, \tau \mid \boldsymbol{z}_t), & \text{otherwise.} \end{cases} \tag{3}$$

Central differences of this hybrid field provide spatial derivatives, and automatic differentiation provides the temporal derivative at the stencil center. The known velocity thus enters the physical residuals at a neighboring predicted location, coupling the label to surrounding predictions (Figure 3).

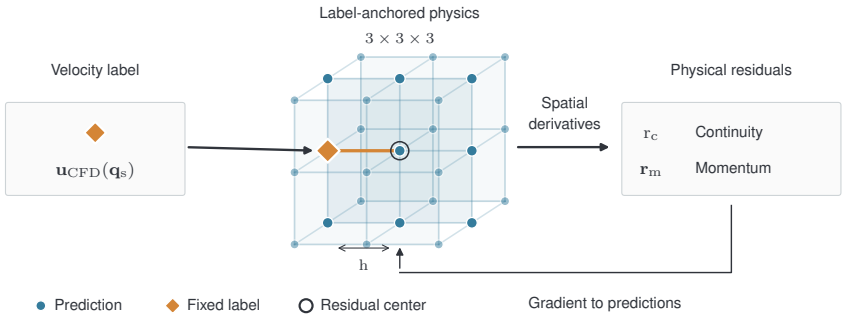

Figure 3: **Label-anchored physics.** A fixed velocity label replaces one axial neighbor in a $3 \times 3 \times 3$ stencil. Local continuity and momentum residuals propagate its supervision to neighboring predicted nodes.

These physical relations are evaluated in the moving frame used for future queries. The model expresses absolute fluid velocity in this frame, so the relative advection velocity is

$$\boldsymbol{c}^* = \hat{\boldsymbol{u}}^* - \boldsymbol{V}^*_c - \boldsymbol{\omega}^*_c \times \boldsymbol{q}^*. \tag{4}$$

Here $\boldsymbol{V}_c$ and $\boldsymbol{\omega}_c$ are the frame translational and angular velocities. Omitting normalization stars, the continuity and momentum residuals are

$$r_c = \nabla \cdot \hat{\boldsymbol{u}}, \tag{5}$$

$$\boldsymbol{r}_m = \partial_\tau \hat{\boldsymbol{u}} + (\boldsymbol{c} \cdot \nabla)\hat{\boldsymbol{u}} + \boldsymbol{\omega}_c \times \hat{\boldsymbol{u}} + \nabla\hat{p} - \nabla \cdot \left[(\nu + \hat{\nu}_t)\boldsymbol{S}(\hat{\boldsymbol{u}})\right]. \tag{6}$$

Here $\hat{p}$ is auxiliary modified kinematic pressure, ν is molecular kinematic viscosity, $\hat{\nu}_t$ is predicted eddy viscosity, and $\boldsymbol{S}$ is the deviatoric strain-rate tensor. Relative advection and the rotating-basis term keep these residuals consistent with the future-query coordinates.

Independent collocation points complement label anchoring by evaluating the same physical residuals through automatic differentiation at unlabeled locations. Both components act on the assembled forecast in Equation (2), updating the learned corrections while keeping Ego-Transport fixed. The total training loss is

$$\mathcal{L} = \mathcal{L}_{\text{future},u} + \lambda_0 \mathcal{L}_{\text{current},u} + (1-\eta)\mathcal{L}_{\text{ind}} + \eta \mathcal{L}_{\text{stencil}}, \qquad \eta = \frac{1}{8}. \tag{7}$$

Current and future velocity terms use masked mean-squared errors; physical terms use weighted Huber penalties (Huber, 1964). At inference, the trained field responds to queries using the available observation history. Discrete operators, validity screening, collocation budgets, and loss weights are given in Appendices A.4–A.6.

3.3 From wind forecasting to preview control

At each forecast update, the current motion state determines future control-query locations, where EgoCast supplies wind preview. Predicted velocities are transformed to world coordinates and expressed relative to the query-point motion induced by body translation and rotation. A calibrated aerodynamic mapping converts these relative velocities into force preview:

$$\begin{aligned} \hat{\boldsymbol{u}}^W_{j\ell} &= Q_j F^u_\theta(\mathcal{H}^{(M)}_t, \boldsymbol{q}_{j\ell}, \tau_j), \\ \hat{\boldsymbol{F}}^W_j &= \mathcal{A}\big(\{\hat{\boldsymbol{u}}^W_{j\ell} - \boldsymbol{v}^W_{j\ell}\}_\ell, Q_j\big)\,. \end{aligned} \tag{8}$$

Here j and ℓ index future times and query sites, Q_j rotates the future local frame into world coordinates, and $\boldsymbol{v}^W_{j\ell}$ is the query-point velocity. The fixed mapping $\mathcal{A}$ comprises coordinate conversion, spatial feature aggregation, and force estimation (Appendix C).

Temporal interpolation between a current force estimate from the latest observations and the predicted future forces provides the MPC force-preview sequence. It enters state prediction as a known disturbance:

$$\boldsymbol{x}_{k+1} = A(\psi_t)\boldsymbol{x}_k + B\delta \boldsymbol{f}_k + \boldsymbol{d}_k, \qquad (\boldsymbol{d}_k)_{v_x,v_y,v_z} = \frac{\Delta t_{\text{plan}}}{m}\hat{\boldsymbol{F}}^W_k. \tag{9}$$

MPC optimizes tracking error and control effort using these predictions and applies the first command. Force previews remain fixed during each solve and are refreshed as new observations arrive. Interpolation rules and controller parameters are given in Appendix C.

For DOB+EgoCast, the disturbance observer subtracts the previous force-preview contribution before estimating the residual disturbance. This feedback correction is added to the time-varying force preview, combining anticipatory preview with reactive disturbance estimation (Appendix C).

4 Experiments

Experimental setup. We evaluate wind forecasting and preview control in two-way coupled simulations using OpenFOAM (OpenCFD Ltd., 2024), preCICE (Chourdakis et al., 2022), and MuJoCo (Todorov et al., 2012). Forecast accuracy is measured by wind-vector RMSE, reported as mean $\pm$ sample SD over three paired training runs. All forecasting comparisons use held-motion queries, which extrapolate query attitude over the prediction horizon from the issue-time angular motion.

Persistence holds current observations; Ego-Transport queries a frozen interpolated field. DeepONet (Lu et al., 2021) and query-adapted SHRED (Williams et al., 2024) provide operator-learning and sparse-history-decoding baselines. EgoCast and its data-only counterpart (No-PDE) share their architecture and velocity supervision. Sensing budget M counts onboard observation sites per frame; supervision budget S counts labelled sites per training window. Data splits, query layouts, and training details are given in Appendix B.2.

4.1 Transport–residual forecasting at future control queries

We first test how observed wind structure supports forecasting at future control queries. Persistence and Ego-Transport establish the non-learned references. DeepONet and query-adapted SHRED

provide related-work baselines. We compare direct full-field and transport–residual prediction under both training objectives, matching networks, motion inputs, and training budgets. Only the additive reference differs; each budget uses the same training and validation layout across learned methods and the same evaluation layout (Appendix B.2).

Table 1: Future-query wind-vector RMSE ($\mathrm{m\,s^{-1}}$; 0.4–2.0 s). Learned models: mean ± sample SD over three paired runs; references: deterministic. Bold: lowest mean per budget.

Method	$M = S = 16$	$M = S = 8$
Persistence	1.6032	1.6032
Ego-Transport	**1.0154**	1.0326
DeepONet	1.9451 ± 0.4038	1.2610 ± 0.0692
Adapted SHRED	1.7274 ± 0.3904	1.1612 ± 0.0163
Direct field, data-only	1.5824 ± 0.3272	1.2716 ± 0.1001
Residual, data-only (No-PDE)	1.2089 ± 0.0616	1.0128 ± 0.0009
Direct field, physical supervision	1.2978 ± 0.0522	1.2233 ± 0.0900
EgoCast (residual, physical supervision)	1.0572 ± 0.0387	$\mathbf{0.9605 \pm 0.0109}$

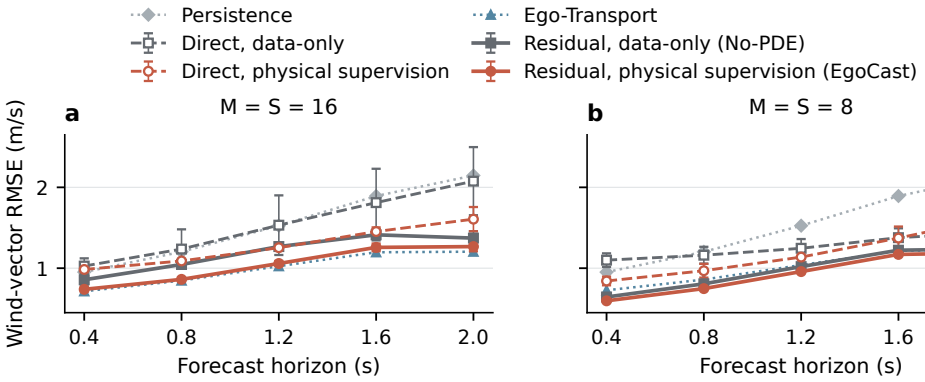

Figure 4: **Transport–residual forecasting versus direct prediction.** **(a)** $M = S = 16$; **(b)** $M = S = 8$. Dashed/open: direct full-field prediction; solid/filled: residual forecasting. Learned curves: gray, data-only; red, physical supervision. Dotted curves: deterministic references. Learned models: mean ± sample SD over three paired runs.

Among the related-work baselines, Adapted SHRED outperforms DeepONet at both budgets. Relative to Adapted SHRED, EgoCast reduces RMSE by 38.8% at $M = S = 16$ and 17.3% at $M = S = 8$.

Ego-Transport improves on Persistence at both budgets, showing the predictive value of current spatial observations and platform motion (Table 1). Relative to direct full-field prediction, transport–residual forecasting reduces mean RMSE by 18.5–23.6%, with lower error in all 12 paired comparisons. Gains under both training objectives show that the complete transport–residual parameterization is more effective than direct full-field prediction under the matched training budget.

EgoCast achieves the lowest mean error among learned methods at both budgets (Figure 4). Ego-Transport remains lower at $M = S = 16$, whereas EgoCast improves on it by 7.0% at $M = S = 8$.

4.2 Physical supervision and sparsity

We use ablations to examine the contributions of physical supervision, then assess how its benefits vary with distance from historical observations and with sensing and supervision sparsity.

Label-anchored physics strengthens supervision. At $M = S = 16$, physical residuals computed entirely from predicted values already improve on the data-only model. Inserting known velocity labels into the stencil further reduces error (Table 2), showing that label-anchored physics adds constraints beyond physical regularization alone.

Table 2: Physical-supervision ablation ($M = S = 16$; Appendix B.4). Future-query vector RMSE ($\mathrm{m\,s^{-1}}$): mean $\pm$ sample SD over three paired runs.

Method	Vector RMSE
No-PDE	1.2089 ± 0.0616
All-predicted stencil neighbors	1.0981 ± 0.0480
EgoCast	1.0573 ± 0.0387

To assess where label anchoring provides additional constraints, we compare all-predicted and label-anchored physical stencils under matched training configurations. Queries are grouped within each forecast horizon by distance to the nearest historical observation. Label anchoring further reduces RMSE by 9.9–13.5% in the farthest group, with improvements in all three training runs at every horizon (Table 3). The nearest group instead shows increased error at the three longest horizons, indicating that the benefit depends on location and is concentrated farther from historical observations. These results support embedding sparse velocity labels in local physical constraints to improve prediction where observation support is weaker.

Table 3: Label-anchoring gains over all-predicted stencils ($M = S = 16$). Mean paired RMSE reduction (%) over three runs; positive is better. Groups contain the nearest and farthest 25% of queries.

Horizon (s)	Nearest	Farthest
0.4	+3.48%	+13.47%
0.8	+0.37%	+11.49%
1.2	−3.87%	+10.26%
1.6	−6.31%	+9.90%
2.0	−6.43%	+10.76%

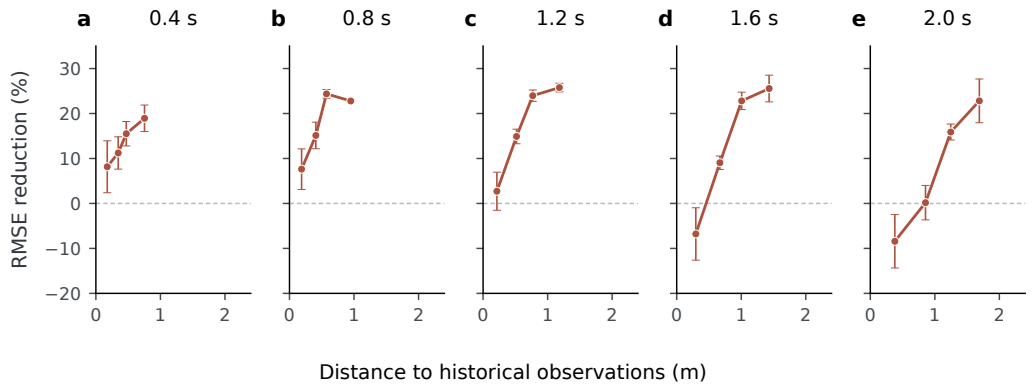

Figure 5: **Prediction gains versus distance to historical observations** ($M = S = 16$). Queries are binned by distance to the nearest historical observation within each delay. Points: bin midpoints and mean paired RMSE reduction versus No-PDE; bars: sample SD over three runs.

Physical supervision improves prediction beyond nearby observations. Within each forecast delay, we group queries by distance to the nearest historical observation. At all five delays, the farthest group benefits more than the nearest, with mean paired RMSE reductions of 18.94–25.77% relative to the data-only model (Figure 5). Physical supervision thus provides additional value at locations farther from available observations.

Benefits of physical supervision under sparse onboard sensing. We vary sensing and supervision budgets separately to examine how physical-supervision gains depend on both forms of sparsity. EgoCast and its data-only counterpart perform similarly at the densest sensing budget, whereas EgoCast improves on its data-only counterpart across all tested supervision budgets at $M = 16$. Overall, EgoCast outperforms its data-only counterpart in 21 of 25 budget combinations (Figure 6), showing that physical constraints can supplement sparse observations across different label budgets.

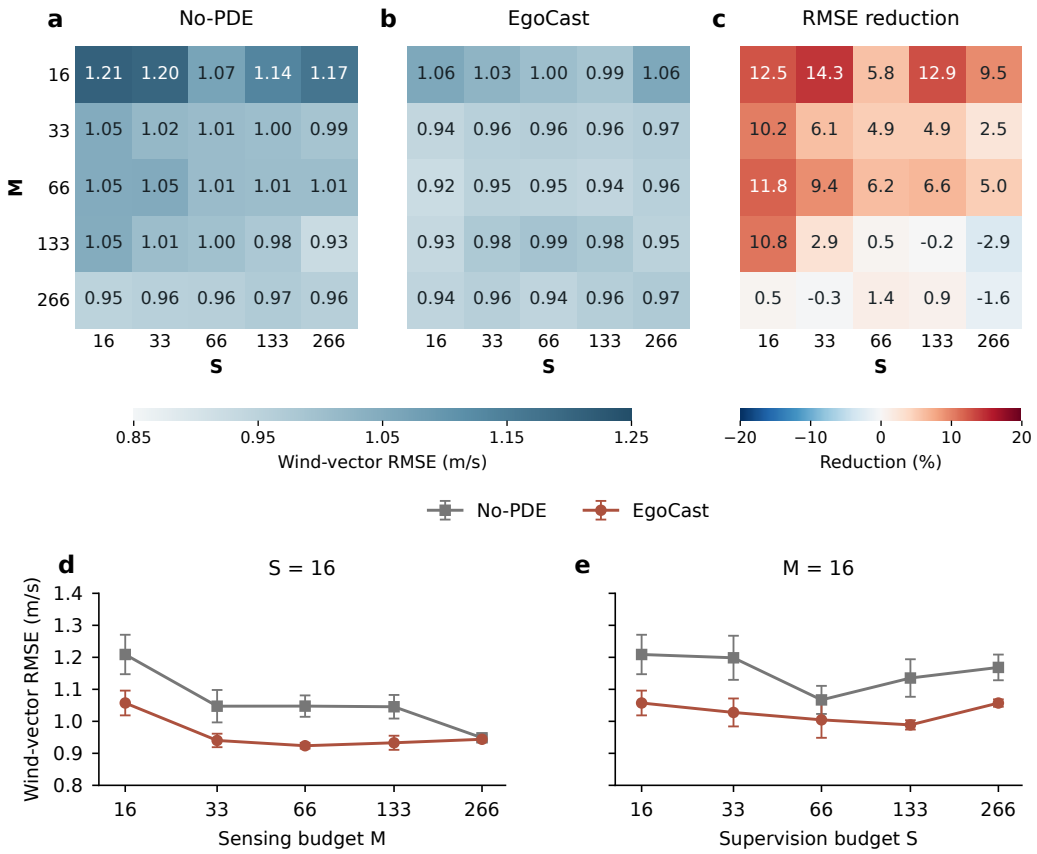

Figure 6: **Sensing and supervision budgets.** M: sensing budget; S: supervision budget. **(a,b)** Five-delay aggregate RMSE, averaged over three runs. **(c)** Reduction $100(1-\bar{E}_{\text{EgoCast}}/\bar{E}_{\text{No-PDE}})$; positive favors EgoCast. **(d,e)** Slices at $S = 16$ and $M = 16$: mean $\pm$ sample SD.

Complete ablations, budget scans, and distance analyses are given in Appendices B.4, B.3, and B.5.

4.3 From wind forecasting to closed-loop tracking

We next test whether improved wind prediction translates through aerodynamic force preview and MPC into better tracking. We first use $M = 16$, $S = 33$, the configuration with the largest physical-supervision gain in the preceding budget scan, then evaluate $M = S = 8$ to assess learned previews beyond transport under sparser sensing and supervision. Each control configuration is evaluated using one trained model for each learned method (Appendix C).

At $M = 16$, $S = 33$, EgoCast reduces future-force RMSE by 29.1–39.3% relative to the data-only model and achieves lower tracking error on all four trajectories (Table 4). The prediction gains from physical supervision therefore carry into force preview and closed-loop control.

Ego-Transport remains a strong baseline: it outperforms EgoCast on the three short trajectories in this configuration and is slightly worse on the long trajectory. At $M = S = 8$, EgoCast outperforms Ego-Transport on all four trajectories, reducing tracking error by 24–50% on the three short trajectories. Both residual predictors also outperform Adapted SHRED on these short trajectories. The data-only model and EgoCast achieve similar tracking errors, indicating that the main benefit in this sparse configuration comes from residual forecasting supplementing the transport reference.

Table 4: Force-preview and closed-loop position RMSE. Force errors pool shared-history predictions; tracking errors use complete trajectories. Values are rounded for display; percentage reductions in the text are computed from unrounded RMSE values.

(a) Force-preview RMSE (N)

Budget	Method	Arc climb	Arc descent	Spatial S	Extended
$M = 16, S = 33$	No-PDE	2.214	2.136	2.334	2.005
	EgoCast	1.343	1.493	1.508	1.423
$M = S = 8$	Ego-Transport	1.433	1.579	1.589	1.477
	No-PDE	0.711	0.972	0.917	1.061
	EgoCast	0.691	0.956	0.894	1.052

(b) Position RMSE (m), $M = 16$**,** $S = 33$

Trajectory	Pure MPC	Ego-Transport	No-PDE	EgoCast	DOB-MPC	DOB+EgoCast
Arc climb (5 s)	0.2272	0.0790	0.1475	0.0850	0.0802	0.0651
Arc descent (5 s)	0.2001	0.0549	0.1032	0.0599	0.0843	0.0695
Spatial S (5 s)	0.2229	0.0690	0.1268	0.0735	0.0848	0.0748
Extended trajectory (20 s)	0.2290	0.0985	0.1283	0.0969	0.0813	0.0698

(c) Position RMSE (m), $M = S = 8$

Trajectory	Pure MPC	Ego-Transport	Adapted SHRED	No-PDE	EgoCast
Arc climb (5 s)	0.2272	0.0885	0.0666	0.0448	0.0442
Arc descent (5 s)	0.2001	0.0614	0.0883	0.0464	0.0466
Spatial S (5 s)	0.2229	0.0764	0.0762	0.0506	0.0505
Extended trajectory (20 s)	0.2290	0.1021	0.0987	0.1008	0.1011

4.4 Control value of future wind information

To assess whether wind preview adds value beyond reactive disturbance compensation, we implement a DOB-MPC baseline using filtered acceleration residuals. The observer estimates current disturbances from vehicle motion, while adding EgoCast supplies time-varying force previews along the MPC horizon. We retain the $M = 16$, $S = 33$ configuration and four trajectories from Section 4.3 (Appendix C).

Adding preview reduces position RMSE by 11.80–18.86% relative to DOB-MPC across all four trajectories (Table 4). These gains show that wind preview improves tracking even with current-disturbance compensation. The combined controller also outperforms preview alone on two trajectories, supporting complementary anticipatory and reactive disturbance compensation.

5 Discussion and Conclusion

This work introduces EgoCast, a physics-informed forecasting framework that transforms sparse moving observations into continuous wind estimates at future control queries. Ego-Transport preserves sensing-supported structure, residual forecasting models unresolved evolution, and the resulting forecasts provide aerodynamic force previews for MPC.

Label-anchored physics embeds sparse velocity measurements into local physical constraints at unobserved locations. Matched comparisons show that residual forecasting improves on direct full-field prediction, while label anchoring strengthens prediction under shifted sensing layouts and farther from historical observations.

Coupled flow–flight simulations show that forecast improvements carry into force preview and trajectory tracking, including additional gains over DOB compensation. Closed-loop evaluation uses a single trained model per method and configuration, and the 0.2 s control query lies below the shortest supervised forecast delay. Validation is limited to the tested simulation conditions; future work will examine broader wind and sensing regimes and real-world flight.

Reproducibility Statement

Appendix A details EgoCast's formulation and implementation and indexes the supplementary evidence supporting the main-text results. Appendix B specifies the simulation setup, data partitions, sensing and supervision layouts, training and model-selection procedures, and evaluation metrics, together with per-run forecasting results and ablation analyses. Appendix C documents the aerodynamic mapping, MPC and disturbance-observer configurations, closed-loop evaluation protocols, and computational costs. Appendix D reports the setup and results of the supplementary current-field reconstruction study. The supplementary input archive provides exact layouts, window indices, reference trajectories, initial states, and force-map coefficients, indexed in its `README.md`.

AI Use Statement

Generative AI tools assisted with writing and language polishing, literature retrieval, method implementation, experiment design, and interpretation of results. The authors reviewed and verified all AI-assisted content and take full responsibility for the final manuscript.

References

Florian Achermann, Thomas Stastny, Bogdan Danciu, Andrey Kolobov, Jen Jen Chung, Roland Siegwart, and Nicholas Lawrance. WindSeer: Real-time volumetric wind prediction over complex terrain aboard a small uncrewed aerial vehicle. *Nature Communications*, 15(1):3507, 2024. doi: 10.1038/s41467-024-47778-4.

Huaguan Chen, Yang Liu, and Hao Sun. PINP: Physics-informed neural predictor with latent estimation of fluid flows. In *International Conference on Learning Representations*, 2025. URL `https://proceedings.iclr.cc/paper_files/paper/2025/hash/7c52ea3df1af5cf63b98c2f667249f82-Abstract-Conference.html`.

Gerasimos Chourdakis, Kyle Davis, Benjamin Rodenberg, Miriam Schulte, Frédéric Simonis, Benjamin Uekermann, Georg Abrams, Hans-Joachim Bungartz, Lucia Cheung Yau, Ishaan Desai, Konrad Eder, Richard Hertrich, Florian Lindner, Alexander Rusch, Dmytro Sashko, David Schneider, Amin Totounferoush, Dominik Volland, Peter Vollmer, and Oguz Ziya Koseomur. preCICE v2: A sustainable and user-friendly coupling library. *Open Research Europe*, 2:51, 2022. doi: 10.12688/openreseurope.14445.2.

Megan R. Ebers, Jan P. Williams, Katherine M. Steele, and J. Nathan Kutz. Leveraging arbitrary mobile sensor trajectories with shallow recurrent decoder networks for full-state reconstruction. *IEEE Access*, 12:97428–97439, 2024. doi: 10.1109/ACCESS.2024.3423679.

Kai Fukami, Romit Maulik, Nesar Ramachandra, Koji Fukagata, and Kunihiko Taira. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nature Machine Intelligence*, 3(11):945–951, 2021. doi: 10.1038/s42256-021-00402-2.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. doi: 10.1214/aoms/1177703732.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. doi: 10.1038/s42254-021-00314-5.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Shengqi Lu and George Papadakis. Forecasting the evolution of three-dimensional turbulent recirculating flows from sparse sensor data. *Journal of Fluid Mechanics*, 1030:A21, 2026. doi: 10.1017/jfm.2026.11251.

D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000. doi: 10.1016/S0005-1098(99)00214-9.

Arthur P. Mendez, James F. Whidborne, and Lejun Chen. Wind preview-based model predictive control of multi-rotor UAVs using LiDAR. *Sensors*, 23(7):3711, 2023. doi: 10.3390/s23073711.

Yaxin Mo and Luca Magri. Reconstructing unsteady flows from sparse, noisy measurements with a physics-constrained convolutional neural network. *Physical Review Fluids*, 10(3):034901, 2025. doi: 10.1103/PhysRevFluids.10.034901.

Michael O'Connell, Guanya Shi, Xichen Shi, Kamyar Azizzadenesheli, Anima Anandkumar, Yisong Yue, and Soon-Jo Chung. Neural-fly enables rapid learning for agile flight in strong winds. *Science Robotics*, 7(66):eabm6597, 2022. doi: 10.1126/scirobotics.abm6597.

OpenCFD Ltd. OpenFOAM v2406. Software release, 2024. URL https://dl.openfoam.com/source/v2406/.

Yusuf Patel, Vincent Mons, Olivier Marquet, and Georgios Rigas. Turbulence model augmented physics-informed neural networks for mean-flow reconstruction. *Physical Review Fluids*, 9(3): 034605, 2024. doi: 10.1103/PhysRevFluids.9.034605.

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. doi: 10.1016/j.jcp.2018.10.045.

Javier E. Santos, Zachary R. Fox, Arvind Mohan, Daniel O'Malley, Hari Viswanathan, and Nicholas Lubbers. Development of the Senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, 5(11):1317–1325, 2023. doi: 10.1038/s42256-023-00746-x.

Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference*, pp. 517–524. ACM, 1968. doi: 10.1145/800186.810616.

Nathaniel Simon, Allen Z. Ren, Alexander Piqué, David Snyder, Daphne Barretto, Marcus Hultmark, and Anirudha Majumdar. FlowDrone: Wind estimation and gust rejection on UAVs using fast-response hot-wire flow sensors. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5393–5399. IEEE, 2023. doi: 10.1109/ICRA48891.2023.10160454.

Abdullah Tasim and Wei Sun. Physics informed reconstruction of four-dimensional atmospheric wind fields using multi-UAS swarm observations in a synthetic turbulent environment. *arXiv preprint arXiv:2601.22111*, 2026. URL https://arxiv.org/abs/2601.22111.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012. doi: 10.1109/IROS.2012.6386109.

Mitchell Torok, Man Ching Melvin Chan, Donglin Sui, and Mohammad Deghat. Online wind mapping for coupled path planning and contouring control of quadrotors. *Sensors*, 26(16):5250, 2026. doi: 10.3390/s26165250.

Mengyun Wang, Bo Wang, Yifeng Niu, and Chang Wang. PI-WAN: A physics-informed wind-adaptive network for quadrotor dynamics prediction in unknown environments. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5413–5419. IEEE, 2025. doi: 10.1109/IROS60139.2025.11247234.

Jan P. Williams, Olivia Zahn, and J. Nathan Kutz. Sensing with shallow recurrent decoder networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 480(2298):20240054, 2024. doi: 10.1098/rspa.2024.0054.

Liwen Xu, Bailing Tian, Cong Wang, Junjie Lu, Dandan Wang, Zhiyu Li, and Qun Zong. Fixed-time disturbance observer-based MPC robust trajectory tracking control of quadrotor. *IEEE/ASME Transactions on Mechatronics*, 30(6):4272–4282, 2025. doi: 10.1109/TMECH.2024.3503062.

Yuan Yin, Matthieu Kirchmeyer, Jean-Yves Franceschi, Alain Rakotomamonjy, and Patrick Gallinari. Continuous PDE dynamics forecasting with implicit neural representations. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=B73niNjbPs.

Yihang Zhou, Chao Lin, Hideki Kikumoto, Ryozo Ooka, and Sibo Cheng. Rooftop wind field reconstruction using sparse sensors: From deterministic to generative learning methods. *Building and Environment*, 296:114480, 2026. doi: 10.1016/j.buildenv.2026.114480.

A METHOD AND IMPLEMENTATION DETAILS

Main-text result index. The held-motion forecasting results are given in Tables 8 and 9; physical-supervision ablations, distance groups, and joint budgets are given in Tables 12, 13, and 11, respectively. Tables 15 and 16 report the control and force-preview results, and Table 18 reports the disturbance-feedback comparison.

A.1 NORMALIZATION AND COORDINATE CONVENTIONS

The reference scales are $L_0 = 1$ m and $U_0 = 6\,\mathrm{m\,s^{-1}}$. For physical coordinates, velocities, pressure, and viscosity,

$$\boldsymbol{q}^* = \boldsymbol{q}/L_0, \qquad \tau^* = \tau U_0/L_0, \qquad \boldsymbol{u}^* = \boldsymbol{u}/U_0, \qquad p^* = p/U_0^2, \tag{10}$$

$$\boldsymbol{V}^* = \boldsymbol{V}/U_0, \qquad \boldsymbol{\omega}^* = \boldsymbol{\omega} L_0/U_0, \qquad \nu^* = \nu/(U_0 L_0), \qquad \nu_t^* = \nu_t/(U_0 L_0). \tag{11}$$

The molecular kinematic viscosity is $1.5 \times 10^{-5}\,\mathrm{m^2\,s^{-1}}$. The output rate in Equation (2) is a rate with respect to normalized time, so multiplying it by τ^* gives a normalized flow increment.

Let $\boldsymbol{p}_t^W$ and R_t denote the issue-time vehicle center and body-to-world rotation, and let $\boldsymbol{V}_t$ and $\boldsymbol{\omega}_t$ be the linear and angular velocities expressed in B_t. Holding their inertial values fixed defines the future query chart:

$$R_\tau = \exp([\boldsymbol{\omega}_t \tau]_\times), \qquad R_{t+\tau}^{\mathrm{hm}} = R_t R_\tau,$$

$$\boldsymbol{x}^W(\boldsymbol{q}, \tau) = \boldsymbol{p}_t^W + R_t(\boldsymbol{V}_t \tau + R_\tau \boldsymbol{q}). \tag{12}$$

The velocity target in this chart is $(R_t R_\tau)^\top \boldsymbol{u}_{\mathrm{CFD}}^W(\boldsymbol{x}^W, t+\tau)$. The chart specifies query coordinates and vector components; the realized trajectory is determined by the coupled flight dynamics.

For observation s acquired at t_h, history alignment gives

$$\widetilde{\boldsymbol{q}}_{h,s}^{B_t} = R_t^\top(\boldsymbol{p}_h^W + R_h \boldsymbol{q}_{h,s}^{B_h} - \boldsymbol{p}_t^W), \tag{13}$$

$$\widetilde{\boldsymbol{u}}_{h,s}^{B_t} = R_t^\top R_h \boldsymbol{u}_{h,s}^{B_h}. \tag{14}$$

Historical linear and angular velocities use the same rotation. Aligned observations retain their acquisition times and share the issue-time coordinate basis.

Historical positions and vectors are transformed into B_t before encoding. A future output instead uses the basis $R_t R_\tau$ defined by Equation (12). Consequently, the same inertial center velocity has components $R_\tau^\top \boldsymbol{V}_t$ at future delay τ. Although $R_\tau^\top \boldsymbol{\omega}_t = \boldsymbol{\omega}_t$ under held angular velocity, this invariance does not apply to a general linear velocity. Spatial derivatives act on future local coordinates; temporal derivatives hold these coordinates and the history context fixed. Since the output is absolute fluid velocity rather than grid-relative velocity, the basis term is $\boldsymbol{\omega}_c \times \boldsymbol{u}$, alongside grid-relative advection.

At each delay, the frame-velocity components are

$$\boldsymbol{V}_c^* = R_\tau^\top \boldsymbol{V}_t^*, \qquad \boldsymbol{\omega}_c^* = R_\tau^\top \boldsymbol{\omega}_t^*.$$

The deviatoric strain-rate tensor in Equation (6) is

$$\boldsymbol{S}(\hat{\boldsymbol{u}}) = \nabla \hat{\boldsymbol{u}} + (\nabla \hat{\boldsymbol{u}})^\top - \tfrac{2}{3}(\nabla \cdot \hat{\boldsymbol{u}}) I.$$

A.2 PRESSURE CONVENTION AND AUXILIARY FIELDS

The pressure input is the sampled CFD kinematic pressure centered on the available observations. Let $\mathcal{O}_{h,p}^{(M)}$ index the selected locations with valid pressure at history time t_h. For a nonempty observation set, the input at location s is

$$p_{h,s}^{\mathrm{in}} = p_{h,s} - \frac{1}{|\mathcal{O}_{h,p}^{(M)}|} \sum_{j \in \mathcal{O}_{h,p}^{(M)}} p_{h,j}, \qquad s \in \mathcal{O}_{h,p}^{(M)}. \tag{15}$$

Centering is performed separately for every historical frame, after the M input locations have been selected. Invalid pressure channels remain masked and use zero-filled input placeholders, including frames with no valid pressure. The centered values are normalized by U_0^2 for both original and motion-aligned observations. Models trained with this convention use the same reference at inference.

The output $\hat{p}$ is an auxiliary modified pressure that absorbs isotropic turbulent stress, distinct from the sampled input pressure. It enters the momentum residual through its spatial gradient, without support-dependent projection. CFD pressure is not inserted into the stencil. The current configuration has no pressure-target, direction, or turbulence-closure supervision. The positive eddy viscosity enters the deviatoric stress. These auxiliary fields support velocity regularization; they are not validated reconstructions of CFD pressure or the SST turbulence closure. The implementation retains a positive turbulent-kinetic-energy head, but it is neither supervised nor used by the source-free residual reported here.

A.3 History encoder and continuous decoder

For Ego-Transport, let $m_{s,k}$ indicate a valid velocity component k at current input location s. The componentwise interpolant is

$$[\mathcal{I}_t^*(\boldsymbol{x}^*)]_k = \frac{\sum_s m_{s,k} w_s(\boldsymbol{x}^*) u_{t,s,k}^*}{\sum_s m_{s,k} w_s(\boldsymbol{x}^*)}. \tag{16}$$

Weights use inverse-square distances with a distance floor of 10^{-6} in normalized coordinates, normalized separately over valid observations for each velocity component. At least one valid current observation is required per component. The reference uses the latest selected observations and issue-time kinematics, with the same held-motion geometry as the forecast targets. Its spatial and temporal dependence remains in the differentiation graph during physical regularization.

Ego-Transport evaluates this frozen interpolant using the held-motion query transform:

$$\boldsymbol{u}_{\mathrm{Ego}}^*(\boldsymbol{q}^*, \tau^*) = R_\tau^\top \mathcal{I}_t^*(\boldsymbol{V}_t^* \tau^* + R_\tau \boldsymbol{q}^*).$$

A Direct token concatenates the four normalized sample channels, four channel masks, and three aligned spatial coordinates. A shared pointwise MLP produces 128-dimensional features, pooled over valid locations by a masked mean. Linear and angular velocity histories enter two six-channel kinematic embeddings, each of width 128, as implemented by the history backbone. The pooled observation feature, kinematic embeddings, and scalar validity summary enter a one-layer GRU with 96-dimensional hidden state. Only the Direct observation stream is active. The frames are processed in temporal order, yielding the shared context

$$\boldsymbol{z}_t = \mathrm{GRU}_\theta(\boldsymbol{s}_{t_0}, \ldots, \boldsymbol{s}_t)_{\mathrm{last}}. \tag{17}$$

The coordinate decoder maps $(\boldsymbol{q}^*, \tau^*)$ through two 128-wide SiLU layers and adds a context projection. Three residual FiLM blocks condition these features on the history:

$$\boldsymbol{f}_{\ell+1} = \boldsymbol{f}_\ell + \mathrm{SiLU}\big[(W_\ell \boldsymbol{f}_\ell + \boldsymbol{b}_\ell) \odot (1 + 0.1 \tanh \boldsymbol{a}_\ell(\boldsymbol{z}_t)) + \boldsymbol{b}_\ell^{\mathrm{ctx}}(\boldsymbol{z}_t)\big]. \tag{18}$$

Writing D_θ for the smooth coordinate decoder and H_0 for its zero-delay correction head, the zero-delay correction is

$$\Delta \boldsymbol{y}_{0,\theta}^*(\boldsymbol{q}^* | \boldsymbol{z}_t) = H_0(D_\theta(\boldsymbol{q}^*, 0; \boldsymbol{z}_t)). \tag{19}$$

The zero-delay correction head uses zero-delay decoder features. The future-rate head uses future features concatenated with τ^*. Each flow head has a 128-wide SiLU hidden layer and four outputs. Their final weights and biases are initialized to zero, so both the zero-delay correction and the future evolution increment initially vanish. Residual forecasters therefore start from the Ego-Transport velocity field. The direct-field variants use the same neural initialization but omit the additive reference, so their initial velocity predictions are zero. The shared parameter initialization thus gives the two parameterizations different initial output fields. The gate is a four-channel affine head followed by a sigmoid, initialized uniformly to 0.1 using zero weights and bias $\log(0.1/0.9)$. The eddy-viscosity head uses softplus with positive floor 10^{-8} and initial normalized value 0.02.

The configuration uses masked mean spatial encoding and no query-to-sensor attention or spatial boundary gate. The sigmoid gate scales the future rate, while the zero-delay correction remains trainable at zero delay. The transport reference is added only to velocity; pressure is the sum of its current-head output and gated increment. Training histories are encoded once per example and reused by all its supervised and residual queries.

A.4 Spatial stencil and discrete moving-frame operator

For each labelled local position $\boldsymbol{q}_s = (q_{s,x}, q_{s,y}, q_{s,z})$, the residual center is

$$\boldsymbol{q}_c = \boldsymbol{q}_s + \sigma_s h \boldsymbol{e}_x, \qquad \sigma_s = \begin{cases} -1, & q_{s,x} \geq 0, \\ +1, & q_{s,x} < 0, \end{cases} \qquad h = 0.125 \text{ m}. \tag{20}$$

The nodes are $\boldsymbol{q}_c + h(i,j,k)$ for $i,j,k \in \{-1,0,1\}$. The original supervised point is one axial neighbor, not the residual center. At each optimizer update, one of the five labelled future delays is selected cyclically and used for all stencils in the labelled batch.

Below, all variables and h are normalized, and subscripts denote stencil offsets about the center. For a scalar or component field f,

$$D_a f = \frac{f_{+\boldsymbol{e}_a} - f_{-\boldsymbol{e}_a}}{2h}, \tag{21}$$

$$D_{aa} f = \frac{f_{+\boldsymbol{e}_a} - 2f_0 + f_{-\boldsymbol{e}_a}}{h^2}, \tag{22}$$

$$D_{ab} f = \frac{f_{+\boldsymbol{e}_a+\boldsymbol{e}_b} - f_{+\boldsymbol{e}_a-\boldsymbol{e}_b} - f_{-\boldsymbol{e}_a+\boldsymbol{e}_b} + f_{-\boldsymbol{e}_a-\boldsymbol{e}_b}}{4h^2}, \qquad a \neq b. \tag{23}$$

Let $\boldsymbol{u}_h$ be the hybrid velocity in Equation (3). Define

$$A_{ia} = D_a u_{h,i}, \qquad d = \operatorname{tr} A, \qquad \boldsymbol{S}_h = A + A^\top - \tfrac{2}{3} d I, \tag{24}$$

$$(\Delta_h \boldsymbol{u}_h)_i = \sum_a D_{aa} u_{h,i}, \qquad (\boldsymbol{b}_h)_a = \sum_b D_{ab} u_{h,b}. \tag{25}$$

Here $\boldsymbol{b}_h$ discretizes $\nabla(\nabla \cdot \boldsymbol{u})$. The variable-viscosity contribution used by the implementation is

$$\boldsymbol{v}_h = (\nu + \hat{\nu}_{t,0}) \left(\Delta_h \boldsymbol{u}_h + \tfrac{1}{3} \boldsymbol{b}_h \right) + \boldsymbol{S}_h D\hat{\nu}_t. \tag{26}$$

This expands the divergence of the deviatoric stress, including the eddy-viscosity gradient.

The residuals at the predicted center are

$$r_{c,h} = d, \tag{27}$$

$$\boldsymbol{r}_{m,h} = \partial_\tau \hat{\boldsymbol{u}}_0 + A(\hat{\boldsymbol{u}}_0 - \boldsymbol{V}_c - \boldsymbol{\omega}_c \times \boldsymbol{q}_c) + \boldsymbol{\omega}_c \times \hat{\boldsymbol{u}}_0 + D\hat{p} - \boldsymbol{v}_h. \tag{28}$$

Automatic differentiation evaluates $\partial_\tau \hat{\boldsymbol{u}}_0$ from the neural field at the center. The spatial differences use detached CFD velocity at the one known node and network predictions elsewhere. Pressure and eddy viscosity are predicted throughout. No temporal interpolation of CFD truth is used. This is a Cartesian finite-difference regularizer, not the native OpenFOAM finite-volume operator.

A.5 Supports, masks, and collocation budgets

The valid target domain is

$$\Omega^W_{\text{valid}}(s) = \Omega^W_{\text{fluid}}(s) \setminus \mathcal{E}^W_{\text{rot}}(s), \tag{29}$$

It excludes solids and the immediate rotor-source region. Queries satisfy $\boldsymbol{x}^W(\boldsymbol{q}, \tau) \in \Omega^W_{\text{valid}}(t + \tau)$ and lie within the local prediction bounds. Vehicle-induced flow in this domain remains part of the prediction target.

Observation selection begins with a 338-point field layout. A body-centered sphere of radius 0.25 m and four body-z-aligned rotor-source cylinders define its exclusion mask. Cylinder centers are $(\pm 0.45, 0, 0.09)$ and $(0, \pm 0.45, 0.09)$ m, each with radius 0.40 m and half-height 0.15 m. The retained layout contains 266 positions. Prediction bounds are the coordinate-wise extent of the retained layout.

The separate S-site supervision and 30-site evaluation arrays are sampled at the world positions defined by held motion. Their masks require finite sampled values and valid overset cell status (finite cell-type value below 1.5). Native target validity and geometric residual screening serve different roles, as specified below.

Residual screening. Training stencils require all 27 nodes inside the local prediction bounds and outside the configured rotor-source cylinders, with valid inserted velocity channels.

Of the $C = 64$ nominal centers, eight use the labelled batch and the shifted stencils. Rejected stencils do not contribute to either continuity or momentum. The other 56 positions are fixed coordinates from the fixed collocation layout, separate from the supervised-site centers. They use training histories and random delays spanning $[0, 2.0]$ s; their residuals are evaluated by automatic differentiation. Momentum additionally uses the configured rotor-source mask. They require no future labels in the optimization objective.

The physics budget counts residual centers, not all decoder calls. An accepted hybrid center requires 27 spatial-node evaluations and a temporal-derivative evaluation at its center. Each regional penalty is normalized by its own valid scalar count. Mask rejection does not change the mixing weights specified in Appendix A.6. A regional penalty with no valid entries contributes zero, without reallocating its weight to the other region. Therefore $C = 64$ is not a guarantee of 64 accepted centers at every update.

A.6 Supervised losses and physics weighting

We denote the sensing budget (input locations per frame) by M, the supervision budget (labelled spatial locations per training window) by S, and nominal residual centers by C. A supervised site is a spatial location; a velocity label is the known target value at that site and a specified time. Each accepted local center requires 27 field evaluations. The architecture-matched data-only model (No-PDE) retains the forecast construction and velocity supervision and uses zero physical penalties (Section 4.1).

For normalized scalar channels with mask m_i, write

$$\mathfrak{M}(\hat{a}, a; m) = \frac{\sum_i m_i(\hat{a}_i - a_i)^2}{\max(\sum_i m_i, 1)}. \tag{30}$$

The supervised objective is the sum of masked velocity MSE over the five future delays and masked current-velocity MSE at the S supervised sites:

$$\mathcal{L}_{\text{sup}} = \mathfrak{M}(\hat{\boldsymbol{u}}^*, \boldsymbol{u}^*_{\text{CFD}}; m_f) + \mathfrak{M}(\hat{\boldsymbol{u}}^*_\theta(\boldsymbol{q}^*, 0), \boldsymbol{u}^*_{\text{CFD},0}; m_0). \tag{31}$$

Each mean pools its valid scalar velocity channels over the batch and spatial support (and future delays for the first term). Current supervision is at the separate supervised sites, not an exact interpolation condition on the M onboard observation sites.

For scalar residual r, the Huber penalty is

$$\rho_\delta(r) = \begin{cases} \frac{1}{2}r^2, & |r| \le \delta, \\ \delta(|r| - \frac{1}{2}\delta), & |r| > \delta, \end{cases} \qquad \delta = 0.05. \tag{32}$$

Let $\langle\cdot\rangle_{\mathcal{C}}$ denote the masked mean over valid scalar residual components in a region. Then

$$\mathcal{L}_{\text{ind}} = \lambda_c\langle\rho_\delta(r_c)\rangle_{\text{ind}} + \lambda_m\langle\rho_\delta(\boldsymbol{r}_m)\rangle_{\text{ind}}, \tag{33}$$

$$\mathcal{L}_{\text{stencil}} = \lambda_c\langle\rho_\delta(r_{c,h})\rangle_{\text{stencil}} + \beta\lambda_m\langle\rho_\delta(\boldsymbol{r}_{m,h})\rangle_{\text{stencil}}. \tag{34}$$

The forecasting runs use $\lambda_0 = 1$, $\eta = 1/8$, $\beta = 10$, and $\lambda_c = \lambda_m = 20$. The factor β scales the Huber-penalized momentum term. The stated velocity, continuity, and momentum loss weights are active from the first epoch.

Fixed-mixture protocol. The fixed-mixture protocol in Equation (7) uses $\eta = 1/8$ independently of S/C:

$$\mathcal{L}_{\text{phys}} = \tfrac{7}{8}\mathcal{L}_{\text{ind}} + \tfrac{1}{8}\mathcal{L}_{\text{stencil}}.$$

The mixture remains fixed across observation and supervision budgets and after validity masking. With $\beta = 10$, the independent and stencil momentum means have coefficients $7\lambda_m/8$ and $10\lambda_m/8$, respectively. The corresponding continuity coefficients are $7\lambda_c/8$ and $\lambda_c/8$. The budget scans retain eight labelled stencils and 56 independent centers per update. For $S > 8$, the stencil anchors rotate across the labelled sites (Appendix B.3).

Both learned arms in the reported physics-loss comparison use the same 54 labelled windows. The known stencil values come only from the labelled batch and reuse its existing future velocity targets. Training runs for 768 updates; model selection uses validation data loss, combining current and future velocity errors, rather than test error.

B Wind Forecasting: Protocols and Supplementary Results

B.1 Held-motion evaluation: seed-level results

Tables 5 and 6 give the paired $M = S = 8$ records underlying the primary held-motion evaluation. Table 5 gives all three paired comparisons. Table 6 reports individual horizon errors. Both arms use identical target arrays, masks, and test window keys. The test set contains 3,884 valid future vectors.

Table 5: Held-motion evaluation at $M = S = 8$: selected epochs and aggregate errors by training seed. RMSE units are $\mathrm{m\,s^{-1}}$ for vector and speed and degrees for direction. Gains compare vector RMSE within seed.

Seed	Method	Epoch	Vector	Speed	Direction	Gain (%)
20260825	No-PDE	1	1.0139	0.4685	28.53	—
	EgoCast	5	0.9497	0.4614	26.66	6.33
20260826	No-PDE	1	1.0125	0.4684	28.74	—
	EgoCast	7	0.9603	0.4650	27.19	5.15
20260827	No-PDE	1	1.0122	0.4710	28.43	—
	EgoCast	6	0.9716	0.4693	27.31	4.01

Table 6: Held-motion evaluation at $M = S = 8$: individual vector RMSE ($\mathrm{m\,s^{-1}}$) at each forecast delay, using the models in Table 5.

Seed	Method	0.4 s	0.8 s	1.2 s	1.6 s	2.0 s
20260825	No-PDE	0.6547	0.8112	1.0181	1.2185	1.2403
	EgoCast	0.6008	0.7477	0.9500	1.1543	1.1664
20260826	No-PDE	0.6357	0.8021	1.0211	1.2271	1.2394
	EgoCast	0.5910	0.7422	0.9549	1.1721	1.1965
20260827	No-PDE	0.6413	0.8064	1.0200	1.2217	1.2386
	EgoCast	0.5976	0.7534	0.9694	1.1856	1.2066

Angular metrics use the three-dimensional included angle after clipping its cosine to $[-1, 1]$. They exclude vectors for which either predicted or target speed is at most $10^{-6}\,\mathrm{m\,s^{-1}}$. All 3,884 valid test vectors pass this additional criterion for all evaluated models. Direction RMSE pools squared angular errors across all valid vectors; it is not the mean absolute angular error. Relative differences of mean RMSEs and means of paired relative differences are distinct summaries; the manuscript uses the latter for paired gains.

B.2 Forecasting setup and supplementary results

CFD data generation. The forecasting campaigns use OpenFOAM v2406 (`overPimpleDyMFoam`) with incompressible k–ω SST and kinematic viscosity $1.5 \times 10^{-5}\,\mathrm{m^2\,s^{-1}}$. The domain is $[-6, 8] \times [-5, 5] \times [-2.75, 8.25]\,\mathrm{m^3}$. Eight cylinders of radius 0.28 m extend from $z = 0.75$ to 4.25 m; their centres lie on a radius-3.9 m ring at angles $22.5° + 45°i$, $i = 0, \ldots, 7$. A body-attached component mesh moves through a stationary background, using inverse-distance overset interpolation. The background starts from $52 \times 36 \times 40$

cells and the component from $20 \times 20 \times 16$ cells over a $2.5 \times 2.5 \times 2.0$ m box. After local refinement and assembly, the mesh contains 225,138 cells (including overset cells).

The prescribed, spatially varying boundary velocity has nominal speed scale $2\,\mathrm{m\,s^{-1}}$, sinusoidal gust amplitude $0.25\,\mathrm{m\,s^{-1}}$ and frequency 0.28 Hz, shear coefficient 0.08, and cross-stream and vertical biases of 0.12 and $0.08\,\mathrm{m\,s^{-1}}$. Circular campaigns rotate the forcing direction in 60° increments. Velocity is prescribed on the five outer patches other than the bottom; the bottom and cylinders are no-slip, and the body uses a moving-wall condition. Pressure is fixed to zero on the patch named outlet and has zero normal gradient on the other physical boundaries.

Time integration is first-order Euler, with upwind convection and corrected central diffusion. Each step uses one PIMPLE outer iteration and two pressure correctors. Serial-explicit preCICE coupling advances CFD and MuJoCo at $\Delta t_{\mathrm{CFD}} = 0.004$ s: vehicle pose and rotor commands determine mesh motion and CFD rotor sources, while CFD surface and rotor loads drive the rigid-body dynamics. Each forecasting campaign lasts 5 s; fields and observation frames are retained every 0.1 s. Target sampling and validity screening then follow the query protocols below.

Campaign splits and default budget. The reported forecasting comparison uses six circular-trajectory campaigns with inflow directions separated by 60° and two figure-eight training campaigns. Campaigns are assigned to training, validation, and testing before temporal windows are assembled (Table 7). With the present history and forecast horizon, each campaign provides 26 admissible windows. Training labels are restricted to nine windows per training campaign. The validation campaign provides 26 available windows; checkpoint selection uses the same fixed nine-window subset for the $8/8$, $16/16$, and $16/33$ models and the matched physics ablations.

Table 7: Campaign and window allocation for future forecasting. Labelled windows provide current and future velocity targets; the physics pool supplies training histories for independent residual evaluation.

Split	Inflow directions (°)	Available windows	Labelled windows
Training	$\{0, 60, 120, 180\}$	104	36
Training (figure-eight)	two campaigns	52	18
Validation	$\{240\}$	26	26
Test	$\{300\}$	26	26

Observations, supervision, and reported queries. The learned models receive M direct velocity–pressure samples per frame at the causal history offsets

$$(-0.2, -0.1, 0.0)\ \mathrm{s},$$

and are trained at five forecast delays,

$$\tau \in \{0.4, 0.8, 1.2, 1.6, 2.0\}\ \mathrm{s}. \tag{35}$$

The input locations are fixed subsets of the retained CFD observation layout. Historical pose aligns the samples, and linear and angular velocities enter the history encoder.

Supervision uses a separate set of S fixed local positions, with current and future targets sampled from saved CFD fields using held-motion geometry. Thus the sensing budget and the label budget refer to different spatial sets. Validation and the held-motion comparison use 30 fixed exterior query locations, distinct from the supervised sites. The query layout is selected using training data and fixed before test evaluation. Layout selection uses a larger training-data pool than the 54 labelled windows; both model arms share the layout.

Invalid CFD targets are omitted using the validity masks. At the five future delays, the held-motion test set contains 3,884 valid vectors out of 3,900 requested window–delay–location combinations. The input, target, and physics-support checks are specified separately in Appendix A.5.

Held-motion control queries. All primary forecasting comparisons use the same 26 test windows and 30 readout offsets. Query positions and attitudes follow the held-motion construction in Equation (12): issue-time linear and angular motion are extrapolated over each forecast delay, without

using recorded future vehicle motion. Targets are sampled from saved CFD fields at these world positions. The common validity mask retains 3,884 of the 3,900 requested window–delay–location vectors.

Persistence holds the nearest current observation over the horizon. Ego-Transport evaluates its frozen current field at the held-motion query locations and rotates vector components into the corresponding query frame. The two sensing budgets share the same query positions and CFD targets; their observation layouts differ according to M. All learned variants use three paired training runs. Aggregate RMSE pools squared vector errors across all delays, and percentage reductions in Section 4.1 use ratios of mean RMSEs.

Table 8: Held-motion future-query vector RMSE ($\mathrm{m\,s^{-1}}$). Learned models report mean $\pm$ sample SD over three paired runs; deterministic references have one value.

Method	Overall	0.4 s	0.8 s	1.2 s	1.6 s	2.0 s
$M = S = 16$						
Persistence	1.6032	0.9515	1.2030	1.5258	1.8926	2.1464
Ego-Transport	1.0154	0.7142	0.8472	1.0248	1.1969	1.2041
DeepONet	1.9451 ± 0.4038	1.8496 ± 0.5062	1.9474 ± 0.3318	1.9651 ± 0.3470	1.9752 ± 0.3979	1.9797 ± 0.4538
Adapted SHRED	1.7274 ± 0.3904	1.6642 ± 0.4801	1.6808 ± 0.4716	1.7143 ± 0.4217	1.7914 ± 0.3409	1.7721 ± 0.2683
Direct-data	1.5824 ± 0.3272	1.0264 ± 0.0955	1.2367 ± 0.2438	1.5320 ± 0.3688	1.8110 ± 0.4187	2.0763 ± 0.4214
No-PDE	1.2089 ± 0.0616	0.8568 ± 0.0615	1.0457 ± 0.0437	1.2676 ± 0.0476	1.4129 ± 0.0693	1.3723 ± 0.0854
Direct-physics	1.2978 ± 0.0522	0.9844 ± 0.0274	1.0897 ± 0.0246	1.2533 ± 0.0238	1.4534 ± 0.0514	1.6065 ± 0.1501
EgoCast	1.0572 ± 0.0387	0.7385 ± 0.0255	0.8617 ± 0.0242	1.0580 ± 0.0317	1.2573 ± 0.0479	1.2672 ± 0.0586
$M = S = 8$						
Persistence	1.6032	0.9515	1.2030	1.5258	1.8926	2.1464
Ego-Transport	1.0326	0.7269	0.8593	1.0390	1.2150	1.2304
DeepONet	1.2610 ± 0.0692	1.0596 ± 0.0856	1.1993 ± 0.1016	1.2770 ± 0.0756	1.3659 ± 0.0484	1.3762 ± 0.0458
Adapted SHRED	1.1612 ± 0.0163	0.9526 ± 0.0187	1.0451 ± 0.0362	1.1435 ± 0.0379	1.2780 ± 0.0432	1.3419 ± 0.0572
Direct-data	1.2716 ± 0.1001	1.0995 ± 0.0840	1.1607 ± 0.1040	1.2462 ± 0.1144	1.3778 ± 0.1154	1.4423 ± 0.0848
No-PDE	1.0128 ± 0.0009	0.6439 ± 0.0098	0.8066 ± 0.0046	1.0197 ± 0.0015	1.2224 ± 0.0043	1.2394 ± 0.0008
Direct-physics	1.2233 ± 0.0900	0.8417 ± 0.0535	0.9686 ± 0.0851	1.1373 ± 0.1242	1.3728 ± 0.1423	1.6298 ± 0.0994
EgoCast	0.9605 ± 0.0109	0.5965 ± 0.0050	0.7477 ± 0.0056	0.9581 ± 0.0101	1.1706 ± 0.0157	1.1898 ± 0.0209

Related-work baselines and transport–residual ablation. DeepONet uses a branch network to encode the causal observation history and a trunk network to encode future space–time queries. Adapted SHRED retains the original LSTM history encoder and replaces the fixed-output decoder with a coordinate–time decoder for changing future control queries. Both baselines predict wind directly, without Ego-Transport or physical supervision, and use the same data split, query targets, and training budget as the learned models in Table 1.

The direct models predict the complete current field plus a gated temporal increment. They retain the history encoder, motion inputs, query decoder, current-field head, temporal head, and gate used by residual forecasting; the additive Ego-Transport reference is removed. Each structural pair has 697,405 trainable parameters and identical initial parameter tensors, labelled data, optimizer updates, and validation selection. Physical-supervision pairs additionally share the same history pool, collocation sequence, stencil geometry, label insertion, and loss weights. All models train for 24 epochs and 768 updates using seeds 20260825–20260827. Both flow heads have zero-initialized final layers: the direct predictor starts from zero velocity, while the residual-forecasting model starts from Ego-Transport. Models are selected by validation data loss. The comparison therefore evaluates the complete forecasting parameterization, including its initial output field, rather than the additive reference in isolation. Direct-data and Direct-physics denote the two full-field variants in Tables 8–9; No-PDE and EgoCast retain the transport reference.

Spatial forecast-error display. Figure 7 provides an $M = S = 8$ qualitative slice using held-motion query geometry, comparing EgoCast with the architecture-matched data-only forecaster. Pointwise error is $e = \|\hat{\boldsymbol{u}} - \boldsymbol{u}_{\mathrm{CFD}}\|_2$. Blue in the difference maps denotes $e_{\text{No-PDE}} > e_{\text{EgoCast}}$, and red denotes the reverse.

The display retains window 4, originally selected for high query-error reduction, and re-evaluates it with held-motion queries using seed 20260826. It illustrates one forecast; aggregate comparisons use all 26 test windows. Both forecasts use history ending at $t = 0.8\,\mathrm{s}$ and the same horizontal slice at $z^W = 2.456\,\mathrm{m}$. CFD velocities are linearly reconstructed from active cell centres on a

Table 9: Individual held-motion vector RMSE ($\mathrm{m\,s^{-1}}$). Seeds 25–27 denote 20260825–20260827.

Budget	Method	Seed	Overall	0.4 s	0.8 s	1.2 s	1.6 s	2.0 s
16/16	DeepONet	25	1.8818	1.6449	1.9305	1.9404	1.9357	1.9404
16/16	DeepONet	26	1.5767	1.4777	1.6245	1.6311	1.5986	1.5469
16/16	DeepONet	27	2.3769	2.4261	2.2873	2.3238	2.3914	2.4519
16/16	Adapted SHRED	25	2.1781	2.2093	2.2204	2.2009	2.1829	2.0731
16/16	Adapted SHRED	26	1.5121	1.4793	1.4744	1.4874	1.5595	1.5578
16/16	Adapted SHRED	27	1.4920	1.3041	1.3476	1.4546	1.6319	1.6855
16/16	Direct-data	25	1.9107	1.1307	1.4849	1.8966	2.2249	2.4997
16/16	Direct-data	26	1.5800	1.0053	1.2277	1.5401	1.8204	2.0723
16/16	Direct-data	27	1.2564	0.9433	0.9976	1.1592	1.3877	1.6570
16/16	No-PDE	25	1.2793	0.9237	1.0952	1.3224	1.4924	1.4702
16/16	No-PDE	26	1.1828	0.8440	1.0290	1.2438	1.3807	1.3331
16/16	No-PDE	27	1.1646	0.8027	1.0127	1.2366	1.3655	1.3134
16/16	Direct-physics	25	1.3070	0.9549	1.0624	1.2445	1.4669	1.6761
16/16	Direct-physics	26	1.2416	0.9890	1.0967	1.2352	1.3966	1.4343
16/16	Direct-physics	27	1.3448	1.0092	1.1100	1.2803	1.4967	1.7091
16/16	EgoCast	25	1.0966	0.7679	0.8882	1.0879	1.3023	1.3269
16/16	EgoCast	26	1.0191	0.7220	0.8408	1.0248	1.2070	1.2098
16/16	EgoCast	27	1.0559	0.7257	0.8561	1.0614	1.2627	1.2650
8/8	DeepONet	25	1.1881	0.9709	1.0941	1.2006	1.3144	1.3243
8/8	DeepONet	26	1.3258	1.1416	1.2968	1.3517	1.4105	1.4107
8/8	DeepONet	27	1.2690	1.0663	1.2071	1.2788	1.3727	1.3937
8/8	Adapted SHRED	25	1.1639	0.9498	1.0071	1.1128	1.2951	1.3949
8/8	Adapted SHRED	26	1.1437	0.9726	1.0792	1.1318	1.2289	1.2813
8/8	Adapted SHRED	27	1.1759	0.9354	1.0489	1.1859	1.3100	1.3495
8/8	Direct-data	25	1.2506	1.0838	1.1333	1.2217	1.3576	1.4247
8/8	Direct-data	26	1.3805	1.1903	1.2757	1.3709	1.5019	1.5346
8/8	Direct-data	27	1.1835	1.0245	1.0731	1.1460	1.2738	1.3677
8/8	No-PDE	25	1.0139	0.6547	0.8112	1.0181	1.2185	1.2403
8/8	No-PDE	26	1.0125	0.6357	0.8021	1.0211	1.2271	1.2394
8/8	No-PDE	27	1.0122	0.6413	0.8064	1.0200	1.2217	1.2386
8/8	Direct-physics	25	1.2648	0.8339	0.9578	1.1573	1.4263	1.7386
8/8	Direct-physics	26	1.1200	0.7924	0.8895	1.0044	1.2115	1.5438
8/8	Direct-physics	27	1.2851	0.8986	1.0587	1.2504	1.4805	1.6070
8/8	EgoCast	25	0.9497	0.6008	0.7477	0.9500	1.1543	1.1664
8/8	EgoCast	26	0.9603	0.5910	0.7422	0.9549	1.1721	1.1965
8/8	EgoCast	27	0.9716	0.5976	0.7534	0.9694	1.1856	1.2066

0.05 m grid; RMSEs use all common valid slice points. On this slice, vector RMSE decreases from 0.942 to $0.844\,\mathrm{m\,s^{-1}}$ at 0.4 s and from 1.126 to $1.036\,\mathrm{m\,s^{-1}}$ at 2.0 s. Circles and diamonds project historical observations and held-motion future query sites, respectively. Dark silhouettes show the recorded future UAV geometry in projection; light gray marks invalid CFD points or body/source exclusions. The star marks the issue-time position; solid, dashed and dotted lines show the recorded past, constant-velocity held-motion future, and recorded future centre paths, respectively.

Models and paired training. The residual forecasters use the explicit Ego-Transport reference in Equation (2). Both learned arms retain motion-aligned histories, a zero-delay correction, a gated future evolution increment, and held-motion target geometry.

Persistence assigns the latest nearest valid observed velocity to each query and repeats it across forecast delays. **Ego-Transport** constructs a frozen field from latest-frame observations by inverse-distance-squared interpolation, then transports queries and rotates velocity components using held motion.

The **data-only model (No-PDE)** is a supervised forecaster conditioned on sparse velocity–pressure observations and vehicle-motion histories. It learns a zero-delay correction and a gated future evolution increment from current and future velocity targets, with no PDE residual loss. **EgoCast** uses the same inputs, forward architecture, and supervised objective, adding the physical regularizer in Equation (7). Future labels are identical in both arms: EgoCast additionally reuses their velocities in the spatial stencils.

Each pair shares initialization, labelled-window sampling, input and target locations, optimizer, and training duration. We use seeds 20260825, 20260826, and 20260827, keeping the campaign split and point layouts fixed. Both arms run for 24 epochs, with 256 sampled labelled examples per epoch and batch size 8, giving 768 optimizer updates. AdamW uses learning rate 3×10^{-4} and weight

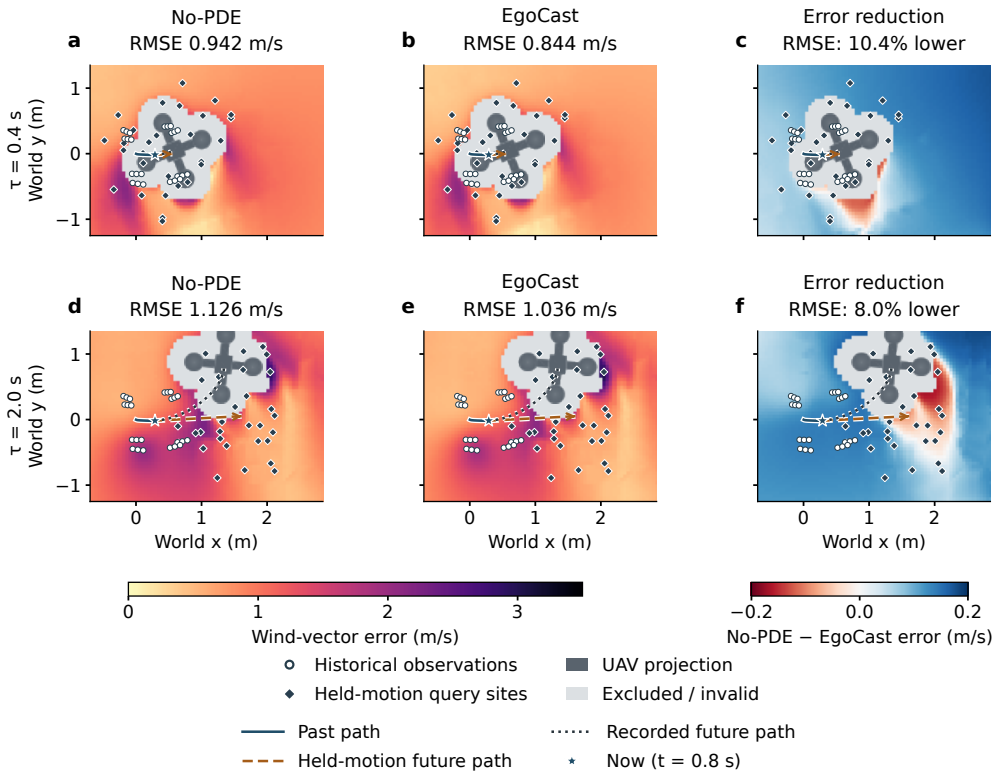

Figure 7: **Spatial errors in future wind prediction.** Prediction errors and their differences at 0.4 and 2.0 s for the $M = S = 8$ qualitative case. Differences are No-PDE minus EgoCast; blue indicates lower EgoCast error. The same fixed test window is shown at both delays.

decay 10^{-5}. The reported model minimizes the validation data loss, combining normalized current and future velocity errors on the 30 validation queries.

EgoCast uses $C = 64$ nominal centers: eight shifted label-neighbor centers and 56 independent locations. The stencil spacing is $h = 0.125$ m. The continuity and momentum weights are both 20, with an additional factor of 10 on stencil momentum. The independent and stencil branch weights are $7/8$ and $1/8$. The fixed-mixture protocol retains these weights independently of S (Appendix A.6). All physics residuals use a Huber threshold of 0.05 in normalized units; gradient-norm balancing is disabled. Appendix A.6 specifies the complete weighting.

Evaluation metrics. Wind-vector RMSE is defined as

$$E_u = \sqrt{\frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \|\hat{\boldsymbol{u}}_i - \boldsymbol{u}_{i,\mathrm{CFD}}\|_2^2}, \tag{36}$$

where $\mathcal{I}$ indexes valid test vectors. Squared errors are pooled before taking the square root, separately for each forecast delay and across all five delays within each training run. Secondary metrics are speed-magnitude RMSE and angular RMSE. Angular errors require predicted and target speeds above $10^{-6}\,\mathrm{m\,s^{-1}}$.

Statistical reporting. The completed repeated forecasting comparisons report mean ± sample standard deviation (SD) of the three seed-level metrics. Paired gains are computed as $100(E_{\mathrm{noPDE}} - E_{\mathrm{EgoCast}})/E_{\mathrm{noPDE}}$ within each seed and then averaged for the distance analysis and one-factor budget tables. Section 4.1 and the budget heatmap instead compute relative reductions from the mean RMSEs across seeds. Queries and overlapping windows are subsamples, not independent replicates. Comparisons are descriptive; no significance tests are performed. These seeds measure training variability on the same held-out inflow campaign, not variability across independent environments.

B.3 Observation and supervision budget scans

The joint matrix in Figure 6 evaluates all 25 combinations of $M, S \in \{16, 33, 66, 133, 266\}$. The supplementary one-factor scans vary either M or S over $\{8, 16, 33, 66, 133, 266\}$ while fixing the other at eight. Both protocols use seeds 20260825–20260827 and the training and evaluation settings described below. The $M = S = 16$ primary comparison, joint matrix, component ablations, and distance analysis use space-filling observation layouts during training and validation, followed by nested observation layouts at evaluation. The $M = S = 8$ primary comparison and the one-factor scans use nested sensing layouts throughout. At $M = 16$, six sites are shared between the training and evaluation layouts. Each method pair receives identical inputs within its comparison. The joint scan therefore varies sensing count and coverage under a change in observation layout.

Nested layouts preserve the original eight sites in their original order, then add sites by farthest-point selection from geometry-only candidates inside the original spatial envelope, excluding the body and rotor regions. Additional candidates are drawn from a deterministic Halton sequence. Observation offsets exclude fixed supervision and readout offsets; supervision offsets exclude fixed observation and readout offsets. These exclusions concern the body-frame layouts. The scans vary spatial coverage as well as count. Figure 8 shows the resulting layouts at four budgets.

Each model uses the same 54 labelled training windows, with 24 epochs and 768 optimizer updates. Model selection uses validation data loss on the same fixed set of nine validation windows. Added supervision sites use velocity labels sampled from saved CFD fields at the training query coordinates; the original eight labels and masks are retained exactly. Pressure and closure supervision remain disabled. Unavailable samples retain their validity masks. Evaluation uses the same 3,884 valid vectors as Table 1.

Table 10 reports both one-factor scans; gains are computed within each seed before averaging.

Table 10: Complete budget scans. Vector RMSE ($\mathrm{m\,s^{-1}}$) is mean $\pm$ sample SD over three seeds. Positive paired gain favors EgoCast; negative values indicate increased error.

Budget	No-PDE	EgoCast	Paired gain (%)
Observation sweep ($S = 8$)			
$M = 8$	1.0128 ± 0.0009	0.9605 ± 0.0109	5.16
$M = 16$	1.0257 ± 0.0125	0.9724 ± 0.0141	5.20
$M = 33$	0.9947 ± 0.0034	0.9466 ± 0.0072	4.83
$M = 66$	1.0147 ± 0.0191	0.9673 ± 0.0176	4.65
$M = 133$	0.9941 ± 0.0169	0.9291 ± 0.0106	6.53
$M = 266$	0.9398 ± 0.0063	0.9251 ± 0.0045	1.56
Supervision sweep ($M = 8$)			
$S = 8$	1.0128 ± 0.0009	0.9605 ± 0.0109	5.16
$S = 16$	0.9960 ± 0.0102	0.9737 ± 0.0150	2.23
$S = 33$	0.9827 ± 0.0081	1.0102 ± 0.0034	-2.81
$S = 66$	0.9900 ± 0.0016	1.0176 ± 0.0044	-2.79
$S = 133$	0.9933 ± 0.0055	0.9960 ± 0.0056	-0.28
$S = 266$	1.0023 ± 0.0049	0.9854 ± 0.0030	1.68

The requested physics budget remains 64 centers per update: 56 independent centers and eight label-anchored stencils, with the same loss weights. For $S > 8$, the eight anchor indices at update k are $(8k + j) \bmod S$, $j = 0, \ldots, 7$. Selection counts differ by at most one across the 768 updates; stencil validity is checked as usual. The $M = S = 8$ models are shared by both scans.

Table 11 reports all 25 combinations in Figure 6. Its gains use ratios of mean RMSEs, matching the main-text matrix.

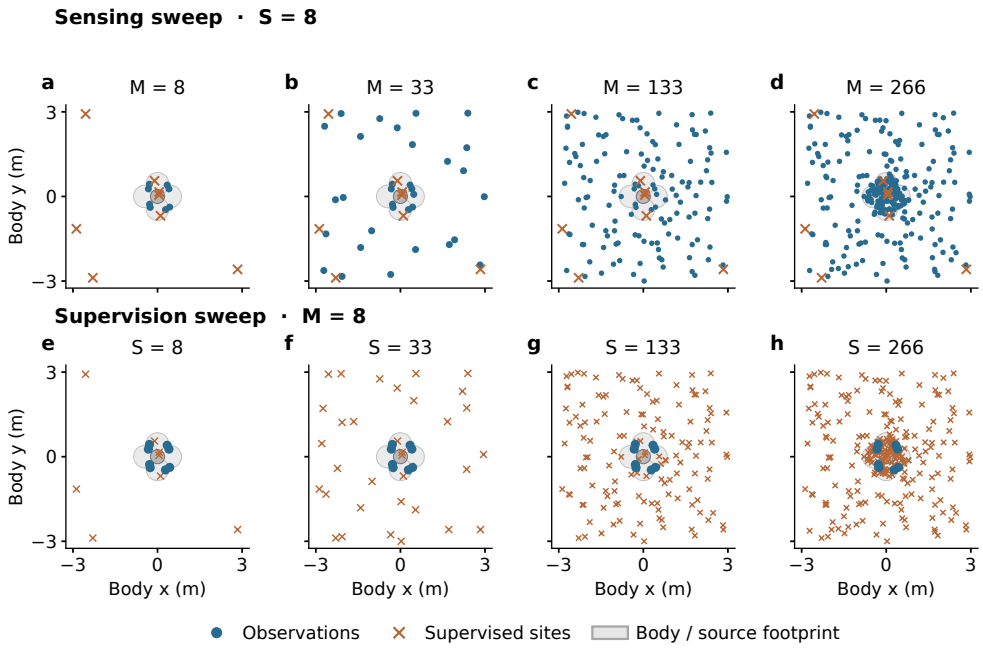

Figure 8: **Nested observation and supervision layouts.** Body-frame x–y projections at four budgets from the six-budget scans. The observation sweep (a–d) fixes $S = 8$; the supervision sweep (e–h) fixes $M = 8$. Every point in each displayed budget is retained, and larger layouts preserve the original eight sites. Blue circles denote observations and orange crosses denote supervised sites. Gray shapes are projected body/source exclusion footprints; points with different heights can overlap these footprints in projection. All panels use the same extent and scale.

Table 11: Complete joint budget matrix. Vector RMSE ($\mathrm{m\,s^{-1}}$) is mean $\pm$ sample SD over three paired runs. Gain is $100(1 - \overline{E}_{\mathrm{EgoCast}}/\overline{E}_{\mathrm{No-PDE}})$.

M	S	No-PDE	EgoCast	Gain (%)
16	16	1.2089 ± 0.0616	1.0573 ± 0.0387	12.54
16	33	1.1985 ± 0.0689	1.0277 ± 0.0437	14.25
16	66	1.0669 ± 0.0439	1.0046 ± 0.0561	5.84
16	133	1.1353 ± 0.0587	0.9889 ± 0.0145	12.90
16	266	1.1685 ± 0.0401	1.0572 ± 0.0115	9.53
33	16	1.0473 ± 0.0507	0.9403 ± 0.0211	10.22
33	33	1.0187 ± 0.0030	0.9567 ± 0.0111	6.09
33	66	1.0105 ± 0.0250	0.9614 ± 0.0082	4.86
33	133	1.0038 ± 0.0231	0.9551 ± 0.0018	4.85
33	266	0.9900 ± 0.0132	0.9654 ± 0.0020	2.49
66	16	1.0476 ± 0.0334	0.9237 ± 0.0090	11.82
66	33	1.0508 ± 0.0525	0.9523 ± 0.0054	9.37
66	66	1.0099 ± 0.0248	0.9477 ± 0.0115	6.16
66	133	1.0072 ± 0.0127	0.9409 ± 0.0067	6.58
66	266	1.0074 ± 0.0654	0.9571 ± 0.0074	5.00
133	16	1.0455 ± 0.0369	0.9330 ± 0.0220	10.77
133	33	1.0074 ± 0.0121	0.9779 ± 0.0162	2.92
133	66	0.9970 ± 0.0052	0.9918 ± 0.0034	0.52
133	133	0.9783 ± 0.0130	0.9798 ± 0.0155	-0.15
133	266	0.9258 ± 0.0118	0.9528 ± 0.0091	-2.92
266	16	0.9485 ± 0.0103	0.9439 ± 0.0087	0.49
266	33	0.9563 ± 0.0093	0.9592 ± 0.0072	-0.30
266	66	0.9579 ± 0.0053	0.9448 ± 0.0235	1.37
266	133	0.9716 ± 0.0430	0.9627 ± 0.0082	0.92
266	266	0.9589 ± 0.0313	0.9740 ± 0.0062	-1.57

Figure 9 reproduces the two budget slices from Figure 6 at a larger scale.

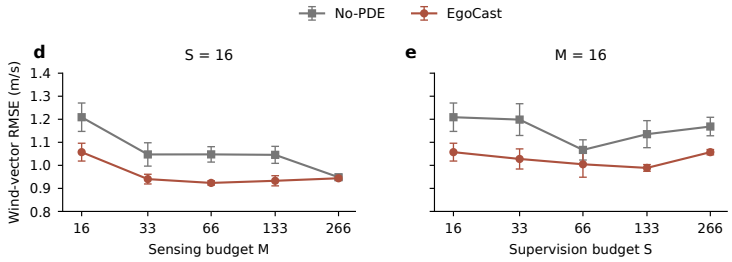

Figure 9: **Slices of the joint budget matrix.** Vector RMSE at $S = 16$ **(d)** and $M = 16$ **(e)**: mean $\pm$ sample SD over three paired runs. Training and evaluation use different sensing layouts, as in Figure 6.

B.4 Matched physics ablations

The component ablations retain the joint-matrix layout protocol (Appendix B.3), three seeds, 54 labelled windows, 24 epochs, 768 updates, validation-data selection, and 56 independent plus eight stencil centers. The frame-motion variant zeroes translation and angular velocity in both physical-residual paths; forward prediction, query geometry, and masks retain actual motion. The all-predicted variant disables velocity-label insertion while retaining stencil centers, spacing, validity masks, and weights. Audits verify 768 evaluations per residual path and identical labelled sampling schedules. Label insertion is zero in the all-predicted arm and positive in the frame-motion arm. The No-PDE and EgoCast baselines are archived joint-matrix checkpoints; Table 1 reports separately trained models under the same layout protocol. Thus the EgoCast means of 1.0573 here and 1.0572 in Table 1 are results from distinct training records, not two roundings of one result. Comparisons within each table use its own matched model set.

Table 12: Per-seed vector RMSE ($\mathrm{m\,s^{-1}}$) for the matched held-motion ablations under the joint-matrix layout protocol. Seed columns abbreviate 20260825–20260827.

Method	25	26	27
No-PDE	1.2793	1.1828	1.1646
EgoCast	1.0966	1.0192	1.0561
Without frame-motion terms	1.0794	1.0670	0.9877
All-predicted neighbors	1.1199	1.0431	1.1315

Removing the frame-motion terms changes the mean RMSE from 1.0573 to $1.0447\,\mathrm{m\,s^{-1}}$, with mixed seed-level differences: removal improves two seeds and worsens one. This ablation does not establish an accuracy gain from the frame-motion terms. They are retained to evaluate the physical residual in the stated moving coordinates.

B.5 Distance from historical observations

For a future control query at world position $\boldsymbol{x}^W$, we use

$$d_{\mathrm{obs}}(\boldsymbol{x}^W; \mathcal{H}_t) = \min_{\boldsymbol{y} \in \mathcal{O}_t^W} \|\boldsymbol{x}^W - \boldsymbol{y}\|_2,$$

where $\mathcal{O}_t^W$ contains all valid observation positions in the three-frame history, expressed in world coordinates. This measures separation from input observation locations, not supervised sites. We evaluate the three paired joint-matrix $M = S = 16$ models under the layout change described above, using the 3,884 valid held-motion queries (Figure 5). Distances use all valid observations from the sixteen-site, three-frame evaluation history. Four distance-quantile bins per delay are fixed before reading prediction errors. Bins with fewer than 100 vectors or five windows would be merged with an adjacent bin; none required merging. All models share these bins and masks, and the last interval includes its upper edge. Table 13 reports every bin; assignment uses full-precision boundaries.

Table 13: Distance-stratified vector RMSE for $M = S = 16$ (mean $\pm$ sample SD, three seeds) and mean paired reduction. N counts valid vectors; W counts distinct windows. Overlapping windows are not independent replicates.

Delay (s)	Distance (m)	No-PDE	EgoCast	Gain (%)	N	W
0.4	0.050–0.302	0.9195 ± 0.0586	0.8424 ± 0.0201	8.16	195	26
0.4	0.302–0.390	0.6803 ± 0.0501	0.6031 ± 0.0333	11.23	195	26
0.4	0.390–0.552	0.8592 ± 0.0675	0.7247 ± 0.0345	15.51	195	26
0.4	0.552–0.955	0.9433 ± 0.0697	0.7633 ± 0.0283	18.94	195	26
0.8	0.027–0.349	1.0815 ± 0.0506	0.9979 ± 0.0340	7.63	195	26
0.8	0.349–0.467	0.8669 ± 0.0455	0.7352 ± 0.0303	15.12	195	26
0.8	0.467–0.680	1.0813 ± 0.0495	0.8176 ± 0.0327	24.37	195	26
0.8	0.680–1.219	1.1326 ± 0.0325	0.8745 ± 0.0213	22.78	195	26
1.2	0.030–0.398	1.2935 ± 0.0549	1.2578 ± 0.0574	2.71	194	26
1.2	0.398–0.638	1.1706 ± 0.0453	0.9959 ± 0.0329	14.90	193	26
1.2	0.638–0.901	1.2103 ± 0.0473	0.9198 ± 0.0227	23.97	193	26
1.2	0.901–1.464	1.3845 ± 0.0442	1.0275 ± 0.0195	25.77	194	26
1.6	0.086–0.507	1.5342 ± 0.0731	1.6371 ± 0.0880	-6.79	194	26
1.6	0.507–0.832	1.2943 ± 0.0472	1.1770 ± 0.0444	9.06	193	26
1.6	0.832–1.182	1.3085 ± 0.0661	1.0092 ± 0.0266	22.81	194	26
1.6	1.182–1.682	1.4971 ± 0.0898	1.1128 ± 0.0229	25.55	194	25
2.0	0.096–0.664	1.3719 ± 0.0648	1.4868 ± 0.0963	-8.42	194	26
2.0	0.664–1.046	1.2870 ± 0.0671	1.2839 ± 0.0578	0.17	193	26
2.0	1.046–1.447	1.3327 ± 0.0791	1.1202 ± 0.0462	15.89	194	26
2.0	1.447–1.932	1.4883 ± 0.1296	1.1446 ± 0.0262	22.81	194	25

The label-anchoring comparison in Table 3 uses these same queries and distance bins. Table 14 reports all bins for the matched physical-supervision variants.

Table 14: Label-anchoring results by distance to historical observations ($M = S = 16$). Vector RMSE (m s^{-1}) and paired relative reduction (%) are mean $\pm$ sample SD over three runs. Positive reduction favors EgoCast. Counts and bin boundaries match Table 13.

Delay (s)	Distance (m)	All-predicted	EgoCast	Reduction (%)
0.4	0.050–0.302	0.8737 ± 0.0455	0.8424 ± 0.0201	3.48 ± 3.06
0.4	0.302–0.390	0.6624 ± 0.0717	0.6031 ± 0.0333	8.60 ± 4.72
0.4	0.390–0.552	0.8036 ± 0.0586	0.7247 ± 0.0345	9.69 ± 2.99
0.4	0.552–0.955	0.8830 ± 0.0488	0.7633 ± 0.0283	13.47 ± 3.22
0.8	0.027–0.349	1.0018 ± 0.0413	0.9979 ± 0.0340	0.37 ± 1.10
0.8	0.349–0.467	0.7481 ± 0.0511	0.7352 ± 0.0303	1.60 ± 2.71
0.8	0.467–0.680	0.9059 ± 0.0532	0.8176 ± 0.0327	9.63 ± 4.32
0.8	0.680–1.219	0.9891 ± 0.0466	0.8745 ± 0.0213	11.49 ± 3.61
1.2	0.030–0.398	1.2112 ± 0.0488	1.2578 ± 0.0574	-3.87 ± 3.83
1.2	0.398–0.638	1.0183 ± 0.0500	0.9959 ± 0.0329	2.12 ± 2.67
1.2	0.638–0.901	0.9987 ± 0.0540	0.9198 ± 0.0227	7.75 ± 4.34
1.2	0.901–1.464	1.1466 ± 0.0587	1.0275 ± 0.0195	10.26 ± 4.18
1.6	0.086–0.507	1.5405 ± 0.0518	1.6371 ± 0.0880	-6.31 ± 5.88
1.6	0.507–0.832	1.1946 ± 0.0501	1.1770 ± 0.0444	1.41 ± 3.91
1.6	0.832–1.182	1.1049 ± 0.0696	1.0092 ± 0.0266	8.46 ± 5.14
1.6	1.182–1.682	1.2370 ± 0.0675	1.1128 ± 0.0229	9.90 ± 4.08
2.0	0.096–0.664	1.3969 ± 0.0404	1.4868 ± 0.0963	-6.43 ± 6.09
2.0	0.664–1.046	1.2977 ± 0.0607	1.2839 ± 0.0578	0.98 ± 5.04
2.0	1.046–1.447	1.2174 ± 0.0800	1.1202 ± 0.0462	7.79 ± 5.62
2.0	1.447–1.932	1.2850 ± 0.0764	1.1446 ± 0.0262	10.76 ± 4.35

C Wind Preview and Closed-Loop Control

The tracking comparisons in Section 4.3 use paired models at $M = 16$, $S = 33$ and $M = S = 8$ (seed 20260826), under the 300° inflow condition. Each model runs in its own two-way-coupled OpenFOAM–preCICE–MuJoCo simulation, with CFD supplying body and rotor loads. Three 5 s reference trajectories combine horizontal motion with an arc climb, an arc descent, or a spatial S-shaped maneuver; a fourth trajectory lasts 20 s. The paired runs share the reference trajectory, controller, and aerodynamic mapping. The Ego-Transport control baseline uses the same current observation sites as the learned models at each budget, along with the same 30 force-query sites, aerodynamic mapping, and controller settings. It evaluates the frozen interpolant with held-motion

compensation, with the learned correction disabled. All four trajectories are evaluated; this deterministic baseline requires no training seed.

At each forecast issue time t, the history and held-motion frame are fixed. Thirty exterior offsets $\boldsymbol{a}_\ell$ define the wind queries. The controller uses 25 planning intervals of 0.04 s and one optimization pass per update. For a general world-space query, the same coordinate conversion is

$$\boldsymbol{q}_{j\ell} = R_{\tau_j}^\top [R_t^\top (\boldsymbol{x}_{j\ell}^W - \boldsymbol{p}_t^W) - \boldsymbol{V}_t \tau_j]. \tag{37}$$

This inverts Equation (12). Control queries in these experiments follow measured held motion: $\boldsymbol{x}_{j\ell}^W = \boldsymbol{p}_t^W + \boldsymbol{v}_t^W \tau_j + Q_j \boldsymbol{a}_\ell$, where $Q_j = R_t R_{\tau_j}$ and $\tau_j \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ s. Thus $\boldsymbol{q}_{j\ell} = \boldsymbol{a}_\ell$ in the held-motion chart. The wind and force previews remain fixed during each MPC solve. The predicted absolute velocity in world coordinates is

$$\hat{\boldsymbol{u}}_{j\ell}^W = R_t R_{\tau_j} F_\theta^u(\mathcal{H}_t^{(M)}, \boldsymbol{q}_{j\ell}, \tau_j). \tag{38}$$

Calibrated three-dimensional aerodynamic mapping. The mapping estimates the body-surface force (F_x^W, F_y^W, F_z^W). The 30 query sites are partitioned into four body-azimuth sectors, centered on the $+x, +y, -x, -y$ axes, containing 8, 8, 7, and 7 sites. Let $H_j \in \mathbb{R}^{3\times 3}$ denote the yaw rotation of Q_j, with the vertical axis unchanged. The mapping holds world translational velocity $\boldsymbol{v}_t^W = R_t \boldsymbol{V}_t$ fixed and expresses the held angular velocity in the extrapolated basis, $\boldsymbol{\omega}_j = Q_j^\top R_t \boldsymbol{\omega}_t$. Its three-dimensional relative-flow input at offset $\boldsymbol{a}_\ell$ is

$$\boldsymbol{r}_{j\ell} = H_j^\top \left[\hat{\boldsymbol{u}}_{j\ell}^W - \boldsymbol{v}_t^W - Q_j(\boldsymbol{\omega}_j \times \boldsymbol{a}_\ell)\right]. \tag{39}$$

For each group G_g, form

$$\begin{aligned} \boldsymbol{\phi}_{jg} &= \frac{1}{|G_g|} \sum_{\ell \in G_g} \begin{bmatrix} r_x & r_y & r_z & |r_x| r_x & |r_y| r_y & |r_z| r_z & r_x r_y & r_x r_z & r_y r_z \end{bmatrix}^\top, \\ \hat{\boldsymbol{F}}_j^W &= H_j \left[W^\top \frac{\boldsymbol{\phi}_j - \boldsymbol{\mu}}{\boldsymbol{\sigma}} + \boldsymbol{b} \right], \end{aligned} \tag{40}$$

where $\boldsymbol{\phi}_j$ concatenates the four groups into 36 features, division is componentwise, and $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ are training feature means and standard deviations. Scales at most 10^{-10} are replaced by one. The map has $W \in \mathbb{R}^{36\times 3}$ and $\boldsymbol{b} \in \mathbb{R}^3$, fitted by mean squared error in heading coordinates with ridge penalty $0.01\|W\|_F^2$ and an unpenalized intercept.

Calibration uses simultaneous CFD velocities, measured kinematics, and unclipped body-surface forces from four circular training conditions (0°, 60°, 120°, 180°). Each condition supplies 26 snapshots, giving 104 calibration samples used to fit a time-independent map that is frozen for both forecasters. These force-calibration labels are separate from the neural $S = 33$ supervision budget. Its output supplies three-dimensional body-force feedforward; rotor loads remain part of the CFD actuator dynamics.

The model is trained at delays $\{0.4, 0.8, 1.2, 1.6, 2.0\}$ s. For control, its continuous forecast is evaluated directly at $\{0.2, 0.4, 0.6, 0.8, 1.0\}$ s using the available history. The 0.2 s query is outside the supervised delay range; 0.6 and 1.0 s are between supervised delays. The 0.4–2.0 s forecast tables therefore do not separately establish accuracy at every control knot. The pooled 0.2–1.0 s force-preview results and closed-loop trajectories provide end-to-end evaluations of this control-query schedule. A current-time force knot uses spatial interpolation of the latest observations. Forces are linearly interpolated between the current and future knots at MPC interval starts. The final force is held for at most 0.1 s between forecast refreshes.

MPC dynamics and optimization. The controller uses state $\boldsymbol{x} = (\boldsymbol{p}^W, \boldsymbol{v}^W, \phi, \theta, \psi, \boldsymbol{\omega}^B) \in \mathbb{R}^{12}$ and four rotor-command increments $\delta \boldsymbol{f} = \boldsymbol{f} - (mg/4)\mathbf{1}$. At each solve, a small-angle model is linearized at the measured yaw ψ_t, then discretized by forward Euler as in Equation (9). The continuous model uses $\dot{\boldsymbol{p}} = \boldsymbol{v}$, $(\dot{\phi}, \dot{\theta}, \dot{\psi}) = \boldsymbol{\omega}^B$, $\dot{v}_x = g(\sin\psi_t\, \phi + \cos\psi_t\, \theta)$, $\dot{v}_y = g(-\cos\psi_t\, \phi + \sin\psi_t\, \theta)$, $\dot{v}_z = K_{1,:}\delta \boldsymbol{f}/m$, and $\dot{\boldsymbol{\omega}}^B = J^{-1} K_{2:4,:} \delta \boldsymbol{f}$. Here $m = 1$ kg, $J = \mathrm{diag}(0.02, 0.02, 0.04)$ kg m^2, $g = 9.81\,\mathrm{m\,s^{-2}}$, and the fixed CFD control mixer is

$$K = \begin{bmatrix} 0.882851 & 0.999371 & 0.899106 & 1.077745 \\ 0.007544 & 0.443111 & 0.003430 & -0.442330 \\ -0.411435 & -0.006233 & 0.383758 & 0.001996 \\ -0.027112 & 0.033834 & -0.028819 & 0.026354 \end{bmatrix}.$$

Its first row maps command increments to collective force and its remaining rows to body moments. For $N = 25$ and $\Delta t_{\text{plan}} = 0.04$ s, the controller solves

$$\begin{aligned} \min_{\delta \boldsymbol{f}_{0:N-1}} \quad & \sum_{k=0}^{N-1} \left(\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_{k+1}^{\text{ref}}\|_{Q_k}^2 + \|\delta \boldsymbol{f}_k\|_R^2 \right), \\ \text{subject to} \quad & \text{Equation (9)}, \quad 0 \le f_{k,i} \le 5.0 \text{ N}, \quad \sum_{i=1}^{4} f_{k,i} \le 20.0 \text{ N}, \end{aligned} \tag{41}$$

where $Q_k = Q$ except $Q_{N-1} = 20Q$, $Q = \text{diag}(20, 20, 30, 5, 5, 10, 2, 2, 1, 0.5, 0.5, 0.2)$, and $R = 0.05 I_4$. Reference positions and velocities sample the prescribed trajectory at prediction endpoints; reference roll, pitch, yaw and angular rates are zero in this fixed-heading configuration. OSQP solves the condensed 100-variable problem with absolute and relative tolerances 10^{-5}, at most 2,000 iterations, and the previous solution as a warm start. The first command is applied every 0.004 s; forecasting refreshes every 0.1 s after the three-frame history becomes available. The disturbance observer is disabled for the EgoCast and data-only preview comparisons; the observer variants are described below. Commands are converted to rotor RPM with a square-law map (900 RPM at hover, capped at 1,700 RPM); the configured motor lag and slew limit are zero. In the CFD rotor model, blade-element relative inflow subtracts the moving rotor's translational and angular velocity from the local fluid velocity.

Paired force-preview evaluation. At both $M = 16$, $S = 33$ and $M = S = 8$, the learned forecasters replay the arrived observation histories from both recorded learned-controller runs for each of the four trajectories. Ego-Transport is additionally evaluated on the same histories at $M = S = 8$. Within each replay, all methods receive identical observations, measured kinematics, query positions, and the frozen aerodynamic mapping. The five future force knots at 0.2–1.0 s are compared with the recorded CFD body-surface loads at $t + \tau + 0.004$ s, accounting for the coupling-step timestamp offset. The current-time knot is excluded from this forecast metric. Only issue times with a complete future interval are retained: 38 per recorded 5 s run and 188 per recorded 20 s run. Squared three-dimensional force errors are pooled over both recorded runs and all five delays before taking the square root. This gives 380 force vectors per short trajectory and 1,880 for the 20 s trajectory, with exactly the same targets for all compared methods. These are force-preview diagnostics on recorded histories; tracking errors come from the separate closed-loop simulations. The $M = S = 8$ replay exactly reproduces each method's recorded online force previews on its own histories. Including the additional Ego-Transport rollout histories preserves the method ranking on all four trajectories.

Complete closed-loop comparisons. Table 15 reports both budgets in Section 4.3; each uses one trained predictor pair from seed 20260826, not an average over the three forecasting seeds. These trajectory results describe the tested runs and do not estimate training-seed variability in closed-loop performance. The Pure MPC reference is shared across budgets. Figure 10 summarizes the corresponding force-preview and tracking comparisons.

Table 15: Full-trajectory position RMSE (m) for both control budgets.

$M = 16,\ S = 33$				
Trajectory	Pure MPC	No-PDE	Ego-Transport	EgoCast
Arc climb (5 s)	0.2272	0.1475	0.0790	0.0850
Arc descent (5 s)	0.2001	0.1032	0.0549	0.0599
Spatial S (5 s)	0.2229	0.1268	0.0690	0.0735
Extended (20 s)	0.2290	0.1283	0.0985	0.0969

$M = S = 8$					
Trajectory	Pure MPC	No-PDE	Ego-Transport	Adapted SHRED	EgoCast
Arc climb (5 s)	0.2272	0.0448	0.0885	0.0666	0.0442
Arc descent (5 s)	0.2001	0.0464	0.0614	0.0883	0.0466
Spatial S (5 s)	0.2229	0.0506	0.0764	0.0762	0.0505
Extended (20 s)	0.2290	0.1008	0.1021	0.0987	0.1011

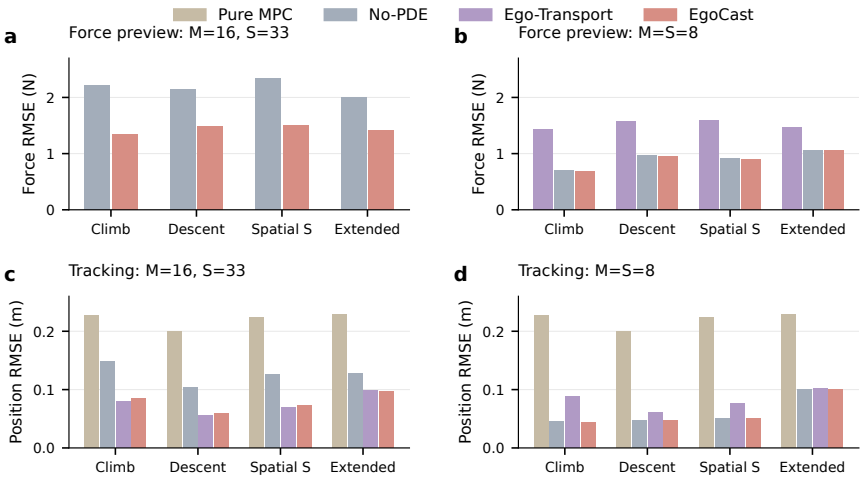

Figure 10: **Preview control across budgets. (a,b)** Shared-history force RMSE over 0.2–1.0 s delays at $M = 16$, $S = 33$ and $M = S = 8$, respectively. **(c,d)** Full-trajectory position RMSE at $M = 16$, $S = 33$ and $M = S = 8$, with matched scales. One trained pair per budget; Pure MPC is shared.

Table 16: Shared-history force-preview RMSE (N) for both budgets. Methods are grouped by the comparison performed at each budget. Values are rounded for display; percentage reductions use unrounded RMSE values.

Budget	Method	Arc climb	Arc descent	Spatial S	Extended
16/33	No-PDE	2.214	2.136	2.334	2.005
	EgoCast	1.343	1.493	1.508	1.423
8/8	Ego-Transport	1.433	1.579	1.589	1.477
	No-PDE	0.711	0.972	0.917	1.061
	EgoCast	0.691	0.956	0.894	1.052

Tracking components. Table 17 separates horizontal and vertical tracking errors and reports the maximum Euclidean position error. All metrics use the complete trajectories reported in Table 15. Figure 11 shows the corresponding 20 s position-error traces at $M = 16$, $S = 33$.

Table 17: Tracking components for the paired $M = 16$, $S = 33$ closed-loop runs. All errors are in metres.

	Horizontal RMSE		Vertical RMSE		Maximum error	
Trajectory	No-PDE	EgoCast	No-PDE	EgoCast	No-PDE	EgoCast
Arc climb	0.1438	0.0814	0.0331	0.0246	0.2126	0.1287
Arc descent	0.0919	0.0439	0.0470	0.0408	0.1303	0.0859
Spatial S	0.1236	0.0707	0.0283	0.0201	0.2048	0.1147
Extended trajectory	0.1164	0.0843	0.0540	0.0478	0.2230	0.1903

Combining wind preview with residual disturbance compensation. Across the four reference trajectories in Table 4, we compare EgoCast alone, DOB-MPC, and DOB+EgoCast using the same MPC settings and the $M = 16$, $S = 33$ predictor for both preview controllers. The observer update, filter constant, and zero initialization described below are shared across all four trajectories. The DOB estimates translational acceleration disturbances from measured velocity increments and the previous thrust command, with a 0.15 s exponential filter. For DOB+EgoCast, the previous EgoCast disturbance contribution is subtracted before filtering the residual; the resulting correction is added throughout the planning horizon while retaining the time-varying EgoCast preview. Angular residual

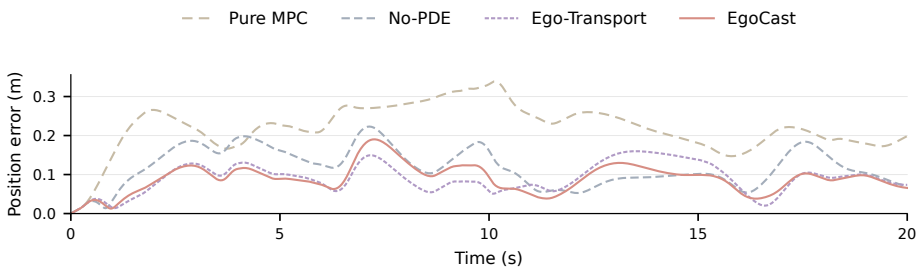

Figure 11: **Tracking over the 20 s trajectory** at $M = 16$, $S = 33$. Euclidean position error at every control step; one trained predictor pair.

estimation is disabled. Specifically, with control sampling interval $h = 0.004$ s, the update is

$$\begin{aligned} \boldsymbol{r}_t &= \frac{\boldsymbol{v}_t^W - \boldsymbol{v}_{t-h}^W}{h} - \boldsymbol{a}_{t-h}^{\mathrm{nom}} - \boldsymbol{a}_{t-h}^{\mathrm{EgoCast}}, \\ \boldsymbol{b}_t &= (1-\alpha)\boldsymbol{b}_{t-h} + \alpha \boldsymbol{r}_t, \qquad \alpha = 1 - \exp[-h/(0.15\,\mathrm{s})], \\ \hat{\boldsymbol{F}}_{t,j}^{\mathrm{DOB+EgoCast}} &= \hat{\boldsymbol{F}}_{t,j}^{\mathrm{EgoCast}} + m\boldsymbol{b}_t. \end{aligned} \tag{42}$$

Here $\boldsymbol{a}_{t-h}^{\mathrm{nom}}$ is the nominal acceleration from the previous state and issued rotor commands, and $\boldsymbol{a}_{t-h}^{\mathrm{EgoCast}}$ is the EgoCast acceleration contribution used for that control interval. The observer starts from $\boldsymbol{b}_0 = \boldsymbol{0}$. The same three-dimensional correction is held across the planning horizon and enters the velocity increments as $\Delta t_{\mathrm{plan}} \boldsymbol{b}_t$. DOB-MPC uses this update with the EgoCast terms set to zero.

Table 18: Full-trajectory position RMSE (m) for the disturbance-feedback comparison at $M = 16$, $S = 33$, corresponding to Table 4. Values are rounded for display; percentage reductions use unrounded RMSE values.

Trajectory	DOB-MPC	EgoCast	DOB+EgoCast
Arc climb (5 s)	0.0802	0.0850	0.0651
Arc descent (5 s)	0.0843	0.0599	0.0695
Spatial S (5 s)	0.0848	0.0735	0.0748
Extended (20 s)	0.0813	0.0969	0.0698

Figure 12 compares the four-trajectory tracking results.

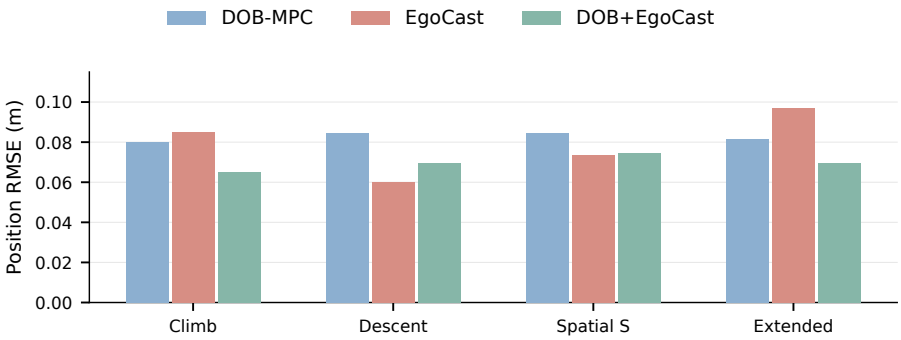

Figure 12: **Anticipatory preview complements reactive disturbance estimation.** Full-trajectory position RMSE for DOB-MPC, EgoCast, and DOB+EgoCast on the four trajectories in Table 4.

For the 20 s trajectory, Table 19 shows that the combined controller improves both horizontal and vertical tracking. Relative to EgoCast alone, horizontal RMSE falls by 17.6% and vertical RMSE by

86.5%. Its total position RMSE is also lower in each of the four consecutive 5 s intervals: 0.0675, 0.0739, 0.0616, and 0.0752 m, compared with 0.0859, 0.1218, 0.0932, and 0.0817 m for EgoCast alone. Thus, the benefit extends across the trajectory.

Table 19: Tracking on the 20 s trajectory with wind preview and disturbance observation. Each run contains 5,000 control steps; errors are in metres. EgoCast and DOB+EgoCast use the same $M = 16$, $S = 33$ predictor and aerodynamic mapping.

Controller	Total RMSE	Horizontal RMSE	Vertical RMSE	Maximum error
EgoCast	0.09692	0.08431	0.04779	0.1903
DOB-MPC	0.08134	0.08063	0.01075	0.1243
DOB+EgoCast	**0.06975**	**0.06945**	**0.00646**	**0.1097**

Closed-loop robustness to velocity noise. We evaluate DOB-MPC and DOB+EgoCast on all four trajectories with zero-mean Gaussian measurement noise at standard deviations of 0.05, 0.10, and $0.20\,\mathrm{m\,s^{-1}}$ (Table 20; Figure 13). Independent samples perturb each component of vehicle velocity at 250 Hz and observed wind velocity at 10 Hz, using separate random streams. Each controller pair shares the vehicle-velocity noise sequence. DOB+EgoCast reduces the mean position RMSE across the four trajectories by 15.01%, 14.17%, and 11.07% at the three noise levels, respectively.

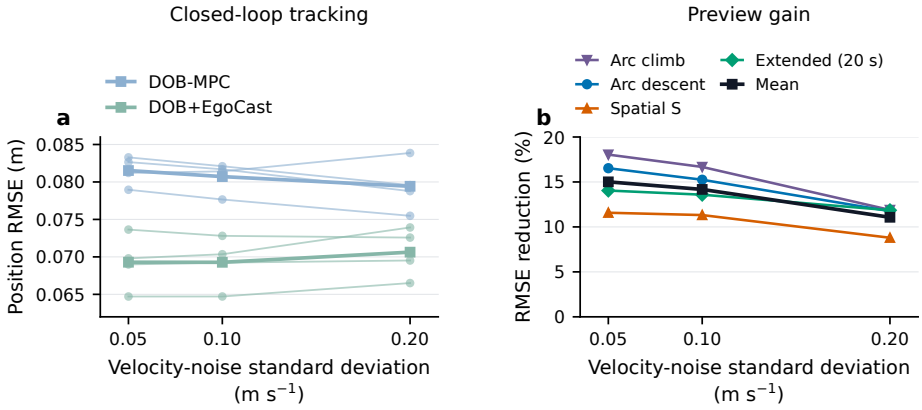

Figure 13: **Closed-loop robustness to velocity noise.** **(a)** Position RMSE for DOB-MPC and DOB+EgoCast across the three noise levels. Thin curves show the four trajectories; squares show their mean. **(b)** Relative RMSE reductions; the mean uses the ratio of mean RMSEs.

Table 20: Position RMSE (m) under velocity noise. Means average four trajectories. Reductions use unrounded RMSEs; mean reductions compare the two controller means.

Noise std. ($\mathrm{m\,s^{-1}}$)	Trajectory	DOB-MPC	DOB+EgoCast	Reduction (%)
0.05	Arc climb	0.0790	0.0647	18.04
	Arc descent	0.0827	0.0690	16.53
	Spatial S	0.0833	0.0736	11.58
	Extended (20 s)	0.0812	0.0698	14.05
	Mean	0.0815	0.0693	15.01
0.10	Arc climb	0.0777	0.0647	16.67
	Arc descent	0.0817	0.0692	15.25
	Spatial S	0.0821	0.0728	11.32
	Extended (20 s)	0.0814	0.0704	13.58
	Mean	0.0807	0.0693	14.17
0.20	Arc climb	0.0755	0.0665	11.89
	Arc descent	0.0788	0.0695	11.75
	Spatial S	0.0796	0.0726	8.79
	Extended (20 s)	0.0839	0.0739	11.85
	Mean	0.0794	0.0706	11.07

Computational cost. For a separate $M = S = 8$ profiling configuration, Table 21 measures the live forecasting code on an AMD Ryzen 7 7840H CPU under Linux, using PyTorch 2.5.1, batch size one and one PyTorch/BLAS thread. Each model replays the same 48 three-frame histories three times after 20 warm-up calls. Each refresh encodes 3×8 observations and queries $30 \times 5 = 150$ future points; the current IDW knot and aerodynamic mapping are included. Eight CFD MPI processes were active on the same machine during profiling.

Table 21: Measured forecast computation time in ms (median / 95th percentile; 144 refreshes per model). Query and mapping includes Ego-Transport and held-motion feature construction.

Stage	No-PDE	EgoCast
History alignment and normalization	0.63 / 3.46	0.72 / 3.90
History encoding	1.07 / 5.97	1.17 / 6.19
Query and aerodynamic mapping	4.50 / 21.57	5.26 / 22.37
Complete forecast refresh	6.25 / 30.95	7.12 / 30.97

For a separate single-solve 20 s profiling configuration, each simulation records 5,000 control updates and 198 forecast refreshes. The MPC control block, including problem assembly and solution, takes median times of 5.30 / 5.63 ms for No-PDE / EgoCast (95th percentiles 8.01 / 8.31 ms). On refresh updates, the sum of forecast and control computation takes 10.78 / 11.44 ms (95th percentiles 14.70 / 16.20 ms). These measurements exclude CFD advancement, coupling waits and logging outside the timed blocks. The 0.004 s control interval is a simulation-time step, rather than a measured real-time deadline.

D Supplementary Current-Field Reconstruction

Task and comparisons. This supplementary study evaluates current-field reconstruction with a task-specific query-local network, distinct from the forecasting model. It complements the future-query experiments by examining how physical regularization affects spatial smoothness and local detail in current-field reconstruction as sensing and supervision become sparse. It serves as a diagnostic of physical regularization under changing spatial budgets, rather than as a forecasting baseline for EgoCast. It compares a current-frame local IDW baseline with paired No-PDE and PINN reconstructions from three raw velocity-observation frames. Observation, supervision, and physics budgets vary jointly as $M = S \in \{1024, 256, 96, 32\}$ and $C = 8S$.

Data and spatial support. The reconstruction study uses retained circular-trajectory CFD data at 150°, 180° and 210° inflow, sampled every $0.1\,\mathrm{s}$. The body-frame region is $[-0.6, 0.6] \times [-0.6, 0.6] \times [-0.4, -0.1]\,\mathrm{m}^3$, with $\|\boldsymbol{q}\| \geq 0.25\,\mathrm{m}$ and the immediate rotor-source exclusion. Valid targets require finite velocity, positive cell volume, an active native CFD cell and no source mask. Fixed observation and supervision sites use nearest native cells with cell width at most $0.0626\,\mathrm{m}$ and sampling distance at most $0.055\,\mathrm{m}$. The two maximal 1024-site layouts map to distinct native cells; smaller budgets use their nested prefixes. Geometric and validity screening used all three retained inflow conditions. This study compares inflow conditions within the existing circular-trajectory data.

Training targets cover $t = 1.0$–$3.5\,\mathrm{s}$ at 150° and 210° (52 windows); validation uses $t = 4.0$–$5.0\,\mathrm{s}$ in those conditions (20 windows). Testing uses 180° at $t = 1.0$–$5.0\,\mathrm{s}$ (41 windows). Every budget is scored on the same valid native cell centres after removing all cells occupied by either maximal 1024-site layout, leaving 7,460–7,466 cells per test frame. The reconstruction metric is

$$E_{u,\mathrm{rec}} = \left[\frac{1}{T} \sum_{t=1}^{T} \frac{\sum_{i \in \mathcal{Q}_t} V_{ti} \|\hat{\boldsymbol{u}}_{ti} - \boldsymbol{u}_{ti,\mathrm{CFD}}\|_2^2}{\sum_{i \in \mathcal{Q}_t} V_{ti}} \right]^{1/2}, \tag{43}$$

where V_{ti} is native cell volume and $\mathcal{Q}_t$ is the common evaluation support. Thus changing S changes training supervision, not the resolution or locations used for scoring.

Spatial regularity and reconstruction detail. Compared with No-PDE, PINN produces fewer irregular local variations on the displayed slice, particularly at intermediate and low budgets (Figure 14). The smoother predictions also attenuate some high-speed structures, indicating a trade-off

Table 22: Current-field reconstruction under jointly reduced budgets. Entries are volume-weighted vector RMSE in $\mathrm{m\,s^{-1}}$ over 41 test frames; neural models report mean $\pm$ sample standard deviation across three training seeds. Local IDW uses $K = 8$ current observations and has no fitted parameters. C counts candidate physics centres, not stencil nodes.

$M = S$	C	Local IDW	No-PDE	PINN
1024	8192	0.310	0.319 ± 0.002	0.426 ± 0.008
256	2048	0.455	0.471 ± 0.002	0.642 ± 0.004
96	768	0.602	0.666 ± 0.008	0.772 ± 0.002
32	256	0.815	1.060 ± 0.020	0.978 ± 0.025

between spatial regularity and local detail recovery. Table 22 reports the complete volume-weighted comparison. Local IDW has the lowest RMSE at all four budgets, while PINN improves over No-PDE only at $M = S = 32$. At this sparsest budget, physical regularization reduces reconstruction RMSE by a mean paired 7.76%, with improvements in all three seeds.

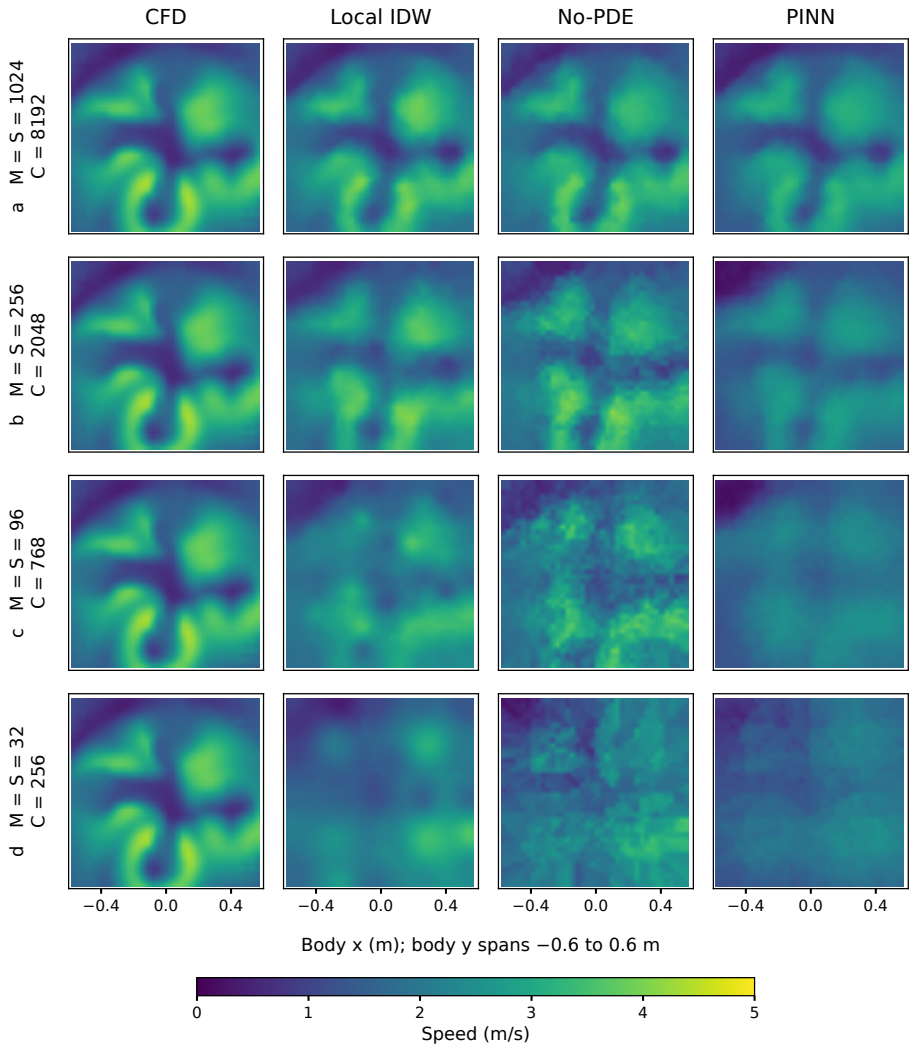

Figure 14: Current-field reconstruction from dense to sparse budgets. Rows (a–d) use the four $M/S/C$ configurations in Table 22; columns show CFD, local IDW, No-PDE and PINN speed on the same body-frame slice at $z = -0.25\,\mathrm{m}$, $t = 5\,\mathrm{s}$ and 180° inflow. Neural results use the fixed seed 20260826. All panels share the same speed scale. Bilinear display interpolation of the 38×38 query grid does not add CFD resolution; the table is scored on native volumetric cells, not these display pixels.

Paired gains and spatial errors. Relative gains are computed within each seed as $100(1 - E_{u,\mathrm{PINN}}/E_{u,\text{No-PDE}})$ and then averaged. At $M = S = 32$, the three gains are 5.89%, 8.47% and 8.92%. Figure 15 localizes vector errors on the same fixed slice as Figure 14. These spatial maps complement the full-volume, 41-frame evaluation in Table 22.

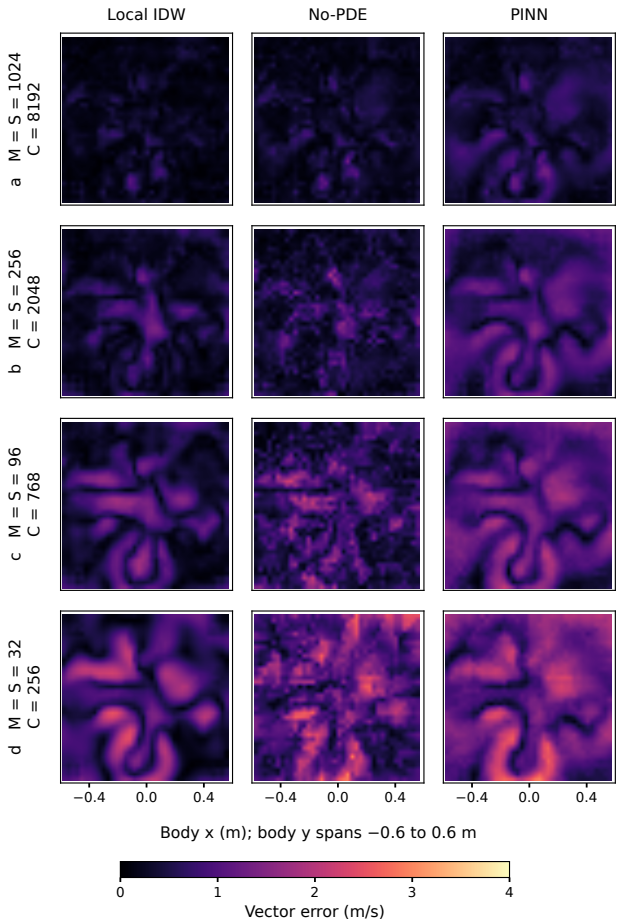

Figure 15: Spatial vector errors for all four reconstruction budgets. Rows (a–d) match Figure 14; columns show local IDW, No-PDE and PINN on the same fixed $z = -0.25\,\mathrm{m}$ slice at $t = 5\,\mathrm{s}$, $180°$ inflow and seed 20260826. Colour denotes $\|\hat{\boldsymbol{u}} - \boldsymbol{u}_{\mathrm{CFD}}\|_2$ with a common scale in $\mathrm{m\,s^{-1}}$. Errors are evaluated at valid queries on the 38×38 slice grid and bilinearly interpolated for display only.

Local architecture and training. Each query selects the nearest $K = 8$ observation sites, ordered by distance. Each neighbour contributes its relative position scaled by $0.1\,\mathrm{m}$ and nine velocity components from three frames, scaled by $U_0 = 6\,\mathrm{m\,s^{-1}}$. The resulting 96-dimensional input passes through a $96 \to 128 \to 96$ encoder. Concatenating the query position scaled by $L_0 = 1\,\mathrm{m}$ gives a 99-dimensional decoder input; the decoder has widths $99 \to 128 \to 128 \to 5$. Hidden layers use Tanh. Outputs are velocity, auxiliary pressure and a non-negative eddy viscosity parameterized as $0.02\ \mathrm{softplus}(a_\nu)$ in normalized units. The final linear layer is zero-initialized. The local encoder reads raw body-frame velocities at each sample time, without pressure input, a GRU, additional history alignment or an additive Ego-Transport reference. Vehicle linear and angular velocities enter the moving-frame physics residual. The local IDW comparator uses inverse-square distance weights over the same eight nearest sites, but only the current frame; it uses neither supervision values nor fitted parameters.

Only the S current velocity targets enter the data loss, a componentwise mean squared error in normalized units. Training uses AdamW with learning rate 3×10^{-4}, weight decay 10^{-5}, batch size 8, and 24 epochs of 32 updates. Seeds 20260825, 20260826 and 20260827 are paired across model

arms and budgets, with matched initialization and batch schedules. Model selection minimizes velocity component MSE on the same 1024 validation sites for every budget. These validation labels do not enter the training loss. Reported standard deviations quantify training-seed variation.

Scaled physics support. The $C = 8S$ candidate centres comprise S known-neighbour centres and $7S$ independent centres, with nested layouts across budgets. Each centre has a $3 \times 3 \times 3$ stencil of spacing $h = 0.03125\,\mathrm{m}$. A known-neighbour stencil reuses one current supervised velocity at an axial neighbour; the centre velocity, pressure and eddy viscosity remain network outputs. Boundary candidates that cannot accommodate a complete stencil are excluded from physics, while their supervised sites remain in the data loss. Validity is required at all 27 current nodes and at the previous-time centre. Every update evaluates all valid selected centres; Table 23 gives their mean counts over the 52 training windows.

Table 23: Reconstruction physics support. Effective counts are per-frame means over 52 training windows after validity checks. Each centre contributes a 27-node stencil.

$M = S$	Nominal C	Valid shared centres	Valid independent centres
1024	8192	914.44	5971.35
256	2048	215.08	1492.94
96	768	79.00	558.96
32	256	25.35	189.60

The moving-frame residual uses the network outputs at t and $t - 0.1\,\mathrm{s}$ for its temporal difference, requiring observations back to $t - 0.3\,\mathrm{s}$ during PINN training. Spatial residuals use molecular viscosity $1.5 \times 10^{-5}\,\mathrm{m^2\,s^{-1}}$ and the predicted eddy viscosity. Both temporal predictions retain gradients; each stencil node selects its own local neighbours. With $\mathcal{H}_c$ and $\mathcal{H}_m$ denoting Huber penalties on normalized continuity and momentum residuals ($\delta = 0.05$), the PINN objective is

$$\mathcal{L}_{\mathrm{rec}} = \mathcal{L}_{\mathrm{data}} + \tfrac{1}{8}\left\langle \mathcal{H}_c + 10\mathcal{H}_m \right\rangle_{\mathrm{shared}} + \tfrac{7}{8}\left\langle \mathcal{H}_c + \mathcal{H}_m \right\rangle_{\mathrm{ind}}, \tag{44}$$

where each group is averaged over its valid centres within each frame and then over the batch. The group weights are fixed across budgets and are not adjusted by the surviving point counts. No-PDE retains only $\mathcal{L}_{\mathrm{data}}$.