

# STCO: Conditional Neural Operators for Time-Dependent PDEs

Xingxin Yang, Zhan Zhang, Juan Li\*

Department of Engineering, King’s College London, London, United Kingdom

\*Correspondence: juan.li@kcl.ac.uk

## Abstract

Neural operators have emerged as efficient surrogates for time-dependent physical systems governed by partial differential equations (PDEs), but their future-state predictions are often conditioned only on observed states and static problem descriptors. For control or optimization, however, body motion, inflow, or forcing are prescribed for the query without being determined solely by the observed state. We introduce the Spatiotemporal Conditional Operator (STCO) for prescribed-condition operator learning (PCOL), a common interface that supplies prescribed target-time condition fields to heterogeneous backbone architectures while retaining their architecture-specific core computation and context pathways. Its condition interface combines Flow-Aware Graph Leaf (FAGL) with Dual-Site Feature-wise Linear Modulation (DSFiLM). Non-learned FAGL uses vorticity from the final observed frame to construct a fixed-cardinality adaptive partition, then co-locates the observed history and target-time condition fields at its regional coordinates. DSFiLM injects separate motion, inflow, and force routes before and after operator computation through current-feature-driven slot- and channel-wise gates. We evaluate twelve matched backbone architectures with different existing physical and temporal inputs. The immersed-boundary computational fluid dynamics (CFD) benchmark spans prescribed motion, inflow disturbances, body-force actuation, and morphology. Across twelve matched backbones, three regimes, and two lead ranges, STCO yields mean paired reductions of 31.1% in relative- $L_2$  field error and 24.7% in normalized pressure-derived load error. It also lowers longer-lead field error for 11 backbones, while interventions on individual condition groups produce measurable prediction changes for every group evaluated.

## Introduction

Neural operators learn mappings between function spaces and provide efficient surrogates for parametric partial differential equations (PDEs) (Li et al. 2021; Lu et al. 2021; Li et al. 2023b). For time-dependent PDEs, they commonly forecast a future field from an initial state or observed trajectory, with lead time or auxiliary trajectories supplied as context (Alkin et al. 2024; Herde et al. 2024; Mousavi et al. 2025; Kassai Koupai et al. 2025). This formulation is effective when the requested future is determined by the observed dynamics and fixed problem descriptors.

Control and optimization often require a different query. The system response must be evaluated under a future sce-

nario specified for the query, such as a gust, target body geometry, or prescribed boundary input (Bartels 2013; Sedky et al. 2022; Han et al. 2021; Hu and Liu 2025). The history describes the current dynamics, while the prescribed condition distinguishes the future to be evaluated. As shown in Figure 1, we formulate this prediction problem as *prescribed-condition operator learning* (PCOL). In PCOL, a conditional neural operator maps an observed history, a lead time, and physical condition fields specified at the target time to the corresponding future response. This query structure is also shared by predictive world models that infer future states from past observations and a supplied action or intervention (NVIDIA et al. 2025; Hafner et al. 2025). PCOL resolves the response as a PDE solution field that can be evaluated in downstream control or optimization.

We examine PCOL through two-dimensional incompressible moving-body flow. Geometry and motion alter the boundary, inflow disturbances modify inlet data, and distributed forces enter the momentum equation, allowing distinct prescribed mechanisms to be varied within one Navier–Stokes system (Mittal and Iaccarino 2005; Sotiropoulos and Yang 2014). This setting exposes three linked challenges. **1) Representation alignment.** Observed and prescribed fields differ in channels, spatial support, and physical meaning. **2) Selective condition modulation.** The model must preserve the roles of geometry, inflow, and forcing while adapting their influence to the current features. **3) Sparse-event supervision.** Brief force and inflow events provide few condition-revealing pairs under uniform temporal sampling.

We introduce the *Spatiotemporal Conditional Operator* (STCO), shown in Figure 2, as a common condition interface for heterogeneous backbone architectures. Its *Flow-Aware Graph Leaf* (FAGL) representation forms a fixed-cardinality, vorticity-aware partition and co-locates historical and prescribed fields at shared regional slots. *Dual-Site Feature-wise Linear Modulation* (DSFiLM) assigns framewise motion, inflow, and force to separate routes, with signed-distance geometry localizing the motion route. Its independently parameterized IN-DSFiLM and OUT-DSFiLM modules act before and after the backbone core, respectively, while current features gate each route by slot and channel. External-activity sampling increases exposure to onset crossings and active intervals. STCO adds target-time conditioning while retaining each backbone’s core computation and existing context

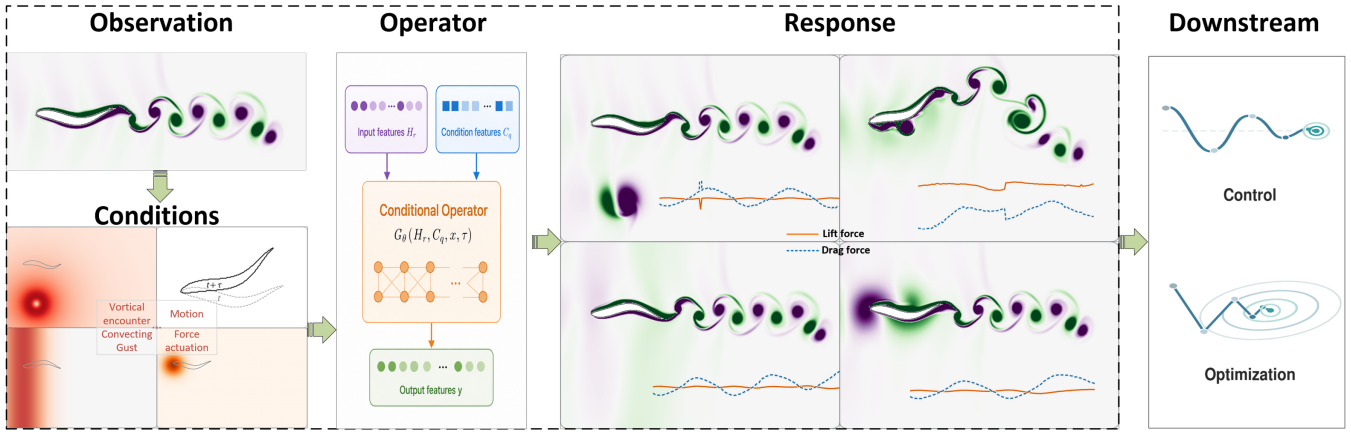


Figure 1: Prescribed-condition operator learning in moving-body flow. The dashed frame marks the PCOL map from observed history and prescribed condition to future response. Examples include a vortical encounter, body motion, a convecting gust, and body-force actuation. Response panels show vorticity and pressure-derived load histories from CFD ground truth (GT). The predicted velocity–pressure field supports downstream control and optimization.

pathways.

We evaluate this interface through twelve matched pairs of baseline (Base) and STCO configurations. Within each pair, FAGL, external-activity sampling, data, backbone, decoder, and training protocol are fixed. The Base models retain their original inputs. Eleven receive observed-frame conditions, and five also encode lead time. STCO preserves these inputs and activates DSFiLM, which combines target-time condition routes with lead-time modulation. The paired differences measure the predictive value of this complete interface under an otherwise matched representation and training protocol. STCO yields a positive mean field gain across the six regime–lead strata for all 12 backbones, lowers the field error beyond the training lead range for 11, and reduces pressure-derived load error overall. *Modulation counterfactual* (MCF) interventions establish prediction sensitivity to every evaluated spatial condition group.

Our contributions are summarized below.

- We formulate PCOL as predicting a future response from observed history and physical conditions prescribed for the query. We address this problem with STCO, a common condition interface that retains the core computation of heterogeneous backbone architectures.
- We introduce vorticity-aware FAGL for adaptive spatial alignment of observed fields and prescribed conditions. DSFiLM modulates separate motion, inflow, and force routes before and after backbone computation, with their influence gated by the current features.
- We construct a moving-body-flow benchmark and evaluate STCO in matched comparisons across twelve backbones. External-activity sampling increases exposure to sparse force and inflow events. MCF measures prediction sensitivity to each evaluated spatial condition group.

## Related Work

**Prescribed response and conditional operators.** Moving-boundary predictors infer the next flow field from recent

flow fields and boundary-position histories (Han et al. 2021). A closely related rigid-body fluid–structure interaction study predicts the terminal fluid state from an initial state and a prescribed motion sequence over the full prediction interval (Zhong and Meidani 2025). It compares direct concatenation, temporal pooling, and sequential coupling. Boundary-control operators map temporal boundary inputs to boundary-output trajectories (Hu and Liu 2025). Other neural operators evaluate candidate controls within an outer optimization loop or predict controller gains and feedback inputs (Hwang et al. 2022; Bhan, Shi, and Krstić 2024). PCOL addresses forward response from observed history and target-time condition fields, leaving control selection downstream. MIONet represents products of function spaces (Jin, Meng, and Lu 2022), while GNOT, GINO, and Unisolver process heterogeneous inputs, geometry, or structured PDE descriptors (Hao et al. 2023; Li et al. 2023b; Zhou et al. 2025). GEPS learns environment conditioning, and time-dependent operators incorporate lead time or context trajectories (Kassaï Koupai et al. 2024; Herde et al. 2024; Mousavi et al. 2025; Kassaï Koupai et al. 2025). STCO organizes multiple target-time physical fields through one interface rather than introducing another backbone architecture.

**Adaptive representation and modulation.** Irregular-domain operators use coordinate deformation, graph–grid transfer, regional graphs, or latent tokens (Li et al. 2023a,b; Mousavi et al. 2025; Alkin et al. 2024). UPT also conditions Transformer and Perceiver blocks on time or velocity context through feature modulation (Alkin et al. 2024). FAGL draws on flow-guided refinement (Kamkar et al. 2011) and graph-based component merging (Felzenszwalb and Huttenlocher 2004) to allocate a fixed slot budget and align fields at shared indices. FiLM applies condition-dependent affine transformations, SPADE makes them spatially varying, and squeeze–excitation derives channel weights from current features (Perez et al. 2018; Park et al. 2019; Hu, Shen, and Sun 2018). DSFiLM combines these principles through sep-

arate spatial condition routes, local state-dependent gates, and independently learned IN-DSFiLM and OUT-DSFiLM modules.

## Methodology

### Problem Formulation

**Setup.** We instantiate PCOL for time-dependent PDEs in incompressible moving-body flow. Let  $\mathbf{x} \in \Omega(t) \subset \mathbb{R}^2$  denote a point in the time-dependent fluid domain and  $\partial\Omega_b(t)$  the moving body boundary. Using body chord  $L$ , reference inflow speed  $U_\infty$ , and constant density  $\rho$  as the length, velocity, and density scales, we scale time, pressure, and acceleration by  $L/U_\infty$ ,  $\rho U_\infty^2/2$ , and  $U_\infty^2/L$ , respectively. After dropping dimensionless superscripts, the velocity  $\mathbf{u} = (u, v)$  and pressure coefficient  $p$  satisfy

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{2} \nabla p + \text{Re}^{-1} \nabla^2 \mathbf{u} + \mathbf{g}, \quad (1)$$

$$\nabla \cdot \mathbf{u} = 0.$$

Here,  $\text{Re} = U_\infty L / \nu$ ,  $\nu$  is the kinematic viscosity, and  $\mathbf{g}(\mathbf{x}, t)$  is a prescribed body-force acceleration. The prescribed boundary conditions are  $\mathbf{u} = \mathbf{u}_b$  on  $\partial\Omega_b(t)$  and  $\mathbf{u} = \mathbf{e}_x + \mathbf{u}_{bc}$  on the inlet  $\Gamma_{\text{in}}$ . Here,  $\mathbf{u}_b$  is the prescribed body velocity,  $\mathbf{e}_x = (1, 0)^\top$  is the nondimensional base inflow, and  $\mathbf{u}_{bc}$  is the prescribed inflow perturbation. Body kinematics define the signed-distance field  $\psi(\mathbf{x}, t)$ , whose zero level set is  $\partial\Omega_b(t)$ . We sample  $T$  stored frames at times  $\{t_n\}_{n=0}^{T-1}$ . At frame  $n$ , the domain is  $\Omega_n = \Omega(t_n)$ , and for  $\mathbf{x} \in \Omega_n$  we define the response and prescribed-condition fields

$$\mathbf{y}_n(\mathbf{x}) = [u_n(\mathbf{x}), v_n(\mathbf{x}), p_n(\mathbf{x})]^\top, \quad (2)$$

$$\mathbf{c}_n(\mathbf{x}) = [\psi_n(\mathbf{x}), \Delta\psi_n(\mathbf{x}), \mathbf{g}_n(\mathbf{x})^\top, \mathbf{u}_{bc,n}(\mathbf{x})^\top]^\top.$$

The framewise signed-distance change is  $\Delta\psi_n = \psi_n - \psi_{n-1}$  for  $n \geq 1$ , with  $\Delta\psi_0 = 0$ . The six condition channels encode geometry and framewise motion through  $\psi_n$  and  $\Delta\psi_n$ , distributed forcing through  $\mathbf{g}_n$ , and inflow perturbation through  $\mathbf{u}_{bc,n}$ .

**Objective.** Let  $n_r$  be the final observed reference index and  $n_q > n_r$  the queried future index. Subscripts  $r$  and  $q$  henceforth denote evaluation at  $n_r$  and  $n_q$ . The observed history is  $\mathcal{H}_r = \{(\mathbf{y}_{n_k}, \mathbf{c}_{n_k})\}_{k=1}^K$ , where  $K$  is the number of observed frames and  $n_1 < \dots < n_K = n_r$ . The query lead is  $\Delta n = n_q - n_r$ , represented by the normalized scalar  $\tau = \Delta n / (T - 1)$ . PCOL seeks an operator that maps the observed history and the condition prescribed for this query to its future response

$$\hat{\mathbf{y}}_q(\mathbf{x}) = \mathcal{G}_\theta(\mathcal{H}_r, \mathbf{c}_q, \mathbf{x}, \tau) \approx \mathbf{y}_q(\mathbf{x}), \quad \mathbf{x} \in \Omega_q, \quad (3)$$

where  $\mathbf{y}_q = \mathbf{y}_{n_q}$ ,  $\mathbf{c}_q = \mathbf{c}_{n_q}$ ,  $\Omega_q = \Omega_{n_q}$ , and  $\theta$  contains the learned parameters.

### STCO Architecture

STCO realizes the PCOL map through adaptive spatial alignment and conditional modulation, as shown in Figure 2(a). FAGL co-locates the observed history and target-time condition fields at shared regional slots. Backbone-specific

adapters connect IN-DSFiLM and OUT-DSFiLM to the selected backbone core while preserving its computation.

Let  $\mathbf{X}_r = [\mathbf{x}_{r,\ell}]_{\ell=1}^{N_\ell} \in \mathbb{R}^{N_\ell \times 2}$  denote the  $N_\ell$  final-observed-frame slot coordinates, and let  $\mathbf{X}_q = [\mathbf{x}_{q,j}]_{j=1}^{N_q} \in \mathbb{R}^{N_q \times 2}$  contain the  $N_q$  requested coordinates, with  $\mathbf{x}_{q,j} \in \Omega_q$ . Let  $\bar{\mathbf{Z}}_r \in \mathbb{R}^{K \times N_\ell \times 9}$  contain the aligned history features of the three response and six condition channels, and let  $\bar{\mathbf{C}}_q \in \mathbb{R}^{N_\ell \times 6}$  contain the target-time condition sampled at the same coordinates. Its  $\ell$ th slot is  $\bar{\mathbf{C}}_{q,\ell} = [\bar{\psi}_{q,\ell}, \bar{\Delta\psi}_{q,\ell}, \bar{\mathbf{g}}_{q,\ell}^\top, \bar{\mathbf{u}}_{bc,q,\ell}^\top]^\top$ . Here, overbars mark aligned values before standardization. Vorticity determines the observed-frame partition but does not enter either tensor. Fixed channelwise maps standardize both quantities after aggregation and alignment, giving  $\mathbf{Z}_r = \mathcal{N}_z(\bar{\mathbf{Z}}_r)$  and  $\mathbf{C}_q = \mathcal{N}_c(\bar{\mathbf{C}}_q)$ . The unstandardized nondimensional signed-distance component  $\bar{\psi}_q \in \mathbb{R}^{N_\ell}$  is retained for boundary localization. The query lead  $\tau$  remains a separate scalar input. Evaluating the prescribed field  $\psi_q$  on  $\mathbf{X}_q$  defines the query-frame fluid mask used for training and evaluation. For backbone  $b$ , let  $\mathcal{E}_b$ ,  $\mathcal{O}_b$ , and  $\mathcal{D}_b$  denote its input adapter, core computation, and dense readout, and let  $\mathcal{M}_b^{\text{in}}$  and  $\mathcal{M}_b^{\text{out}}$  denote its IN-DSFiLM and OUT-DSFiLM maps. STCO computes

$$\begin{aligned} \mathbf{h}_0 &= \mathcal{E}_b(\mathbf{Z}_r, \mathbf{X}_r, \tau), & \mathbf{h}_1 &= \mathcal{M}_b^{\text{in}}(\mathbf{h}_0, \mathbf{C}_q, \bar{\psi}_q, \tau), \\ \mathbf{h}_2 &= \mathcal{O}_b(\mathbf{h}_1, \mathbf{X}_r, \tau), & \mathbf{h}_3 &= \mathcal{M}_b^{\text{out}}(\mathbf{h}_2, \mathbf{C}_q, \bar{\psi}_q, \tau), \\ \hat{\mathbf{y}}_q(\mathbf{X}_q) &= \mathcal{D}_b(\mathbf{h}_3, \mathbf{X}_r, \mathbf{X}_q). \end{aligned} \quad (4)$$

Here,  $\mathbf{h}_0, \mathbf{h}_1 \in \mathbb{R}^{N_\ell \times D_{\text{in}}}$  and  $\mathbf{h}_2, \mathbf{h}_3 \in \mathbb{R}^{N_\ell \times D_{\text{out}}}$ , where  $D_{\text{in}}$  and  $D_{\text{out}}$  are the respective interface widths. For the shared DSFiLM form,  $(\mathbf{h}_{\text{in}}, \mathbf{h}_{\text{out}}) = (\mathbf{h}_0, \mathbf{h}_1)$  for IN-DSFiLM and  $(\mathbf{h}_2, \mathbf{h}_3)$  for OUT-DSFiLM. The two modulation maps have independent parameters.

**Flow-Aware Graph Leaf.** FAGL maps each observed grid to  $N_\ell$  adaptive regional slots through graph construction, partition, and remapping, as shown in Figure 2(b). **Graph.** At observed frame  $n$ , let  $\omega_n(\mathbf{x}) = \partial_x v_n(\mathbf{x}) - \partial_y u_n(\mathbf{x})$  denote the scalar vorticity and  $\mu_n(\mathbf{x}) = |\omega_n(\mathbf{x})|$  its magnitude. An eight-neighbor grid graph assigns edge dissimilarity from spatial separation and contrasts in  $\mu_n$  and  $\|\nabla \mu_n\|_2$ . Processing edges in increasing dissimilarity merges adjacent components when the connecting edge is small relative to their internal variation (Felzenszwalb and Huttenlocher 2004). This retains sharp vorticity changes for refinement.

**Partition.** Each eligible component  $C$  in refinement batch  $\mathcal{B}$  is ranked by

$$\mathcal{P}_n(C) = \frac{|C|}{2} + \frac{\sum_{\mathbf{x}_i \in C} \mu_n(\mathbf{x}_i)}{\bar{\mu}_B + \varepsilon_\mu} \quad (5)$$

where  $|C|$  is its number of grid points,  $\bar{\mu}_B$  is the mean of  $\mu_n$  over  $\mathcal{B}$ , and  $\varepsilon_\mu = 10^{-6}$  stabilizes the denominator. The two terms preserve spatial coverage and direct more slots to integrated vorticity activity. Inspired by feature-driven refinement for vortex-dominated flows (Kamkar et al. 2011), selected components are median-split along the best of  $x$ ,  $y$ , and signed  $\omega_n$ ; the signed candidate separates opposing

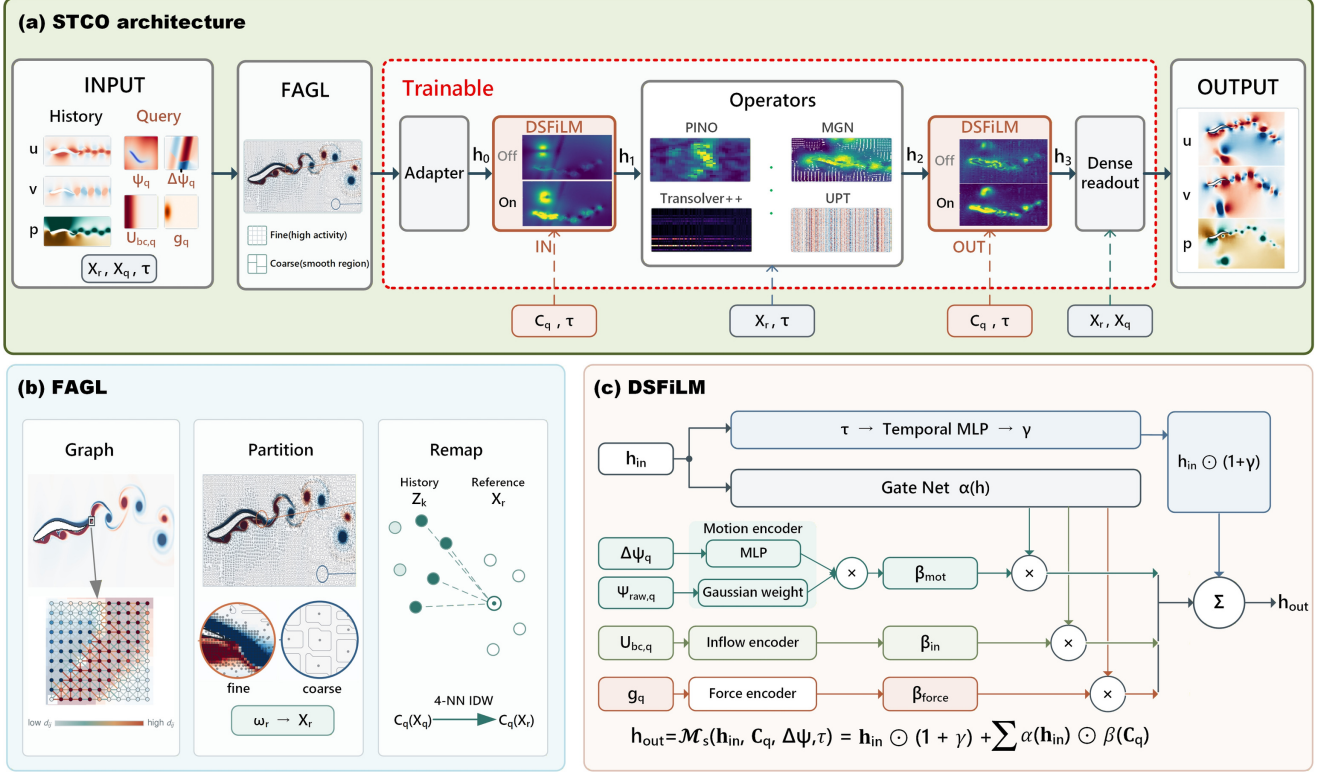


Figure 2: STCO architecture. (a) Overall pipeline from observed history and prescribed target-time conditions to the velocity–pressure response on  $X_q$ . FAGL provides regional coordinates  $X_r$ , while IN-DSFiLM and OUT-DSFiLM bracket a selected backbone core. (b) FAGL constructs an observed-vorticity-aware partition and aligns historical aggregates and target-time conditions on  $X_r$  by four-neighbor inverse-distance weighting. (c) DSFiLM applies lead-time scaling and state-gated motion, inflow, and force shifts, with signed distance localizing motion near the body.

rotations when selected. Refinement ends at  $N_\ell$  or when no eligible split remains, after which the collection is normalized to exactly  $N_\ell$  slots. **Remap.** For the grid points  $S_{n,\ell}$  in slot  $\ell$ , let  $\mathcal{A}_{n,\ell}[f]$  be the mean of field  $f$  and  $x_{n,\ell}$  their centroid. Historical aggregates and the target-time condition are aligned to  $X_r$  by

$$\bar{z}_{k,\ell} = \text{IDW}_4\left(\{x_{n_k,j}, \mathcal{A}_{n_k,j}[y_{n_k}, c_{n_k}]\}_{j=1}^{N_\ell}, x_{r,\ell}\right), \quad (6)$$

$$\bar{c}_{q,\ell} = \text{IDW}_4[c_q](x_{r,\ell}).$$

Here,  $\text{IDW}_4$  uses normalized inverse-distance weights over four nearest sources; the second line interpolates directly from the target-time condition mesh. Stacking  $\bar{z}_{k,\ell}$  gives  $\bar{Z}_r$ . The supplementary material specifies the merge, refinement, interpolation, and slot-normalization rules.

**Dual-Site Feature-wise Linear Modulation.** Spatial alignment places prescribed mechanisms on common slots but leaves their influence on current features unresolved. As detailed in Figure 2(c), DSFiLM combines feature-wise affine conditioning in FiLM (Perez et al. 2018), spatially varying modulation in SPADE (Park et al. 2019), and feature-dependent channel gating in squeeze–excitation (Hu, Shen, and Sun 2018). IN-DSFiLM acts before the backbone core

and OUT-DSFiLM before dense readout. The two sites have independent parameters.

Consider a slot at either site and omit site and slot indices. Let  $h_{in} \in \mathbb{R}^D$  be the feature entering that module and  $h_{out}$  its modulated output. Three independent two-layer networks encode standardized local  $\Delta\psi_q$ ,  $u_{bc,q}$ , and  $g_q$  channels as the shifts  $\beta_{mot}$ ,  $\beta_{inflow}$ , and  $\beta_{force} \in \mathbb{R}^D$ . Motion acts near the body, so its shift is multiplied by  $w = \exp[-\bar{\psi}_q^2/(2\delta^2)]$ .  $\bar{\psi}_q$  denotes the unstandardized signed distance at that slot and  $\delta > 0$  is a learned bandwidth. A fourth two-layer network maps the sinusoidal encoding of  $\tau$  to a lead-time scale  $\gamma(\tau) \in \mathbb{R}^D$ . This scale is shared across slots. A sigmoid affine projection of  $h_{in}$  produces three channel-wise gates  $\alpha_{mot}$ ,  $\alpha_{inflow}$ , and  $\alpha_{force} \in (0, 1)^D$ . The modulation is

$$h_{out} = h_{in} \odot [1 + \gamma(\tau)] + \alpha_{mot} \odot (w\beta_{mot}) + \alpha_{inflow} \odot \beta_{inflow} + \alpha_{force} \odot \beta_{force}, \quad (7)$$

where  $\mathbf{1} \in \mathbb{R}^D$  is the all-ones vector and  $\odot$  denotes elementwise multiplication. Inspired by the identity-preserving zero-initialization principle of adaLN-Zero (Peebles and Xie 2023), the final scale and shift layers are initialized at zero, so Equation 7 begins with  $h_{out} = h_{in}$ .



## External-Activity Sampling

Brief body-force and inflow events yield few condition-revealing pairs under uniform reference–query sampling. For simulation  $m$ , let  $a_m(n) \in [0, 1]$  be the spatial mean of  $\|\mathbf{g}_{m,n}\|_2^2 + \|\mathbf{u}_{bc,m,n}\|_2^2$  over grid points that are fluid in at least one stored frame, normalized by its maximum over that simulation. For simulations without external force or inflow,  $a_m(n) = 0$  for every  $n$ . Within the selected simulation, an admissible reference–query pair is sampled according to

$$\Pr(n_r, \Delta n \mid m) \propto \eta_m(n_r, \Delta n) \left[ \varepsilon + \sum_{j=0}^{\Delta n} a_m(n_r + j) \right]. \quad (8)$$

The sum accumulates external activity from the reference frame through the query frame. The floor  $\varepsilon = 0.1$  gives every admissible pair nonzero probability. The onset factor  $\eta_m(n_r, \Delta n) = 10$  when an interval crosses the first frame where any force or inflow component exceeds  $10^{-3}$  in magnitude. It is 1 otherwise. Motion remains part of  $c_q$  but is excluded from  $a_m$ , so the activity weighting targets force and inflow events. Every matched Base–STCO pair receives identical temporal and spatial samples. The supplementary material gives the complete rule.

## Experiments

### Experimental Setup

**Datasets.** The benchmark contains 142 two-dimensional simulations of incompressible moving-body flow generated with `WaterLily.jl` (Weymouth and Font 2025). Each simulation contains 256 stored frames, giving 36,352 frames in total. All simulations use  $\text{Re} = 5000$ . The benchmark spans five swimmer morphologies, an undisturbed baseline, and six prescribed-condition families. Three families use analytic instantiations of Lamb–Oseen vortex encounters, convecting finite-width velocity gusts, and localized transverse gusts. Their physical settings are motivated by prior experimental and computational studies (Hufstedler and McKeon 2019; Bartels 2013; Sedky et al. 2022). The remaining families comprise localized Gaussian body-force actuation, prescribed pose or undulation changes, and compound motion–disturbance cases. The held-out simulations vary in amplitude, onset, placement, duration, waveform, and morphology. The training, validation, and test sets contain 92, 8, and 42 simulations, respectively, with no simulation shared across splits. The reference swimming kinematics, analytic condition profiles, waveform variants, parameter values, and case allocation are given in the supplementary material.

**Backbones.** We evaluate the same STCO interface on twelve architecture cores adapted to a common FAGL-to-query contract: MGN (Pfaff et al. 2021), RIGNO (Mousavi et al. 2025), PINO (Li et al. 2024), Poseidon (Herde et al. 2024), GAOT (Wen et al. 2025), GINO (Li et al. 2023b), Transolver++ (Luo et al. 2025), Unisolver (Zhou et al. 2025), MPP (McCabe et al. 2024), CALM-PDE (Hagnberger, Musekamp, and Niepert 2025), UPT (Alkin et al. 2024), and GEPS (Kassaï Koupaï et al. 2024). They span

graph, spectral, geometry-aware, transformer, and latent operator architectures. The labels denote adapted architecture cores rather than reproductions of the original training recipes. All configurations are trained from scratch with the common data objective. PINO uses its FNO core without the physics-informed loss, and Poseidon uses its ScOT core. MPP denotes a Transformer core without multiple-physics pretraining, while GEPS uses low-rank context layers without inference-time adaptation.

**Implementation.** Vorticity-aware FAGL represents each grid with 12,000 adaptive slots, approximately 9.4% of the mean fluid-grid cardinality. Across five representative simulations, mean channelwise reconstruction errors range from 1.5% to 11.8%. With one observed frame, each epoch samples 10,000 reference–query pairs with replacement at  $\Delta n = 1$ –20. Each pair resamples 15,000 supervision coordinates, nominally comprising 80% FAGL-region representatives and 20% additional nonrepresentative locations. After framewise outlier removal and fixed standardization, all models minimize the fluid-masked channel-mean relative- $L_2$  loss over standardized  $u$ ,  $v$ , and  $p$ . Training runs for 80 epochs on one H200 GPU using AdamW with an initial learning rate of  $5 \times 10^{-4}$  and batch size 32. In paired experiments, Base and STCO share the split, samples, FAGL, retained backbone context, adapters, decoder, query geometry, and optimization. Base bypasses both DSFiLM sites, whereas STCO activates the complete interface.

**Evaluation.** A predefined set of 5,004 unique reference–query pairs from all 42 test simulations is shared across configurations. Its six 834-pair strata combine three regimes (pre-event, motion transition, and external onset) with ID-Lead ( $\Delta n = 1$ –20) or OOD-Lead ( $\Delta n = 21$ –40). All predictions are scored on target-frame fluid points of the complete  $514 \times 514$  mesh. Within each stratum,  $E_y$  pools queries and fluid points per eligible simulation to compute channelwise relative- $L_2$  errors, then macro-averages over  $u$ ,  $v$ ,  $p$ , and simulations. The ID-Lead, OOD-Lead, and overall scores average three, three, and six strata, respectively.  $E_{F_p}$  macro-averages per-simulation joint pressure-derived drag–lift RMSE normalized by fixed training-set RMS scales. Both metrics use inverse-standardized predictions. Latency is the batch-one H200 forward-pass time with dense decoding but without FAGL preprocessing.

Modulation counterfactual (MCF) sensitivity measures functional dependence on  $r \in \{\tau, \psi, \Delta\psi, \mathbf{g}, \mathbf{u}_{bc}\}$ . It is evaluated on the same fixed set  $\mathcal{I}$  of 5,004 pairs shared by all model cells. Let  $\hat{\mathbf{Y}}_{i,j}$  and  $\hat{\mathbf{Y}}_{i,j}^{(r)}$  denote the original and intervened inverse-standardized full-mesh predictions for channel  $j$ . The reported score is

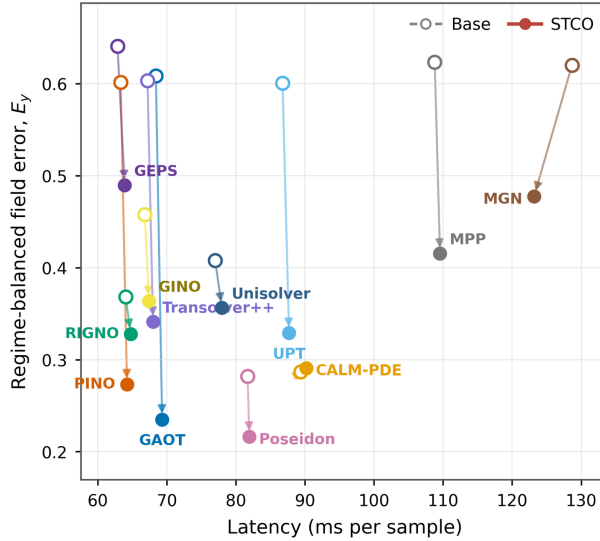
$$\text{MCF}_r = \frac{1}{3|\mathcal{I}|} \sum_{i \in \mathcal{I}} \sum_{j \in \{u,v,p\}} \frac{\|\mathbf{M}_i \odot (\hat{\mathbf{Y}}_{i,j} - \hat{\mathbf{Y}}_{i,j}^{(r)})\|_2}{\max\{\|\mathbf{M}_i \odot \hat{\mathbf{Y}}_{i,j}\|_2, 10^{-6}\}}. \quad (9)$$

Here,  $\mathbf{M}_i$  is the fixed fluid mask. Each intervention holds the observed history, other condition groups, slot and query coordinates, fluid mask, and random state fixed. MCF measures functional input dependence, complementing  $E_y$  and  $E_{F_p}$ , which quantify accuracy against paired CFD ground

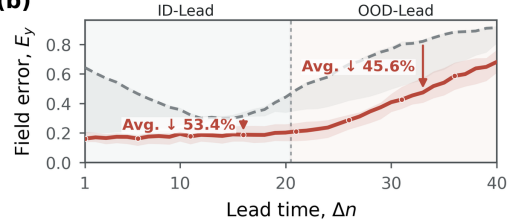
| Backbone       | Field $E_y$ , ID |              |              | Field $E_y$ , OOD |              |              | Load $E_{F_p}$ , ID |              |              | Load $E_{F_p}$ , OOD |              |              | Field gain          |       | MCF sensitivity |  |
|----------------|------------------|--------------|--------------|-------------------|--------------|--------------|---------------------|--------------|--------------|----------------------|--------------|--------------|---------------------|-------|-----------------|--|
|                | Base             | STCO         | $\Delta$ (%) | Base              | STCO         | $\Delta$ (%) | Base                | STCO         | $\Delta$ (%) | Base                 | STCO         | $\Delta$ (%) | Mean $\Delta_y$ (%) | Lead  | Spatial         |  |
| MGN            | 0.438            | <b>0.242</b> | +44.8        | 0.802             | <b>0.713</b> | +11.1        | 0.772               | <b>0.606</b> | +21.5        | 0.774                | <b>0.613</b> | +20.8        | +26.8               | 0.386 | 0.293           |  |
| RIGNO          | 0.243            | <b>0.211</b> | +13.3        | 0.493             | <b>0.445</b> | +9.8         | 0.597               | <b>0.592</b> | +0.9         | <b>0.564</b>         | 0.606        | -7.6         | +11.4               | 0.318 | 0.352           |  |
| PINO           | 0.438            | <b>0.194</b> | +55.7        | 0.765             | <b>0.353</b> | +53.9        | 0.879               | <b>0.559</b> | +36.5        | 0.776                | <b>0.437</b> | +43.6        | +53.6               | 0.211 | 0.374           |  |
| Poseidon       | 0.206            | <b>0.154</b> | +25.4        | 0.358             | <b>0.279</b> | +22.1        | 0.572               | <b>0.544</b> | +4.9         | 0.490                | <b>0.434</b> | +11.5        | +24.2               | 0.293 | 0.323           |  |
| GAOT           | 0.431            | <b>0.139</b> | +67.7        | 0.786             | <b>0.331</b> | +57.9        | 0.783               | <b>0.521</b> | +33.4        | 0.811                | <b>0.479</b> | +40.9        | +61.5               | 0.232 | 0.393           |  |
| GINO           | 0.310            | <b>0.273</b> | +12.1        | 0.605             | <b>0.455</b> | +24.8        | 0.690               | <b>0.636</b> | +7.8         | 0.538                | <b>0.478</b> | +11.2        | +18.3               | 0.254 | 0.622           |  |
| Transolver++   | 0.434            | <b>0.157</b> | +63.7        | 0.773             | <b>0.525</b> | +32.0        | 0.791               | <b>0.528</b> | +33.2        | 0.775                | <b>0.571</b> | +26.4        | +46.9               | 0.030 | 0.484           |  |
| Unisolver      | 0.257            | <b>0.204</b> | +20.6        | 0.558             | <b>0.509</b> | +8.9         | 0.672               | <b>0.596</b> | +11.3        | 0.615                | <b>0.480</b> | +22.1        | +14.4               | 0.272 | 0.516           |  |
| MPP            | 0.466            | <b>0.247</b> | +47.0        | 0.781             | <b>0.584</b> | +25.3        | 0.817               | <b>0.612</b> | +25.1        | 0.747                | <b>0.562</b> | +24.7        | +35.3               | 0.141 | 0.445           |  |
| CALM-PDE       | 0.191            | <b>0.135</b> | +29.4        | <b>0.383</b>      | 0.447        | -16.8        | 0.522               | <b>0.496</b> | +4.9         | 0.488                | <b>0.463</b> | +4.9         | +5.1                | 0.019 | 0.413           |  |
| UPT            | 0.426            | <b>0.136</b> | +68.2        | 0.775             | <b>0.522</b> | +32.6        | 0.775               | <b>0.490</b> | +36.7        | 0.756                | <b>0.530</b> | +29.9        | +49.6               | 0.010 | 0.419           |  |
| GEPS           | 0.475            | <b>0.300</b> | +36.8        | 0.807             | <b>0.680</b> | +15.8        | 0.810               | <b>0.682</b> | +15.8        | 0.783                | <b>0.651</b> | +16.9        | +25.6               | 0.356 | 0.391           |  |
| <b>Average</b> | 0.360            | <b>0.199</b> | +40.4        | 0.657             | <b>0.487</b> | +23.1        | 0.723               | <b>0.572</b> | +19.3        | 0.676                | <b>0.525</b> | +20.4        | +31.1               | 0.210 | 0.419           |  |

Table 1: Regime-balanced accuracy and MCF sensitivity across twelve matched backbones. ID/OOD denote  $\Delta n = 1-20/21-40$ ; errors macro-average three regimes. Positive  $\Delta$  favors STCO, and bold marks the lower paired error. The Average row macro-averages backbone-level entries, including paired reductions. Mean  $\Delta_y$  averages six field gains; MCF reports lead-time and mean spatial-condition sensitivity.

(a)



(b)



(c)

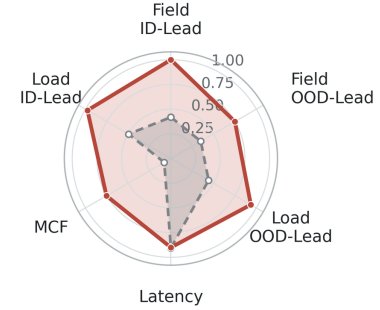


Figure 3: Cross-backbone accuracy, condition dependence, and inference cost. (a) H200 median dense-query latency at  $N_q = 264,196$  and batch size one versus six-stratum regime-balanced  $E_y$ ; arrows connect matched Base and STCO configurations. (b) Cross-backbone median and interquartile range of  $E_y$  at each lead. Annotations average the relative reductions of exact-lead medians within ID-Lead and OOD-Lead, whereas Mean  $\Delta_y$  in Table 1 averages six stratum-level gains per backbone. (c) Cross-backbone means of ID/OOD  $E_y$ ,  $E_{F_p}$ , MCF, and latency under shared normalization; MCF equally weights lead, moving-boundary, force, and inflow responses. Accuracy and MCF use the same fixed population of 5,004 reference-query pairs, with MCF applying controlled input interventions to each pair. Outward denotes lower error or latency and stronger MCF; polygon area is not an aggregate score.

truth. The supplementary material specifies the load metric, complete aggregation, intervention distributions, and deterministic seeding.

## Main Results

STCO improves accuracy for PCOL across twelve heterogeneous backbones (Table 1). The matched gains evaluate the complete condition interface, rather than an individual route

or modulation site.

STCO lowers  $E_y$  in 68/72 regime-lead comparisons (mean: 31.1%) and yields a positive six-stratum mean field gain for all 12 backbones. This includes all 11 backbones retaining observed-frame conditions and all 5 with existing lead-time inputs. The complete interface therefore adds predictive value alongside the retained observed-condition and lead-time pathways. Across the same comparisons,  $E_{F_p}$  de-

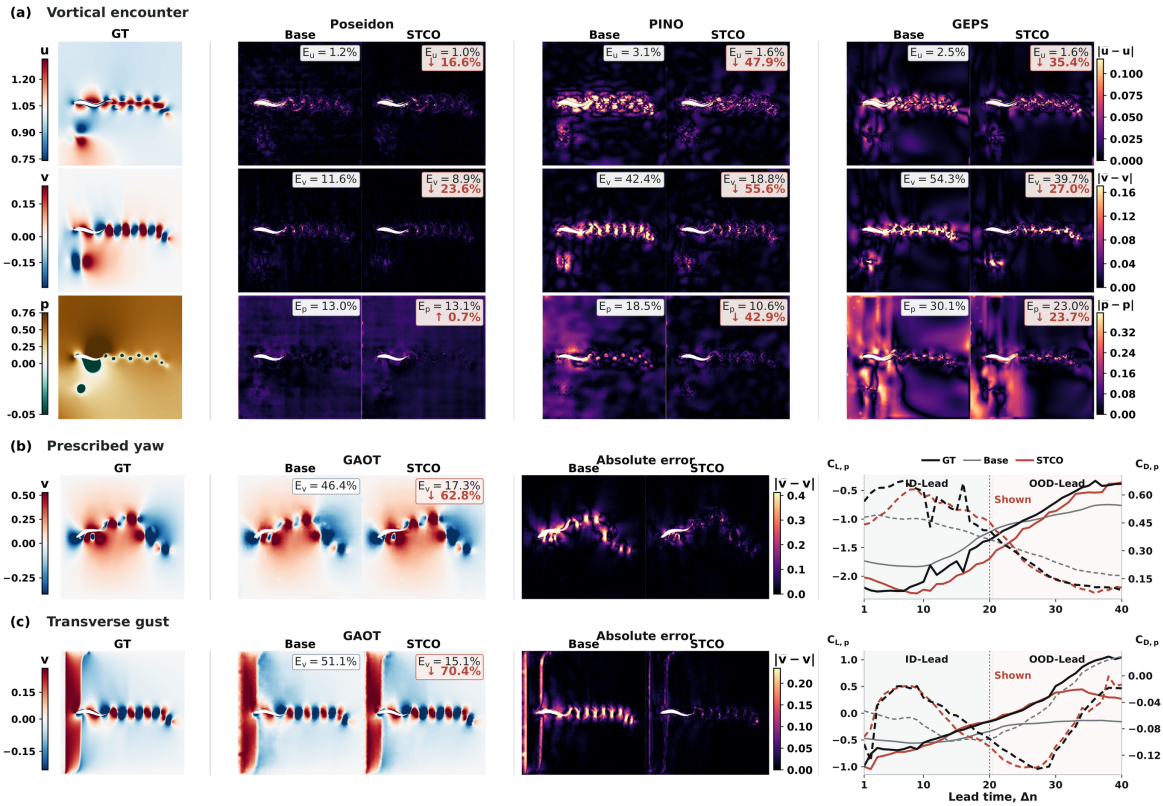


Figure 4: Prescribed-condition velocity–pressure responses and pressure-derived loads. **(a)** CFD GT and matched Base/STCO absolute errors for Poseidon, PINO, and GEPS on a vortical encounter at  $n_r = 128$ ,  $\Delta n = 10$ ; rows show  $u$ ,  $v$ , and  $p$ . **(b,c)** CFD GT, GAOT Base/STCO predictions, absolute  $v$  errors, and independent-query loads for prescribed yaw ( $n_r = 128$ ) and transverse gust ( $n_r = 126$ ) at  $\Delta n = 20$ . Backgrounds mark ID/OOD-Lead, and vertical lines mark displayed fields. Annotations give relative- $L_2$  errors; color limits are shared within comparisons.

creases by 24.7% on average and improves in 63 cases. The cross-backbone field gain is 40.4% in ID-Lead and 23.1% in OOD-Lead, while load reductions remain comparable at 19.3% and 20.4%.

**Resolved physical responses.** Figure 4 links the aggregate gains to resolved fields and pressure-derived loads. On the common vortical query, STCO reduces eight of nine displayed channel errors, with lower residuals around the body, shear layers, and wake. For prescribed yaw and transverse gust, the transverse-velocity error decreases by 62.8% and 70.4%, respectively, at  $\Delta n = 20$ . Across the 40 direct queries in each case, lift and drag RMSE reductions span 33.6%–75.1%. Together, the motion and inflow examples connect improved condition-specific fields to pressure-derived responses relevant to downstream evaluation.

**Longer leads, condition dependence, and cost.** 11 backbones retain lower OOD-Lead  $E_y$  beyond the training range. In Figure 3(b), the cross-backbone median remains below Base at all forty leads, with mean reductions of 53.4% in ID-Lead and 45.6% in OOD-Lead. Although both medians rise overall across OOD-Lead, the advantage persists. MCF is nonzero for every evaluated spatial condition group and STCO backbone, establishing condition dependence along-

side the paired CFD accuracy metrics. Spatial MCF peaks for GINO, while field gain peaks for GAOT, separating sensitivity from accuracy. Figure 3(a) shows that these accuracy gains are achieved at comparable inference latency across matched pairs.

**Limitations.** The evaluation uses one seed and one observed frame in a two-dimensional moving-body Navier–Stokes system from one CFD solver.  $E_{F_p}$  excludes viscous shear, and the largest yaw excursions remain underestimated. The load curves assemble independent target-time queries. Complete condition trajectories, cross-time-consistent prediction, broader PDEs, and closed-loop action selection remain future work.

## Conclusion

PCOL formulates future-response prediction from observed history and prescribed target-time fields. STCO couples vorticity-aware FAGL, DSFiLM, and activity sampling across heterogeneous backbones. Across twelve matched pairs, it reduces mean field and load error within and beyond the training lead range, with condition dependence at comparable inference cost.

## Supplementary Material

This supplement details the benchmark, FAGL implementation, training, evaluation, and extended results.

### Benchmark

Each simulation contains  $T = 256$  stored frames generated with the immersed-boundary computational fluid dynamics (CFD) solver `WaterLily.jl` (Weymouth and Font 2025). The CFD response fields serve as ground truth (GT). At frame  $n$ , the response and prescribed-condition fields are  $\mathbf{y}_n = [u_n, v_n, p_n]^\top$  and  $\mathbf{c}_n = [\psi_n, \Delta\psi_n, \mathbf{g}_n^\top, \mathbf{u}_{bc,n}^\top]^\top$ . The reference and query indices satisfy  $n_q = n_r + \Delta n$ , with normalized lead  $\tau = \Delta n / (T - 1)$ . All experiments use the one-frame history  $\mathcal{H}_r = \{(\mathbf{y}_r, \mathbf{c}_r)\}$ .

### Simulation Families and Splits

Figure 5 illustrates the undisturbed baseline and six prescribed-condition families through representative condition fields, CFD responses, and load histories. Table 2 gives the 92/8/42 simulation-level train/validation/test allocation.

### Flow, Kinematics, and Prescribed Conditions

All simulations use the same base inflow,  $\text{Re} = 5000$ , NACA 0012 section, and solver configuration. The  $514 \times 514$  grid resolves one chord with  $L_{\text{pix}} = 128$  lattice units, and the nondimensional variables set  $L = U_\infty = \rho = 1$ . For body-frame chordwise coordinate  $x$ , define  $s = \text{clip}(x/L, 0, 1)$ . The reference centerline is

$$y_b(s, t) = LA_b(s) \cos\left(\frac{2\pi s}{\lambda^+} - \omega_{\text{kin}}t + \phi\right),$$

$$A_b(s) = A_{0,b} + A_{1,b}s + A_{2,b}s^2, \quad (10)$$

$$\omega_{\text{kin}} = \frac{2\pi \text{St} U_\infty}{2A_{\text{max}}L}.$$

Here,  $b$  indexes morphology,  $A_b(s)$  is its dimensionless amplitude envelope,  $\lambda^+$  is the nondimensional body wavelength, and  $\phi$  is the phase. Table 3 lists five amplitude envelopes with tail amplitude  $A_{\text{max}} = A_b(1) = 0.1$ . All cases use Strouhal number  $\text{St} = 0.4$ ,  $\lambda^+ = 1$ , and  $\phi = 0$ . The kinematic period is  $T_{\text{cyc}} = 2\pi/\omega_{\text{kin}}$ .

$$v_{\text{gust}}(x, t) = A_g F(\xi), \quad \xi = \frac{U_\infty(t - t_0) - (x - x_0)}{w}. \quad (11)$$

Here,  $v_{\text{gust}}$  is the prescribed transverse-velocity perturbation,  $t_0$  is the onset time,  $x_0$  and  $w$  are the streamwise reference position and profile scale, and  $A_g$  is the peak amplitude. Convecting gusts use Gaussian, sine, or one-minus-cosine profiles, and transverse gusts use top-hat, sine-squared, or trapezoidal profiles (Bartels 2013; Andreu-Angulo et al. 2020; Sedky, Biler, and Jones 2022); the Gaussian profile is  $F(\xi) = e^{-4 \ln 2 (\xi - 1/2)^2}$ .

The Lamb–Oseen vortical inflow is

$$\mathbf{u}_{\text{vrt}}(\mathbf{x}, t) = \frac{\Gamma}{2\pi r^2} \left(1 - e^{-\alpha r^2/r_c^2}\right) \mathbf{r}^\perp. \quad (12)$$

| Label    | Condition            | Train | Val. | Test | Total |
|----------|----------------------|-------|------|------|-------|
| baseline | Undisturbed          | 4     | 1    | 0    | 5     |
| vrt      | Vortical encounter   | 15    | 2    | 7    | 24    |
| cvg      | Convecting gust      | 16    | 1    | 9    | 26    |
| tg       | Transverse gust      | 17    | 1    | 8    | 26    |
| act      | Body-force actuation | 18    | 1    | 10   | 29    |
| motion   | Prescribed motion    | 19    | 1    | 4    | 24    |
| mix      | Motion + disturbance | 3     | 1    | 4    | 8     |
| Total    |                      | 92    | 8    | 42   | 142   |

Table 2: Composition of the moving-body-flow benchmark. Training, validation, and test contain 92, 8, and 42 simulations. The test set covers all six prescribed-condition families.

| Morphology $b$ | $A_{0,b}$ | $A_{1,b}$ | $A_{2,b}$ |
|----------------|-----------|-----------|-----------|
| Anguilliform   | 0.0367    | 0.0323    | 0.0310    |
| Carangiform    | 0.0200    | -0.0825   | 0.1625    |
| Subcarangiform | 0.0200    | 0.0300    | 0.0500    |
| Thunniform     | 0.0050    | -0.0200   | 0.1150    |
| Ostraciiform   | 0.0020    | -0.0350   | 0.1330    |

Table 3: Centerline-amplitude coefficients in Equation 10. All five profiles satisfy  $A_b(1) = A_{\text{max}} = 0.1$  while varying the chordwise amplitude envelope.

Here,  $\mathbf{x} = (x, y)$ ,  $\Gamma$  is the circulation,  $r_c$  is the core radius, and  $(x_0, y_0)$  is the vortex center at  $t = t_0$ . With  $\mathbf{x}_c(t) = (x_0 + U_\infty(t - t_0), y_0)$ , define  $\mathbf{r} = \mathbf{x} - \mathbf{x}_c(t)$ ,  $r = \|\mathbf{r}\|_2$ , and  $\mathbf{r}^\perp = (-r_y, r_x)$ . The vortex is zero for  $t < t_0$ , and  $\alpha = 1.25643$  places the peak tangential velocity at  $r = r_c$ .

Body-force actuation uses

$$\mathbf{g}(\mathbf{x}, t) = A_f(t) e^{-\|\mathbf{x} - \mathbf{x}_f\|_2^2 / (2\sigma^2)} \mathbf{e}_y. \quad (13)$$

Here,  $\mathbf{x}_f$  and  $\sigma$  are the forcing center and spatial scale, and  $\mathbf{e}_y$  is the transverse unit vector. The one-cosine amplitude is  $A_f(t) = \frac{1}{2}A[1 - \cos(2\pi(t - t_0)/d)]$  on  $[t_0, t_0 + d]$  and zero otherwise, where  $A$  and  $d$  are its peak amplitude and duration.

Prescribed motion uses

$$\chi(t) = \chi_\infty \mathcal{R}\left(\frac{t - t_0}{d}\right). \quad (14)$$

Here,  $\chi(t)$  is the displacement or yaw angle,  $\chi_\infty$  is its terminal value, and  $d$  is the ramp duration.  $\mathcal{R}(\vartheta)$  equals 0 for  $\vartheta \leq 0$ ,  $\frac{1}{2}(1 - \cos \pi \vartheta)$  for  $0 < \vartheta < 1$ , and 1 for  $\vartheta \geq 1$ . At target frame  $q$ , the gust and vortex fields define  $\mathbf{u}_{bc,q}$ , actuation defines  $\mathbf{g}_q$ , and prescribed motion determines  $\psi_q$  and  $\Delta\psi_q = \psi_q - \psi_{q-1}$ . Table 4 summarizes the varied factors and profiles.

### Frames, Candidate Pairs, and Preprocessing

With 256 frames per simulation, the benchmark contains  $142 \times 256 = 36,352$  frames, including  $92 \times 256 = 23,552$  for training. The 216 reference indices  $n_r = 20, \dots, 235$  and 20 training leads  $\Delta n = 1, \dots, 20$  give  $92 \times 216 \times 20 = 397,440$  admissible training pairs. Each epoch samples 10,000 pairs with replacement.

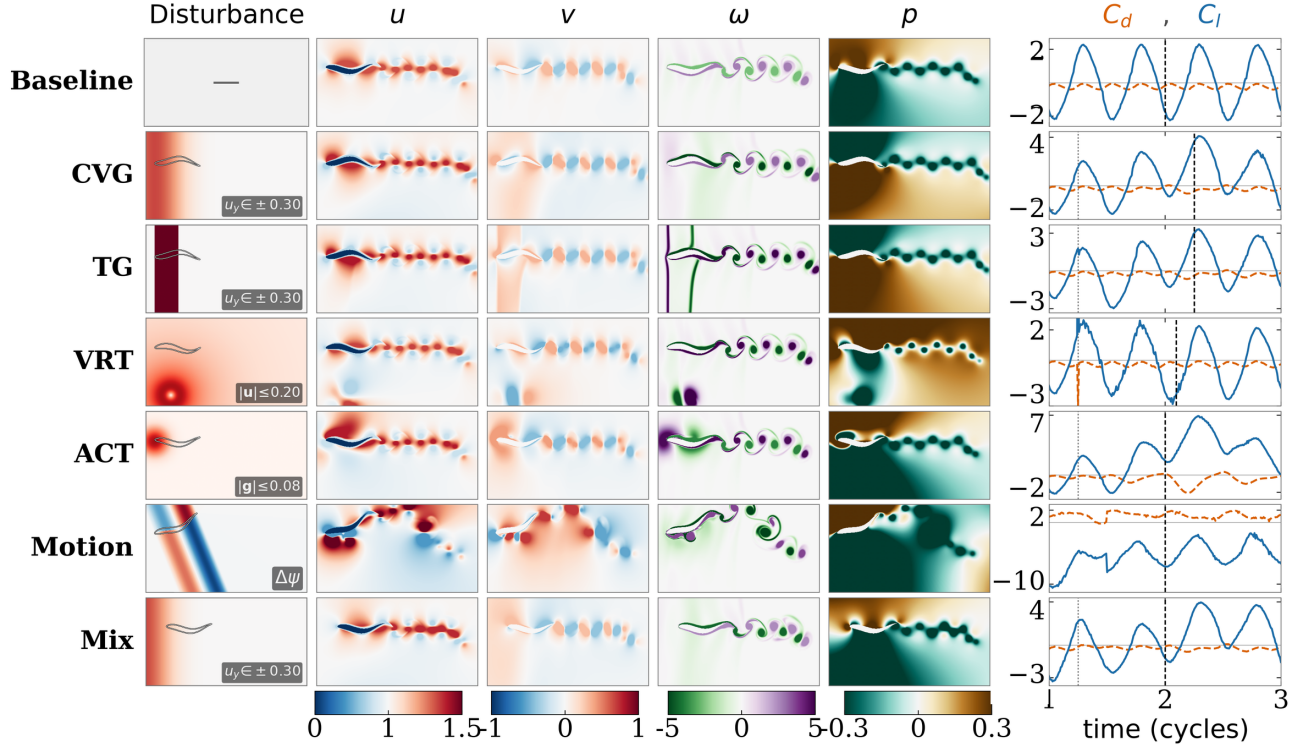


Figure 5: Representative benchmark simulations. Rows correspond to the undisturbed, convecting-gust, transverse-gust, vortical-encounter, actuation, motion, and compound families. The first column shows prescribed condition fields, with motion represented by the one-frame signed-distance change  $\Delta\psi$ . The remaining columns show CFD GT fields and CFD load histories. Dashed markers identify the displayed frames. The learned target is the fluid-region velocity–pressure field  $(u, v, p)$ .

After framewise response outlier removal, fixed channel-wise maps standardize the aligned response and prescribed-condition inputs. Metrics use inverse-standardized response fields.

### FAGL Implementation

FAGL derives the observed vorticity  $\omega_r = \partial_x v_r - \partial_y u_r$  and its magnitude  $\mu_r = |\omega_r|$  from  $\mathbf{y}_r$ . For simulation  $m$ , let  $\mathcal{V}_m$  denote the time-union fluid grid determined by its prescribed geometry sequence. The observed-frame regional slots are

$$\{\mathcal{S}_{r,\ell}\}_{\ell=1}^{N_\ell} = \Pi_{N_\ell}(\mu_r, \mathcal{V}_m), \quad N_\ell = 12,000, \quad (15)$$

where  $\Pi_{N_\ell}$  denotes the FAGL partitioner and  $\mathcal{S}_{r,\ell}$  contains the grid points assigned to slot  $\ell$ . Observed response and condition fields are averaged within each slot. The prescribed target-time fields  $(\psi_q, \Delta\psi_q, \mathbf{g}_q, \mathbf{u}_{bc,q})$  are aligned to the slot centroids by four-neighbor inverse-distance interpolation.

FAGL combines graph merging with bounded refinement to form a fixed-cardinality representation. On the eight-neighbor grid graph, an edge  $(i, j)$  receives dissimilarity  $d_{ij}$  from grid distance, the contrast in observed vorticity magnitude  $\mu_r$ , and the contrast in  $\|\nabla\mu_r\|_2$ . The three cues are normalized within each frame.

Edges are scanned once in ascending  $d_{ij}$ . Components  $A$

and  $B$  merge when

$$d_{ij} \leq \min\left\{\text{Int}(A) + \frac{\kappa}{|A|}, \text{Int}(B) + \frac{\kappa}{|B|}\right\}, \quad (16)$$

where  $\text{Int}(C)$  is the largest accepted edge within  $C$  and  $\kappa$  is the frame-adaptive merge threshold.

If the merge returns fewer than  $N_\ell$  regions, refinement prioritizes

$$\mathcal{P}_r(C) = \frac{|C|}{2} + \frac{\sum_{\mathbf{x}_i \in C} \mu_r(\mathbf{x}_i)}{\bar{\mu}_B + 10^{-6}}. \quad (17)$$

Here,  $|C|$  is the number of points in region  $C$ , and  $\bar{\mu}_B$  is the pointwise mean of  $\mu_r$  over refinement batch  $B$ . Regions with at least four points are median-split along the best of  $x$ ,  $y$ , and signed  $\omega_r$ , selected by balance and vorticity contrast. Refinement stops at  $N_\ell$  regions or when no eligible split remains, after which the collection is normalized to exactly  $N_\ell$  slots. Counts above  $N_\ell$  are reduced by retaining the largest regions; counts below  $N_\ell$  are padded with empty observed slots centered on the grid before target conditions are aligned.

Each nonempty slot stores the mean coordinate and observed fields of its member points.

All interpolation and dense decoding use four-neighbor



| Family | Varied benchmark factors                      | Profile or command                     |
|--------|-----------------------------------------------|----------------------------------------|
| vrt    | Circulation, lateral placement, and onset     | Lamb–Oseen                             |
| cvg    | Amplitude, streamwise placement, and onset    | Gaussian, sine, one-minus-cosine       |
| tg     | Amplitude, streamwise placement, and onset    | Top-hat, sine-squared, trapezoid       |
| act    | Amplitude, placement, onset, and duration     | One-cosine, step, ramp-and-hold        |
| motion | Surge, sway, and yaw magnitude                | Smooth ramp                            |
| mix    | Motion–disturbance pairing and relative onset | cvg+surge, act+surge, vrt+yaw, tg+sway |

Table 4: Prescribed-condition families and varied benchmark factors. Compound cases pair one motion command with one disturbance.

inverse-distance weighting,

$$\begin{aligned}
\text{IDW}_4[\mathbf{f}](\mathbf{x}) &= \sum_{j \in \mathcal{N}_4(\mathbf{x})} w_j(\mathbf{x}) \mathbf{f}_j, \\
w_j(\mathbf{x}) &= \frac{\rho_j(\mathbf{x})^{-1}}{\sum_{k \in \mathcal{N}_4(\mathbf{x})} \rho_k(\mathbf{x})^{-1}}, \\
\rho_j(\mathbf{x}) &= \max\{\|\mathbf{x} - \mathbf{x}_j\|_2, 10^{-8}\}.
\end{aligned} \tag{18}$$

Here,  $\mathcal{N}_4(\mathbf{x})$  indexes the four nearest source coordinates and  $\mathbf{f}_j$  is the source value at  $\mathbf{x}_j$ . FAGL uses  $\omega_r$  only to construct the observed-frame partition. The learned operator receives regional means of  $(u, v, p, \psi, \Delta\psi, \mathbf{g}, \mathbf{u}_{bc})$ . Figure 6 reports reconstruction across slot counts.

## Training and Evaluation

To isolate the predictive contribution of the target-time condition interface, each matched Base–STCO pair uses the same split, temporal and spatial samples, FAGL representation, retained backbone context, adapters, decoder, query geometry, and optimization. Base bypasses both DSFiLM sites, whereas STCO activates the complete interface with  $\mathbf{c}_q$  and  $\tau$ .

## External-Activity Sampling and Optimization

External force and inflow disturbances are active over only part of each trajectory, so uniform pair sampling would draw many intervals without external activity. We therefore sample each training example in two stages: first a simulation, and then a reference–query pair within that simulation.

For simulation  $m$ , let  $\mathcal{V}_m$  denote its time-union fluid grid,  $|\mathcal{V}_m|$  its number of points, and  $h$  a grid-point index. The fields  $\mathbf{g}_{m,n,h}$  and  $\mathbf{u}_{bc,m,n,h}$  are the prescribed body-force acceleration and inflow perturbation at frame  $n$  and point  $h$ . We define the case-normalized external activity at frame  $n$  as

$$\begin{aligned}
\tilde{a}_m(n) &= \frac{1}{|\mathcal{V}_m|} \sum_{h \in \mathcal{V}_m} (\|\mathbf{g}_{m,n,h}\|_2^2 + \|\mathbf{u}_{bc,m,n,h}\|_2^2), \\
a_m^* &= \max_{n'} \tilde{a}_m(n'), \\
a_m(n) &= \begin{cases} \tilde{a}_m(n)/a_m^*, & a_m^* > 10^{-12}, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{19}$$

Here,  $\tilde{a}_m(n)$  is the spatially averaged raw activity,  $a_m^*$  is its maximum over frames  $n'$ , and  $a_m(n) \in [0, 1]$  is the normalized activity trace. The  $10^{-12}$  threshold avoids division

by a numerically zero maximum. Prescribed body motion is excluded from this sampling score and remains represented by  $\psi$  and  $\Delta\psi$  in the model input.

At the first stage, undisturbed and zero-external-activity simulations receive weight  $b_m = 0.3$ , while all others receive  $b_m = 1$ . We sample  $m$  with  $\Pr(m) = b_m / \sum_{m' \in \mathcal{M}_{\text{train}}} b_{m'}$ , where  $\mathcal{M}_{\text{train}}$  is the training simulation set and  $m'$  is its summation index.

At the second stage, the admissible pairs are  $\mathcal{A}_m = \{(n_r, \Delta n) : n_r \in \{20, \dots, 235\}, \Delta n \in \{1, \dots, 20\}\}$ .  $n_r$  is the reference frame and  $\Delta n$  is the forecast lead in stored frames. If simulation  $m$  contains an external-activity onset, let  $n_{\text{on},m}$  be the first frame at which the absolute value of any component of  $\mathbf{g}$  or  $\mathbf{u}_{bc}$  exceeds  $10^{-3}$  anywhere on the grid, and define

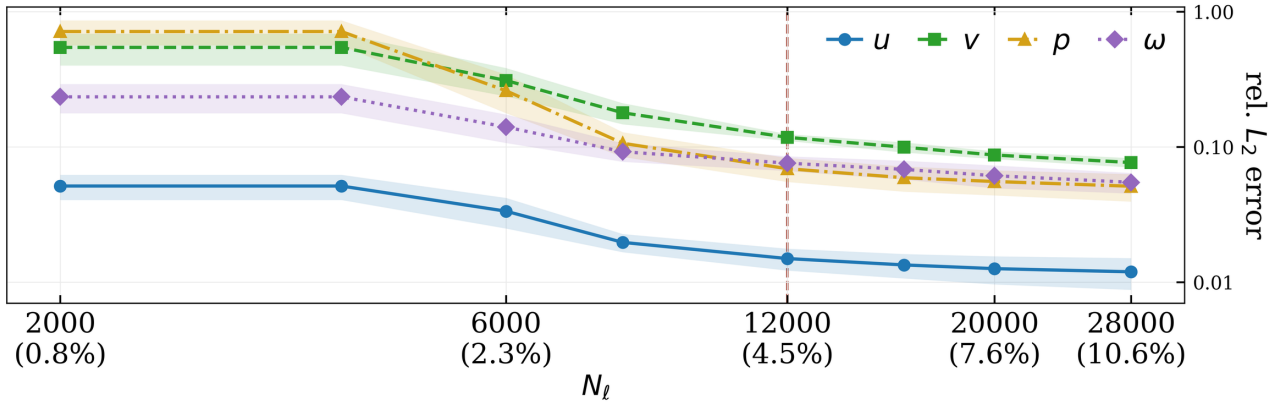
$$\begin{aligned}
\eta_m(n_r, \Delta n) &= \begin{cases} 10, & n_r < n_{\text{on},m} \leq n_r + \Delta n, \\ 1, & \text{otherwise,} \end{cases} \\
w_m(n_r, \Delta n) &= \eta_m(n_r, \Delta n) \left[ 0.1 + \sum_{j=0}^{\Delta n} a_m(n_r + j) \right], \\
\Pr(n_r, \Delta n \mid m) &= \frac{w_m(n_r, \Delta n)}{\sum_{(n'_r, \Delta n') \in \mathcal{A}_m} w_m(n'_r, \Delta n')}.
\end{aligned} \tag{20}$$

Here,  $\eta_m$  is the onset-crossing multiplier,  $w_m$  is the unnormalized pair weight,  $j$  indexes frames after  $n_r$ , and  $(n'_r, \Delta n')$  indexes candidate pairs in the normalizing sum. For simulations without an external-activity onset,  $\eta_m = 1$  for every pair. The activity sum covers the complete reference-to-query interval; the constant 0.1 retains nonzero probability for quiescent intervals, and  $\eta_m$  increases the probability of intervals that cross the first onset. The complete sampling distribution is  $\Pr(m, n_r, \Delta n) = \Pr(m) \Pr(n_r, \Delta n \mid m)$ . Figure 7 illustrates the resulting activity and onset weighting.

Each sampled reference–query pair draws 15,000 supervised coordinates. The sampler targets 80% unique representatives of observed-frame FAGL regions. The remaining nominal budget comprises 12% nonrepresentative locations weighted by observed  $|\omega_r|$  and 8% sampled uniformly. All evaluation metrics use the full mesh.

Training uses batch size 32, 80 epochs, and AdamW with learning rate  $5 \times 10^{-4}$ , weight decay  $10^{-3}$ ,  $(\beta_1, \beta_2) = (0.9, 0.95)$ , gradient clipping at 1, and cosine decay to  $10^{-5}$ . Let  $\tilde{\mathbf{y}}_j, \hat{\mathbf{y}}_j \in \mathbb{R}^{B \times N_q}$  collect the standardized target and pre-

**(a) Slot count vs reconstruction error**



**(b)  $\omega$  reconstruction with leaf outlines**

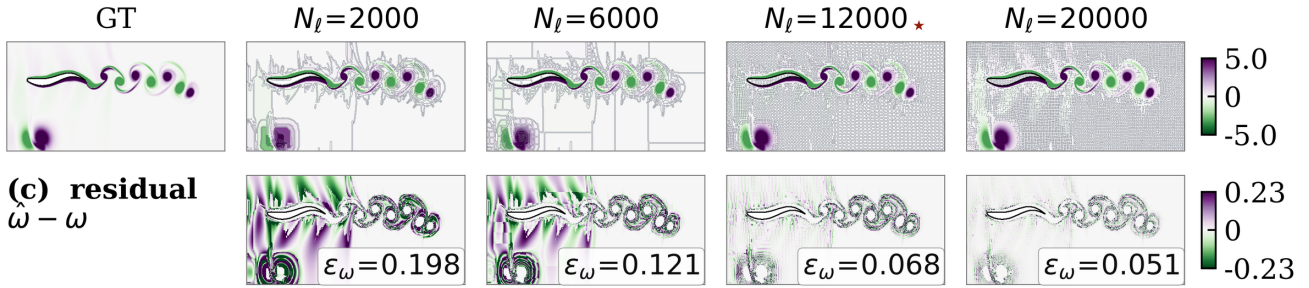


Figure 6: FAGL reconstruction across slot counts. In panel (a), parenthetical labels give  $N_l$  as a percentage of the complete  $514 \times 514$  mesh. (a) Fluid-region relative- $L_2$  error for constant-region reconstructions. Curves and bands give the mean and one standard deviation over five simulations at frame 128. At  $N_l = 12,000$ , the mean errors are 1.5%, 11.8%, 6.9%, and 7.6% for  $u$ ,  $v$ ,  $p$ , and  $\omega_r$ . (b) GT vorticity and representative regional reconstructions with region boundaries. (c) Signed residuals  $\hat{\omega}_r - \omega_r$ . FAGL uses  $\omega_r$  for partitioning, while the operator predicts  $(u, v, p)$ .

diction for channel  $j$  over a batch, and let  $\mathbf{M} \in \{0, 1\}^{B \times N_q}$  be the corresponding query-frame mask. Here,  $B$  is the batch size,  $N_q$  is the number of supervised coordinates, and  $\theta$  denotes the trainable parameters. The loss is

$$\mathcal{L}(\theta) = \frac{1}{3} \sum_{j \in \{u, v, p\}} \frac{\|\mathbf{M} \odot (\hat{\mathbf{y}}_j - \tilde{\mathbf{y}}_j)\|_2}{\|\mathbf{M} \odot \tilde{\mathbf{y}}_j\|_2}. \quad (21)$$

With incomplete batches dropped, the 10,000 training pairs give 312 optimization steps per epoch. Each matched Base-STCO pair uses the same training and validation pair-sampling seeds. Every five epochs, validation evaluates 1,000 sampled pairs from the eight validation simulations at  $\Delta n = 1-20$ , using 15,000 coordinates per pair. Each result uses the checkpoint with the lowest validation relative- $L_2$  error. Each of the 24 configurations is trained from scratch on a single NVIDIA H200 GPU.

### Predefined Evaluation Set

Evaluation scores all target-frame fluid points on the complete  $514 \times 514$  mesh. The predefined evaluation set contains 5,004 unique reference-query pairs from all 42 test simulations, divided into six 834-pair regime-lead strata and shared

| Regime            | ID-Lead<br>pairs/cases | OOD-Lead<br>pairs/cases |
|-------------------|------------------------|-------------------------|
| Pre-event         | 834/38                 | 834/8                   |
| Motion transition | 834/4                  | 834/4                   |
| External onset    | 834/37                 | 834/37                  |

Table 5: Composition of the predefined 5,004-pair evaluation set. Each regime-lead stratum contains 834 pairs. Metrics macro-average the contributing simulations within each stratum. A simulation may contribute pairs to more than one regime.

across all configurations. ID-Lead is  $\Delta n = 1-20$ , and OOD-Lead is  $\Delta n = 21-40$ . Pairs are assigned to the pre-event regime when  $n_r + \Delta n < n_{\text{event}}$ , to motion transition when  $n_{\text{start}} \leq n_r < n_r + \Delta n \leq n_{\text{end}}$ , and to external onset when  $n_r < n_{\text{event}} \leq n_r + \Delta n$ . Here,  $n_{\text{event}}$  is the annotated disturbance onset, while  $n_{\text{start}}$  and  $n_{\text{end}}$  delimit the prescribed motion transition. Table 5 gives the stratum composition.

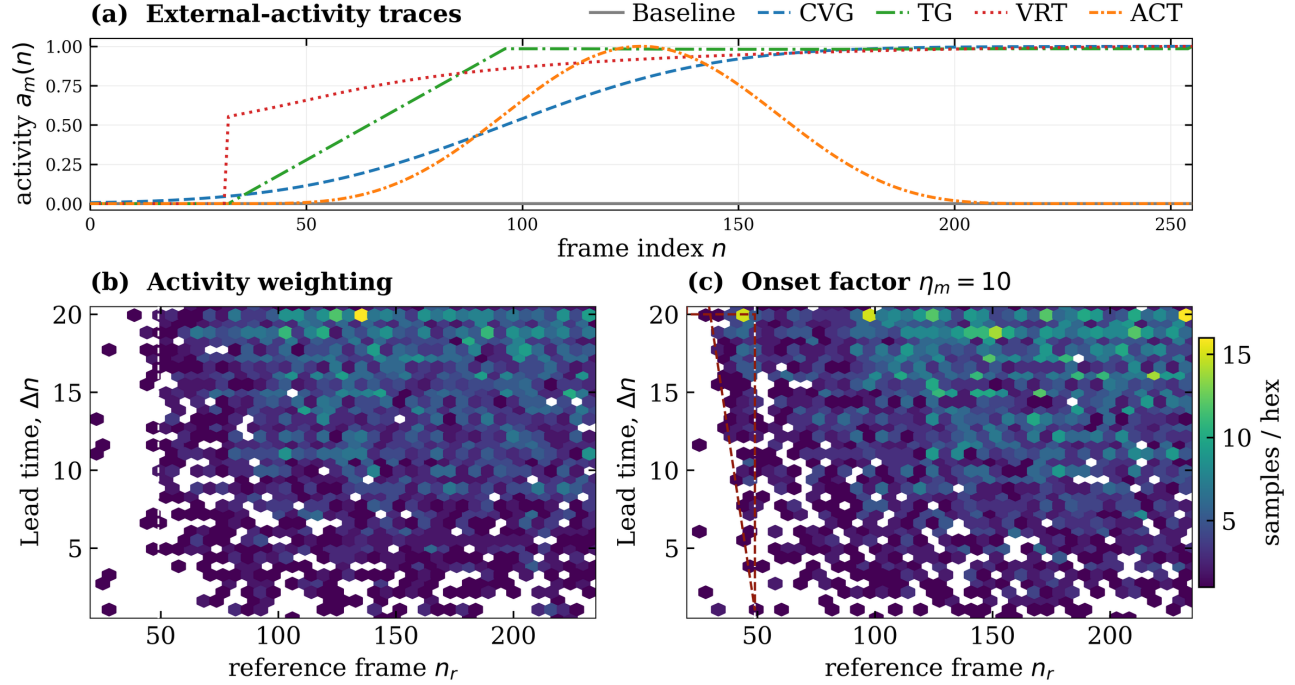


Figure 7: External-activity sampling. (a) Case-normalized traces for the undisturbed baseline, convecting gust (CVG), transverse gust (TG), Lamb–Oseen vortex (VRT), and body-force actuation (ACT). (b,c) Four thousand pairs sampled from one transverse-gust simulation. Activity weighting favors active intervals, and the onset factor further favors onset-crossing pairs.

### Field and Pressure-Derived Load Metrics

Let  $\varsigma$  index a regime–lead stratum,  $\mathcal{C}_\varsigma$  its contributing simulations, and  $\mathcal{Q}_{m,\varsigma}$  the selected pair indices from simulation  $m$ . For channel  $j \in \{u, v, p\}$  and mesh point  $h$ , let  $y_{m,i,h,j}$  and  $\hat{y}_{m,i,h,j}$  be the GT and predicted target values, and let  $M_{m,i,h}$  be the query-frame fluid mask. For values  $z_{m,i,h}$  on these pairs, define the masked norm  $\|z\|_{m,\varsigma}^2 = \sum_{i \in \mathcal{Q}_{m,\varsigma}} \sum_h M_{m,i,h} |z_{m,i,h}|^2$ . The field metric is

$$E_{m,\varsigma,j} = \frac{\|\hat{y}_{m,j} - y_{m,j}\|_{m,\varsigma}}{\max\{\|y_{m,j}\|_{m,\varsigma}, 10^{-6}\}}, \quad (22)$$

$$E_y^{(\varsigma)} = \frac{1}{3|\mathcal{C}_\varsigma|} \sum_{m \in \mathcal{C}_\varsigma} \sum_{j \in \{u,v,p\}} E_{m,\varsigma,j}.$$

The pressure-derived drag and lift coefficients are reconstructed as

$$\begin{aligned} \mathbf{C}_p(t) &= [C_{D,p}(t) \quad C_{L,p}(t)]^\top \\ &= - \int_{\partial\Omega_b(t)} p(\mathbf{x}, t) \mathbf{n}(\mathbf{x}, t) ds. \end{aligned}$$

Here,  $\partial\Omega_b(t)$  is the prescribed body boundary,  $\mathbf{n}$  its outward unit normal, and  $ds$  the boundary arc-length element. GT and predicted loads use the same query-frame signed-distance contour and quadrature. The stored field is the pressure coefficient  $p$  defined in the main paper. For  $k \in \{D, L\}$ , the

normalized pressure-derived load error is

$$E_{m,\varsigma,F_p}^2 = \frac{1}{2|\mathcal{Q}_{m,\varsigma}|} \sum_{i \in \mathcal{Q}_{m,\varsigma}} \sum_{k \in \{D,L\}} \left( \frac{\hat{C}_{m,i,k,p} - C_{m,i,k,p}}{s_k} \right)^2, \quad (23)$$

$$E_{F_p}^{(\varsigma)} = \frac{1}{|\mathcal{C}_\varsigma|} \sum_{m \in \mathcal{C}_\varsigma} E_{m,\varsigma,F_p},$$

where  $C_{m,i,k,p}$  is pressure-derived component  $k$  for pair  $i$  from simulation  $m$ , and  $s_D = 0.1187$  and  $s_L = 0.7959$  are RMS scales over all 397,440 admissible training pairs.

For either  $E \in \{E_y, E_{F_p}\}$ , band and overall errors are

$$\begin{aligned} E^{\text{ID}} &= \frac{1}{3} \sum_{\varsigma \in \mathcal{S}_{\text{ID}}} E^{(\varsigma)}, \quad E^{\text{OOD}} = \frac{1}{3} \sum_{\varsigma \in \mathcal{S}_{\text{OOD}}} E^{(\varsigma)}, \\ E^{\text{all}} &= \frac{1}{6} \sum_{\varsigma} E^{(\varsigma)}. \end{aligned} \quad (24)$$

Here,  $\mathcal{S}_{\text{ID}}$  and  $\mathcal{S}_{\text{OOD}}$  contain the three regime strata in their respective lead bands. The aggregation gives equal weight to contributing simulations and regime strata, while  $E_y$  also weights the three response channels equally. Paired gain is  $100(E_{\text{Base}} - E_{\text{STCO}})/E_{\text{Base}}$ .

## Extended Results

| Backbone     | Metric | ID-Lead gain (%) |        |          | OOD-Lead gain (%) |        |          | Route-level MCF |        |              |       |          |
|--------------|--------|------------------|--------|----------|-------------------|--------|----------|-----------------|--------|--------------|-------|----------|
|              |        | Pre-event        | Motion | External | Pre-event         | Motion | External | $\tau$          | $\psi$ | $\Delta\psi$ | $g$   | $u_{bc}$ |
| MGN          | Field  | 63.9             | 24.4   | 43.0     | 21.2              | -7.4   | 15.9     | 0.386           | 0.239  | 0.156        | 0.172 | 0.510    |
|              | Load   | 77.3             | 8.4    | 9.6      | 48.2              | 3.9    | 30.5     |                 |        |              |       |          |
| RIGNO        | Field  | 11.2             | 8.2    | 19.1     | 8.8               | 3.8    | 17.3     | 0.318           | 0.343  | 0.178        | 0.270 | 0.526    |
|              | Load   | 11.3             | -3.3   | 2.2      | -11.6             | -8.8   | -2.5     |                 |        |              |       |          |
| PINO         | Field  | 70.1             | 39.1   | 54.0     | 61.2              | 39.1   | 58.1     | 0.211           | 0.339  | 0.186        | 0.401 | 0.458    |
|              | Load   | 83.7             | 33.5   | 18.7     | 55.9              | 33.1   | 52.7     |                 |        |              |       |          |
| Poseidon     | Field  | 35.2             | 6.4    | 36.7     | 21.0              | 4.4    | 41.3     | 0.293           | 0.255  | 0.145        | 0.383 | 0.385    |
|              | Load   | 42.2             | -0.2   | 5.2      | 18.9              | 2.4    | 34.8     |                 |        |              |       |          |
| GAOT         | Field  | 83.2             | 51.5   | 63.5     | 65.3              | 44.5   | 61.3     | 0.232           | 0.225  | 0.209        | 0.505 | 0.455    |
|              | Load   | 88.6             | 28.2   | 15.6     | 63.4              | 24.7   | 52.2     |                 |        |              |       |          |
| GINO         | Field  | 6.3              | 10.2   | 18.0     | 25.6              | 20.3   | 29.4     | 0.254           | 1.298  | 0.216        | 0.569 | 0.539    |
|              | Load   | 31.0             | 8.5    | 3.7      | -13.1             | 21.0   | -1.6     |                 |        |              |       |          |
| Transolver++ | Field  | 80.0             | 48.0   | 58.7     | 33.0              | 23.5   | 38.4     | 0.030           | 0.462  | 0.356        | 0.508 | 0.534    |
|              | Load   | 86.8             | 32.3   | 14.5     | 27.4              | 23.2   | 31.2     |                 |        |              |       |          |
| Unisolver    | Field  | 16.5             | 17.6   | 25.8     | 1.0               | 3.5    | 22.0     | 0.272           | 0.598  | 0.189        | 0.618 | 0.535    |
|              | Load   | 24.1             | 19.1   | 4.0      | 28.3              | 23.2   | 15.3     |                 |        |              |       |          |
| MPP          | Field  | 65.2             | 26.8   | 45.9     | 29.1              | 11.7   | 32.9     | 0.141           | 0.614  | 0.367        | 0.219 | 0.626    |
|              | Load   | 80.0             | 9.5    | 14.1     | 54.6              | 6.3    | 34.1     |                 |        |              |       |          |
| CALM-PDE     | Field  | 39.3             | 15.6   | 34.3     | -51.9             | -10.0  | 3.1      | 0.019           | 0.451  | 0.390        | 0.368 | 0.450    |
|              | Load   | 22.7             | 6.9    | 2.9      | -28.0             | 17.2   | -12.8    |                 |        |              |       |          |
| UPT          | Field  | 84.0             | 53.9   | 62.5     | 33.6              | 28.9   | 34.7     | 0.010           | 0.417  | 0.362        | 0.364 | 0.504    |
|              | Load   | 88.5             | 40.3   | 16.4     | 35.1              | 25.7   | 33.4     |                 |        |              |       |          |
| GEPS         | Field  | 54.9             | 17.3   | 37.4     | 24.2              | -2.9   | 22.8     | 0.356           | 0.389  | 0.224        | 0.236 | 0.630    |
|              | Load   | 61.6             | 5.0    | 8.5      | 25.7              | 9.0    | 23.9     |                 |        |              |       |          |

Table 6: Stratified paired gains and route-level MCF on the predefined 5,004-pair evaluation set. Positive gains indicate lower STCO error, bold marks negative gains, and “External” denotes external onset. STCO improves 68/72 field cells and 63/72 pressure-derived load cells, with mean gains of 31.1% and 24.7%. For MCF,  $\tau$  is replaced by another same-band lead,  $\psi$  and  $\Delta\psi$  are drawn from fixed channelwise Gaussian marginals, and  $g$  and  $u_{bc}$  use active benchmark fields. Deterministic draws are shared across models; unchanged-input repeats define the numerical floor. GAOT has the largest mean field gain, and GINO the largest spatial MCF.

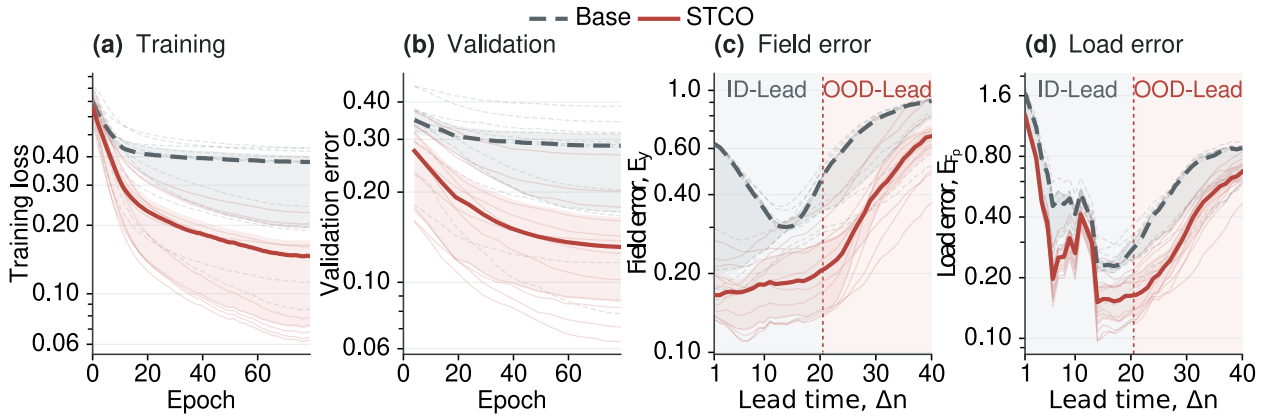


Figure 8: Learning and lead-time dynamics across twelve matched Base-STCO pairs. (a) Training objective and (b) validation relative- $L_2$  error over 80 epochs. (c) Velocity-pressure field error  $E_y$  and (d) pressure-derived load error  $E_{F_p}$  over the predefined 5,004-pair evaluation set. Thin curves show individual backbones; thick curves and bands show the cross-backbone median and interquartile range. Each lead is an independent target-time query. Gray and red backgrounds denote ID-Lead and OOD-Lead. Before smoothing, STCO lowers field error in 466/480 and load error in 454/480 backbone-lead comparisons.

## Prescribed yaw

$\Delta n = 20$  fields;  $\Delta n = 1\text{--}40$  loads

— GT — Base — STCO —  $C_{L,p}$  - - -  $C_{D,p}$

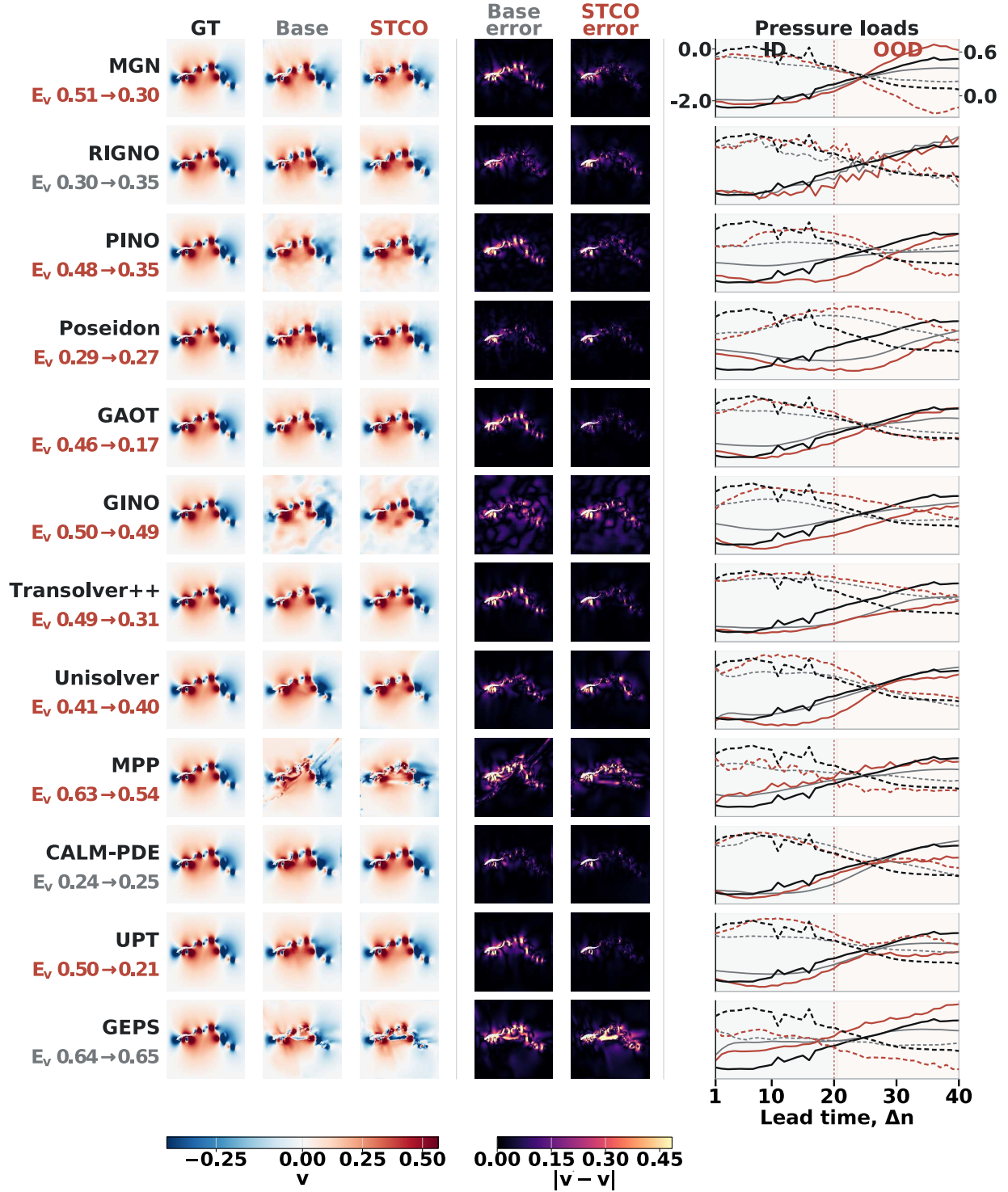


Figure 9: Prescribed-yaw response across twelve matched backbones. At  $n_r = 128$  and  $\Delta n = 20$ , rows show GT, Base/STCO transverse velocity  $v$ , absolute errors, and fluid-region relative- $L_2$  error  $E_v$ , followed by pressure-derived lift (solid) and drag (dashed) over independent target-time queries at  $\Delta n = 1\text{--}40$ . Field, error, and load-axis scales are shared across rows. Gray/red backgrounds mark ID/OOD-Lead. STCO lowers  $E_v$  for 9/12 backbones, with a mean reduction of 18.9%. GAOT and UPT show the largest reductions, 62.8% and 57.5%.



# Transverse gust

$\Delta n = 20$  fields;  $\Delta n = 1\text{--}40$  loads

— GT — Base — STCO —  $C_{L,p}$  - - -  $C_{D,p}$

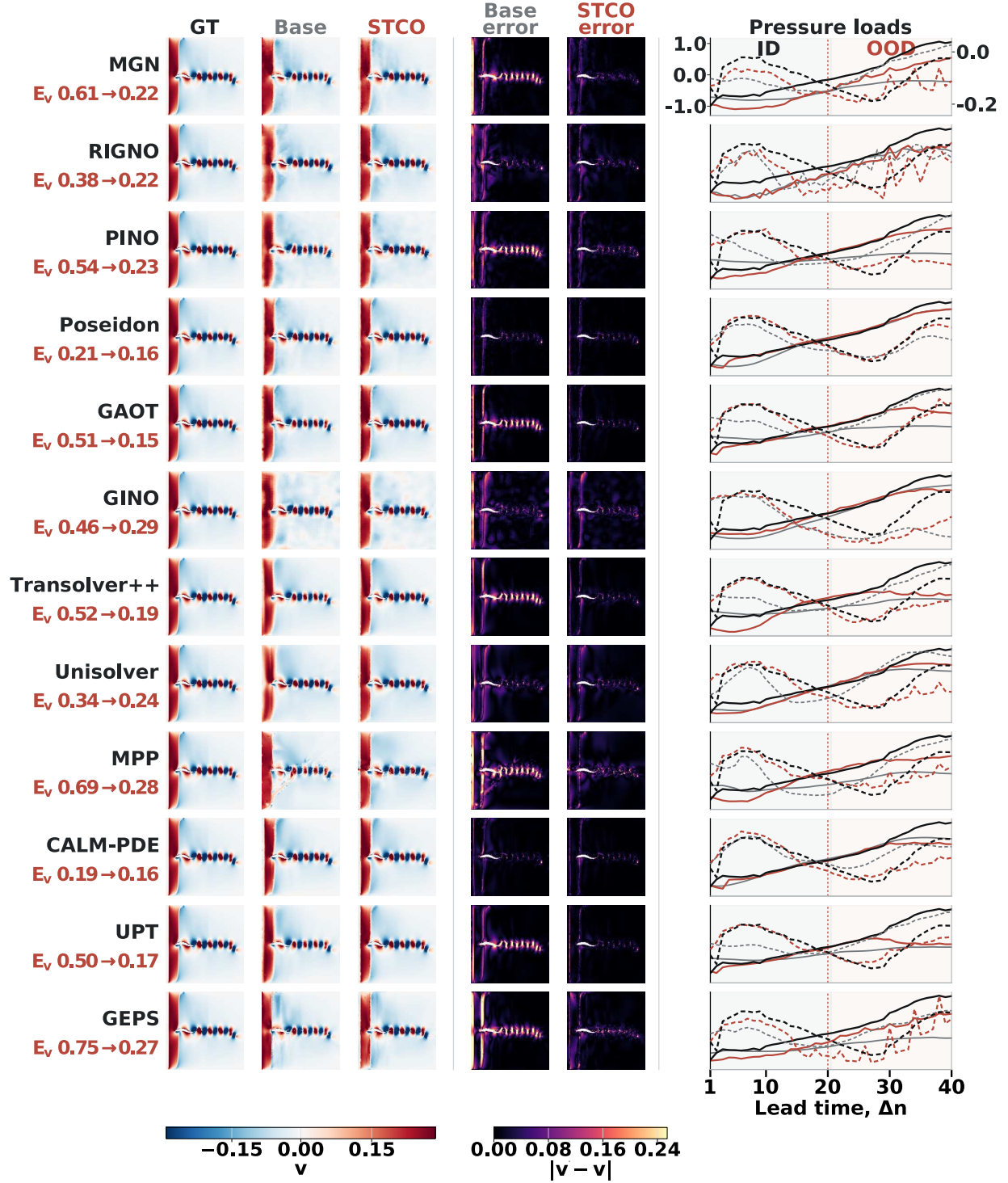


Figure 10: Transverse-gust response across twelve matched backbones at  $n_r = 126$  and  $\Delta n = 20$ . Rows use the field and pressure-derived load layout of Figure 9. The load curves span independent target-time queries at  $\Delta n = 1\text{--}40$ . STCO lowers  $E_v$  for all twelve backbones, with a median reduction of 58.5%. It lowers the pressure-derived load error of Equation 23 over the forty leads for 11/12 backbones, with a median reduction of 34.3%.

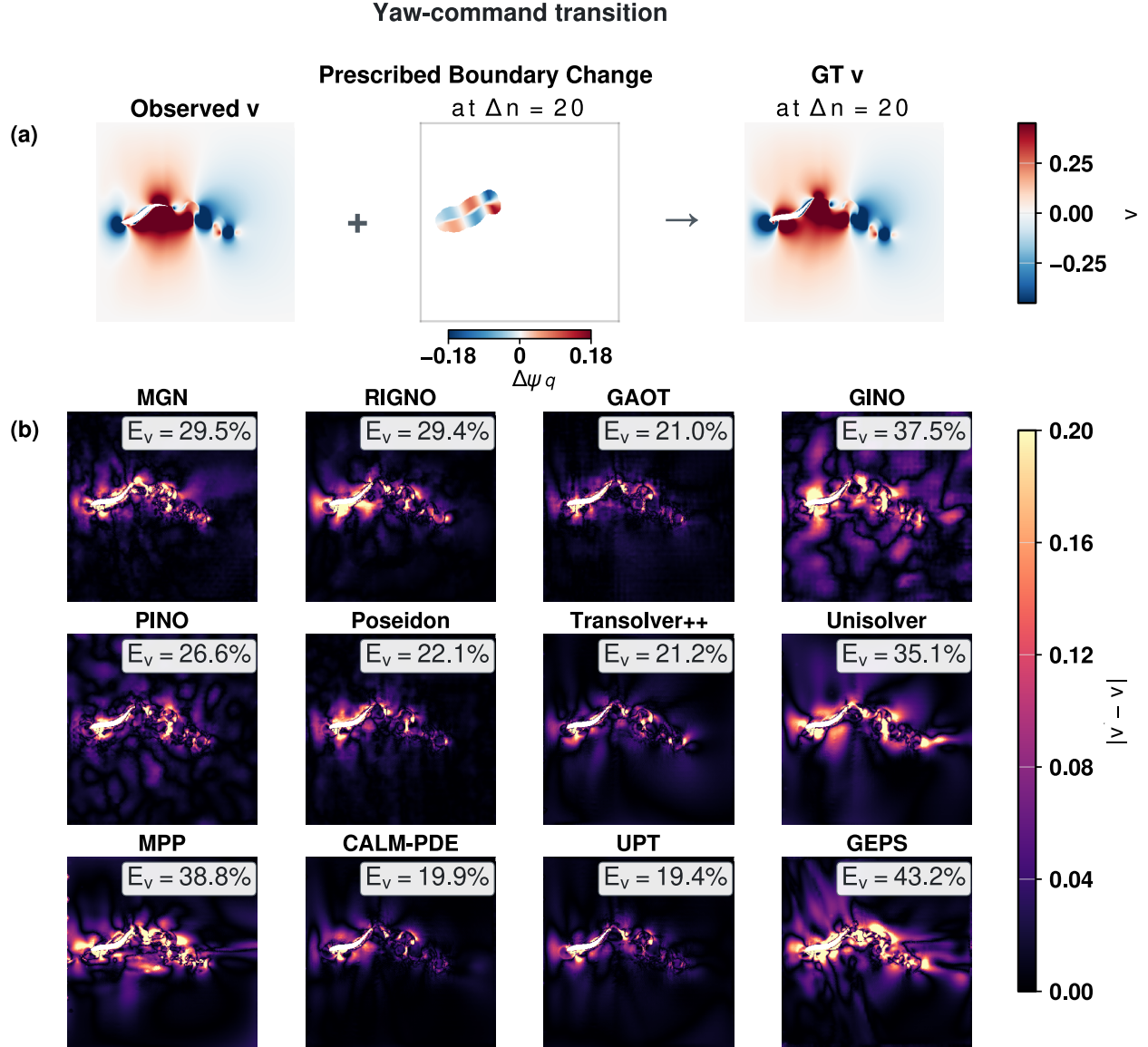


Figure 11: Prescribed-yaw transition across twelve STCO configurations. (a) Transverse velocity  $v$  at the observed frame  $n_r = 44$ , target-time signed-distance change  $\Delta\psi_q$ , and GT  $v$  at  $\Delta n = 20$ . (b) Pointwise absolute errors  $|\hat{v} - v|$ . Labels report full-mesh fluid-region relative- $L_2$  error  $E_v$ . Observed and GT  $v$  share one color limit, and all error panels share another. The median  $E_v$  across the twelve models is 0.280. Errors concentrate near the moving body and wake.

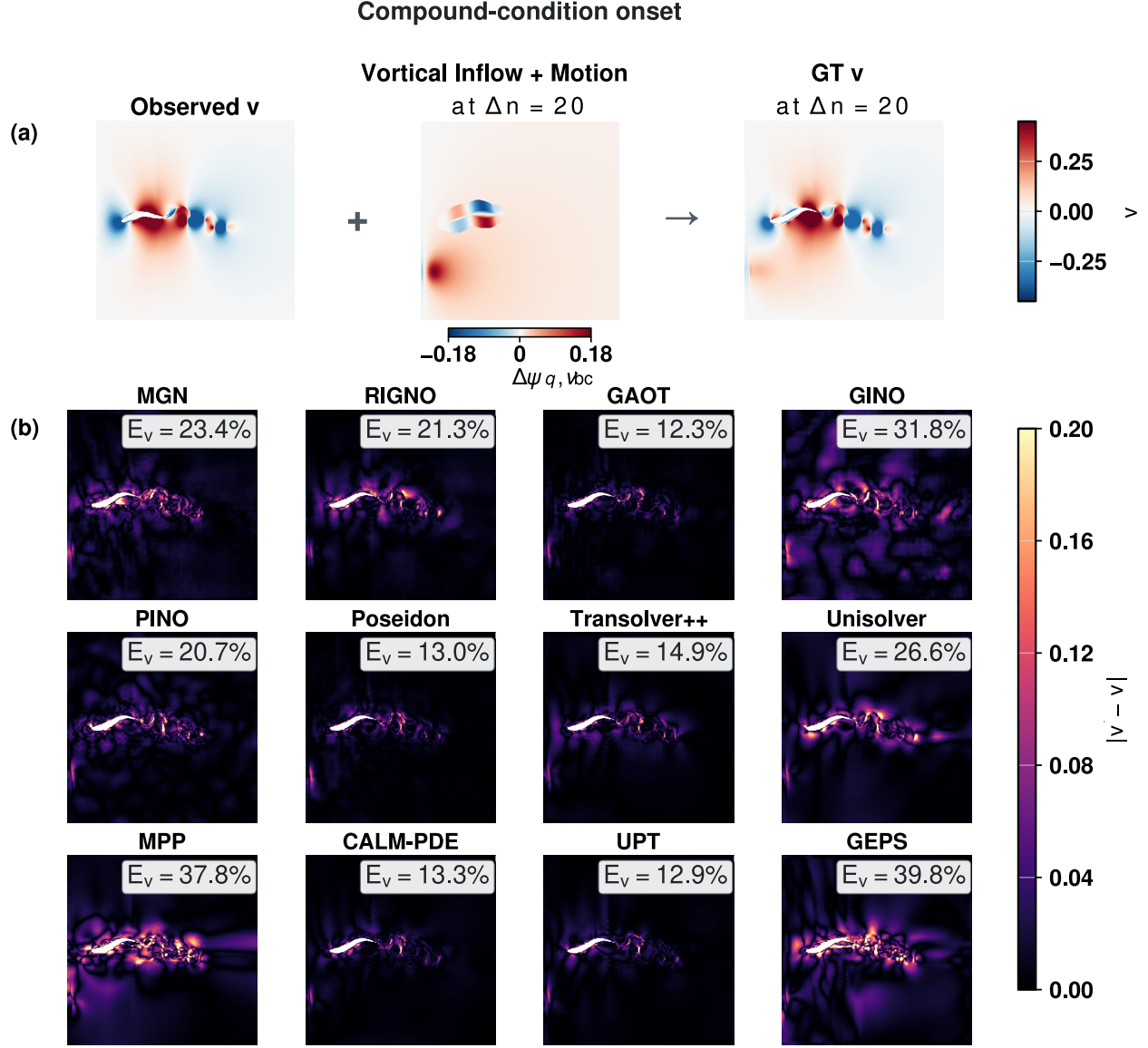


Figure 12: Compound vortical-inflow and motion response across twelve STCO configurations. (a) Transverse velocity  $v$  at the observed frame  $n_r = 22$ , target-time  $v_{bc,q}$  overlaid with  $\Delta\psi_q$ , and GT  $v$  at  $\Delta n = 20$ . The two condition fields share the range  $[-0.18, 0.18]$ . (b) Pointwise absolute errors  $|\hat{v} - v|$ . Labels report full-mesh fluid-region relative- $L_2$  error  $E_v$ . The median  $E_v$  across the twelve models is 0.210. Errors concentrate near the moving body and disturbed wake.

## Acknowledgments

This research is funded by Engineering Start-up Grant of King's College London, and the Daiwa Anglo-Japanese Foundation through Daiwa Foundation Awards (14465/15310). The authors acknowledge the use of King's Computational Research, Engineering and Technology Environment (CREATE) in conducting this research (King's College London 2026).

## References

- Alkin, B.; Fürst, A.; Schmid, S.; Gruber, L.; Holzleitner, M.; and Brandstetter, J. 2024. Universal physics transformers: A framework for efficiently scaling neural operators. In *Advances in Neural Information Processing Systems*, volume 37.
- Andreu-Angulo, I.; Babinsky, H.; Biler, H.; Sedky, G.; and Jones, A. R. 2020. Effect of transverse gust velocity profiles. *AIAA Journal*, 58(12):5123–5133.
- Bartels, R. E. 2013. Developing an accurate CFD based gust model for the truss braced wing aircraft. In *31st AIAA Applied Aerodynamics Conference*, AIAA Paper 2013-3044.
- Bhan, L.; Shi, Y.; and Krstić, M. 2024. Neural operators for bypassing gain and control computations in PDE backstepping. *IEEE Transactions on Automatic Control*, 69(8):5310–5325.
- Felzenszwalb, P. F.; and Huttenlocher, D. P. 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181.
- Hagnberger, J.; Musekamp, D.; and Niepert, M. 2025. CALM-PDE: Continuous and adaptive convolutions for latent space modeling of time-dependent PDEs. In *Advances in Neural Information Processing Systems*, volume 38.
- Hafner, D.; Pasukonis, J.; Ba, J.; and Lillicrap, T. 2025. Mastering diverse control tasks through world models. *Nature*, 640:647–653.
- Han, R.; Zhang, Z.; Wang, Y.; Liu, Z.; Zhang, Y.; and Chen, G. 2021. Hybrid deep neural network based prediction method for unsteady flows with moving boundary. *Acta Mechanica Sinica*, 37(10):1557–1566.
- Hao, Z.; Wang, Z.; Su, H.; Ying, C.; Dong, Y.; Liu, S.; Cheng, Z.; Song, J.; and Zhu, J. 2023. GNOT: A general neural operator transformer for operator learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 12556–12569.
- Herde, M.; Raonić, B.; Rohner, T.; Käppeli, R.; Molinaro, R.; de Bézenac, E.; and Mishra, S. 2024. Poseidon: Efficient foundation models for PDEs. In *Advances in Neural Information Processing Systems*, volume 37.
- Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141.
- Hu, H.; and Liu, C. 2025. Safe PDE boundary control with neural operators. In *Proceedings of the 7th Annual Learning for Dynamics & Control Conference*, volume 283 of *Proceedings of Machine Learning Research*, 513–526.
- Hufstedler, E. A. L.; and McKeon, B. J. 2019. Vortical gusts: Experimental generation and interaction with wing. *AIAA Journal*, 57(3):921–931.
- Hwang, R.; Lee, J. Y.; Shin, J. Y.; and Hwang, H. J. 2022. Solving PDE-constrained control problems using operator learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4):4504–4512.
- Jin, P.; Meng, S.; and Lu, L. 2022. MIONet: Learning multiple-input operators via tensor product. *SIAM Journal on Scientific Computing*, 44(6):A3490–A3514.
- Kamkar, S. J.; Wissink, A. M.; Sankaran, V.; and Jameson, A. 2011. Feature-driven Cartesian adaptive mesh refinement for vortex-dominated flows. *Journal of Computational Physics*, 230(16):6271–6298.
- King's College London. 2026. King's Computational Research, Engineering and Technology Environment (CREATE). Retrieved August 20, 2026, from <https://doi.org/10.18742/rnvf-m076>.
- Kassai Koupai, A.; Mifsut Benet, J.; Yin, Y.; Vittaut, J.-N.; and Gallinari, P. 2024. GEPS: Boosting generalization in parametric PDE neural solvers through adaptive conditioning. In *Advances in Neural Information Processing Systems*, volume 37.
- Kassai Koupai, A.; Le Boudec, L.; Serrano, L.; and Gallinari, P. 2025. ENMA: Tokenwise autoregression for continuous neural PDE operators. In *Advances in Neural Information Processing Systems*, volume 38.
- Li, Z.; Kovachki, N.; Azizzadenesheli, K.; et al. 2021. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*.
- Li, Z.; Huang, D. Z.; Liu, B.; and Anandkumar, A. 2023. Fourier neural operator with learned deformations for PDEs on general geometries. *Journal of Machine Learning Research*, 24(388):1–26.
- Li, Z.; Kovachki, N.; Choy, C.; et al. 2023. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems*, volume 36.
- Li, Z.; Zheng, H.; Kovachki, N.; Jin, D.; Chen, H.; Liu, B.; Azizzadenesheli, K.; and Anandkumar, A. 2024. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27.
- Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; and Karniadakis, G. E. 2021. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229.
- Luo, H.; Wu, H.; Zhou, H.; Xing, L.; Di, Y.; Wang, J.; and Long, M. 2025. Transolver++: An accurate neural solver for PDEs on million-scale geometries. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 41432–41449.
- McCabe, M.; Régaldo-Saint Blancard, B.; Parker, L.; Ohana, R.; Cranmer, M.; et al. 2024. Multiple physics pretraining for spatiotemporal surrogate models. In *Advances in Neural Information Processing Systems*, volume 37.
- Mittal, R.; and Iaccarino, G. 2005. Immersed boundary methods. *Annual Review of Fluid Mechanics*, 37:239–261.
- Mousavi, S.; Wen, S.; Lingsch, L.; Herde, M.; Raonić, B.; and Mishra, S. 2025. RIGNO: A graph-based framework for robust and accurate operator learning for PDEs on arbitrary domains. In *Advances in Neural Information Processing Systems*, volume 38.
- NVIDIA; Agarwal, N.; Ali, A.; Bala, M.; et al. 2025. Cosmos world foundation model platform for physical AI. arXiv:2501.03575.
- Park, T.; Liu, M.-Y.; Wang, T.-C.; and Zhu, J.-Y. 2019. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2337–2346.
- Peebles, W.; and Xie, S. 2023. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4195–4205.

- Perez, E.; Strub, F.; de Vries, H.; Dumoulin, V.; and Courville, A. 2018. FiLM: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1):3942–3951.
- Pfaff, T.; Fortunato, M.; Sánchez-González, A.; and Battaglia, P. 2021. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*.
- Sedky, G.; Biler, H.; and Jones, A. R. 2022. Experimental comparison of a sinusoidal and trapezoidal transverse gust. *AIAA Journal*, 60(5):3347–3351.
- Sedky, G.; Gementzopoulos, A.; Andreu-Angulo, I.; Lagor, F. D.; and Jones, A. R. 2022. Physics of gust response mitigation in open-loop pitching manoeuvres. *Journal of Fluid Mechanics*, 944:A38.
- Sotiropoulos, F.; and Yang, X. 2014. Immersed boundary methods for simulating fluid–structure interaction. *Progress in Aerospace Sciences*, 65:1–21.
- Wen, S.; Kumbhat, A.; Lingsch, L.; Mousavi, S.; Zhao, Y.; Chandrashekar, P.; and Mishra, S. 2025. Geometry-aware operator transformer as an efficient and accurate neural surrogate for PDEs on arbitrary domains. In *Advances in Neural Information Processing Systems*, volume 38.
- Weymouth, G. D.; and Font, B. 2025. WaterLily.jl: A differentiable and backend-agnostic Julia solver for incompressible viscous flow around dynamic bodies. *Computer Physics Communications*, 315:109748.
- Zhong, W.; and Meidani, H. 2025. Benchmarking of neural operators for rigid-body fluid–structure interaction. In *Machine Learning and the Physical Sciences Workshop at NeurIPS*.
- Zhou, H.; Ma, Y.; Wu, H.; Wang, H.; and Long, M. 2025. Uni-solver: PDE-conditional transformers towards universal neural PDE solvers. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 79061–79088.