STABILITY-AWARE PARAMETRIC NEURAL ODES FOR BIFURCATING DYNAMICAL SYSTEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Parametric dynamical surrogates must recover both the stability of an equilibrium and the nonlinear oscillations that emerge when it loses stability. Near a Hopf bifurcation, weak growth or decay can be difficult to resolve alongside the dominant oscillation, while joint fitting allows nonlinear training to alter the equilibrium spectrum. We introduce a two-stage parametric neural ODE that couples shared operator identification with exact Jacobian preservation. A shared matrix polynomial in the operating parameter is identified jointly across training conditions from normalised early-window trajectory error, with stability penalties, and then frozen. A neural coefficient matrix supplies the nonlinear correction through an amplitude gate. The resulting residual is third order at the equilibrium, so the deployed model's equilibrium Jacobian is exactly the identified operator at every parameter value, independently of residual training. We evaluate the construction on a thermoacoustic Rijke tube, a cylinder wake and a flexible aircraft wing. The method recovers the correct number of stability crossings on all three systems and accurately locates their boundaries, with errors within the reference resolution on the wake and wing and within a tenth of the grid spacing on the tube. It combines this stability recovery with accurate prediction of nonlinear trajectories and limit-cycle geometry at unseen parameter conditions. Mean per-seed median test trajectory nRMSE is 0.108, 0.169 and 0.147, respectively. Objective comparisons and architectural ablations justify the choices of trajectory fitting, polynomial operator parameterisation, freezing and gating.

1 INTRODUCTION

A wing operates across angles of attack, a cylinder wake across flow speeds, and a combustor across flame delays. A useful surrogate must capture the resulting family of vector fields and changes in equilibrium stability. At a Hopf bifurcation, disturbances change from decay to growth and can develop into sustained nonlinear oscillations (Guckenheimer & Holmes, 1983; Kuznetsov, 2004). Some families recover stability at a second boundary. The surrogate must therefore capture these transitions, transient growth and nonlinear saturation at unseen parameter values.

Near an oscillatory boundary, the growth rate σ is small relative to the frequency ω, making its derivative contribution weak. Trajectory fitting can resolve accumulated growth, but nonlinear saturation can dominate its loss. Joint fitting also lets nonlinear training change the equilibrium Jacobian and move its stability boundaries. We therefore separate identification of weak growth and decay from nonlinear learning.

We address this challenge with the two-stage construction in Figure 1. First, a shared matrix polynomial in the parameter is identified jointly across training conditions from normalised early-window trajectories with stability-label regularisation, then frozen. Second, a neural residual learns nonlinear evolution through a gate that makes its contribution third order at the equilibrium. The contribution is the coupling of identification across operating conditions with exact preservation of the full equilibrium Jacobian during nonlinear learning. Spectral accuracy is evaluated against the reference systems. The guarantee concerns preservation of the identified operator.

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.

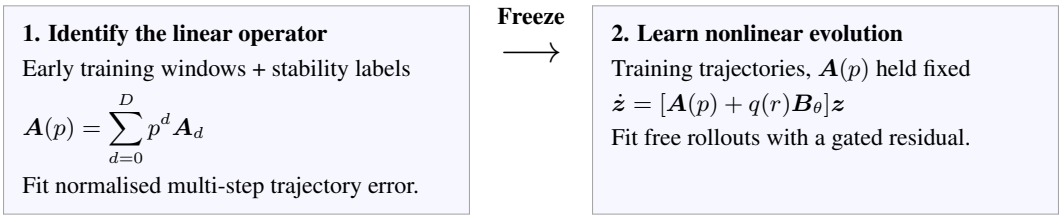


Figure 1: Two-stage construction. The gate makes the nonlinear residual $O(\|\boldsymbol{z}\|^3)$ near the supplied equilibrium. Stage-two training therefore preserves the identified equilibrium Jacobian exactly. Agreement with the physical system is evaluated separately.

Contributions.

- A parametric neural ODE that jointly identifies a shared polynomial equilibrium operator across conditions, then preserves it while learning nonlinear evolution.
- Exact preservation of the identified equilibrium Jacobian (Proposition 1), including its full spectrum and stability crossings.
- Evaluation on three physical systems, recovering correct crossing counts and the lowest neural test trajectory error (tied on the wing). The polynomial retains comparable tube trajectory accuracy and improves it on the wake relative to a frozen neural operator map with 94 and 27 times as many operator parameters, respectively.

2 Related work

Parametric identification. Dynamic mode decomposition (DMD) identifies linear dynamics from snapshots (Schmid, 2010). Parametric DMD (Andreuzzi et al., 2023) and reduced-order interpolation (Amsallem & Farhat, 2008; Benner et al., 2015) extend prediction across operating conditions. Optimised DMD fits exponential dynamics to trajectory snapshots by variable projection (Askham & Kutz, 2018). Its standard formulation does not learn dependence on an operating parameter. Here a shared matrix polynomial is fitted jointly across conditions and preserved as the Jacobian during nonlinear learning. Operator inference fits reduced polynomial operators (Peherstorfer & Willcox, 2016; McQuarrie et al., 2023), while sparse regression selects terms from a library (Brunton et al., 2016; Kaiser et al., 2018), with parameter dependence introduced through monomials or group sparsity (Rudy et al., 2019; Nicolaou et al., 2023; Schaeffer et al., 2017). In equilibrium-centred coordinates, their linear coefficient gives the Jacobian, making them natural comparisons here. Their literature addresses noise-sensitive differentiation, unstable rollouts and stabilising constraints (Messenger & Bortz, 2021; Loiseau & Brunton, 2018; Kaptanoglu et al., 2021; McQuarrie et al., 2021; Sawant et al., 2023; Kramer, 2021; Goyal et al., 2025).

Learned bifurcation models. Forecasting methods estimate flutter boundaries from measured damping (Yamasaki & Epureanu, 2017; García Pérez et al., 2023; 2024) or post-critical oscillations from pre-critical transients (Ghadami & Epureanu, 2018). Learned representations also describe bifurcations through normal forms, topology, reduced dynamics and Koopman coordinates (Kalia et al., 2021; Moriel et al., 2024; Ricci et al., 2025; Lusch et al., 2018), or combine mechanistic and learned components (Lee et al., 2023; Li et al., 2025). The closest parametric flow models combine autoencoders with sparse identification and use continuation to recover bifurcations (Conti et al., 2023; Tomada et al., 2025). Thus recovering bifurcation structure from a learned parametric model is established. Our distinction is identifying and preserving the equilibrium operator during nonlinear learning. We compare parametric AE-SINDy on the cylinder benchmark using both full and cropped records approximating the prior-work setting (Section 4.2, Appendix F.7).

Structured neural dynamics. Neural ODEs learn a vector field directly (Chen et al., 2018; Rubanova et al., 2019). Hybrid and constrained variants add physical structure or learned corrections (Kaheman et al., 2019; Yin et al., 2021; Takeishi & Kalousis, 2021; Levine & Stuart, 2022), including explicit linear components, Lyapunov functions, delays and topological constraints (Linot & Graham, 2022; Manek & Kolter, 2019; Zhu et al., 2021; Tang, 2024). Universal differential equa-

tions (Rackauckas et al., 2020) encompass learned and hybrid fields. Our contribution specifies how a parameter-dependent equilibrium Jacobian is identified and preserved. Linot et al. (2023) combine linear operators with nonlinear neural terms for stabilised forecasting at fixed physical settings, jointly learning both terms or prescribing the linear term. Their models do not condition on an operating parameter. Nonlinear system identification also builds on a best linear approximation (Paduart et al., 2010; Schoukens & Ljung, 2019). For an additive field $\boldsymbol{A}(p)\boldsymbol{z} + \boldsymbol{g}_\theta(\boldsymbol{z}; p)$, fixing $\boldsymbol{A}(p)$ does not force $D_{\boldsymbol{z}}\boldsymbol{g}_\theta(\boldsymbol{0}; p) = \boldsymbol{0}$. We identify $\boldsymbol{A}(p)$ jointly across conditions. Freezing and the gate preserve it as the full equilibrium Jacobian during nonlinear training. Ungated and jointly trained ablations test these choices (Section 4.3).

3 METHOD

Setting and deployed model. The operating parameter $p \in \mathbb{R}$ is scalar and constant within a trajectory. The state $\boldsymbol{z}(t) \in \mathbb{R}^n$ is measured relative to a supplied equilibrium, so that $\boldsymbol{z} = \boldsymbol{0}$ is the equilibrium at every condition. We learn

$$\dot{\boldsymbol{z}} = \boldsymbol{F}_\theta(\boldsymbol{z}; p) = [\boldsymbol{A}(p) + q(r)\, \boldsymbol{B}_\theta(\boldsymbol{f}(p), \boldsymbol{z}/\boldsymbol{s}, r)]\, \boldsymbol{z}, \qquad r = \|\boldsymbol{z}/\boldsymbol{s}\|/\sqrt{n}, \tag{1}$$

where $\boldsymbol{A}(p)$ is an identified linear operator and $\boldsymbol{B}_\theta$ is a neural coefficient matrix. The vector $\boldsymbol{s}$ contains fixed positive state scales, division is componentwise, and $\boldsymbol{f}(p)$ contains centred parameter powers. The amplitude gate is

$$q(r) = \frac{r^2}{r^2 + r_0^2}, \qquad r_0 > 0. \tag{2}$$

It suppresses the residual near the equilibrium and approaches one at large amplitude. The coefficient matrix depends on the state, so the complete vector field is nonlinear.

Stage 1: identify the equilibrium operator. We use a matrix polynomial,

$$\boldsymbol{A}(p) = \sum_{d=0}^{D} p^d \boldsymbol{A}_d, \qquad \boldsymbol{A}_d \in \mathbb{R}^{n\times n}, \tag{3}$$

with degree D fixed per system (Table 2). Its coefficients are fitted jointly across the K training conditions:

$$\min_{\{\boldsymbol{A}_d\}} \frac{1}{K} \sum_{k=1}^{K} \frac{\| \exp(\boldsymbol{A}(p_k)t)\boldsymbol{z}_0^{(k)} - \boldsymbol{z}(t; p_k)\|^2_{L^2(0,T_k)}}{\|\boldsymbol{z}(t; p_k)\|^2_{L^2(0,T_k)}} + \lambda_{\text{stab}}\, \mathcal{L}_{\text{stab}}. \tag{4}$$

Here $\boldsymbol{z}(t; p_k)$ is the recorded trajectory from initial state $\boldsymbol{z}_0^{(k)}$, and $\exp(\boldsymbol{A}(p_k)t)\boldsymbol{z}_0^{(k)}$ is the linear model's prediction (Moler & Van Loan, 2003). The norm is evaluated at the sampling instants of the window $[0, T_k]$. Each record is normalised by its own squared size so that small-amplitude conditions retain weight. Windows end before nonlinear saturation. Their system-specific rules and numerical lengths are given in Appendix B.1.

Trajectory fitting is an established identification approach (Askham & Kutz, 2018). Its sensitivity to weak growth motivates its use here. Consider $\boldsymbol{z}(t) = e^{\sigma t}(\cos \omega t, \sin \omega t)$ near an oscillatory boundary. At a fixed observed state, perturbing the growth rate by δ changes the predicted derivative by relative order $|\delta|/|\omega|$, but changes the trajectory after time T by order $|\delta|T$ for $|\delta|T \ll 1$. The sensitivity ratio scales as $|\omega|T$, proportional to the number of observed cycles. Objective and window-length comparisons test this empirically (Section 4.3).

The basic penalty on stable training records is

$$\mathcal{L}_{\text{stab}} = \sum_{k\in\mathcal{S}} \max\Big(0,\ \max_i \operatorname{Re} \lambda_i\big(\boldsymbol{A}(p_k)\big) + m\Big)^2, \tag{5}$$

where $\mathcal{S}$ indexes stable training records, λ_i denotes an eigenvalue and $m > 0$ is a fixed margin. Stability labels derived from training records select the penalty conditions. Removing the penalties gives the trajectory control in Table 5. System-specific weights and the wake's separate oscillatory- and real-mode penalties are specified in Appendix B.

Stage 2: learn the gated nonlinear residual. With $\boldsymbol{A}(p)$ frozen, a fully connected network with three tanh hidden layers emits the $n \times n$ matrix $\boldsymbol{B}_\theta$. Its weights minimise the trajectory error of free rollouts of equation 1, together with the system-specific auxiliary losses in Appendix B. The fixed radius r_0 sets the amplitude at which the gate reaches half strength. The matrix representation is not unique away from the equilibrium, and no spectral interpretation is claimed there.

The following direct consequence of the gate states the preservation property used in the evaluation.

Proposition 1 (Equilibrium-Jacobian preservation)**.** *Let $\boldsymbol{F}_\theta(\cdot\,; p)$ be the deployed vector field of equation 1, and write $D_{\boldsymbol{z}}\boldsymbol{F}_\theta(\boldsymbol{0}; p)$ for its Jacobian at the equilibrium. Assume $\boldsymbol{B}_\theta$ is bounded in a neighbourhood of $\boldsymbol{z} = \boldsymbol{0}$ and that $q(r) = O(r^2)$ as $r \to 0$, both of which hold for a tanh network and for equation 2. No differentiability of $\boldsymbol{B}_\theta$ is assumed. Then the residual term of equation 1 is $O(\|\boldsymbol{z}\|^3)$ and*

$$D_{\boldsymbol{z}}\boldsymbol{F}_\theta(\boldsymbol{0}; p) = \boldsymbol{A}(p) \qquad \textit{for every } p.$$

Hence the deployed model preserves the frozen operator's entire equilibrium spectrum, including the locations and count of its stability crossings.

Proof. Fix p and write $\boldsymbol{R}_\theta(\boldsymbol{z}; p) = \boldsymbol{F}_\theta(\boldsymbol{z}; p) - \boldsymbol{A}(p)\boldsymbol{z}$. Fixed positive scales imply $r \le c_s\|\boldsymbol{z}\|$. Local boundedness of $\boldsymbol{B}_\theta$ and $q(r) = O(r^2)$ therefore give $\|\boldsymbol{R}_\theta(\boldsymbol{z}; p)\| \le C_p\|\boldsymbol{z}\|^3$ near the origin. Since $\boldsymbol{R}_\theta(\boldsymbol{0}; p) = \boldsymbol{0}$, $\|\boldsymbol{R}_\theta(\boldsymbol{z}; p)\|/\|\boldsymbol{z}\| \to 0$ as $\boldsymbol{z} \to \boldsymbol{0}$. Thus $D_{\boldsymbol{z}}\boldsymbol{R}_\theta(\boldsymbol{0}; p) = \boldsymbol{0}$ and $D_{\boldsymbol{z}}\boldsymbol{F}_\theta(\boldsymbol{0}; p) = \boldsymbol{A}(p)$. □

The guarantee preserves the *identified* Jacobian through residual training. Agreement with the physical system's spectrum is an empirical question evaluated in Section 4.2. The proposition fixes the operator's crossing locations and count. It does not determine whether a Hopf bifurcation is supercritical or subcritical, which depends on nonlinear terms (Appendix B.7).

4 RESULTS

Systems and protocol. The Rijke tube permits comparison against characteristic roots of known delay equations. The cylinder wake connects the evaluation to parametric AE-SINDy, while the flexible wing tests coupled nonlinear aeroelastic dynamics (Table 1). Each record starts from a disturbance of a supplied equilibrium. The tube and wing lose and then recover stability. The wake loses stability once (Appendix A). Wake and wing POD bases use training snapshots only. All conditions are projected onto these fixed bases. Development conditions select hyperparameters, and test conditions are reserved for evaluation. Figures show the reported parameter grid with training conditions marked. Test-only trajectory errors are reported separately. The stable-records-only boundary extrapolation experiment is in Appendix F.8.

Table 1: Evaluation setup. Counts refer to the reported parameter range. Training, development and test conditions are disjoint. The full grids, reference crossing brackets and data-generation details are in Appendix A.

System	States	Parameter	Train / dev. / test	True crossings
Rijke tube	6	τ	14 / 4 / 37	2
Cylinder wake	8	Re	19 / 4 / 44	1
Flexible wing	9	α (deg)	20 / 4 / 35	2

Training. Stage one uses L-BFGS on early windows before saturation. Stage two uses AdamW and a rollout-horizon curriculum (Table 2). Neural comparisons share the Runge–Kutta integrator and loss implementation, with curriculum and tuning differences detailed in Appendix B.3. The proposed model's stage-two stability penalty is constant because its operator is frozen.

Comparisons. The external baselines are parametric SINDy, AE-SINDy following Conti et al. (2023) and Tomada et al. (2025), delay-augmented SINDy, and operator inference (OpInf), evaluated on the same records, coordinates and splits in their own formulations (Appendix B.4). A controlled sequence of neural architectures tests the plain neural ODE, coefficient-matrix form, free neural operator and proposed polynomial operator. Three further ablations remove the gate, release the frozen operator, or train both components jointly from scratch (Appendix B.3).

Table 2: Reported model and training settings. D is the operator's polynomial degree, r_0 is the gate radius, and rollout horizons are step counts in curriculum order. Detailed window rules, losses, optimisation schedules and selection provenance are in Appendix B.

Setting	Rijke tube	Cylinder wake	Flexible wing
Polynomial degree D	1	4	2
Gate radius r_0	0.3	0.2	0.45
Residual parameters	39,076	42,944	45,137
Rollout horizons	100, 300, 500, 800, 1600	48, 160, 480, 1600, 3200	300, 800, 1600, 3000, 3000, 3000, 3000

Neural results use 5 random seeds, except the wing's jointly trained and frozen free operators, which use four and three, respectively.

Metrics. Trajectory nRMSE is the free-rollout root-mean-square error relative to the reference trajectory's root-mean-square magnitude. Remaining at the equilibrium scores one. A model's growth rate is the largest real part of its equilibrium Jacobian's eigenvalues. A crossing is a sign change of that rate. Crossing error is the distance from a reference crossing to the nearest predicted one, normalised by the unstable interval width (tube, wing) or critical Reynolds number (wake), in percent. Crossing counts are reported alongside location errors so that spurious or missing crossings remain visible (Appendices B.5 and C).

4.1 Trajectory and limit-cycle prediction

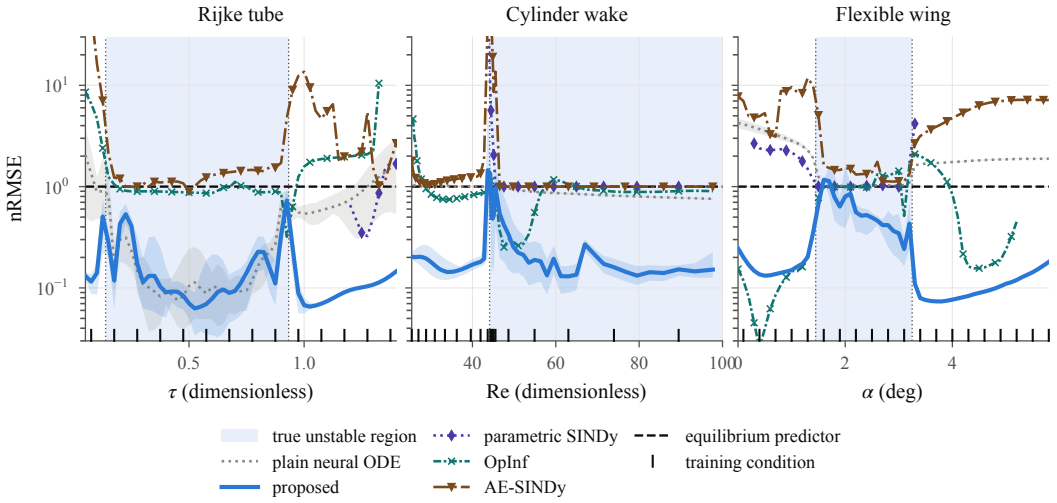


Figure 2: Trajectory prediction across the parameter range. nRMSE is shown on a logarithmic axis, where lower is better. Bands span the neural training seeds. AE-SINDy is the mean over seeds, and other external methods use their reported configuration. Shading marks the true unstable interval, bottom ticks mark training conditions, and the dashed line is the equilibrium predictor. Values above the axis limit are clipped. Failed or unavailable rollouts are not connected. Architectural comparisons are in Figure 19.

Figure 2 compares trajectory errors in stable and unstable regimes. The proposed model's mean per-seed median test nRMSE is 0.108, 0.169 and 0.147 for the tube, wake and wing, respectively, the lowest among the neural architectures, with a tie to the ungated model on the wing (Table 3). Regime-wise comparisons pooled across all conditions and seeds, including two cells slightly favouring ablations, are reported in Table 9.

The frozen free operator shares the Jacobian-preservation guarantee. It matches the proposed model on the tube within seed variation, has higher test error on the wake, and fails the boundedness criterion at most wing conditions. Its operator map uses 94 and 27 times as many scalar parameters as the polynomial on the tube and wake, respectively. These ratios concern the operator, not the

Table 3: Generalisation to unseen parameter conditions. Test-only trajectory nRMSE (lower is better), excluding training and development conditions: 37 tube, 44 wake and 35 wing conditions. Entries are medians over test conditions, averaged over 5 seeds (four and three for the wing's jointly trained and frozen free operators). The proposed row is bold. An equilibrium predictor scores one. The complete statistics are in Table 8.

Architecture	Rijke tube	Cylinder wake	Flexible wing
Plain neural ODE	0.441	1.001	1.804
with the full recipe	1.265	1.108	1.811
Coefficient-matrix form	0.293	0.591	1.021
Free operator, trained jointly	0.303	0.672	7.945
Free operator, frozen	0.115	0.233	66.406
Proposed	**0.108**	**0.169**	**0.147**
Proposed, without the gate	0.270	0.430	0.147
Proposed, operator released	0.147	0.375	0.534
Proposed, trained jointly from scratch	0.489	1.230	6.347

complete model. Errors peak near crossings, where weak growth or decay changes the full record (Appendix D).

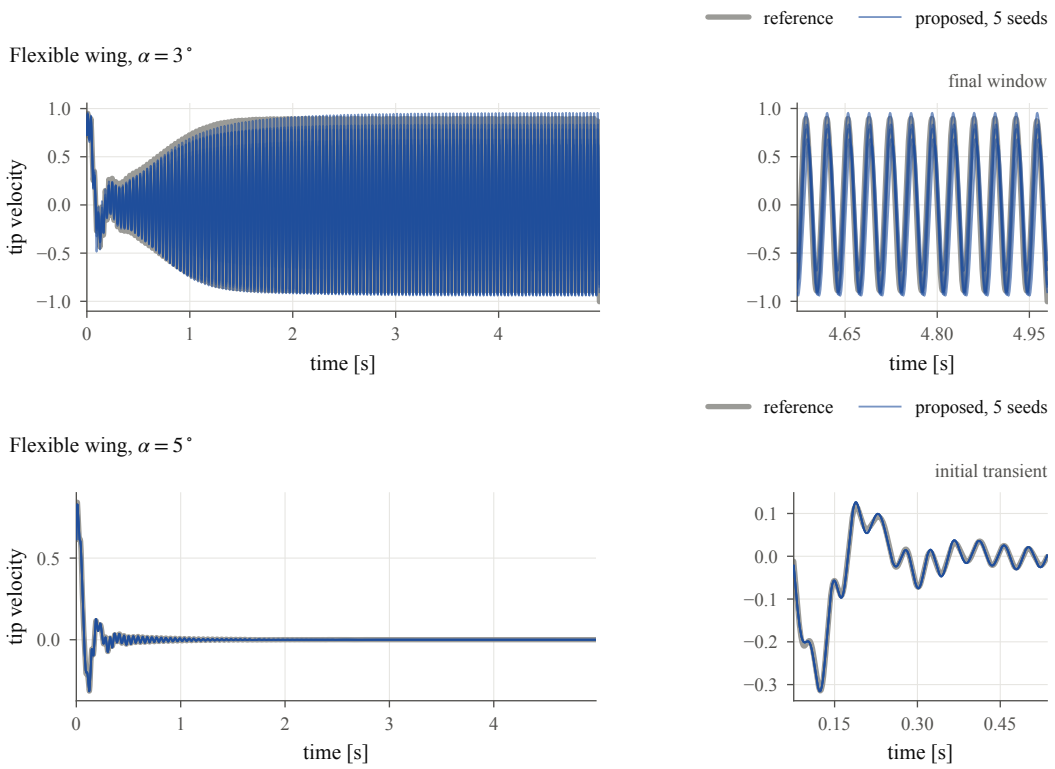


Figure 3: Free rollouts at two test angles of the flexible wing: $\alpha = 3.0^\circ$, within the flutter interval (top), and $\alpha = 5.0^\circ$, in the stable regime above it (bottom). Left: tip velocity over the complete record. Right: final oscillations at 3.0° and the initial decay at 5.0°. Grey denotes the reference and blue all 5 seeds of the proposed model. Neither angle is used for training or development.

Test-wing rollouts reproduce sustained oscillations at 3.0° and stable decay at 5.0° (Figure 3). Growth is slower than the reference near onset, contributing to the error peak in Figure 2. Appendix D shows additional test conditions, all 5 seeds and wake field reconstructions.

The settled orbits show nonlinear geometry across the parameter range (Figure 4). The proposed model follows the reference amplitude on the tube and wake. On the wing, its median amplitude and rollout frequency errors, averaged over seeds, are 17.6 and approximately 0.12 percent. The other members of the architectural sequence miss the amplitude by 77 to 120 percent (Table 11).

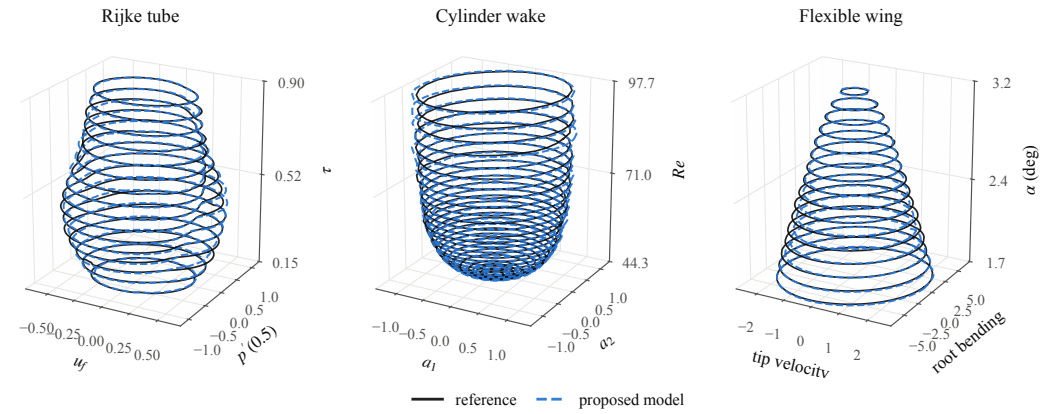


Figure 4: Phase-averaged limit-cycle geometry. Black curves show the reference. Blue dashed curves show one seed of the proposed model. Each ring is the settled tail of a free rollout at one parameter value, projected onto two observables and phase-averaged, including both training and unseen conditions. This comparison assesses orbit shape and amplitude. Figure 2 assesses the full transient trajectory, including accumulated phase error.

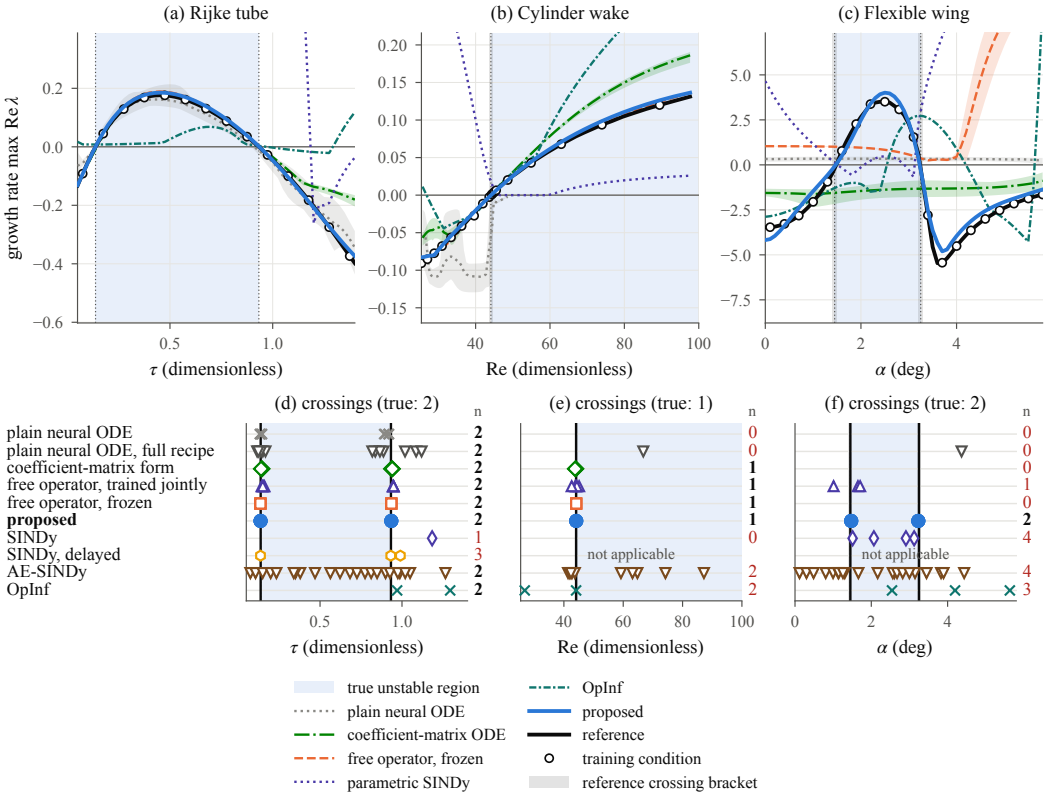


Figure 5: Stability prediction. Top: model growth rates, reference curves (black) and training conditions (circles), with the reference unstable region shaded. Bands span reported seeds (Table 4). Appendix B.5 describes the reference estimators and interpolated wing values. Bottom: every seed's predicted crossings against reference locations (vertical lines) and brackets (grey strips). Counts at the right are medians (upper medians for even seed counts), bold when correct.

4.2 Stability Recovery

Among the models in Table 4, the proposed model alone recovers the correct number of stability crossings on all three systems (Figure 5). Its crossing errors are at most 1.26 percent. On the wake

and wing the predicted crossings lie within the reference resolution. On the tube, where the reference is exact, they lie within a tenth of the grid spacing (Appendix F.4). Both wing crossings fall within the correct 0.1° reference brackets, between training angles 0.3° apart. The top row of Figure 5 also compares growth-rate profiles across the range. On the wake and wing the reference curves are finite-window estimates subject to the qualifications in Appendix B.5.

Location accuracy alone can conceal an incorrect stability structure. OpInf's nearest wake crossing has error < 0.01 percent, below the reference resolution, but it predicts 2 crossings there. The matched-projection SINDy and AE-SINDy baselines also miss the count on at least one system. AE-SINDy's wake crossing locations vary with the seed.

Comparison with parametric AE-SINDy. On the cylinder benchmark, Conti et al. (2023) use 19 training Reynolds numbers and two tests, $Re = 43, 55$, after discarding the first 15 s. They show trajectory overlays and spatial errors, but no aggregate trajectory-error metric. At the stable test condition, their Figure 9(a,b) shows the predicted residual oscillation decaying faster than the reference. The article links no public implementation or dataset and offers data on request. Our independent, development-selected implementation (Appendix B.4) is evaluated on complete disturbance-response records at 44 test conditions and on a larger cropped-record sweep approximating their later-time setting. The proposed model has lower trajectory error under both protocols. On the latter, its mean seed-median finite-error nRMSE is 0.12 against 0.91 for AE-SINDy, although it predicts 3 crossings instead of one (Appendix F.7). These comparisons use our simulator and records.

With only stable training records on both sides of the unstable interval, the tube operator locates crossings within 0.117–0.836 percent of its width, versus 1.60–7.89 percent distance to the nearest record. Appendix F.8 reports all three systems and comparison fits.

Table 4: Stability prediction. Crossing error in percent at the lower and upper edge of the unstable interval (tube, wing) or at the single crossing (wake), sign score in percent, and crossing count n against a true two, one and two. *None*, no seed returns the crossing, n/a not defined for the system, a dash not recorded or, for the sign, undefined because the field has no equilibrium at most conditions. Every row is read at its own equilibrium. Blocks are the neural architectures (mean of 5 seeds, four and three for the wing's free-operator rows), the existing methods (AE-SINDy the mean of 5 seeds), two linear fits to the stage-one windows, and the better constant predictor.

	Rijke tube				Cylinder wake			Flexible wing			
method	lower	upper	sign	n	crossing	sign	n	lower	upper	sign	n
plain neural ODE	0.58	2.99	97.5	2	none	—	0	none	none	31.5	0
with the full recipe	2.00	10.27	92.0	2	51.35	—	0	162.00	62.00	—	0
coefficient-matrix form	0.83	0.83	100.0	2	0.33	100.0	1	none	none	69.5	0
free operator, trained jointly	1.81	1.82	98.5	2	1.40	97.6	1	17.73	105.04	35.2	1
free operator, frozen	0.20	0.28	100.0	2	0.08	100.0	1	none	none	30.5	0
proposed	**0.08**	**0.25**	**100.0**	**2**	**0.09**	**100.0**	**1**	**1.26**	**1.04**	**100.0**	**2**
SINDy	131.61	31.61	74.5	1	none	29.9	0	3.34	7.05	18.6	4
with delayed terms	0.13	0.14	69.1	3	n/a	n/a	n/a	n/a	n/a	n/a	n/a
AE-SINDy	32.43	33.03	50.5	2	3.79	73.9	2	26.88	18.45	40.7	4
OpInf	104.73	4.73	81.8	2	< 0.01	92.5	2	60.71	39.29	62.7	3
parametric linear control, stage-one windows	none	none	58.2	0	none	46.3	0	37.57	32.76	78.0	2
linear OpInf, same windows	137.23	37.23	70.9	1	4.16	89.6	2	68.50	—	—	3
better constant predictor	—	—	*58.2*	0	—	*53.7*	0	—	—	*69.5*	0

4.3 Mechanism and ablations

Trajectory objective and stability information. Table 5 compares first-stage objectives. Trajectory fitting retains the correct crossing counts without stability penalties. The penalty improves the wing's lower crossing. The wake's penalised derivative fit differs in degree and initialisation (Appendix E.5). Separately, our identified operators give better boundary recovery and higher test sign accuracy than the tested parametric optimised-DMD constructions on all three systems (Table 13). These compare first-stage procedures, including parameter coupling and regularisation. Nonlinear residuals were not trained on the alternative operators.

Table 5: First-stage controls: crossing errors (%) and predicted count n. The true counts are two, one and two. Rows use the full fitting windows. The symbols $+$ and $-$ denote inclusion and removal of the stability penalty on stable training records. The wake uses degree four except the penalised derivative fit, marked †, which has effective degree two and uses its own initialisation. These are independent refits under the reported recipe. Full sign scores and fitting details are in Table 12.

	Rijke tube			Cylinder wake		Flexible wing		
Objective / penalty	lower	upper	n	crossing	n	lower	upper	n
Trajectory / $+$	0.17	0.18	2	0.15	1	1.39	0.97	2
Trajectory / $-$	0.08	0.34	2	0.04	1	8.40	1.28	2
Derivative / $+$	8.15	108.15	1	$0.93^{\dagger}$	1	2.87	3.56	2
Derivative / $-$	137.23	37.23	1	4.16	2	68.50	—	3

Freezing and gating. For a given first-stage operator, Proposition 1 guarantees that residual training cannot move the equilibrium spectrum. Repeated first-stage fits return the same operator. Ablations separate these two observations by allowing the second stage to change the Jacobian.

Removing the gate introduces seed variation in the crossing: 0.102 percent on the tube and 0.212 in Reynolds number on the wake. On the wing only the upper crossing moves. Releasing the operator or training both parts jointly also changes the boundaries (Figure 24 and Table 17). Removing the gate raises stable-regime trajectory error from 0.10 to 0.32 on the tube and from 0.17 to 0.45 on the wake. On the wing the gated and ungated models tie in pooled median trajectory error. The gated model additionally guarantees Jacobian preservation. Releasing the operator can improve settled amplitude at the cost of crossing accuracy (Table 11). First-stage sensitivity to initialisation and training data is reported in Appendix G, which distinguishes crossing ranges, deviations from a reference fit and errors against the physical reference.

5 Scope and limitations

Experiments cover Hopf bifurcations in noise-free simulations with a supplied equilibrium at each condition. Stability labels derived from training records regularise identification. They are not required for the architecture or Jacobian preservation. Short or noisy records can obscure these labels, and noise robustness remains untested. Only a scalar operating parameter is tested. The cubic residual excludes quadratic terms at the supplied equilibrium in the chosen coordinates, including the wing's, although it can approximate their effects away from equilibrium. The fixed equilibrium branch excludes its creation or destruction in a saddle-node bifurcation. The missing quadratic term excludes the usual transcritical normal form, while pitchfork bifurcations are representable.

6 Conclusion

We presented a parametric neural ODE that couples shared identification of the equilibrium operator with exact preservation during nonlinear learning. Freezing the polynomial operator and gating the residual preserve the full equilibrium Jacobian at every parameter value, so learning nonlinear motion cannot move the identified stability boundaries.

Across a thermoacoustic tube, a cylinder wake and a coupled aeroelastic wing, the proposed model is the only one in Table 4 to recover the correct number of crossings on all three systems. It accurately locates these boundaries and reproduces nonlinear trajectories and limit-cycle geometry at unseen conditions, with the lowest test trajectory error among the compared neural architectures, with a tie on the wing.

The polynomial operator retains comparable tube accuracy and improves wake accuracy relative to a frozen neural operator map with 94 and 27 times as many operator parameters, respectively. The resulting surrogate combines accurate nonlinear prediction with a compact, directly inspectable equilibrium operator. Its identified stability structure remains intact throughout nonlinear learning.

Ethics statement

This work develops and evaluates a modelling method on simulated physical systems. It involves no human subjects, personal data or sensitive information, and the data are produced by numerical solvers. The authors have read and adhere to the ICLR Code of Ethics and see no ethical concerns arising from this work.

Reproducibility statement

The quantitative figures and tables are prepared from stored run outputs and result summaries. The appendix describes the three systems, the training settings and the model selection procedure. Code and data will be released upon publication.

Use of large language models

Large language models were used to assist with drafting and polishing the text, preparing figures, checking citations, and debugging code. The authors reviewed all content and take full responsibility for it.

REFERENCES

David Amsallem and Charbel Farhat. Interpolation method for adapting reduced-order models and application to aeroelasticity. *AIAA Journal*, 46(7):1803–1813, 2008.

Francesco Andreuzzi, Nicola Demo, and Gianluigi Rozza. A dynamic mode decomposition extension for the forecasting of parametric dynamical systems. *SIAM Journal on Applied Dynamical Systems*, 22(3):2432–2458, 2023.

Travis Askham and J. Nathan Kutz. Variable projection methods for an optimized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 17(1):380–416, 2018.

Or Avin, Daniella E. Raveh, Ariel Drachinsky, Yaron Ben-Shmuel, and Moshe Tur. Experimental aeroelastic benchmark of a very flexible wing. *AIAA Journal*, 60(3):1745–1768, 2022.

Koushik Balasubramanian and R. I. Sujith. Thermoacoustic instability in a Rijke tube: Non-normality and nonlinearity. *Physics of Fluids*, 20(4):044103, 2008.

Dwight Barkley. Linear analysis of the cylinder wake mean flow. *Europhysics Letters*, 75(5):750–756, 2006.

Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

Kathleen Champion, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116 (45):22445–22451, 2019.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems 31*, 2018.

Paolo Conti, Giorgio Gobat, Stefania Fresca, Andrea Manzoni, and Attilio Frangi. Reduced order modeling of parametrized systems through autoencoders and SINDy approach: continuation of periodic solutions. *Computer Methods in Applied Mechanics and Engineering*, 411:116072, 2023.

Alfonso del Carre, Arturo Muñoz-Simón, Norberto Goizueta, and Rafael Palacios. SHARPy: A dynamic aeroelastic simulation toolbox for very flexible aircraft and wind turbines. *Journal of Open Source Software*, 4(44):1885, 2019.

Ariel Drachinsky, Or Avin, Daniella E. Raveh, Yaron Ben-Shmuel, and Moshe Tur. Flutter tests of the Pazy wing. *AIAA Journal*, 60(9):5414–5421, 2022.

Alexander J. Eaton, Chris Howcroft, Etienne B. Coetzee, Simon A. Neild, Mark H. Lowenberg, and Jonathan E. Cooper. Numerical continuation of limit cycle oscillations and bifurcations in high-aspect-ratio wings. *Aerospace*, 5(3):78, 2018.

Jesús García Pérez, Leonardo Sanches, Amin Ghadami, Guilhem Michon, and Bogdan I. Epureanu. A physics-informed data-driven approach for forecasting bifurcations in dynamical systems. *Nonlinear Dynamics*, 111(13):11773–11789, 2023.

Jesús García Pérez, Amin Ghadami, Leonardo Sanches, Bogdan I. Epureanu, and Guilhem Michon. Data-driven bifurcation analysis of experimental aeroelastic systems using preflutter measurements. *AIAA Journal*, 62(5):1906–1914, 2024.

Amin Ghadami and Bogdan I. Epureanu. Forecasting critical points and post-critical limit cycles in nonlinear oscillatory systems using pre-critical transient responses. *International Journal of Non-Linear Mechanics*, 101:146–156, 2018.

Irina Ginzburg, Frederik Verhaeghe, and Dominique d'Humières. Two-relaxation-time lattice Boltzmann scheme: about parametrization, velocity, pressure and mixed boundary conditions. *Communications in Computational Physics*, 3(2):427–478, 2008.

Pawan Goyal, Igor Pontes Duff, and Peter Benner. Guaranteed stable quadratic models and their applications in SINDy and operator inference. *Physica D: Nonlinear Phenomena*, 483:134893, 2025.

John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, 1983.

C. P. Jackson. A finite-element study of the onset of vortex shedding in flow past variously shaped bodies. *Journal of Fluid Mechanics*, 182:23–45, 1987.

Matthew P. Juniper and R. I. Sujith. Sensitivity and nonlinearity of thermoacoustic oscillations. *Annual Review of Fluid Mechanics*, 50:661–689, 2018.

Kadierdan Kaheman, Eurika Kaiser, Benjamin Strom, J. Nathan Kutz, and Steven L. Brunton. Learning discrepancy models from experimental data. *arXiv preprint arXiv:1909.08574*, 2019.

Eurika Kaiser, J. Nathan Kutz, and Steven L. Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A*, 474(2219): 20180335, 2018.

Manu Kalia, Steven L. Brunton, Hil G. E. Meijer, Christoph Brune, and J. Nathan Kutz. Learning normal form autoencoders for data-driven discovery of universal, parameter-dependent governing equations. *arXiv preprint arXiv:2106.05102*, 2021.

Alan A. Kaptanoglu, Jared L. Callaham, Aleksandr Aravkin, Christopher J. Hansen, and Steven L. Brunton. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *Physical Review Fluids*, 6(9):094401, 2021.

Boris Kramer. Stability domains for quadratic-bilinear reduced-order models. *SIAM Journal on Applied Dynamical Systems*, 20(2):981–996, 2021.

Yuri A. Kuznetsov. *Elements of Applied Bifurcation Theory*. Springer, 3rd edition, 2004.

K. H. Lee, D. A. W. Barton, and L. Renson. Modelling of physical systems with a Hopf bifurcation using mechanistic models and machine learning. *Mechanical Systems and Signal Processing*, 191:110173, 2023.

Matthew E. Levine and Andrew M. Stuart. A framework for machine learning of model error in dynamical systems. *Communications of the American Mathematical Society*, 2(7):283–344, 2022.

Xin Li, Chengli Zhao, Xue Zhang, and Xiaojun Duan. Symbolic neural ordinary differential equations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18511–18519, 2025.

Alec J. Linot and Michael D. Graham. Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations. *Chaos*, 32(7):073110, 2022.

Alec J. Linot, Joshua W. Burby, Qi Tang, Prasanna Balaprakash, Michael D. Graham, and Romit Maulik. Stabilized neural ordinary differential equations for long-time forecasting of dynamical systems. *Journal of Computational Physics*, 474:111838, 2023.

Jean-Christophe Loiseau and Steven L. Brunton. Constrained sparse Galerkin regression. *Journal of Fluid Mechanics*, 838:42–67, 2018.

Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, 2018.

Gaurav Manek and J. Zico Kolter. Learning stable deep dynamics models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Shane A. McQuarrie, Cheng Huang, and Karen E. Willcox. Data-driven reduced-order models via regularised operator inference for a single-injector combustion process. *Journal of the Royal Society of New Zealand*, 51(2):194–211, 2021.

Shane A. McQuarrie, Parisa Khodabakhshi, and Karen E. Willcox. Nonintrusive reduced-order models for parametric partial differential equations via data-driven operator inference. *SIAM Journal on Scientific Computing*, 45(4):A1917–A1946, 2023.

Daniel A. Messenger and David M. Bortz. Weak SINDy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021.

Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):3–49, 2003.

Noa Moriel, Matthew Ricci, and Mor Nitzan. Let's do the time-warp-attend: Learning topological invariants of dynamical systems. In *International Conference on Learning Representations*, 2024.

Zachary G. Nicolaou, Guanyu Huo, Yihui Chen, Steven L. Brunton, and J. Nathan Kutz. Data-driven discovery and extrapolation of parameterized pattern-forming dynamics. *Physical Review Research*, 5(4):L042017, 2023.

Bernd R. Noack, Konstantin Afanasiev, Marek Morzyński, Gilead Tadmor, and Frank Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497:335–363, 2003.

Johan Paduart, Lieve Lauwers, Jan Swevers, Kris Smolders, Johan Schoukens, and Rik Pintelon. Identification of nonlinear systems using polynomial nonlinear state space models. *Automatica*, 46(4):647–656, 2010.

Alessandro Pecile, Nicola Demo, Marco Tezzele, Gianluigi Rozza, and Dimitri Breda. Data-driven discovery of delay differential equations with discrete delays. *Journal of Computational and Applied Mathematics*, 461:116439, 2025. doi: 10.1016/j.cam.2024.116439.

Benjamin Peherstorfer and Karen Willcox. Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*, 306:196–215, 2016.

Michel Provansal, Christian Mathis, and Luc Boyer. Bénard-von Kármán instability: transient and forced regimes. *Journal of Fluid Mechanics*, 182:1–22, 1987.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

Matthew Ricci, Guy Pelc, Zoe Piran, Noa Moriel, and Mor Nitzan. TRENDy: Temporal regression of effective nonlinear dynamics. In *International Conference on Learning Representations*, 2025.

Cristina Riso and Carlos E. S. Cesnik. Low-order geometrically nonlinear aeroelastic modeling and analysis of the Pazy wing experiment. In *AIAA SciTech Forum*, number AIAA 2022-2313, 2022.

Yulia Rubanova, Ricky T. Q. Chen, and David Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems 32*, 2019.

Samuel H. Rudy, Alessandro Alla, Steven L. Brunton, and J. Nathan Kutz. Data-driven identification of parametric partial differential equations. *SIAM Journal on Applied Dynamical Systems*, 18(2): 643–660, 2019.

Antoine Sandoz, Verena Ducret, Georg A. Gottwald, Gilles Vilmart, and Karl Perron. SINDy for delay-differential equations: application to model bacterial zinc response. *Proceedings of the Royal Society A*, 479(2269):20220556, 2023. doi: 10.1098/rspa.2022.0556.

Nihar Sawant, Boris Kramer, and Benjamin Peherstorfer. Physics-informed regularization and structure preservation for learning stable reduced models from data with operator inference. *Computer Methods in Applied Mechanics and Engineering*, 404:115836, 2023.

Hayden Schaeffer, Giang Tran, and Rachel Ward. Learning dynamical systems and bifurcation via group sparsity. *arXiv preprint arXiv:1709.01558*, 2017.

Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

Johan Schoukens and Lennart Ljung. Nonlinear system identification: A user-oriented road map. *IEEE Control Systems*, 39(6):28–99, 2019.

Lawrence Sirovich. Turbulence and the dynamics of coherent structures. I. Coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, 1987.

Naoya Takeishi and Alexandros Kalousis. Physics-integrated variational autoencoders for robust and interpretable generative modeling. In *Advances in Neural Information Processing Systems 34*, 2021.

Wentao Tang. Data-driven bifurcation analysis via learning of homeomorphism. In *Proceedings of the 6th Annual Learning for Dynamics and Control Conference*, volume 242 of *Proceedings of Machine Learning Research*, pp. 1149–1160, 2024.

Lorenzo Tomada, Moaad Khamlich, Federico Pichi, and Gianluigi Rozza. Sparse identification for bifurcating phenomena in computational fluid dynamics. *Computers & Fluids*, 302:106841, 2025. doi: 10.1016/j.compfluid.2025.106841.

Charles H. K. Williamson. Vortex dynamics in the cylinder wake. *Annual Review of Fluid Mechanics*, 28:477–539, 1996.

Hiroshi Yamasaki and Bogdan I. Epureanu. Forecasting supercritical and subcritical Hopf bifurcations in aeroelastic systems. *International Journal of Non-Linear Mechanics*, 94:400–405, 2017.

Yuan Yin, Vincent Le Guen, Jérémie Dona, Emmanuel de Bézenac, Ibrahim Ayed, Nicolas Thome, and Patrick Gallinari. Augmenting physical models with deep networks for complex dynamics forecasting. In *International Conference on Learning Representations*, 2021.

Qunxi Zhu, Yao Guo, and Wei Lin. Neural delay differential equations. In *International Conference on Learning Representations*, 2021.

A APPLICATIONS

The applications test identification of a parametric family with one or two stability transitions. The Rijke tube has known delay equations, allowing comparison with characteristic roots. The cylinder wake has been modelled parametrically across its Hopf bifurcation by Conti et al. (2023), with a Hopf transition in a sudden-expansion channel studied by Tomada et al. (2025), and at fixed Reynolds number by sparse regression (Brunton et al., 2016) and a constrained sparse Galerkin model (Loiseau & Brunton, 2018). The flexible wing is a reduction of a geometrically nonlinear fluid-structure interaction solver with a bounded flutter interval. Its governing vector field is not supplied to the learner. Low-order physics-based models of such wings (Riso & Cesnik, 2022) and numerical continuation of their limit cycles (Eaton et al., 2018) both require the governing equations.

In this section, we describe the three applications, their dynamics and the data each was trained on. They were chosen to place different demands on a parametric model while sharing one qualitative structure, an equilibrium that loses stability as a physical parameter varies. The wing and the tube regain it and the wake does not. They also span three levels of fidelity, as Table 6 summarises, from analytically specified delay differential equations to a reduction of a high-fidelity solver whose vector field the learner never sees.

Table 6: Systems, grids and reference bifurcation points, in the units of each system's parameter.

	Flexible wing	Rijke tube	Cylinder wake
generator	FSI solver	Galerkin delay system	lattice Boltzmann
states / parameter	9 / α (deg)	6 / τ	8 / Re
conditions reported	59	55	67
split, train / development / test	20 / 4 / 35	14 / 4 / 37	19 / 4 / 44
true bifurcation points	1.45, 3.25	0.1377, 0.9326	44.11
recovered bifurcation points	1.47, 3.23	0.1371, 0.9346	44.15

A.1 FLEXIBLE WING, A GEOMETRICALLY NONLINEAR FLUID-STRUCTURE INTERACTION PROBLEM

The third system is a highly flexible cantilever wing of span $550\,\mathrm{mm}$ and chord $100\,\mathrm{mm}$, simulated with SHARPy (del Carre et al., 2019). The model stems from an experimental highly flexible wing (Avin et al., 2022; Drachinsky et al., 2022). SHARPy couples a geometrically exact displacement-based beam to an unsteady vortex-lattice model. Large deflections make the structural response geometrically nonlinear, while the feedback between wing motion and unsteady aerodynamic loading produces flutter and finite-amplitude oscillations over an interval of angle of attack.

Structural and aerodynamic coordinates. The simulated beam velocities and internal forces/moments are expressed in intrinsic coordinates and projected onto the structural linear normal modes. Retaining N modes gives the approximations

$$\boldsymbol{x}_1(s,t) \approx \sum_{j=1}^{N} \boldsymbol{\Phi}_{1j}(s) q_{1j}(t), \qquad \boldsymbol{x}_2(s,t) \approx \sum_{j=1}^{N} \boldsymbol{\Phi}_{2j}(s) q_{2j}(t),$$

where $\boldsymbol{x}_1$ contains translational and angular velocities, $\boldsymbol{x}_2$ contains internal forces and moments, and $\boldsymbol{q}_1, \boldsymbol{q}_2 \in \mathbb{R}^N$ are their modal amplitudes.

The aerodynamic coordinates are obtained from snapshots of the distributed forces and moments $\boldsymbol{F}(s,t)$. A global snapshot matrix pools training conditions and times. Proper orthogonal decomposition supplies an orthonormal spatial basis $\boldsymbol{\Phi}_3$, giving

$$\boldsymbol{F}(s,t) \approx \boldsymbol{\Phi}_3(s)\boldsymbol{q}_3(t), \qquad \boldsymbol{q}_3(t) = \int_0^L \boldsymbol{\Phi}_3(s)^\top \boldsymbol{F}(s,t)\,\mathrm{d}s \in \mathbb{R}^r.$$

Only training conditions enter basis estimation. The same fixed basis projects development, test and other non-training records.

Coupled aeroelastic dynamics. The structural modal equations and an autonomous closure for the aerodynamic coordinates give the approximate coupled reduced description

$$\begin{aligned}\dot{\boldsymbol{q}}_1 &= \Omega \boldsymbol{q}_2 - q_{1\ell}\,\Gamma_1^\ell \boldsymbol{q}_1 - q_{2\ell}\,\Gamma_2^\ell \boldsymbol{q}_2 + \boldsymbol{g}(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3; \alpha),\\ \dot{\boldsymbol{q}}_2 &= -\Omega \boldsymbol{q}_1 + q_{2\ell}\,(\Gamma_2^\ell)^\top \boldsymbol{q}_1,\\ \dot{\boldsymbol{q}}_3 &= \boldsymbol{f}(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3; \alpha).\end{aligned} \tag{6}$$

Repeated indices are summed. The diagonal matrix $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_N)$ contains the structural natural frequencies. The matrices $\Gamma_1^\ell, \Gamma_2^\ell \in \mathbb{R}^{N\times N}$ are slices of the coupling tensors obtained by Galerkin projection of the intrinsic beam equations. Their quadratic terms represent geometric nonlinearity in the intrinsic formulation. The function $\boldsymbol{g}$ represents aerodynamic loading projected onto the structural modes, including its dependence on wing deformation. The function $\boldsymbol{f}$ describes the evolution of the unsteady aerodynamic coordinates coupled to the structural motion. Here α varies at fixed flow speed. Projection onto a finite set of coordinates does not by itself guarantee an exact autonomous closure for $\boldsymbol{q}_3$.

We retain $N = 3$ structural modes and $r = 3$ aerodynamic modes, giving

$$\boldsymbol{x} = [\boldsymbol{q}_1^\top, \boldsymbol{q}_2^\top, \boldsymbol{q}_3^\top]^\top \in \mathbb{R}^9.$$

These nine coordinates are shared by all compared methods. The proposed model identifies the full vector field from their equilibrium-centred trajectories. The structural operators in equation 6 are not supplied to the learner. Tip velocity, a root-bending proxy and lift are fixed linear projections of these coordinates, rather than additional learned quantities.

Both ends of the interval are changes of the stability of the equilibrium. Each record starts from a small disturbance of the static equilibrium. The early envelope provides an empirical indicator of growth or decay. Finite-window rates can also reflect mixtures of modes. The rate from the tip signal over the early part of each record changes sign between adjacent angles at both ends of the interval, from -0.41 to $+0.09$ per second between $1.4°$ and $1.5°$ and from $+0.64$ to -0.67 per second between $3.2°$ and $3.3°$, while the oscillation frequency is continuous across each end, at 34 and 29 Hz respectively. At the upper end the record at $3.3°$ decays over 1.3 decades with a fit of $R^2 = 1.00$, so the observed disturbance decays. Together with the nonzero oscillation frequency, these sign changes are consistent with Hopf transitions, including restabilisation at the upper boundary. They do not establish all Hopf nondegeneracy conditions or exclude coexisting attractors. Each crossing is bracketed by the grid interval between the last stable and the first unstable angle. Rollouts of the learned model from small and from large initial states approach the same long-time regime at every angle tested. The equilibrium itself is a geometrically nonlinear static deflection that varies with the angle of attack. Low-order aeroelastic models of wings of this class are given by Riso & Cesnik (2022), and numerical continuation of the limit cycles and their bifurcations on wings of this class by Eaton et al. (2018).

At the flow speed used here, $U = 57.5$ m/s, the wing is stable up to $\alpha = 1.4°$, holds a limit cycle from $1.5°$ to $3.2°$, and is stable again from $3.3°$. The unstable interval is therefore bounded, but its endpoints are bracketed by the grid rather than resolved, so the nominal onset $1.45°$ and offset $3.25°$ carry an uncertainty of half the grid spacing, $0.05°$. Tip-displacement amplitude rises from 0.09 at $1.5°$ to its peak of 2.49 at $1.6°$, one grid step above the onset, and then falls monotonically to 0.41 at $3.2°$, so the branch is strongly asymmetric with a steep rise at onset and a slow decline towards the offset. The oscillation frequency drifts downward across the window, from 33.8 Hz at $1.5°$ to 29.0 Hz at $3.2°$, a fall of about 14%, so amplitude and frequency vary together and a model that matches one does not thereby match the other.

Training specifications. Trajectories are generated at 61 angles of attack from $0.0°$ to $6.0°$ in steps of $0.1°$, each of 4,980 samples at $\Delta t = 1$ ms. Training uses 20 conditions, fourteen at which the equilibrium is stable, $\alpha \in \{0.1, 0.4, \ldots, 1.3\} \cup \{3.4, 3.7, \ldots, 5.8\}$, and six inside the window at which it grows, $\alpha \in \{1.6, 1.9, 2.2, 2.5, 2.8, 3.1\}$. Of the 59 angles up to $5.8°$, the last training angle, 20 are training, four are development conditions used only for model selection, $\alpha \in \{0.5, 2.0, 2.9, 4.5\}$, and 35 are test. The two angles above it are not scored. The test condition $0.0°$ lies just below the lowest training angle, $0.1°$. The remaining reported conditions lie within the training range. The operator is fitted on the window of each record that precedes saturation of the amplitude. Every model, including every baseline, sees the same twenty records.

A.2 RIJKE TUBE, THERMOACOUSTIC INSTABILITY WITH A TIME DELAY

A Rijke tube is a duct containing a compact heat source. Acoustic velocity fluctuations at the heater modulate the heat release, the released heat drives the acoustic field, and when the phase relation between the two is favourable the coupling extracts energy from the mean flow and an initially quiet duct develops a self-sustained oscillation. The mechanism is the canonical model problem for combustion instability (Juniper & Sujith, 2018) and the non-normal transient behaviour of the linearised operator is well documented (Balasubramanian & Sujith, 2008).

Expanding the acoustic field in duct eigenmodes gives modal velocity and pressure amplitudes obeying

$$\dot{\eta}_j = j\pi\mu_j, \tag{7}$$

$$\dot{\mu}_j = -j\pi\eta_j - \zeta_j\mu_j - 2Q(t)\sin(j\pi x_f), \qquad j = 1, \dots, N, \tag{8}$$

$$Q(t) = K\left(\sqrt{|u_0 + u_f(t-\tau)|} - \sqrt{u_0}\right), \tag{9}$$

$$u_f(t) = \sum_j \eta_j(t)\cos(j\pi x_f), \tag{10}$$

with modal damping $\zeta_j = c_1 j + c_2\sqrt{j}$, heat release following King's law with a time lag τ, and $(x_f, c_1, c_2, K, u_0) = (0.25, 0.28, 0.16, 1, 1/3)$. The parameter swept is that lag. Because the delayed velocity $u_f(t-\tau)$ appears on the right-hand side, the true system is a delay differential equation and therefore infinite-dimensional. Any finite-dimensional Markovian model, including the one identified here, is a surrogate for its dominant dynamics rather than an exact representation. Current-state polynomial libraries also incur this closure approximation. Delay-augmented libraries can represent history dependence. The truncation keeps $N = 3$ modes, giving $\boldsymbol{x} = [\eta_1, \eta_2, \eta_3, \mu_1, \mu_2, \mu_3]^\top \in \mathbb{R}^6$, and the evaluated observables are the flame velocity $u_f = (\eta_1 - \eta_3)/\sqrt{2}$, the mid-duct pressure $p'(0.5) = -\mu_1 + \mu_3$ and the mid-duct velocity $u(0.5) = -\eta_2$.

The crossings are at $\tau = 0.1377$ and $\tau = 0.9326$, so the window is closed. Inside it the growth rate is a single arch peaking at $+0.176$ near $\tau = 0.45$ and never exceeding 0.2, so the identification problem is one of resolving a weak positive rate rather than a strong one. Flame-velocity amplitude rises from zero at onset, peaks at 0.653 near $\tau = 0.5$ and returns to zero at the offset, while the oscillation frequency falls monotonically from 0.53 to 0.45 across the window. Beyond $\tau \approx 1.1$ a second family of characteristic roots overtakes the fundamental one, and the reference follows the fundamental family, which carries both crossings and is the branch a finite-dimensional operator represents. The evaluation range is $\tau \leq 1.4$.

Training specifications. The grid is 71 conditions uniformly spaced by 0.025 over $\tau \in [0.05, 1.80]$, each a $100\,\mathrm{s}$ record at $\Delta t = 0.025$, giving an array of shape $(71, 4000, 6)$. Fourteen conditions, $\tau = 0.075, 0.175, \dots, 1.375$, are used for training, four $\{0.25, 0.45, 0.70, 0.85\}$ for model selection only, and 53 for test. The reference unstable interval is $0.1377 < \tau < 0.9326$ and both crossings fall inside training gaps. Results are reported over the 55 conditions with $\tau \leq 1.4$, split 14 training, 4 development and 37 test. The operator is fitted on the window of each record that precedes saturation of the amplitude.

A.3 CYLINDER WAKE, A REDUCTION OF A PARTIAL DIFFERENTIAL EQUATION

Two-dimensional incompressible flow past a circular cylinder is the canonical bluff-body wake. Below a critical Reynolds number the flow settles to a steady symmetric pair of recirculation bubbles. Above it the wake becomes unstable to an antisymmetric global mode and saturates into the periodic von Kármán vortex street. The transition is a supercritical Hopf bifurcation whose amplitude follows the Stuart-Landau square-root law near onset (Provansal et al., 1987; Jackson, 1987), and the wake and its low-order descriptions are among the most studied objects in fluid mechanics (Williamson, 1996; Noack et al., 2003; Barkley, 2006). Unlike the other two systems the equilibrium does not restabilise over the reported range, so this application falls outside the closed-window setting.

Trajectories are produced by a D2Q9 lattice Boltzmann solver with a two-relaxation-time collision operator at the parametrisation $\Lambda = 3/16$, which fixes the effective wall position of the bounce-back

condition independently of the relaxation time and therefore prevents the geometry from drifting as the Reynolds number changes (Ginzburg et al., 2008). The cylinder is resolved by 20 lattice nodes across its diameter on a 461×241 grid extending $8D$ upstream, $15D$ downstream and $\pm 6D$ laterally, with a free-slip lateral boundary and a zero-gradient outlet. Lengths are measured in diameters, velocities in the inlet speed and time in convective units D/U. Above the critical Reynolds number the base flow is an unstable equilibrium, so it is obtained on the half domain with a symmetry plane that forbids the antisymmetric mode and then mirrored, which converges onto the unstable branch rather than falling onto the limit cycle.

Snapshots are cropped and strided to 80×140 nodes per velocity component. The base flow computed at each Reynolds number is subtracted, so the reduced origin is an exact equilibrium at every condition and the model is never required to learn where the equilibrium sits. The fluctuations of the training conditions only are pooled into a single proper orthogonal decomposition basis, computed by the method of snapshots (Sirovich, 1987), and then held fixed. Development, test and other non-training snapshots do not enter basis estimation. All conditions are projected onto the same fixed training basis. The state is the vector of the leading $r = 8$ coefficients,

$$\boldsymbol{z}(t) = \boldsymbol{\Phi}^{\top}[\boldsymbol{u}(\cdot, t) - \bar{\boldsymbol{u}}_{Re}] \in \mathbb{R}^{8}.$$

A shared training basis gives consistent coordinates across parameters. Retaining $r = 8$ POD modes captures 95.66% of the fluctuation energy. The omitted energy limits reconstruction of the full field, not prediction of the retained coordinates. For an orthonormal basis, the squared full-field error is the sum of the squared coordinate error and the squared projection residual. Thus the two errors coincide only when the reference field has also been projected. The wake's trajectory scores in this paper use the retained coordinates. They have no error floor imposed by the discarded POD energy.

The wake has a single crossing at $Re_c = 44.11$. That value is below the 46 to 47 usually quoted for the two-dimensional cylinder, because it is the critical Reynolds number of this lattice Boltzmann discretisation at this resolution, bracketed from its own records, and not the continuum value. The reference and the model are compared on the same discretisation, and 39 of the 75 conditions lie above it. Near onset the settled amplitude follows the Stuart-Landau law closely, with the squared amplitude linear in $Re - Re_c$ at a measured slope of about 0.12, giving 0.24 at $Re = 44.6$ and 0.90 at $Re = 51.0$. Further out the branch saturates rather than continuing to grow, reaching 1.18 near $Re = 71$ and declining slightly to 1.02 by $Re = 120$. The shedding frequency rises monotonically with Reynolds number over the whole range, from 0.132 just above onset to 0.176 at $Re = 120$, so amplitude and frequency are not in a fixed relation and a model has to track both. Half of the total amplitude change occurs within the first 4% of the supercritical range.

Training specifications. The original 71-condition grid runs from $Re = 25.42$ to $Re = 154.21$, uniform in $\mu = 1/Re$ with spacing 4.69×10^{-4} and a spread of 2.1×10^{-7} across the intervals. Four additional training conditions at $Re = 44.29$, 44.44, 44.85 and 45.11 refine the grid near onset, giving 75 conditions in total. Of these, 36 lie below onset and 39 above. Of the full sweep, 21 conditions are used for fitting, 4 for model selection and 50 for test. Records are stored every 0.4 convective times and the model step is $\Delta t = 0.399$. The reference critical value is $Re_c = 44.11$. Onset falls between the two original grid points 43.65 and 44.57. The condition $Re = 44.57$ is held out of training. The sweep contains 75 conditions, of which 21 are used for training and the remaining conditions are reserved for development or testing. Results are reported over the 67 conditions with $Re \leq 100$, which is the operating range the surrogate is built for. Of those 67, 19 are training conditions, 4 are development conditions used for model selection only, at $Re = 29.68, 38.18, 52.21 and 82.55$, and 44 are test conditions. The training conditions above $Re = 100$ inform the fit and are not scored. The operator is fitted on the window of each record that precedes saturation of the amplitude.

B Architecture and training

This section describes the model and training settings. It is organised as the two stages of the proposed model, the common recipe of the controlled comparison, the external baselines, and the provenance of every reference quantity.

B.1 Stage one, the operator

Stage one minimises the normalised multi-step trajectory error of equation 4 with L-BFGS on the linear window of each record, the part that precedes saturation of the amplitude, plus the one-sided stability penalty equation 5. The window is fixed by a rule on the record's own envelope. On the tube the window of a growing record ends where the root-mean-square envelope first exceeds 40 percent of its settled level, and a decaying record contributes its first 2000 steps. On the wake it ends where the oscillatory amplitude, the norm of the state after removing a running mean over one shedding period, first reaches 30 percent of the record's settled amplitude, a record whose settled amplitude is within three times its early minimum is treated as decaying and contributes the whole record, and every window is capped at 2400 steps. On the wing the window of a saturating record ends where the 120-sample root-mean-square envelope first exceeds 40 percent of its settled level after the post-transient trough, with a floor of 400 samples, and every other record contributes its whole 4980 samples. On the tube and the wing the constant term is initialised from a least-squares fit and the higher coefficients start at zero. On the wake the reported operator is the last link of a continuation in the polynomial degree, each degree being initialised from the converged operator of the degree below. The margin m and weight λ_{stab} are listed in Table 7. The stability penalty uses stable training records. On the wake it is implemented as a hinge with margin m on the oscillatory pair and a penalty of weight 2 with margin 0.05 on the real eigenvalues. The growth rate reported and scored is the largest real part over all eigenvalues (Section 4).

The degree of the matrix polynomial is 1 on the tube, 2 on the wing and 4 on the wake. On the tube it is the lowest degree, and raising it to two or three does not improve the growth-rate error on the development conditions (Appendix F.5). On the wake the degree was selected on the trajectory error of the development conditions among operators of degree two to five, each reached by continuation from the degree below, fitting degree two then three then four from the previous solution. On the wing the degree is two, and an operator of degree three fitted with the same recipe also returns exactly two crossings, with lower- and upper-crossing errors of 1.95 and 0.10 percent of the interval against 1.26 and 1.04 percent at degree two (Table 4). Both degrees recover the count, with different location errors. This is a sensitivity check rather than degree independence.

B.2 Stage two, the residual

Stage two trains the gated residual of equation 1 on free-running, amplitude-normalised trajectory segments with AdamW and unit gradient clipping, under the system-specific learning-rate schedules and horizons in Table 7. The objective combines trajectory error with system-specific auxiliary terms. The terms and their nonzero weights differ by system, as listed below. Weight magnitudes alone do not measure their relative influence because the terms have different scales.

- Trajectory error, the relative ℓ_2 discrepancy of equation 11 between the rollout and the record over each segment, states pooled.
- Log-envelope error, comparing the running envelope of the rollout with that of the record so that growth and decay rates are penalised directly rather than through phase.
- Settled-amplitude error at training conditions that develop a limit cycle.
- Stability hinge, the penalty of equation 5 applied to the deployed equilibrium Jacobian at the training conditions. For the proposed model that Jacobian is the frozen operator, so the term is inactive. It is retained because the same recipe is applied to every architecture of the controlled comparison, where it acts on the Jacobian of that architecture.
- Application-specific terms. The tube adds to the second stage a term that anchors the model's vector field to the finite-difference derivative of the record at states sampled from the training records, annealed over the curriculum and absent from the plain neural ODE row.

Weights. On the tube the second-stage weights are 1 on the trajectory term, 0.09 on the observable, 0.3, 0.9, 0.6 and 0.6 on the envelope, transient-envelope, envelope-outline and modal-envelope terms, 0.9 on the instantaneous frequency, 120 on the settled frequency, 7.5 on the settled level, 0.9 on the settled mean, 1 on the vector-field anchor and 5 on the stability hinge with margin 0.02, the auxiliary terms scaled by 0.3. On the wake they are 1 on the trajectory term, 0.1 on the phase term, 0 on the amplitude term and 5 on the stability hinge with margin 0.004. On the wing they are 0.22 on

the pointwise term, 4 on the envelope, 10 on the settled amplitude, 5 on the traverse term, 80 on the carrier frequency, 6 on the carrier phase, 0 on the relative term and 1.5 on the trajectory growth-rate error, with no stability hinge in the second stage since the operator is frozen. The first-stage stability hinge has weight 5 and margin 0.5 per second.

Table 7: Implementation summary for the proposed model. D is the degree of the matrix polynomial in equation 3, r_0 the gate radius of equation 2, and horizons are step counts listed in training order.

	Flexible wing	Rijke tube	Cylinder wake
state dimension	9	6	8
parameter	α	τ	$\mu = 1/Re$
parameter degree D	2	1	4
features, operator	raw $\alpha^{0..2}$	raw $\tau^{0..1}$	$t^d, t = \mu/\mu_{\max}$
features, residual	$\hat{\alpha}^{1..3}, \hat{\alpha} = (\alpha - 3)/3$	$[t, t^2, t^3]$, $t = (\tau - 0.925)/0.85$	3 polynomial features
stage-1 window rule	linear window before saturation	linear window before saturation	linear window before saturation
stage-1 margin m, weight λ_{stab}	0.5, 5	0.02, 5	0.004, 5
residual network	3×128, tanh	3×128, tanh	3×128, tanh
residual parameters	45,137	39,076	42,944
gate radius r_0	0.45	0.3	0.2
time step	0.001 s	0.025 s	0.399 ct
residual horizons	300, 800, 1600, 3000, 3000, 3000, 3000	100, 300, 500, 800, 1600	48, 160, 480, 1600, 3200
iterations per stage	80, 80, 80, 120, 90, 90, 120	100, 100, 120, 500, 400	300, 400, 400, 400, 300
learning rate	10^{-3} to 10^{-4}	10^{-3} to 1.2×10^{-4}	5×10^{-4} to 3×10^{-5}
optimiser, stage 1 / 2	LBFGS/AdamW	LBFGS/AdamW	LBFGS/AdamW
seeds	5	5	5
tuning configurations	9, plain neural ODE rows	9 per architecture	9 per architecture
diverged runs excluded	0	0	0
selection rule	development whole-record error, median	development pointwise error at the final horizon	development free-rollout error, mean plus half the maximum

B.3 The controlled comparison

Within each system the neural architectures use a common training implementation and deviation coordinates. The conventional plain neural ODE omits the curriculum and auxiliary terms. The remaining rows use the system's common recipe, subject to the tuning differences stated below. The neural residuals have parameter counts within five percent, while free-operator models add their own parameters. The wing residual uses parameter and state inputs. r enters its gate but is not a separate network input. Every architecture in this comparison, including the proposed model, uses fourth-order Runge–Kutta integration, so the time-stepping method is held fixed. Every learned output layer is initialised to zero, so that each architecture starts at its operator. The equilibrium Jacobian of each architecture is verified against automatic differentiation, and the verification is itself tested by releasing the operator, removing the gate or randomising one component and confirming that it then fails.

The rollout-horizon schedules are specified in Table 7. The tube schedule uses recorded saturation times. Model selection uses the development conditions only, by the rule in Table 7, and a configuration is eligible only if every stage of the curriculum completed. On the tube and the wake the learning rate and one multiplier on the weights of the auxiliary losses are each varied over three values at one seed for every architecture, nine configurations each, and the plain neural ODE with the full recipe also varies its output scale over three values, for a total of twenty-seven configurations. The remaining seeds are run at the selected configuration. On the wing the three plain neural ODE rows are tuned over nine configurations each, the learning rate and the longest rollout horizon over three values, while the proposed model and its ablations run at the fixed curriculum without a search. No tuning run was excluded for divergence, because a diverged run records an infinite score.

Successive architectures are compared on the per-seed median error using seeds available for both architectures. Table 8 reports how often the later architecture has the smaller error, with the mean difference.

B.4 External baselines

Parametric sparse regression. Monomials in the state to degree 3 tensored with monomials in the parameter, to degree 3 on the tube and 2 on the wake, built through the SINDy-CP construction

end to end. Derivatives are finite differences on the state records. Sequentially thresholded least squares with the threshold swept on a logarithmic grid, 14 configurations on the tube and 6 on the wake, selected on development rollout error.

AE-SINDy. An independent implementation of parametric AE-SINDy following Conti et al. (2023) and Tomada et al. (2025), with architecture and library selected on our development conditions. The cylinder benchmark is shared with Conti et al., but our simulator, records and architecture differ. We compare the methods on common data and splits without claiming to reproduce the authors' implementation or published performance. An encoder and a decoder, each a two-layer network of width 64, map the state to a latent space of dimension two, three or four, in which the dynamics are a sparse polynomial of degree two or three in the latent state tensored with monomials of degree one, two or three in the parameter. The three networks are trained jointly with the loss of Champion et al. (2019), namely reconstruction, matching of the latent derivative through the encoder and of the state derivative through the decoder, and an ℓ_1 penalty with sequential thresholding at 10^{-3}, 10^{-2} or 10^{-1}, for 2000 epochs of Adam. The 54 configurations are selected on the development conditions by boundedness and then rollout error, and the selected configuration is run at 5 seeds. Rollouts are integrated in the latent space from the encoded initial state and decoded, and the stability of the model is read from the Jacobian of the latent field at its fitted equilibrium, found by Newton iteration from the encoded equilibrium of the data. The selected configuration has latent dimension three, latent degree three and parameter degree three, threshold 10^{-3} on the tube (217 active terms), latent dimension two, latent degree two and parameter degree three, threshold 10^{-3} on the wake (39) and latent dimension four, latent degree three and parameter degree one, threshold 10^{-3} on the wing (252).

Delay-augmented sparse regression. Implemented following Pecile et al. (2025) and Sandoz et al. (2023), with the library augmented by delayed samples of the state and the delay supplied rather than searched. Polynomial degrees 2, 3, 4, 5, 7 were swept over 120 configurations, and a single model was selected by the authors' own sparsity rule on the derivative cost, with no test metric consulted. This implementation uses supplied delays but no polynomial parameter library, limiting its parameter dependence relative to the parametric SINDy baseline. The ground-truth heat-release term was never supplied to it.

Operator inference. The reduced polynomial form $\dot{\boldsymbol{z}} = \boldsymbol{c}(p) + \boldsymbol{A}(p)\boldsymbol{z} + \boldsymbol{H}(p)(\boldsymbol{z} \otimes \boldsymbol{z})$ with each coefficient polynomial in the parameter, using unique quadratic monomials rather than the full Kronecker product, which is rank deficient. The quadratic state form is motivated by the quadratic terms in the wake's Galerkin dynamics and the wing's intrinsic beam equations. It need not close the reduced aerodynamic dynamics or represent the tube's delayed heat release exactly. On the tube the configurations span two derivative schemes, three degrees of the parameter dependence, and a free or a zero constant term. On the wing a cubic state form was run as well, and the configuration reported is the one the selection rule returns. Regularisation is Tikhonov with separate weights on the linear and quadratic blocks, selected on development conditions by boundedness first and trajectory error among the bounded. A quadratic correction has zero derivative at the origin. When the constant term is constrained to zero, the origin is an equilibrium and its Jacobian is therefore the linear block. These models are a close structural comparison, but their jointly fitted coefficients do not implement the separate identification and subsequent preservation used here.

Parametric linear control. On the tube and wake, the DMD control in Table 4 fits a linear operator at each training condition by snapshot-to-snapshot DMD on the first-stage windows, then interpolates operator entries across the parameter (Schmid, 2010; Amsallem & Farhat, 2008). The wing row instead uses a joint least-squares fit of a polynomial operator across the 20 training angles, without per-angle DMD. Its first-stage-window control uses this same estimator on the first-stage windows. The full-record row uses all 4980 samples. These controls use no exponential trajectory-fitting objective or stability penalty. Optimised-DMD-based comparisons are reported separately in Appendix E.6.

Linear operator inference on the first-stage windows. The second linear control. It fits one linear operator that is polynomial in the parameter, of the same degree as the proposed operator, by least squares on the state derivatives over the same windows, with each condition weighted by its

own norm as in the first stage. It is the same model class as the first stage of the proposed model fitted by derivative matching instead of the trajectory objective, again with no stability penalty.

Parametric neural ordinary differential equations. The two unstructured architectures of the controlled comparison. The conventionally trained variant uses the full record length from the start with no curriculum and is reported for contrast.

B.5 Reference quantities

The reference crossings, growth rates and frequencies are obtained as follows. For the tube, these quantities are obtained from the characteristic root of the linearised delay system that carries both crossings, computed exactly as a function of τ. It is the rightmost root up to $\tau \approx 1.11$, beyond which a second family overtakes it while both decay, and no root is unstable anywhere beyond the offset. For the wake, each record is projected onto a left eigenvector of an estimated parameter-dependent linear operator, and a straight line is fitted to the logarithm of the projected amplitude. The stable-record window starts at the initial amplitude peak and ends before the amplitude falls below three times the estimated noise floor. The unstable-record window starts after the initial kick has decayed and ends at 40 percent of the settled amplitude. The projection reduces contamination by other modes of the estimated operator. It does not guarantee isolation of a mode of the reference system. Because the projection direction is learned from the records, these rate estimates are not fully independent of the fitted linear model.

For the wing, the plotted stable-branch rates use the logarithmic slope of the full-state, scale-normalised envelope over 0.2–1.2 seconds. Inside the unstable interval the plot uses tip-signal envelope rates, with quadratic interpolation at $1.6°$–$2.1°$. These are finite-window diagnostics using different observables, not direct measurements of an asymptotic leading eigenvalue. In particular, the stable records reach the noise floor before that asymptotic rate can be resolved. On the wake and wing the reported crossing locations are nominal centres of brackets between adjacent sampled regimes. Their resolution is half the bracket width, $0.05°$ on the wing and 0.46 in Reynolds number on the wake. It is not a statistical confidence interval. Crossing errors are reported alongside that resolution. Settled amplitudes and limit-cycle frequencies are measured from the final quarter of each record as defined in Appendix C.

B.6 Compute

Training and available rollout timings are reported separately. On the tube the reference is a six-state delay integration costing 0.62 seconds per condition, against 169 seconds for the first stage and a median of 15,498 seconds for the second. On the wake, a 371-step rollout of the trained surrogate costs 642 milliseconds. This is not the lattice Boltzmann solver's cost. The two training stages take 33 and 131 minutes. These tube and wake timings use one CPU core without a GPU. On the wing the second stage took 73 to 407 minutes across the 5 seeds on a shared cluster. Those wall times include hardware and scheduling variability and do not establish a solver speed-up.

B.7 Two remarks on Proposition 1

First, the proposition fixes the linear spectrum, and hence the location and count of every crossing, but not its character. Whether a Hopf point is supercritical or subcritical is decided by the cubic terms of the deployed field, and those are supplied by the residual. Second, on smoothness. The network takes the amplitude r as an input, and r is not differentiable at the equilibrium. The proposition needs only that the network is bounded near the equilibrium, which holds for any input it is given, and its proof uses nothing else. The deployed field is nevertheless continuously differentiable at the equilibrium, because $q'(r) = O(r)$, so the derivative of the residual term $q(r)\boldsymbol{B}_\theta \boldsymbol{z}$ tends to zero with $\boldsymbol{z}$ and the Jacobian is continuous there. A linear gate $q(r) = r/(r + r_0)$ would make the residual $O(\|\boldsymbol{z}\|^2)$ and preserve the Jacobian equally well. The quadratic gate is used because r^2 is a quadratic form in $\boldsymbol{z}$, so the gate itself is a smooth function of the state. No smoothness beyond the first derivative is claimed at the equilibrium.

C METRICS

Trajectory metrics use free rollouts driven only by the initial state and parameter. Spectral metrics use the model's equilibrium Jacobian. Write $y_{\mathrm{ref}}(t_k)$ and $y_{\mathrm{pred}}(t_k)$, $k = 1, \dots, K$, for the reference and predicted scalar observable of a condition, sampled at the model step. Attractor quantities use the settled window $k \in \mathcal{T}$, the final quarter of the record, while trajectory quantities use the whole record.

Normalised error. Trajectory and cycle-waveform errors use the relative ℓ^2 discrepancy

$$\mathcal{N}(u, v) = \frac{\|u - v\|_2}{\max\big(\|v\|_2,\, \varepsilon\big)}, \qquad \varepsilon = 10^{-12}, \tag{11}$$

with v the reference. It is the normalised root-mean-square error, written nRMSE elsewhere in the paper. It is zero when the two agree and one when the prediction is identically zero, so unity is the score of a model that remains at the equilibrium.

Analytic signal, envelope and settled amplitude. Let $\mathcal{H}$ denote the Hilbert transform and $z(t) = \big(y(t) - \bar{y}\big) + \mathrm{i}\,\mathcal{H}\big[y - \bar{y}\big](t)$ the analytic signal of a mean-removed observable y, with $\bar{y}$ its mean over the record. The envelope is $|z|$ averaged over a moving window of one carrier period, W samples long,

$$a(t_k) = \frac{1}{|\mathcal{W}_k|} \sum_{m \in \mathcal{W}_k} \big|z(t_m)\big|, \tag{12}$$

where $\mathcal{W}_k$ indexes the samples in the smoothing window. The settled amplitude is its median over the settled window, $A = \mathrm{median}_{k \in \mathcal{T}}\, a(t_k)$.

Amplitude error. The relative difference of the two settled amplitudes,

$$e_{\mathrm{amp}} = \left| \frac{A_{\mathrm{pred}}}{A_{\mathrm{ref}}} - 1 \right|. \tag{13}$$

It is reported only at conditions whose reference develops a limit cycle, since $A_{\mathrm{ref}} \to 0$ elsewhere.

Cycle-folded error. Let $\phi(t) = \arg z(t) \bmod 2\pi$ be the instantaneous phase of a signal and partition $[0, 2\pi)$ into $N_\phi = 64$ equal bins $B_1, \dots, B_{N_\phi}$. The cycle-averaged waveform is the mean of the observable in each bin, with $\mathbb{1}[\cdot]$ the indicator function,

$$f_n = \frac{\sum_{k \in \mathcal{T}}\, y(t_k)\, \mathbb{1}[\phi(t_k) \in B_n]}{\max\Big(\sum_{k \in \mathcal{T}} \mathbb{1}[\phi(t_k) \in B_n]\,,\, 1\Big)}, \qquad n = 1, \dots, N_\phi, \tag{14}$$

and the cycle error is $e_{\mathrm{cyc}} = \mathcal{N}\big(f_{\mathrm{pred}}, f_{\mathrm{ref}}\big)$. Each signal is folded onto *its own* phase, so a constant phase offset between prediction and reference does not enter, while amplitude, waveform asymmetry and harmonic content all do.

Pointwise error. The pointwise trajectory error is $e_{\mathrm{pw}} = \mathcal{N}\big(y_{\mathrm{pred}}, y_{\mathrm{ref}}\big)$ over the whole record. The scored coordinates depend on the system, as specified below. Over many carrier cycles, a small frequency error accumulates phase drift and can raise the pointwise error even when amplitude and waveform are accurate. Figure 20 gives it per condition over the full record, pooled over the nine states for the wing and the eight latent coordinates for the wake and taken on the flame velocity for the tube. The wing median is 0.147, rising from 0.13 outside the unstable interval to 0.54 inside it and peaking at 1.169 at $\alpha = 1.60°$, just past onset.

Carrier-frequency error. The instantaneous frequency of a signal is $\omega(t) = \mathrm{d}\varphi/\mathrm{d}t$ with $\varphi = \arg z$ the unwrapped phase of the analytic signal, and the carrier frequency is its average over the settled window weighted by the envelope a, so that the quiet parts of a cycle do not dominate,

$$\bar{\omega} = \frac{\sum_{k \in \mathcal{T}} \omega(t_k)\, a(t_k)}{\max\big(\sum_{k \in \mathcal{T}} a(t_k),\, \varepsilon\big)}. \tag{15}$$

The reported error is $e_{\mathrm{freq}} = |\bar{\omega}_{\mathrm{pred}}/\bar{\omega}_{\mathrm{ref}} - 1|$. For the proposed wing model, the carrier is estimated from the settled tip-velocity rollout and reference by zero crossings and FFT, which agree at the reported precision. Its per-seed error is the median over the 14 angles from 1.7° to 3.0°. The daggered wing entries in Table 11 instead compare the model's equilibrium eigenfrequency with the reference carrier. They do not measure nonlinear rollout frequency.

Boundedness. A condition counts as bounded when the rollout stays finite and within twenty times the largest reference excursion,

$$\max_k \left|y_{\mathrm{pred}}(t_k)\right| \;\le\; 20 \max_k \left|y_{\mathrm{ref}}(t_k)\right|. \tag{16}$$

Stability sign and bifurcation points. The predicted growth rate at parameter p is $\sigma(p) = \max_i \mathrm{Re}\,\lambda_i\big(\boldsymbol{A}(p)\big)$, the rightmost eigenvalue of the frozen operator, which by equation 2 is also the deployed model's equilibrium Jacobian. The stability sign is correct at a condition when $\mathrm{sign}\,\sigma(p)$ matches the reference regime, and the sign column of Table 4 is the fraction of conditions at which it is. Recovered bifurcation points are the roots of σ, obtained by linear interpolation between the grid points that bracket each sign change. Their error is normalised by the width of the true unstable interval for the wing and the tube, and by the critical Reynolds number for the wake,

$$e_{\mathrm{bif}} = \frac{|\hat{p} - p^*|}{p^*_{\mathrm{off}} - p^*_{\mathrm{on}}}, \qquad e_{\mathrm{bif}}^{\mathrm{wake}} = \frac{|\widehat{Re}_c - Re_c|}{Re_c}. \tag{17}$$

D TRAJECTORY PREDICTIONS

Free rollouts are shown at five additional wing angles, eight tube delays and eight wake Reynolds numbers. All are test conditions, excluded from training and development. Each rollout is determined by the initial state and parameter only. The thick grey line is the reference and the blue lines show all 5 seeds of the proposed model. Left panels show the complete record. Right panels show the initial decay for stable conditions and the final oscillations for unstable conditions, separating local waveform and amplitude agreement from accumulated transient error. The near-crossing tube and wake examples also expose variability in growth and phase across seeds.

D.1 FLEXIBLE WING

Each condition shows tip velocity, root bending moment (RBM) and lift, from top to bottom. These are fixed linear projections of the reduced state.

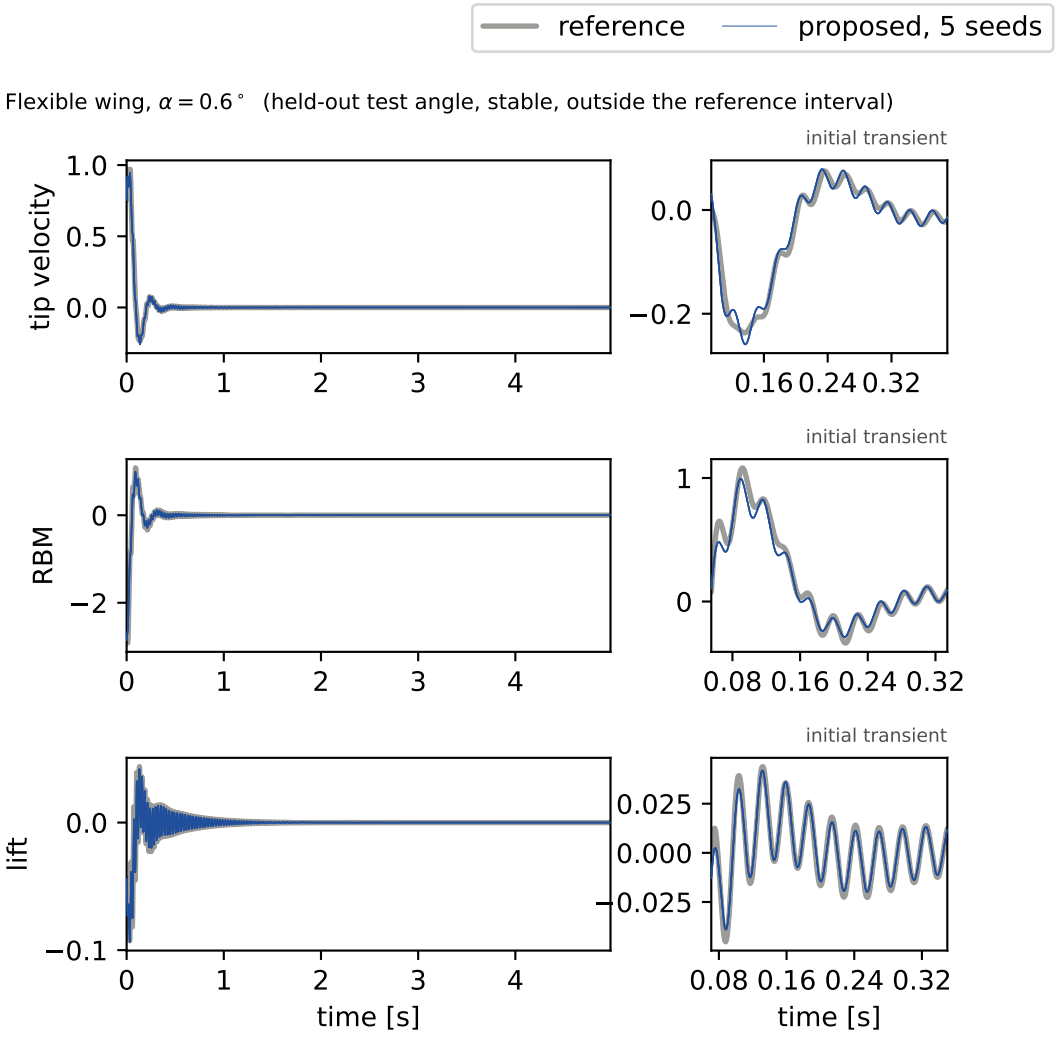


Figure 6: Wing at test angle $\alpha = 0.6°$, below flutter onset. Grey denotes the reference and blue all 5 seeds. The right column shows the initial decay.

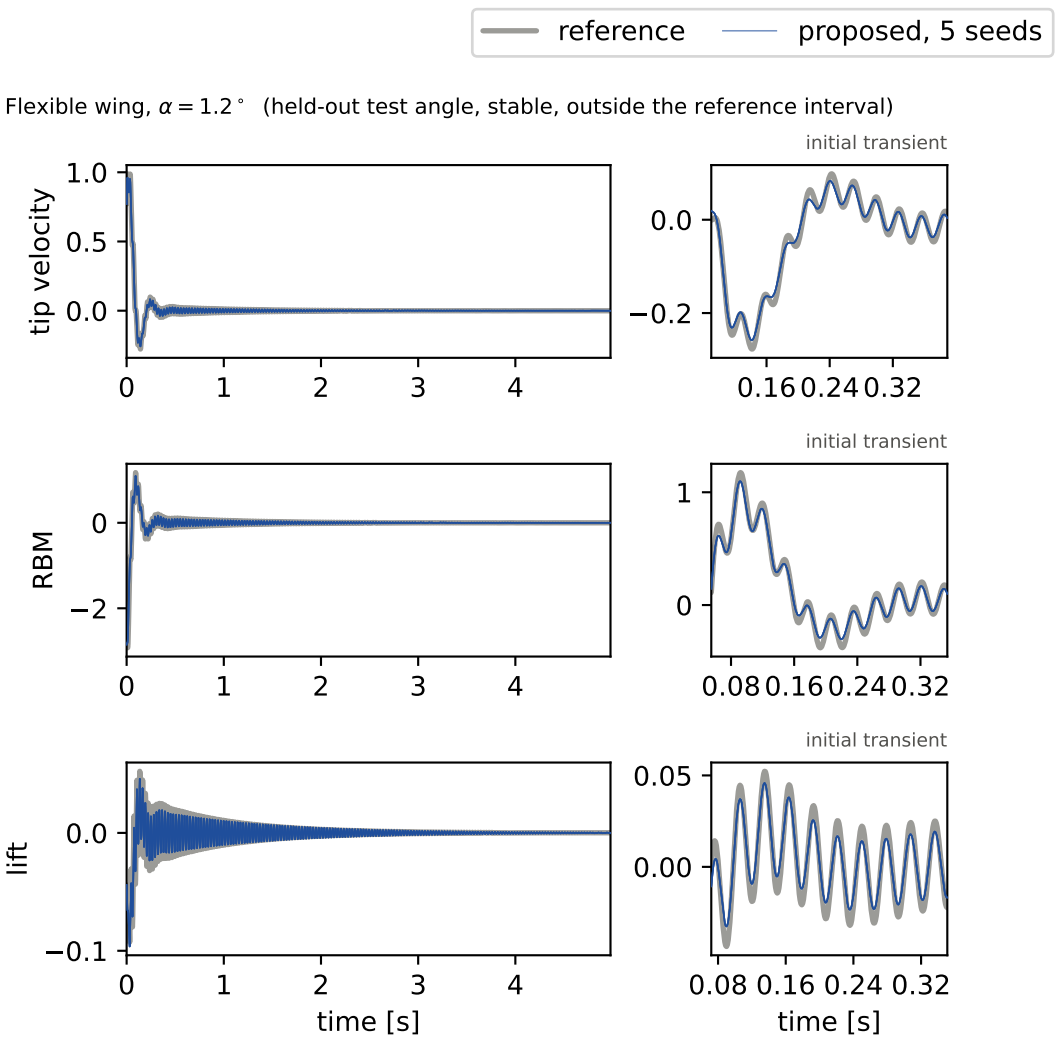


Figure 7: Wing at test angle $\alpha = 1.2°$, below flutter onset. Grey denotes the reference and blue all 5 seeds. The right column shows the initial decay.

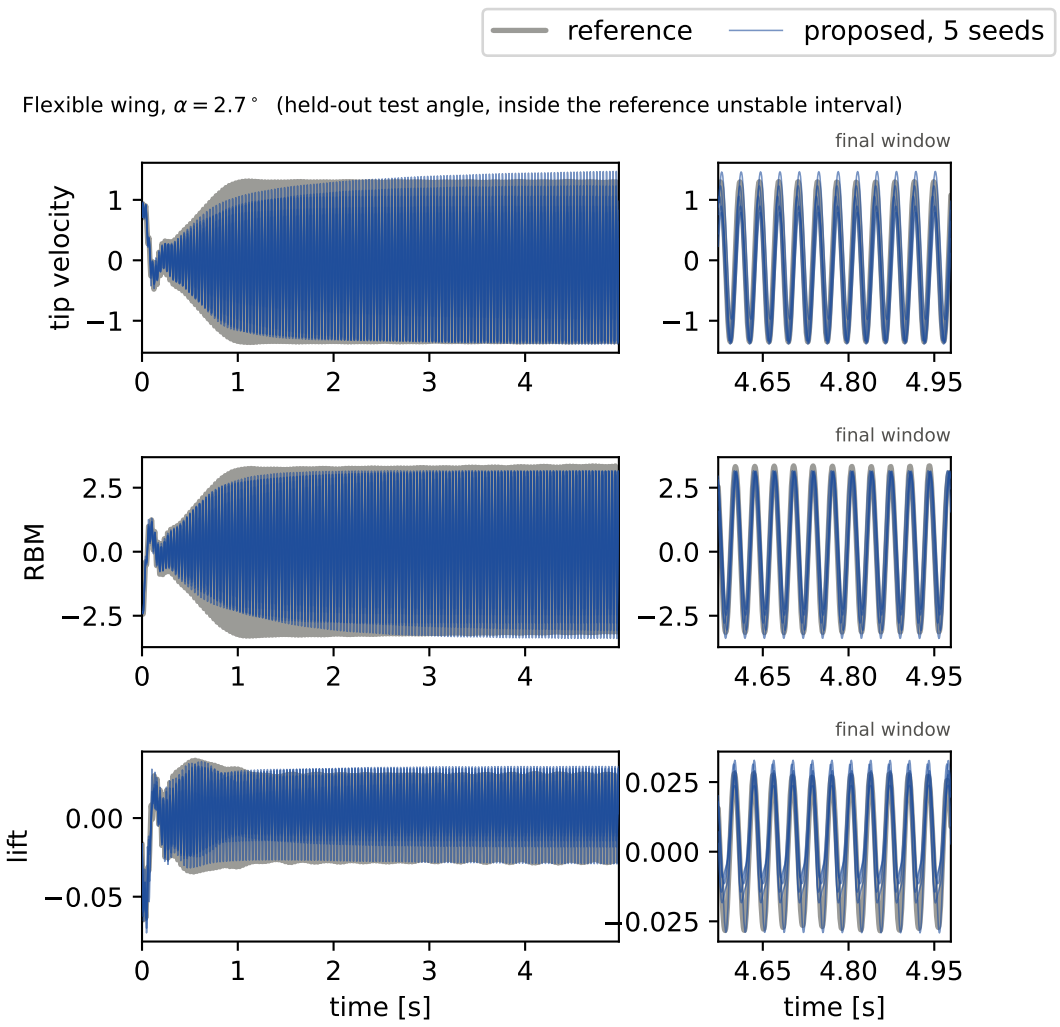


Figure 8: Wing at test angle $\alpha = 2.7°$, inside the flutter interval. Grey denotes the reference and blue all 5 seeds. The right column shows the final oscillations.

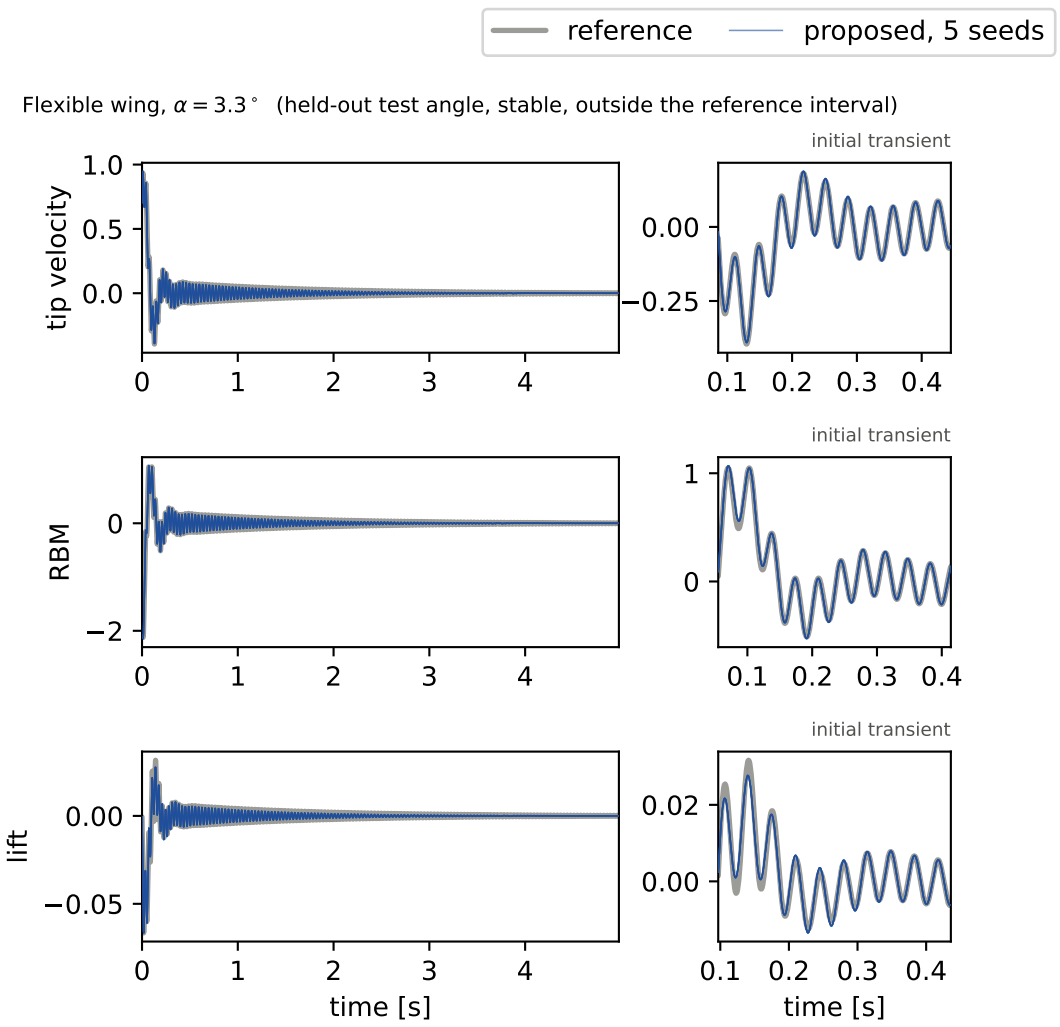


Figure 9: Wing at test angle $\alpha = 3.3°$, immediately above the upper flutter boundary. Grey denotes the reference and blue all 5 seeds. The right column shows the initial decay.

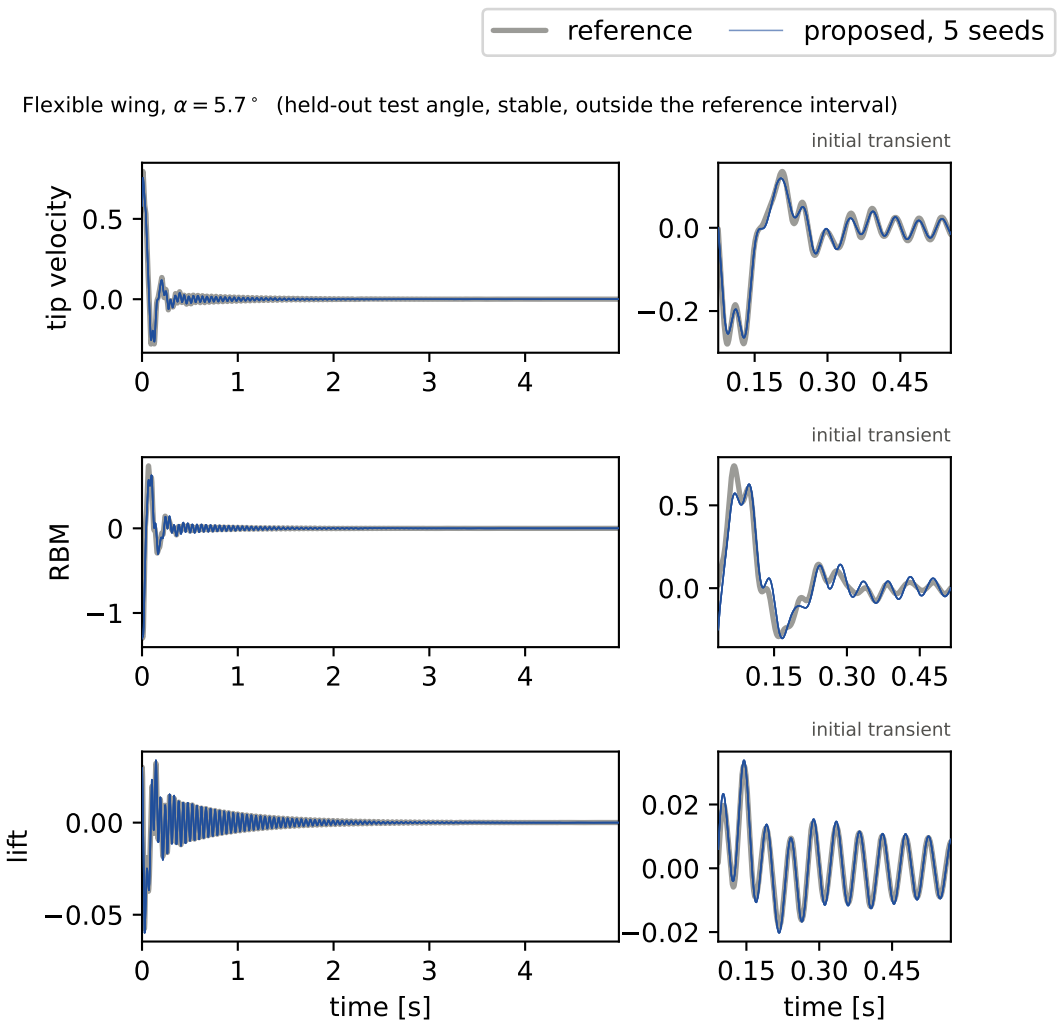


Figure 10: Wing at test angle $\alpha = 5.7°$, in the stable regime above the flutter interval. Grey denotes the reference and blue all 5 seeds. The right column shows the initial decay.

D.2 RIJKE TUBE

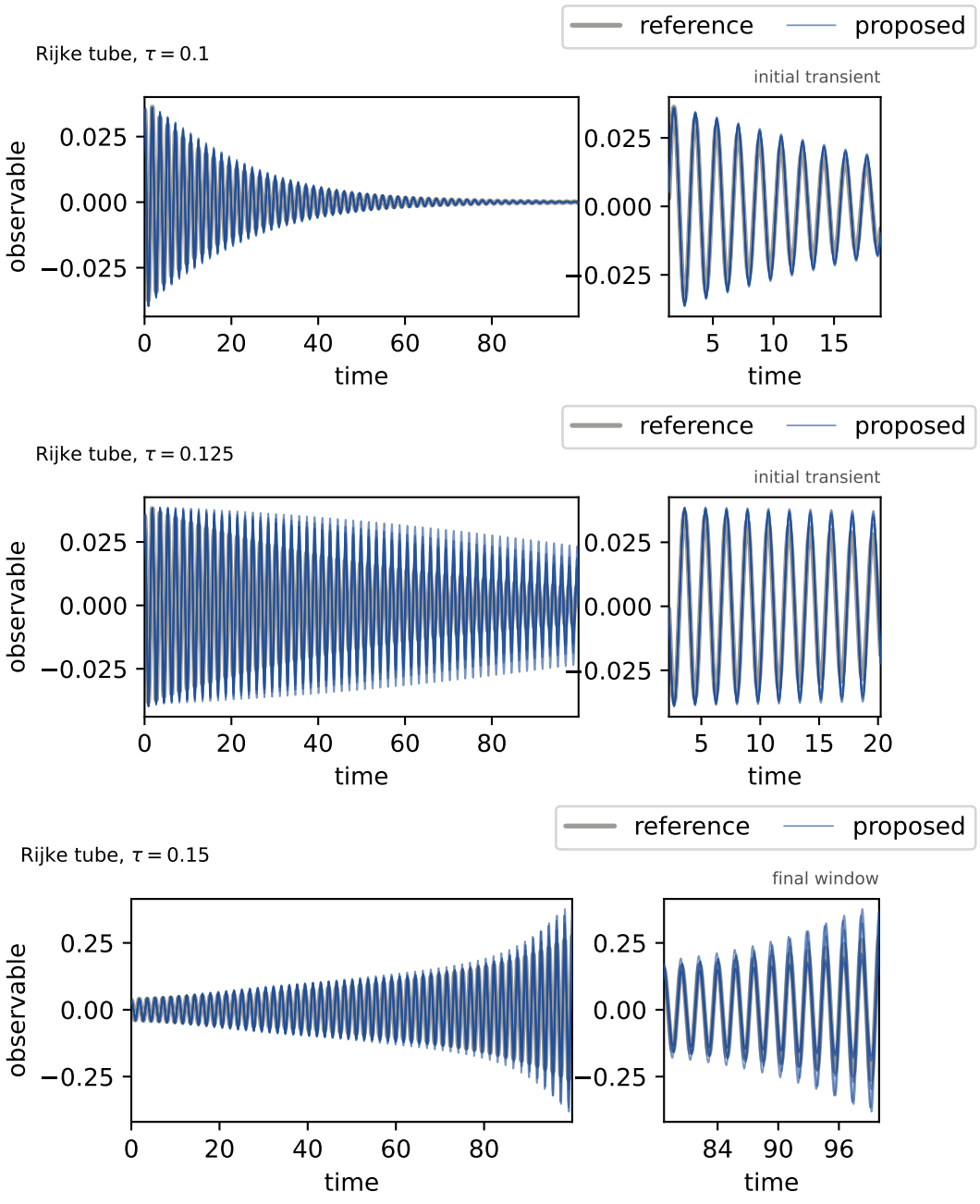


Figure 11: Tube near the lower stability boundary at test delays $\tau = 0.100$, 0.125 and 0.150, from top to bottom. The first two are stable. The third lies just inside the oscillatory interval. Grey denotes the reference and blue all 5 seeds.

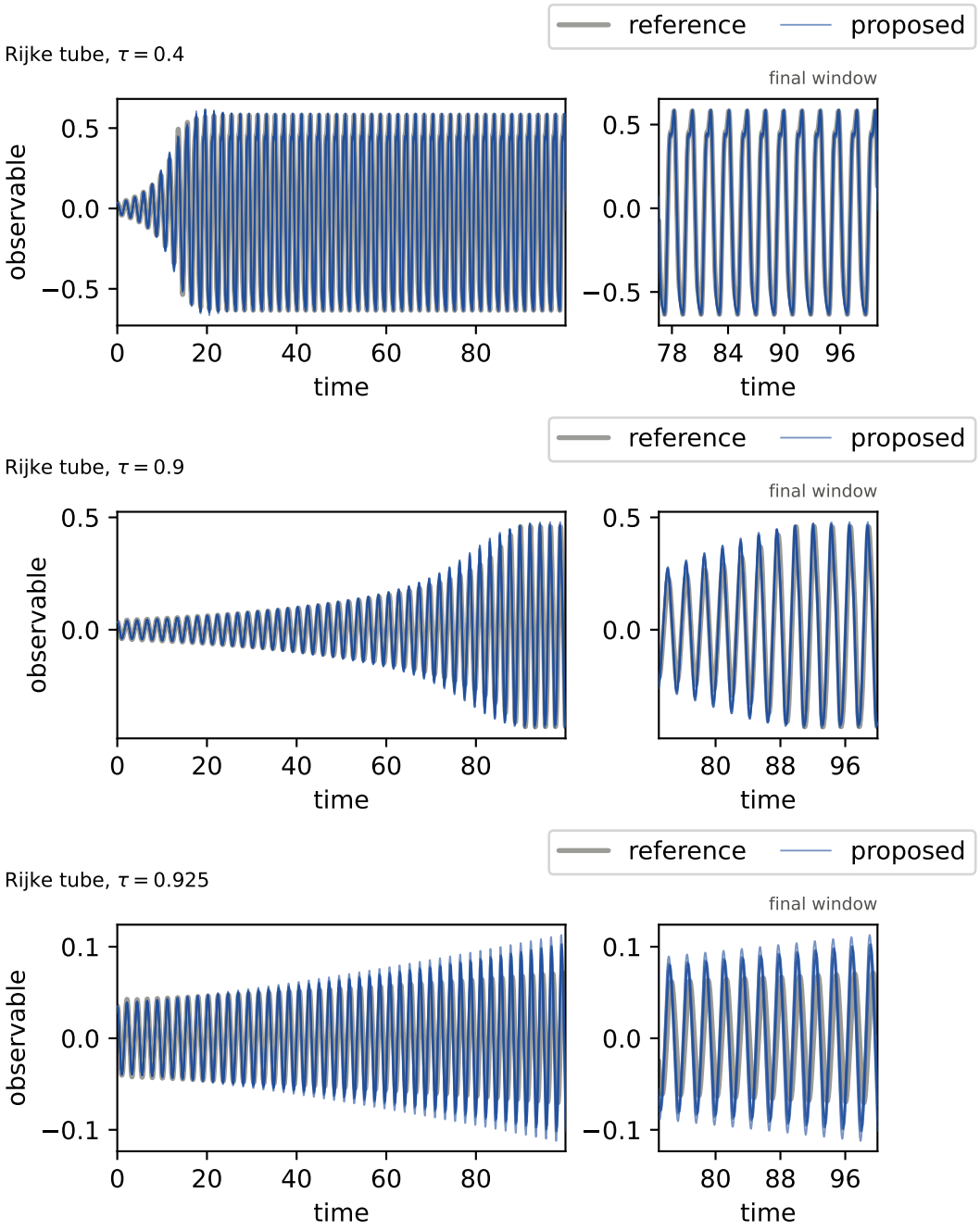


Figure 12: Tube inside the unstable interval at test delays $\tau = 0.400$, 0.900 and 0.925, from top to bottom. The latter two approach the upper crossing. Growth and phase discrepancies remain visible near this boundary. Grey denotes the reference and blue all 5 seeds.

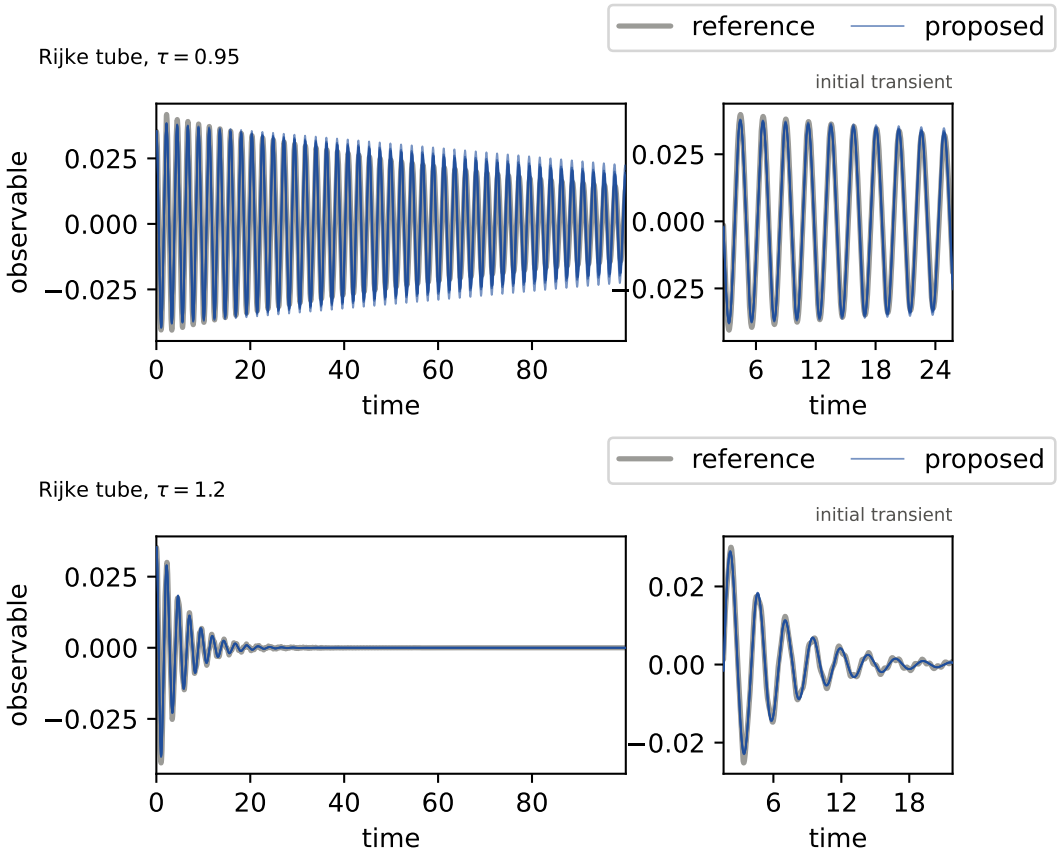


Figure 13: Tube above the upper stability boundary at test delays $\tau = 0.950$ (top) and 1.200 (bottom). The right panels resolve the initial decay. Grey denotes the reference and blue all 5 seeds.

D.3 CYLINDER WAKE

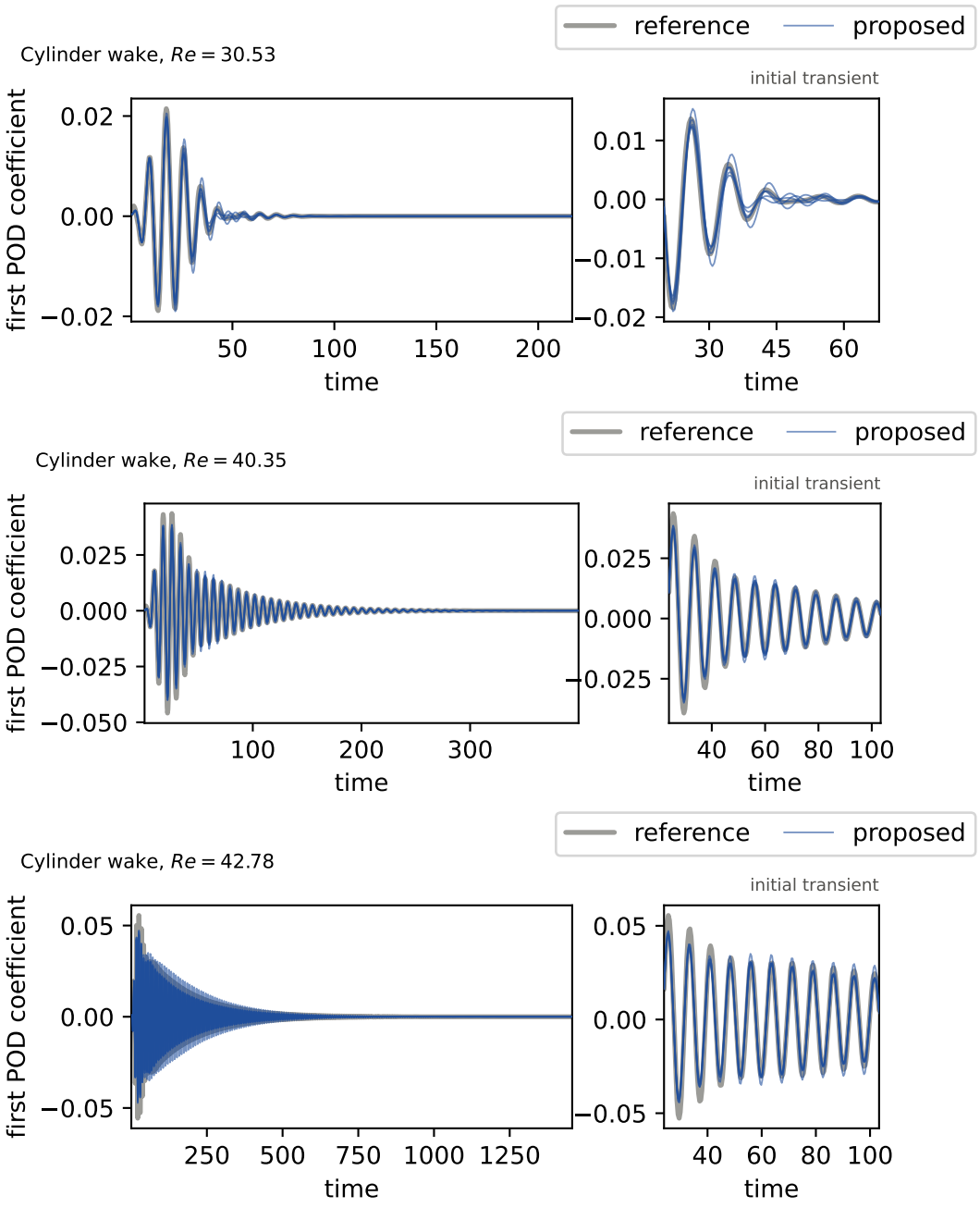


Figure 14: Wake below onset at test conditions $Re = 30.53$, 40.35 and 42.78, from top to bottom. The right panels resolve the decay of the first POD coefficient. Grey denotes the reference and blue all 5 seeds.

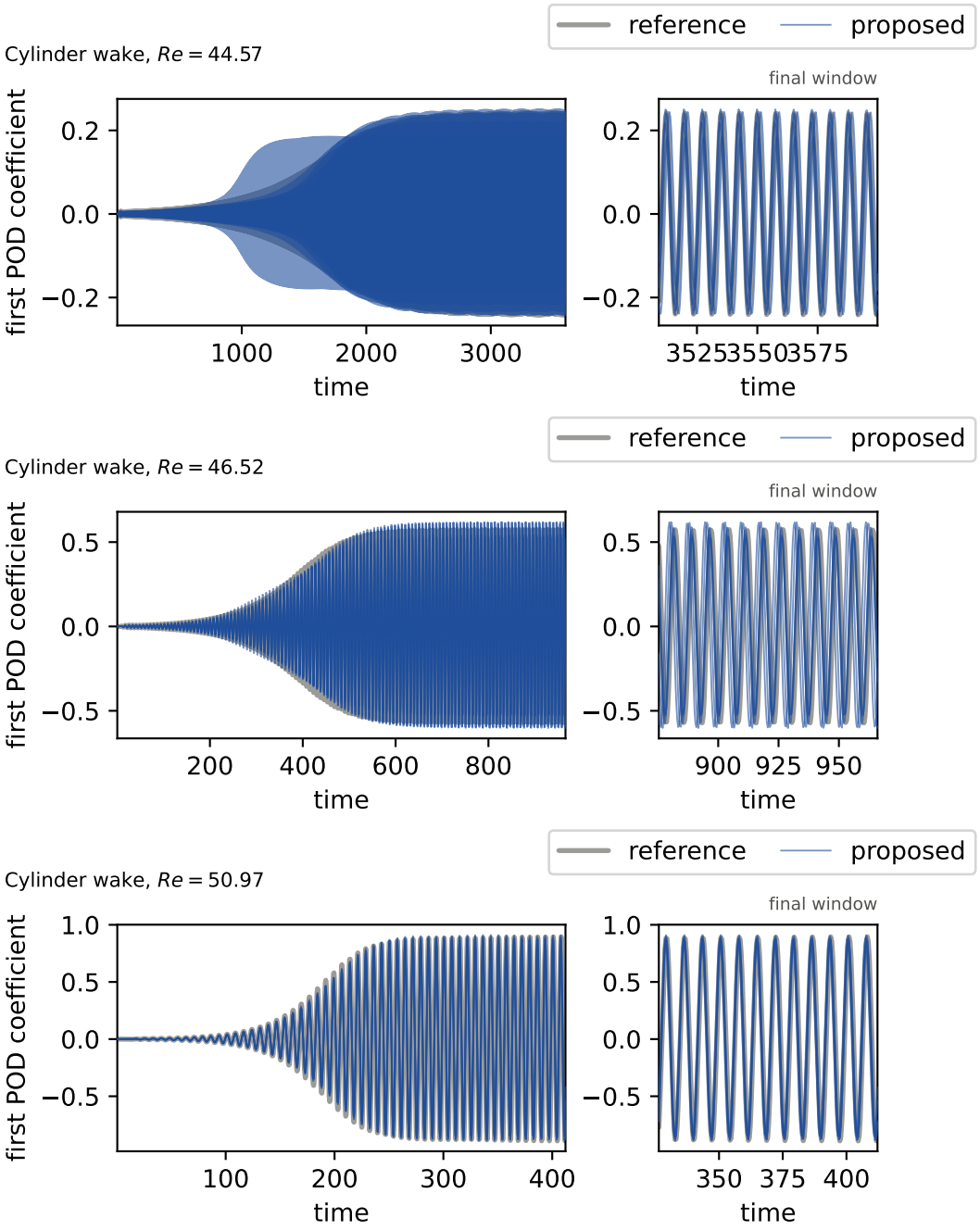


Figure 15: Wake above onset at test conditions $Re = 44.57$, 46.52 and 50.97, from top to bottom. Near onset, the whole-record panels reveal variability in growth timing even when the final oscillations agree closely. Grey denotes the reference and blue all 5 seeds.

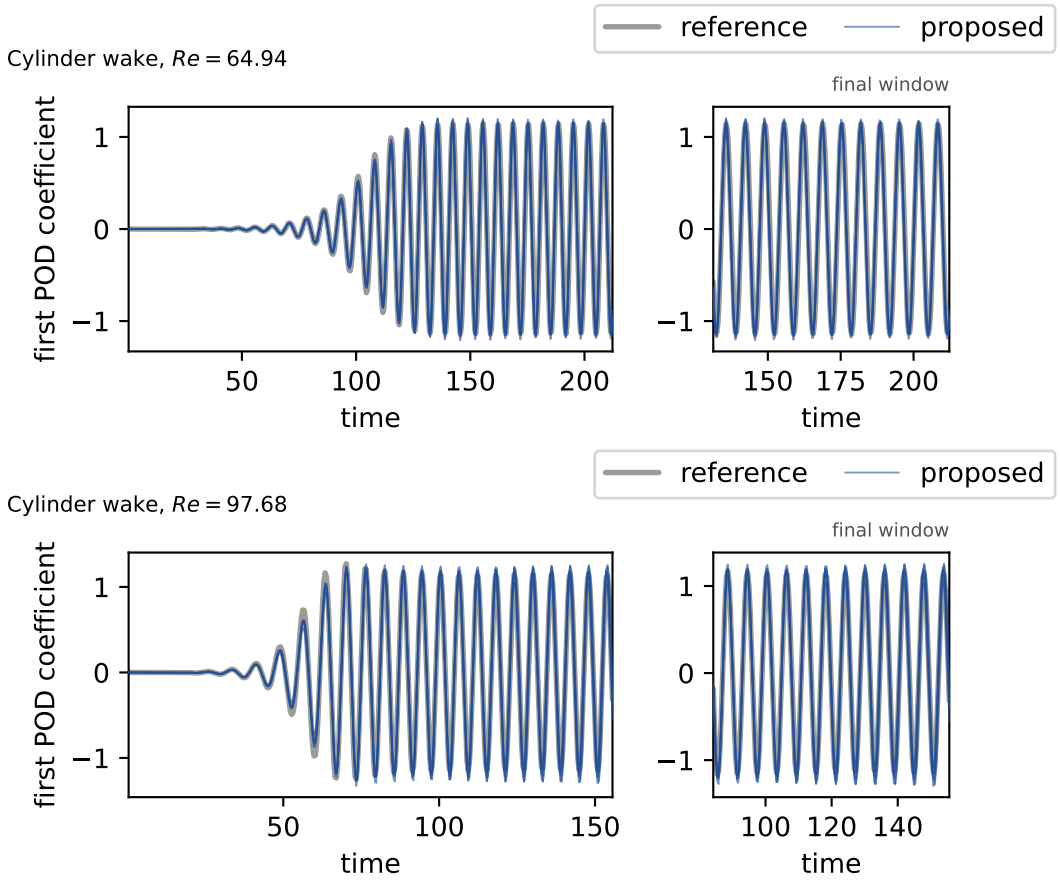


Figure 16: Wake with developed shedding at test conditions $Re = 64.94$ (top) and 97.68 (bottom). Grey denotes the reference and blue all 5 seeds.

D.4 WAKE FIELD RECONSTRUCTIONS

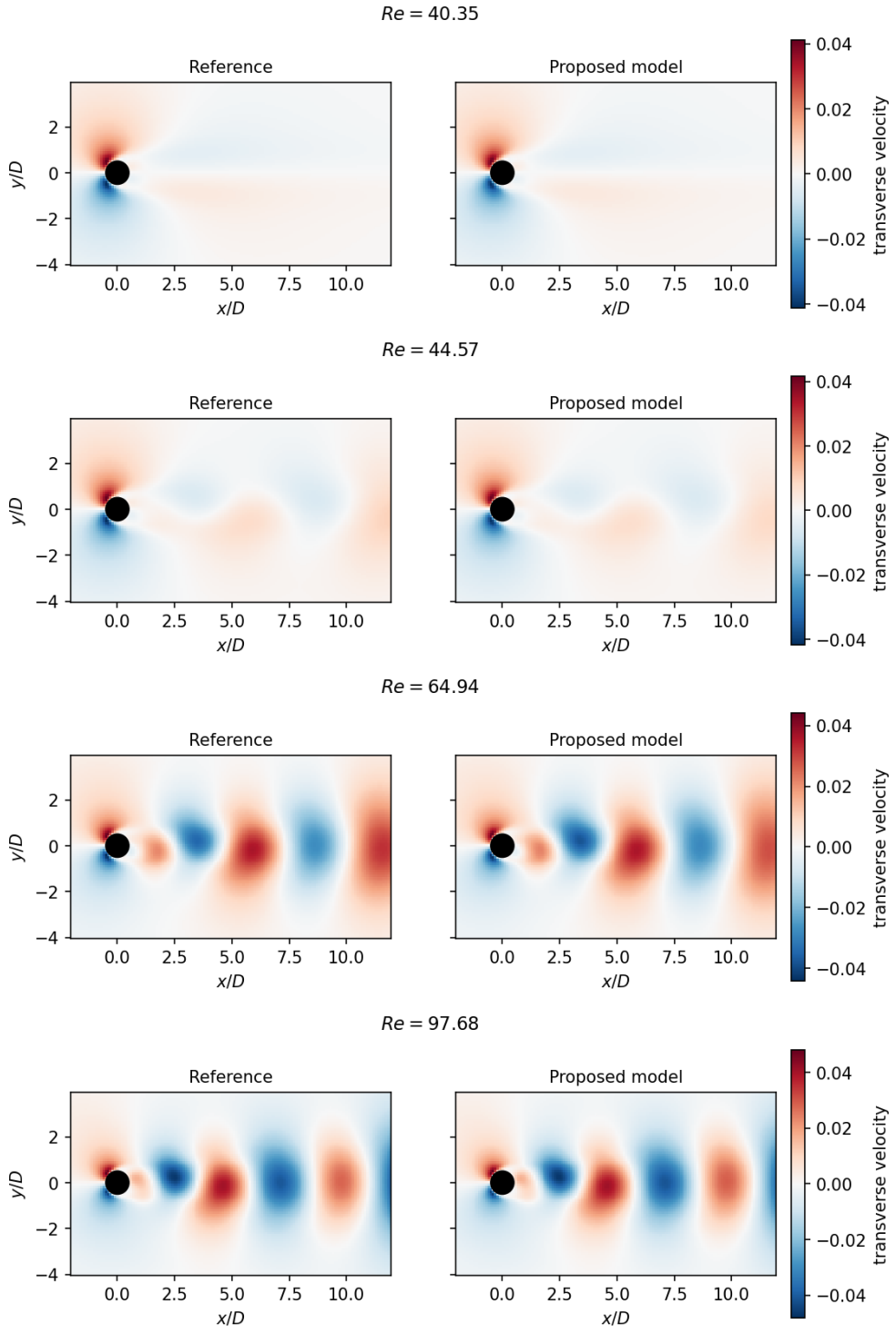


Figure 17: Transverse-velocity reconstructions at four test Reynolds numbers. Each row compares the reference (left) and proposed model, seed 0 (right), using the base flow and eight POD modes, with a shared colour scale within that row. The reference is at the final instant of its record. The model instant is selected within the final shedding period to minimise its state distance to that reference, aligning phase. These panels assess spatial structure after phase alignment. The trajectory panels above show the unshifted predictions.

E SUPPORTING RESULTS FOR SECTION 4.1, TRAJECTORY PREDICTION

This appendix supports Section 4.1. It gives the error pooled over all conditions for every method (Figure 18 and Table 8), the error inside and outside the unstable interval (Appendix E.1), the error at every condition for every architecture (Appendix E.2), the limit-cycle metrics (Appendix E.3), and two checks on whether the advantage of the proposed model is attributable to its architecture rather than to its training recipe or its fitting objective. The plain neural ODE trained conventionally, without the recipe, occupies the first lane of Figure 18 and fails in the same way, with whole-record errors of 1.001, 0.990 and 2.059, showing that poor trajectory scores also occur under conventional training.

The pooled result is in Table 8. The proposed model has the lowest error among the neural architectures (tied with the ungated model on the wing), 0.109 on the tube, 0.174 on the wake and 0.147 on the wing, and likewise on the held-out conditions alone, at 0.108, 0.169 and 0.147. In the sequence of architectures the two steps that lower the error on every system are the step to the coefficient-matrix form and the final step to the polynomial operator, and on the tube and the wake freezing the free operator lowers it as well. Read against the level of one, which is the error of a model that stays at the equilibrium, the plain neural ODE has comparable or larger trajectory error on the wake and wing under either recipe, as do both free-operator variants on the wing. This metric alone does not establish whether a limit cycle was learned, because phase drift can also produce errors above one. The three ablations in the second block remove one element of the construction each. Removing the gate leaves the pooled error unchanged on the wing, for the reason given in Appendix E.1, and releasing the operator or training jointly from scratch raises it on every system.

Table 8: Trajectory error of the neural architectures. The column *error* is the median over all conditions of the whole-record error of a free rollout, averaged over 5 seeds (four and three for the wing's free-operator rows), *held out* the same statistic over the test conditions alone, neither training nor development (37, 44 and 35 conditions), and a model that remains at the equilibrium scores one. The first block follows the architectural sequence, with training and tuning differences specified in Appendix B.3. The column *seeds won* gives the number of seeds on which that row has the smaller error than the row above, out of the seeds the two rows share, with the mean difference in brackets, positive when the lower row is better, a tie counting as not won and *identical* marking a tie on every seed. In the second block each row is an ablation of the proposed model and is compared with the proposed row in the same way. The lowest error on each system is printed in bold.

	Rijke tube			Cylinder wake			Flexible wing		
architecture	error	held out	seeds won	error	held out	seeds won	error	held out	seeds won
plain neural ODE	0.340	0.441		0.999	1.001		1.804	1.804	
with the full recipe	1.284	1.265	0/5 (-0.94)	1.047	1.108	0/5 (-0.05)	1.807	1.811	2/5 (-0.003)
coefficient-matrix form	0.284	0.293	5/5 (+1.00)	0.587	0.591	5/5 (+0.46)	1.021	1.021	5/5 (+0.79)
free operator, trained jointly	0.295	0.303	3/5 (-0.01)	0.662	0.672	1/5 (-0.07)	7.765	7.945	0/4 (-6.74)
free operator, frozen	0.116	0.115	5/5 (+0.18)	0.240	0.233	5/5 (+0.42)	66.406	66.406	0/2 (-49.22)
polynomial operator (proposed)	**0.109**	**0.108**	3/5 (+0.01)	**0.174**	**0.169**	5/5 (+0.07)	**0.147**	**0.147**	3/3 (+66.26)
proposed, without the gate	0.255	0.270	0/5 (-0.15)	0.423	0.430	0/5 (-0.25)	0.147	0.147	identical
proposed, operator released	0.145	0.147	0/5 (-0.04)	0.362	0.375	0/5 (-0.19)	0.534	0.534	0/5 (-0.39)
proposed, trained jointly from scratch	0.471	0.489	0/5 (-0.36)	1.125	1.230	0/5 (-0.95)	6.268	6.347	0/5 (-6.12)

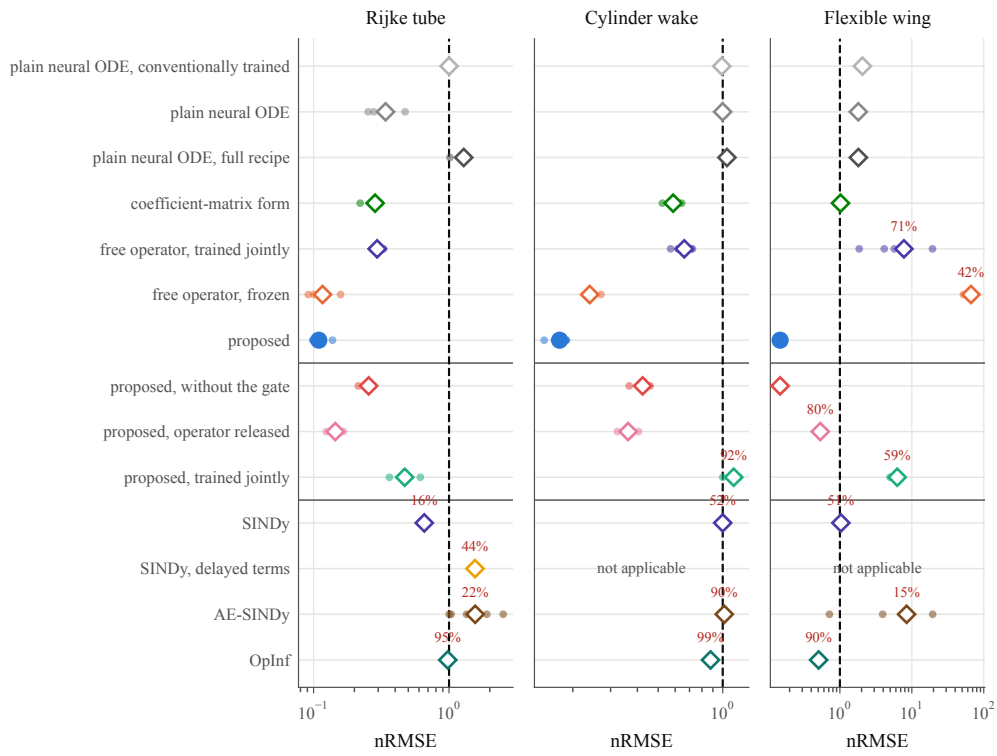


Figure 18: Normalised trajectory error (nRMSE) of every method on the three systems, on a logarithmic axis. Each lane is one method. A small filled marker is one training seed, and the large marker is the mean over seeds. The value for one seed is the median over all conditions of the whole-record error of a free rollout. The dashed vertical line at one is the error of a model that remains at the equilibrium. The upper block contains the sequence of neural architectures, with the plain neural ODE also shown under conventional training in its first lane, the middle block the three ablations of the proposed model, and the lower block the existing methods, which are deterministic and have a single value, except the autoencoder sparse regression, which is drawn at its 5 seeds. A red percentage above a marker is the share of conditions at which the rollout of that method remains bounded, printed only where it is below 100, and the error of such a method is computed over the bounded conditions only, which favours it.

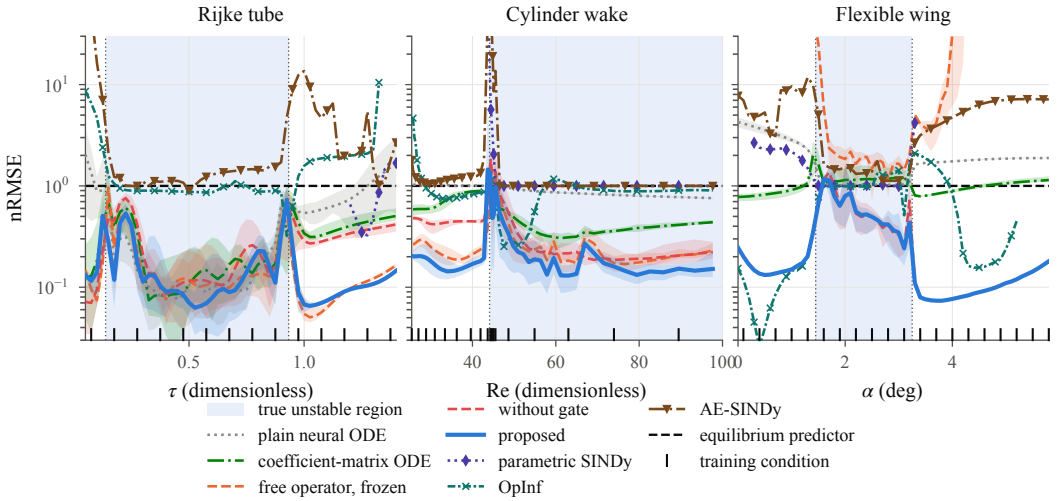


Figure 19: Complete per-condition trajectory comparison, including the coefficient-matrix, frozen free-operator and ungated architectural variants. Colours and metrics match Figure 2. Bands span the training seeds. All curves use the same archived results as the main figure.

E.1 Error inside and outside the unstable interval

Table 9: Median normalised trajectory error inside and outside the true unstable interval, pooling training, development and test conditions and all seeds. Inside the interval the reference settles onto a limit cycle, and outside it the reference decays.

	Rijke tube		Cylinder wake		Flexible wing	
	inside	outside	inside	outside	inside	outside
plain neural ODE	0.12	0.71	0.87	1.00	1.02	1.87
with the full recipe	1.35	1.13	1.00	1.16	1.15	1.88
coefficient-matrix form	0.15	0.38	0.41	0.66	1.07	0.93
free operator, trained jointly	0.12	0.40	0.43	0.72	1.58	12.38
free operator, frozen	0.13	0.10	0.24	0.23	1.93	79.34
proposed	**0.11**	**0.10**	**0.18**	**0.17**	**0.54**	**0.13**
proposed, without the gate	0.15	0.32	0.24	0.45	0.52	0.13

Table 10: The split of Table 9 over the test conditions alone, neither training nor development (37, 44 and 35 conditions), all seeds pooled.

	Rijke tube		Cylinder wake		Flexible wing	
	inside	outside	inside	outside	inside	outside
plain neural ODE	0.12	0.69	0.86	1.00	1.02	1.87
with the full recipe	1.34	1.19	1.00	1.17	1.15	1.88
coefficient-matrix form	0.15	0.38	0.38	0.65	1.07	0.93
free operator, trained jointly	0.12	0.41	0.39	0.72	1.67	11.89
free operator, frozen	0.13	0.10	0.23	0.23	2.01	76.62
proposed	**0.13**	**0.10**	**0.16**	**0.17**	**0.55**	**0.13**
proposed, without the gate	0.15	0.31	0.21	0.45	0.55	0.13

On the wing the model without the gate has the same median over all conditions as the proposed model, 0.147, on every seed. The median over the conditions of the wing falls on a stable condition. Agreement of these medians does not imply equality of the two vector fields or their Jacobians. Inside the unstable interval the two differ, with 0.52 against 0.54. On the tube and the wake the median falls inside the unstable interval and the two models differ on every seed.

E.2 Error at every condition for all architectures

Figure 20 resolves the pooled numbers by condition. Three features recur on all three systems. The error of the proposed model is smallest away from the crossings and largest at the first conditions past onset, where the growth is slowest and a small error in the rate changes the whole record. The unstructured models sit at or above one across the unstable interval of the wake and wing, indicating large trajectory errors. On the wing the jointly trained model and both free-operator models exceed the plotted error range at some conditions. Boundedness is evaluated separately. Crossing an axis limit is not itself a divergence test.

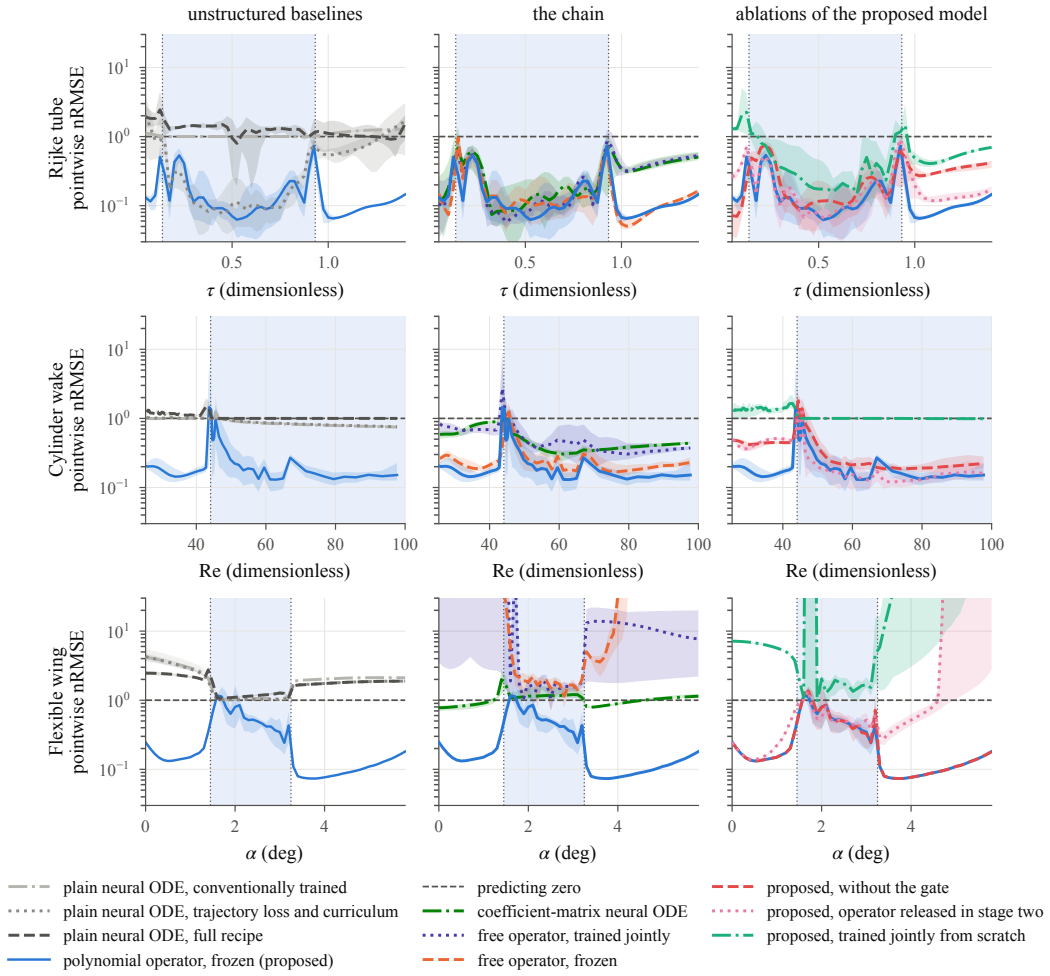


Figure 20: Normalised trajectory error at every condition, on a logarithmic axis, with the range over seeds. The rows are the three systems and the columns are the unstructured baselines, the sequence of architectures and the ablations of the proposed model. The dashed line at one is a model that remains at the equilibrium. The vertical axis is truncated: values above its limit exceed the displayed error range. Boundedness uses the separate criterion in Appendix C.

E.3 Limit-cycle metrics

The whole-record error accumulates phase, so Appendix C defines three metrics of the limit cycle that do not, namely the error of the settled amplitude, the error of the phase-averaged cycle and the relative error of the oscillation frequency. Table 11 gives them for every architecture. They separate two outcomes that the whole-record error does not distinguish. A model with a small amplitude error and a frequency error of a few percent, such as the plain neural ODE with the full recipe on the tube, has found the limit cycle and drifts in phase. A model whose amplitude and frequency errors are both close to one, such as every unstructured model on the wake, has not found it.

Table 11: Limit-cycle metrics and frequency diagnostics, as the mean over seeds. The amplitude and cycle errors are normalised so that a model that remains at the equilibrium scores one. Frequency errors use nonlinear rollouts except for daggered wing entries, which compare an equilibrium eigenfrequency with the reference carrier and are separate linear diagnostics. A dash denotes an unavailable or undefined quantity. Wing values are medians over the reported unstable-band conditions. The proposed wing model's rollout frequency error is approximately 0.12 percent (0.09–0.16 across seeds), rounded to 0.001 in relative units.

	Rijke tube			Cylinder wake			Flexible wing		
	amp.	cycle	freq.	amp.	cycle	freq.	amp.	cycle	freq.
plain neural ODE	0.031	0.045	< 0.001	0.967	0.993	0.958	0.968	0.998	—
with the full recipe	0.068	0.179	0.029	0.983	0.999	0.959	0.786	0.960	—
coefficient-matrix form	0.007	0.027	0.001	0.169	0.201	0.005	0.879	0.892	$0.391^{\dagger}$
free operator, trained jointly	0.010	0.035	0.001	0.167	0.194	0.006	0.769	0.763	$0.007^{\dagger}$
free operator, frozen	0.010	0.037	0.001	0.047	0.083	0.002	1.197	1.227	$0.022^{\dagger}$
proposed	**0.015**	**0.037**	**0.001**	**0.034**	**0.064**	**0.002**	**0.176**	**0.185**	**0.001**
proposed, without the gate	0.014	0.048	0.001	0.046	0.082	0.002	0.067	0.082	$0.066^{\dagger}$
proposed, operator released	0.009	0.035	0.001	0.024	0.055	0.001	0.106	0.115	$0.029^{\dagger}$
proposed, trained jointly from scratch	0.021	0.075	0.002	0.993	0.995	0.188	0.618	0.614	$0.014^{\dagger}$
OpInf	—	—	—	0.615	0.639	0.233	—	—	—
AE-SINDy	0.719	0.782	0.137	0.997	0.998	0.003	0.804	—	0.400
remaining at the equilibrium	1.000	1.000	—	1.000	—	—	—	—	—

E.4 Training recipe or architecture

The plain neural ODE comparisons test whether supplying the training recipe alone reproduces the proposed model's trajectory accuracy. Under the full recipe, the unstructured network still has errors comparable to or larger than the equilibrium predictor on the wake and wing (Table 8). Thus the recipe alone is insufficient. This comparison does not rule out interactions between architecture and optimisation, or improvements under other training choices.

E.5 First-stage objective and stability information

We examine how the fitting objective, window length and stability penalty influence the recovered crossings. The wake controls have the degree and initialisation differences described below, so their comparison is not a fully controlled change of objective alone.

Window length. Figure 21 refits the operator at 6, 7 and 6 fitting-window lengths on the three systems, spanning 1.4 to 146 carrier cycles. It reports the lower crossing under the convention of Table 4, together with the crossing count. On the tube, trajectory fitting returns both crossings at every length. Its shortest- and full-window errors are 0.74 and 0.17 percent of the unstable interval. Derivative matching returns no crossing at the 2 shortest windows and never both crossings. Its full-window error is 8.13 percent.

On the wake, derivative matching is competitive through roughly twelve cycles. At the full window, the trajectory and derivative errors are 0.09 and 0.93 percent of the critical Reynolds number. Both full-window operators have effective degree two, whereas the trajectory refit in Table 12 has degree four. At the 3.2-cycle window the derivative operator has nonzero degree-three and degree-four blocks. The full-window derivative fit starts from its own least-squares fit, while a fit from the shared initialisation gives a 1.15% derivative error.

On the wing, the full-window trajectory and derivative errors are 1.39 and 2.87 percent, respectively. Derivative matching also predicts a spurious third crossing at one shorter window. Across the three systems the errors are not monotonic in window length. The full-window results are consistent with the sensitivity argument of Section 3, but do not establish a general monotonic benefit from longer windows or a trajectory advantage at every length.

Stability information. Table 12 crosses the objective with the stability penalty at the full window. Removing the penalty omits the stability regularisation on stable training records. The derivative controls have the degree and initialisation differences listed in the caption. Without the penalty the trajectory objective still returns both crossings of the tube, at 0.08 and 0.34 percent, the wake

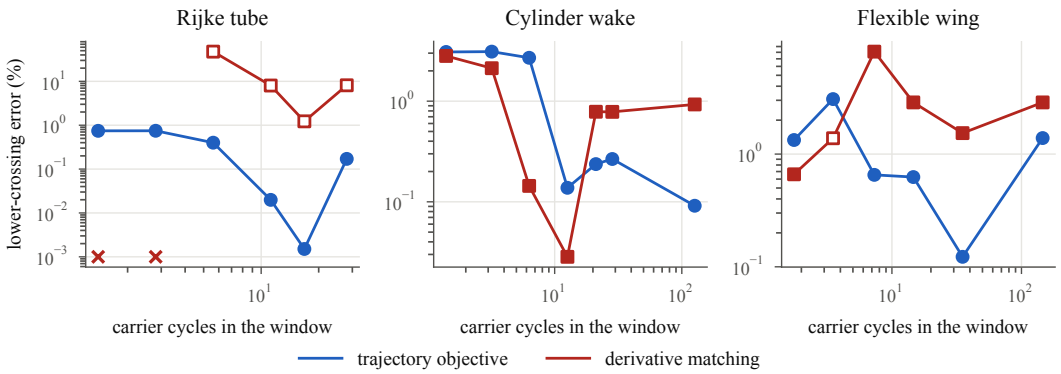


Figure 21: Lower-crossing error against the number of carrier cycles in the first-stage fitting window, for the trajectory objective (circles) and derivative matching (squares). Filled markers report the true number of crossings, hollow markers a wrong count, and crosses no crossing at all. The wake's 3.2-cycle derivative operator includes degree-three and degree-four blocks. Its full-window derivative fit uses its own least-squares initialisation.

crossing at 0.04 percent, and both crossings of the wing, with the lower one 8.40 percent off against 1.39 with the penalty. In these refits the penalty has little effect on crossing accuracy on the tube and wake, and improves the wing's lower crossing. The reported derivative fits have larger crossing-location errors where both entries are available. Without the penalty they return an incorrect crossing count on all three systems.

The penalised trajectory rows are independent refits under the reported recipe and serve as controls for the other rows. Their crossings agree with the reported operators in Table 4 to within 0.09 percent of the tube interval and 0.06 percent of the critical Reynolds number on the wake.

Table 12: First-stage objective against stability information, at the full window. Crossing error in percent at the lower and upper edge of the unstable interval (tube, wing) or at the single crossing (wake), sign score in percent and crossing count, as in Table 4. The penalty regularises stable training records. The penalised derivative fits share the trajectory initialisation on the tube and wing. The wake uses its own derivative initialisation. The derivative fits without the penalty are the closed-form linear fits of Table 4. On the wake the two trajectory fits and the derivative fit without the penalty have degree four, and the penalised derivative fit has effective degree two.

	Rijke tube				Cylinder wake			Flexible wing			
first-stage objective	lower	upper	sign	n	crossing	sign	n	lower	upper	sign	n
trajectory, with the penalty	0.17	0.18	100.0	2	0.15	100.0	1	1.39	0.97	100.0	2
trajectory, without	0.08	0.34	100.0	2	0.04	100.0	1	8.40	1.28	96.6	2
derivative, with the penalty	8.15	108.15	40.0	1	0.93	—	1	2.87	3.56	96.6	2
derivative, without	137.23	37.23	70.9	1	4.16	89.6	2	68.50	—	—	3

E.6 Comparison with optimised-DMD-based parametric estimators

Optimised DMD fits exponential trajectories by variable projection (Askham & Kutz, 2018). We compare our identified operator with parametric estimators built from such fits on the training conditions and first-stage windows (Table 13). The fits use the same state coordinates and initialise continuous-time eigenvalues from conventional DMD. They use six, eight and nine modes on the tube, wake and wing, respectively, with conjugate pairs enforced. On the tube and wake, the resulting operator entries are fitted by a polynomial in the operating-parameter coordinate. The wing's primary comparator instead regresses a shared polynomial operator jointly across conditions using derivatives obtained from the fitted exponentials. The per-condition-operator interpolation is reported separately. These estimators use no stability penalty. The trajectory control without the penalty in Table 5 instead optimises a shared matrix polynomial jointly across conditions through equation 4 with all stability penalties removed, whereas the estimators in Table 13 combine per-condition variable-projection fits with subsequent parameter coupling. Thus this comparison evaluates complete first-stage procedures, including their parameter coupling and regularisation. It does not isolate the trajectory objective alone.

Fitting and convergence. The tube uses Levenberg–Marquardt variable projection, with function and step tolerances 10^{-10} and a cap of 400 function evaluations per parameter. All 14 fits meet the stopping criteria. Polynomial degrees one to three are reported both with the known acoustic block and structural mask used by the proposed operator, and without those constraints ("raw"). The wake uses real sine–cosine exponential bases and trust-region reflective least squares, with function, step and gradient tolerances 10^{-12} and a cap of 20,000 evaluations. Twenty of its 21 fits meet the tolerances. The remaining fit, at $Re = 154.21$, reaches the cap: its relative residual changes from 0.005862 at 400 evaluations to 0.005818 at 20,000. This condition lies outside the reported $Re \leq 100$ range but contributes to the parameter fit. Both degree four, matching the proposed operator, and degree one are reported. All 20 wing variable-projection fits meet the stopping criteria. Its joint regression selects degree three from degrees one to three using development rollout error, whereas the proposed operator has degree two. For the interpolation diagnostic, all three degrees diverge on development conditions, so degree does not change that outcome and degree one is shown. No mode-count or regularisation sweep is performed for these optimised-DMD-based estimators.

Table 13: First-stage comparison with optimised-DMD-based parametric estimators. D is the polynomial degree. Crossing errors are percentages of the unstable-interval width (tube and wing) or critical Reynolds number (wake), each reference crossing scored against the nearest prediction. n is the total predicted crossing count (true counts: two, one and two). A dash denotes no predicted crossing or no upper reference crossing. Sign accuracy is over test conditions only: 37 tube, 44 wake and 35 wing. Rate error is the median absolute growth-rate difference on those conditions, in inverse record-time units (tube), inverse convective-time units (wake) and s^{-1} (wing). The wing rate score uses only the 25 stable test angles with reference envelope slopes and is a finite-window diagnostic. All tube variants tested are shown. The wing interpolation row is a diagnostic alternative to its joint regression. Protocols and limitations are described in Appendix E.6.

System	First-stage estimator	D	Lower	Upper	n	Test sign (%)	Rate error
Rijke tube	Proposed operator	1	0.08	0.25	2	100.0	0.00506
	OptDMD + polynomial, structured	1	126.87	26.87	2	73.0	0.139
	OptDMD + polynomial, raw	1	129.78	29.78	2	70.3	0.117
	OptDMD + polynomial, structured	2	126.11	26.11	1	75.7	0.156
	OptDMD + polynomial, raw	2	127.23	27.23	1	75.7	0.0991
	OptDMD + polynomial, structured	3	0.62	14.39	3	89.2	0.106
	OptDMD + polynomial, raw	3	2.69	5.61	3	91.9	0.052
Cylinder wake	Proposed operator	4	0.09	—	1	100.0	0.00218
	OptDMD + polynomial	4	25.03	—	2	43.2	3.32
	OptDMD + polynomial	1	—	—	0	43.2	2.5
Flexible wing	Proposed operator	2	1.26	1.04	2	100.0	0.477
	OptDMD-derived joint regression	3	43.85	56.15	4	51.4	2.4
	OptDMD + polynomial (diagnostic)	1	—	—	0	28.6	474

Results and interpretation. The proposed operator gives better boundary recovery and higher test sign accuracy than the tested alternatives. On the tube, increasing the polynomial degree substantially improves the optimised-DMD-based estimators, but both degree-three variants predict three crossings instead of two. On the wake, the degree-four estimator reports crossings at $Re = 33.0127$ and 33.0695, against the reference 44.11, and correctly classifies 19 of 44 test conditions. The proposed operator correctly classifies all 44. These two nearby roots come from the estimator's dense scan. Neither is resolved by the coarser record grid. The wing joint fit predicts four crossings and correctly classifies 18 of 35 test conditions (31 of 59 over the full reporting grid). Its first two reported roots are 0.414853° and 2.239238°. The remaining sign changes lie in $[4.7, 4.8]^\circ$ and $[5.5, 5.6]^\circ$ and do not change either nearest-crossing error. The proposed operator correctly classifies all 35 test conditions.

These results support better stability recovery by the proposed identification procedure for the tested constructions. They do not establish superiority over optimised DMD in general, identify parameter interpolation as the sole source of error, or compare nonlinear models trained with alternative first-stage operators. No nonlinear residual was trained on these baseline operators. Reference-resolution qualifications remain as in Appendix F.4.

F SUPPORTING RESULTS FOR SECTION 4.2, PREDICTION OF THE STABILITY BOUNDARY

This appendix supports Section 4.2. It gives the eigenvalue locus of the frozen operator, the constant predictors against which the sign score is read, the sign score on held-out conditions and on the conditions nearest the crossings, the sources of variation of the reported crossings, and two points on how the reported crossings were obtained and scored.

F.1 ROOT LOCUS OF THE FROZEN OPERATOR

Because the equilibrium Jacobian of the deployed model is the explicit operator $\boldsymbol{A}(p)$, its spectrum can be computed directly at every parameter value, including values not used for fitting, without integrating a trajectory. Figure 22 draws the eigenvalues in the complex plane as the parameter is swept. Each Hopf crossing appears as a complex-conjugate pair passing through the imaginary axis, the frequency at the crossing is read from the imaginary part, and the remaining eigenvalues show which modes stay damped across the range. On the tube and the wing one pair crosses into the right half-plane and returns, which is the closed unstable interval, and on the wake one pair crosses once and does not return. This is the standard object of a linear stability analysis, obtained here from a model identified from data alone. The equilibrium Jacobian of an unstructured neural ODE is available at each parameter value as well, by differentiation after training. What distinguishes the operator is that its spectrum was identified separately, before the residual, and is unchanged by the residual's training (Proposition 1), so the locus is fixed once the first stage ends.

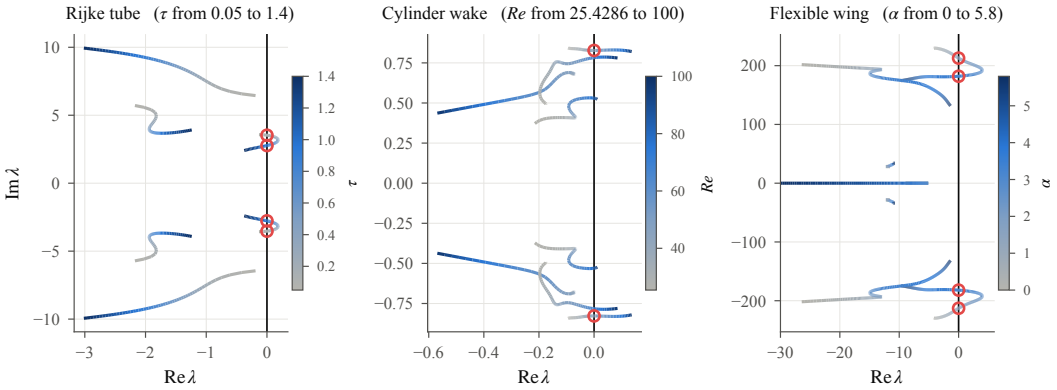


Figure 22: Eigenvalues of the frozen operator in the complex plane as the physical parameter is swept, coloured by the parameter. Crossings of the imaginary axis by the rightmost eigenvalue are circled. The whole locus is available before any rollout is computed, and at parameter values that were never fitted, because the spectrum is a component of the model rather than a quantity measured from its output. The wing panel is limited to the modes near the imaginary axis, and the remaining, more strongly damped eigenvalues lie to the left of it.

F.2 CONSTANT PREDICTORS

The trajectory error is read against a model that remains at the equilibrium, which scores one. The sign score needs a corresponding reference level, which depends on the share of stable conditions in each system. On the tube an always-stable predictor classifies 41.8 percent of the conditions correctly, and a model that calls every condition unstable scores 58.2 percent. On the wake the sign is scored over the 67 conditions of the reported range, up to a Reynolds number of 100, and the 8 conditions of the sweep above it are excluded. A model that calls every condition stable scores 53.7 percent, and one that calls every condition unstable scores 46.3 percent. On the 48 held-out conditions of that range the better constant predictor scores 56.25 percent. The subsets are not balanced, so one half is not the level against which a sign score on this system is read. The wing has 59 scored conditions, of which 18 are unstable and 41 are stable, so a model that calls every condition stable scores 69.5 percent and one that calls every condition unstable scores 30.5 percent.

Read against these levels, several models have not learned the stability of the system. On the wake the plain neural ODE, in all three of its variants, has the trajectory error of a model that remains at the equilibrium and the sign score of a constant predictor on every seed. Of the 14 models trained on the wing, 4 fit the data, 7 do not, and for 3 the seeds disagree. On that system no operator other than the frozen polynomial one yields a model that fits on every seed.

Trajectory accuracy and spectral accuracy fail independently. On the tube the proposed model trained jointly from scratch has a moderate trajectory error, 0.471, while its sign score of 41.8 percent equals that of a constant predictor on every seed and it reports no crossing. A trajectory-only score would not expose its incorrect stability structure. Conversely, on the tube the plain neural ODE with the full recipe scores above one on the whole-record error while reporting two crossings and a sign score of 92.0 percent. Its limit-cycle metrics show that it fits the data, with an amplitude error of 0.068 and a frequency error of 0.029, and that its whole-record error is the accumulation of phase described in Appendix C. Both properties therefore have to be checked. The same holds for the wing model trained jointly from scratch, whose whole-record error is 6.268 while its amplitude and cycle errors are 0.618 and 0.614, against 0.176 and 0.185 for the proposed model.

The wing's joint parametric linear control behaves in the opposite way. Its trajectory error is 1.005, that of a model that remains at the equilibrium, while it classifies 86.4 percent of the conditions correctly. It recovers part of the stability structure but does not improve on the equilibrium predictor's aggregate trajectory error.

F.3 Sign score on held-out conditions and near the crossings

The sign score of Table 4 is computed over every condition of the reported range. Because the stability penalty acts at the training conditions, the score is also computed over the held-out conditions only. There, the sign score of the proposed model is 100.0, 100.0 and 100.0 percent on the three systems, and the largest change for any other method is 5.9, 12.3 and 4.2 percentage points. On the wing the unstructured models score 30.8 percent on the held-out conditions, which is the score of a model that calls every condition unstable. No conclusion of Section 4.2 changes.

The score is also computed over the 12 conditions nearest the crossings (6 on the wake), with equal numbers on either side of each crossing, so that a constant predictor scores exactly one half. A model that gives a constant prediction therefore moves to one half on this subset, which reflects the change of reference level and not an improvement of the model.

F.4 Sources of variation of the reported crossings

Proposition 1 fixes the equilibrium spectrum conditional on the identified first-stage operator. Changing the identification data, initialisation or polynomial degree can change the estimated crossings. Table 14 collects diagnostics of these effects alongside the reference resolution. The entries use different statistics, as specified below, and do not form an additive error budget.

F.5 Selection of the polynomial degree

The degree of the matrix polynomial is fixed per system, 1 on the tube, 2 on the wing and 4 on the wake, and the following checks assess sensitivity to that choice (Appendix B.1). The tube check uses development growth-rate error. The wake comparison uses development trajectory error among the candidate operators. On the tube, operators of degrees one, two and three all report the two crossings, with errors of 0.084, 0.037 and 0.138 percent at the lower edge and 0.188, 0.051 and 0.367 percent at the upper edge, and the error of the measured growth rate on the development conditions differs by 1.1 percent across the three. On the wake, the operator of degree one and the reported operator of degree 4 both report the single crossing, at 0.0607 and 0.0936 percent of the critical Reynolds number, and their trajectory errors differ by less than the spread over the seeds, the degree-one operator having the smaller error on 4 of 5 seeds. On the wing, an operator of degree three fitted with the reported recipe returns two crossings, with lower- and upper-crossing errors of 1.95 and 0.10 percent of the interval against 1.26 and 1.04 percent at the reported degree two.

The linear operator-inference control of Table 4 uses on the tube a loss that is relative per condition, as the first stage does. With every sample weighted equally the longest and largest records dominate

Table 14: Crossing diagnostics, normalised by the unstable-interval width (tube, wing) or critical Reynolds number (wake), in percent. Stage-two seed variation is zero for a fixed operator. For starts perturbed by 5 percent, the tube entry is the range of onset predictions. The wake and wing entries are maximum deviations from the unperturbed refit, using the upper crossing on the wing. For varied training sets, the tube reports the largest reference error across 4 leave-one-out refits, the wake the onset range across 7 training sets, and the wing the onset range across 20 leave-one-out refits. The degree row gives the change in absolute onset error between degrees one and two (tube), or four and one (wake), rather than crossing displacement. The wing entry compares absolute onset errors at degrees three and two. Reference resolution is the reference bracket's half-width. The tube reference is exact. The final row gives the reported lower-crossing error.

diagnostic	Rijke tube	Cylinder wake	Flexible wing
random seed, stage two	0.000	0.000	0.000
perturbed start	0.036	0.048	2.50
training set varied	0.52	0.62	4.63
polynomial degree	0.047	0.033	1.95 vs. 1.26
reference resolution	0	1.04	2.78
reported crossing error	*0.08*	*0.09*	*1.26*

the fit. With the relative weighting its fit of degree three reaches 0.61 and 3.13 percent at the two best of its 4 crossings, and at the reported degree it reports 1 crossing where there are two. The values in Table 4 are those of the relative weighting.

F.6 Scoring of methods that report several crossings

The crossing error is the distance from a true crossing to the nearest predicted crossing. A method that reports several crossings is therefore scored at its best one, and spurious crossings carry no penalty in that column. This favours operator inference on the wake, whose crossing error in Table 4 is small while it reports a crossing that does not exist. For this reason the crossing count is reported beside the crossing error in Table 4, and the conclusions of Section 4.2 rest on the count.

F.7 Cropped records approximating the prior-work setting

This experiment compares the methods on records approximating the later-time setting of the closest prior work, Conti et al. (2023). Their cylinder example omits the initial 15 s. Here the wake records are cut to remove the initial small-disturbance phase. The cut is the first point where the smoothed state-norm envelope reaches 95% of its maximum, snapped to the largest raw norm within w samples, where the smoothing width is $w = \max(5, \lfloor N/200 \rfloor)$ for a record of N samples. Stable records retain transient decay, while unstable records begin near the developed oscillation. All methods are refitted on these common records using the original model classes and record splits, with development selection repeated for AE-SINDy. This experiment uses the full 75-condition sweep, $25.4 \leq Re \leq 154.2$, comprising 21 training, 4 development and 50 test conditions. It extends beyond the $Re \leq 100$ range of the main tables.

The proposed model has mean seed-median test nRMSE 0.12, compared with 0.91 for AE-SINDy (Table 15). It remains bounded at every condition on every seed. AE-SINDy remains bounded at 53.3 percent of conditions on average, and its seed-median test errors range from 0.67 to 1.08. Thus the trajectory advantage persists when the initial growth phase is excluded. This is an aggregate comparison across conditions. AE-SINDy can track individual settled cycles, as illustrated by seed-zero restarts at $Re = 50.97$ and 64.94, with nRMSE 0.11 and 0.12, respectively.

The error statistic uses finite test errors and assigns no extra penalty to failed rollouts. AE-SINDy supplies 50, 28, 38, 32, 19 finite errors out of 50 test conditions across its five seeds, while the proposed model supplies all 50 on every seed. Boundedness is reported separately because a finite error can still violate that criterion. The SINDy row summarises six sparsity thresholds rather than random seeds. Only two yield finite test-error medians, from 30 and 17 test conditions, and their mean is reported. Its stability and boundedness columns include all six configurations. OpInf uses the configuration of Table 4, refitted on the cropped records.

The proposed model's sign accuracy is 85.3 percent, with the nearest crossing 2.7 percent from the reference. AE-SINDy has 33 to 51 percent sign accuracy and 5, 2, 1, 2, 2 crossings across seeds. Developed-cycle observations do not directly probe the growth of small disturbances at the equilibrium. When no growth window exists on a cropped unstable record, stage one fits its full retained record. The main experiment includes both the initial transient and subsequent nonlinear motion. Cropping preserves good trajectory accuracy but degrades the recovered stability structure.

Table 15: Wake, cropped records. nRMSE averages per-seed median finite errors on 50 test conditions. Missing-score coverage and SINDy's threshold aggregation are described above. Sign accuracy and boundedness (percent) use all 75 sweep conditions, versus 67 in the main tables. n is the median crossing count (true one). Error averages nearest-crossing errors in percent of the critical Reynolds number. Cropped-record errors are not directly comparable with Table 8.

method	nRMSE	sign	n	error	bounded
plain neural ODE	0.96	48.0	0	none	100.0
with the full recipe	1.31	48.5	0	none	100.0
coefficient-matrix form	0.25	91.7	2	1.5	100.0
free operator, trained jointly	0.19	88.8	2	1.5	100.0
free operator, frozen	0.12	74.7	5	2.2	100.0
proposed	**0.12**	**85.3**	**3**	**2.7**	**100.0**
proposed, without the gate	0.13	83.7	2	1.4	100.0
proposed, operator released	0.12	85.9	2	1.6	100.0
proposed, trained jointly from scratch	1.12	67.7	1	14.1	99.7
SINDy (threshold sweep)	2.31	58.0	2	0.9	15.6
AE-SINDy	0.91	39.7	2	30.0	53.3
OpInf	0.40	85.3	3	0.1	86.7
remaining at the equilibrium	1.00	*48.0*	0	—	100.0

F.8 Crossings beyond the labelled range

The labels do not fix the crossing locations. The main comparison evaluates stability transitions within the range of training parameters. Here the first stage is refitted on stable records only, so that no training record lies on the unstable side of the crossing it has to find. On the tube this interpolates across a gap in stable-only data. The wake and each wing edge require extrapolation beyond the corresponding stable training subset. The operator is evaluated as in Section 4.2. Figure 23 shows the resulting growth rate against the reference. On the tube, fitted on the stable records on both sides of the interval, the operator reproduces the arch of the growth rate across the interval it never saw and places both edges. On the wake, fitted on records below the crossing only, it continues past its last record and crosses zero at the reference crossing. On the wing each edge is found from the records on its own side. The same model class fitted by linear least squares on derivatives, drawn in red, does not cross at the reference on any of the three. Table 16 gives the crossing error against the distance from the nearest training record to the crossing, over training sets that leave progressively wider gaps. The error grows with the distance and stays within one percent of the natural scale until the nearest record is about eight percent of the tube's interval and thirteen percent of the wake's critical value away.

Table 16: Crossings located from stable records only. For each training set, the distance from the training record nearest the crossing to the crossing, and the error of the crossing the operator predicts, both in percent of the width of the unstable interval (tube, wing) or of the critical Reynolds number (wake). The wing rows use 15 and 5 training angles below the interval.

Rijke tube		Cylinder wake		Flexible wing	
distance	error	distance	error	distance	error
1.60	0.117	1.02	0.052	2.78	2.72
4.74	0.517	4.92	0.107	8.33	0.61
7.89	0.836	13.44	0.533		
		21.84	6.091		

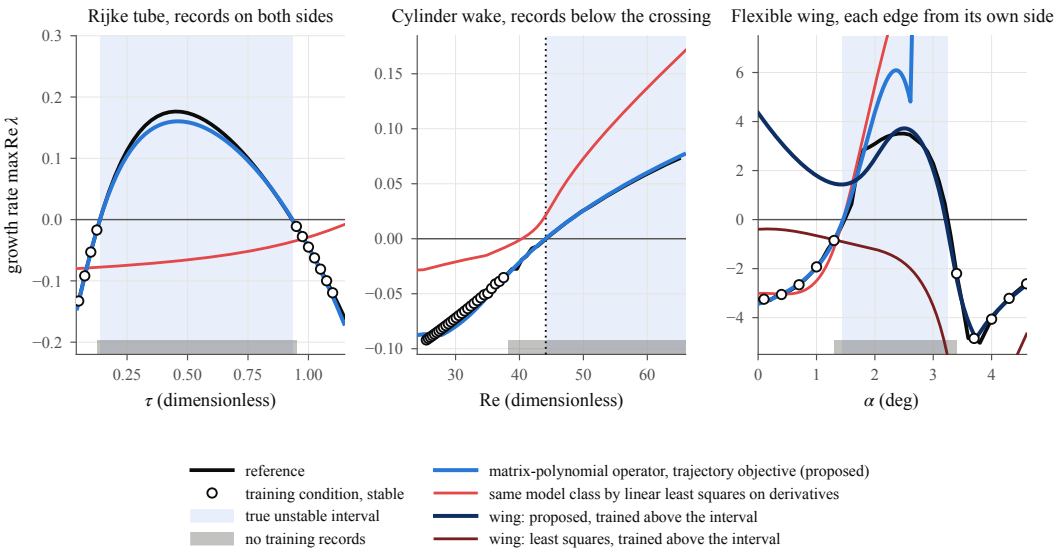


Figure 23: Growth rate of the first-stage operator fitted on stable records only (blue) against the reference (black), with the stable training conditions as circles, the true unstable interval shaded and the range without training records in grey. Left, the tube, fitted on both sides of the interval. Centre, the wake, fitted on records below the crossing (dotted line). Right, the wing, each edge fitted from the records on its own side. The red curves are the same model class fitted by linear least squares on the state derivatives.

G Supporting results for Section 4.3, the role of the frozen operator and the gate

This appendix supports Section 4.3. It gives the variation of the predicted crossing over training seeds for every architecture, the robustness of the first stage, and a statement of what the absence of seed variation does and does not establish.

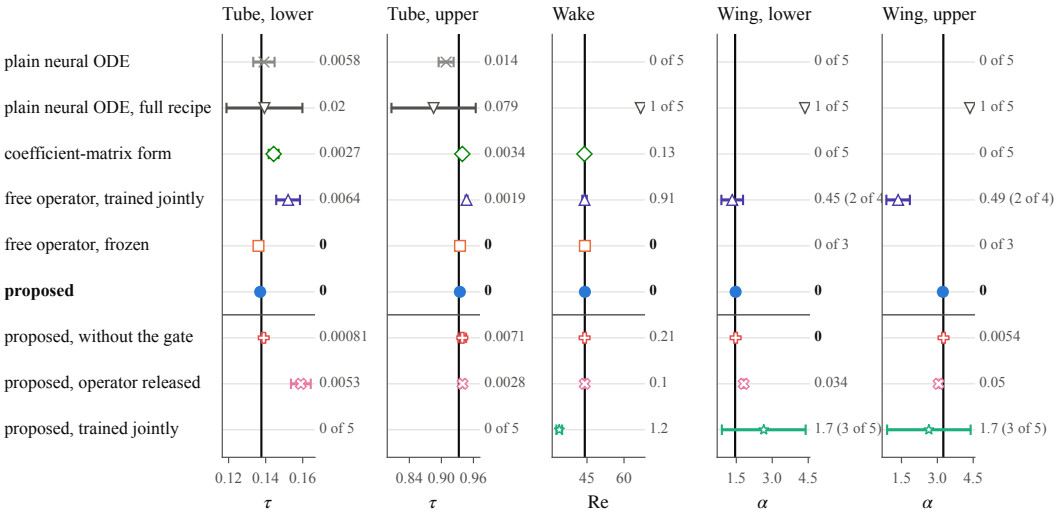


Figure 24: Predicted crossings of the nine neural architectures, one panel per true crossing. The marker is the mean over seeds returning the crossing, the bar is one standard deviation, and the vertical line is the true crossing. The number at the right of each lane is the standard deviation in the units of the axis over the seeds that return the crossing. Where fewer than all available seeds return it, their number is given, and no standard deviation is defined for fewer than two. For the two rows with a frozen operator the first stage is fitted once and shared across that row's seeds, so their zero is by construction. Total seed counts are reported in Table 8.

Table 17: Variation of the predicted first crossing across training seeds (total counts in Table 8). For each system the column *seeds* is the number of seeds that return a crossing and the column *s.d.* is the standard deviation of the predicted crossing over those seeds, in percent of the width of the unstable interval for the tube, in Reynolds number for the wake and in degrees for the wing. A dash means that fewer than two seeds return a crossing, so that no spread is defined. The first block is the sequence of architectures and the second block contains the three ablations of the proposed model.

	Rijke tube		Cylinder wake		Flexible wing	
architecture	seeds	s.d.	seeds	s.d.	seeds	s.d.
plain neural ODE	5	0.383	0	—	0	—
with the full recipe	5	1.310	1	—	1	—
coefficient-matrix form	5	0.334	5	0.128	0	—
free operator, trained jointly	5	0.806	5	0.912	2	—
free operator, frozen	5	0.000	5	0.000	0	—
proposed	**5**	**0.000**	**5**	**0.000**	**5**	**0.000**
proposed, without the gate	5	0.102	5	0.212	5	0.000
proposed, operator released	5	0.671	5	0.105	5	0.034
proposed, trained jointly from scratch	0	—	5	—	3	—

G.1 Robustness of the first stage

Since the seeds do not vary the first stage, its sensitivity is tested by refitting it, from a perturbed starting point and on varied training sets.

Perturbed starting point. On the tube, perturbing the starting point by 5 percent moves the predicted lower crossing by at most 0.036 percent of the interval. On the wake, 5 starts perturbed by 5 percent move the crossing by at most 0.021 in Reynolds number, which is 0.048 percent of the critical value, with a standard deviation of 0.021 percent. Every start converges, every start returns exactly one crossing, and the sign score over the reported range is 100.0 percent for every start. On the wing, 5 starts perturbed by 5 percent all converge and all return two crossings with a sign score of 100.0 percent. They move the lower crossing by at most 1.32 percent of the interval and the upper by at most 2.50 percent, with standard deviations of 0.46 and 0.79 percent, and every start places both crossings inside the grid interval that brackets the reference.

Varied training set. On the tube, removing one training condition at a time gives a maximum lower-crossing error against the physical reference of 0.52 percent over 4 refits. On the wake, refitting on each of 7 training sets moves the crossing over a range of 0.62 percent of the critical value. On the wing, 20 refits with one training angle removed move the lower crossing over a range of 4.63 percent of the interval and the upper over a range of 2.74 percent.

Gate radius. The gate radius of each system was swept with the operator frozen and only the second stage retrained, and the value with the lowest second-stage trajectory error on the development conditions was kept. The count and the location of the crossings are identical at every radius on every system, as Proposition 1 requires.

G.2 What the absence of seed variation establishes

For the proposed model, the first stage is fitted once and shared by the 5 seeds, so a seed varies the second stage only. The absence of seed variation in the reported crossings follows from the construction. Repeating the deterministic first-stage fit with the same data, initialisation and settings returns the same operator in the reported runs. Sensitivity to changes in those choices is evaluated separately above. The seeds here test residual training: trajectory errors vary while crossings remain fixed. Sharing an operator alone would not ensure this, because a residual with a nonzero equilibrium derivative can change the deployed Jacobian. This is what is observed when the gate is removed (Section 4.3). The robustness of the crossing itself is given by the refits of the first stage above.