SALTO: SOLVER-ANCHORED LIFTED TRANSFER OPERATORS FOR NONLINEAR EVENT INFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Learned dynamical models can accurately predict trajectories while producing unreliable sensitivities when differentiated or transposed, limiting their use in inverse problems and scientific decision-making. Classical adjoints provide exact local derivatives, but a single adjoint cannot represent the nonlinear response of a system to a family of finite-amplitude perturbations. We introduce SALTO (*Solver-Anchored Lifted Transfer Operators*), a Koopman-based framework for inferring finite-amplitude sources. In the lifted space, the nonlinear dynamics are approximated by a linear propagator, and its transpose carries each sensor backward in a single solve. The result is a reusable lifted adjoint that predicts the measurement produced by any candidate localized event, such as a source of given position and amplitude, with one encoder pass and an inner product, so that the search needs no repeated nonlinear simulation. To give the model the simulator's exact local behavior, SALTO imposes three structural constraints: it anchors the physical-state dynamics to the simulator's linearized propagator, centers the learned observables so that they vanish to second order at the reference state, and uses a block-triangular lifted propagator. These constraints guarantee that the model reduces to the simulator's tangent dynamics and classical adjoint as the amplitude vanishes, while the learned observables carry only the finite-amplitude correction. On three nonlinear PDE systems, SALTO predicts the measurements of finite-amplitude sources to within 2–8% where the classical adjoint is in error by 41–165%, recovers source position and intensity from sparse sensors nearly as accurately as inversion through the solver, and samples the joint posterior of two interacting sources, matching the solver's posterior on reaction–diffusion at a twentieth of its cost. One trained model serves new sensor layouts and measurement times with a single backward solve, and its exact local structure keeps both measurements and their gradients correct as the amplitude vanishes, where unconstrained Koopman models do not. Enforcing exact local structure thus turns a learned lift into a reliable instrument for nonlinear sensitivity and inference.

1 INTRODUCTION

Many problems in the physical sciences ask us to identify the cause of an observed signal. When a contaminant is detected at a downstream sensor, or a disturbance is recorded in a flow, we want to know which event produced it, where and when it occurred, and how strong it was. Such source localization problems arise in environmental monitoring, leak detection and flow diagnosis, and they all depend on how sensitive a measurement is to a perturbation of the state. The standard tool for computing this sensitivity is the adjoint of the tangent linear model. A single backward solve gives the derivative of a measurement with respect to the state at every earlier time, which is why adjoints underlie source localization, data assimilation, sensor placement and gradient-based design (Errico, 1997; Giles & Pierce, 2000; Luchini & Bottaro, 2014; Zaki, 2025). The adjoint, however, has a structural limitation. Because it linearizes the dynamics about a base trajectory, the measurement it predicts for a source is linear in the amplitude of that source. For a source of finite amplitude the measurement is a nonlinear functional of the source, and no single adjoint, whether about the base state or about any one nonlinear trajectory, can represent it. Inference therefore needs a sensitivity defined over a whole family of sources rather than a single vector.

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.

This paper obtains that functional from a single linear adjoint solve by means of a Koopman lift (Koopman, 1931; Mezić, 2005; Williams et al., 2015; Brunton et al., 2022). We augment the state with a set of learned nonlinear observables, so that the augmented, or lifted, state evolves linearly under an approximate Koopman operator. A measurement at the final time is then a linear functional of the lifted state, and carrying it one step backward in time amounts to applying the transpose of that operator. Repeating this back to the time of the source gives what we call the *lifted adjoint* (Fig. 1). Because the lifted adjoint does not depend on the state, it is computed once for each sensor and reused for every candidate source. The measurement predicted for a candidate source is then an inner product between the lifted adjoint and the lifted source. This inner product is a nonlinear functional of the source that costs a single encoder pass to evaluate, and it is exactly what finite-amplitude source inference needs.

The difficulty is that forward accuracy does not make a learned lift a reliable sensitivity. An unconstrained lift must learn the linear dynamics along with the nonlinear ones, even though the solver already provides them exactly, and any error in this linear part enters every sensitivity the lift predicts, including those of vanishingly weak sources. Such errors barely affect trajectory predictions at the training amplitudes, which is why a lift can forecast well and still transpose poorly. We therefore take the linear part from the solver and learn only the nonlinear correction, and we enforce this by construction with three constraints. The learned observables are centered so that they vanish to second order at the base state (C1), the lifted propagator uses the solver's linearized propagator for the physical state (C2), and it is block-triangular so that the learned observables evolve on their own (C3). The lifted model then reduces exactly to the solver's tangent dynamics and the classical adjoint as the amplitude vanishes, and the learned observables carry only the finite-amplitude correction.

Contributions. (i) **A solver-anchored lifted adjoint that is exact at small amplitude.** We formulate finite-amplitude sensitivity with a Koopman lift whose state-independent adjoint, obtained from one backward solve, turns inference over a prescribed source family into inner products (§3). We impose three constraints that make its transpose usable, namely solver anchoring, closed-form quadratic centering and a block-triangular propagator, and show that they make the tangent linear model and the classical adjoint exact in the infinitesimal limit, leave a nonlinear remainder of second order in the amplitude, and give the learned blocks a well-posed, closed-form fit (§4). (ii) **One backward solve for every candidate source.** One backward solve per sensor yields a lifted adjoint whose physical-state part is exactly the classical adjoint. The linear term of the functional is therefore the classical positional embedding (You & Wang, 2026), and the learned channels add the finite-amplitude correction. Predicting the measurement of any candidate source, computing its gradient or sampling a full posterior then costs encoder passes instead of nonlinear solves. (iii) **Accurate, fast inference on three nonlinear PDE systems.** We test the framework on the viscous Burgers equation about an evolving base flow, a two-dimensional reaction–diffusion system, and two-dimensional Kolmogorov flow about a laminar shear, each with a prescribed low-dimensional source family. Against the solver, the classical adjoint and its quadratic correction, unconstrained lifts, structural ablations, a neural surrogate and networks trained on measurements, the lift is exact at small amplitude in both the measurement and its gradient, infers sources nearly as accurately as the solver, reproduces the solver's joint posterior on reaction–diffusion at a twentieth of the cost, and serves new sensor layouts without retraining (§5).

2 Related work

Three lines of research are relevant to inferring finite-amplitude sources from sparse sensors. **Koopman and transfer operators.** Koopman theory represents nonlinear dynamics by a linear operator acting on observables of the state (Koopman, 1931; Mezić, 2005), and its data-driven approximations are reviewed by Brunton et al. (2022). Extended dynamic mode decomposition fits a finite-dimensional Koopman matrix on a prescribed dictionary (Williams et al., 2015), and Koopman autoencoders learn the dictionary together with the matrix (Lusch et al., 2018; Otto & Rowley, 2019). The transpose of the EDMD matrix acts on the coefficients of observables, and together with the dictionary's Gram matrix it gives the Galerkin approximation of the Perron–Frobenius operator (Klus et al., 2016). Physics-informed DMD constrains the fitted operator to known structure such as symmetries and conservation laws (Baddoo et al., 2023), and hybrid models add a learned correction to a known physical model (Yin et al., 2021; Um et al., 2020). Carleman linearization

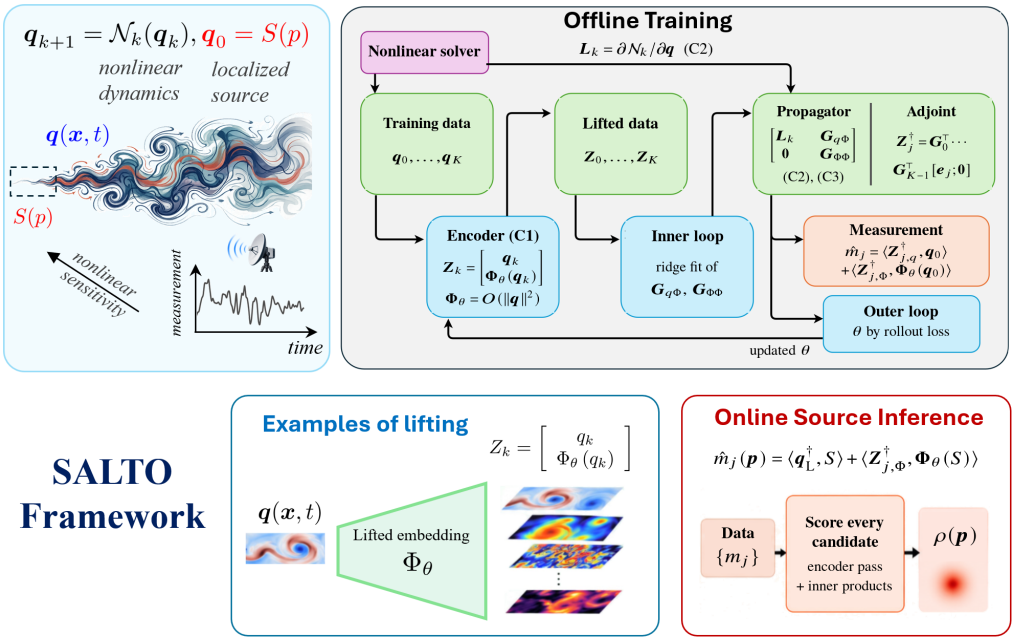

Figure 1: SALTO in one picture. Localized sources $S(\boldsymbol{p})$ seed a nonlinear flow observed by a few probes. Offline, the solver supplies $\boldsymbol{L}_k$ and the data supply only the finite-amplitude correction, so the lifted model is exact at small amplitude. Online, one backward solve per sensor gives the lifted adjoint $\boldsymbol{Z}_j^\dagger$, and every candidate source is scored by an encoder pass and inner products.

builds a linear embedding of a polynomial system from its equations, at the cost of an infinite hierarchy that must be truncated (Carleman, 1932; Kowalski & Steeb, 1991). **Adjoints and learned surrogates.** Adjoint sensitivity is classical in fluid mechanics and control, where one backward solve gives the gradient of a measured functional with respect to every input (Errico, 1997; Giles & Pierce, 2000; Luchini & Bottaro, 2014). In turbulent flows the adjoint itself degrades over long horizons, which has motivated regularized and latent-space assimilation (Ke et al., 2026; Wang & Zaki, 2026; Gutierrez & Wang, 2026a; Cleary et al., 2025). Differentiable surrogates provide gradients through learned models (Holl et al., 2020; Kochkov et al., 2021), although these gradients are rarely validated against the solver. Sobolev training supervises the derivatives of a network directly, at the cost of tangent solves during training (Czarnecki et al., 2017), and derivative-informed neural operators train on derivatives for inverse problems (O'Leary-Roseberry et al., 2024). **Source inference.** Adjoint-based methods localize sources from sparse sensors with a model that is linear in the source amplitude, and a quadratic correction extends this linear embedding to second order (Wang & You, 2026). Amortized simulation-based inference instead learns the posterior from simulations (Cranmer et al., 2020), and other reconstruction methods exploit the domain of dependence of each measurement (You et al., 2026; Wang & Zaki, 2025; Gutierrez & Wang, 2026b; Du & Wang, 2026; Domínguez-Vázquez et al., 2025).

3 Lifted Sensitivity

This section poses the inference problem and shows how a Koopman lift turns its nonlinear measurement functional into an inner product with a lifted adjoint. Let $\boldsymbol{q}(t) \in \mathbb{R}^N$ be the discretized perturbation of the state about a base solution. A nonlinear solver advances it over a snapshot interval Δs, and we write this one-step map as $\boldsymbol{q}_{k+1} = \mathcal{N}_k(\boldsymbol{q}_k)$, $k = 0, \ldots, K-1$, with $T = K\Delta s$. We call an initial perturbation a source and write it as $\boldsymbol{q}_0 = S(\boldsymbol{p})$, where $\boldsymbol{p}$ collects a few parameters such as position, intensity, width, orientation and injection time. Because the dynamics are written for the perturbation about a base solution, the system is homogeneous, so a zero source produces a zero response at all times. We assume throughout that the base solution is known and that the solver can reproduce it. The measurements are linear functionals of the final state, $m_j = \langle \boldsymbol{e}_j, \boldsymbol{q}_K \rangle$

for $j = 1, \ldots, N_m$, where $\boldsymbol{e}_j$ is a point or footprint kernel. The solver also provides the linearized one-step propagators $\boldsymbol{L}_k = \partial \mathcal{N}_k / \partial \boldsymbol{q}$ about the base trajectory, which together form the discrete tangent linear model. The *linear adjoint* $\boldsymbol{q}_{\mathrm{L}}^{\dagger}$ carries a kernel backward through $\boldsymbol{L}_k^{\top}$, and it is what a classical adjoint code returns (Wang et al., 2019a; 2022; 2019b). It is a single vector, so it cannot express the measurement $m_j(\boldsymbol{p})$, which is a nonlinear function of $\boldsymbol{p}$ across the source family. We form the lifted adjoint once, about the base state, because the perturbation itself is the unknown that the inverse problem seeks. Section 5 tests the functional and its gradient away from the base.

Let $\boldsymbol{\Phi}_\theta : \mathbb{R}^N \to \mathbb{R}^{cN}$ be a learned map to c channels, and let $\boldsymbol{\Psi}(\boldsymbol{q}) = [\boldsymbol{q}; \boldsymbol{\Phi}_\theta(\boldsymbol{q})] \in \mathbb{R}^d$, with $d = N(1+c)$, be the lifted state. The lifted linear model is $\boldsymbol{Z}_{k+1} = \boldsymbol{G}_k \boldsymbol{Z}_k$ with the block-triangular propagator

$$\boldsymbol{G}_k = \begin{bmatrix} \boldsymbol{L}_k & \boldsymbol{G}_{q\Phi} \\ \boldsymbol{0} & \boldsymbol{G}_{\Phi\Phi} \end{bmatrix}, \tag{1}$$

where $\boldsymbol{L}_k$ is the solver's propagator, and $\boldsymbol{G}_{q\Phi}$ and $\boldsymbol{G}_{\Phi\Phi}$ are learned, time-invariant blocks that couple the features to the state and advance the features. Fitting all of $\boldsymbol{G}_k$ by least squares on snapshot pairs would amount to extended dynamic mode decomposition (Williams et al., 2015). Here, by contrast, two of its blocks are imposed rather than fitted (§4). The lifted model starts from the encoded source, $\boldsymbol{Z}_0 = \boldsymbol{\Psi}(\boldsymbol{q}_0)$, and predicts the measurement as a linear functional of the lifted state, $\hat{m}_j = \langle [\boldsymbol{e}_j; \boldsymbol{0}], \boldsymbol{Z}_K \rangle$. Carrying this functional backward in time gives the *lifted adjoint*

$$\begin{aligned} \boldsymbol{Z}_j^{\dagger}(\tau_k) &= \boldsymbol{G}_{K-k}^{\top} \cdots \boldsymbol{G}_{K-1}^{\top} \begin{bmatrix} \boldsymbol{e}_j \\ \boldsymbol{0} \end{bmatrix}, \\ \hat{m}_j &= \Big\langle \begin{bmatrix} \boldsymbol{e}_j \\ \boldsymbol{0} \end{bmatrix}, \boldsymbol{Z}_K \Big\rangle = \big\langle \boldsymbol{Z}_j^{\dagger}(\tau_k), \boldsymbol{Z}_{K-k} \big\rangle \end{aligned} \tag{2}$$

for every backward time $\tau_k = k\Delta s$. The lifted adjoint holds the coefficients of the observable whose value on the lifted state at time $T - \tau$ equals the final measurement, just as the classical adjoint field is the sensor kernel carried backward by the transposed tangent model. The identity equation 2 holds exactly for the lifted rollout, so the only error of the lifted model relative to the solver is the error of that rollout, which Section 5 measures. Three properties follow. First, $\boldsymbol{Z}_j^{\dagger}$ does not depend on the state, so it is computed once per sensor and reused for every source. Second, setting $\tau = T$ and $\boldsymbol{Z}_0 = \boldsymbol{\Psi}(S(\boldsymbol{p}))$ in equation 2 gives the predicted measurement of a finite source directly,

$$\hat{m}_j(\boldsymbol{p}) = \big\langle \boldsymbol{Z}_{j,q}^{\dagger}(T), S(\boldsymbol{p}) \big\rangle + \big\langle \boldsymbol{Z}_{j,\Phi}^{\dagger}(T), \boldsymbol{\Phi}_\theta(S(\boldsymbol{p})) \big\rangle, \tag{3}$$

a nonlinear functional of the source, which we call the *nonlinear duality relation*. Its first term is the linear positional embedding of the classical adjoint, and its second term is the finite-amplitude correction. Evaluating it for a candidate costs one encoder pass and two inner products, with no time stepping. Third, by (C2) and (C3) the state block satisfies $\boldsymbol{Z}_{j,q}^{\dagger}(\tau_k) = \boldsymbol{L}_{K-k}^{\top} \cdots \boldsymbol{L}_{K-1}^{\top} \boldsymbol{e}_j = \boldsymbol{q}_{\mathrm{L}}^{\dagger}$, which is the linear adjoint itself. Every inference in this paper scores candidates by evaluating the duality relation equation 3 directly. Only the two-source point estimates of Appendix B.3 are afterwards refined, by gradient descent on the same functional.

Posterior density of a source. The duality relation equation 3 scores one candidate at a time, but the experiments also ask how precisely a source can be located, which a point estimate cannot tell us. We write $p = (x_s, y_s, I_s)$ for the parameters of one source and $\widehat{\boldsymbol{m}}(p_1, \ldots, p_{n_s})$ for the measurement that a model predicts when n_s sources are present together. Every model, including the solver, is evaluated on the combined field $\sum_a S(p_a)$, so that the prediction includes the interaction between the sources. With independent Gaussian noise of width σ and a flat prior,

$$\begin{aligned} \mathbb{P}(p_1, \ldots, p_{n_s}) \;&\propto\; \exp\Big(-\tfrac{1}{2\sigma^2} \big\| \boldsymbol{m}_{\mathrm{obs}} - \widehat{\boldsymbol{m}}(p_1, \ldots, p_{n_s}) \big\|^2 \Big), \\ \rho(x_s, y_s) \;&\propto\; \int \mathrm{d}I_s \int \prod_{b \geq 2} \mathrm{d}p_b \, \mathbb{P}\big((x_s, y_s, I_s), p_2, \ldots, p_{n_s}\big), \end{aligned} \tag{4}$$

where ρ is the density shown in the figures. Because the sources are exchangeable, ρ has one peak per resolved source, and we estimate it from Markov chain Monte Carlo samples (Appendix A).

The width σ accounts for more than the observation noise. A surrogate's own error in predicting the measurement adds to the noise in the residual, and when this error dominates, a likelihood based on

the noise alone is overconfident and can be confidently wrong. We therefore set $\sigma^2 = \sigma_n^2 + \sigma_{\mathrm{mod}}^2$, where σ_{mod} is the model's root-mean-square discrepancy from the solver on an independent set of combined pair fields. This makes models of different accuracy comparable, and for the solver $\sigma_{\mathrm{mod}} = 0$.

4 TRUSTWORTHY STRUCTURE

The lifted adjoint of Section 3 is only as reliable as the transpose of the learned propagator. A lift fitted only for forward accuracy leaves its transpose unconstrained away from the states visited in training, and it overwrites the linear dynamics that the solver already knows, so it need not reduce to the classical adjoint at small amplitude. On Burgers about a quiescent base, for example, an unconstrained lift of the same size matches our forward accuracy, yet its rollout error stays between 4% and 53% as the amplitude vanishes instead of vanishing with it (Appendix C). The three constraints below remove this failure by construction rather than through training.

(C1) Quadratic centering. The learned channels must be at least second order in the source perturbation, or they would corrupt the linear response that the solver already gives exactly. A nonzero $\boldsymbol{\Phi}_\theta(\mathbf{0})$ injects a constant through the coupling block, and a nonzero Jacobian at the base state adds a spurious linear term to the adjoint that does not vanish with the amplitude. We therefore require $\boldsymbol{\Phi}_\theta(\mathbf{0}) = \mathbf{0}$ and $\partial\boldsymbol{\Phi}_\theta/\partial\boldsymbol{q}|_{\mathbf{0}} = \mathbf{0}$, so that $\boldsymbol{\Phi}_\theta(\boldsymbol{q}) = O(\|\boldsymbol{q}\|^2)$. Given any smooth network $\widetilde{\boldsymbol{\Phi}}_\theta$, which is the encoder before centering, we set

$$\boldsymbol{\Phi}_\theta(\boldsymbol{q}) = \widetilde{\boldsymbol{\Phi}}_\theta(\boldsymbol{q}) - \widetilde{\boldsymbol{\Phi}}_\theta(\mathbf{0}) - \boldsymbol{J}_0\boldsymbol{q}, \qquad \boldsymbol{J}_0 = \partial\widetilde{\boldsymbol{\Phi}}_\theta/\partial\boldsymbol{q}|_{\mathbf{0}}, \tag{5}$$

which is exact and differentiable in θ. Both correction terms have closed forms for a tanh network (Appendix A). For a translation-equivariant convolutional encoder, $\widetilde{\boldsymbol{\Phi}}_\theta(\mathbf{0})$ is one constant per channel and $\boldsymbol{J}_0$ is a fixed convolution, so the centering needs no extra pass of the encoder.

(C2) Solver anchoring. The state row of equation 1 uses the exact linearized solver $\boldsymbol{L}_k$.

(C3) Block triangularity. The feature rows depend on the features alone. The lower-left block is $\mathbf{0}$.

Identities. Under (C1)–(C3), four identities hold for every k. (i) The k-step state block of the lifted propagator is exactly the solver's, $(\boldsymbol{G}_{k-1}\cdots\boldsymbol{G}_0)_{qq} = \boldsymbol{L}_{k-1}\cdots\boldsymbol{L}_0$, so at every horizon the linearization of the lifted model at the base state is the tangent linear model. (ii) The lifted rollout from $\boldsymbol{q}_0 = \epsilon\boldsymbol{v}$ equals the tangent linear rollout up to $O(\epsilon^2)$, and the learned channels contribute only the $O(\epsilon^2)$ remainder. (iii) The state block of the lifted adjoint is the linear adjoint, $\boldsymbol{Z}^\dagger_{j,q} = \boldsymbol{q}^\dagger_{\mathrm{L}}$, so the lifted functional reduces to the classical positional embedding as $\epsilon \to 0$, since $\partial\boldsymbol{\Phi}_\theta/\partial\boldsymbol{q} = O(\epsilon)$. (iv) The fit of the learned blocks is a constrained ridge least-squares problem with a closed-form solution. Each proof follows in one line from the block structure and from $\boldsymbol{\Phi}_\theta = O(\|\boldsymbol{q}\|^2)$, and we verify the identities numerically to machine precision relative to the supplied $\boldsymbol{L}_k$, which on Burgers and Kolmogorov flow are central-difference Jacobians of the solver. None of (i)–(iii) holds for an unconstrained lift, whose linearization at the base state is a fitted matrix that matches the tangent model only where the data constrain it.

Closed-form fit and bilevel training. Given snapshot pairs $(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ along the training trajectories, with $\boldsymbol{L}_k\boldsymbol{q}_k$ supplied by the solver, the learned blocks solve

$$\begin{aligned} \boldsymbol{G}_{q\Phi} &= \arg\min \textstyle\sum_k \|\boldsymbol{q}_{k+1} - \boldsymbol{L}_k\boldsymbol{q}_k - \boldsymbol{G}_{q\Phi}\boldsymbol{\Phi}_\theta(\boldsymbol{q}_k)\|^2 + \lambda\|\cdot\|^2, \\ \boldsymbol{G}_{\Phi\Phi} &= \arg\min \textstyle\sum_k \|\boldsymbol{\Phi}_\theta(\boldsymbol{q}_{k+1}) - \boldsymbol{G}_{\Phi\Phi}\boldsymbol{\Phi}_\theta(\boldsymbol{q}_k)\|^2 + \lambda\|\cdot\|^2, \end{aligned} \tag{6}$$

Each is a single Gram solve on the feature block, with rows scaled to unit RMS and a ridge parameter λ set relative to the trace. The data only have to explain the residual $\boldsymbol{q}_{k+1} - \boldsymbol{L}_k\boldsymbol{q}_k$, because the linear transport is not learned. We train the encoder parameters θ by gradient descent on the multi-step rollout error of the lifted model, differentiating through the closed-form solve of the inner loop, and we lengthen the rollout gradually during training. Every problem and every ablation uses the same trainer.

Baselines with the same code. Removing (C1) leaves raw, uncentered features (anchoring only). Removing (C2) and (C3) fits the full $\boldsymbol{G}$ by ridge EDMD (centering only). Removing all three gives

case	model	accuracy ε_m (%)	accuracy δ_x (cells)	accuracy δ_I (%)	cost seconds
Burgers *self-advection*	linear	72.9	6.00	15.2	0.003
	quadratic	96.1	3.00	8.9	4.1
	lifted (ours)	**1.8**	**1.00**	**3.6**	9.6
	solver	–	1.00	3.4	167
reaction–diffusion *saturation*	linear	164.5	0.63	40.2	0.005
	quadratic	174.5	0.91	37.1	1.9
	lifted (ours)	**2.4**	**0.55**	**14.2**	11.9
	solver	–	0.46	14.4	21.3
Kolmogorov flow *sheared advection*	linear	41.3	0.63	7.2	0.008
	quadratic	22.6	0.45	6.3	31.0
	lifted (ours)	**7.7**	**0.41**	**3.9**	11.8
	solver	–	0.41	2.8	214

Table 1: Accuracy and cost on the three systems, with the $c = 24$ lift on Burgers and the $c = 4$ lifts on the two-dimensional systems. Each block uses one common draw of test sources. Bold marks the best of the three models, the solver serves as the reference, and timings are wall-clock times.

the unconstrained Koopman autoencoder. A neural surrogate uses the same encoder with a linear head as a residual one-step map, $\boldsymbol{q}_{k+1} = \boldsymbol{L}_k \boldsymbol{q}_k + \boldsymbol{W}\widetilde{\boldsymbol{\Phi}}_\theta(\boldsymbol{q}_k) + \boldsymbol{b}$, trained on the same rollout loss. Its gradient is exact by automatic differentiation, but scoring each candidate requires a rollout.

5 Experiments

We evaluate SALTO on three systems, each showing a different way in which a nonlinearity defeats a linear adjoint. The first has self-advection that carries the response away from where the linear model places it, the second has a saturating reaction, and the third has a vortex that advects itself inside a sheared Navier–Stokes flow. Each system has a prescribed low-dimensional source family, and we compare against the solver and two classical models. The linear embedding predicts a measurement as $\langle \boldsymbol{q}_{\mathrm{L}}^{\dagger}, S(\boldsymbol{p})\rangle$, and the quadratic correction adds the second-order term of the expansion about the base, computed with second-order tangent and adjoint solves (Wang & You, 2026). Throughout, ε_m is the relative error of a model's predicted measurements at the top of the training range. The median position error δ_x and the median relative intensity error δ_I describe the one-shot inversion at 1% noise, with δ_x in units of the candidate spacing so that it compares across domains. The cost is the time for one estimate, which is a single pass over the candidate grid. Because candidate measurements do not depend on the data, a dictionary of solver measurements could also be cached once per sensor layout and horizon. The lift's cost advantage therefore lies in what no finite dictionary covers, such as the continuous joint posterior of several sources. Table 1 collects accuracy and cost, with bootstrap intervals in Appendix B, for one lift per system, with $c = 24$ channels on Burgers and $c = 4$ on the two-dimensional systems. The paper uses three kinds of lift, namely the $c = 4$ dense lifts of Table 1, Fig. 4 and the reuse study, the $c = 2$ dense lifts of the ablations, evaluated on a separate draw that extends down to zero intensity, and an eight-channel translation-equivariant lift trained on random fields for Fig. 3. Appendix A gives the equations, discretizations, source families, sensors, architectures and hardware.

5.1 Self-advection: Burgers about an evolving base flow

The first system is the viscous Burgers equation perturbed about an evolving base flow (Appendix A). The linearized solver misses both the steepening and the slowing of the front (Fig. 2). Its error at T grows from about a tenth of the response for the weakest source to four fifths for the strongest, whereas the lifted model keeps it below a few per cent. When we invert (x_s, I_s) in one shot from the five sensors by a grid search with 1% noise, the lifted functional matches brute-force inversion through the solver in intensity in every band and is at least as accurate in position (Appendix B). The quadratic correction is the stronger classical baseline for inference. It halves the position error of the linear embedding and reduces its intensity error from 15% to 9%, although its

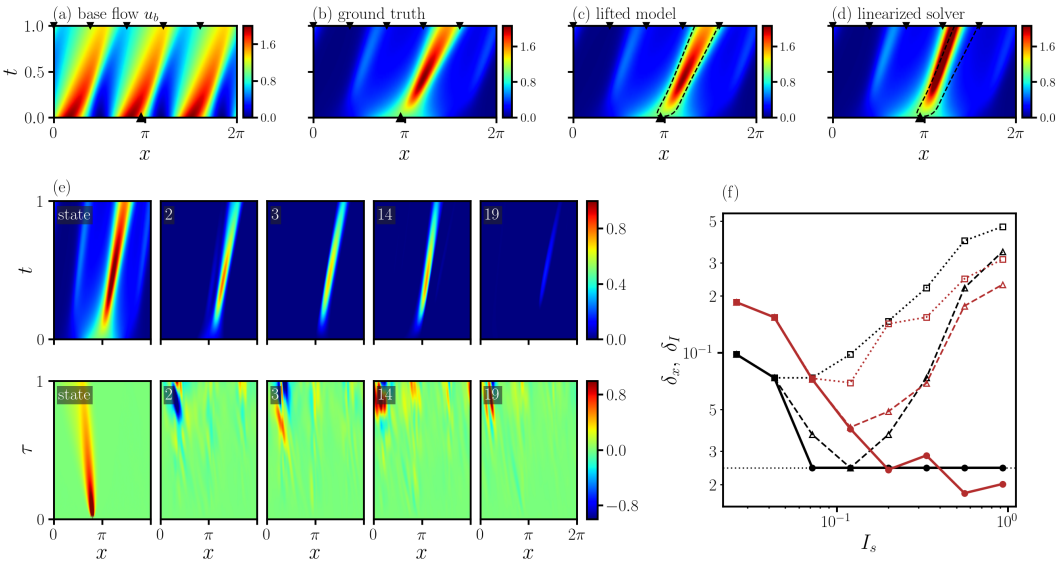

Figure 2: **(a–d)** Space–time evolution of the base flow and of one release, from the solver, the lifted model and the linearized solver. The dashed line is the half-maximum contour of the ground truth. **(e)** Five representative blocks of the lift, forward in time (top) and adjoint in backward time τ (bottom). **(f)** Median inversion error for 800 sources at 1% noise, with δ_x in black and δ_I in red, and linear dotted, quadratic dashed and lifted solid. The dotted horizontal line marks one grid cell.

measurement error at the top of the training range is larger than the linear embedding's. It still falls short of the lifted functional, because it expands about the base, and the front carries the release beyond the amplitude at which that expansion is accurate.

5.2 Saturation: two-dimensional reaction–diffusion

The second system saturates rather than advects, and we use it to study inference in full, from a single source to the joint posterior of two (Appendix A). **Forward, measurement, small amplitude.** The four-channel lifted functional predicts the nine measurements to within 2.4% at the top of the training range and 4% beyond it, where the error of the linear embedding exceeds the measurement itself. As the amplitude decreases, the error of the two-channel ablation lift decreases in proportion, reaching 0.008% at $I_s = 10^{-3}$, below the linear embedding's 0.06%. There the unconstrained lift is off by 23%, and the neural surrogate by more than the measurement itself (Appendix B).

Inference and cost. We infer each source by a grid search over $13{,}824$ candidates, scoring each with the duality relation equation 3, with 1% noise added to the measurements. Over 60 test sources the lifted functional recovers the position to 0.55 cells at the median, close to the 0.46 cells of brute-force inversion through the solver, and matches its intensity error of 14%. It places 48 of the 60 sources within one cell of the truth, against 49 for brute force and 39 for the linear embedding (Appendix B). Both classical baselines fall behind. The linear embedding cannot represent saturation and underestimates the intensity by 40% at the median, reading a saturated response as a weaker source, and the quadratic correction helps only within its radius of convergence, since the expansion diverges for the stronger sources. On this problem, scoring by inner products is the most accurate of the surrogates but not the cheapest. It takes twelve seconds per estimate, against two for the quadratic correction and twenty-one for the solver (Table 1).

Two sources at once. When two sources overlap, one to three source widths apart, their joint measurement differs from the sum of their separate measurements by about a sixth (Appendix B.3). The lift used here is more general than the one reported in the tables. Its learned blocks are translation-equivariant (Appendix A.3), it has eight channels, and it is trained on 1600 band-limited positive random fields rather than on the blob family that the inference searches over, so it never sees a source of the kind it is asked to find. The training fields must stay positive, because the reaction $-\kappa u - \gamma u^2$ drives u without bound once $u < -\kappa/\gamma$.

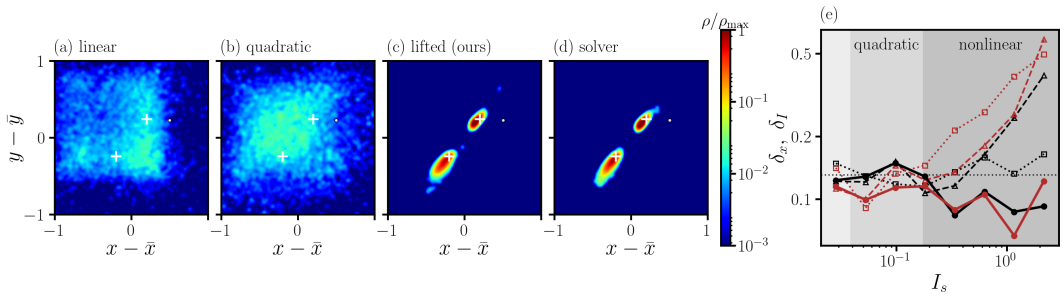

Figure 3: Two-dimensional reaction–diffusion, with the eight-channel equivariant lift. The lifted posterior reproduces the solver's, while the classical ones spread over the whole window. **(a–d)** Joint posterior density of source position (Eq. 4) for a pair 1.4 source widths apart, on a logarithmic scale. Each model is evaluated on the combined field, and its likelihood is widened by its own discrepancy. Crosses mark the true sources and the dot marks a probe. **(e)** Error of the one-shot inversion against the true intensity, with δ_x in black and δ_I in red, and linear dotted, quadratic dashed and lifted solid. Shading marks where each classical model still predicts the measurements to within three per cent, and the narrow unlabelled band at the left is where the linear embedding does.

A point estimate cannot tell whether a pair has been resolved, so we sample the joint posterior of both sources (Eq. 4), applying every model to the combined field (Fig. 3a–d). Over 17 overlapping pairs the lifted posterior reproduces the solver's. Its mean lies a median of 0.4 solver posterior standard deviations from the solver's mean. The central 95% interval of each of the six parameters contains the true value in 87% of the 102 cases, against 85% for the solver under the same sampler, a difference within the binomial standard error of about 3.5 points. Each lifted posterior costs 13 seconds, against 295 for the solver. The linear and quadratic posteriors are 13 to 27 times wider, with means 3.6 and 5.6 solver standard deviations away, so their full coverage only reflects that they are uninformative. Although this lift was trained on generic random fields and never on the blob family, it matches the solver on that family.

5.3 Sheared advection: Kolmogorov flow

The third system is the two-dimensional Navier–Stokes equations in vorticity form about a steady laminar shear flow, with a localized injection of vorticity as the source (Appendix A). The linearized propagator already carries the injected vorticity downstream and tilts it with the shear, so the lift only has to learn how the injected vortex advects itself (Fig. 4e). At the top of the training range, missing this self-advection costs the linear embedding 41% of the measurement and the quadratic correction 23%, against 7.7% for the lifted functional (Table 1).

Inference and cost. The one-shot inversion of the 60 test sources matches the solver in position (0.41 cells) and nearly matches it in intensity (3.9% against 2.8%), whereas the linear embedding reaches 0.63 cells and 7.2% (Table 1, Fig. 4f). Scoring the 13,824 candidates takes 12 seconds with the lifted functional, against 214 for the solver and 31 for the quadratic correction. For separated strong pairs the lift places both sources within one cell in 38% of cases, against 13% for the linear embedding and all pairs for brute force (Appendix B.3).

Ablations. The constraints make the lift exact at small amplitude, where every unconstrained variant fails, while keeping its finite-amplitude accuracy competitive (Table 2). At $I_s = 10^{-3}$ on reaction–diffusion the unconstrained lift is off by 23% against 0.008% in the current study, because its fitted linear part does not reduce to the solver's. On Kolmogorov flow ours is the most accurate learned model in every column. Its measurement gradient is also within 0.7% of the solver's for the weakest sources, which no other learned model achieves.

What the lifted dynamics add. The lift learns its features through the dynamics, and these features carry the finite-amplitude response. On Burgers about a quiescent base, a Carleman lift built on twelve POD modes is five times less accurate than the learned lift (Appendix C). Trained on the same simulations, a network of the same architecture fitted to the measurements alone is less accurate in every band on both systems, by up to a factor of twenty, and a network from the source parameters

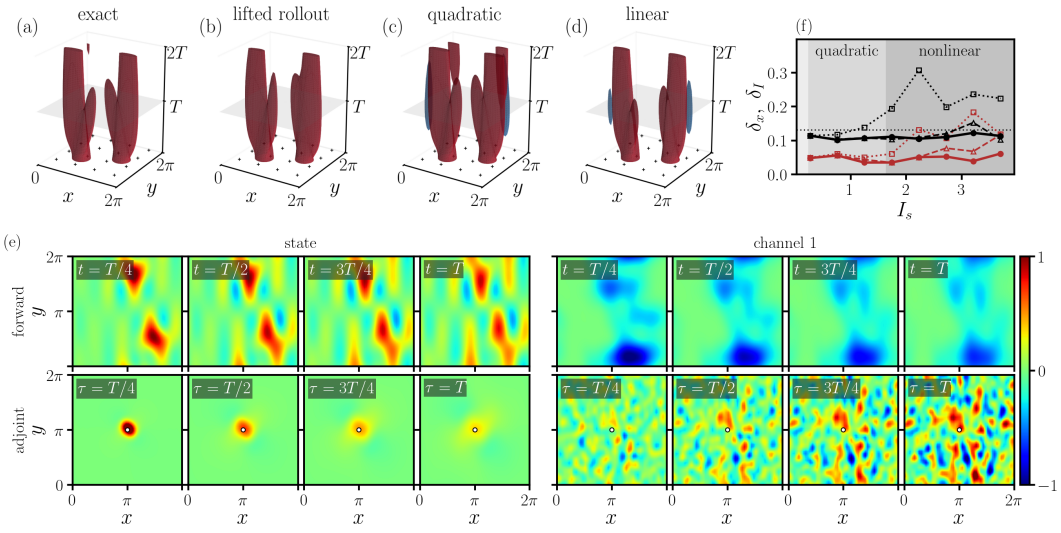

Figure 4: Kolmogorov flow. **(a–d)** Perturbation vorticity at $\pm 35\%$ of its initial maximum, time upward to $2T$, for the most strongly interacting pair, from the solver, the lifted rollout, the quadratic expansion and the linear superposition. **(e)** The state and one learned channel along that pair's solver trajectory (top), and the same blocks of the centre probe's lifted adjoint $\boldsymbol{Z}^\dagger$ (circle) in backward time (bottom), each on its own scale. The state block is exactly $\boldsymbol{q}_\mathrm{L}^\dagger$. The channel block is spread out because a measurement uses only its projection on the channels. **(f)** Inversion error over 600 further sources, as in Fig. 3e on linear axes; the narrow band is where the linear embedding holds.

Table 2: What the constraints buy (two-channel lifts). Only the anchored, centered lift is exact as the amplitude vanishes, in the measurement at $I_s = 10^{-3}$ (a ratio) and in the gradient for $I_s \in [0.01, 0.1)$, and on Kolmogorov flow it is the most accurate learned model throughout. The neural surrogate did not reach the lift's training loss on reaction–diffusion. Bold marks the best per column.

	reaction–diffusion				Kolmogorov flow			
	ε_m (%)	ε_m	gradient (%)		ε_m (%)	ε_m	gradient (%)	
model	top band	$I_s{=}10^{-3}$	weak	train	top band	$I_s{=}10^{-3}$	weak	train
lifted (ours)	6.0	8×10^{-5}	**0.7**	21	**13**	9×10^{-5}	**0.6**	**17**
unconstrained	**3.2**	0.23	4.9	**7.8**	14	0.67	16	25
centering only	4.8	3×10^{-2}	2.9	17	14	6×10^{-2}	10	20
anchoring only	9.2	1.9	11	23	22	8.2	17	27
neural surrogate	51	124	66	94	34	9	18	37
linear embedding	147	6×10^{-4}	5.1	205	40	2×10^{-4}	1.1	43

is off by 13–130% for weak sources and cannot represent two sources (Appendix B). Because the propagator is shared, one lift serves any sensor layout and time with a single backward solve, even beyond its training horizon, where at $1.5T$ it stays within 13% on reaction–diffusion and 17% on Kolmogorov flow against up to 196% and 63% for the linear embedding (Appendix B).

6 Discussion and Outlook

We have proposed SALTO, which performs linear sensitivity analysis in the lifted Koopman space to obtain a reusable nonlinear sensitivity in the original space, exact at small amplitude by construction, and which opens the way to efficient and accurate causal inference. Open questions include how many lifted channels and which architecture a problem requires, how the lift converges as it grows, and how far the idea extends to chaotic turbulent flows, which are closer to real applications.

AI USE STATEMENT

Generative AI tools were used to assist with editing and polishing the manuscript, including improving clarity, grammar, presentation, and refining the visualization and drawing part of the schematics. All AI-assisted text was reviewed and verified by the authors, who take full responsibility for the final content of the paper.

REFERENCES

Peter J. Baddoo, Benjamin Herrmann, Beverley J. McKeon, J. Nathan Kutz, and Steven L. Brunton. Physics-informed dynamic mode decomposition. *Proceedings of the Royal Society A*, 479(2271): 20220576, 2023.

Steven L Brunton, Marko Budišić, Eurika Kaiser, and J Nathan Kutz. Modern Koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, 2022.

Torsten Carleman. Application de la théorie des équations intégrales linéaires aux systèmes d'équations différentielles non linéaires. *Acta Mathematica*, 59:63–87, 1932.

Andrew Cleary, Qi Wang, and Tamer A Zaki. Latent-space variational data assimilation in two-dimensional turbulence. *arXiv preprint arXiv:2512.15470*, 2025.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

Wojciech M Czarnecki, Simon Osindero, Max Jaderberg, Grzegorz Swirszcz, and Razvan Pascanu. Sobolev training for neural networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Daniel Domínguez-Vázquez, Nicolas Escobar-Castaneda, Qi Wang, and Gustaaf B Jacobs. Inference of inertial particle dynamics from limited measurements. *Physics of Fluids*, 37(4), 2025.

Yifan Du and Qi Wang. Inverse reconstruction of free-stream vortices using surface pressure on a joukovsky airfoil. In *AIAA SCITECH 2026 Forum*, pp. 1112, 2026.

Ronald M. Errico. What is an adjoint model? *Bulletin of the American Meteorological Society*, 78 (11):2577–2591, 1997.

Michael B. Giles and Niles A. Pierce. An introduction to the adjoint approach to design. *Flow, Turbulence and Combustion*, 65(3):393–415, 2000.

Leopoldo Gutierrez and Qi Wang. Mixed lagrangian-eulerian data assimilation for turbulent channel flows. In *AIAA SciTech 2026 Forum*, pp. 0682, 2026a.

Leopoldo Gutierrez and Qi Wang. Reconstructing dangerous flow events using the domain of dependence from surface pressure data. In *AIAA SCITECH 2026 Forum*, pp. 0510, 2026b.

Philipp Holl, Vladlen Koltun, and Nils Thuerey. Learning to control PDEs with differentiable physics. In *International Conference on Learning Representations*, 2020.

Hongyi Ke, Zejian You, and Qi Wang. Preconditioned adjoint data assimilation for two-dimensional decaying and forced turbulence. *arXiv preprint arXiv:2602.14016*, 2026.

Stefan Klus, Péter Koltai, and Christof Schütte. On the numerical approximation of the Perron–Frobenius and Koopman operator. *Journal of Computational Dynamics*, 3(1):51–79, 2016.

Dmitrii Kochkov, Jamie A. Smith, Ayya Alieva, Qing Wang, Michael P. Brenner, and Stephan Hoyer. Machine learning-accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21):e2101784118, 2021.

Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

Krzysztof Kowalski and Willi-Hans Steeb. *Nonlinear Dynamical Systems and Carleman Linearization*. World Scientific, 1991.

Paolo Luchini and Alessandro Bottaro. Adjoint equations in stability analysis. *Annual Review of fluid mechanics*, 46:493–517, 2014.

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

Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, 2005.

Thomas O'Leary-Roseberry, Peng Chen, Umberto Villa, and Omar Ghattas. Derivative-informed neural operator: an efficient framework for high-dimensional parametric derivative learning. *Journal of Computational Physics*, 496:112555, 2024.

Samuel E Otto and Clarence W Rowley. Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal on Applied Dynamical Systems*, 18(1):558–593, 2019.

Kiwon Um, Robert Brand, Yun Raymond Fei, Philipp Holl, and Nils Thuerey. Solver-in-the-loop: Learning from differentiable physics to interact with iterative PDE-solvers. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6111–6122, 2020.

Mengze Wang, Qi Wang, and Tamer A Zaki. Discrete adjoint of fractional-step incompressible navier-stokes solver in curvilinear coordinates and application to data assimilation. *Journal of Computational Physics*, 396:427–450, 2019a.

Qi Wang and Zejian You. Localization of sources in weakly nonlinear fluid systems using linear and quadratic sensitivity analysis. *Physical Review Fluids*, 11(5):054901, 2026.

Qi Wang and Tamer A Zaki. Domain of dependence for wall-pressure measurements in high-speed boundary layers. *Journal of Fluid Mechanics*, 1009:A67, 2025.

Qi Wang and Tamer A Zaki. Mitigating adjoint chaos in wall turbulence. *Journal of Fluid Mechanics*, 1039:A48, 2026.

Qi Wang, Yosuke Hasegawa, and Tamer A Zaki. Spatial reconstruction of steady scalar sources from remote measurements in turbulent flow. *Journal of Fluid Mechanics*, 870:316–352, 2019b.

Qi Wang, Mengze Wang, and Tamer A. Zaki. What is observable from wall data in turbulent channel flow? *Journal of Fluid Mechanics*, 941:A48, 2022. doi: 10.1017/jfm.2022.295.

Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25:1307–1346, 2015.

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.

Zejian You and Qi Wang. Positional embeddings of scalar sources in turbulent channel flow from linear adjoint sensitivity. *in preparation*, 2026.

Zejian You, Qi Wang, and Xiaowei Zhu. Scalar source localization using multi-sensor domains of dependence in turbulent channel flow. In *AIAA SCITECH 2026 Forum*, pp. 2343, 2026.

Tamer A. Zaki. Turbulence from an observer perspective. *Annual Review of Fluid Mechanics*, 57:311–334, 2025. ISSN 1545-4479. doi: https://doi.org/10.1146/annurev-fluid-030424-114735. URL https://www.annualreviews.org/content/journals/10.1146/annurev-fluid-030424-114735.