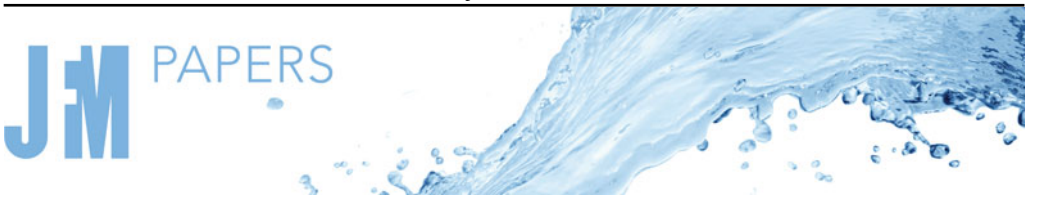
J. Fluid Mech. (2026), *vol.* 1027, A24, doi:10.1017/jfm.2025.11077



Using optimal transport aligned latent embeddings for separated flow analysis

Jonathan Tran[1], **Chi-An Yeh**[2] and **Kunihiko Taira**[1]

[1]Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095, USA

[2]Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC 27695, USA

Corresponding author: Jonathan Tran, jqtranus@g.ucla.edu

(Received 8 September 2025; revised 10 November 2025; accepted 13 December 2025)

Quantifying differences between flow fields is a key challenge in fluid mechanics, particularly when evaluating the effectiveness of flow control or other problem parameters. Traditional vector metrics, such as the Euclidean distance, provide straightforward pointwise comparisons but can fail to distinguish distributional changes in flow fields. To address this limitation, we employ optimal transport (OT) theory, which is a mathematical framework built on probability and measure theory. By aligning Euclidean distances between flow fields in a latent space learned by an autoencoder with the corresponding OT geodesics, we seek to learn low-dimensional representations of flow fields that are interpretable from the perspective of unbalanced OT. As a demonstration, we utilise this OT-based analysis on separated flows past a NACA 0012 airfoil with periodic heat flux actuation near the leading edge. The cases considered are at a chord-based Reynolds number of 23 000 and a free-stream Mach number of 0.3 for two angles of attack (AoA) of 6° and 9°. For each angle of attack, we identify a two-dimensional embedding that succinctly captures the different effective regimes of flow responses and control performance, characterised by the degree of suppression of the separation bubble and secondary effects from laminarisation and trailing-edge separation. The interpretation of the latent representation was found to be consistent across the two AoA, suggesting that the OT-based latent encoding was capable of extracting physical relationships that are common across the different suites of cases. This study demonstrates the potential utility of optimal transport in the analysis and interpretation of complex flow fields.

Key words: machine learning, big data, separated flows

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

1. Introduction

The analysis of flow fields and the design of effective control strategies require an exhaustive comparison of flows across typically large parameter spaces. This frequently involves substantial qualitative and quantitative analyses of a large number of high-fidelity simulations, experimental data or a combination of both. Given the high-dimensional, nonlinear and multi-scale nature of fluid flows, it is often advantageous to attempt to understand differences in key features of the flows through low-order models that extract relevant features (Lumley, Yaglom & Tatarski 1967; Schmid 2010; Taira *et al.* 2017).

Recently, the use of autoencoder (or encoder-decoder) methods for dimensionality reduction has grown in popularity due to their ability to handle complex nonlinear problems as opposed to traditional linear techniques (Murata, Fukami & Fukagata 2020; Vinuesa & Brunton 2022; Zeng, Linot & Graham 2022; Fukami & Taira 2023; Racca, Doan & Magri 2023; Smith *et al.* 2024; Tran *et al.* 2024; Mousavi & Eldredge 2025). These models seek to approximate a continuously differentiable and invertible transformation between a manifold in some high-dimensional space and a subset of some low-dimensional latent space. Murata *et al.* (2020) demonstrated the use of a convolutional autoencoder to compress the dynamics of flow past a cylinder to nonlinear modes, exhibiting superior compression performance and robustness to noise compared with proper orthogonal decomposition (POD). Zeng *et al.* (2022) utilised a low-order space learned by an autoencoder to perform reinforcement learning for the control of chaotic systems in a tractable manner. Additionally, Tran *et al.* (2024) exhibited the effectiveness of a combined POD and neural network autoencoder approach for performing a data-driven aerodynamic design optimisation of automobile geometries for drag reduction.

Although autoencoders are a powerful method to compress and extract features from fluid flows, they come with a few drawbacks. While methods such as POD may perform worse in compressing highly complex nonlinear systems, the low-order coordinates and modes have intuitive physical explanations and have reproducible results (Taira *et al.* 2017), which is not explicitly true for most machine-learning-based approaches. Fundamentally, autoencoders operate on the principle that high-dimensional datasets exist on low-dimensional manifolds (Gorban & Tyukin 2018). The autoencoder learns a latent space which gives curvilinear coordinates that describe a parametric surface fit to the data (Arvanitidis, Hansen & Hauberg 2018; Magri & Doan 2022). Often, these models are prone to overfitting, requiring extensive regularisation and hyperparameter tuning to improve their generalisation capabilities, especially when the underlying manifold is not densely sampled (Kingma 2013; Lee *et al.* 2022; Kvalheim & Sontag 2023). The problem of learning the manifold and latent space coordinates is inherently ill posed as any solution is only unique up to a diffeomorphism (Syrota, Moreno-Muñoz & Hauberg 2024). The non-uniqueness of the learned latent space coordinates is a major contributing reason to why obtaining a consistent interpretation of the latent coordinates is difficult outside of qualitative observations. This is especially an issue in the context of fluid mechanics, where it is desirable to use the learned latent space to perform downstream tasks that rely on geometric properties of the latent coordinates, such as dynamics modelling, flow field comparison, generative modelling or interpolation.

Previous studies have sought to address some of these issues of interpretability. Studies by Magri & Doan (2022) and Kelshaw & Magri (2024) interpret the latent space geometry of autoencoders using proper latent decomposition, which extends POD approaches to the non-Euclidean geometries learned by autoencoders. Fukami & Taira (2023) utilise an augmented autoencoder approach to analyse gust–airfoil interactions in which a model

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

is simultaneously trained to reconstruct flow-field information and an estimate of the lift coefficient from the compressed latent representation. This encourages the model to learn a low-order representation where there is a qualitative relationship between the latent space coordinates and estimates of the lift. Smith *et al.* (2024) employed persistent homology, a method from topological data analysis, to design an autoencoder that provides a simple representation of the topology of the state dynamics of large-amplitude gust encounters in the latent space representation.

In this study, we seek an alternative data-driven approach to imposing structure in the low-order coordinates for the analysis of fluid flows by incorporating information of pairwise distances or similarities into the learned latent coordinates. This is motivated by the fact that, when evaluating the effect of actuation on the dynamics of a flow, it is often necessary to quantify a degree of similarity or dissimilarity between two distinct flow fields. This typically involves the use of metrics or distances. One example is the widely adopted L^2 metric (or Euclidean, ℓ^2, for finite-dimensional vectors). However, while simple to compute, in some scenarios the Euclidean distance may not be the most informative measurement of dissimilarity as it ignores the underlying structural and geometric information, only facilitating pointwise comparisons of the flow field across corresponding spatial locations. This may be misleading in some scenarios, such as in the case where we seek to compare fields in which structures have been displaced.

Here, we utilise optimal transport (OT), which offers an intuitive framework for quantifying dissimilarities between flow fields. The class of OT distances, which includes the Wasserstein distance or earth mover's distance, has been studied extensively for various problems, including those in economics (Kantorovich 1960; Vaserstein 1969), generative modelling (Arjovsky, Chintala & Bottou 2017), image science (Rubner, Tomasi & Guibas 1998; Peyré & Cuturi 2019) and partial differential equation theory (Villani 2009; Mainini 2012*b*). Rather than only considering pointwise differences, OT distances effectively quantify a cost of transporting some resource or measure from one distribution to another (Villani 2009). This makes OT particularly well suited for scenarios in which the comparison of the spatial arrangement of flow features – such as vortices, coherent structures or transported quantities – is of interest. Additionally, OT distances have been demonstrated to be robust to noise and outliers in measurements (Peyré & Cuturi 2019).

To demonstrate the potential of OT applied to fluid flows, we consider the analysis of separated aerodynamic flows. Separated flow is a phenomenon accompanied by often undesirable behaviour, such as lift reduction and drag increase (Lissaman 1983; Mueller & DeLaurier 2003). Accordingly, engineers have long sought to understand and mitigate or reduce separation in aerodynamic flows through active and passive flow control techniques (Prandtl 1925; Lachmann 1961; Wu *et al.* 1998; Seifert & Pack 1999; Greenblatt & Wygnanski 2000; Joslin & Miller 2009; Kotapati *et al.* 2010; Yeh & Taira 2019).

Flow separation typically occurs due to the emergence of an adverse pressure gradient, in which the boundary layer is decelerated and detaches from the surface, forming a shear layer (Lachmann 1961; Mueller & DeLaurier 2003; Marxen, Lang & Rist 2013). Separation can also appear in the presence of sharp gradients in the surface geometry (Lachmann 1961; Häggmark *et al.* 2000) or when induced by nearby vortical structures (Harvey & Perry 1971). In the case of laminar separation, the separated shear layer is unstable, resulting in flow unsteadiness and/or a transition to turbulent flow (Dovgal, Kozlov & Michalke 1994). For high Reynolds number flows, the separated shear layer becomes more receptive to instabilities and becomes highly unsteady, eventually fully transitioning to turbulence around which the Reynolds number is of the order of 10^4-10^5. In such scenarios, the Kelvin–Helmholtz instabilities in the shear layer lead to the

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

roll-up of spanwise coherent vortical structures, which drive momentum mixing between the free stream and the reverse flow region (Tani 1964; Yarusevych, Sullivan & Kawall 2009; Kotapati *et al.* 2010; Yeh & Taira 2019). Unsteady mixing facilitates entrainment of high-momentum fluid from the free stream to enable the separated shear layer to reattach to the airfoil surface further downstream, provided the boundary layer is sufficiently energised (Lachmann 1961; Lissaman 1983; Mueller & DeLaurier 2003; Jaroslawski *et al.* 2023).

When the flow reattaches, the time-averaged flow exhibits a closed region of recirculating flow, forming a separation bubble. Depending on various factors, such as chord-based Reynolds number or the angle of attack, the characteristics of this separated flow can vary substantially. The size, shape and position of the separation bubble directly influence the performance of the airfoil, including stall characteristics and lift-to-drag ratio (Tani 1964; Lissaman 1983; Mueller & DeLaurier 2003; Klewicki, Schenkman & Spedding 2025). At Reynolds numbers of the order of 10^5, we may observe a long separation bubble that occupies over 20 %–30 % of the chord, resulting in noticeable changes in the pressure distribution over the suction side (Lissaman 1983), while at even higher Reynolds numbers, a short separation bubble may be observed. These shorter bubbles are generally characterised by a linear increase of lift with the angle of attack, with the bubble bursting at the onset of stall (Lissaman 1983; Mueller & DeLaurier 2003). However, a long separation bubble can also lead to stall without bursting if it spans a sufficient amount of the airfoil surface.

The unsteady separation bubble has been described as a self-excited flow structure maintained by a feedback loop that occurs due to interactions between the Kevin–Helmholtz instability in the shear layer and the vortex shedding in the wake (Zaman, McKinzie & Rumsey 1989; Kiya, Shimizu & Mochizuki 1997; Greenblatt & Wygnanski 2000; Yarusevych *et al.* 2009; Kotapati *et al.* 2010). Various studies have been performed to investigate how to leverage these instability mechanisms that contribute to the sustainment of the separation bubble. In particular, periodic forcing has been demonstrated to be an effective method of unsteady boundary layer control, as it can reach similar control authority as steady blowing or suction while requiring orders of magnitude smaller momentum input (Wygnanski 1997; Greenblatt & Wygnanski 2000). Amitay & Glezer (2002) performed an experimental parametric study using surface-mounted fluidic actuators to observe the flow response over a range of forcing frequencies, finding that unsteady periodic actuation can support complete flow reattachment. Yeh & Taira (2019) explored the use of resolvent analysis to guide the design of a periodic thermal flux actuator for the suppression of flow separation over a NACA 0012 airfoil. The resolvent formulation was used to gain insights into the effects of variation of the spatial and temporal actuation frequencies of a heat flux actuator on the separation dynamics.

In what follows, we analyse the effect of unsteady thermal actuation on separated flow over a NACA 0012 airfoil in a data-driven manner from the perspective of OT. We leverage an OT distance-based embedding learned by an autoencoder to analyse the controlled flows in a low-dimensional manner. We align pairwise Euclidean distances in the latent space with dissimilarities computed with OT, which attempts to associate the geometric structure of the latent space with the spatial distribution of fluid structures according to the OT distance. This restriction on the latent space coordinates allows relative distances in the latent space to encode information of similarity in terms of the spatial distribution of structures in the flow response. We find that the distribution of flow responses to periodic thermal actuation is reducible to an interpretable two-dimensional latent representation that encodes the behaviour of the separation bubble in the time-averaged flows. This demonstrates the potential utility of OT in the analysis of fluid flows.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

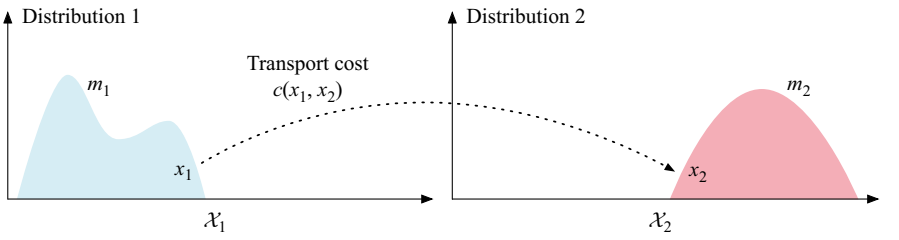


Figure 1. Pedagogical schematic of OT between a supply and demand distribution representative of two different flow fields. The OT distance is the minimum total cost to move the supply to the demand distribution.

2. Problem set-up and approach

2.1. *Optimal transport for comparison of fluid flows*

We utilise OT distances to quantify the change in the distribution of structures in flow fields. In this section, we give a brief description of OT and its application in computing dissimilarities between flow fields. While our focus is on a high-level explanation, additional details can be found in the appendices. For a more comprehensive review of the relevant subjects and formalisms from probability theory and OT, we refer the reader to Williams (1991), Villani (2009) and Tao (2011).

Optimal transport is naturally framed within the language of probability theory and measures. To define a measure, we first start with the concept of a measurable space, which comprises of a set $\mathcal{X}$ and a σ-algebra on $\mathcal{X}$ denoted as $\mathcal{A}$, which is a non-empty collection of subsets of $\mathcal{X}$ that is closed under complementation, countable unions and countable intersections. A measure is then a set function that takes measurable sets from $\mathcal{A}$ and maps them to an element of the extended real numbers. This measurement must satisfy certain properties, such as being zero for an empty set and also being countably additive. A physical interpretation could be the measurement of the mass, which prescribes a numerical value (mass) to a physical region of space.

Loosely speaking, the distribution of a measure describes how it is organised over the measurable space $(\mathcal{X}, \mathcal{A})$. For example, if $\mathcal{X}$ represents a physical domain, a distribution can describe the spatial arrangement of measures like the density of a material, the concentration of a substance or the probability of an event occurring in a region. In our discussion of fluid flows, we can consider distributions of physical measures of interest, where the measure corresponds to a quantity such as the fluid density or velocity magnitude.

Optimal transport distances provide a mathematical framework for comparing distributions of quantities while accounting for spatial structure. Originally developed to address problems of efficient resource allocation, the original OT problem seeks the most cost-effective way to redistribute a given resource from one configuration to another (Kantorovich 1960; Vaserstein 1969). The classical formulation considers a set of supply (or source) locations over which resources are distributed, and a set of demand (or target) locations, each requesting a specified amount of resources. The cost of transportation between locations depends on both the ground cost, in this case, the distance travelled, and the quantity of resources transported. The OT distance then corresponds to the minimum cost required to transport the resources from the supply distribution to the demand distribution. For these cases in which the total supply and demand are the same (balanced), the source and target distributions can be represented as probability measures, and OT can be used to define a metric between them. An illustration of OT for this pedagogical problem is shown in figure 1(*a*).

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

Following these examples, one can intuitively understand OT distances as a standard way to lift metrics (or distances) defined on some ground measurable space to a distance between distributions of measures (Villani 2009; Chizat *et al.* 2018; Peyré & Cuturi 2019). Optimal transport enables us to leverage the underlying geometric structure to compare distributions. In the context of fluid flows, the compared distributions may represent different flow fields, with each distribution describing the spatial variation of physical quantities such as density, velocity, or vorticity.

To define the OT distance between two measures, we start with spaces corresponding to the spatial domain of the two distributions, which are depicted in figure 1 by $\mathcal{X}_1$ and $\mathcal{X}_2$. We then choose a lower semi-continuous cost function that maps between the product space of $\mathcal{X}_1$ and $\mathcal{X}_2$ to the non-negative extended reals $C : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}_{\geqslant 0} \cup \{+\infty\}$. This cost function C describes the cost of transport from point $x_1 \in \mathcal{X}_1$ to another point $x_2 \in \mathcal{X}_2$. For this work, we choose C to be the Euclidean distance between points in $\mathcal{X}_1$ and $\mathcal{X}_2$. However, the relevant choice of the cost function is problem-dependent. We note that the inclusion of $+\infty$ in the range is to account for possible obstacles that may be present in the flow domain (e.g. a solid submerged body). Given $\mathcal{M}_+(\cdot)$, the set of all non-negative measures over some measurable space, we have that for positive measures $m_1 \in \mathcal{M}_+(\mathcal{X}_1)$ and $m_2 \in \mathcal{M}_+(\mathcal{X}_2)$, the OT cost is defined as

$$\mathcal{J}(\Gamma) := \int_{\mathcal{X}_1 \times \mathcal{X}_2} C(x_1, x_2)\, \mathrm{d}\Gamma(x_1, x_2). \tag{2.1}$$

With this cost, the balanced OT distance is computed with the following constrained minimisation:

$$d_{OT}(m_1, m_2) := \inf_{\Gamma \in \mathcal{M}_+(\mathcal{X}_1 \times \mathcal{X}_2)} \left\{ \mathcal{J}(\Gamma) : P_{\#}^{\mathcal{X}_1} \Gamma = m_1,\ P_{\#}^{\mathcal{X}_2} \Gamma = m_2 \right\}. \tag{2.2}$$

Here Γ, called the coupling (or transport plan), is a non-negative measure over the product space $\mathcal{X}_1 \times \mathcal{X}_2$ and describes the transport of resources between the distributions of m_1 and m_2. The joint distribution of Γ over $\mathcal{X}_1 \times \mathcal{X}_2$ quantifies how much resource is moved between the points x_1 and x_2. The measures $P_{\#}^{\mathcal{X}_1}\Gamma$ and $P_{\#}^{\mathcal{X}_2}\Gamma$ denote the first and second marginals of Γ, respectively. These are defined identically to the familiar notion of marginal probability, given by $P_{\#}^{\mathcal{X}_1}\Gamma = \int_{x_2 \in \mathcal{X}_2} \mathrm{d}\Gamma(\,\cdot\,, x_2)$ and $P_{\#}^{\mathcal{X}_2}\Gamma = \int_{x_1 \in \mathcal{X}_1} \mathrm{d}\Gamma(x_1, \,\cdot\,)$. The constraints enforce that the total measure must be the same for both distributions.

Naturally, the distributions of measures in different fluid flow fields need not integrate to the same amount. However, the traditional balanced OT formulation requires both distributions to have the same total measure. To account for this imbalance, we utilise a modification to the OT cost using Csiszár divergence functionals. These divergences are functionals that compare two distributions of measures in a pointwise manner (like the Kullback–Leibler divergence) and can be formally built with entropy functions in the context of probability and information theory. Further details on how these divergences are computed for the general measures in the current work can be found in Appendix A. The total cost of $\mathcal{J}(\Gamma)$ is modified to include an extra divergence functional term, replacing the equality constraint on the marginals in the optimisation. In other words, instead of strictly enforcing equality of the total measures, the divergences only softly penalise imbalance between the distributions, giving us an unbalanced OT distance.

Thus, to allow for a possible imbalance between the two flow fields, we additionally consider $\mathcal{D}_{\varphi_1}$ and $\mathcal{D}_{\varphi_2}$, which are Csiszár divergences over $\mathcal{X}_1$ and $\mathcal{X}_2$. Given $\mathcal{M}_+(\cdot)$, the set of all non-negative measures over some measurable space, we have that for positive measures $m_1 \in \mathcal{M}_+(\mathcal{X}_1)$ and $m_2 \in \mathcal{M}_+(\mathcal{X}_2)$, the unbalanced optimal transport (UOT) cost

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

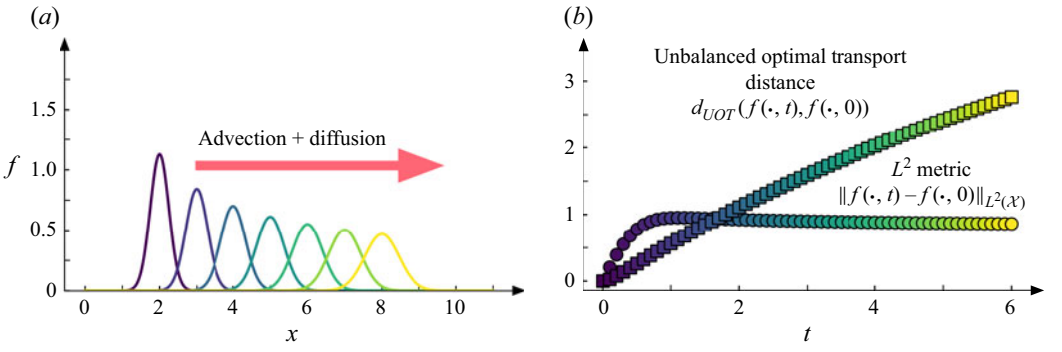


Figure 2. Comparison of unbalanced OT distance d_{UOT} with the L^2 metric. (*a*) Gaussian pulse undergoing advection and diffusion. (*b*) Comparison of distances between the evolving pulse and the initial condition.

is defined as

$$\mathcal{J}(\Gamma) := \int_{\mathcal{X}_1\times\mathcal{X}_2} C(x_1, x_2)\,\mathrm{d}\Gamma(x_1, x_2) + \rho\big[\mathcal{D}_{\varphi_1}(P^{\mathcal{X}_1}_{\#}\Gamma|m_1) + \mathcal{D}_{\varphi_2}(P^{\mathcal{X}_2}_{\#}\Gamma|m_2)\big]. \quad (2.3)$$

Here, the term $\rho > 0$ is described as a characteristic radius of transport which balances the contribution of the transport and divergence terms (Séjourné *et al.* 2019). The UOT distance is then defined as the infimum of this cost over $\Gamma \in \mathcal{M}_+(\mathcal{X}_1 \times \mathcal{X}_2)$

$$d_{UOT}(m_1, m_2) := \inf_{\Gamma\in\mathcal{M}_+(\mathcal{X}_1\times\mathcal{X}_2)} \mathcal{J}(\Gamma). \quad (2.4)$$

The first integral term describes the cost associated with a change in the spatial distribution. For example, this term would capture if a flow structure, such as a vortex or a shear layer, is at a different position in two flow fields. The second term that includes the divergences penalises how different the marginals of Γ are from the input measures m_1 and m_2, which allows for imbalance in the total amount of the two input measures.

To compute the divergence functionals, we use the Kullback–Leibler entropy and we set the characteristic transport radius $\rho = 1$ (Villani 2009; Chizat *et al.* 2018; Peyré & Cuturi 2019). To solve the optimisation problem, we use the Sinkhorn–Knopp algorithm (Chizat *et al.* 2018). The Sinkhorn–Knopp algorithm is an iterative matrix-scaling procedure that efficiently computes an entropy-regularised OT plan. It alternates between normalising the rows and columns of a transport matrix until convergence, making it well suited for large-scale problems, as it only requires scaling operations instead of solving the full linear system. This yields an approximation $S^{\varepsilon}(m_1, m_2) \approx d_{UOT}(m_1, m_2)$. We utilise the Python Optimal Transport library, which efficiently implements the Sinkhorn–Knopp matrix-scaling algorithm (Flamary *et al.* 2021).

For two flow fields V_1 and V_2 we define measures m_1 and m_2, corresponding to the distribution of some physical variable. To allow for the case of signed measures, such as vorticity, we compute the UOT distance for the positive and negative parts of the flow field separately and sum to obtain a total distance as follows:

$$d_{field}(V_1, V_2) = S^{\varepsilon}(m_1^+, m_2^+) + S^{\varepsilon}(m_1^-, m_2^-). \quad (2.5)$$

To consider multiple measures corresponding to the same field (e.g. distributions of multiple components of velocity), we can compute a total dissimilarity $\tilde{d}_{field}(\boldsymbol{V}_1, \boldsymbol{V}_2)$ by some permutation invariant aggregation of the dissimilarities for each variable $d_{field}(V_1^{(i)}, V_2^{(i)})$. A detailed explanation of how the flow-field dissimilarity $\tilde{d}_{field}(\boldsymbol{V}_1, \boldsymbol{V}_2)$ is computed can be found in Appendix B.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

To build further physical intuition, consider figure 2(*a*), which shows selected time steps of the solution to the advection–diffusion equation, with a Gaussian initial condition. Presented in figure 2(*b*) are the L^2 metric and the unbalanced OT distance d_{UOT} from the initial condition. Note that, for this particular example, one can also opt to use the balanced OT distance, given that the shown transport process is conservative. Here, by L^2 metric, we mean the metric that is induced by the L^2 norm on the Lebesgue space of square integrable functions on some domain $\mathcal{X}$ (as opposed to the ℓ^2 metric for vectors in $\mathbb{R}^n$). The L^2 distance from the initial condition is given by

$$\|f(\cdot, t) - f(\cdot, 0)\|_{L^2(\mathcal{X})} = \left\{ \int_{\mathcal{X}} \left[f(x, t) - f(x, 0) \right]^2 \, \mathrm{d}x \right\}^{1/2}, \tag{2.6}$$

which captures pointwise dissimilarities. As a result, it increases rapidly with even just a small translation but then saturates, classifying most subsequent time steps equally dissimilar once there is little overlap with the initial condition. Unlike the L^2 metric, the unbalanced OT distance correlates with the displacement of the Gaussian pulse even when the support does not overlap with the initial condition. As a result, the unbalanced OT distance provides a characterisation of dissimilarity that better aligns with our intuitive understanding of how the distribution evolves over time.

2.2. *Physical problem description*

As a demonstrative example, we consider the effects of active flow control on separated flows over a NACA 0012 airfoil at angles of attack (AoA) $\alpha \in \{6°, 9°\}$ with chord-based Reynolds number $Re_{L_c} \equiv u_\infty L_c/\nu_\infty = 23\,000$ and free-stream Mach number $M_\infty \equiv u_\infty/a_\infty = 0.3$. Here, u_∞ is the free-stream velocity, L_c is the chord length, a_∞ is the free-stream speed of sound and ν_∞ is the kinematic viscosity.

The flow fields were obtained via large eddy simulation using the *CharLES* finite-volume compressible flow solver (Khalighi *et al.* 2011; Brés *et al.* 2017) with the Vreman subgrid model (Vreman 2004). We used a C-grid mesh following the set-up in Yeh & Taira (2019) which has been examined for convergence in the flow field and aerodynamic forces with refinements in the near field. The computational domain is within $(x/L_c, y/L_c, z/L_c) \in [-19, 26] \times [-20, 20] \times [-0.1, 0.1]$ with the leading edge of the airfoil being placed at the origin. The solution was computed using a constant time step of $\Delta t u_\infty/L_c = 4.14 \times 10^{-5}$ corresponding to a maximum Courant–Friedrichs–Lewy number of 0.84. The statistics of the aerodynamic forces were computed using over 80 convective time units. Additional details of the computational set-up and grid are provided in Yeh *et al.* (2017*a*) and Yeh & Taira (2019).

The specific heat ratio was specified as $\gamma = 1.4$ with a Prandtl number $Pr \equiv \nu/\kappa = 0.7$, where κ is the thermal diffusivity. The dynamic viscosity was specified as a function of temperature in the form of a power law, given by $\mu(T) = \mu_\infty (T/T_\infty)^{0.76}$ for $T/T_\infty \in [0.5, 1.7]$ with μ_∞ and T_∞ denoting the free-stream dynamic viscosity and temperature, respectively (Garnier, Adams & Sagaut 2009). For all of the flows in this study, we observed that the maximum temperature fluctuation in the flow is within 42 % of T_∞, falling within the allowable range of the viscosity model (Yeh & Taira 2019).

For the far-field boundary, a Dirichlet boundary condition was specified as $(\rho, u_x, u_y, u_z, T) = (\rho_\infty, u_\infty, 0, 0, T_\infty)$. A sponge layer (Freund 1997) was placed along the outlet boundary with a target state being the running-averaged flow over 10 acoustic time units. Over the airfoil surface, a no-slip boundary condition was prescribed. The airfoil surface was also specified to be adiabatic except where the periodic heat-flux actuator was placed at the leading edge.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

The thermal actuation set-up employed in this study followed the methodology outlined in Yeh & Taira (2019), where a thermal actuator was placed across the span near the leading edge with prescribed frequency and spanwise profile. The actuator was implemented in the energy equation as a non-homogeneous, time-dependent Neumann boundary condition. The actuator model was expressed using a Hann window with compact spatial support given by

$$\phi^+(\omega^+, k_z^+) = \frac{1}{4}\hat{\phi}\sin(\omega^+ t)\left[1+\cos(k_z^+ z)\right]\left\{1+\cos\left[\frac{2\pi}{\sigma_a}(x-x_a)\right]\right\}, \tag{2.7}$$

where $(x - x_a)/\sigma_a \in [-0.5, 0.5]$. The actuator was centred at $x_a/L_c = 0.03$ on the suction surface with width $\sigma_a/L_c = 0.04$. The amplitude $\hat{\phi}$ was fixed according to the normalised total actuation power

$$E^+ = \frac{(1/4)\hat{\phi}\sigma_a}{(1/2)\rho_\infty v_\infty^3 (L_c \sin\alpha)} = 0.0902, \tag{2.8}$$

which is consistent with the set-up of previous studies (Corke, Enloe & Wilkinson 2010; Sinha *et al.* 2012; Yeh & Taira 2019). The thermal input from the actuator manifests in an oscillatory surface vorticity flux and baroclinic torque (Yeh *et al.* 2017*b*; Yeh & Taira 2019).

From this actuator, we have a two-dimensional parameter space of control inputs $(k_z^+, \omega^+) \in 10\pi\mathbb{Z} \times \mathbb{R}_{\geqslant 0}$. The control parameter pairs are characterised by a dimensionless wavenumber $k_z^+ L_c$ and chord-based actuation Strouhal number $St^+ := \omega^+ L_c/(2\pi u_\infty)$, where ω is the prescribed actuation frequency. For each angle of attack, we considered actuation wavenumbers of $k_z^+ L_c \in \{0, 10\pi, 20\pi, 40\pi\}$ with around 30 cases varying $St^+ \in [0, 18]$ for a total of around 120 cases per angle of attack.

An assortment of wake responses resulting from changes in the input parameters are shown in figure 3. Despite varying only two parameters in the shown cases, the resulting flow fields exhibit remarkably complex and diverse behaviour. These differences manifest in several key aspects of the wake dynamics. For example, the extent and nature of flow separation deviate greatly, with the size and position of the separation bubble differing significantly across different controlled cases. Additionally, the degree of laminarisation and turbulent mixing over the airfoil surface and in the downstream wake also changes noticeably between cases, with some cases exhibiting partial or global laminarisation. The wide array of flow field responses motivates the use of a data-driven analysis to extract relevant features in the flow response that relate to the control performance.

2.3. *Optimal transport dissimilarity based coordinate identification*

We seek to utilise a modified autoencoder to learn a low-dimensional representation of fluid flows while preserving dissimilarity between inputs in the learned latent representation. In other words, the goal is to learn not just a compressed representation of our flow fields, but to learn a representation with some geometric structure informed by OT that allows the latent coordinates to be interpreted as providing relative similarities between flow fields.

Consider some instance of discretised flow-field data $\boldsymbol{q} \in \mathbb{R}^{N\times N_s}$, where N refers to the number of variables and N_s refers to the size of the spatial discretisation. A typical autoencoder consists of an encoder–decoder pair $(\mathcal{E}, \mathcal{D})$ which can be constructed via neural networks (Hinton & Salakhutdinov 2006). The encoder $\mathcal{E}: \mathbb{R}^{N\times N_s} \to \mathbb{R}^n$ maps the input from a physical space to a latent space representation, with coordinates $\boldsymbol{\xi} \in \mathbb{R}^n$. The decoder $\mathcal{D}: \mathbb{R}^n \to \mathbb{R}^{N\times N_s}$ takes a latent space coordinate and produces a

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

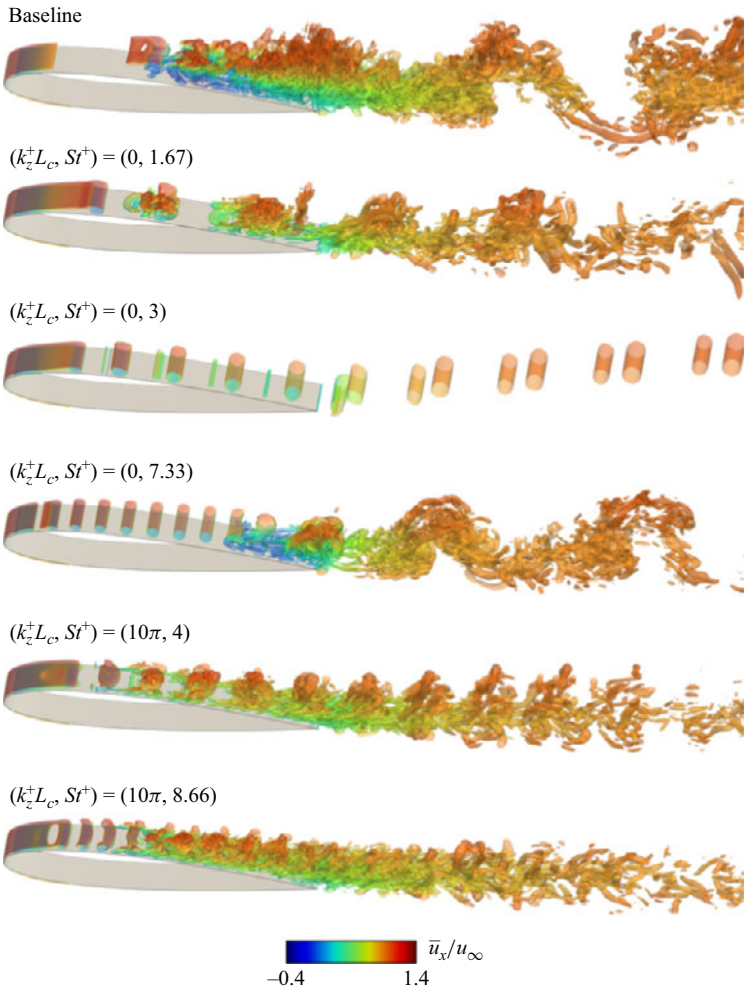


Figure 3. Examples of baseline flow and various responses of a separated wake past a NACA 0012 airfoil at $\alpha = 6°$ to a heat flux actuator input at the leading edge (Yeh & Taira 2019). The Q-criterion ($QL_c^2/u_\infty^2 = 50$) isosurface coloured by the normalised streamwise velocity is shown.

reconstruction of the input $\tilde{\boldsymbol{q}} = \mathcal{D} \circ \mathcal{E}(\boldsymbol{q}) \in \mathbb{R}^{N \times N_s}$. In this study, the encoder network consists of convolutional layers followed by dense layers. Each layer is followed by batch normalisation and a Rectified linear unit (ReLU) activation function. The decoder is constructed in a reverse fashion with dense layers followed by convolutional layers. Residual skip connections are placed around each convolutional block to stabilise training and reduce overfitting (Ioffe & Szegedy 2015; He *et al.* 2016). The parameters for the autoencoder architecture are given in table 1. In the present work, the number of filters are increased with depth (LeCun *et al.* 2002). All variables have been min–max normalised prior to training.

Here, we consider the time and spanwise-averaged streamwise velocity, vertical velocity and the turbulent kinetic energy, k, as the input and output, i.e. $\boldsymbol{q} = (\bar{u}_x, \bar{u}_y, k)$. We note that, since $\bar{\omega}_z = \partial \bar{u}_y/\partial x - \partial \bar{u}_x/\partial y$, we can also reconstruct $\bar{\omega}_z$ using the reconstructed velocities for assessment of the reconstruction performance. For the autoencoder input and output as well as the flow-field OT-dissimilarity computation, we consider

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

Encoder		Decoder	
Layer	Data size	Layer	Data size
Input ($\bar{u}_x$, $\bar{u}_y$, k)	(250, 100, 3)	Dense	(64)
Conv2D + BN (3,3,16)	(250, 100, 16)	Dense	(128)
MaxPooling2D	(125, 50, 16)	Dense	(256)
Conv2D + BN (3,3,32)	(125, 50, 32)	Dense	(1536)
MaxPooling2D	(63, 25, 32)	Reshape	(8, 3, 64)
Conv2D + BN (3,3,32)	(63, 25, 32)	Conv2D + BN (3,3,32)	(8, 3, 32)
MaxPooling2D	(32, 13, 32)	UpSampling2D	(16, 6, 32)
Conv2D + BN (3,3,64)	(32, 13, 64)	Conv2D + BN (3,3,32)	(16, 6, 32)
MaxPooling2D	(16, 6, 64)	Conv2D + BN (3,3,32)	(16, 6, 32)
Conv2D + BN (3,3,64)	(16, 6, 64)	UpSampling2D	(32, 12, 32)
MaxPooling2D	(8, 3, 64)	Conv2D + BN (3,3,16)	(32, 12, 16)
Conv2D (3,3,64)	(8, 3, 64)	Conv2D + BN (3,3,16)	(32, 12, 16)
Reshape	(1536)	UpSampling2D	(64, 24, 16)
Dense	(256)	Conv2D + BN (3,3,16)	(64, 24, 16)
Dense	(128)	Conv2D + BN (3,3,16)	(64, 24, 16)
Dense	(64)	UpSampling2D	(128, 48, 16)
Dense (Latent vector)	(2)	Conv2D + BN (3,3,8)	(128, 48, 8)
—	—	Conv2D + BN (3,3,8)	(128, 48, 8)
—	—	UpSampling2D	(250, 100, 8)
—	—	Output ($\bar{u}_x$, $\bar{u}_y$, k)	(250, 100, 3)

Table 1. Network architecture of the encoder and decoder models. The activation function used is ReLU.

a two-dimensional spatial region around the airfoil given by $\mathcal{X} = [-0.5L_c, 2L_c] \times [-0.5L_c, 0.5L_c]$, with $n_x = 250$ and $n_y = 100$.

The model's weights θ are optimised via the following objective function:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \ (\mathcal{L}_1 + \lambda \mathcal{L}_2), \tag{2.9}$$

$$\mathcal{L}_1 = \frac{1}{MNN_s} \sum_{i=1}^{M} \left\| \boldsymbol{q}_i - (\mathcal{D} \circ \mathcal{E})(\mathbf{q}_i) \right\|_F^2, \tag{2.10}$$

$$\mathcal{L}_2 = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \left(\|\mathcal{E}(\boldsymbol{q}_i) - \mathcal{E}(\boldsymbol{q}_j)\|_2 - \tilde{d}_{field}(\boldsymbol{q}_i, \boldsymbol{q}_j) \right)^2, \tag{2.11}$$

where M is the dataset size. The first term in the objective $\mathcal{L}_1$ is the mean square error of the flow-field reconstruction, which seeks to make the autoencoder approximate the identity function such that $\mathcal{D} \approx \mathcal{E}^{-1}$.

We introduce the OT-based embedding loss term $\mathcal{L}_2$, defined as the squared difference comparing pairwise Euclidean distances of latent points with the OT-based flow-field dissimilarities of their corresponding flow fields. With this term, we seek to impose a geometric structure on the embedded flow fields in which straight lines in the latent space should correspond to an approximate OT geodesic between inputs. If structures in the flow field are moved between two different inputs, resulting in a higher flow-field dissimilarity, the model would essentially learn to embed these fields further apart in the latent space. Conversely, if two inputs have a small flow-field dissimilarity, the model should learn to embed them close to one another in the latent space. This prevents the autoencoder from placing latent points arbitrarily relative to one another. With such an embedding, geometric properties of the latent space can be used to intuit physical trends or

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

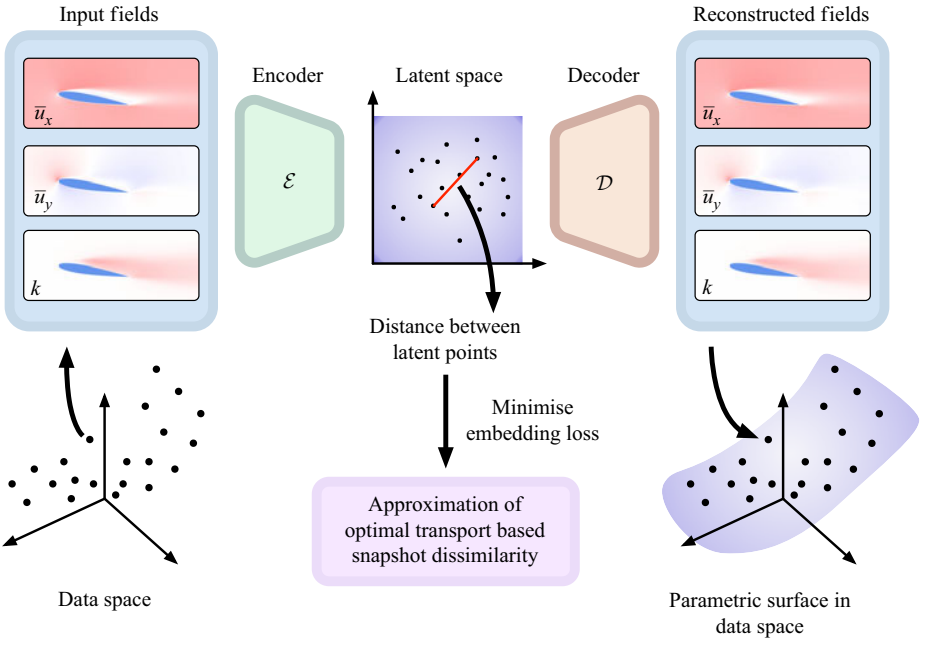


Figure 4. Illustration of the augmented autoencoder problem set-up. During training, the model learns to associate Euclidean distances in the latent space (red line) with flow-field dissimilarities computed with OT.

relationships between different controlled cases. We note that, if necessary for downstream tasks, we can also include a secondary decoder $\mathcal{F}: \mathbb{R}^n \to \mathbb{R}^m$ which maps from latent coordinates to estimates of physical parameters (Fukami & Taira 2023; Tran *et al.* 2024). However, for our particular demonstration, we found that using the relevant performance metrics $\boldsymbol{p} = (\bar{C}_D, \bar{C}_L) \in \mathbb{R}^2$ as an auxiliary output did not have a significant impact on the overall results.

We utilise the AdamW optimiser (Kingma 2014) with an initial learning rate of 1×10^{-4} and weight decay parameter of 10^{-2} for up to 4000 total epochs. An 80 : 20 split is performed between training and test data. The hyperparameter λ determines the relative influence of the embedding loss term. We choose this parameter by observing the tradeoff between the two loss terms as we vary λ. The tradeoff takes the form of an L-curve, or Pareto front, and we choose λ to be around the elbow of this curve (Hansen & O'Leary 1993). An illustration of the autoencoder based approach is depicted in figure 4.

3. Results

We examine the effectiveness of the present OT-based approach in identifying low-dimensional coordinates that succinctly capture the effects of the open-loop thermal control on the separated flow for controlled cases with $\alpha = 6°$ and 9°. In the following analyses, we consider separate autoencoders trained only on each angle of attack. For the choice of the $\mathcal{L}_2$ loss weight λ, an L-curve analysis is shown in figure 5(a) with $10^{-4} \leqslant \lambda \leqslant 10^4$. We choose λ to be the value around the elbow of the curve, which occurs approximately when $\lambda = 0.1$. In figure 5(b), we report the total loss for both training and test sets with respect to the choice of the latent space dimension with the chosen weight of $\lambda = 0.1$. The mean and standard deviation of the loss are computed using the last 500 epochs of training, where the loss has stabilised. For both AoA, we find that the loss

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

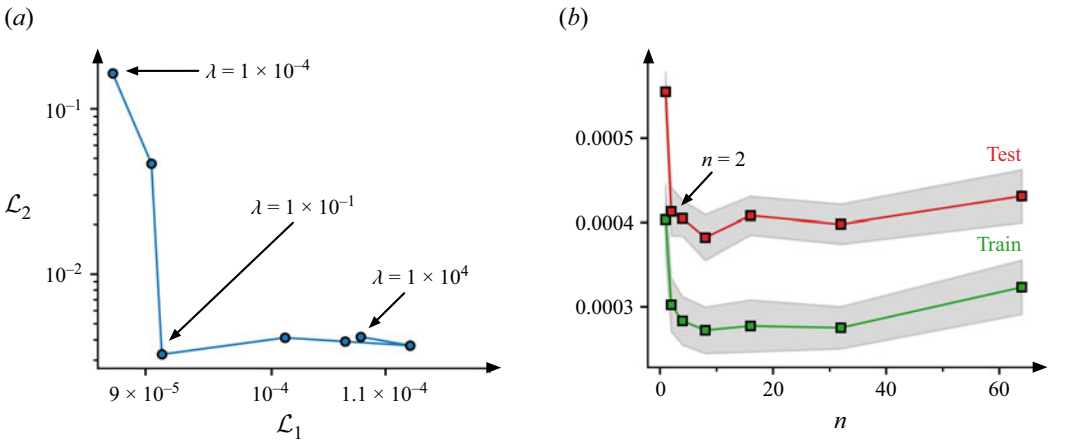


Figure 5. Example parameter study of the OT-based autoencoder for $\alpha = 9°$. (*a*) The *L*-curve showing the trade-off between $\mathcal{L}_1$ and $\mathcal{L}_2$ for the test set with $10^{-4} \leqslant \lambda \leqslant 10^{4}$. (*b*) Variation of the total loss $\mathcal{L}_1 + \lambda\mathcal{L}_2$ with respect to the latent space dimension for $\lambda = 0.1$. Standard deviation for the last 500 epochs is coloured in grey.

sharply drops as we go from $n = 1$ to $n = 2$, however, the loss for the test set seems to vary little after this point. We show a representative reconstruction from the OT autoencoder using a two-dimensional latent space for both AoA in figure 6. We report the reconstruction error for a given flow-field image f with $\varepsilon = 100 \times \| f_{rec.} - f_{original} \|_F / \| f_{original} \|_F$ for each reconstructed field. To additionally assess the autoencoder performance, we compare ω_z computed from the input and reconstructed velocities. We observe that the OT autoencoder is capable of qualitatively reconstructing the flow field from just a two-dimensional latent space, successfully recovering the general profile of the separation bubble and shear layer.

The performance of the two models (with $n = 2$) for both AoA is summarised in table 2. The average errors for the flow fields f are reported using the Frobenius norm error metric. Note that, if the discretised flow fields are flattened from matrices into vectors, this is equivalent to the vector ℓ^2 norm, which is usually reported. For the OT embedding term, error is measured using a normalised ℓ^2 error metric between the embedded latent distances and the OT-based dissimilarity of the corresponding flow fields

$$\varepsilon_d = 100 \times \sqrt{\frac{\sum_{i=1}^{M} \sum_{j=i+1}^{M} \left(\| \mathcal{E}(\boldsymbol{q}_i) - \mathcal{E}(\boldsymbol{q}_j) \|_2 - \tilde{d}_{field}(\boldsymbol{q}_i, \boldsymbol{q}_j) \right)^2}{\sum_{i=1}^{M} \sum_{j=i+1}^{M} \| \mathcal{E}(\boldsymbol{q}_i) - \mathcal{E}(\boldsymbol{q}_j) \|_2^2}}, \tag{3.1}$$

which in this context is referred to as the multi-dimensional scaling (MDS) stress (Kruskal & Wish 1978; Davison & Sireci 2000; Borg & Groenen 2005). This is a measurement of how distorted the embedded distances are from the original dissimilarities, and is invariant under translation and uniform stretching of the dissimilarities and latent coordinates. An MDS stress value that is smaller than 15 % is typically described as acceptable, although this threshold is heuristic and can be problem-dependent (Kruskal 1964; Kruskal & Wish 1978; Borg & Groenen 2005). Additional details of the training performance for the training and test sets are reported in Appendix C.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

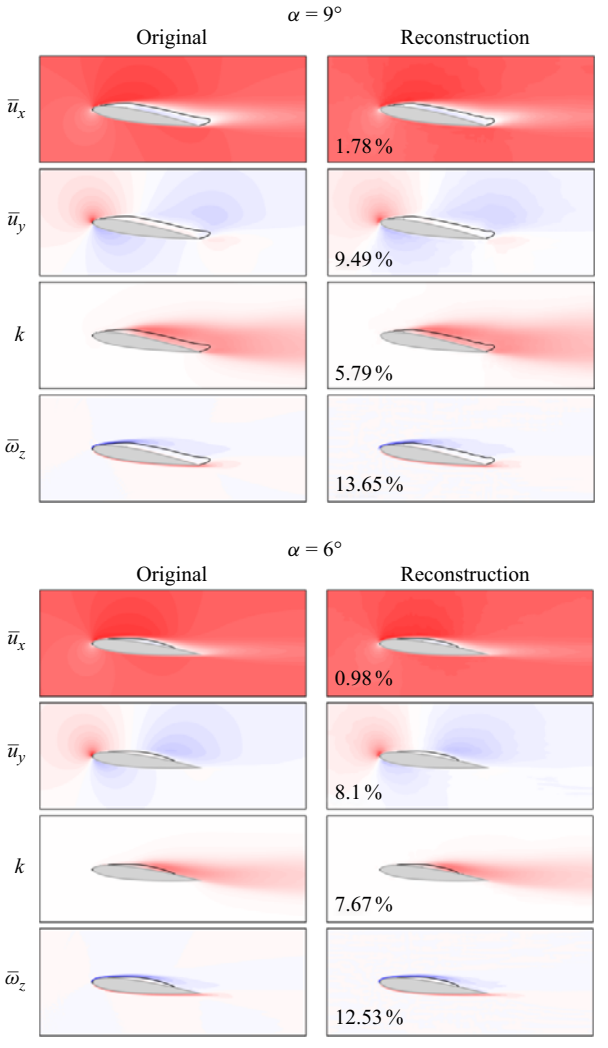


Figure 6. Reconstructions of baseline flow fields by OT autoencoder. The $\bar{u}_x = 0$ isocontour is shown in black for all fields. The reconstructed vorticity is obtained by central differencing of the reconstructed velocity fields. The per cent Frobenius norm reconstruction error is reported.

3.1. *The $\alpha = 9°$ latent space cases*

We begin by examining the $\alpha = 9°$ cases, which exhibit comparatively simpler flow responses than the $\alpha = 6°$ cases, primarily due to the absence of global laminarisation in the flow responses (Yeh & Taira 2019). In particular, we first observe the effect of the OT-based embedding term on the learned latent space representation. We consider latent spaces learned by two representative autoencoder models as shown in figure 7: (*a*) a standard autoencoder trained solely with reconstruction loss ($\mathcal{L}_1$) and (*b*) the OT autoencoder which includes the OT embedding term ($\mathcal{L}_1$ and $\mathcal{L}_2$). For ease of discussion, we rotate all latent spaces to align with their principal component axes so that we can describe how the flow fields vary along axes where the OT distances are estimated to vary the most. This transformation does not affect the loss or the interpretation of relative dissimilarity as the reconstruction loss $\mathcal{L}_1$ remains unchanged since the rotation can be reversed without loss, and the $\mathcal{L}_2$ term is inherently invariant to uniform rotation of the

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

AoA	Quantity	Reg. AE Err.		OT AE Err.	
		Train	Test	Train	Test
6°	$\bar{u}_x$	1.11	1.13	1.25	1.37
	$\bar{u}_y$	9.99	10.08	8.75	9.05
	k	7.86	9.35	8.45	10.36
	$\bar{\omega}_z$	13.26	14.36	15.09	15.29
	OT Emb.	—	—	12.77	13.52
9°	$\bar{u}_x$	1.38	1.14	1.42	1.47
	$\bar{u}_y$	8.59	8.50	8.50	8.48
	k	6.62	7.18	8.28	8.71
	$\bar{\omega}_z$	16.60	16.48	14.89	14.76
	OT Emb.	—	—	13.82	15.00

Table 2. Comparison of average per cent error for field variables using different autoencoder architectures (with two latent variables) at AoA $\alpha = 6°$ and 9°. Field variable errors are reported as per cent Frobenius norm error. The embedding loss is reported using the MDS stress.

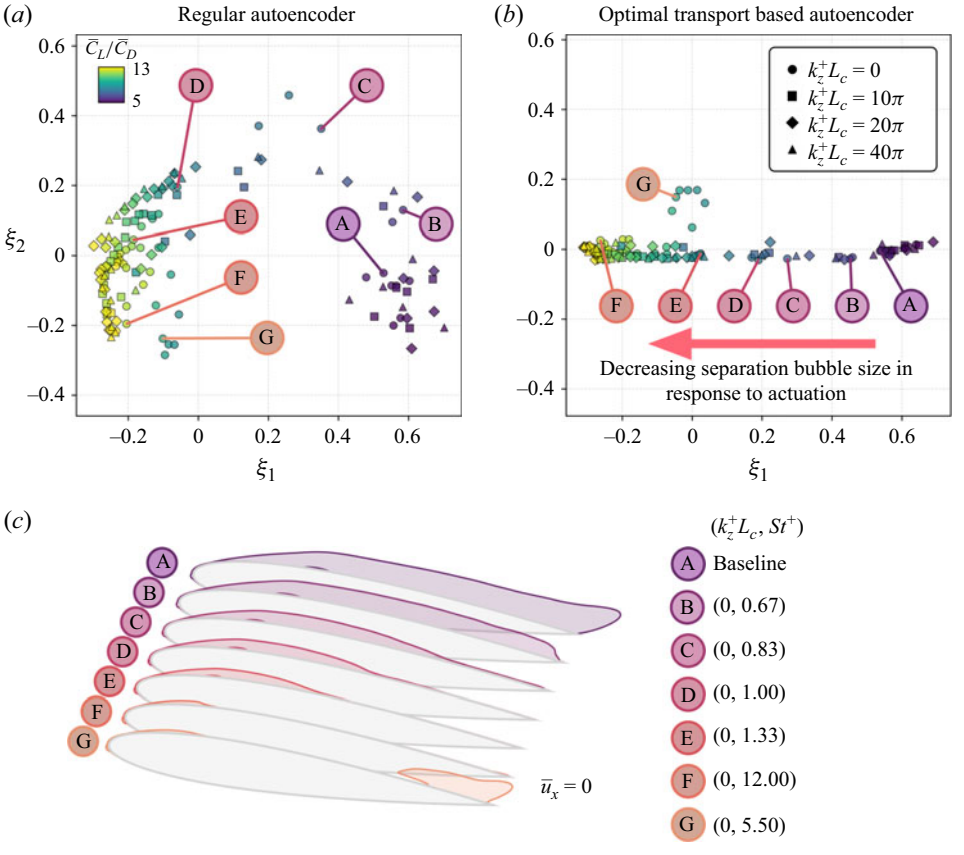


Figure 7. Comparison of learned latent spaces with $\alpha = 9°$ coloured by aerodynamic performance $\bar{C}_L/\bar{C}_D$ for (*a*) standard autoencoder ($\mathcal{L}_1$ loss only), (*b*) OT autoencoder ($\mathcal{L}_1$ and $\mathcal{L}_2$ losses). (*c*) Isocontours of time-averaged streamwise velocity $\bar{u}_x = 0$ shown for different representative cases labelled in (*a*) and (*b*).

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

latent space. We note that both models achieve comparable reconstruction performance, indicating that the additional loss term does not severely compromise the ability to accurately recover the flow fields.

Also depicted in figure 7 are the $\bar{u}_x = 0$ isocontours for representative input cases, which exhibit the variation in the separation bubble behaviour as we traverse the latent space. Case A in figure 7 shows the baseline flow, with a long separation bubble that spans the entire suction side. As we move from case A to case F, we observe that the separation bubble shrinks towards the leading edge. Conversely, case G presents a qualitatively different flow regime. Separation occurs near the trailing edge as spanwise rollers merge and break down near the trailing edge, forming a recirculation zone, and there is a pronounced increase in turbulent kinetic energy near the trailing edge. This is more clearly shown by case 1-2 in figure 8.

For both the regular autoencoder and the OT-based latent space, we see a sequential progression from case A to case F as the separation bubble shrinks, which correlates with the control performance in terms of the average lift-to-drag ratio. However, for the standard autoencoder in figure 7(*a*), while we can make out a qualitative trend of the lift-to-drag ratio in the latent space, the latent representations are unconstrained in geometry, and there is no inherent interpretation to the distance between latent points. We again emphasise that the relative positioning of the latent points for the standard autoencoder latent space in figure 7(*a*) is arbitrary up to diffeomorphism, which inhibits our ability to judge how similar the flow fields are based solely on their latent positions. In fact, the distinction between the outlier cases in which separation occurs near the trailing edge (case G) is not clearly made in the regular autoencoder latent space. Additionally, in the latent space learned by the regular autoencoder, the best-performing and worst-performing cases begin to approach each other, despite being the most qualitatively 'dissimilar' cases.

In the OT-based latent space shown in figure 7(*b*), we find that the first latent coordinate ξ_1 clearly corresponds with the variation in size of the separation bubble. As ξ_1 decreases, the separation bubble size reduces and eventually vanishes, as seen in cases A–F. The second latent coordinate ξ_2 captures changes in the flow field associated with the partial laminarisation of the flow and the trailing-edge separation, which is specific to the $k_z^+ = 0$ configurations, depicted by case G. Most cases are distributed along the ξ_1 axis (i.e. first principal component), suggesting that the separation bubble size is the dominant mode of variation in the flow response according to the OT-based latent geometry. We once again emphasise that the ability to derive an interpretation for the principal components in the latent space arises because Euclidean distances in the latent space serve as surrogates for the OT-based snapshot dissimilarity.

From figure 7(*b*), we additionally observed that ξ_1 is also strongly correlated with the control performance in terms of the average lift-to-drag ratio. Lower values of ξ_1 are associated with reduced flow separation (hence a smaller separation bubble) and an increased lift-to-drag ratio. To understand the variation of the control performance in terms of the control input parameters, in figure 8 we show the OT-based latent space coloured by the actuation frequency with marker shapes corresponding to the spanwise wave number of actuation. We find that the ξ_1 coordinate or separation bubble size primarily depends on St^+, having a weaker dependence on $k_z^+ > 0$.

We also depict several representative cases from both two and three-dimensional actuation at various frequencies in figure 8, with their corresponding latent embeddings labelled. Starting from the baseline (case 0), increasing the forcing frequency within the range $0 \lesssim St^+ \lesssim 4.5$ (e.g. cases 1-1, 2-1 and 3-1) leads to a rapid decrease of ξ_1, corresponding to an improved lift-to-drag ratio as was seen in figure 7(*b*). In this regime, flow separation is effectively suppressed, and the turbulent kinetic energy in the

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

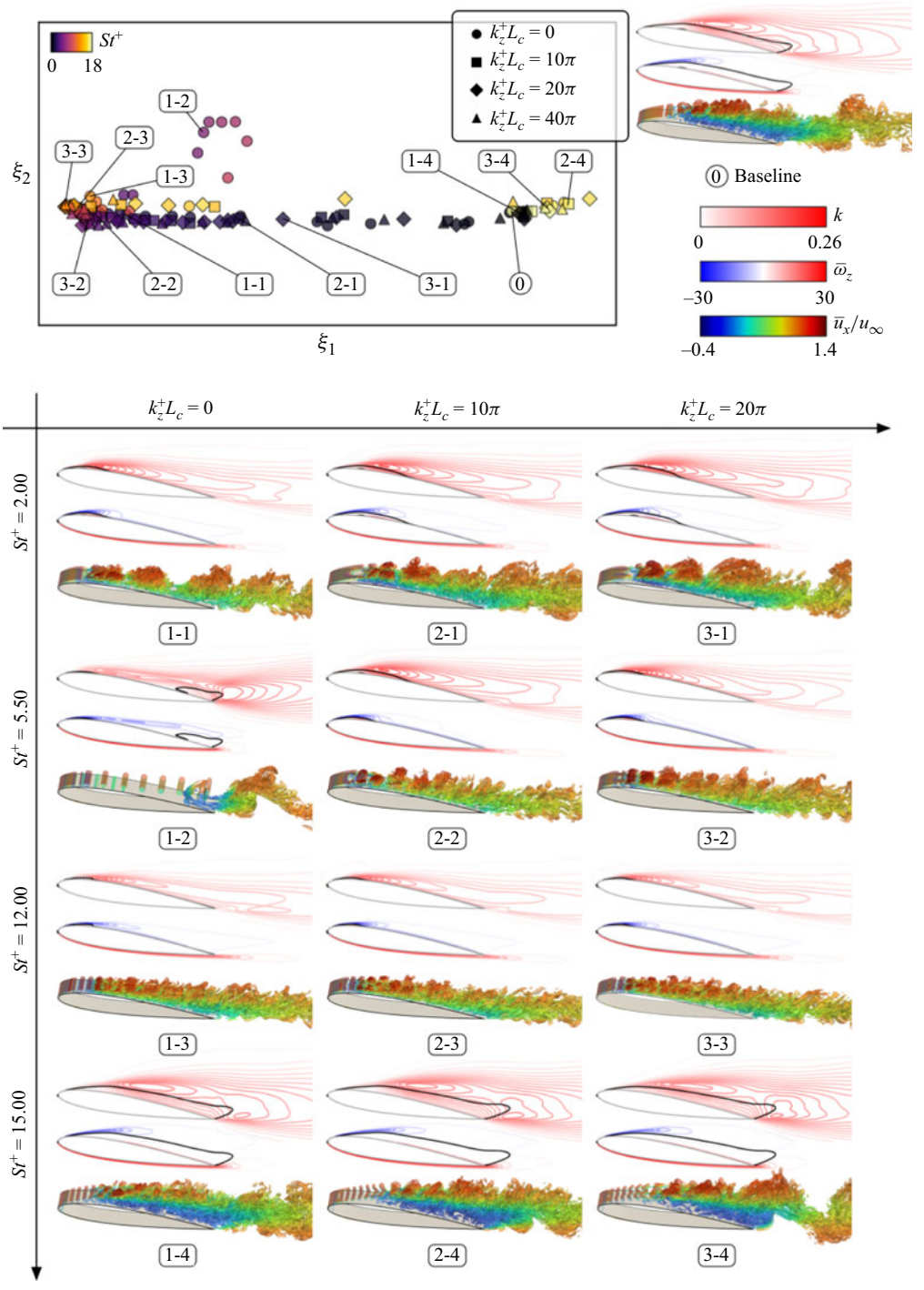


Figure 8. Plot of latent embeddings highlighting the influence of actuation parameters for $\alpha = 9°$. Labelled are example actuation cases. Examples of average turbulent kinetic energy, average vorticity fields, as well as instantaneous Q-criterion ($QL_c^2/u_\infty^2 = 50$) coloured by streamwise velocity are shown for the labelled cases. The $\bar{u}_x = 0$ isocontour is shown for all average fields.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

wake becomes lower, reflecting a reduction in the size of the separation bubble and the associated turbulent mixing region. We see that the two-dimensional actuation initially sees a faster rate of reduction of the separation bubble size with respect to increasing the actuation frequency, as seen in both the flow fields. This is also reflected in the relative placement of cases 1-1, 2-1 and 3-1 along ξ_1.

For $5 \lesssim St^+ \lesssim 8$, when $k_z^+ = 0$, there is an increase of ξ_1 and ξ_2 toward the cluster of outlier points (represented by case 1-2). Here, performance deteriorates for the two-dimensional actuation due to the reappearance of separation near the trailing edge, leading to a reduction lift-to-drag ratio (increase of ξ_1). In this range, two-dimensional actuation induces partial laminarisation over the suction surface, characterised by spanwise vortical structures that originate near the leading edge and break down further downstream, causing the trailing-edge separation. For $k_z^+ > 0$, trailing-edge separation is not observed, and there is no substantial increase of ξ_1 during this intermediate range of forcing frequencies. We see that the OT-based embedding clearly distinguishes the $k_z^+ = 0$ cases with partial laminarisation from the other control cases.

As the forcing frequency increases further to $9 \lesssim St^+ \lesssim 13$, the lift-to-drag ratio continues to improve, with the trailing-edge separation suppressed for $k_z^+ = 0$ (case 1-3), and we return to the left side of the ξ_1 axis in latent space. It is during this range of forcing frequencies that we are the furthest in the negative ξ_1 direction, and we observe the local optimal lift-to-drag ratio. This occurs at around $St^+ \approx 12$ for $k_z^+ = 10\pi$ and 20π (cases 2-3 and 3-3). For $St^+ \gtrsim 12$, ξ_1 increases sharply, indicating a loss of actuator effectiveness. As a result, flow separation reappears, and lift-to-drag ratio declines, returning to baseline levels (cases 1-4, 2-4, 3-4).

3.2. *The $\alpha = 6°$ latent space cases*

Next, we analyse the latent embedding for the $\alpha = 6°$ cases which exhibits a more diverse range of behaviours compared with $\alpha = 9°$. To visualise the effect of the OT embedding for $\alpha = 6°$, in figure 9 we show the latent embeddings for the $\alpha = 6°$ cases corresponding to the regular and OT-based autoencoders in addition to separation profiles for representative cases. Just as in the $\alpha = 9°$ case, while see that we can interpret a trend of the lift-to-drag ratio for the regular autoencoder in figure 9(*a*), we again emphasise that the relative placement of cases is arbitrary, and thus the cases are scattered in the latent space without any regard to any notion of similarity between the flow fields. We see that the OT-based embedding shown in figure 9(*b*) more succinctly captures the separation bubble behaviour (cases A–E) while distinguishing cases with laminarisation and trailing-edge separation (e.g. the branches containing cases F and G). In contrast to the $\alpha = 9°$ latent space, while we see that the lift-to-drag ratio visibly correlates with ξ_1, we also note that there is some dependence on ξ_2 as seen with the performance of some of these laminarised cases.

Again, as shown in figure 9(*b*), the ξ_1-coordinate captures the dominant mode of variation in the latent space with most cases being positioned along this axis. As was the case for the $\alpha = 9°$ cases, a decrease in ξ_1 generally corresponds to a shrinking bubble that moves toward the leading edge, before disappearing completely. Likewise, the ξ_2-coordinate identifies control cases in which the flow exhibits different behaviour in the wake due to laminarisation. However, a major distinguishing feature of the $\alpha = 6°$ latent space compared with the $\alpha = 9°$ one is that, for the cases with leading-edge separation, the separation bubble size alone does not as effectively characterise distinct cases. For example, cases E and F have very similar separation bubble profiles; however, they clearly exhibit different control performance and placement in the latent space. To understand why this is the case, we note that, unlike the $\alpha = 9°$ cases, which only experience partial

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

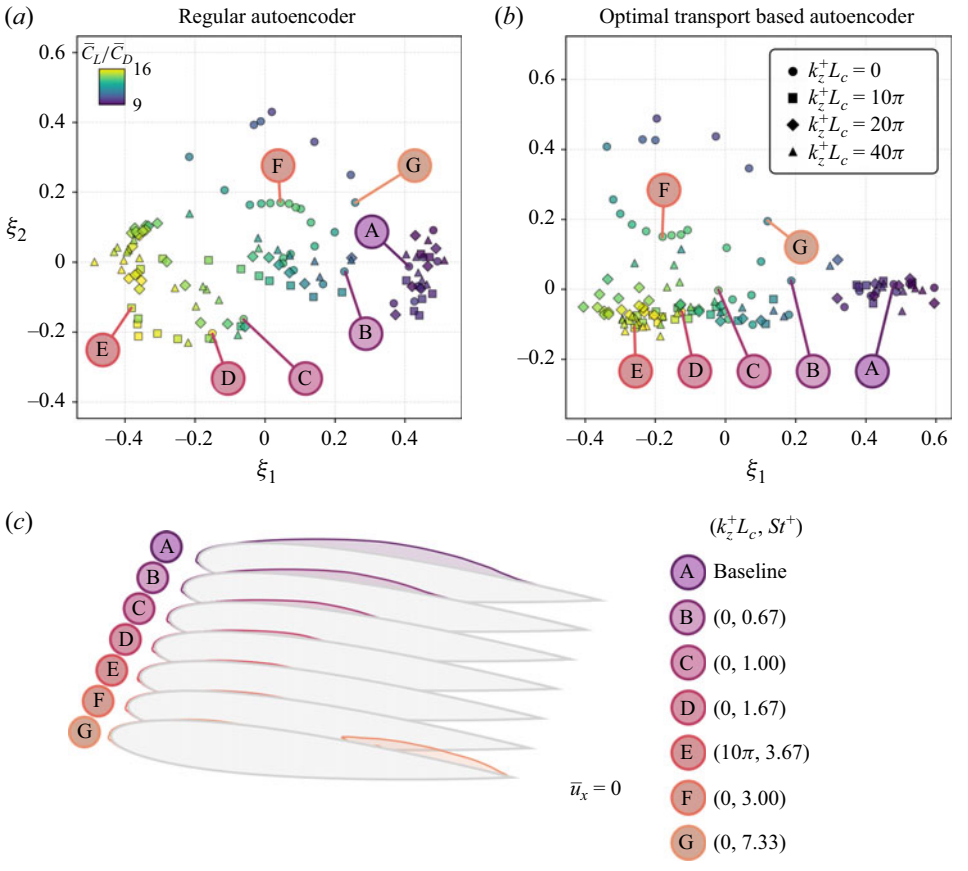


Figure 9. Comparison of learned latent spaces with $\alpha = 6°$ coloured by control performance $\bar{C}_L/\bar{C}_D$ for (*a*) standard autoencoder ($\mathcal{L}_1$ loss only), (*b*) OT autoencoder ($\mathcal{L}_1$ and $\mathcal{L}_2$ losses). (*c*) Isocontours of time-averaged streamwise velocity $\bar{u}_x = 0$ shown for different representative cases labelled in (*a*) and (*b*).

laminarisation, the $\alpha = 6°$ latent space includes cases in which separation is suppressed and the flow field is globally laminarised (case F). To make this point clearer, in figure 8, we again show the variation of the latent space with respect to the control parameters in addition to example average and instantaneous flow fields corresponding to labelled cases. For cases with global laminarisation, spanwise vortical structures that originate from the leading edge indicate the suppression of three-dimensional flow structures as they advect downstream. These cases are distinguished from partially laminarised fields that have trailing-edge separation due to the breakdown of these spanwise structures (e.g. case G in figure 9(*c*) or case 1-3 in figure 8). This manifests in cases with very similar separation profiles, but different wake behaviour. Additional effects associated with the laminarisation of the flow captured in the ξ_2 coordinate are seen in the variation of the turbulent kinetic energy fields, which also play a significant role in the estimation of the lift-to-drag ratio for $\alpha = 6°$.

Despite the aforementioned differences, the overall behaviour within the latent space when the forcing frequency is varied parallels that of the $\alpha = 9°$ cases. If we start at the baseline flow (case 0) in figure 10, increasing the forcing frequency within $0 \lesssim St^+ \lesssim 2$ results in an initial decrease of ξ_1 corresponding to the suppression of separation and improvement of the control performance (e.g. cases 1-1, 2-1, 3-1).

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

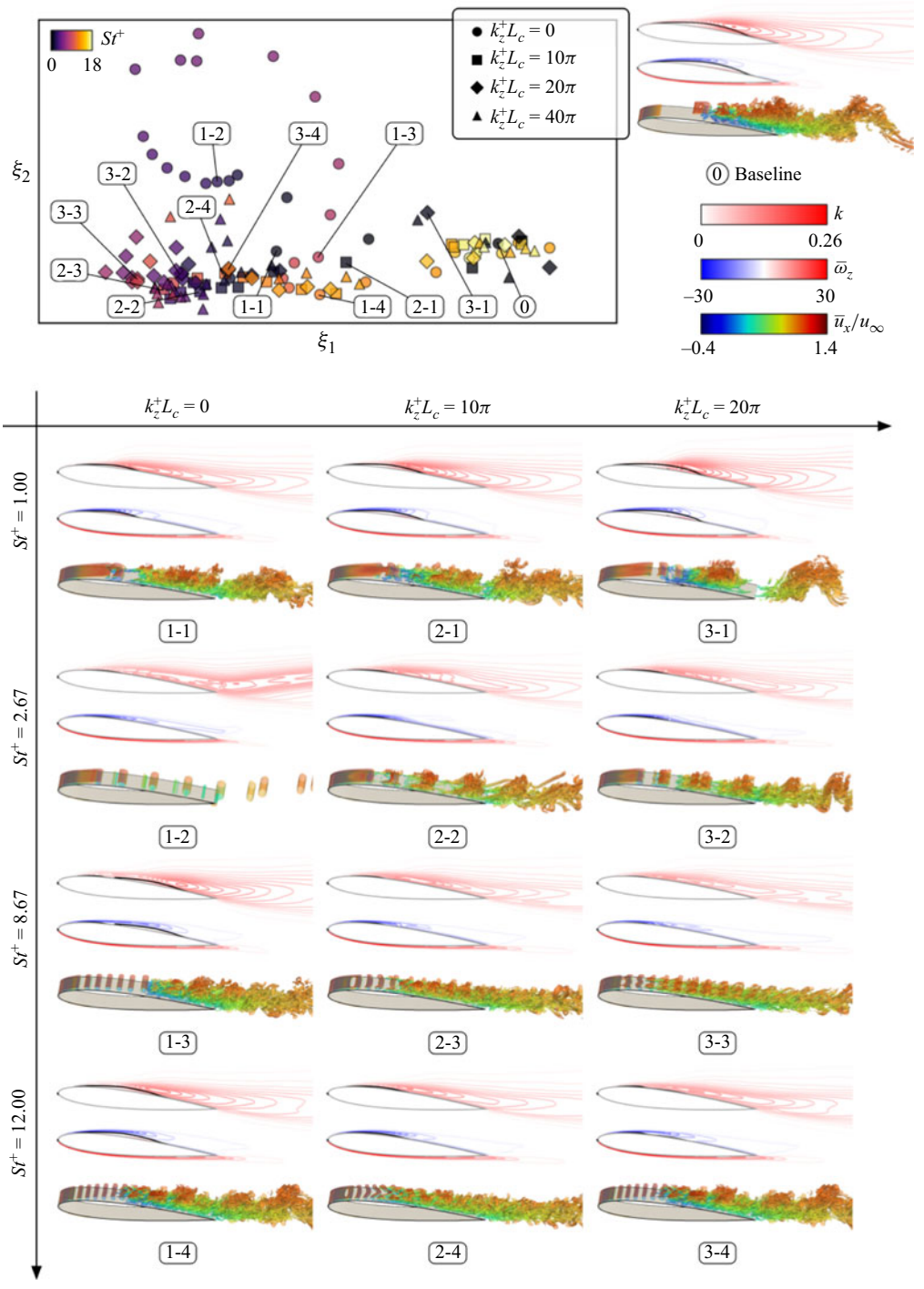


Figure 10. Plot of latent embeddings highlighting the influence of actuation parameters for $\alpha = 6°$. Labelled are example actuation cases. Examples of average turbulent kinetic energy, average vorticity fields, as well as instantaneous Q-criterion ($QL_c^2/u_\infty^2 = 50$) coloured by streamwise velocity are shown for the labelled cases. The $\bar{u}_x = 0$ isocontour is shown for all average fields.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

For cases with $k_z^+ > 0$, increasing St^+ past this range would result in moving further left along ξ_1 until the local maximum in the lift-to-drag ratio is reached in the range of $2 \lesssim St^+ \lesssim 12$ as separation is eventually completely suppressed (e.g. cases 2-2, 3-2). For $k_z^+ = 0$, while ξ_1 is decreased, reflecting the suppression of separation, around $St^+ \approx 2$ ($\xi_1 \approx -0.1$), there is a vertical jump in ξ_2 corresponding to cases experiencing global laminarisation (case 1-2). Here we see that the turbulent kinetic energy field shows a distinctively different appearance in the separated flow. The wake is directed upwards and is elongated towards the trailing edge, reflecting the advection of the pairs of spanwise rollers downstream. In this globally laminarised regime, even though there is a reattachment of the wake, there is a decrease in the lift-to-drag ratio when compared with cases with similar separation bubble sizes (similar ξ_1 position).

When $k_z^+ = 0$ and $4 \lesssim St^+ \lesssim 8$, the vortical structures merge and break down near the trailing edge (e.g. case G in figure 9), resulting in separation and a further decrease in performance. Eventually, the separation bubble begins to move back up towards the leading edge and rejoins the main distribution of cases along the ξ_1 axis (case 1-3). Moving between cases 1-2 and 1-3, we see that ξ_2 primarily characterises differences in the trailing-edge behaviour. These cases observed strengthened turbulent kinetic energy mixing near the trailing edge and the wake, while cases along the ξ_1 axis primarily have the strongest turbulent kinetic energy at the rear edge of the separation bubble over the mid-chord and leading edge. For $St^+ \gtrsim 12$, ξ_1 increases for all k_z^+ as we again return to baseline performance as the controller has diminished effectiveness (e.g. cases 1-4, 2-4, 3-4).

In summary, these findings suggest that with the use of OT, the wide array of complex flow responses to thermal actuation can be effectively characterised by a low-dimensional representation that is consistent across the two AoA studied. For both $\alpha = 6°$ and 9°, the dominant mode of variation in response to control ξ_1 captures variations in separation bubble size and shows a strong correlation with the lift-to-drag ratio. Secondary effects are included in ξ_2, which encodes the behaviour of the turbulent kinetic energy wake response corresponding to laminarization of the flow and the position of separation.

4. Conclusion

We introduced an autoencoder framework that incorporates unbalanced OT distances to learn latent representations of flow fields that encode a physically interpretable notion of similarity. Unlike standard autoencoders, which may organise latent variables arbitrarily, provided reconstruction accuracy is obtained, the OT-based approach imposes a notion of geometric structure by attempting to align pairwise distances in latent space with OT geodesics in the input space. When applied to flows over a NACA 0012 airfoil subject to unsteady thermal actuation, we are capable of obtaining an interpretable representation of the wide array of responses to control in a low-dimensional space. The use of OT as a dissimilarity metric in this problem setting allows us to capture displacements of flow structures and aerodynamic performance trends in response to different control inputs.

The current results demonstrate that the flow response to thermal actuation can be succinctly represented by two latent variables with the use of the OT-based embedding. The first latent coordinate captures the primary response of the flow to control. Along this coordinate, we explicitly observed a correlation between the separation bubble size and the control performance in terms of the lift-to-drag ratio. The second coordinate was found to contain information on changes in the wake due to laminarisation of the flow

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

and trailing-edge separation. This interpretation of the learned latent coordinates in the OT-based approach was consistent across the two AoA studied. This consistency was not observed with the standard autoencoder formulation, suggesting the OT-based autoencoder uncovered shared physical relationships across the two sets of cases despite being trained separately for each angle of attack.

While this study employed a relatively simple scheme based on matching Euclidean distances in the latent space to OT-based dissimilarities in the input space, one may incorporate the intrinsic non-Euclidean geometry of the learned manifold. Although our analysis focused on a two-dimensional control parameter space, (k_z^+, St^+), the present framework can be applied for higher-dimensional parameter spaces where interpretability becomes increasingly difficult and a naive exploration of parameters is expensive. Additionally, while the current study analysed time-averaged flow fields, an extension to time-resolved or subsampled spatial domains (Fukami & Taira 2024) presents a promising direction for the utilisation of the OT-based approach for analysing flow structures and control mechanisms. Additionally, it may be of interest to consider OT distances between state trajectories, rather than individual flow fields, using ground cost metrics such as the Hausdorff distance between attractors (Ishar *et al.* 2019), the Fréchet distance (Alt & Godau 1995) or dynamic time warping (Berndt & Clifford 1994). Given that we have also learned a relationship between the latent coordinates and the control performance, future efforts can also address design optimisation in addition to uncertainty quantification. Additionally, the present approach may assist in other downstream tasks such as generative modelling, given that interpolants in the latent space are trained to coincide with OT geodesics, which may improve the quality of interpolation. With this in mind, the current results demonstrate the utility of OT in the representation and interpretation of complicated fluid physics.

Acknowledgements. J.T. and K.T. acknowledge support from the U.S. Air Force Office of Scientific Research (FA9550-23-1-0715) and the U.S. Department of Defense Vannevar Bush Faculty Fellowship (N00014-22-1-2798). The authors thank Dr L. Mathelin for insightful discussions on optimal transport.

Declaration of interests. The authors report no conflict of interest.

Appendix A. Definition of Csiszár divergence functionals

Here, we provide a high-level summary of Csiszár divergence functionals as described in Chizat *et al.* (2018), Liero, Mielke & Savaré (2018) and Nguyen (2025). To compute a Csiszár divergence functional, one must start with an entropy function. An admissible entropy function is a function $\varphi : \mathbb{R} \to \mathbb{R}_{\geqslant 0} \cup \{\infty\}$ which is convex, lower semicontinuous, $\mathrm{dom}(\varphi) \subset [0, \infty)$ and $\mathrm{dom}(\varphi) \cap (0, \infty) \neq \emptyset$.

From the entropy function, the Csiszár divergence functional can be defined for two measures $m_1, m_2 \in \mathcal{M}(\mathcal{X})$ as

$$\mathcal{D}_\varphi(m_1 \mid m_2) := \int_{\mathcal{X}} \varphi\left(\frac{\mathrm{d}m_1}{\mathrm{d}m_2}\right) \mathrm{d}m_2 + \varphi'_\infty m_1^\perp(\mathcal{X}), \tag{A1}$$

when m_1 and m_2 are non-negative (∞ otherwise). This expression relies on the Lebesgue decomposition, which decomposes m_1 into a part absolutely continuous with respect to m_2, denoted m_1^{ac}, and a singular part $m_1^\perp$ such that

$$m_1 = m_1^{ac} + m_1^\perp, \quad m_1^{ac} \ll m_2, \quad m_1^\perp \perp m_2. \tag{A2}$$

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

The Radon–Nikodym derivative $\mathrm{d}m_1/\mathrm{d}m_2$ is defined for the absolutely continuous part m_1^{ac}. The recession constant φ'_∞ is given by

$$\varphi'_\infty := \lim_{s\to\infty} \frac{\varphi(s)}{s} \in \mathbb{R} \cup \{\infty\}. \tag{A3}$$

In the present work, we compute the divergence using the Kullback–Leibler (KL) entropy φ_{KL} given by:

$$\varphi_{KL}(s) = \begin{cases} s\log(s) - s + 1 & \text{for } s > 0, \\ 1 & \text{for } s = 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{A4}$$

Note that, when the total measures are equal (e.g. probability measures) and $m_1^{\perp}(\mathcal{X}) = 0$, $\mathcal{D}_\varphi(m_1 \mid m_2)$ with φ_{KL} reduces to the familiar KL divergence from statistics.

Appendix B. Computation of the unbalanced optimal transport-based flow-field dissimilarity

As mentioned in § 2.1, the unbalanced OT distance d_{UOT} is defined as the infimum of the OT cost $\mathcal{J}$ over all possible transport plans $\Gamma \in \mathcal{M}_+(\mathcal{X}_1 \times \mathcal{X}_2)$

$$\mathcal{J}(\Gamma) := \int_{\mathcal{X}_1\times\mathcal{X}_2} C(x_1, x_2)\,\mathrm{d}\Gamma(x_1, x_2) + \rho\big[\mathcal{D}_{\varphi_1}(P_\#^{\mathcal{X}_1}\Gamma \mid m_1) + \mathcal{D}_{\varphi_2}(P_\#^{\mathcal{X}_2}\Gamma \mid m_2)\big], \tag{B1}$$

$$d_{UOT}(m_1, m_2) := \inf_{\Gamma\in\mathcal{M}_+(\mathcal{X}_1\times\mathcal{X}_2)} \mathcal{J}(\Gamma). \tag{B2}$$

However, this optimisation problem normally exhibits poor tractability. To compute the OT distance for histograms of dimension N_h, the computational cost scales at least in $\mathcal{O}(N_h^3 \log(N_h))$, in the general case where no restrictions are placed upon the metric used to define the OT cost (Pele & Werman 2009; Cuturi 2013). Consequently, rather than directly solving this constrained optimisation problem, it is common to replace the non-negativity constraint on Γ with an entropic regularisation term to obtain an approximate solution. The entropic regularisation changes the original linear programming problem to a strictly convex problem, which can be solved more efficiently using the Sinkhorn–Knopp matrix-scaling algorithm (Cuturi 2013; Chizat *et al.* 2018).

To define the entropically regularised problem, we consider the entropy of the transport plan. The negative entropy is given by

$$\mathcal{H}(\Gamma) = \int_{\mathcal{X}_1\times\mathcal{X}_2} r(\log r - 1)d(x_1, x_2), \tag{B3}$$

where the coupling Γ is assumed to admit a density r with respect to the reference Lebesgue product measure on $\mathcal{X}_1 \times \mathcal{X}_2$.

The entropically regularised UOT distance between measures m_1 and m_2 is the following convex optimisation problem:

$$d^{\varepsilon}_{UOT}(m_1, m_2) = \min_{\Gamma\in\mathcal{M}(\mathcal{X}_1\times\mathcal{X}_2)} (\mathcal{J}(\Gamma) + \varepsilon\mathcal{H}(\Gamma)), \tag{B4}$$

where $\varepsilon > 0$ is a small regularisation constant. Here, the inclusion of the entropic regularisation term $\mathcal{H}(\Gamma)$ follows the maximum-entropy principle and serves as a relaxation of the non-negativity constraint on Γ (Cuturi 2013; Chizat *et al.* 2018).

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

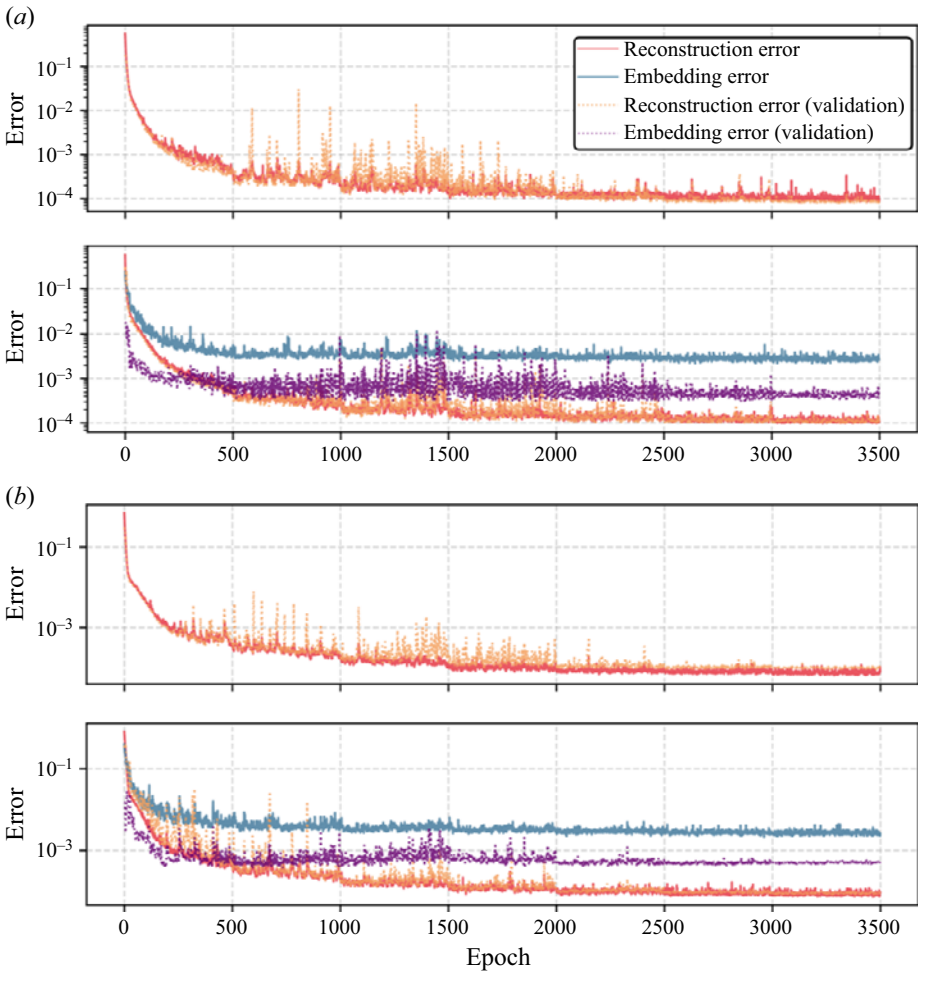


Figure 11. Plot of loss curves for (*a*) $\alpha = 9°$ and (*b*) $\alpha = 6°$ cases of both regular (top) and OT-based autoencoders (bottom).

It is important to note that the introduction of the entropic regularisation results in a non-zero bias even when the input measures are identical (Genevay *et al.* 2019; Séjourné *et al.* 2019). Because of this the following normalisation is performed to obtain what is called the Sinkhorn divergence:

$$S^{\varepsilon}(m_1, m_2) = d^{\varepsilon}_{UOT}(m_1, m_2) - \frac{1}{2}\left[d^{\varepsilon}_{UOT}(m_1, m_1) + d^{\varepsilon}_{UOT}(m_2, m_2)\right] + \frac{\varepsilon}{2}\left(\int_{\mathcal{X}_1} \mathrm{d}m_1 - \int_{\mathcal{X}_2} \mathrm{d}m_2\right)^2, \tag{B5}$$

which ensures that $S^{\varepsilon}(m_1, m_2) = 0$ when $m_1 = m_2$. We note that since we choose the cost function C to be the ℓ^2 distance between two spatial points and only consider a compact spatial domain $\mathcal{X}$, we have that for all $\varepsilon > 0$ the Sinkhorn divergence is positive definite and convex in each argument because the Euclidean distance is symmetric, 1-Lipschitz with respect to the essential supremum norm, and $k_\varepsilon = e^{-C(\cdot,\cdot)/\varepsilon}$ is a positive universal kernel (Séjourné *et al.* 2019).

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

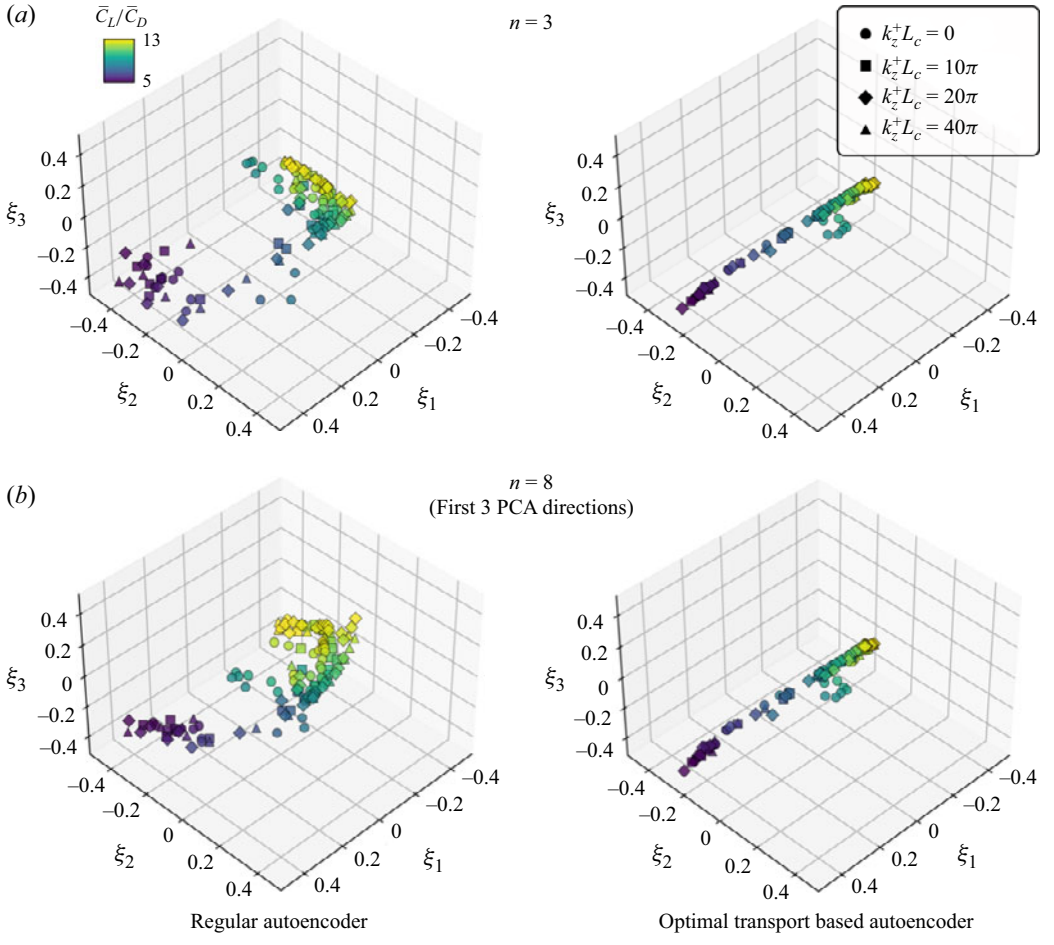


Figure 12. Plot of $\alpha = 9°$ latent space for both regular and OT-based autoencoder with (*a*) $n = 3$ and (*b*) $n = 8$ (showing the first 3 principal component analysis axes).

To compute the flow-field dissimilarity for signed measures (such as vorticity), we decompose the flow field into positive and negative parts. Consider a single flow-field variable over a given compact spatial domain represented by a function $V : \mathcal{X} \to \mathbb{R}$ assumed to be measurable. Using this function V, we define a signed measure with $\mathrm{d}m = V(x, y)\mathrm{d}x\mathrm{d}y$ (here, we use a two-dimensional domain $\mathcal{X}$ for notational simplicity, but the results are extensible to arbitrary spatial dimensions). This measure can be uniquely decomposed into two mutually singular positive measures m^+ and m^-, in this case given by

$$\mathrm{d}m^+ = V^+(x, y)\mathrm{d}x\mathrm{d}y, \tag{B6}$$

$$\mathrm{d}m^- = V^-(x, y)\mathrm{d}x\mathrm{d}y, \tag{B7}$$

where $V^+(x, y) = \max(\{V(x, y), 0\})$ and $V^-(x, y) = -\min(\{V(x, y), 0\})$ are the Radon–Nikodym derivatives of measures m^+ and m^- with respect to the reference Lebesgue measure. These define the density of the positive and negative parts of the flow field at a given point in the spatial domain. We can interpret m^+ as the 'amount' of V^+

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

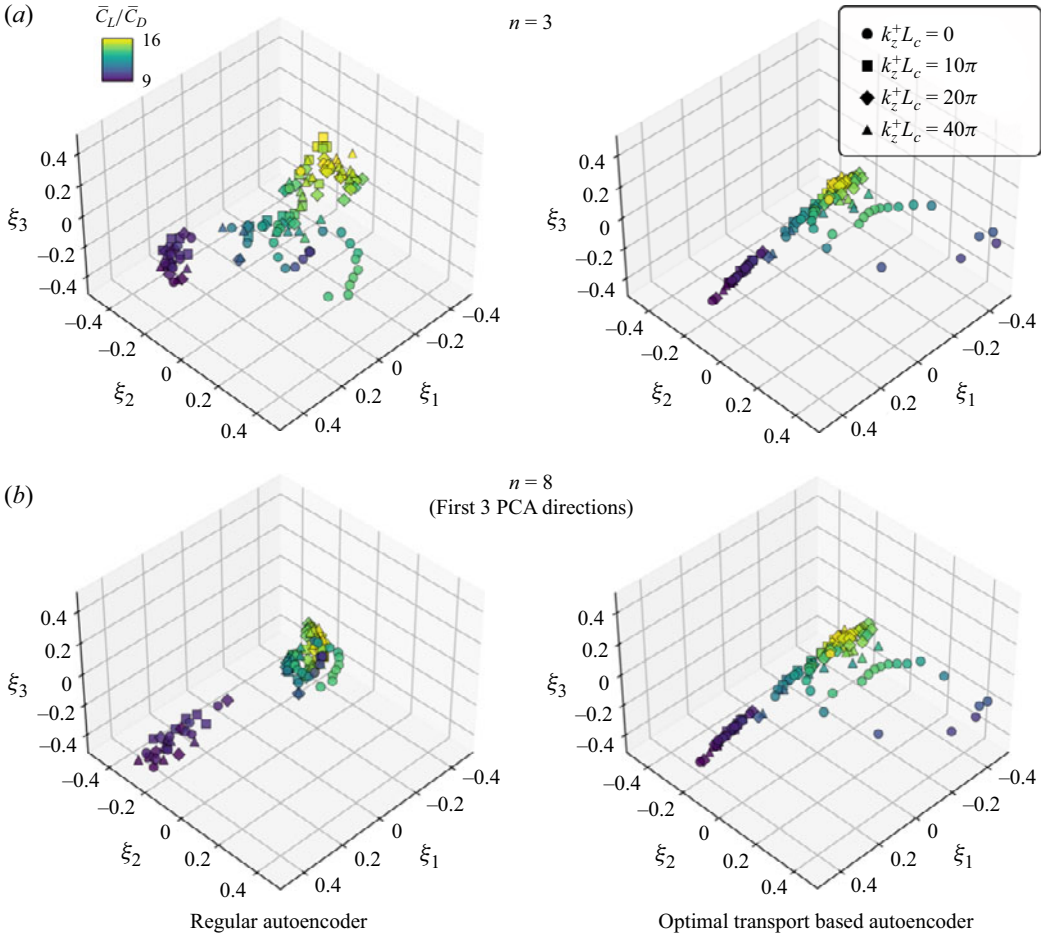


Figure 13. Plot of $\alpha = 6°$ latent space for both regular and OT-based autoencoder with (*a*) $n = 3$ and (*b*) $n = 8$ (showing the first 3 principal axes).

contained in some area and the densities as the spatial distribution of V^+ (and analogously for m^- and V^-).

We perform this decomposition for a field V_1, yielding measures m_1^+ and m_1^- as well as for a second field V_2, yielding a second set of measures which we call m_2^+ and m_2^-. We define a dissimilarity by considering the Sinkhorn divergences between the positive and negative parts of each field separately

$$d_{field}(V_1, V_2) = S^\varepsilon(m_1^+, m_2^+) + S^\varepsilon(m_1^-, m_2^-), \tag{B8}$$

inspired by Mainini (2012*a*,*b*).

Now, if we have two sets of fields of N variables, $\boldsymbol{V}_1 = \{V_1^{(i)}\}_{i=1}^N$ and $\boldsymbol{V}_2 = \{V_2^{(i)}\}_{i=1}^N$, we compute the OT-based dissimilarity for each variable and consider a total dissimilarity using

$$\tilde{d}_{field}(\boldsymbol{V}_1, \boldsymbol{V}_2) = \bigoplus_{i=1}^{N} d_{field}\big(V_1^{(i)}, V_2^{(i)}\big), \tag{B9}$$

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

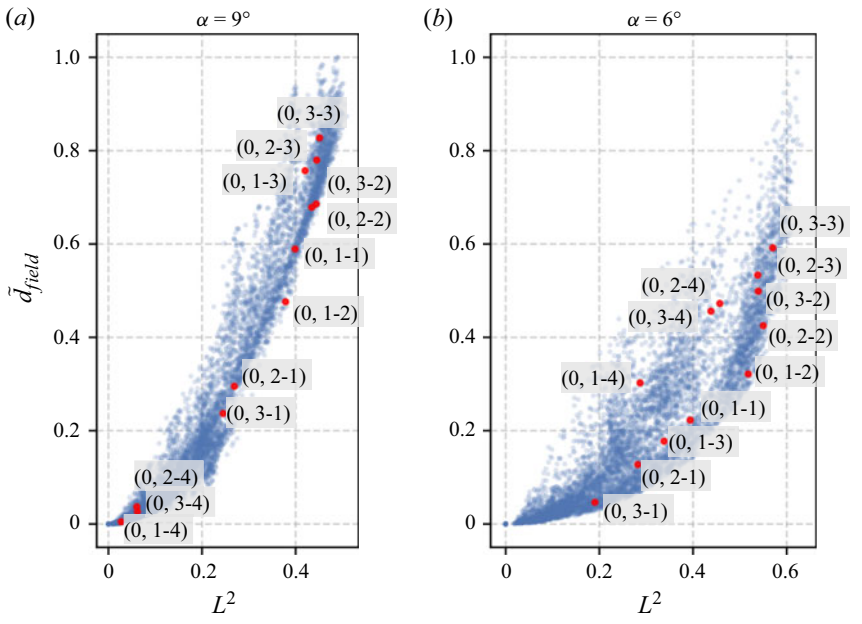


Figure 14. Plot of pairwise UOT-based distances ($\tilde{d}_{field}$) vs. pairwise L^2 distances for discretised flow fields for (*a*) $\alpha = 9°$ and (*b*) $\alpha = 6°$. Both axes have been normalised by the maximum pairwise distance. Red points denote example pairwise distances from the baseline flow corresponding to the cases in figures 8 and 10.

where $\bigoplus$ denotes some permutation invariant aggregation operator, such as a (possibly weighted) sum, mean or a vector ℓ^2 norm. For this work, we take the ℓ^2 norm of a vector consisting of the field dissimilarities in each variable.

Appendix C. Autoencoder model details

In the present analysis, we compare the learned latent spaces for two autoencoder architectures: a standard autoencoder that only uses reconstruction loss ($\mathcal{L}_1$ loss only) and the OT-based autoencoder ($\mathcal{L}_1$ and $\mathcal{L}_2$ losses). We note that both models can qualitatively reconstruct the flow fields at similar levels of accuracy. We find that the inclusion of the OT-based embedding term does not have any noticeable impacts on the training process in practice, as depicted by the example loss curves in figure 11.

During the parameter study, we find that, when increasing the latent space dimension, the reduction in the loss slowed after $n = 2$, prompting the use of a two-dimensional latent space. Additionally, we find that, for higher-dimensional latent spaces, while the regular autoencoder produced mostly arbitrary representations of the latent geometry, the OT-based autoencoder consistently produced the same two-dimensional structure as seen in figures 12 and 13 for $\alpha = 9°$ and 6°, respectively.

Appendix D. Pairwise distance comparison

To illustrate the difference between the OT-based distance and the L^2 distance we plot pairwise UOT-based distances ($\tilde{d}_{field}$) vs. pairwise L^2 distances for discretised flow fields normalised by the maximum distance in the dataset in figure 14. Similar to the case with the previous example of the advecting and diffusing gaussian in figure 2, while the distances appear correlated, the L^2 distance appears to saturate at around half the

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

maximum value of $\tilde{d}_{field}$ after the separation bubble has been suppressed. This means that the UOT distance is capable of more effectively distinguishing cases as the separation bubble is suppressed. This intuitively manifests as a change in curvature, in which most cases are flattened along the ξ_1 axis in the OT based autoencoder latent space.

References

Alt, H. & Godau, M. 1995 Computing the fréchet distance between two polygonal curves. *Intl J. Comput. Geometry Applics* **5** (01n02), 75–91.

Amitay, M. & Glezer, A. 2002 Role of actuation frequency in controlled flow reattachment over a stalled airfoil. *AIAA J.* **40** (2), 209–216.

Arjovsky, M., Chintala, S. & Bottou, L. 2017 Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research*, vol. 70, pp. 214–223. PMLR.

Arvanitidis, G., Hansen, L.K. & Hauberg, S. 2018 Latent space oddity: on the curvature of deep generative models. In *Proceedings of International Conference on Learning Representations*. ICLR.

Berndt, D.J. & Clifford, J. 1994 Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, pp. 359–370. AAAI Press.

Borg, I. & Groenen, P.J.F. 2005 *Modern Multidimensional Scaling: Theory and Applications*. Springer.

Brés, G.A., Ham, F.E., Nichols, J.W. & Lele, S.K. 2017 Unstructured large-eddy simulations of supersonic jets. *AIAA J.* **55** (4), 1164–1184.

Chizat, L., Peyré, G., Schmitzer, B. & Vialard, F.-X. 2018 Scaling algorithms for unbalanced transport problems. *Maths Comput.* **87** (314), 2563–2609.

Corke, T.C., Enloe, C.L. & Wilkinson, S.P. 2010 Dielectric barrier discharge plasma actuators for flow control. *Annu. Rev. Fluid Mech.* **42** (1), 505–529.

Cuturi, M. 2013 Sinkhorn distances: lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pp. 26. NeurIPS.

Davison, M.L. & Sireci, S.G. 2000 Multidimensional scaling. In *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, pp. 323–352. Elsevier.

Dovgal, A.V., Kozlov, V.V. & Michalke, A. 1994 Laminar boundary layer separation: instability and associated phenomena. *Prog. Aerosp. Sci.* **30**, 61–94.

Flamary, R., *et al.* 2021 Pot: Python optimal transport. *J. Machine Learning Res.* **22** (78), 1–8.

Freund, J.B. 1997 Proposed inflow/outflow boundary condition for direct computation of aerodynamic sound. *AIAA J.* **35** (4), 740–742.

Fukami, K. & Taira, K. 2023 Grasping extreme aerodynamics on a low-dimensional manifold. *Nat. Commun.* **14**, 6480.

Fukami, K. & Taira, K. 2024 Single-snapshot machine learning for super-resolution of turbulence. *J. Fluid Mech.* **1001**, A32.

Garnier, E., Adams, N. & Sagaut, P. 2009 *Large Eddy Simulation for Compressible Flows*. Springer Science & Business Media.

Genevay, A., Chizat, L., Bach, F., Cuturi, M. & Peyré, G. 2019 Sample complexity of sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1574–1583. PMLR.

Gorban, A.N. & Tyukin, I.Y. 2018 Blessing of dimensionality: mathematical foundations of the statistical physics of data. *Phil. Trans. R. Soc. A Math. Phys. Engng Sci.* **376** (2118), 20170237.

Greenblatt, D. & Wygnanski, I.J. 2000 The control of flow separation by periodic excitation. *Prog. Aerosp. Sci.* **36** (7), 487–545.

Häggmark, C.P., Bakchinov, A.A. & Alfredsson, P.H. 2000 Experiments on a two–dimensional laminar separation bubble. *Phil. Trans. R. Soc. Lond. A* **358** (1777), 3193–3205.

Hansen, P.C. & O'Leary, D.P. 1993 The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.* **14** (6), 1487–1503.

Harvey, J.K. & Perry, F.J. 1971 Flowfield produced by trailing vortices in the vicinity of the ground. *AIAA J.* **9** (8), 1659–1660.

He, K., Zhang, X., Ren, S. & Sun, J. 2016 Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.

Hinton, G.E. & Salakhutdinov, R.R. 2006 Reducing the dimensionality of data with neural networks. *Science* **313** (5786), 504–507.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

IOFFE, S. & SZEGEDY, C. 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pp. 448–456. PMLR.
ISHAR, R., KAISER, E., MORZYŃSKI, M., FERNEX, D., SEMAAN, R., ALBERS, M., MEYSONNAT, P.S., SCHRÖDER, W. & NOACK, B.R. 2019 Metric for attractor overlap. *J. Fluid Mech.* **874**, 720–755.
JAROSLAWSKI, T., FORTE, M., VERMEERSCH, O., MOSCHETTA, J.-M. & GOWREE, E.R. 2023 Disturbance growth in a laminar separation bubble subjected to free-stream turbulence. *J. Fluid Mech.* **956**, A33.
JOSLIN, R.D. & MILLER, D.N. 2009 *Fundamentals and Applications of Modern Flow Control*. American Institute of Aeronautics and Astronautics.
KANTOROVICH, L.V. 1960 Mathematical methods of organizing and planning production. *Manage. Sci.* **6** (4), 366–422.
KELSHAW, D. & MAGRI, L. 2024 Proper latent decomposition. In *Proceedings of the Summer Program*, pp. 175–184. Center for Turbulence Research, Stanford University.
KHALIGHI, Y., HAM, F., NICHOLS, J., LELE, S. & MOIN, P. 2011 Unstructured large eddy simulation for prediction of noise issued from turbulent jets in various configurations. In *17th AIAA/CEAS Aeroacoustics Conference (32nd AIAA Aeroacoustics Conference)*. AIAA Paper 2011-2886. AIAA.
KINGMA, D.P. 2013 Auto-encoding variational Bayes. arXiv:1312.6114.
KINGMA, D.P. 2014 Adam: a method for stochastic optimization. arXiv:1412.6980.
KIYA, M., SHIMIZU, M. & MOCHIZUKI, O. 1997 Sinusoidal forcing of a turbulent separation bubble. *J. Fluid Mech.* **342**, 119–139.
KLEWICKI, C., SCHENKMAN, A. & SPEDDING, G. 2025 Aerodynamic control of laminar separation on a wall-bounded airfoil at transitional Reynolds numbers. arXiv:2504.15191.
KOTAPATI, R.B., MITTAL, R., MARXEN, O., HAM, F., YOU, D. & CATTAFESTA III, L.N. 2010 Nonlinear dynamics and synthetic-jet-based control of a canonical separated flow. *J. Fluid Mech.* **654**, 65–97.
KRUSKAL, J.B. 1964 Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29** (2), 115–129.
KRUSKAL, J.B. & WISH, M. 1978 *Multidimensional Scaling*. Sage.
KVALHEIM, M.D. & SONTAG, E.D. 2023 Why should autoencoders work?. In *Transactions on Machine Learning Research*. JMLR.
LACHMANN, G.V. 1961 *Boundary Layer and Flow Control: Its Principles and Application*. Pergamon Press.
LECUN, Y., BOTTOU, L., BENGIO, Y. & HAFFNER, P. 2002 Gradient-based learning applied to document recognition. *Proc. IEEE* **86** (11), 2278–2324.
LEE, Y., YOON, S., SON, M. & PARK, F.C. 2022 Regularized autoencoders for isometric representation learning. In *International Conference on Learning Representations*. ICLR.
LIERO, M., MIELKE, A. & SAVARÉ, G. 2018 Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Invent. Math.* **211** (3), 969–1117.
LISSAMAN, P.B.S. 1983 Low-Reynolds-number airfoils. *Annu. Rev. Fluid Mech.* **15** (1), 223–239.
LUMLEY, J.L., YAGLOM, A.M. & TATARSKI, V.I. 1967 Atmospheric turbulence and radio wave propagation. *J. Comput. Chem.* **23** (13), 1236–1243.
MAGRI, L. & DOAN, A.K. 2022 On interpretability and proper latent decomposition of autoencoders. In *Proceedings of the Summer Program*, pp. 107–115. Center for Turbulence Research, Stanford University.
MAININI, E. 2012*a* A description of transport cost for signed measures. *J. Math. Sci.* **181**, 837–855.
MAININI, E. 2012*b* Well-posedness for a mean field model of Ginzburg–Landau vortices with opposite degrees. *Nonlinear Diff. Equ. Applics* **19**, 133–158.
MARXEN, O., LANG, M. & RIST, U. 2013 Vortex formation and vortex breakup in a laminar separation bubble. *J. Fluid Mech.* **728**, 58–90.
MOUSAVI, H. & ELDREDGE, J.D. 2025 Low-order flow reconstruction and uncertainty quantification in disturbed aerodynamics using sparse pressure measurements. *J. Fluid Mech.* **1013**, A41.
MUELLER, T.J. & DELAURIER, J.D. 2003 Aerodynamics of small vehicles. *Annu. Rev. Fluid Mech.* **35** (1), 89–111.
MURATA, T., FUKAMI, K. & FUKAGATA, K. 2020 Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *J. Fluid Mech.* **882**, A13.
NGUYEN, K. 2025 An introduction to sliced optimal transport: foundations, advances, extensions, and applications. *Foundations Trends® Comput. Graph. Vis.* **17** (3–4), 171–406.
PELE, O. & WERMAN, M. 2009 Fast and robust earth mover's distances. In *2009 IEEE 12th International Conference on Computer Vision*, pp. 460–467. IEEE.
PEYRÉ, G. & CUTURI, M. 2019 Computational optimal transport. *Foundations Trends Mach. Learning* **11** (5–6), 55–607.
PRANDTL, L. 1925 Magnuseffekt und Windkraftschiff. *Sci. Nat.* **13** (6), 93–108.
RACCA, A., DOAN, N.A.K. & MAGRI, L. 2023 Predicting turbulent dynamics with the convolutional autoencoder echo state network. *J. Fluid Mech.* **975**, A2.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press

RUBNER, Y., TOMASI, C. & GUIBAS, L.J. 1998 A metric for distributions with applications to image databases. In *Sixth International Conference on Computer Vision*. IEEE.

SCHMID, P.J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28.

SEIFERT, A. & PACK, L.G. 1999 Oscillatory control of separation at high Reynolds numbers. *AIAA J.* **37** (9), 1062–1071.

SÉJOURNÉ, T., FEYDY, J., VIALARD, F.-X., TROUVÉ, A. & PEYRÉ, G. 2019 Sinkhorn divergences for unbalanced optimal transport. arXiv:1910.12958.

SINHA, A., ALKANDRY, H., KEARNEY-FISCHER, M., SAMIMY, M. & COLONIUS, T. 2012 The impulse response of a high-speed jet forced with localized arc filament plasma actuators. *Phys. Fluids* **24** (12), 125104.

SMITH, L., FUKAMI, K., SEDKY, G., JONES, A. & TAIRA, K. 2024 A cyclic perspective on transient gust encounters through the lens of persistent homology. *J. Fluid Mech.* **980**, A18.

SYROTA, S., MORENO-MUÑOZ, P. & HAUBERG, S. 2024 Decoder ensembling for learned latent geometries. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM), Proceedings of Machine Learning Research*, vol. 251, pp. 277–285. PMLR.

TAIRA, K., BRUNTON, S.L., DAWSON, S.T.M., ROWLEY, C.W., COLONIUS, T., MCKEON, B.J., SCHMIDT, O.T., GORDEYEV, S., THEOFILIS, V. & UKEILEY, L.S. 2017 Modal analysis of fluid flows: an overview. *AIAA J.* **55** (12), 4013–4041.

TANI, I. 1964 Low-speed flows involving bubble separations. *Prog. Aerosp. Sci.* **5**, 70–103.

TAO, T. 2011 *An Introduction to Measure Theory*. American Mathematical Society.

TRAN, J., FUKAMI, K., INADA, K., UMEHARA, D., ONO, Y., OGAWA, K. & TAIRA, K. 2024 Aerodynamics-guided machine learning for design optimization of electric vehicles. *Commun. Engng* **3** (1), 174.

VASERSTEIN, L.N. 1969 Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii* **5** (3), 64–72.

VILLANI, C. 2009 *Optimal Transport: Old and New*. Springer.

VINUESA, R. & BRUNTON, S.L. 2022 Enhancing computational fluid dynamics with machine learning. *Nat. Comput. Sci.* **2** (6), 358–366.

VREMAN, A.W. 2004 An eddy-viscosity subgrid-scale model for turbulent shear flow: algebraic theory and applications. *Phys. Fluids* **16** (10), 3670–3681.

WILLIAMS, D. 1991 *Probability with Martingales*. Cambridge University Press.

WU, J.-Z., LU, X.-Y., DENNY, A.G., FAN, M. & WU, J.-M. 1998 Post-stall flow control on an airfoil by local unsteady forcing. *J. Fluid Mech.* **371**, 21–58.

WYGNANSKI, I. 1997 Boundary layer and flow control by periodic addition of momentum. In *4th Shear Flow Control Conference*, pp. 2117. AIAA.

YARUSEVYCH, S., SULLIVAN, P.E. & KAWALL, J.G. 2009 On vortex shedding from an airfoil in low-Reynolds-number flows. *J. Fluid Mech.* **632**, 245–271.

YEH, C.-A., MUNDAY, P. & TAIRA, K. 2017*a* Use of local periodic heating for separation control on a NACA 0012 airfoil. In *55th AIAA Aerospace Sciences Meeting*, AIAA Paper 2017-1451. AIAA.

YEH, C.-A., MUNDAY, P.M. & TAIRA, K. 2017*b* Laminar free shear layer modification using localized periodic heating. *J. Fluid Mech.* **822**, 561–589.

YEH, C.-A. & TAIRA, K. 2019 Resolvent-analysis-based design of airfoil separation control. *J. Fluid Mech.* **867**, 572–610.

ZAMAN, K.B.M.Q., MCKINZIE, D.J. & RUMSEY, C.L. 1989 A natural low-frequency oscillation of the flow over an airfoil near stalling conditions. *J. Fluid Mech.* **202**, 403–442.

ZENG, K., LINOT, A.J. & GRAHAM, M.D. 2022 Data-driven control of spatiotemporal chaos with reduced-order neural ODE-based models and reinforcement learning. *Proc. R. Soc. Lond. A* **478** (2267), 20220297.

https://doi.org/10.1017/jfm.2025.11077 Published online by Cambridge University Press