

From snapshots to manifolds – a tale of shear flows

E. Farzamnik^{1,†}, A. Ianiro^{2,†}, S. Discetti², N. Deng^{3,4}, K. Oberleithner⁵, B.R. Noack^{3,6} and V. Guerrero¹

¹Department of Statistics, Universidad Carlos III de Madrid, 28903 Getafe, Spain

²Department of Aerospace Engineering, Universidad Carlos III de Madrid, 28911 Leganés, Spain

³School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, PR China

⁴Institute of Mechanical Sciences and Industrial Applications, ENSTA-Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France

⁵Laboratory for Flow Instabilities and Dynamics, Technische Universität Berlin, 10623 Berlin, Germany

⁶Hermann-Föttinger-Institut, Technische Universität Berlin, 10623 Berlin, Germany

(Received 28 February 2022; revised 25 November 2022; accepted 30 November 2022)

We propose a novel nonlinear manifold learning from snapshot data and demonstrate its superiority over proper orthogonal decomposition (POD) for shedding-dominated shear flows. Key enablers are isometric feature mapping, Isomap, as encoder and, K -nearest neighbours (KNN) algorithm as decoder. The proposed technique is applied to numerical and experimental datasets including the fluidic pinball, a swirling jet and the wake behind a couple of tandem cylinders. Analysing the fluidic pinball, the manifold is able to describe the pitchfork bifurcation and the chaotic regime with only three feature coordinates. These coordinates are linked to the vortex-shedding phases and the force coefficients. The manifold coordinates of the swirling jet are comparable to the POD mode amplitudes, yet allow for a more distinct and less noise-sensitive manifold identification. A similar observation is made for the wake of two tandem cylinders. The tandem cylinders are aligned and located at a streamwise distance which corresponds to the transition between the single bluff body and the reattachment regimes of vortex shedding. Isomap unveils these two shedding regimes while the Lissajous plot of the first two POD mode amplitudes features a single circle. The reconstruction error of the manifold model is small compared with the fluctuation level, indicating that the low embedding dimensions contain the coherent structure dynamics. The proposed Isomap– KNN manifold learner is expected

† Email addresses for correspondence: ehsan.farzamnik@gmail.com, aianiro@ing.uc3m.es

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

to be of great importance in estimation, dynamic modelling and control for a large range of configurations with dominant coherent structures.

Key words: wakes, low-dimensional models, machine learning

1. Introduction

The chaotic nature of turbulent flows and their importance in physical and engineering systems has motivated countless studies attempting to obtain simplified models for control purposes in engineering applications. In particular, unbounded shear flows such as jets and wakes have received unabated interest due to their importance for drag reduction, control of unsteady loads, mixing enhancement, etc.

Despite their chaotic nature, such flows are characterized by recurrent flow patterns that are typically referred to as coherent structures. The beauty of the coherent structures has fascinated scientists since Leonardo da Vinci (Marusic & Broomhall 2021) and has hinted at the possibility that the flow dynamics can be represented as a system evolving on a low-dimensional attractor. Proper orthogonal decomposition (POD) (Berkooz, Holmes & Lumley 1993), also called principal component analysis in statistics, has received significant attention since it allows decomposing a flow field into orthogonal modes sorted according to their contribution to the variance of the quantity to be analysed. One of the most exploited advantages of POD in fluid mechanics is its capability to simplify the Navier–Stokes equations into a system of ordinary quadratic differential equations employing Galerkin projections (Noack *et al.* 2003). Low-order models obtained from POD open a door to a vast space of applications such as flow control (Brunton & Noack 2015) and also crisp least-order models for bifurcations and interactions of coherent structures (Deng *et al.* 2020).

A myriad of POD studies hint at low-dimensional manifolds describing turbulent shear flows. In the case of oscillatory flows, two-dimensional manifolds have been identified from laminar two-dimensional cylinder flows (Noack *et al.* 2003) to experimental turbulent wakes behind finite cylinders at high Reynolds numbers (Bourgeois, Martinuzzi & Noack 2013). These manifolds are the cornerstone of mean-field Galerkin models. Even flows with several frequencies may live on a mean-field manifold (Luchtenburg *et al.* 2009). The pioneering POD model of Aubry *et al.* (1988) derives such a manifold for the turbulent boundary layer from the Reynolds equations. For more complex flows, the energy spectrum of POD typically reveals $O(10)$ distinct most-energetic eigenvalues associated with physically interpretable modes. This distinct spectrum is usually followed by a steadily decreasing eigenvalue distribution with less interpretable and increasingly fine-scaled modes. More rigorously, if states have finite variance, the covariance operator for POD is positive trace class, hence compact, (see, e.g. theorem 4.1 in the book by Minh & Murino (2017)) and so its non-zero spectrum consists of isolated eigenvalues that are either a finite set or converge to 0 (see, e.g. theorem 6.8 in the book by Brezis (2011)). The distinct POD mode amplitudes can be conceptualized as the conductors of a large ‘slaved’ turbulence orchestra (Callaham, Brunton & Loiseau 2022).

While the focus of POD is on obtaining reduced-order models optimal in terms of energy, the field of statistical learning provides a vast number of tools for dimensionality reduction (Franklin 2005). Multidimensional scaling (MDS), for instance, is based on the singular value decomposition of the data distance matrix and allows representing a dataset in a low-dimensional space preserving the distance between the snapshots in the

high-dimensional space (Torgerson 1952; Kruskal 1964). Multidimensional scaling has been used in fluid mechanics mainly for visualization purposes and effectively captures some hidden features of the flows (Kaiser *et al.* 2014; Foroozan *et al.* 2021).

The capability of obtaining low-dimensional representations is tempting to identify embedded manifolds of the flows under study. Manifold learning attempts to recognize a low-dimensional surface, the manifold, near which the dataset actually resides. In statistical language, the dataset can be described to lie on or near a manifold in a low-dimensional space in which the manifold expresses some basic features of it. In this sense the manifold is in fact the set of relations that connect snapshots to each other. Interestingly, often high-dimensional systems appear to evolve on low-dimensional manifolds, thus simplifying their modelling if the manifold can be identified. These aspects pushed the development of the manifold learning techniques in the last decades. Remarkable examples are locally linear embedding (Roweis & Saul 2000), isometric mapping (Tenenbaum, de Silva & Langford 2000) and diffusion maps (Coifman & Lafon 2006).

The traditional dimensionality-reduction methods are structured on linear models and thus fail in capturing the manifolds when a nonlinear structure is present in the data. Turbulent flows exhibit a nonlinear behaviour which motivates the investigation of nonlinear models for manifold learning. As reported by Gorban & Karlin (2005), manifold learning falls in the field of model engineering, which attracts a wide interest among physics, mathematics and engineering, as is also evident from the three-volume book on model-order reduction by Benner *et al.* (2020*a,b*, 2021). For instance, autoencoders are increasingly used in fluid mechanics as a nonlinear dimensionality reduction architecture (Lee & Carlberg 2020). The fluid mechanics community has exploited the approximation of the Koopman operator to obtain nonlinear embeddings of fluid flows (Mezic 2013; Otto & Rowley 2021). Koopman modes are, for nonlinear dynamics, analogous to normal modes in linear problems, provided that the nonlinear dynamics are represented in the infinite-dimensional space of all possible observable measurements of the system. Giannakis *et al.* (2018) reported an insightful application of Koopman eigenfunction analysis on a complex three-dimensional flow such as a turbulent Rayleigh–Bénard convection cell. However, the identification of finite-dimensional coordinate systems in which the dynamics appear approximately linear remains an open challenge (Brunton *et al.* 2022). As shown by Morton *et al.* (2018) low-dimensional nonlinear dynamical models can be able to predict the cylinder wake and eventually be employed for model predictive control (resulting in simple, interpretable control laws).

Tenenbaum *et al.* (2000) have shown that the dimensionality reduction based on geodesic distances can be a powerful tool in preserving the actual behaviour of nonlinear datasets. This technique is referred to as isometric feature mapping, or Isomap. Despite the importance of identifying the manifold dimension for the modelling, prediction, control and understanding of fluid flows, surprisingly, the application of Isomap in fluid mechanics is minimal. Tauro, Grimaldi & Porfiri (2014) successfully employed Isomap to identify manifolds from flow-visualization data while others used it for combustion (Bansal, Mascarenhas & Chen 2011), and design optimization (Franz *et al.* 2014). Recently Otto & Rowley (2022) discussed the limitation of the linear methods in the case of selection and placement of sensors in a flow field.

This manuscript introduces a framework of manifold learning as an encoder for unbounded shear flows with a K -nearest neighbours (KNN) decoder. The input snapshots, which can be obtained either from a simulation or an experiment, are encoded using Isomap as the primary tool. The high-dimensional space transforms to a low-dimensional

space to identify the hidden embedding manifold of the dataset. In this new space, the manifold is interpreted to unravel the relationship between the manifold low-dimensional characteristics and the main features of the flow dynamics. We can reproduce the snapshots in the high-dimensional space with a KNN decoder using this new, easy-to-understand space and fast computing. This whole encoder–decoder model provides a robust framework to analyse shear flows and then implement applications (such as designing flow control systems) based on it.

Four datasets with different features have been used to investigate the framework's performance. The selected datasets vary from direct numerical simulations (DNS) of wake flows to fully turbulent experimental datasets with measurement noise. The simulation datasets are based on the wake of the fluidic pinball which in recent years has been shown to be a suitable test-bed configuration to study general flow phenomena like bifurcations and flow control (Deng *et al.* 2020). To study different flow regimes, the results from the simulations at $Re = 80$ and $Re = 130$ are reported and discussed, allowing us to identify the manifold learning capabilities both in a simpler bifurcation and in a more complex chaotic environment. The first experimental dataset consists of particle image velocimetry (PIV) measurements in a highly functional swirling jet configuration. This configuration has a wide application in modern gas turbine combustors and aerodynamically stabilizes lean premixed flames (Lückoff *et al.* 2017, 2021). Both the turbulent regime and the measurement noise challenge the encoder–decoder framework. The last tested dataset relates to the flow in the wake of two tandem cylinders. Tandem cylinders are characterized by several working regimes depending on the streamwise cylinder distance. The proposed dataset is at the intersection of two regimes, however, by using POD in a previous work, Raiola, Ianiro & Discetti (2016) could not unveil a regime switch.

The paper is organized as follows: in § 2 after the introduction, a detailed description of the developed framework is provided; the datasets and flow configurations employed are described in § 3; the most important outcomes from the analysis using the encoder–decoder framework are presented in § 4; and, finally, the conclusion and the possible future steps have been put in § 5. Two appendices describe a criterion for the choice of Isomap parameters and discuss possible criteria for the definition of the manifold residual variance.

2. Isomap – KNN manifold learner

In this work, a manifold learner methodology for fluid data is developed. The proposed approach consists of three steps. First, data is gathered either from simulations or experiments. Second, the thus-obtained data is embedded into a low-dimensional space using isometric feature mapping, Isomap, (Tenenbaum *et al.* 2000). This encoding part, which is fully data-driven, is carried out with the aim of revealing a hidden manifold that allows us to relate the new coordinates to physical features of the flow such as, for instance, force coefficients. Finally, a decoding part that enables return to the high-dimensional space and reconstructs the original flow field is developed. The proposed decoder is based on KNN and linear interpolation. Figure 1 shows the three stages of our procedure, which are described in detail in what follows.

Let us consider that N flow field snapshots have been observed, either from an experimental setting or a simulation. Considering that each snapshot is a matrix of P elements, the vectorized version of each snapshot is an observation (point) in the high-dimensional space \mathbb{R}^P , where each dimension (feature) contains information about a point of the field. Let $\mathbf{X} \in \mathbb{R}^{N \times P}$ be the data matrix containing the stated information and $\mathbf{x}_i \in \mathbb{R}^P$ be each of its rows, i.e. the flow fields for $i = 1, \dots, N$. The dataset in \mathbf{X} is

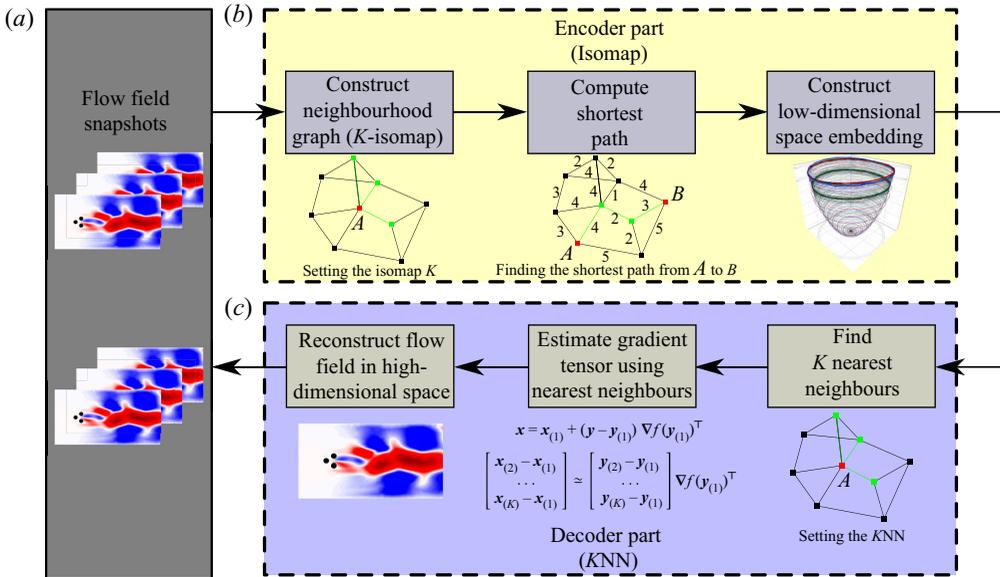


Figure 1. Encoder–decoder procedure: (a) obtaining flow field snapshots from simulations or experiments; (b) encoder part, representing the Isomap method application on input snapshots to identify the embedding manifold in low-dimensional space; (c) decoder part which reconstruct the flow field snapshots from low-dimensional space coordinates.

complex by nature and being able to extract a meaningful small number of coordinates that capture the main characteristics of the flow is challenging.

Isometric feature mapping is a nonlinear dimensionality reduction technique that finds a low-dimensional embedding of the data points that best preserve the geodesic distances measured in the high-dimensional input space. In order to estimate these geodesic distances, the shortest paths in a graph-connecting neighbouring points are employed. These distances are then used as an input in classical MDS (Torgerson 1952) to construct the low-dimensional embedding so that the Euclidean pairwise distances resemble those in the neighbouring graph. Therefore, the Isomap algorithm runs as follows. First, the Euclidean distances $d_{\mathbf{X}}(i, j)$ between flow fields x_i and x_j , corresponding to the i th and j th rows of \mathbf{X} , are computed for all $i, j = 1, \dots, N$. Second, for $i = 1, \dots, N$, $\mathcal{N}_{\mathbf{X}}^k(i)$, is defined as the set of the k closest observations to x_i . Based on these neighbourhoods, the neighbouring graph \mathbf{G} is defined over these data points such that two nodes (flow fields) i and j are connected by an edge of weight $d_{\mathbf{X}}(i, j)$ if they are neighbours, i.e. there is an edge between i and j if $x_j \in \mathcal{N}_{\mathbf{X}}(i)$. Observe that \mathbf{G} approximates the high-dimensional manifold containing the observed data. See Appendix A for a discussion about the choice of the number of neighbours k to build \mathbf{G} . Third, the shortest paths between all pair of vertices in \mathbf{G} are computed, yielding $d_{\mathbf{G}}(i, j)$ for all $i, j = 1, \dots, N$, using Floyd’s algorithm (Floyd 1962). Let $\mathbf{D}_{\mathbf{G}}$ be the matrix containing these shortest path distances. Finally, obtain the low-dimensional embedding $\mathbf{\Gamma} \in \mathbb{R}^{N \times p}$, $p \ll P$, using MDS. The new coordinates for the N samples are then found so that their pairwise Euclidean distance resembles $d_{\mathbf{G}}(i, j)$. This is equivalent to finding $\mathbf{\Gamma}$ which minimizes the cost function

$$\left\| \mathbf{\Gamma} \mathbf{\Gamma}^T - \mathbf{B} \right\|_F^2, \quad (2.1)$$

where $\mathbf{B} = -\frac{1}{2}\mathbf{H}^\top(\mathbf{D}_G \odot \mathbf{D}_G)\mathbf{H}$ is the Gram matrix in the input space, with $\mathbf{H} = \mathbf{I}_N - (1/N)\mathbb{1}_N$ being the centring matrix, \mathbf{I}_N the identity matrix of dimension N , $\mathbb{1}_N$ the all-ones matrix of dimension N , \odot the Hadamard (element-wise) product and $\|\cdot\|_F$ the Frobenius norm.

Fixing a dimension p for the low-dimensional embedding, the value of Γ minimizing the quantity in (2.1) is the matrix made up of the p eigenvectors $\mathbf{y}_1, \dots, \mathbf{y}_p$ corresponding to the p largest (positive) eigenvalues of the matrix \mathbf{A} arising from the eigendecomposition of \mathbf{B} , namely $\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ and $\mathbf{\Gamma} = \mathbf{V}_p$.

The aforementioned Isomap algorithm admits the choice of other norms different from the Euclidean, different ways of identifying the neighbours to construct \mathbf{G} , other shortest path algorithms or a non-classical approach to MDS. However, the choices made in our methodology are motivated by the implemented version of Isomap in the RDRTtoolbox in the R software (R Core Team 2020; Bartenhagen 2021), which has been used to carry out our analyses.

In order to assess the performance of Isomap, Tenenbaum *et al.* (2000) proposed using the definition of residual variance as in (2.2). Let \mathbf{D}_Γ be the matrix of Euclidean distances between each pair of points in the low-dimensional embedding. Then, the residual variance is defined as one minus the squared correlation coefficient between the vectorization of the distance matrices \mathbf{D}_G and \mathbf{D}_Γ , yielding

$$1 - R^2(\text{vec}(\mathbf{D}_G), \text{vec}(\mathbf{D}_\Gamma)), \tag{2.2}$$

where R^2 refers to the squared correlation coefficient and ‘vec’ is the vectorization operator. Observe that the result in (2.2) is a number between 0 and 1 which accounts for the amount of information that remains unexplained by the low-dimensional embedding of the original data. Therefore, the lower the value in (2.2) the better. For a discussion about the definition of residual variance to assess the performance of Isomap against other dimensionality-reduction methods such as POD we refer the reader to Appendix B.

In order to provide a decoder to create a correspondence between Isomap coordinates $\mathbf{y}_1, \dots, \mathbf{y}_p$ and the ones in the high-dimensional space, namely \mathbb{R}^P , we employ a purely data-driven approach. Any flow field $\mathbf{x}_i \in \mathbb{R}^P$ has its low-dimensional counterpart $\mathbf{y}_i \in \mathbb{R}^p$, $i = 1, \dots, N$. Then, let $f : \mathbb{R}^p \rightarrow \mathbb{R}^P$ be the unknown mapping which transforms the flow fields in the low-dimensional space onto the high-dimensional ones. To reconstruct the flow field for any $\mathbf{y} \in \mathbb{R}^p$ we assume that its K -nearest neighbours $\mathbf{y}_{(1)}, \dots, \mathbf{y}_{(K)}$ and their high-dimensional counterparts, namely $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(K)}$, are identified. Therefore, the reconstruction (or decoding) of \mathbf{y} , denoted as \mathbf{x} , can be obtained as a first-order Taylor expansion starting from the nearest neighbour to be mapped back to the original space, i.e. $\mathbf{x}_{(1)}$, as

$$\mathbf{x} = \mathbf{x}_{(1)} + (\mathbf{y} - \mathbf{y}_{(1)})\nabla f(\mathbf{y}_{(1)})^\top, \tag{2.3}$$

where the gradient tensor in $\mathbf{y}_{(1)}$, namely $\nabla f(\mathbf{y}_{(1)}) = (\partial f/\partial \mathbf{y}_1(\mathbf{y}_{(1)}), \dots, \partial f/\partial \mathbf{y}_p(\mathbf{y}_{(1)}))$ is estimated assuming an orthogonal projection of the $K - 1$ directions provided by the K -nearest neighbours in \mathbb{R}^p to those in \mathbb{R}^P . This is

$$\begin{bmatrix} \mathbf{x}_{(2)} - \mathbf{x}_{(1)} \\ \dots \\ \mathbf{x}_{(K)} - \mathbf{x}_{(1)} \end{bmatrix} \simeq \begin{bmatrix} \mathbf{y}_{(2)} - \mathbf{y}_{(1)} \\ \dots \\ \mathbf{y}_{(K)} - \mathbf{y}_{(1)} \end{bmatrix} \nabla f(\mathbf{y}_{(1)})^\top, \tag{2.4}$$

which yields $\nabla f(\mathbf{y}_{(1)})^\top = (\Delta \mathbf{Y}^\top \Delta \mathbf{Y})^{-1} \Delta \mathbf{Y}^\top \Delta \mathbf{X}$ if least squares minimization is used to approximate it, and $\Delta \mathbf{X}$ and $\Delta \mathbf{Y}$ are the left-hand side and the first term in the right-hand side of (2.4), respectively.

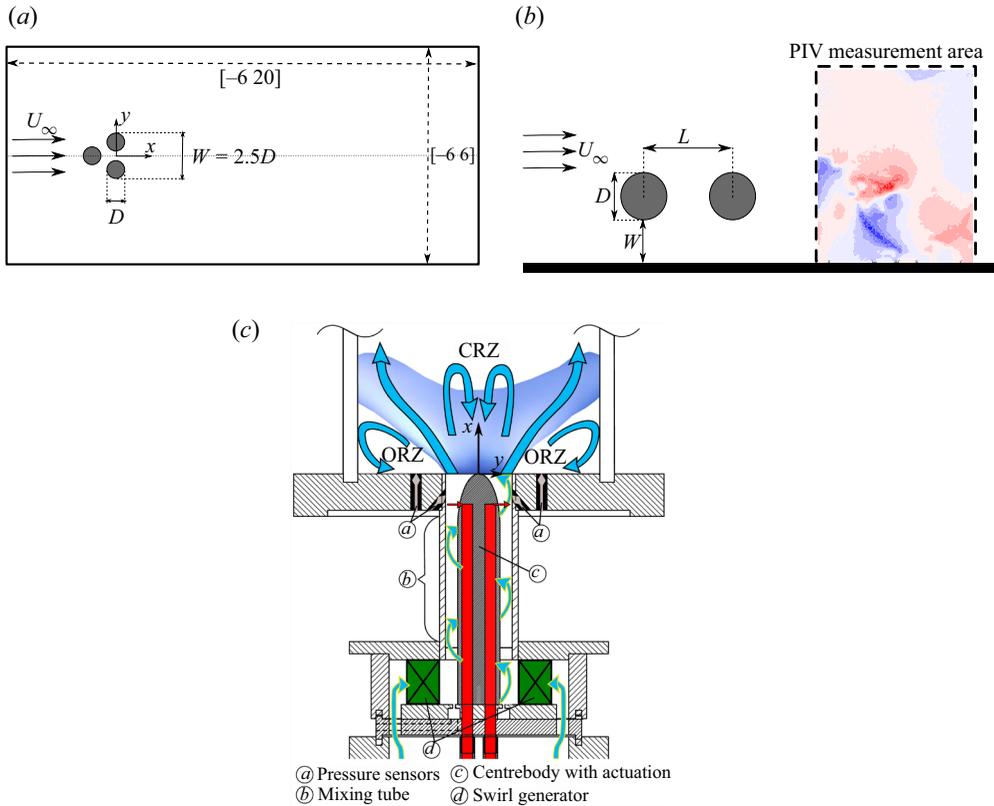


Figure 2. Dataset configurations: (a) fluidic pinball; (b) tandem cylinders; (c) swirling jet (Reprinted from Lückoff *et al.* (2021) with permission from Elsevier).

3. Datasets

In this section we describe the datasets that have been used to test our methodology. Three configurations are considered, which yield different flow fields and regimes under both experimental and simulation set-ups.

3.1. Fluidic pinball dataset

The fluidic pinball is a flow configuration consisting of three rotatable cylinders of equal diameter D whose axes are located in the vertices of an equilateral triangle, as sketched in figure 2(a). The triangle has a centre-to-centre side length $3D/2$ and is immersed in a viscous incompressible flow with a uniform upstream velocity U_∞ . The Reynolds number for this set-up is defined as $Re = U_\infty D / \nu$, where ν is the kinematic viscosity of the fluid. The wake flow undergoes a set of interesting transitions at different values of the Reynolds number. This allows exploration of reduced-order modelling and flow control strategies in a wide range of scenarios. In the recently published literature, the fluidic pinball has been used as a benchmark configuration for testing the mean-field modelling (Deng *et al.* 2020), cluster-based network modelling (Deng *et al.* 2022) and machine learning control (Raibaudo *et al.* 2020; Cornejo Maceda *et al.* 2021).

The numerical results of the fluidic pinball are investigated by employing a software developed to study multiple-input multiple-output flow control by Noack

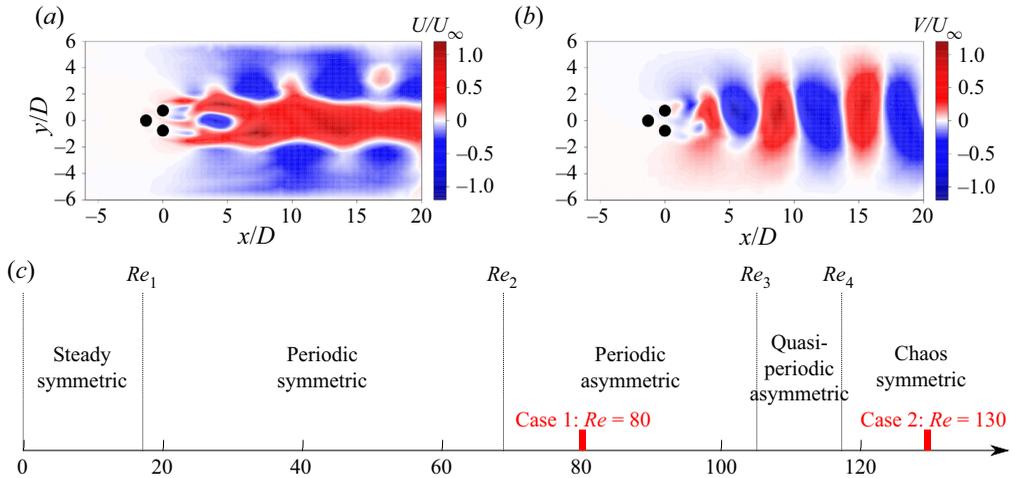


Figure 3. An example snapshot of the fluidic pinball at $Re = 130$ after subtracting the steady solution: (a) the contour of the streamwise fluctuating velocity component U ; and (b) the contour of the crosswise fluctuating velocity component V , normalized with the upstream velocity U_∞ . (c) Transitions of different flow regimes with varying Reynolds numbers.

& Morzyński (2017). Direct numerical simulations of the non-dimensionalized incompressible Navier–Stokes equations are used to compute the two-dimensional viscous wake behind the pinball configuration, where the variables are scaled with length D , velocity U_∞ , time D/U_∞ and density ρ . As shown in figure 2(a), the computational domain $[-6D, 20D] \times [-6D, 6D]$, excluding the interior of the cylinders, is described by a Cartesian coordinate system whose origin is located in the middle of the rear cylinders. As the rotation of the cylinders is not considered in this study, a no-slip condition on the cylinders, the far-field velocity U_∞ and a no-stress condition applied at the outlet of the domain are considered as the boundary conditions. The unsteady Navier–Stokes solver is based on third-order fully implicit time integration using an iterative Newton–Raphson approach and second-order finite-element method discretization on an irregular grid structure with 4225 T6 triangles and 8633 vertices (Deng *et al.* 2020). In order to simplify the distance calculation between snapshots, data is interpolated on a high-resolution uniform grid containing a total of 77 679 points. The steady solution, used as the initial condition at each corresponding Reynolds number, is calculated by the solver for the steady Navier–Stokes equations in the same way. In figures 3(a) and 3(b), the fluctuating streamwise and crosswise velocity field of an example snapshot at $Re = 130$ are shown after subtracting the symmetric steady solution.

With increasing Reynolds number, the flow experiences a transition from a laminar flow to periodic vortex shedding and finally to chaos (Deng *et al.* 2020). Five different regimes have been identified, as summarized in figure 3(c). The transition from a steady symmetric flow to a periodic symmetric vortex shedding occurs at $Re_1 \approx 18$ (following a Hopf bifurcation (Andronov *et al.* 1971; Strogatz *et al.* 1994)). The symmetry of the vortex shedding vanishes at $Re_2 \approx 68$ (pitchfork bifurcation (Strogatz *et al.* 1994)) thus entering a periodic asymmetric regime, presenting as an asymmetric vortex shedding with the base-bleeding jet deflected upward or downward. A secondary frequency appears with a higher Reynolds number, and the flow experiences another transition to a quasiperiodic asymmetric regime at $Re_3 \approx 104$ (Neimark–Säcker bifurcation (Kuznetsov

& Sacker 2008)), where the jet oscillates with the lower frequency. Finally, at $Re_4 \approx 115$, the flow enters into a chaotic symmetric regime with the jet oscillating randomly.

In this study, we focused on two different flow states at the selected Re , representative of the two most complex flow states identified by Deng *et al.* (2022). At $Re = 80$ for the periodic asymmetric regime, there exist a total of six invariant sets in the system state space: three unstable steady solutions; one unstable limit cycle; and two stable limit cycles, resulting from the primary Hopf bifurcation and the secondary pitchfork bifurcation. The unsteady flow is continuously sampled into snapshots with a fixed time step equal to 0.1. A time horizon is chosen long enough to ensure the flow data contain the transient and post-transient dynamics from the unstable steady state to the asymptotic state. The dataset at $Re = 80$ includes the snapshots from the simulations starting from the symmetric steady solution and its mirror-conjugated snapshots for a time horizon of 1500, as well as the snapshots from the simulations starting from the two mirror-conjugated asymmetric steady solutions for a time horizon of 1000. In this case, it is able to ensure that the dataset contains the complete manifold with six invariant sets. At $Re = 130$ for the chaotic symmetric regime, three unstable steady solutions and one chaotic attracting set can be found. The dataset considered at $Re = 130$ only contains the snapshots from the simulation starting from the symmetric steady solution for a time horizon of 1500, since we are interested in the transition from the unstable invariant set to the chaotic attracting set in this case. For the present analysis we limited the total number of snapshots to be analysed to 4000 for both Reynolds numbers. Snapshots are selected homogeneously from the whole range of the simulation covering complete evolution of the flow field until the final limit cycles. Therefore, simulations are performed at Re equal to 80 and 130, thus covering the typical complex wake dynamics regime with multiple invariant sets and chaotic attracting set.

3.2. Swirling jet dataset

Swirling jets have a wide variety of applications in modern gas turbine combustors and aerodynamically stabilize lean premixed flames. In the present work we analyse an experimental dataset obtained with stereoscopic PIV by Lückoff *et al.* (2021). Figure 2(c) shows a schematic of the swirling nozzle and of the measurement domain employed in the work by Lückoff *et al.* (2021). This configuration consists of a feeding line which provides the mass flow rate to a jet issued from a swirling nozzle. The swirling flow is produced using a radial swirl generator and the swirl number Sw , defined as the ratio of the axial flux of tangential momentum to the axial flux of axial momentum, can range between 0 and 1.5. The nozzle exit has a diameter of $D = 55$ mm and has a centrebody at the centre with diameter $D_{CB} = 35$ mm thus the hydraulic diameter of the mixing tube has a diameter of $D_h = 20$ mm. The Reynolds number is defined using the hydraulic diameter and the experimental facility can provide jets with a Reynolds number in the range [13 000, 32 000]. The present dataset has been generated for $Re = 20\,000$ and $Sw = 0.7$.

The measurement domain is located at the nozzle exit to analyse the flow in the combustion chamber. The combustion chamber is a cylinder with an inner diameter of $D_{CCh} = 200$ mm and a length of $L_{CCh} = 300$ mm and is made of quartz glass to enable flow measurement using time-resolved stereoscopic PIV. For the present dataset, 2183 snapshots were captured at a frame rate of 2500 f.p.s. The stereoscopic PIV system consists of a Nd:YLF diode pumped laser, which is synchronized with two high-speed CMOS cameras at 1024×1024 pixels image resolution. One camera was mounted perpendicular to the streamwise field of view and one was mounted with an angle of 40° to the

measurement plane. The two camera views were aligned using a multilevel calibration target. A laser light sheet of approximately 1 mm thickness was generated to illuminate the measurement area. For PIV seeding, heat-resistant solid titanium dioxide (TiO_2) particles of a nominal diameter of 2 μm were introduced to the flow far upstream using a brush-based seeding generator. The acquired particle snapshots were processed using a commercial PIV software which employs a correlation scheme with multigrid refinement. The final window size was set to 16×16 pixels with an overlap of 50 % in combination with spline-based image deformation and subpixel peak fitting. Finally, the estimated velocity fields were filtered for removing outliers, which were always less than 1 %. The velocity at the actuators outlet has been measured using hot-wire sensors. The reader can refer to the work by Lückoff *et al.* (2021) for further details on the measurement set-up and PIV image processing.

The main coherent structure in this kind of flow is a helical structure known as the precessing vortex core (PVC) (Syred 2006), which is generated due to a global self-excited instability (Müller *et al.* 2020). Although the origin of the PVC is well understood, its impact on combustion performance, especially flame stability, is still a matter for study. The existence of a dominant coherent structure in this flow encourages the idea that manifold learning can be successfully implemented to study the behaviour of the PVC under different conditions.

3.3. Tandem cylinders dataset

The third dataset employed in this work consists of flow field measurements in the wake of tandem cylinders near a wall. The data refer to the work by Raiola *et al.* (2016); the experimental configuration is summarized here for completeness. As sketched in figure 2(b), this configuration consists of two equal cylinders located in a cross-flow which are separated by a ratio of L/D denoted as the longitudinal pitch ratio (with L being the longitudinal distance between two cylinders centres and D the diameter of the cylinders equal to 32 mm). Both cylinders are placed at a similar distance to a wall with a ratio of W/D denoted as the wall gap ratio (with W being the distance of the cylinders from the wall). The wind-tunnel velocity U_∞ is set constant and equal to 2.3 m s^{-1} in order to achieve a Reynolds number of 4900, based on cylinder diameter (Raiola *et al.* 2016). In this study, the longitudinal pitch ratio is set to $L/D = 1.5$ and the wall gap ratio is set to $W/D = 3$.

An ensemble of 1800 flow field snapshots is employed for the present study. Velocity field measurements are performed with digital planar PIV. Di-ethyl-hexyl-sebacate (DEHS) droplets of approximately 1 μm diameter are employed to seed the flow. The acquisition is performed at 10 Hz with a TSI PowerView™ Plus 2MP Camera (with an array of 1600×1200 pixels) with a spatial resolution of approximately 7.2 pixels mm^{-1} . The light source employed is a Big Sky Laser CFR400 ND:Yag (230 mJ pulse $^{-1}$, pulse duration 3 ns). Image quality is improved by removing laser reflections and illumination background (Mendez *et al.* 2017). The interrogation strategy employed is an iterative multistep (Soria 1996) image deformation (Scarano 2001) algorithm, with a final interrogation window size of 16×16 pixels, with 50 % overlap resulting in a final vector spacing of approximately $0.035D$. Outlier vectors are identified with the universal median test by Westerweel & Scarano (2005) on a 3×3 vector kernel using a threshold equal to 2. Outliers are replaced with a distance-weighted average of valid neighbours.

While Raiola *et al.* (2016) reported that a gap ratio of $W/D = 3$ is sufficient to have negligible wall-interaction effects, three main flow behaviours in the wake of

tandem cylinders can be identified based on the Reynolds number and the distances between the two cylinders. Zdravkovich (1997) classified the flow around tandem cylinders with identical diameters into three major regimes (extended body; reattachment; and coshedding), depending on the longitudinal pitch ratio. At low longitudinal distance ($L/D < 1.5$), the vortex shedding for the upstream cylinder is suppressed, and the system acts as a unified bluff-body, which is categorized as the extended body regime or single bluff-body regime. By increasing the longitudinal distance ($1.5 < L/D < 4$), the flow starts to show a more complex behaviour which mainly can be characterized by the reattachment of the separated free shear layers from the upstream cylinder on the surface of the downstream cylinder. This regime is referred to as the ‘reattachment’ regime. Furthermore, by increasing the longitudinal distance, both cylinders feature the typical characteristics of the von Kármán vortex street. This regime is often defined the ‘coshedding’ regime. Alam *et al.* (2018) report the existence of a transitional L/D range between the reattachment and coshedding regimes also referred to as ‘critical’ or ‘bistable’ flow spacing. While it can be argued that a similar bistable regime should occur also between the extended body and the reattachment regimes, Raiola *et al.* (2016) did not identify such a feature for $L/D = 1.5$. This dataset appears to be especially suited to discover whether nonlinear manifold learning could unveil such a kind of bistable regime.

4. Results

This section presents and discusses the performance of the proposed encoder–decoder algorithm. The first subsection is dedicated to the performance of the encoder part. We discuss its strengths in unravelling the physical characteristics of the flows distilling the manifolds. In the second subsection, the decoder’s ability to reconstruct the original flow fields from the obtained low-dimensional coordinates is analysed.

4.1. Encoder’s capabilities

In this section, the embedding manifolds obtained from Isomap are presented for the datasets described in § 3.

In order to compute the dimensionality of the datasets at hand, the residual variance as defined in (2.2) is obtained for each number of dimensions. The choice of p is then made based on the elbow method following Tenenbaum *et al.* (2000). The dimension beyond which the residual variance experiences negligible variation can be identified as the true dimensionality of the dataset. This method is also widely used to compute the proper number of the clusters in the clustering techniques (Kaufman & Rousseeuw 1990). The ‘true’ dimensionality is the proper place in a trade-off between the simplicity of the embedded manifold and the loss of information due to truncation.

An example of residual variance as a function of the number of selected dimensions is reported in figure 4(a) for the case of the wake of the pinball at $Re = 80$ and $Re = 130$. For the lowest Reynolds number, it can be seen that three dimensions are sufficient to describe the bulk of the variance since the resulting residual variance for dimensions higher than three remains approximately the same. Note that due to its definition, the residual variance might not be monotonically decreasing with an increasing number of coordinates; for more details the reader is referred to Appendix B. On the other hand, truncating at three dimensions at $Re = 130$ appears to still be acceptable although it induces a larger error in terms of explained variance.

It must be remarked that, from now on, we are limited in plotting the projection of the manifold on the first three dimensions; nonetheless, the number of dimensions needed to

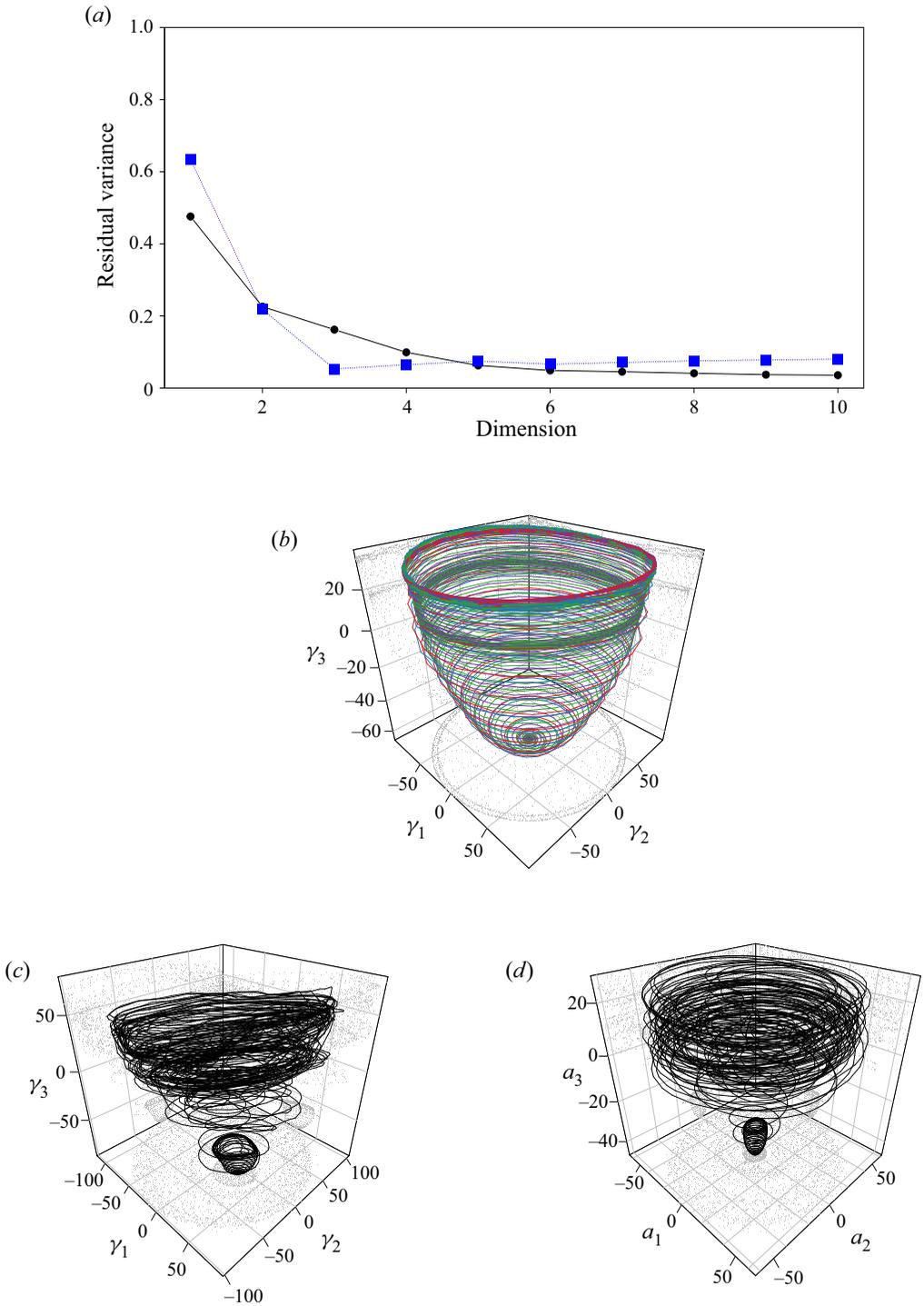


Figure 4. Isomap and POD results for the pinball datasets. (a) Residual Variance: blue, $Re = 80$; black, $Re = 130$. (b) Perspective view of the Isomap embedded manifold of $Re = 80$: purple, symmetric steady solution; green, flipped symmetric steady solution; red, asymmetric upward steady solution; blue, asymmetric downward steady solution. (c) Perspective view of the POD embedded manifold of $Re = 130$. (d) Perspective view of the Isomap embedded manifold of $Re = 130$.

represent accurately the manifold might be larger, depending on the complexity of the dynamics.

4.1.1. Fluidic pinball

Figure 4(a) shows the residual variances of the pinball configuration dataset for $Re = 80$ (blue) and $p = 1, \dots, 10$. For this configuration, the chosen value of k is equal to 8; for more details about the choice of k the reader is referred to Appendix A. The elbow is attained for $p = 3$ which is considered to be the true dimensionality of this problem. The residual variance is approximately the same for $p > 3$, thus, it can be argued that the manifold of input data, embedded in a higher-dimensional state space, has three key dimensions.

The same procedure has been done for the $Re = 130$ case, employing $k = 12$. The residual variance (see the black curve in figure 4a) decreases monotonically for an increasing number of dimensions and is already below 20% after three-dimensions ($p = 3$). However, the residual variance value at $p = 3$ increases by increasing the Reynolds number to higher values and entering the more chaotic regimes. This indicates that in the chaotic regime the ‘true’ dimensionality is higher due to the arising of a more complex dynamics. Therefore, it is reasonable to expect that, for increasing Reynolds number, the number of dimensions needed to explain the bulk of the variance should increase. This behaviour is not surprising, and it is observed in virtually all dimensionality-reduction techniques.

In terms of manifold shape, both for $Re = 80$ and 130, the data lies on a paraboloid (figures 4b and 4c accordingly) with the first two coordinates (γ_1 and γ_2) being representative of the periodic vortex shedding and the third coordinate being representative of a shift-mode characteristic of the transient dynamics from the onset of vortex shedding to the periodic von Kármán wake, analogously to what was found for the cylinder wake by Noack *et al.* (2003). A similar paraboloid shape is reported for the fluidic pinball at $Re = 30$ by Deng *et al.* (2020). It is remarkable that, when analysing all the solutions at $Re = 80$, the manifold correctly identifies the first unstable limit cycle for the symmetric unsteady solution and is able to identify the differences between the asymmetric upward and downward limit cycle. As well for $Re = 130$, the chaotic nature of the data shows a less smooth manifold, which still reveals the characteristic paraboloid shape. When employing POD and plotting the first three temporal modes, similar shapes could be obtained although less clear, especially at $Re = 130$, as evident from the comparison of figures 4(c) and 4(d). Plotting the POD results at $Re = 80$, the resulting manifold shows a similar behaviour as depicted in figure 4(b) and thus it is omitted for the sake of brevity.

Although the dimensions identified by Isomap do not necessarily have a physical meaning, it is a useful exercise to establish whether there exists some correlation between such coordinates and relevant flow quantities. Figure 5(a) shows a clear correlation between the drag coefficient and the coordinate γ_3 . This correlation is expected, since we observed that the third coordinate is representative of the shift mode. Some pairs of coordinates are related to higher-order harmonics of the flow, as is the case for $\gamma_1 - \gamma_2$, $\gamma_4 - \gamma_5$, and $\gamma_6 - \gamma_7$. Interestingly enough, if we extend the analysis to higher-order coordinates, it is possible to identify a high degree of correlation between the γ_8 and the lift coefficient C_L , as shown in figure 5(b). Although these interpretations are case-sensitive and can be affected by changes in the flow configuration and starting conditions in the Isomap algorithm, we have identified situations in which some coordinates in the low-dimensional space could be related to the main flow features. While it is outside of the scope of this paper to assess the interpretability of the coordinates identified by Isomap,

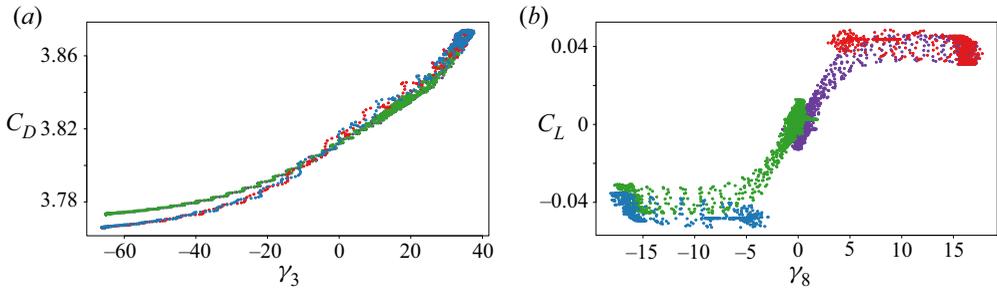


Figure 5. Relation between the coordinates of the embedded manifold and the force coefficients for the fluidic pinball at $Re = 80$: (a) the drag coefficient (C_D) versus third coordinate; (b) the lift coefficient (C_L) versus eighth coordinate; purple, symmetric steady solution; green, flipped symmetric steady solution; red, asymmetric upward steady solution; blue, asymmetric downward steady solution.

we spotlight the possibility of the existence of such a kind of correlation. This could be a powerful catalyst for the extension of the encoder–decoder framework presented here, and it will be object of future study.

4.1.2. Swirling jet

In the more challenging experimental case of the swirling jet, the same encoder procedure based on Isomap using $k = 8$ has been carried out. Here POD and Isomap performance as encoders is compared.

Figure 6(a) shows the residual variances of Isomap (black dots) and POD (green triangles), which is measured as stated in Appendix B, for different numbers of dimensions. In this case, the values of residual variance are significantly larger compared with those in the simulation cases studied before, most likely due to the turbulent nature of the flow and possibly due to measurement noise in the experimental data. It is worth noting that the residual variances of Isomap are lower than those of POD for all the dimensions depicted, thus indicating that Isomap is a better manifold learner in this case to preserve the geometry of the high-dimensional dataset. To investigate this advantage, we can compare the resulting embedding in both cases for three dimensions, which accounts for a reasonable amount of residual variance. Figures 6(b) and 6(d) show the encoded data by Isomap, whereas figures 6(c) and 6(e) illustrate the POD results, respectively. Although for both methods the general shape of the embedded manifold is similar to a hollow cylinder, the one obtained by Isomap shows a more clearly defined shape. Furthermore, the diameter of the hollow cylinder in POD is smaller and less circular with more spread points. Thus, the Isomap encoding provides a more helpful base to interpret the manifold and eventually relate the low-dimensional coordinates to the physical features of the flow.

4.1.3. Tandem cylinders

Regarding the tandem cylinder dataset, figure 7(a) shows that an appropriate number of dimensions in terms of the residual variance for both Isomap (black dots) and POD (green triangles) is two. Furthermore, as in the previous case, Isomap outperforms POD in terms of residual variance. As in the swirling-jet case discussed above, the Isomap results in figure 7(b) show a clearer manifold compared with the one obtained by POD (figure 7c). Furthermore, it also shows some separate groups of snapshots related to some physical features of the flow while the manifold resulted from POD ultimately fails to capture this behaviour of the system. Setting the number of groups to three, the results

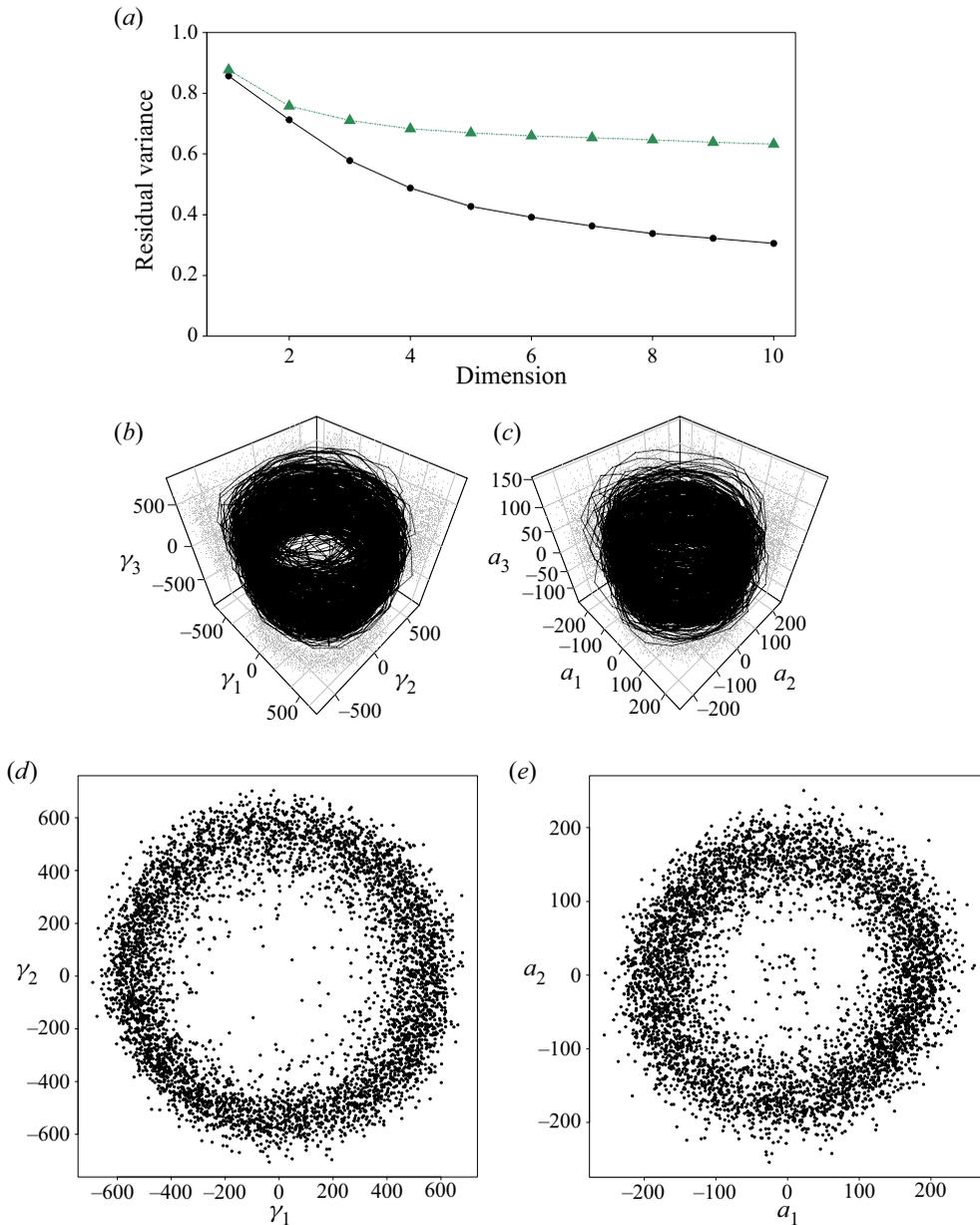


Figure 6. Resulted manifold from Isomap and POD for the swirling jet case. (a) Residual variance: green, POD; black, Isomap. (b) Three-dimensional Isomap. (c) Three-dimensional POD. (d) Top view, Isomap. (e) Top view, POD.

of the classification in the polar coordinates are shown by means of different colours in figure 7(b). The flow fields of each group close to the $\gamma_1 = 0$ line are plotted in figures 7(d)–7(f) from the outer group (Group 1) to the inner one (Group 3), as the representative of each group. We investigate each group by looking at these representative snapshots, which are highlighted with larger, numbered symbols in figure 7(b). As we move outward from the centre ($\gamma_1 = \gamma_2 = 0$), the distance between two consecutive

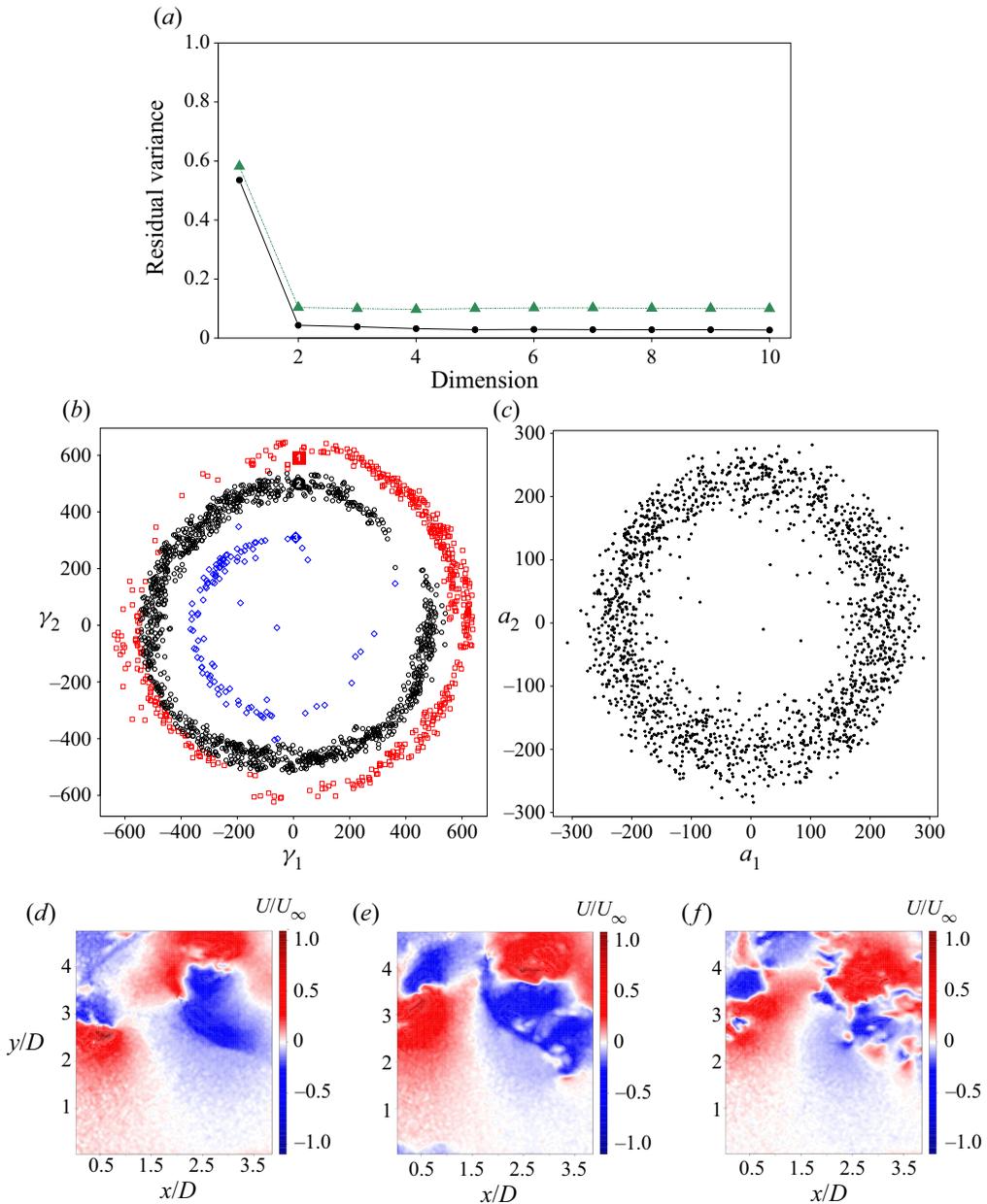


Figure 7. Resulted manifold from Isomap and POD for the tandem cylinders case. (a) Residual variance: green, POD; black, Isomap. (b) Isomap grouped embedding manifold: red, Group 1; black, Group 2; blue, Group 3. (c) Proper orthogonal decomposition. (d–f) Sample snapshots for Groups 1, 2 and 3 corresponding to the points labelled as 1, 2 and 3 on the manifold in panel (b).

vortices in the wake decreases, and the vortices appear less intense. In other words, moving from the groups from the outer part to the inner one, a transition between two different vortex shedding regimes is found which is conjectured to correspond to the bluff-body regime and the reattachment regime of the tandem cylinders. As described in § 3.3, previous studies suggest that this configuration of the flow lies on the bluff-body

Case	Re	N	P	p	k	K	Average NMSE
Pinball	80	4000	17 266	3	8	6	2.10
Pinball	130	4000	17 266	3	12	6	7.25
Tandem cylinders	4900	1800	30 414	2	3	4	37.50
Swirling jet	20 000	5396	26 004	3	8	6	41.05

Table 1. Manifold representation errors by using 70 per cent of the dataset as training dataset.

regime, as most of the snapshots in the low-dimensional space classify in the outer and middle groups, and the results of our encoding procedure are consistent with previous studies done by using POD (Raiola *et al.* 2016). However, Isomap can capture that even in this configuration, some behaviour of the next regime can coexist with the dominant one, suggesting that we cannot define a specific number for L/D as the classifier of the flow regimes, and by increasing L/D the flow smoothly changes its behaviours.

4.2. Decoder’s performance

This section assesses the quality of the decoder approach described in § 2. The normalized mean squared error (NMSE) in a test set of observations \mathcal{T} is computed as

$$NMSE = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|^2}{\|\mathbf{x}\|^2}, \quad (4.1)$$

where $\hat{\mathbf{x}}$ is the decoder’s reconstruction of the flow field $\mathbf{x} \in \mathbb{R}^P$.

Table 1 shows the average NMSE obtained for the different cases studied (last column), namely pinball for $Re = 80$ and $Re = 130$, the swirling jet flow and the wake of tandem cylinders. The sample data has been split into 70 % training and 30 % testing. The number of neighbours considered in Isomap (k) and the number of neighbours considered in the decoding stage (K) are reported in the sixth and seventh columns, respectively. The first column in table 1 depicts the case study, the second the value of Re , the third the number of samples, fourth the number of features in the high-dimensional space and fifth the number of dimensions considered for Isomap. Note that the NMSEs reported in table 1 are actually representation errors, i.e. the error incurred when reconstructing the original datasets by means of our encoder–decoder methodology. These representation errors show that the decoder has an excellent performance on the simulation datasets and has a reasonably good performance on both turbulent experimental datasets. Figure 8 compares reconstructed and actual snapshots for a random point in the test subset. As expected from the reported errors, the differences between the two snapshots in the pinball case are hardly noticeable. In the case of the tandem cylinders, although there are some noticeable differences between reconstructed and actual snapshots, the general behaviour of the flow is well preserved in the reconstructed snapshots.

To investigate further the decoder performance, a comparison using as input low-dimensional coordinates resulting from Isomap and POD has been carried out. The NMSE of the Isomap decoder is calculated using the same 70 % training dataset. The reconstruction NMSE with POD modes is calculated as the reconstruction error employing the first two or three modes, depending on the true dimensionality p of the dataset. Results are summarized in figure 9 and show a clear superiority of the KNN decoder

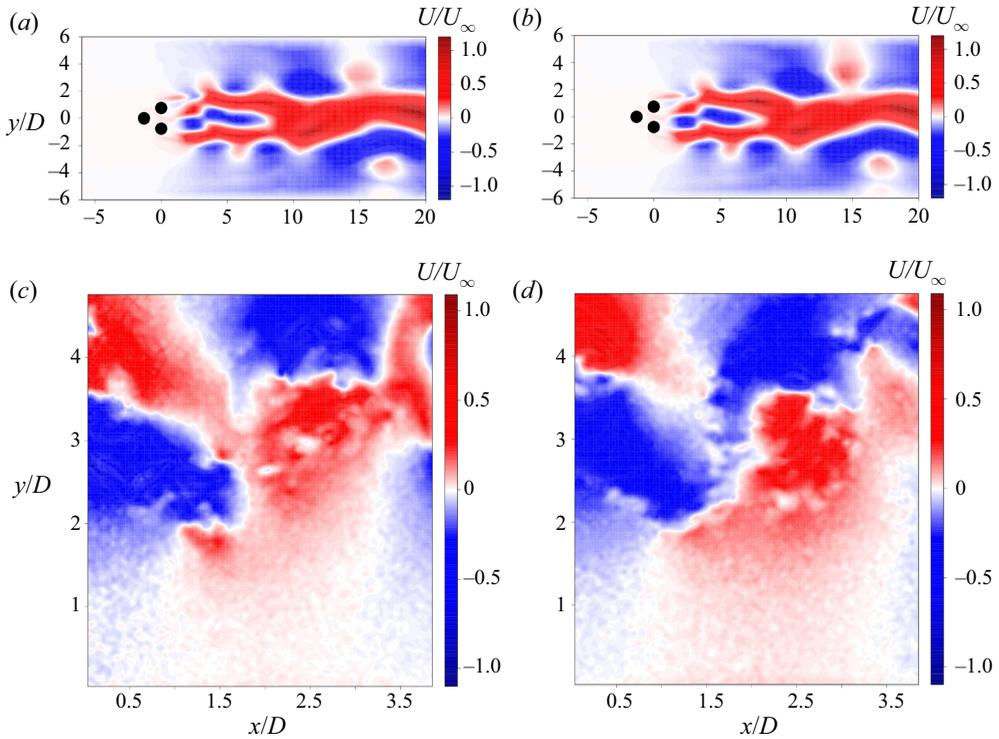


Figure 8. Reconstructed flow field using KNN decoder on the embedded manifold of the Isomap. (a) Reconstructed, pinball $Re = 130$. (b) Actual, pinball $Re = 130$. (c) Reconstructed, tandem cylinders. (d) Actual, tandem cylinders.

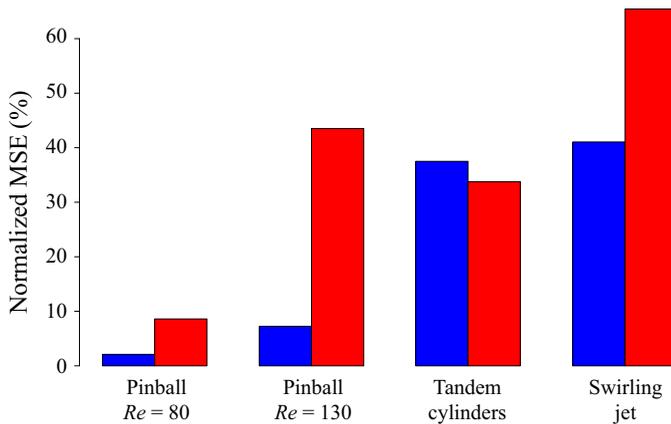


Figure 9. Reconstruction Error of Isomap versus POD: blue, Isomap; red, POD.

in all the cases except for the tandem cylinders. In this case, the performance of the two methods is similar, due to the rather small size of the input experimental dataset and to the sparsity of the manifold leading to large errors for isolated points, as expectable from figure 7(b).

This argument is further confirmed by performing a parametric study reducing the size of the dataset for the pinball test case at $Re = 80$. Reducing the dataset size to the 1/2,

1/4 and 1/8 of the original data set size leads to an increasing NMSE equal to 10 %, 15 % and 43 %.

5. Conclusions

In this paper we have developed an encoder–decoder framework based on manifold learning techniques to tackle the problem of understanding shear flows. The proposed manifold learner is Isomap and it is coupled with a KNN-based decoder. We show that flows which can be described with a limited set of coherent structures are suitable candidates for manifold learning. In the applications proposed in this manuscript we have chosen phenomena whose snapshots correctly sampled all the events defining the system ‘clock’, i.e. jet and wake flows in which the measurement domain samples the evolution of the main vortical features. We have shown that Isomap correctly identifies the true dimensionality of a given dataset and that the manifolds unravelled using the Isomap encoder are representative of meaningful physical quantities and are suitable for a reduced-order modelling of shear flows. The pure physics-uninformed results in the fluidic pinball case also have a correlation with physical properties like vortex-shedding phases or the force coefficients and open ground to use this technique to model wake flows and design flow-control systems.

We have shown that when handling experimental cases with complex behaviour, despite the presence of acceptable measurement noise, the new encoder–decoder tool outperforms the classical dimensionality-reduction techniques like POD in terms of clarity and interpretability of the identified manifolds, and it is less sensitive to experimental noise. In such cases, not only the identified manifolds are more reliable, but they also distil some physical information that POD is not able to catch, including the coexistence of the two shedding regimes and transition between them in the case of the wake of two tandem cylinders.

Finally, the developed decoder proved to have outstanding capabilities in reconstructing the original flow fields from the identified manifold, allowing for the instantaneous identification of the flow state with applications to closed-loop flow control.

The proposed manifold learner may significantly reduce the dimension of the state space as compared with POD expansions with similar representation error, particularly for transient flows. Hence, the autoencoder methodology may be used for full-state estimation and control design. For both tasks, every unnecessary coordinate acts as a noise amplifier and constitutes a danger for the system to get ‘off track’. The authors actively pursue this direction.

Funding. This work has been supported by the Madrid Government (Comunidad de Madrid) under the Multiannual Agreement with Universidad Carlos III de Madrid in the line of ‘Fostering Young Doctors Research’ (PITUFLOW-CM-UC3M), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation). This work has also been partially supported by the project ARTURO, ref. PID2019-109717RBI00/AEI/10.13039/501100011033, funded by the Spanish State Research Agency. B.R.N. and N.D. acknowledge funding by the National Science Foundation of China (NSFC) through grants 12172109 and 12172111 and 12202121, by the Guandong province, China, via the Natural Science and Engineering grant 2022A1515011492, by the Shenzhen Research Foundation for Basic Research, China, via grant JCYJ20220531095605012, and HangHua company (Dalian, China) for their scientific support. The authors warmly thank Dr F. Lückoff and Dr M. Raiola for providing the swirling jet and the tandem cylinder data sets. Funding for APC: Universidad Carlos III de Madrid (Read and Publish Carlos III University of Madrid).

Data availability statement. The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

Author ORCID*s*.

- ① E. Farzamnik <https://orcid.org/0000-0002-5411-8448>;
- ① A. Ianiro <https://orcid.org/0000-0001-7342-4814>;
- ① S. Discetti <https://orcid.org/0000-0001-9025-1505>;
- ① N. Deng <https://orcid.org/0000-0001-6847-2352>;
- ① K. Oberleithner <https://orcid.org/0000-0003-0964-872X>;
- ① B.R. Noack <https://orcid.org/0000-0001-5935-1962>;
- ① V. Guerrero <https://orcid.org/0000-0002-6610-7455>.

Appendix A. Selection of the number of neighbours k in Isomap

As discussed in § 2, selecting a proper number of neighbours k to construct the neighbouring graph \mathbf{G} in Isomap is a crucial decision. The unknown geodesic distances between every pair of observations in the high-dimensional manifold are approximated by the shortest paths distances between the corresponding nodes in the graph \mathbf{G} , which depends on k . On the one hand, a too small k may cause a splitting of the manifold into disjoint ones and thus losing its real structure. On the other hand, if k is too large then points which are far according to the real geodesic distance may become close by using its approximation by their shortest path in \mathbf{G} due to an undue number of connections (edges) and/or the existence of holes in the manifold. This phenomenon is known as short circuiting.

Different approaches have been proposed in the literature to cope with the selection of k in Isomap. In this work, we follow the methodology presented by Samko, Marshall & Rosin (2006) to determine a valid range of values $[k_{min}, k_{max}]$ to perform the search. The lower bound of the interval, k_{min} , is selected as the smallest k so that the neighbouring graph \mathbf{G} is connected. Regarding the upper bound, k_{max} , the largest value of k is picked so that the following equation holds:

$$\frac{2E}{N} \leq k + 2, \tag{A1}$$

where E is the number of edges and N is the number of nodes in \mathbf{G} . Once the valid range of k is found, Samko *et al.* (2006) propose to pick $k \in [k_{min}, k_{max}]$ so that the residual variance is minimum.

In all our studies in § 4 using any value in the valid range selected this way results in a low residual variance. We refer the reader to figure 10 for an illustration of the impact of the choice of k on the residual variance for the pinball dataset for $Re = 130$.

Appendix B. Different approaches to define the residual variance

There exist numerous approaches to assess the fits provided by different dimensionality reduction techniques. Regarding Isomap, the so-called residual variance as stated in (2.2) is the common choice.

Regarding POD, the classical definition of residual to assess its performance is given by

$$1 - \frac{\sum_{j=1}^p \text{var}(\mathbf{a}_j)}{\sum_{j=1}^P \text{var}(\mathbf{a}_j)}, \tag{B1}$$

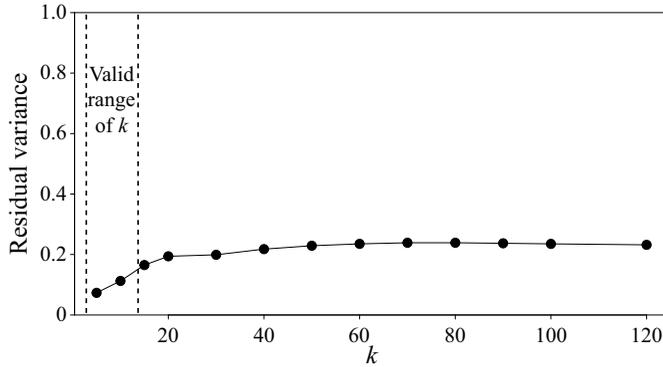


Figure 10. Residual variance of Isomap embedding to three-dimensional space versus different Isomap k , for pinball dataset $Re = 130$ with the valid range of Isomap's k according to Samko *et al.* (2006).

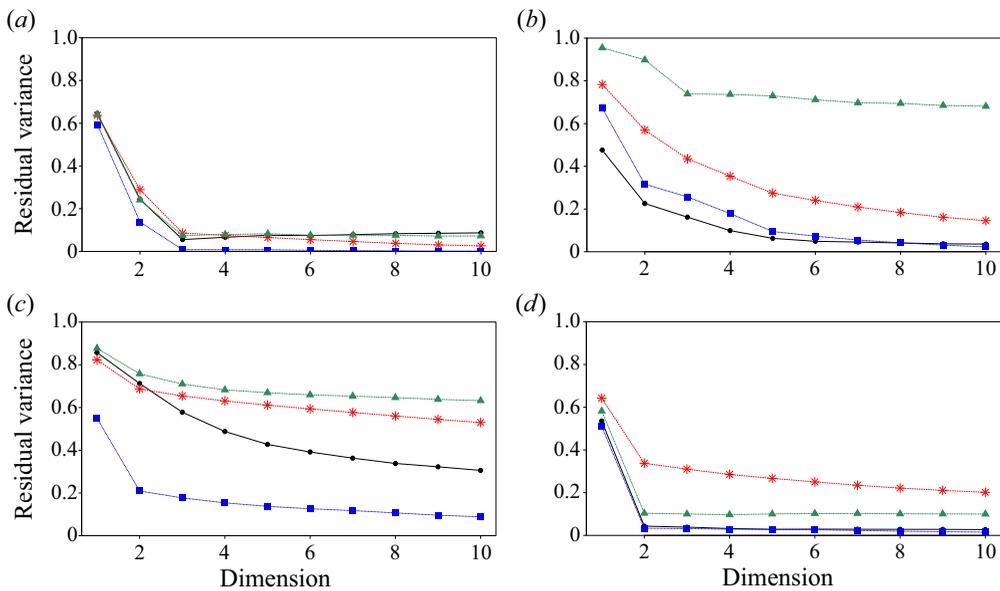


Figure 11. Different methods for defining residual variance: (a) pinball, $Re = 80$; (b) pinball, $Re = 130$; (c) swirling jet; (d) tandem cylinders. Colour codes: black, (2.2); red, POD (B1); blue, POD (B2); green, POD (B3).

where ‘var’ stands for the variance and a_j are the mode amplitudes, $j = 1, \dots, P$.

The ways of measuring the fits in Isomap and POD given by (2.2) and (B1) are not comparable. In order to evaluate Isomap and POD up to the same standard, Tenenbaum *et al.* (2000) proposed to replace the geodesic distances approximated by D_G in (2.2) by the pairwise Euclidean distances in the input high-dimensional space, D_X , where the element in row i and column j in D_X is $\|x_i - x_j\|_2$, $i, j = 1, \dots, n$. Then a related definition of residual variance for POD is

$$1 - R^2(\text{vec}(D_X), \text{vec}(D_Z)), \tag{B2}$$

where \mathbf{D}_Z is the matrix of Euclidean distances considering the retained principal components and R^2 refers to the squared correlation coefficient.

Nevertheless, the definition of residual variance in (B2) does not capture the ability of POD of reproducing the geodesic distances in the high-dimensional space. To do so, a different but related proposal to measure the residual variance for POD combines (2.2) and (B2) as

$$1 - R^2(\text{vec}(\mathbf{D}_G), \text{vec}(\mathbf{D}_Z)). \quad (\text{B3})$$

Finally, we point out that the residual variance as defined in (2.2), (B2) and (B3) may not decrease strictly when the number of dimensions increases. In other words, the correlation between a distance matrix obtained from a set of points in \mathbb{R}^d , \mathbf{D}_d , and another distance matrix, \mathbf{D} , does not necessarily decrease if the distance between the embedded points in \mathbb{R}^{d-1} by dropping one of the dimensions, \mathbf{D}_{d-1} , is considered instead.

Figure 11 shows the values for the different definitions of residual variances presented in this work in each of the case studies in § 4 for different numbers of dimensions. The black line corresponds to (2.2), the red one is (B1), the blue is (B2) and the green is (B3). For all the dimensions considered, the residual variance of Isomap surpasses POD using definitions (2.2) and (B2).

REFERENCES

- ALAM, M.M., ELHIMER, M., WANG, L., LO JACONO, D. & WONG, C.W. 2018 Vortex shedding from tandem cylinders. *Exp. Fluids* **59** (3), 60.
- ANDRONOV, A., LEONTOVICH, E., GORDON, I. & MAIER, A. 1971 *Theory of Bifurcations of Dynamical Systems on a Plane*. Israel Program for Scientific Translations.
- AUBRY, N., HOLMES, P., LUMLEY, J.L. & STONE, E. 1988 The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.* **192**, 115–173.
- BANSAL, G., MASCARENHAS, A.A. & CHEN, J.H. 2011 Identification of intrinsic low dimensional manifolds in turbulent combustion using an Isomap based technique. *Tech. Rep.* No. SAND2011-0709C Sandia National Lab. (SNL-CA), Livermore, CA, USA.
- BARTENHAGEN, C. 2021 RDRToolbox: a package for nonlinear dimension reduction with Isomap and LLEn. R package version 1.42.0.
- BENNER, P., GRIVET-TALOCIA, S., QUARTERONI, A., ROZZA, G., SCHILDERS, W. & SILVEIRA, L.M. 2021 *Model Order Reduction: Volume 1: System-and Data-Driven Methods and Algorithms*. De Gruyter.
- BENNER, P., SCHILDERS, W., GRIVET-TALOCIA, S., QUARTERONI, A., ROZZA, G. & SILVEIRA, L.M. 2020a *Model Order Reduction: Volume 2: Snapshot-Based Methods and Algorithms*. De Gruyter.
- BENNER, P., SCHILDERS, W., GRIVET-TALOCIA, S., QUARTERONI, A., ROZZA, G. & SILVEIRA, L.M. 2020b *Model Order Reduction: Volume 3: Applications*. De Gruyter.
- BERKOOZ, G., HOLMES, P. & LUMLEY, J.L. 1993 The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **25** (1), 539–575.
- BOURGOIS, J.A., MARTINUZZI, R.J. & NOACK, B.R. 2013 Generalised phase average with applications to sensor-based flow estimation of the wall-mounted square cylinder wake. *J. Fluid Mech.* **736**, 316–350.
- BREZIS, H. 2011 *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer.
- BRUNTON, S.L., BUDISIC, M., KAISER, E. & KUTZ, J.N. 2022 Modern Koopman theory for dynamical systems. *SIAM Rev.* **64** (2), 229–340.
- BRUNTON, S.L. & NOACK, B.R. 2015 Closed-loop turbulence control: progress and challenges. *Appl. Mech. Rev.* **67** (5), 050801.
- CALLAHAM, J.L., BRUNTON, S.L. & LOISEAU, J.-C. 2022 On the role of nonlinear correlations in reduced-order modelling. *J. Fluid Mech.* **938**, A1.
- COIFMAN, R.R. & LAFON, S. 2006 Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** (1), 5–30.
- CORNEJO MACEDA, G.Y., LI, Y., LUSSEYRAN, F., MORZYŃSKI, M.K. & NOACK, B.R. 2021 Stabilization of the fluidic pinball with gradient-enriched machine learning control. *J. Fluid Mech.* **917**, A42.
- DENG, N., NOACK, B.R., MORZYŃSKI, M. & PASTUR, L.R. 2022 Cluster-based hierarchical network model of the fluidic pinball – cartographing transient and post-transient, multi-frequency, multi-attractor behaviour. *J. Fluid Mech.* **934**, A24.

- DENG, N., NOACK, B.R., MORZYŃSKI, M. & PASTUR, L.R. 2020 Low-order model for successive bifurcations of the fluidic pinball. *J. Fluid Mech.* **884**, A37.
- FLOYD, R.W. 1962 Algorithm 97: shortest path. *Commun. ACM* **5** (6), 345.
- FOROOZAN, F., GUERRERO, V., IANIRO, A. & DISCETTI, S. 2021 Unsupervised modelling of a transitional boundary layer. *J. Fluid Mech.* **929**, A3.
- FRANKLIN, J. 2005 The elements of statistical learning: data mining, inference and prediction. *Math. Intell.* **27** (2), 83–85.
- FRANZ, T., ZIMMERMANN, R., GÖRTZ, S. & KARCHER, N. 2014 Interpolation-based reduced-order modelling for steady transonic flows via manifold learning. *Intl J. Comput. Fluid Dyn.* **28**, 106–121.
- GIANNAKIS, D., KOLCHINSKAYA, A., KRASNOV, D. & SCHUMACHER, J. 2018 Koopman analysis of the long-term evolution in a turbulent convection cell. *J. Fluid Mech.* **847**, 735–767.
- GORBAN, A.N. & KARLIN, I.V. 2005 *Invariant Manifolds for Physical and Chemical Kinetics*. Springer.
- KAISER, E., NOACK, B.R., CORDIER, L., SPOHN, A., SEGOND, M., ABEL, M., DAVILLER, G., ÖSTH, J., KRAJNOVIĆ, S. & NIVEN, R.K. 2014 Cluster-based reduced-order modelling of a mixing layer. *J. Fluid Mech.* **754**, 365–414.
- KAUFMAN, L. & ROUSSEEUW, P.J. 1990 *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- KRUSKAL, J.B. 1964 Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29** (2), 115–129.
- KUZNETSOV, Y.A. & SACKER, R.J. 2008 Neimark–Sacker bifurcation. *Scholarpedia* **3** (5), 1845.
- LÜCKOFF, F., SIEBER, M., PASCHEREIT, C.O. & OBERLEITHNER, K. 2017 Characterization of different actuator designs for the control of the precessing vortex core in a swirl-stabilized combustor. *Trans. ASME J. Engng Gas Turbines Power* **140** (4), 041503.
- LEE, K. & CARLBERG, K.T. 2020 Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *J. Comput. Phys.* **404**, 108973.
- LUCHTENBURG, D.M., GÜNTER, B., NOACK, B.R., KING, R. & TADMOR, G. 2009 A generalized mean-field model of the natural and actuated flows around a high-lift configuration. *J. Fluid Mech.* **623**, 283–316.
- LÜCKOFF, F., KAISER, T.L., PASCHEREIT, C.O. & OBERLEITHNER, K. 2021 Mean field coupling mechanisms explaining the impact of the precessing vortex core on the flame transfer function. *Combust. Flame* **223**, 254–266.
- MARUSIC, I. & BROOMHALL, S. 2021 Leonardo da Vinci and fluid mechanics. *Annu. Rev. Fluid Mech.* **53** (1), 1–25.
- MENDEZ, M.A., RAIOLA, M., MASULLO, A., DISCETTI, S., IANIRO, A., THEUNISSEN, R. & BUCHLIN, J.-M. 2017 Pod-based background removal for particle image velocimetry. *Exp. Therm. Fluid Sci.* **80**, 181–192.
- MEZIC, I. 2013 Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45** (1), 357–378.
- MINH, H.Q. & MURINO, V. 2017 Covariances in computer vision and machine learning. *Synth. Lect. Comput. Vis.* **7** (4), 1–170.
- MORTON, J., JAMESON, A., KOCHENDERFER, M.J. & WITHERDEN, F. 2018 Deep dynamical modeling and control of unsteady fluid flows. *Adv. Neural Inf. Process. Syst.* **31**.
- MÜLLER, J.S., LÜCKOFF, F., PAREDES, P., THEOFILIS, V. & OBERLEITHNER, K. 2020 Receptivity of the turbulent precessing vortex core: synchronization experiments and global adjoint linear stability analysis. *J. Fluid Mech.* **888**, A3.
- NOACK, B.R., AFANASIEV, K., MORZYŃSKI, M., TADMOR, G. & THIELE, F. 2003 A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.* **497**, 335–363.
- NOACK, B.R. & MORZYŃSKI, M. 2017 The fluidic pinball – a toolkit for multiple-input multiple-output flow control (version 1.0). *Tech. Rep.* 02/2016 Chair of Virtual Engineering, Institute of Combustion Engines and Transport, Poznan University of Technology, Poland.
- OTTO, S.E. & ROWLEY, C.W. 2021 Koopman operators for estimation and control of dynamical systems. *Annu. Rev. Control Rob. Autonom. Syst.* **4**, 59–87.
- OTTO, S.E. & ROWLEY, C.W. 2022 Inadequacy of linear methods for minimal sensor placement and feature selection in nonlinear systems: a new approach using secants. *J. Nonlinear Sci.* **32** (5), 69.
- R CORE TEAM 2020 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- RAIBAUDO, C., ZHONG, P., NOACK, B.R. & MARTINUZZI, R.J. 2020 Machine learning strategies applied to the control of a fluidic pinball. *Phys. Fluids* **32** (1), 015108.

- RAIOLA, M., IANIRO, A. & DISCETTI, S. 2016 Wake of tandem cylinders near a wall. *Exp. Therm. Fluid Sci.* **78**, 354–369.
- ROWEIS, S.T. & SAUL, L.K. 2000 Nonlinear dimensionality reduction by locally linear embedding. *Science* **290** (5500), 2323–2326.
- SAMKO, O., MARSHALL, A.D. & ROSIN, P.L. 2006 Selection of the optimal parameter value for the Isomap algorithm. *Pattern Recognit. Lett.* **27** (9), 968–979.
- SCARANO, F. 2001 Iterative image deformation methods in PIV. *Meas. Sci. Technol.* **13** (1), R1.
- SORIA, J. 1996 An investigation of the near wake of a circular cylinder using a video-based digital cross-correlation particle image velocimetry technique. *Exp. Therm. Fluid Sci.* **12** (2), 221–233.
- STROGATZ, S., FRIEDMAN, M., MALLINCKRODT, A.J. & MCKAY, S. 1994 Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. *Comput. Physics* **8** (5), 532–532.
- SYRED, N. 2006 A review of oscillation mechanisms and the role of the precessing vortex core (PVC) in swirl combustion systems. *Prog. Energy Combust. Sci.* **32** (2), 93–161.
- TAURO, F., GRIMALDI, S. & PORFIRI, M. 2014 Unraveling flow patterns through nonlinear manifold learning. *PLoS One* **9** (3), e91131.
- TENENBAUM, J.B., DE SILVA, V. & LANGFORD, J.C. 2000 A global geometric framework for nonlinear dimensionality reduction. *Science* **290** (5500), 2319–2323.
- TORGERSON, W.S. 1952 Multidimensional scaling: I. Theory and method. *Psychometrika* **17** (4), 401–419.
- WESTERWEEL, J. & SCARANO, F. 2005 Universal outlier detection for PIV data. *Exp. Fluids* **39** (6), 1096–1100.
- ZDRAVKOVICH, M.M. 1997 *Flow Around Circular Cylinders: Volume 1: Fundamentals*. Oxford University Press.