




Dynamics of a data-driven low-dimensional model of turbulent minimal pipe flow

C. Ricardo Constante-Amores¹ , Alec J. Linot²  and Michael D. Graham³ 

¹Department of Mechanical Science and Engineering, Grainger College of Engineering, University of Illinois, Urbana Champaign, IL 61801, USA

²Department of Chemical Engineering, University of Massachusetts Amherst, Amherst, MA, USA

³Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095, USA

Corresponding author: Cristian Ricardo Constante Amores, crconsta@illinois.edu

(Received 6 August 2024; revised 8 September 2025; accepted 9 September 2025)

The simulation of turbulent flow requires many degrees of freedom to resolve all the relevant time and length scales. However, due to the dissipative nature of the Navier–Stokes equations, the long-term dynamics is expected to lie on a finite-dimensional invariant manifold with fewer degrees of freedom. In this study, we build low-dimensional data-driven models of pressure-driven flow through a circular pipe. We impose the ‘shift-and-reflect’ symmetry to study the system in a minimal computational cell (e.g. the smallest domain size that sustains turbulence) at a Reynolds number of 2500. We build these models by using autoencoders to parametrise the manifold coordinates and neural ordinary differential equation to describe their time evolution. Direct numerical simulations (DNSs) typically require of the order of $\mathcal{O}(10^5)$ degrees of freedom, while our data-driven framework enables the construction of models with fewer than 20 degrees of freedom. Remarkably, these reduced-order models effectively capture crucial features of the flow, including the streak breakdown. In short-time tracking, these models accurately track the true trajectory for one Lyapunov time, as well as the leading Lyapunov exponent, while at long-times, they successfully capture key aspects of the dynamics such as Reynolds stresses and energy balance. The model can quantitatively capture key characteristics of the flow, including the streak breakdown and regeneration cycle. Additionally, we report new exact coherent states found in the DNS with the aid of these low-dimensional models. This approach leads to the discovery of seventeen previously

unknown solutions within the turbulent pipe flow system, notably featuring relative periodic orbits characterised by the longest reported periods for such flow conditions.

Key words: pipe flow boundary layer, transition to turbulence, machine learning

1. Introduction

The prevalence of wall turbulence in a diverse range of industrial and everyday applications has attracted substantial interest, as approximately 25 % of the energy consumed by industry is dedicated to transporting fluids through pipes and channels, and about one quarter of this energy is dissipated due to turbulence occurring near walls (Jiménez & Moin 1991; Jiménez 2013; Avila, Barkley & Hof 2023). Pipe flow has been the subject of extensive research since the groundbreaking experiments conducted by Reynolds nearly two centuries ago (Reynolds 1883). Flow regimes are solely governed by the Reynolds number, which characterises the ratio of inertial forces to viscous forces. Studies of the transition of pipe flow from laminar to turbulent have resulted in numerous comprehensive reviews (see e.g. Eckhardt *et al.* 2007, Mullin 2011, Smits, McKeon & Marusic 2011, Avila *et al.* 2023).

The focus of the present work is the extent to which modern data-driven methods can capture the nonlinear dynamics of turbulent pipe flows near transition (Darbyshire & Mullin 1995; Eckhardt 2009). Because of its geometric and dynamical simplicity, the ‘minimal flow unit’ (MFU) of pipe flow has been previously presented (Willis *et al.* 2013, 2016; Budanur *et al.* 2017; Constante-Amores & Graham 2024; Kaszás & Haller 2024). The MFU represents the smallest domain size capable of sustaining turbulence, described by Jiménez & Moin (1991) in the context of plane Couette flow. It encapsulates the essential elements of the turbulent dynamics, particularly the ‘self-sustaining process’ (SSP) described by Hamilton, Kim & Waleffe (1995). In this process, low-speed streaks near the wall become unstable and wavy, leading to their breakdown and the formation of rolls. These rolls then lift fluid from the walls, thereby regenerating the streaks and perpetuating the cycle.

To understand the intricate nonlinear dynamics of turbulence, researchers have adopted a dynamical systems perspective. The turbulent nonlinear dynamics of fluids is governed by the (infinite-dimensional) Navier–Stokes equations (NSE). Despite this infinite-dimensionality, the long-time dynamics is expected to lie on a finite-dimensional invariant manifold within state space (Hopf 1948; Temam 1989; Cvitanovic *et al.* 2005) (we discuss this point in more detail below). From this viewpoint, turbulence can be seen as a chaotic attractor of the NSE. Turbulent flows can display persistent patterns in space and time, commonly known as exact coherent states (ECSs) (Kawahara, Uhlmann & van Veen 2012; Graham & Floryan 2021). There are several ECS types: steady or equilibrium solutions, periodic orbits, travelling waves and relative periodic orbits. A trajectory on the attractor picks up characteristics of underlying unstable ECSs as it approaches them but is ultimately moved away along unstable manifolds. When many of these ECSs are characterised, they can be used to approximate the statistical properties of the attractor, such as work by Nagata (1990), Kawahara & Kida (2001), Chandler & Kerswell (2013) and Page *et al.* (2024). However, a fixed-point ECS cannot capture the dynamics entirely; periodic orbits can represent key aspects of the nonlinear turbulent dynamics, such as bursting behaviour (Cvitanović 2013). In the realm of pipe flow, early studies focused on ECS and their role in the transition to turbulence (Faisst &

Eckhardt 2003; Wedin & Kerswell 2004; Pringle & Kerswell 2007; Duguet, Willis & Kerswell 2008; Willis & Kerswell 2008; Pringle, Duguet & Kerswell 2009; Viswanath & Cvitanović 2009). The first set of ECSs discovered, presented by Faisst & Eckhardt (2003) and Wedin & Kerswell (2004), were travelling-wave solutions, which were also observed in experimentally (Hof *et al.* 2004; de Lozar *et al.* 2012). To date, several studies have focused on the discovery and classification of ECSs according to the value of their average dissipation, wave speed and spatial symmetry (Pringle *et al.* 2009; Willis, Cvitanović & Avila 2013; Budanur *et al.* 2017). Willis *et al.* (2013) and Willis, Short & Cvitanović (2016) reported 29 solutions and visualised these solutions in state space using symmetry reduction, showing the connections of relative periodic orbits and the turbulent attractor. Budanur *et al.* (2017) reported 49 new relative periodic orbits and 10 travelling-wave solutions. Their findings further supported the view that turbulence wanders around ECSs.

However, identifying ECSs remains challenging due to the high-dimensionality of the state space. Traditional Newton–Raphson methods can be employed to locate those solutions, but more advanced techniques, such as the Jacobian-free Newton–Krylov method, are more effective because they avoid explicit computation and inversion of the Jacobian, which is expensive for high-dimensional systems. In the latter, the Jacobian matrix is not explicitly calculated, as detailed in Viswanath (2007). Good initial conditions are important due to non-convexity mostly and are also computationally expensive. Favourable initial conditions help narrow the search area and increase the likelihood of finding ECSs. Lan & Cvitanović (2004) proposed a variational method to find unstable periodic orbits in the Kuramoto–Sivashinsky equation, demonstrating that their method can converge to a broader set of solutions compared with traditional shooting methods.

A promising avenue for studying turbulence is the development of reduced-order models, which simplify the complex dynamics of turbulent flows while retaining essential features. Among the nonlinear approaches to model reduction, invariant-manifold-based frameworks, particularly spectral submanifold (SSM) methods, have emerged as powerful tools. The SSM facilitates the construction of invariant manifolds near known stationary points, to which the dynamics of a system can be reduced (Li, Jain & Haller 2022; Kogelbauer & Haller 2018). The SSMs represent the smoothest nonlinear extensions of the spectral subspaces of the linearised system near a stationary state, such as a fixed point or a periodic orbit. Recently, Kaszás & Haller (2024) employed SSM to successfully identify the invariant manifold capturing the edge of chaos in pipe flow. This manifold serves as a crucial boundary, demarcating the transition from the laminar state to turbulence within the phase space of the NSE. We note that SSM must be ‘anchored’ to a known stationary point. In contrast, the framework we adopt in this work does not rely on such anchoring, allowing for a more flexible exploration of the attractor.

The accurate simulation of MFU pipe flow requires a large state space to resolve all the relevant spatial and temporal scales. For instance, Willis *et al.* (2013) and Budanur *et al.* (2017) required of the order of $\mathcal{O}(10^5)$ degrees of freedom to capture the complex, nonlinear turbulent dynamics. Performing data-driven modelling in this full-state space is computationally challenging. However, due to the dissipative nature of the NSE, it is expected that viscosity attenuates the high wavenumber modes, confining the long-term dynamics to an invariant manifold with fewer degrees of freedom than the full-state dimension (Temam 1989; Zelik 2022). The exact dimension of this invariant manifold is not known beforehand and must be estimated from data. The most common method for linear dimension reduction is principal component analysis (PCA), also known as proper orthogonal decomposition (POD) in the fluid dynamics community. Principal component analysis works by projecting the state onto the set of orthogonal modes that captures

the maximum variance or energy in the data (Jolliffe 1986; Abdi & Williams 2010; Holmes *et al.* 2012). However, PCA assumes a flat manifold because it is an inherently linear technique, which makes it a poor approximation for complex nonlinear problems. To address this, nonlinear techniques for dimension reduction have emerged such as autoencoders. Autoencoders consist of a pair of neural networks in which one network maps from a high-dimensional space to a low-dimensional space, and the other maps back (Kramer 1991; Milano & Koumoutsakos 2002; Hinton & Salakhutdinov 2006). For very high-dimensional systems, it can be advantageous to perform an initial linear dimension reduction step with PCA, followed by further nonlinear dimension reduction using an autoencoder (Linot & Graham 2023; Young *et al.* 2023; Constante-Amores *et al.* 2024a). Additionally, combining PCA and an autoencoder in parallel allows for capturing both linear and nonlinear features of the data, with the autoencoder refining the representation beyond what PCA alone can provide (Linot & Graham 2020).

Once we have a low-dimensional representation of the full state, we can proceed in data-driven modelling of the dynamics in manifold coordinates (i.e. the intrinsic variables that describe the key behaviour of the system in the low-dimensional representation). The goal is to learn a vector field that governs the evolution of the system in this low-dimensional representation. This approach has been successfully applied to chaotic systems, including the one-dimensional Kuramoto–Sivashinsky equation (Linot & Graham 2020; Liu, Axås & Haller 2024), two-dimensional Kolmogorov flow (Pérez-De-Jesús & Graham 2023; Constante-Amores *et al.* 2024b) and Couette flow (Kaszás *et al.* 2022; Linot & Graham 2023). Linot & Graham (2020) presented the framework known as DManD which stands for ‘data-driven manifold dynamics’. In DManD, an autoencoder finds a low-dimensional representation of the full state, and then a neural ordinary differential equation (ODE) (NODE) learns an evolution equation of this low-dimensional representation. The NODE is a neural network that parameterises the vector field of the latent space (e.g. low-dimensional coordinate representation found by the autoencoder) (Chen *et al.* 2019; Linot & Graham 2022). It is important to highlight that DManD is highly advantageous because, like the underlying turbulent systems, it is Markovian in nature (where predictions of the next state only depend on the current state) and a continuous-time formulation.

In this work, we address data-driven modelling for turbulent MFU pipe flow at Reynolds number of $Re = 2500$. We note that, while our approach shares methodological similarities with the recent work of Linot & Graham (2023) on Couette flow, specifically the use of POD, autoencoders and NODE, the focus of the present study is on pipe flow, which poses fundamentally different physical challenges. Unlike the planar, zero-mean shear profile of Couette flow, pipe flow features a non-zero, radially varying mean velocity and geometric curvature, resulting in a richer dynamics and more intricate turbulent structures. This work thus applies manifold-based data-driven modelling techniques to a more practically relevant and dynamically complex shear flow system. We show that the essential dynamics of pipe flow evolves on a low-dimensional manifold, enabling accurate reconstruction of both short-time trajectory evolution and long-time statistical properties. We compute the Lyapunov spectrum on the manifold and compare the leading Lyapunov exponent with that obtained from the direct numerical simulation (DNS). The good agreement indicates that the model successfully captures the dominant dynamics, suggesting that only a few degrees of freedom are required. In addition, we identify ECS in the latent space and successfully converged them in the DNS, leading to the discovery of previously unreported solutions in this flow configuration. We also acknowledge that, in Constante-Amores & Graham (2024), the authors constructed data-driven models using pipe flow data restricted to a single relative periodic orbit, whereas the present study focuses on learning a low-dimensional model from trajectories embedded within the full

attractor, leading to a much more general data-driven model (which is needed to discover new ECSs).

The rest of this paper is structured as follows. In § 2, we describe the framework for dimension reduction and time evolution. In § 3, we present the results that include the dimension reduction and the predictions of the DManD model for short- and long-time statistics, ECS identification and new ECSs found in the DNS using converged ECSs from the model as initial conditions. Finally, in § 4, we summarise the concluding results.

2. Framework

2.1. Dimension reduction

While the state space of a partial differential equation is formally infinite-dimensional, the NSE, which govern the motion of fluids, are dissipative in nature, and therefore solutions are expected to converge to a finite-dimensional invariant manifold, denoted as \mathcal{M} in this context (Hopf 1948; Foias, Manley & Temam 1988; Temam 1989; Zelik 2022). This manifold \mathcal{M} exhibits a local Euclidean structure, implying that each point within it possesses a nearby region that can be bi-directionally mapped to and from a Euclidean space, denoted as $\mathbb{R}^{d_{\mathcal{M}}}$, where $d_{\mathcal{M}}$ (with $d_{\mathcal{M}} \leq d_h$) represents the dimension of the manifold; in this work d_h is a higher dimension, in which the manifold can be embedded. This fact is also what allows for the global coordinate representation, since the dynamics is learned in d_h rather than $d_{\mathcal{M}}$. To effectively characterise this manifold and consequently understand the underlying system dynamics, only $d_{\mathcal{M}}$ independent coordinates are necessary, at least within local contexts. As \mathcal{M} remains unchanged by the system dynamics, the vector field which describes the dynamics on \mathcal{M} is always tangential to the manifold, resulting in a deterministic, memoryless dynamics confined to \mathcal{M} . Then, this dynamics is governed by an ordinary differential equation defined by this tangential vector field.

In this work, there are four distinct representations of the system state. Let \mathcal{H} denote the infinite-dimensional solution space of the NSE. The DNS produces trajectories in a finite-dimensional subspace $\mathbb{R}^d \subset \mathcal{H}$, which we refer to as the ‘full state’. This full state is projected onto a d_{POD} -dimensional subspace via POD, yielding a linear mapping $\mathbf{P} : \mathbb{R}^d \rightarrow \mathbb{R}^{d_{POD}}$. The POD reduced representation is then mapped to a d_h -dimensional coordinate system via a nonlinear mapping $\mathcal{E} : \mathbb{R}^{d_{POD}} \rightarrow \mathbb{R}^{d_h}$, obtained from a trained autoencoder.

We consider a system that is characterised by a deterministic, Markovian dynamics, so if $\mathbf{u} \in \mathbb{R}^d$ represents the full space state, then the dynamics can be represented by an ODE as

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(\mathbf{u}). \quad (2.1)$$

Here, \mathbf{u} represents the full-state space. In practice, \mathbf{u} is obtained from DNS. In this work, we find a mapping to a lower-dimensional representation

$$\mathbf{h} = \boldsymbol{\chi}(\mathbf{u}), \quad (2.2)$$

where $\mathbf{h} \in \mathbb{R}^{d_h}$ is the low-dimensional representation of the full-state space, along with an approximation of its inverse

$$\tilde{\mathbf{u}} = \tilde{\boldsymbol{\chi}}(\mathbf{h}), \quad (2.3)$$

so that the full-state space may be recovered (e.g. ideally $\mathbf{u} \approx \tilde{\mathbf{u}}$). In this work, we opt to parameterise $\boldsymbol{\chi}$, $\tilde{\boldsymbol{\chi}}$ with an autoencoder, referred to as a hybrid autoencoder in Linot & Graham (2020). This hybrid autoencoder is based on the idea of using neural networks to

learn the corrections from the leading POD coefficients

$$\mathbf{h} = \chi(\mathbf{u}; \theta_{\mathcal{E}}) = \mathbf{U}_{d_h}^T \mathbf{u} + \mathcal{E}(\mathbf{U}_{d_{POD}}^T \mathbf{u}, \theta_{\mathcal{E}}), \quad (2.4)$$

where $\mathbf{U}_k \in \mathbb{R}^{d \times k}$ corresponds to a matrix containing the first k POD modes ordered by variance, and \mathcal{E} corresponds to the encoder in the neural network (e.g. § 3.2.1 presents the framework for the linear reduction with POD). In this way, the first term ($\mathbf{U}_{d_h}^T$) is the projection onto the leading d_h POD modes, and the second term is the corrections provided by the neural network. The mapping back to the full space is given by

$$\tilde{\mathbf{u}} = \check{\chi}(\mathbf{h}; \theta_{\mathcal{E}}) = \mathbf{U}_{d_{POD}}([\mathbf{h}, 0]^T + \mathcal{D}(\mathbf{h}; \theta_{\mathcal{D}})). \quad (2.5)$$

Here, $[\mathbf{h}, 0]^T$ is the \mathbf{h} vector zero padded to the correct size, and \mathcal{D} is a neural network. The first term is the POD mapping back to the full space, and the second term is a neural network correction. The weight parameters $\theta_{\mathcal{E}}, \theta_{\mathcal{D}}$ are trained to minimise the loss

$$\begin{aligned} L = & \frac{1}{dK} \sum_{i=1}^K \|\mathbf{u}(t_i) - \check{\chi}(\chi(\mathbf{u}(t_i); \theta_{\mathcal{E}}); \theta_{\mathcal{D}})\|^2 \\ & + \frac{1}{d_h K} \sum_{i=1}^K \kappa \|\mathcal{E}(\mathbf{U}_r^T \mathbf{u}(t_i); \theta_{\mathcal{E}}) + \mathcal{D}_{d_h}(\mathbf{h}(t_i); \theta_{\mathcal{D}})\|^2. \end{aligned} \quad (2.6)$$

Here, the first term corresponds to the mean-squared error of the reconstruction $\tilde{\mathbf{u}}$, while the second term corresponds to a penalty term to enhance the accurate representation of the leading d_h POD coefficients (e.g. \mathcal{D}_{d_h} , denotes the leading d_h elements of the decoder output). Here, κ is a penalty term. This penalty does not directly reduce the magnitude of the encoder's correction; instead, it promotes its removal by the decoder. Throughout, the norm is defined as the L_2 norm, $\|\mathbf{q}\|^2$. The prefactor in front of each term accounts for averaging over the vector components and the batch size K . In § 3.2, we also use standard autoencoders, which can be seen as $\mathbf{h} = \chi(\mathbf{u}; \theta_{\mathcal{E}}) = \mathcal{E}(\mathbf{U}_{POD}^T \mathbf{u}, \theta_{\mathcal{E}})$ and $\tilde{\mathbf{u}} = \check{\chi}(\mathbf{h}; \theta_{\mathcal{E}}) = \mathbf{U}_{POD}^T \mathcal{D}(\mathbf{h}; \theta_{\mathcal{D}})$, to highlight the effectiveness of using the hybrid autoencoders to find an accurate representation of the manifold coordinates. We note that this hybrid autoencoder has been used successfully for the Kuramoto–Sivashinsky equation, chaotic Kolmogorov flow and MFU plane Couette flow (Linot & Graham 2020, 2023; Pérez-De-Jesús & Graham 2023).

To train both the hybrid and standard autoencoders, we use the Adam optimiser to minimise the loss function presented in (2.6), utilising the POD coefficients as inputs (as explained in § 3.2.1). The training process spans 500 epochs, and we incorporate a learning rate scheduler that reduces the learning rate from 10^{-3} to 10^{-4} after the initial 300 epochs. This adjustment is made based on our observation that no significant improvements in reconstruction error occur beyond this number of epochs. For the hybrid autoencoder approach, we set the hyperparameter $\kappa = 0.1$, while for the standard autoencoder, $\kappa = 0$ (indicating that this term is not included). All relevant details of the neural network architectures and their hyperparameters (e.g. number of layers, neurons per layer, activation functions) are summarised in table 1, to ensure reproducibility of the results. The specific network parameters were determined through a meticulous trial and error search, exploring variations in the network's architecture and activation functions. We remark that our goal was to achieve the lowest reconstruction error while avoiding excessive computational costs.

Function	Shape	Activation	Learning Rate
χ	1014/2500/1000/500/100/ d_h	ReLU/ReLU/ReLU/ReLU/lin	$[10^{-3}, 10^{-4}]$
$\tilde{\chi}$	$d_h/100/500/1000/2500/1014$	ReLU/ReLU/ReLU/ReLU/lin	$[10^{-3}, 10^{-4}]$
\mathbf{g}_{NN}	$d_h/250/250/250/250/d_h$	sig/sig/sig/sig/lin	$[10^{-2}, 10^{-3}]$

Table 1. Neural network architectures for the autoencoder and NODE. ‘Shape’ represents the dimension of each layer, ‘Activation’ refers to the types of activation functions used, where ‘rectified linear unit (ReLU)’, ‘sig’ and ‘lin’ stand for ReLU, sigmoid and linear activation functions, respectively. ‘Learning Rate’ represents the learning rate at various times during training.

2.2. Time evolution: neural ODEs

We use a stabilised neural ODE framework for state-space modelling in the latent space. Rather than (2.1), we use a slight modification

$$\frac{d\mathbf{h}}{dt} = \mathbf{g}(\mathbf{h}) - \mathbf{A}\mathbf{h}. \quad (2.7)$$

Here, $\mathbf{A} = \gamma \delta_{ij} \sigma_i(\mathbf{h})$, where $\sigma_i(\mathbf{h})$ stands for the standard deviation of the i th component of \mathbf{h} , γ is a fixed parameter and δ_{ij} is the Kronecker delta. This modification, with a small value of γ , has been found to stabilise the dynamics against spurious growth of fluctuations without compromising the accuracy of predictions (Linot & Graham 2023; Linot *et al.* 2023). Linot & Graham (2023) demonstrated the importance of this damping term in detail for MFU plane Couette flow.

Next, we approximate \mathbf{g} using a neural ODE. For training \mathbf{g} , we integrate (2.7) forward in time from t_i to $t_i + \tau$ resulting in the prediction

$$\tilde{\mathbf{h}}(t_i + \tau) = \mathbf{h}(t_i) + \int_{t_i}^{t_i + \tau} \mathbf{g}(\mathbf{h}(t); \theta_g) - \mathbf{A}\mathbf{h}(t) dt. \quad (2.8)$$

We determine the parameters θ_g by minimising the difference between the true state $\mathbf{h}(t_i + \tau)$ and the predicted state $\tilde{\mathbf{h}}(t_i + \tau)$, as

$$J = \frac{1}{d_h K} \sum_{i=1}^K \|\mathbf{h}(t_i + \tau) - \tilde{\mathbf{h}}(t_i + \tau)\|^2. \quad (2.9)$$

To calculate the derivatives of \mathbf{g} with respect to the neural networks parameters θ_g , we make use of automatic differentiation. We train the stabilised NODE to predict the system evolution over one time unit, using data from which the temporal mean has been subtracted, by employing the Adam optimiser in PyTorch to minimise the loss function described in (2.9) (Paszke 2019). The training process incorporates a learning rate scheduler, which decreases the learning rate at three evenly spaced intervals, continuing until the learning curve stabilises. The specific details of this neural network are provided in table 1. The architectural choices are empirical, determined through trial and error by varying the number of nodes and layers.

Once \mathbf{g} is determined, an arbitrary initial condition can be mapped into the low-dimensional coordinates with χ . Then, the state evolution of \mathbf{h} at arbitrary points in time can be computed as the solution to (2.7), and finally the solution can be mapped back to the full space with $\tilde{\chi}$.

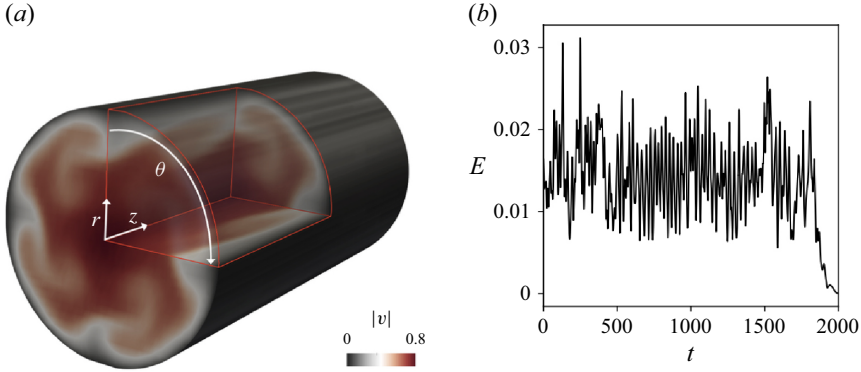


Figure 1. Schematic representation of the three-dimensional pipe flow system. Panel (a) shows a snapshot of the magnitude of the velocity field. For visualisation purposes, the entire pipe is shown, although calculations in this work is restricted in the shift-and-reflect symmetry subspace with $m_p = 4$ (whose boundaries are highlighted with solid red lines). Panel (b) represents the energy in the axially dependent modes only (k non-zero). This quantity decays rapidly after relaminarisations.

3. Results

In this section, we provide a detailed description of the dataset for MFU pipe flow, present the results of dimensionality reduction, evaluate the performance of the reduced models as we vary their dimension and introduce new ECSs. Figure 1(a) shows a three-dimensional representation of the MFU pipe configuration used in the current research. For visualisation purposes, the entire pipe is shown, although calculations in this work are confined to the shift-and-reflect symmetry subspace (which is the highlighted area with opacity).

3.1. Description of pipe flow data

We perform DNS of an incompressible viscous fluid moving inside of a pipe with a circular cross-section. We consider flow with a constant max flux, thus, the dimensionless forms of the NSE are expressed as

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} + \mathbf{U} \cdot \nabla \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{U} + \mathbf{v} \cdot \nabla \mathbf{v} &= -\nabla p + 32 \frac{\beta}{Re} \mathbf{z} + \frac{1}{Re} \nabla^2 \mathbf{v}, \\ \nabla \cdot \mathbf{v} &= 0. \end{aligned} \quad (3.1)$$

The equations are solved in cylindrical coordinates (r, θ, z) which refer to the radial, azimuthal and streamwise (axial) directions, respectively. Here, \mathbf{v} and p stand for the velocity and the pressure, respectively. The Reynolds number Re is defined as $Re = U \mathcal{D} / \nu$, where U , \mathcal{D} and ν are the mean flow velocity, the pipe diameter and the kinematic viscosity, respectively. Lengths and velocities are made non-dimensional using \mathcal{D} and U as characteristic values, and hence, time will be made non-dimensional using \mathcal{D} / U . The velocity $\mathbf{v} = (v_r, v_\theta, v_z)$ represents the deviation from laminar Hagen–Poiseuille flow equilibrium $\mathbf{U}(r) = 2(1 - (2r)^2)\mathbf{z}$. To maintain constant mass flux, a pressure gradient is required, and the excess pressure needed is measured by the feedback variable $\beta = \beta(\mathbf{v})$; thus the total dimensionless pressure gradient is $(1 + \beta)(32/Re)$, and $\beta = 0$ for laminar flow.

In the NSE, symmetries appear in the form of continuous and discrete symmetry groups. For the former, the cylindrical wall in pipe flow limits rotational symmetry around the z -axis and restricts translational movement along it. Let $g(\phi, \ell)$ represent the shift operator, where $g(\phi, 0)$ signifies an azimuthal rotation by ϕ about the pipe axis, and $g(0, \ell)$

indicates the streamwise translation by ℓ . Let σ represent the reflection about the $\theta = 0$ azimuthal angle. Thus

$$\left. \begin{aligned} g(\phi, \ell)[v_r, v_\theta, v_z, p](r, \theta, z) &= [v_r, v_\theta, v_z, p](r, \theta - \phi, z - \ell) \\ \sigma[v_r, v_\theta, v_z, p](r, \theta, z) &= [v_r, -v_\theta, v_z, p](r, -\theta, z) \end{aligned} \right\}. \quad (3.2)$$

Apart from azimuthal reflection, the NSE also have additional discrete symmetries in both azimuthal and streamwise directions across the computational cell Ω . The symmetry group of streamwise periodic pipe flow is $SO(2)_z \times O(2)_\theta$. In this paper, we restrict the dynamics to the 'shift-and-reflect' symmetry subspace

$$S = \{e, \sigma_\theta g_s\}, \quad (3.3)$$

where g_s represents a streamwise shift by $L/2$, i.e. flow fields of (3.1) that satisfy

$$[v_r, v_\theta, v_z, p](r, \theta, z) = [v_r, -v_\theta, v_z, p](r, -\theta, z - L/2). \quad (3.4)$$

This symmetry couples the streamwise translations with the azimuthal reflection. By imposing the shift-and-reflect symmetry, we eliminate the continuous phase along the azimuthal rotations. In this work, we do not factor out the continuous symmetry in the streamwise direction. Factoring out the streamwise symmetry reduces the manifold dimension by only one degree of freedom, and our focus is on developing a general framework that does not rely on symmetry reduction.

To perform the DNS of the incompressible turbulent pipe flow under the assumption of shift-and-reflect symmetry, we use the pseudo-spectral code Openpipeflow (Willis 2017). Fourier discretisation is used for the periodic axial (z) and azimuthal (θ) directions, with K and M representing the number of Fourier modes, i.e.

$$\{v\}(r, \theta, z, t) = \sum_{m=-M}^M \sum_{k=0}^K B_{mk}(r, t) e^{im_p m \theta} e^{ik\alpha z}, \quad (3.5)$$

where $B_{mk}(r, t)$ represents a three vector of Fourier amplitudes, m_p is a parameter to control the azimuthal shift-and-reflect subspace we work in and $\alpha = 2\pi/L$ is a parameter that controls the length of the pipe. In the radial direction, a Chebyshev grid is used that clusters points near the wall to effectively resolve the velocity gradients. No-slip and no-penetration boundary conditions are enforced at the wall. For a more detailed description of the numerical method, we refer to Willis (2017).

The simulation of the entire cross-sectional pipe ($m_p = 1$) presents a naturally periodic azimuthal boundary condition, while other values of m_p result in v repeating itself in the azimuthal direction. In this work, we construct models for MFU pipe flow at $Re = 2500$, with $m_p = 4$ ('shift-and-reflect' invariant subspace) and $\alpha = 1.7$, as in previous work by Willis *et al.* (2016) and Budanur *et al.* (2017). Then, the size of the computational cell is described by

$$\Omega = [1/2, 2\pi/m_p, \pi/\alpha] \equiv (r, \theta, z) \in [0, 1/2] \times [0, 2\pi/m_p] \times [0, \pi/\alpha]. \quad (3.6)$$

Thus, the domain size in wall units for the wall-normal, azimuthal and streamwise dimensions is $\Omega^+ \approx [100, 160, 370]$, respectively, which compares well with the MFUs for Couette flow and plane Poiseuille flow (i.e. $\Omega^+ \sim [68, 128, 190]$ and $\Omega^+ \approx [40, 100, 250-300]$, respectively). This domain size is similar to that used in the minimal box simulation by Jiménez & Moin (1991) and Willis *et al.* (2013), and it is sufficiently large to exhibit complex chaotic behaviour.

Data were generated with $\delta t = 0.01$ on a grid $(N_r, M, K) = (64, 10, 14)$. To eliminate aliasing errors in the evaluation of nonlinear terms, the 3/2 rule is applied. This rule

increases the number of grid points in each periodic direction by a factor of $3/2$, after converting the number of complex modes to the corresponding number of real physical grid points. Since each complex Fourier mode requires two real degrees of freedom in physical space, the number of grid points in each direction becomes $N_\theta = 2M$ and $N_z = 2K$. Therefore, the velocity field is evaluated on a $64 \times 30 \times 42$ grid in physical space, with three velocity components. The total number of degrees of freedom is $d = N_r \times N_\theta \times N_z \times 3 = 241,920$, so $\mathbf{u} \in \mathbb{R}^{241,920}$. In this grid size, $(\Delta\theta D/2)^+ \approx 5.3$ and $\Delta z^+ \approx 8.8$ (which is consistent with grid sizes used by Jiménez & Moin 1991). The resolution was tested to ensure mesh-independent results, confirmed by a drop in the energy spectra by at least 4 orders of magnitude.

We initiated simulations from random divergence-free initial conditions. The solutions were evolved forward in time for either 10 000 time units or until relaminarisation occurred. The initial 100 time units were excluded as transient data, and the final 200 time units were excluded if relaminarisation had occurred. Figure 1(b) shows the energy in the axially dependent modes only (i.e. modes with non-zero axial wavenumber k), which decays rapidly after relaminarisation, indicating the end of the simulation. Relaminarisation is identified by monitoring this energy, and following (Willis 2017), the simulation is terminated once it falls below 10^{-5} . This process was repeated with new initial conditions until we accumulated 96 921 time units of data, sampled at one time unit intervals (i.e. $\tau = 1$). Consequently, all data lie on the attractor. We divided this dataset, allocating 80 % for training and 20 % for testing purposes.

In terms of energy balance, the intermittent nature of relaminarisation results in that the energy balance does not necessarily hold true, especially when patching together trajectories from different simulations. While relaminarisation events temporarily disrupt this balance, the entire dataset is collected from regions where the flow remains on the attractor, where energy input and dissipation balance should hold. Thus, although the system is not strictly stationary due to relaminarisation, all of the data ultimately represent the dynamics within the chaotic saddle. This approach, while not ideal for strictly steady-state analysis, provides a robust basis for exploring the turbulent dynamics across a variety of conditions. Future work to avoid this problem would either increase the Reynolds number or simulate the full pipe without symmetry restrictions.

3.2. Learning of manifold coordinates

In this section, we present our approach to dimensionality reduction. We first apply linear reduction using POD, and then proceed with nonlinear reduction using autoencoders.

3.2.1. Linear dimension reduction with POD: from $\mathcal{O}(10^5)$ to $\mathcal{O}(10^3)$

The first step in constructing the low-dimensional model is to apply POD on the original dataset as a preprocessing step. This aims to reduce the dimension of the problem from approximately $\mathcal{O}(10^5)$ degrees of freedom to $\mathcal{O}(10^3)$, while preserving the essential characteristics of the turbulent flow system. The POD tries to find the function Φ that maximises

$$\frac{\langle |(\mathbf{v}', \Phi)_E|^2 \rangle}{\|\Phi\|_E^2}, \quad (3.7)$$

where $\mathbf{v}'(\mathbf{x}) = \mathbf{v}(\mathbf{x}) - \bar{\mathbf{v}}(\mathbf{x})$ is the fluctuating component of the velocity field, and $\bar{\mathbf{v}}$ is the mean velocity, obtained by averaging over both space and time, $\langle \cdot \rangle$ is the ensemble average

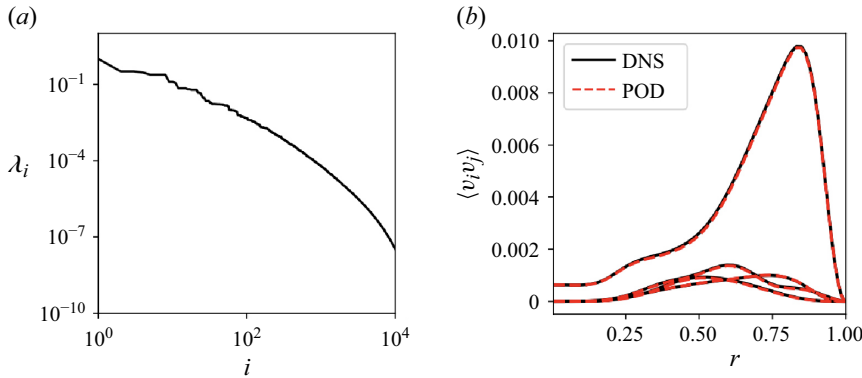


Figure 2. (a) Eigenvalues of the POD modes sorted in descending order. (b) Reconstruction of four components of the Reynolds stresses from the DNS and the data projected onto 512 POD modes. The curves correspond to $\langle u_z^2 \rangle$, $\langle u_\theta^2 \rangle$, $\langle u_r u_z \rangle$ and $\langle u_\theta^2 \rangle$, from top to bottom, respectively.

and the inner product is defined to be

$$(\mathbf{q}_1, \mathbf{q}_2)_E = \int \int \int \mathbf{q}_1 \cdot \mathbf{q}_2 \mathrm{d}\mathbf{x}, \quad (3.8)$$

with the corresponding energy norm $\|\mathbf{q}\|_E^2 = (\mathbf{q}, \mathbf{q})_E$. Solutions $\Phi^{(n)}$ to this problem satisfy the eigenvalue problem

$$\sum_{j=1}^3 \int_0^L \int_0^{2\pi/m_p} \int_0^R \langle v'_i(\mathbf{x}, t) v'_j(\mathbf{x}', t) \rangle \Phi^{(n)}(\mathbf{x}') r' \mathrm{d}r' \mathrm{d}\theta' \mathrm{d}z' = \lambda_i \Phi_i^{(n)}(\mathbf{x}). \quad (3.9)$$

The eigenvalue problem described by (3.9) becomes $d \times d$. To reduce the computational cost of this eigenvalue problem, and preserve symmetries, we treat the POD modes as Fourier modes in both the azimuthal and streamwise directions. This approach has been previously applied by Duggeby *et al.* (2007) and Linot & Graham (2023) for turbulent pipe and plane Couette flow, respectively. Holmes *et al.* (2012) showed that for translation-invariant directions, Fourier modes are the optimal POD modes. Thus, the eigenvalue problem becomes

$$\sum_{j=1}^3 \int_0^R \langle \hat{v}'_i(r', k_\theta, k_z, t) \hat{v}'_j^*(r', k_\theta, k_z, t) \rangle \varphi_{jk_\theta k_z}^{(n)}(r') r' \mathrm{d}r' = \lambda_{k_\theta k_z}^{(n)} \varphi_{ik_\theta k_z}^{(n)}(r), \quad (3.10)$$

where $*$ denotes the complex conjugate. Thus, the eigenvalue problem is reduced from a $d \times d$ to a $3N_r \times 3N_r$ problem for each pair of wavenumbers (k_θ, k_z) in the Fourier coefficients. This analysis gives us POD modes represented by

$$\Phi_{k_\theta k_z}^{(n)}(r, \theta, z) = \varphi_{k_\theta k_z}^{(n)}(r) e^{ik_\theta \theta} e^{i2\pi k_z z/L}, \quad (3.11)$$

and eigenvalues $\lambda_{k_\theta k_z}^{(n)}$. The projection onto these modes results in complex values unless both k_θ and k_z are zero. We arrange the modes in descending order of their eigenvalues (λ_i) , and we select the leading 512 modes, resulting in a vector of POD coefficients $(\mathbf{a}(t))$. Most of these modes are characterised by being complex valued (i.e. they have 2 degrees of freedom), so projecting onto these modes results in a 1014-dimensional system, i.e. $d_{\text{POD}} = 1014$. In figure 2(a), we display the eigenvalues, revealing a rapid decline followed by a long tail that contributes minimally to the energy content. By dividing the sum of the

eigenvalues of the leading 512 modes by the total sum, we find that these modes account for 99.44 % of the energy. Illustratively, [figure 2\(b\)](#) displays the reconstruction of Reynolds stresses for the components $\langle v_z^2 \rangle$, $\langle v_\theta^2 \rangle$, $\langle v_r v_z \rangle$ and $\langle v_r^2 \rangle$, from top to bottom, respectively. This reconstruction is obtained using those 512 modes with 5000 snapshots. We observe an excellent agreement between the DNS and the flow field obtained after truncating to the leading POD modes.

3.2.2. Nonlinear dimension reduction with autoencoders: from $\mathcal{O}(10^3)$ to $\mathcal{O}(10^1)$

After projecting the data to the leading POD modes, and selectively retaining only the high-energy coefficients, our next step involves a nonlinear reduction of the data using autoencoders. As a crucial preprocessing step, we normalise the POD coefficients by subtracting the mean and dividing by the maximum standard deviation, rather than normalising each component by its own standard deviation. Without normalisation, the lower-order coefficients with larger magnitudes would dominate training, potentially causing instability and poor gradient updates. The mean subtracted corresponds to the time-averaged mean flow field from which the POD coefficients are derived, ensuring the data are centred before normalisation.

[Figure 3\(a\)](#) shows the relative error on the test data of the POD coefficients using standard and hybrid autoencoders when varying d_h from 10 to 20. We have also added the corresponding values from a linear projection onto an equivalent number of POD coefficients. The POD projection onto the leading (complex) coefficients can be expressed as $\mathbf{a} = \mathbf{U}_r^T \mathbf{u}$. For each latent dimension d_h two autoencoders are trained independently to reduce the effects of the inherently stochastic nature of neural network training, including random weight initialisation and mini-batch sampling during optimisation, which can lead to variability in performance across training runs. To mitigate this, multiple models are trained, and the one with the lowest validation error is selected to represent performance at that dimension. The same architectures are used for the standard and hybrid autoencoders (see [table 1](#)).

In [figure 3\(a\)](#), the nonlinear reduction leads to nearly one order of magnitude decrease in the value of mean-squared error (MSE) compared with its equivalent with POD for the same dimensionality. Notably, a small reduction in error is observed beyond a threshold of $d_h > 17$. For POD, the relative error appears to plateau beyond this point, indicating convergence to a low-dimensional representation. In contrast, the autoencoders exhibit a more gradual reduction in error, which resembles a power-law decay rather than a distinct plateau. This implies that, with dimensions as few as 17, the autoencoders provide a good coordinate transformation from the full space (e.g. as we will show in [§ 3.3](#)). In all considered cases, the hybrid autoencoder consistently produces slightly better results in terms of MSE compared with the standard autoencoders. In [figure 3\(b\)](#), we compare the performance of both autoencoders by plotting the mean-squared POD coefficient amplitudes for the test data, denoted as $\langle \|a_n\|^2 \rangle$, for the low-dimensional representation with $d_h = 20$ (see [figure 3\(b\)](#)). The hybrid autoencoder exhibits limitations in capturing the amplitude of higher-order coefficients beyond $a > 30$, while the standard autoencoder struggles to accurately represent coefficients beyond the first two. This discrepancy arises from the fact that the hybrid autoencoder prioritises the reconstruction of the leading d_h POD coefficients. [Figure 3\(c\)](#) illustrates the Reynolds stresses for both types of autoencoders with $d_h = 20$ for 5000 snapshots of the test data, showing that the hybrid autoencoder outperforms the standard autoencoder. Finally, [figure 3\(d\)](#) displays field snapshots in the $z - \theta$ plane ($r = 0.496$) at random times showing qualitatively that hybrid autoencoders with $d_h = 20$ can accurately reconstruct the data. This result agrees with

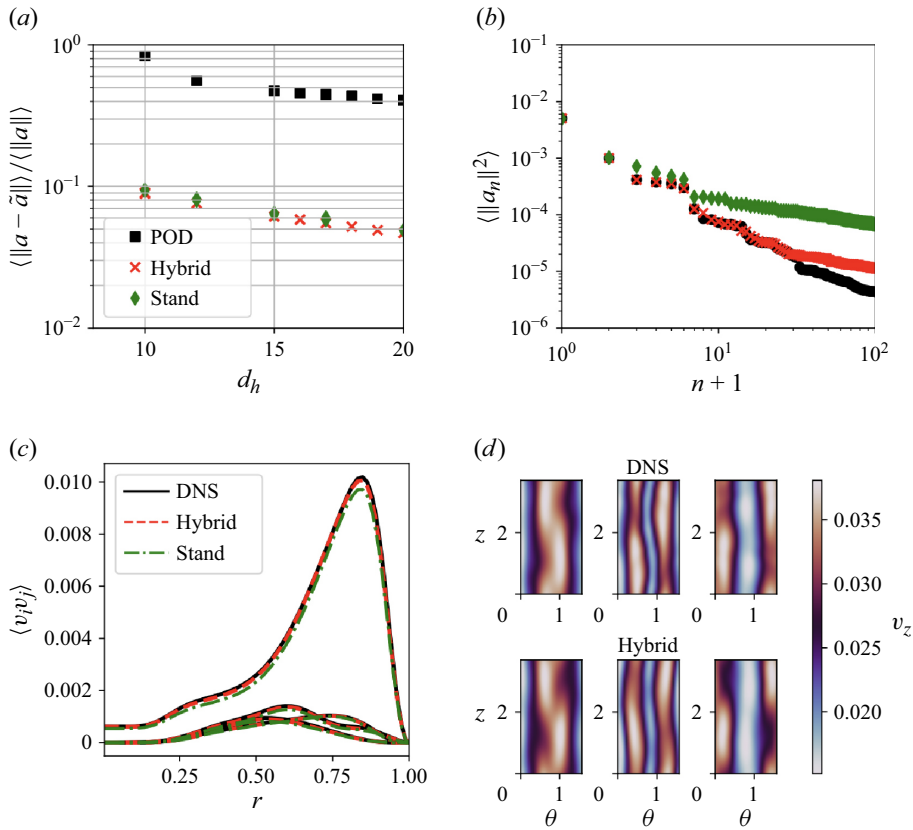


Figure 3. Nonlinear reduction with autoencoders: (a) relative error on test data for POD coefficients, standard and hybrid autoencoders as a function of the latent dimension d_h . For each dimension, results for two standard and two hybrid autoencoders are reported. (b) Reconstruction of $\langle \|a_n\|^2 \rangle$ (mean-squared POD coefficient amplitudes) for the test data from 512 POD modes and the standard and hybrid autoencoders at $d_h = 20$. (c) Components of the Reynolds stresses from the DNS and using autoencoders with $d_h = 20$. (d) Two-dimensional representation of the flow field in a $z - \theta$ plane ($r = 0.496$) with u_z for the DNS and reconstructed using the hybrid autoencoder at $d_h = 20$.

the findings from Kreilos & Eckhardt (2012), who showed in plane Couette flow that high-dimensional systems can live in low-dimensional manifolds.

Lastly, it is important to note that ideally the relative error value would plateau after determining the ‘right’ manifold dimension yet it is not always feasible, primarily due to computational limitations introduced during the training process. Our goal is to identify the optimal dimension for constructing DManD models that faithfully capture both short-time tracking and long-time statistical characteristics of turbulent pipe flow. In the upcoming section, we will systematically build models with varying latent dimension sizes (e.g. size of the low-dimensional space learned by the encoder).

We compared models using 512 and 1024 POD modes and found that increasing to 1024 resulted in less than 0.02 % improvement in prediction accuracy (with $d_h = 20$), despite a significant increase in training time. Since 512 modes already capture over 99.44 % of the total energy in comparison with 99.91 % for 1024 modes, we use 512 modes for efficiency without loss of accuracy.

Although a deep autoencoder could, in theory, learn a low-dimensional representation directly from high-dimensional input (e.g. $\mathcal{O}(10^5)$), training dense networks at this scale

is computationally expensive and prone to overfitting (Goodfellow *et al.* 2016). To address this, we first apply POD to reduce the input dimensionality while preserving the dominant flow structures, enabling efficient and stable training. Alternatively, convolutional neural networks (CNNs) offer a scalable solution that can learn directly from high-dimensional data by exploiting local structure, potentially eliminating the need for POD. While we focus on POD-based preprocessing here, CNN-based models are the alternative (Fukami, Fukagata & Taira 2019).

3.3. Modelling in manifold coordinates

Following the training of the autoencoders, a comprehensive exploration of the damping parameter γ is conducted to prevent the dynamics from drifting away from the attractor during modelling. This is motivated by the concept of an inertial manifold: in dissipative systems such as the NSE, the long-term dynamics is expected to collapse onto a lower-dimensional attracting set. In this spirit, the damping parameter acts as a regulariser that promotes stability in the latent dynamics by encouraging trajectories to remain close to the learned manifold. To find the optimal value of γ , trials are performed with $d_h = 20$, varying γ within the range $0.1 < \gamma < 0.5$. Empirical findings consistently point towards the fact that $\gamma = 0.25$ yields superior outcomes, with respect to the long-term dynamics of the system, as assessed by comparing the norm of the latent representation obtained from the autoencoder with that predicted by the NODE. We note that the significance of γ has been rigorously investigated by Linot & Graham (2023), in the context of MFU plane Couette flow, who showed that without the damping term almost all models become unstable for longer runs. The training objective of the NODE is focused on predicting one time unit ahead ($\tau = 1$), as described by (2.8). As a preprocessing step for training the NODE, we subtract the mean of the autoencoder's latent representations to centre the data. This centralisation ensures that the linear damping effectively guides trajectories towards the origin.

Unless otherwise specified, the results presented showcase the top-performing model at each dimension, with the lowest relative error averaged across all considered statistics. It is crucial to note that, for all DNS versus DManD model comparisons at each d_h , identical initial conditions are applied in the models. From the perspective of the DManD models, this involves encoding the initial condition from the DNS, and subsequently evolving it forward in time in the latent space to generate a time series of \mathbf{h} . This time series is then decoded to the full-state space for comparative analysis.

3.3.1. Short-time tracking

In this section, we evaluate the performance of the DManD models in reconstructing short-time trajectories. Figure 4(a) shows the time evolution of the kinetic energy of the system for the DNS and DManD model with $d_h = 20$. Here, the kinetic energy of the system is given by the L^2 inner product

$$E = \frac{1}{\mathcal{V}} \int_0^{2\pi/\alpha} \int_0^{2\pi/m_p} \int_0^R \frac{1}{2} \mathbf{v} \cdot \mathbf{v} \, r \, dr d\theta dz, \quad (3.12)$$

where \mathcal{V} corresponds to the volume of the cylindrical flow domain. The results displayed in figure 4(a) are normalised by the kinetic energy of the laminar state. We have selected $d_h = 20$ as it represents the minimum dimension that yields superior results for both short-term and long-time measures, as will be demonstrated in this and the subsequent sections. We note that $d_h = 20$ may not necessarily correspond to the exact dimension of the manifold,

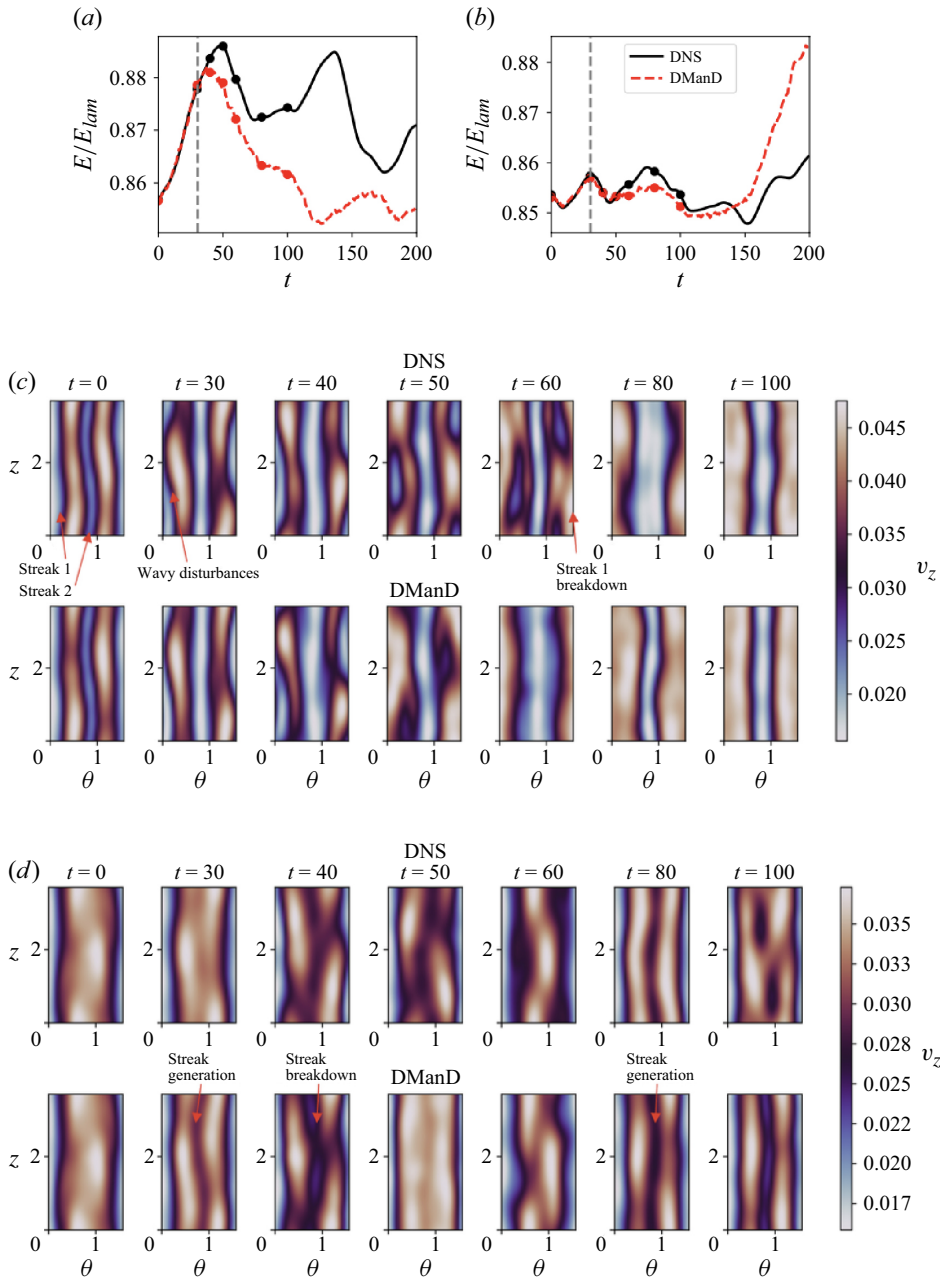


Figure 4. Normalised kinetic energy of the system for the DNS and DManD model with $d_h = 20$ up to $t = 200$ shown for two random initial conditions, corresponding to panels (a) and (b), respectively. Panels (c) and (d) represent two-dimensional representations of the dynamics in the $z-\theta$ plane ($r = 0.496$) with u_z for the DNS and DManD model for initial condition (IC) corresponding to panels (a) and (b), respectively. The vertical dashed line marks one Lyapunov time. We refer the reader to the supplemental materials to view a video of the trajectory corresponding to panel (a).

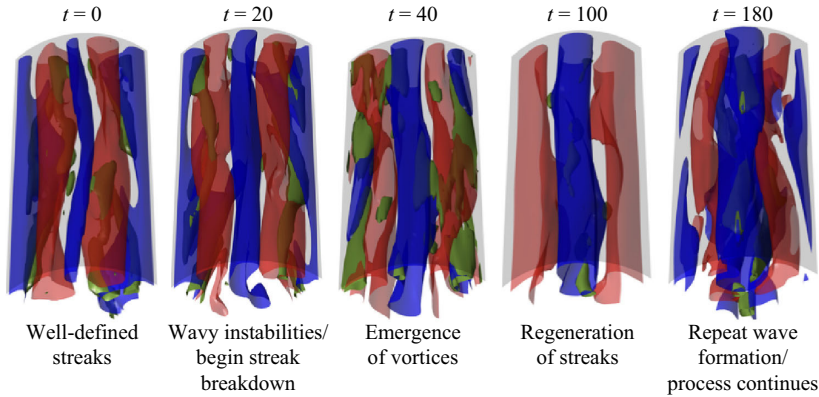


Figure 5. Self-sustaining process in DManD model corresponding to the initial condition shown in figure 4(a). Isosurfaces of the streamwise velocity fluctuations are displayed for $u'_z = 0.05$ (blue, representing fast streaks) and $u'_z = -0.05$ (red, representing low-speed streaks). Additionally, each snapshot includes isosurfaces of the λ_2 criterion with a threshold of $\lambda_2 = 0.1$, shown in green, to highlight vortical structures. Only a quarter of the domain is shown.

but this corresponds to the smallest dimension that can faithfully capture the nonlinear dynamics of the turbulent pipe flow in the present modelling framework. We reiterate that we have started from an initial state dimension of $\mathcal{O}(10^5)$, and developed a data-driven model of $\mathcal{O}(10)$ without substantial loss of accuracy. Figure 4(a,b) show that the DManD model can generate predictions that capture the true dynamics of the system up to $t \sim 50$, which corresponds to slightly more than one Lyapunov time ($t_L = 30.43$, this value is calculated later from the DNS). To provide a more qualitative representation of the model dynamics, figure 4(c,d) displays a two-dimensional representation of the dynamics in the $z - \theta$ plane ($r = 0.496$) with v_z for the DNS and DManD model for IC corresponding to figures 4(a) and 4(b), respectively. We observe a good agreement between the model and the DNS. We refer the reader to the supplementary materials to view a movie from the trajectory for figure 4(a).

The initial conditions for figure 4 are chosen to demonstrate that DManD qualitatively captures the dynamics observed in the true data. These two initial conditions are representative of the SSP, in which streamwise rolls, streaks and wave-like disturbances mutually reinforce one another, thereby counteracting viscous decay. This behaviour is further supported by the results shown in figure 5. In figure 4(c), the true system (top panels), streaks (velocity fluctuations below the mean) near the wall (see blue colour at the edges of the panel at $t = 0$) are affected by azimuthal wavy disturbances (at $t = 30-40$), leading to their breakdown (at $t = 50$) and the formation of rolls (after $t = 80$). The DManD qualitatively captures this dynamics (see the corresponding bottom panels) with a slightly accelerated breakdown of the rolls into streaks. In figure 4(d), we track a second initial condition that also highlights some parts of the SSP dynamics. At $t = 30$, streaks are observed near the centre of the domain. These streaks undergo a breakdown and temporarily vanish by $t = 50$, before gradually re-emerging by $t = 80$. This cyclical pattern observed in panels a and b, of decay followed by regeneration, reflects the characteristic SSP interplay between streamwise rolls, streaks and wave-like instabilities. The DManD reconstruction (bottom rows) qualitatively reproduces this sequence of events, capturing the key transitions present in the DNS (top rows).

Figure 5 illustrates the dynamics of the SSP observed in the DManD model using isosurfaces for the streaks and vortical structures. The blue and red isosurfaces represent

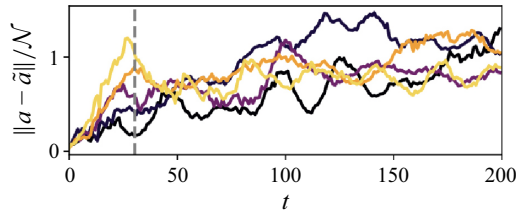


Figure 6. Normalised short-time-tracking error for five arbitrary initial conditions using DManD with $d_h = 20$ up to $t = 200$. The vertical dashed line marks one Lyapunov time. Two of the lines corresponds to the initial conditions shown in [figure 4](#).

high ($u'_z = 0.05$) and low ($u'_z = -0.05$) speed streaks, respectively, while the green isosurfaces correspond to regions of high vorticity, identified using the λ_2 criterion. At $t = 0$, the flow exhibits well-defined streamwise streaks. Due to the imposed periodic boundary conditions, only one quarter of the domain is shown, effectively capturing a pair of streaks. By $t = 20$, the low-speed streaks near the lateral walls begin to exhibit sinuous (wavy) instabilities, which amplify and lead to streak breakdown. This breakdown gives rise to streamwise vortices, clearly visible at $t = 40$ through the emergence of green λ_2 criterion structures. As the cycle progresses, these vortical structures redistribute momentum and energy, leading to the regeneration of coherent streaks by $t = 100$. At $t = 180$, a new breakdown event is underway, completing one full SSP cycle. This sequence captures the essential feedback loop of streak formation, instability, breakdown and regeneration that underpins sustained turbulence in wall-bounded flows.

While [figure 4](#) only shows the trajectory for two selected initial conditions, [figure 6](#) represents the tracking error for five random initial trajectories at $d_h = 20$. Here, we plot $\|\mathbf{a}(t) - \tilde{\mathbf{a}}(t)\|_2^2 / \mathcal{N}$, where \mathcal{N} denotes the error of true solutions at random times t_i and t_j , i.e. $\mathcal{N} = \langle \|\mathbf{a}(t_i) - \mathbf{a}(t_j)\| \rangle$. To enhance computational efficiency, we opted for comparisons in POD coefficients space rather than reconstructing full velocity fields. This decision was motivated by the computationally intensive nature (i.e. memory usage) of reconstructing fields from POD coefficients. Moreover, the POD coefficients enable the capture of the 99.83 % of the energy within the system (with 512 POD modes). [Figure 6](#) shows that, for certain initial conditions, the error remains relatively low until $t = 50$, after which it increases before stabilising at unity at longer times. This behaviour is expected for chaotic systems, where small initial errors grow exponentially and eventually lead to complete decorrelation between predicted and true trajectories. Once this occurs, the normalised error saturates at its maximum possible value, indicating a loss of predictive capability. We note the recent work by Vela-Martín & Avila (2024), who demonstrated that in Kolmogorov flow, the short-term predictability limit for a given uncertainty in initial conditions depends strongly on the location of the initial condition within the attractor. This supports our observation in [figure 6](#) that predictive accuracy varies across different initial conditions due to the intrinsic structure of the chaotic attractor.

Next, we conduct a parametric study on the normalised ensemble-averaged tracking error for DManD models by varying the dimension of the latent space (see [figure 7a](#)). We used 500 random initial conditions evolved over 100 time units. Notably, we observe that, when the dimensionality surpasses 17 ($d_h > 17$), the tracking errors converge, indicating that a dimension of at least 17 is required for a better field reconstruction.

To understand the short-time tracking and the correlation of the models with different initial conditions, we plot the autocorrelation of fluctuations of the kinetic energy. We define its fluctuating part as $k(t) = E(t) - \langle E \rangle$. In [figure 7](#), we plot the temporal autocorrelation of k with respect to its corresponding initial condition for 6000 random

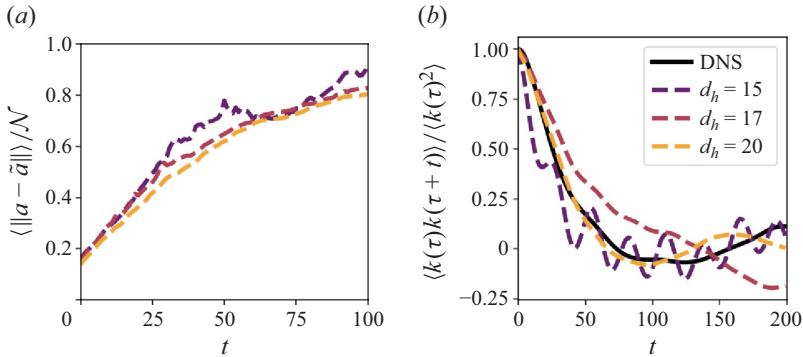


Figure 7. Short-time-tracking performance: (a) ensemble average of 500 random initial conditions as a function of d_h . (b) Temporal autocorrelation of fluctuations in the kinetic energy as a function of d_h . The black solid line represents the temporal autocorrelation calculated in the DNS. For representation purposes, we only show results for $d_h = [15, 17, 20]$.

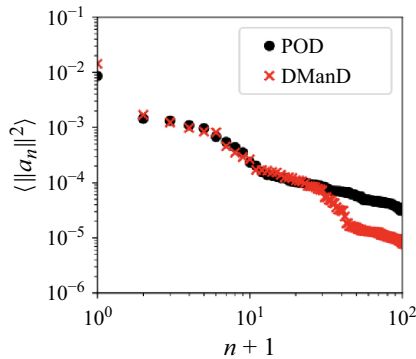


Figure 8. Comparison of $\langle \|a_n\|^2 \rangle$ for the DNS and DManD at $d_h = 20$ for the same initial condition evolved 3000 time units.

initial conditions evolved up to $t = 200$. It is not until $d_h = 20$ that the predicted autocorrelation matches reasonably well up to $t \sim 120$ with respect to the true data.

3.3.2. Long-time statistics

This section is dedicated to presenting the long-time statistics predictions for the DManD models. To evaluate the long-time performance of the DManD models, we first use the $\langle |a_n|^2 \rangle$ metric, plotting it for a long trajectory (up to 3000 time units) in both the DNS solution and the predicted trajectories from DManD at $d_h = 20$ (see figure 8). We observe that DManD shows good agreement with the true solution up to the first 30 leading POD coefficients of the true solution. This suggests that DManD effectively captures the attractor structure, and subsequently the temporal prediction of the nonlinear dynamics of the system over an extended time span. To benchmark the performance of our framework, we draw parallels with classical methods for reduced-order models proposed by Gibson (2002) for Couette flow (i.e. POD Galerkin). Gibson (2002) required between 512 (~ 1000 degrees of freedom) and 1024 modes (~ 2000 degrees of freedom) to achieve a reliable prediction of the leading 30 POD coefficients for plane Couette flow.

We turn our attention to the predictions of various models concerning the Reynolds stresses, as illustrated in figure 9. The streamwise velocity component $\langle u_z^2 \rangle$ is the most

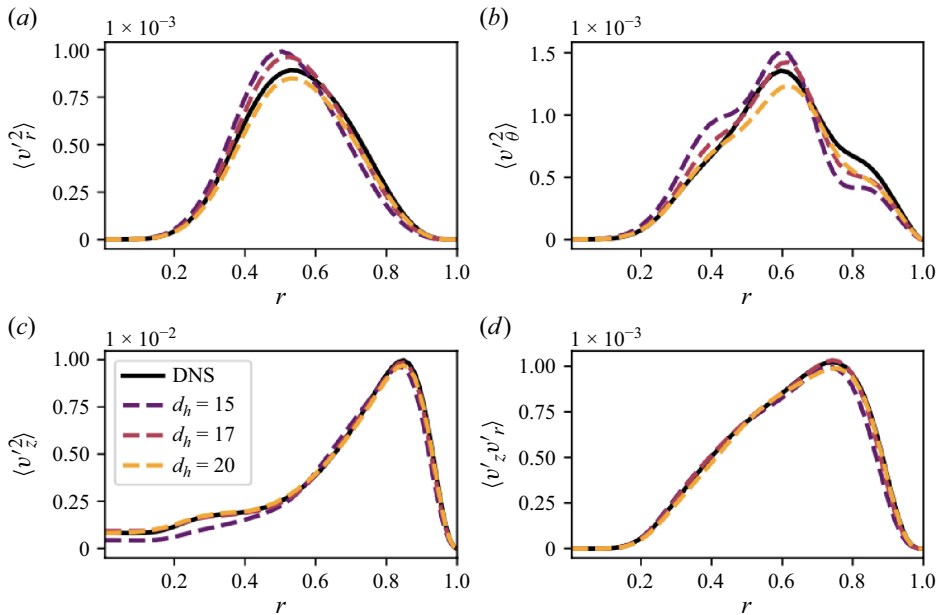


Figure 9. Long-time statistics: components of the Reynolds stresses with increasing dimension for DManD models of various dimensions. The black solid lines represent the Reynolds stresses calculated in the DNS. For clarity, we only show results for $d_h = [15, 17, 20]$.

important component with a peak near the wall, being one order of magnitude bigger than the other components. We observe that our DManD models perform well in predicting the Reynolds stresses of the system beyond $d_h > 17$. At $d_h = 20$, in particular, DManD exhibits exceptional performance by closely matching three out of the four displayed components. However, minor discrepancies are observed in the component corresponding to $\langle u_\theta^2 \rangle$.

Next, we assess how well the DManD models are capable of reconstructing the energy transfer rates at long times by examining the joint probability density functions of power input (I) and dissipation (D). The power input required to maintain constant mass flux and dissipation due to viscosity are defined as

$$I = \frac{1}{\mathcal{A}} \int_0^{2\pi/\alpha} \int_0^{2\pi/m_p} \left| \frac{\partial v_z}{\partial r} \right|_{r=R} d\theta dz, \quad (3.13)$$

$$D = \frac{1}{\mathcal{V}} \int_0^{2\pi/\alpha} \int_0^{2\pi/m_p} \int_0^R |\nabla \times \mathbf{v}|^2 r dr d\theta dz. \quad (3.14)$$

Here, \mathcal{V} and \mathcal{A} stand for the volume of the cylindrical flow domain and area of the pipe, respectively. In the results shown in this paper, the energy input and dissipation values are normalised with respect to their laminar values. An energy balance can be derived from the inner product $\langle \mathbf{v}, \partial \mathbf{v} / \partial t \rangle$ (Waleffe 2001). Then, the energy input rate, the dissipation rate and the kinetic energy $\bar{E} = (I - D)/Re$, which must average to zero over long times. It is imperative to ensure that this quantity averages to zero over extended periods, signifying equilibrium. This assessment is important for assessing the accuracy of DManD models in maintaining energy balance. Specifically, we define the energy balance deviation as $EB = \langle |I(t) - D(t)| \rangle / Re$, where $I(t)$ and $D(t)$ represent the cumulative energy input and dissipation up to time t . Over a trajectory of 5000 time units, the average deviation remains

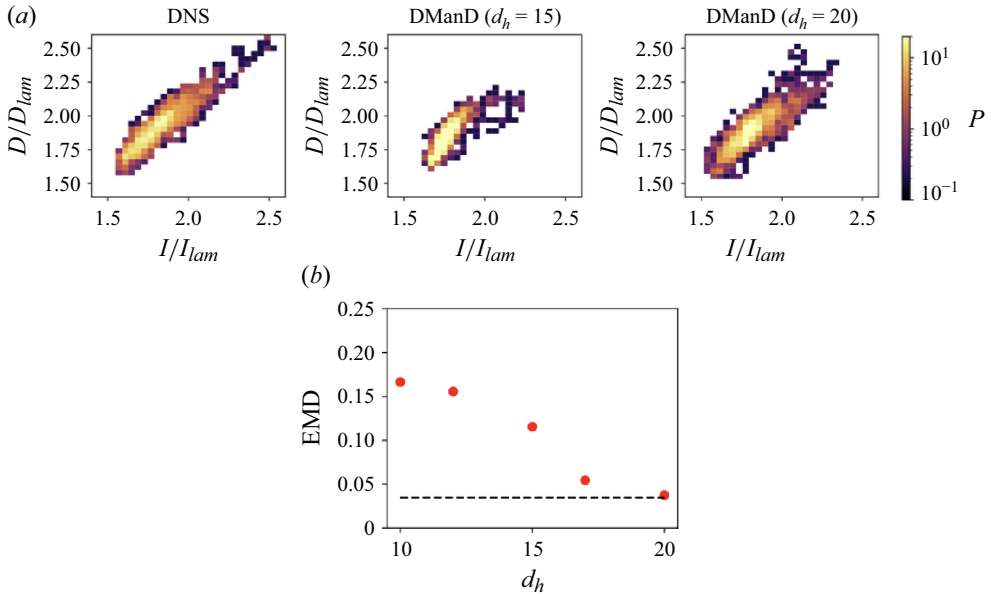


Figure 10. Energy balance: (a) joint probability density functions (PDFs) of the dissipation (D) and power input (I) for the true system, and the DManD models at $d_h = 15$ and $d_h = 20$, corresponding to columns one to three, respectively. (b) Earth movers distance (EMD) between the PDF from the DNS and the PDF predicted by the DManD model at various dimensions d_h . The dashed line represents the error between two PDFs generated from DNS trajectories of the same length but with different initial conditions.

below 1 %, demonstrating that the model faithfully captures the essential energy transfer and dissipation mechanisms without explicitly enforcing energy conservation.

Figure 10 displays the joint PDFs of the normalised D and I from the DNS, and DManD models with $d_h = 15$ and $d_h = 20$, generated from a longtime trajectory evolved up to 3000 time units (with the same initial condition). For $d_h = 15$, the model fails to capture the intrinsic nonlinear dynamics of the system, but for $d_h = 20$, the model accurately captures the core region of these projections, including the excursions occurring at high dissipation rates that are also present in the DNS results, indicating that high-dissipation bursts are preserved in the learned manifold when d_h increases. This suggests that the model is capable of encoding rare events that are observed in the DNS. Overall, we can conclude that DManD can effectively predict accurately the long-time statistics of this complex system in the coordinates of the low-dimensional representation.

To further quantify the divergence between the PDFs from the DNS and DManD, we calculate the EMD as a function of the dimension of the low-dimensional model. The EMD measures the distance between two PDFs by framing the true PDF as the ‘supplies’ and the DManD model PDF as the ‘demands’ (Peyré & Cuturi 2020). We use the EMD as a robust and interpretable metric of similarity between probability distributions. Unlike Kullback–Leibler divergence, EMD is a true distance that captures both the magnitude and spatial displacement of probability mass, features that are especially relevant in turbulent flows, where dissipation can exhibit heavy tails or abrupt shifts across regimes. The EMD seeks to minimise the effort required to transport the supplies to meet the demands, essentially solving a transportation problem. We find the flow f_{ij} that minimises $\sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}$ subject to the constraints

$$f_{ij} \geq 0, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n, \quad (3.15)$$

$$\sum_{j=1}^n f_{ij} = p_i \quad 1 \leq i \leq m, \quad (3.16)$$

$$\sum_{i=1}^m f_{ij} = q_j \quad 1 \leq j \leq n. \quad (3.17)$$

Here, p_i represents the probability density at the i th bin in the model PDF, and q_j is the probability density at the j th bin in the DNS PDF, where both PDFs are discretised into n and m bins (in this scenario, $n = m$). Additionally, d_{ij} denotes the cost associated with moving between bins, with the L_2 distance between bins i and j serving as the measure (where $d_{ij} = 0$ for $i = j$). After solving the minimisation problem to determine the optimal flow f_{ij}^* , the EMD is calculated as

$$\text{EMD} = \frac{\sum_{i=1}^m \sum_{j=1}^n f_{ij}^* d_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}^*}. \quad (3.18)$$

Figure 10(b) shows the EMD values depending on the latent dimensions, with all DManD models starting from the same initial condition and evolved up to 3000 time units. Additionally, we include the error when comparing two PDFs generated from the DNS with different conditions (dashed line). Figure 10(b) demonstrates that the addition of latent dimensions results in an enhancement of the EMD value. Specifically, for $d_h = 20$, the DManD model reaches a level of comparability to the DNS. This observation, combined with short-time tracking, supports the assertion that only 20 degrees of freedom are necessary to create low-dimensional models that faithfully capture both the short-time tracking and long-time statistics of the nonlinear turbulent dynamics of MFU pipe flow at $Re = 2500$. It is important to note that we are not asserting a specific dimension for the manifold, but rather identifying the minimum dimension needed to produce accurate models. Similar results have been observed for plane MFU Couette flow (Linot & Graham 2023; Constante-Amores *et al.* 2024a) and Kolmogorov flow (Pérez-De-Jesús & Graham 2023).

Finally, we examine the leading Lyapunov exponents of the DManD models depending on d_h . The methods used are those described in Sandri (1996), with publicly available code from Rozbeda (2017). Simulations were run for over 1000 time units to ensure convergence of the Lyapunov exponents. Figure 11(a) shows a representative spectrum of Lyapunov exponents λ_n^{LE} as a function of time, obtained from the DManD model with $d_h = 20$ for a single initial condition. We effectively see three positive exponents, we also observe two exponents near zero due to the spatial translational symmetries in θ and z . Figure 11(b) displays the exponents, averaged over five different initial conditions, as we vary the dimension of the DManD model. Increasing the model dimension leads to enhancements in the estimation of these Lyapunov exponents. At low dimensions ($d_h \leq 12$), we do not observe any positive Lyapunov exponents (i.e. models land in a fixed point). We also report the Lyapunov spectrum for $d_h = 22$ to demonstrate that the spectrum converges with increasing latent dimension d_h . Furthermore, the leading Lyapunov exponent computed from the DManD model closely matches that of the DNS (shown below), indicating that the dominant chaotic dynamics is well captured.

To enable comparison with DNS, we calculate the leading Lyapunov exponent (LLE) using the DNS solver. To do this calculation, we numerically evolve two nearby trajectories in Openpipeflow, and the divergence rate of their separation over time is calculated. As the trajectories separate, the difference between them (or the perturbation) can grow or shrink significantly. To prevent numerical errors and ensure consistent tracking, the

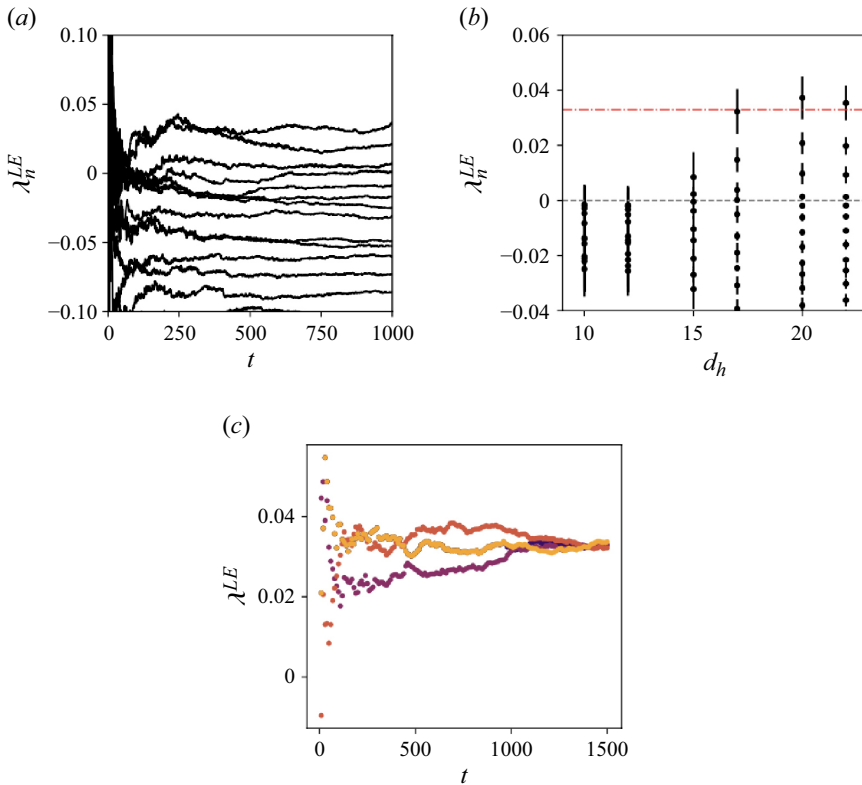


Figure 11. (a) Lyapunov exponents for the DManD models from a single trial with $d_h = 20$. (b) Lyapunov exponents for the DManD models at various dimensions with error bars representing the results from five different trials. The grey dashed line identifies $\lambda_n^{LE} = 0$, while the red dashed line represents the leading Lyapunov exponent of the DNS. (c) Leading Lyapunov exponent for the DNS for four different initial conditions.

separation is rescaled every 10 time units to a small, fixed value (10^{-9}) while maintaining its direction. This allows us to measure the exponential divergence rate reliably. The LLE is calculated as

$$\lambda^{LE} = \frac{A}{t}, \quad (3.19)$$

where A is the cumulative sum of the logarithmic growth of the separation, and t is the total time. After each rescaling, A is updated based on the ratio of the norms before and after rescaling.

To ensure the calculation of the Lyapunov exponent is meaningful, we evolve four different initial conditions that lie on the attractor (see figure 11c). The system is evolved for a sufficiently long time, and the LLE is updated incrementally until convergence with a tolerance of 10^{-6} . The average of these independent calculations yields a LLE of $\lambda^{LE} = 0.0329$. This value compares well with the LLE predicted by DManD.

3.4. Finding ECSs in the model and DNS

In this section, we leverage the DManD model with a dimension of $d_h = 20$ to explore the state space of the low-dimensional representation and discover new ECSs in the DNS. The primary goal is to use DManD to identify optimal initial conditions that can be fed into

Openpipeflow, which is equipped with an ECS solver for the full-state space. It is essential to highlight that discovering suitable initial conditions is pivotal to the success of any ECS solver for high-dimensional systems (as we described in the introduction).

First, we summarise the approach that we use to find ECSs within the context of DManD. This method follows the framework outlined by Cvitanovic *et al.* (2005), and has been previously used by Linot & Graham (2023). To identify ECSs in the full-state space (which, despite being high-dimensional due to the numerical discretisation of the infinite-dimensional NSE, is still finite), our objective is to find an initial condition that leads to a trajectory repeating over a defined time interval T . Thus, we aim to search for solutions where the trajectory's behaviour is periodic, essentially involving the identification of zeros in

$$\mathbf{F}(\mathbf{v}, T) = \mathbf{F}_T(\mathbf{v}) - \mathbf{v}, \quad (3.20)$$

where $\mathbf{F}_T(\mathbf{v})$ refers to the flow map T time units from \mathbf{v} (i.e. $\mathbf{F}_T(\mathbf{v}(t)) = \mathbf{v}(t + T)$). In manifold coordinates, this equation is expressed as

$$\mathbf{H}(\mathbf{h}, T) = \mathbf{G}_T(\mathbf{h}) - \mathbf{h}, \quad (3.21)$$

where $\mathbf{G}_T(\mathbf{h})$ is the flow map T time units from \mathbf{h} (i.e. $\mathbf{G}_T(\mathbf{h}(t)) = \mathbf{h}(t + T)$). To compute, \mathbf{G}_T , we use (2.8). Solving (3.21) requires finding both a point \mathbf{h}^* on the periodic orbit and the period T^* , such as $\mathbf{H}(\mathbf{h}^*, T^*) = 0$. We use a Newton–Raphson method to determine \mathbf{h}^* and T^* .

By taking a Taylor series expansion of \mathbf{H} , we find that near the fixed point \mathbf{h}^*, T^* of \mathbf{H} , such that

$$\left. \begin{aligned} \mathbf{H}(\mathbf{h}^*, T^*) - \mathbf{H}(\mathbf{h}, T) &\approx \mathbf{D}_h \mathbf{H}(\mathbf{h}, T) \Delta \mathbf{h} + \mathbf{D}_T \mathbf{H}(\mathbf{h}, T) \Delta T, \\ \mathbf{H}(\mathbf{h}, T) &\approx \mathbf{D}_h \mathbf{H}(\mathbf{h}, T) \Delta \mathbf{h} + \mathbf{g}(\mathbf{G}_T(\mathbf{h})) \Delta T \end{aligned} \right\}, \quad (3.22)$$

where \mathbf{D}_h and \mathbf{D}_T represent the Jacobians of \mathbf{H} with respect to \mathbf{h} and the period T , respectively. Additionally, $\Delta T = T^* - T$ and $\Delta \mathbf{h} = \mathbf{h}^* - \mathbf{h}$. We impose the constraint that the updates of $\Delta \mathbf{h}$ are orthogonal to the vector field \mathbf{h} (i.e. $\mathbf{g}(\mathbf{h})^T \Delta \mathbf{h} = 0$). At a Newton step (i), the system of equations becomes

$$\begin{bmatrix} \mathbf{D}_h \mathbf{H}(\mathbf{h}^{(i)}, T^{(i)}) & \mathbf{g}(\mathbf{G}_{T^{(i)}}(\mathbf{h}^{(i)})) \\ \mathbf{g}(\mathbf{h}^{(i)})^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{h}^{(i)} \\ \Delta T^{(i)} \end{bmatrix} = - \begin{bmatrix} \mathbf{H}(\mathbf{h}^{(i)}, T^{(i)}) \\ 0 \end{bmatrix}, \quad (3.23)$$

where the standard Newton–Raphson method updates the guesses $\mathbf{h}^{(i+1)} = \mathbf{h}^{(i)} + \Delta \mathbf{h}^{(i)}$ and $T^{(i+1)} = T^{(i)} + \Delta T^{(i)}$.

A Newton scheme can be used to find ECSs within high-dimensional data, the computational challenge posed by the Jacobian calculations has prompted the development of various solutions (Page & Kerswell 2020; Page, Brenner & Kerswell 2021; Parker & Schneider 2022; Linot & Graham 2023; Yasuda & Lucas 2024). Openpipeflow addresses this issue by using a Jacobian-free Newton–Krylov solver with a hookstep-trust-region approach, as detailed by Willis (2019a). This solver efficiently bypasses the need for explicitly calculating the Jacobian when evaluating the objective function \mathbf{F} . For a more detailed understanding of the ECS solver in the Openpipeflow solver, readers are referred to Willis (2019b). The advantage of constructing a low-dimensional model that accurately captures the dynamics becomes pivotal in the search for new ECSs, as demonstrated in this section. Leveraging the inherent low-dimensionality of DManD models, we choose to compute the Jacobian $\mathbf{D}_h \mathbf{H}(\mathbf{h}, T)$ directly using the automatic differentiation tools employed during the training of the NODE.

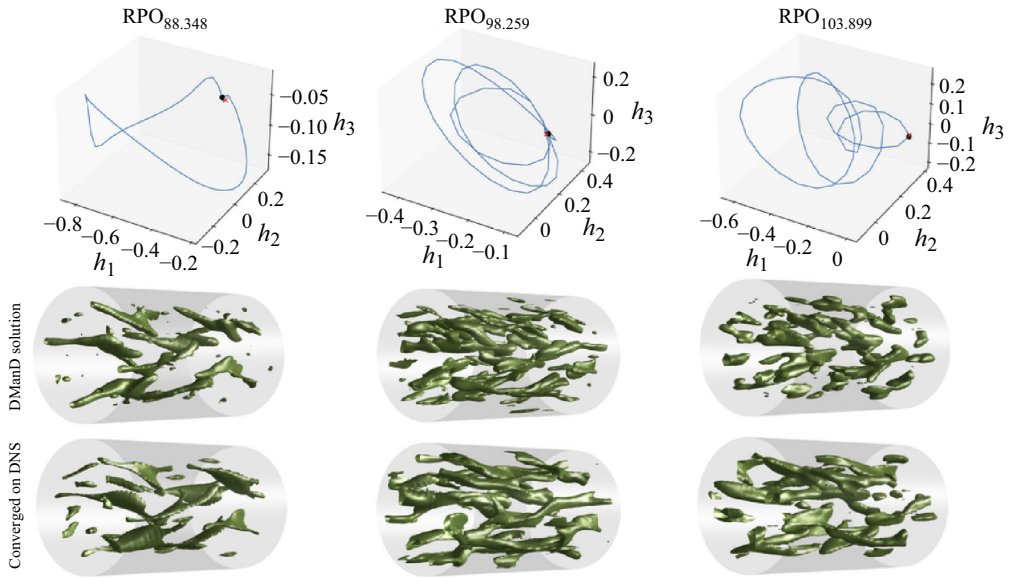


Figure 12. The top panels show the converged ECS by DManD projected onto the first three manifold coordinates. The black dot and red diamond indicate the starting and ending points of the trajectory, respectively. The middle and bottom panels show snapshots of the vorticity flow field using the λ_2 criterion with a threshold of $\lambda_2 = 0.1$. The middle panels represent the ECS from the DManD search, while the bottom panels display the converged state from the ECS solver of Openpipeflow. Note that for representation purposes, we display $m_p = 1$ instead of $m_p = 4$.

Our approach involves randomly selecting 100 initial conditions and exploring five distinct periods $T = [5, 20, 40, 90, 100]$ (e.g. 500 guesses in total). In the reduced model, a Newton residual threshold of 10^{-3} identifies converged ECS candidates efficiently. These provide accurate initial guesses for the full system, where exact convergence is achieved with residuals between 10^{-4} and 10^{-6} , as used in previous works such as Willis *et al.* (2013).

As an illustrative example, the top panels of figure 12 depict the trajectories of three relative periodic orbits discovered by DManD, projected onto the three leading manifold coordinates (h_1, h_2, h_3) . We also display the associated vorticity fields by displaying the λ_2 criterion (Jeong & Hussain 1995) for the state of the converged ECS from DManD (middle panels) and the converged ECS in the DNS (bottom panels). There is a notable qualitative agreement between these solutions, underscoring the effectiveness of using converged ECS from DManD as robust initial conditions for the DNS ECS solver.

Table 2 presents 17 new ECSs of pipe flow identified via the ECS solver embedded in Openpipeflow using the converged solutions obtained via DManD. The RPOs are labelled by their periods T (in units of \mathcal{D}/U), while travelling waves (TWs) are labelled by their average dissipation rate (\bar{D}). Additionally, table 2 includes information regarding the linear stability of each ECS, described by its Floquet multipliers $\Lambda = \exp(\mu_j T + i\theta_j)$. We have performed a systematic cross-check of these ECSs against all ECSs from Willis *et al.* (2013, 2016) and Budanur *et al.* (2017). We note that many of our newly discovered ECSs exhibit significantly longer periods than those previously reported (i.e. Budanur *et al.* (2017) reported RPOs with periods up to $T \approx 68$). We note that the DManD solver demonstrates its capability to accurately capture the orbit lengths of the ECSs, with errors in their periods consistently below 1 %.

Solution	\bar{D}	\bar{c}	μ^{max}
TW _{1.332}	1.3325	0.93571	0.04056
TW _{1.798}	1.7985	1.6619	0.08803
TW _{1.448}	1.4484	1.8009	0.04892
TW _{1.575}	1.5759	2.7052	0.12259
RPO _{4.891}	1.4740	2.7430	0.06686
RPO _{19.631}	2.0324	3.30392	0.03060
RPO _{19.631}	2.0172	3.30392	0.03605
RPO _{24.956}	1.8644	2.4739	0.03784
RPO _{35.042}	1.7582	1.6211	0.04585
RPO _{36.157}	1.7503	1.8426	0.03465
RPO _{38.654}	2.0475	1.142	0.04499
RPO _{88.348}	1.7985	0.83400	0.04039
RPO _{94.891}	1.9147	4.01616	0.03011
RPO _{98.259}	2.0036	1.5157	0.02991
RPO _{102.683}	1.9793	0.98244	0.02988
RPO _{103.899}	1.9036	6.1943	0.01088
RPO _{105.466}	1.8270	3.0258	0.01653

Table 2. List of new invariant solutions for pipe flow at $Re = 2500$ using initial conditions from the DManD model. The travelling waves are labelled with their dissipation rate \bar{D} , whereas the relative periodic orbits (RPOs) are labelled by their period T . For each solution, we report the average rate of dissipation \bar{D} , average downstream velocity \bar{c} , and the real part of the largest stability eigenvalue/Floquet exponent μ^{max} .

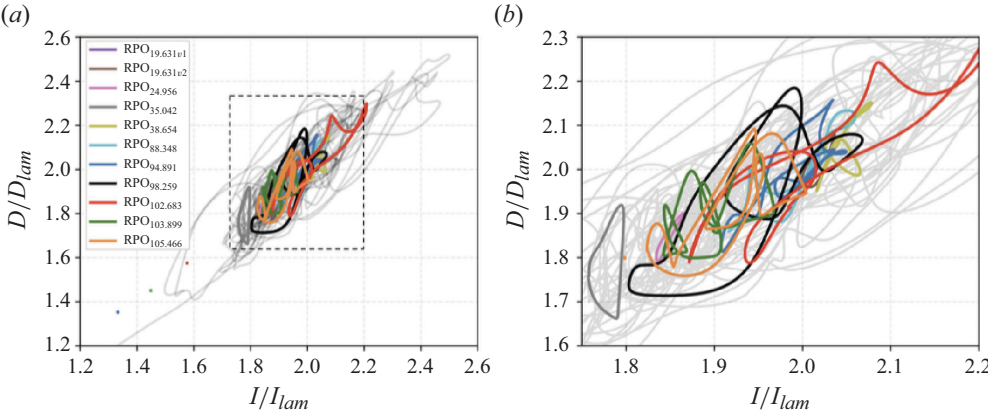


Figure 13. The normalised dissipation versus power input for the collection of invariant solutions displayed in [table 2](#) with a long trajectory of the DNS turbulence plotted in the background. The dashed box in the left panel outlines the region that is magnified in the right panel for a closer view.

One useful method for visualising of the high-dimensional space of these new invariant solutions is by projecting them onto the dissipation versus energy input plane. [Figure 13](#) illustrates the input and dissipation projection of these new ECSs, with the shaded region indicates the area in the input-dissipation projection where a long turbulent trajectory predominantly resides. It is expected that turbulence should explore more of the phase space and visit simple invariant solutions, then these ECSs appear to be embedded in this projection. While these solutions appear to reside within the central region of the ID projection, determining their precise location within the turbulent attractor would require a more detailed analysis, in the full-state space, such as the work by Krygier,

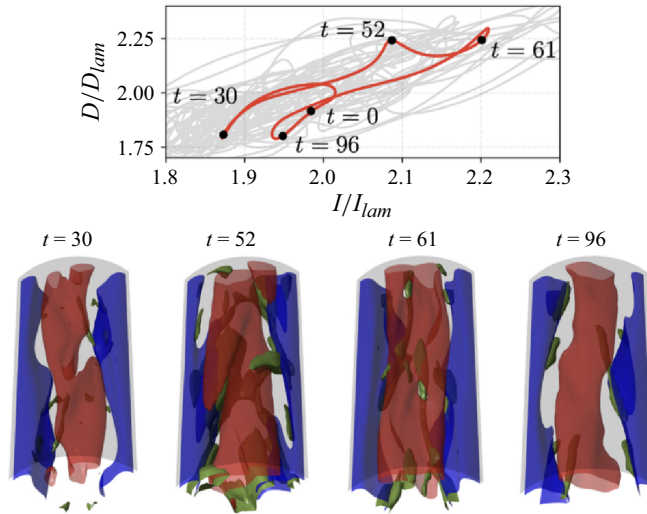


Figure 14. The normalised dissipation versus power input plot for $\text{RPO}_{102.683}$ is depicted, with a long trajectory of the DNS turbulence plotted in the background. The black dot points denote the times along the orbit at which three-dimensional snapshots are shown below. Each snapshot displays isosurfaces of the streamwise velocity fluctuations for $u'_z = 0.075$ (blue, representing fast streaks) and $u'_z = -0.075$ (red, representing low-speed streaks). Additionally, each snapshot includes isosurfaces of the λ_2 criterion with a threshold of $\lambda_2 = 0.1$, shown in green, to highlight vortical structures. Note that for representation purposes, only a quarter of the domain is shown.

Pughe-Sanford & Grigoriev (2021). But we leave these avenues for future work. This finding is consistent with previous assertions suggesting that the attractor is guided by ECS Chandler & Kerswell (2013), Cvitanović (2013), Budanur *et al.* (2017). The condition $D = I$ signifies dissipation balancing out energy input, which is the essential requirement for any equilibria or TW (essentially an equilibrium in a co-moving frame of reference). We observe the discovery of four TWs. The shaded region shows that the DNS predominantly stays within the region of $1.6 < D/D_{lam} < 2.2$ and $1.7 < I/I_{lam} < 2.2$. Most of the discovered ECSs remain within this region. In the right panel of figure 13, we present a magnified view. This image reveals that many of these RPOs have complicated ID curves, whereas the periodic orbits exhibit simple loops.

In figures 14 and 15, we focus on elucidating the state space of $\text{RPO}_{102.683}$ and $\text{RPO}_{103.899}$, respectively. The selection of $\text{RPO}_{102.683}$ is based on its characteristic complex ID curve, which spans both high and low-dissipation regions within the state space. To understand the flow dynamics associated with this RPO, the bottom panels of figure 14 also shows blue and red isosurfaces representing high ($u'_z = 0.075$) and low ($u'_z = -0.075$) speed streaks, respectively, while the green isosurfaces correspond to regions of high vorticity, identified using the λ_2 criterion with $\lambda_2 = 0.1$. We observe that the flow exhibits well-defined streamwise streaks during the entire cycle of the RPO, with some wavy disturbances at $t = 30$, $t = 52$ and $t = 96$. At the points of higher dissipation, ($t = 52$) and ($t = 61$), the vorticity appears to undergo significant shearing, leading to vortex breakup (see the intense isosurfaces associated with λ_2 criterion). High rates of dissipation correspond to intense velocity gradients, and subsequently, this leads to high values of the strain rate tensor. This indicates that these times are characterised by strong and turbulent flow structures. At the lowest point of dissipation, see $t = 30$ and $t = 96$, the vortex structures are much weaker and less pronounced. This weakening of the vortex structures

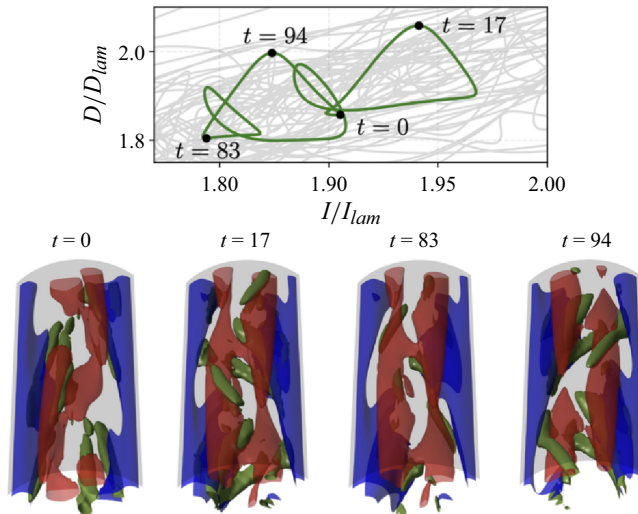


Figure 15. The normalised dissipation versus power input plot for $RPO_{103,899}$ is depicted, with a long trajectory of the DNS turbulence plotted in the background. The black dot points denote the times along the orbit at which three-dimensional snapshots are shown below. Each snapshot displays isosurfaces of the streamwise velocity fluctuations for $u'_z = 0.075$ (blue, representing fast streaks) and $u'_z = -0.075$ (red, representing low-speed streaks). Additionally, each snapshot includes isosurfaces of the λ_2 criterion with a threshold of $\lambda_2 = 0.1$, shown in green, to highlight vortical structures. Note that for representation purposes, only a quarter of the domain is shown.

correlates with lower dissipation rates. Then, the dynamics stabilises due to the increases in the power input ($t = 96$), in comparison with $t = 52$. It is worth noting that for these three-dimensional snapshots only a quarter of the domain is shown. Finally, we turn our attention to $RPO_{103,899}$. The top panel of figure 15 presents the ID projection for this RPO, which intriguingly resembles the shadow of two shorter orbits, similar to those described by Budanur *et al.* (2017) (see their figure 11). We also observe well-formed fast streaks in the domain. This results in the dynamics at the highest point of dissipation, at $t = 17$ and $t = 94$, exhibiting significant shearing, leading to intense vortex interactions and high dissipation rates. These moments are characterised by extreme shearing, events affecting the slow streaks. Conversely, at the lowest point of dissipation, for example $t = 83$, the vortex structures are notably weaker. This reduced dissipation corresponds to a more stable and less turbulent flow state.

4. Conclusions

In this study, we have built data-driven models for pressure-driven fluid flow through a circular pipe. To reduce the computational requirements, we impose the shift-and-reflect symmetry to study the system in a minimal computational cell at $Re = 2500$. Nonetheless, this computational size is capable of maintaining the chaotic nonlinear dynamics of turbulence. To build these data-driven models, we employed DManD, an invariant-manifold-based method. The DManD model is based on the idea of the modelling of turbulence from a dynamical systems approach in which the long-time dynamics of the dissipative NSE is expected to live in a finite-dimensional invariant manifold. Thus, DManD allows the parameterisation of the invariant manifold with vastly fewer degrees of freedom compared with the original data. For learning these manifold coordinates, we first perform a linear dimension reduction with POD, and then a nonlinear dimension reduction

via autoencoders which are capable of accurately predicting the low POD coefficients. Finally, we use a state-space approach with NODEs within these learned coordinates to model the dynamics. This combination of linear and nonlinear techniques allows for a compact and efficient representation of the turbulent flow dynamics. Our framework, solely driven by data, enables us to construct models with fewer than 20 degrees of freedom, a significant reduction compared with a fully resolved DNS that requires of the order of $\mathcal{O}(10^5)$. In short-time tracking, the model accurately track the true trajectory for one Lyapunov time. Additionally, the LLE estimated from DManD closely matches that obtained from the DNS, confirming that our approach captures the chaotic dynamics and the short-term predictability of the flow. In the long term, the models successfully capture key aspects of the nonlinear dynamics such as the Reynolds stresses and the probability distribution in ID space.

We have also identified seventeen previously unknown ECSs for turbulent pipe flow at $Re = 2500$. The success in discovering these new ECSs lies in using converged ECSs from DManD at $d_h = 20$ as effective initial conditions for the ECS solver in Openpipeflow. This approach has led to the reporting of RPOs with the longest periods observed for three-dimensional turbulent pipe flow to date, to the best of our knowledge. These periodic orbits are situated within the core of the state space traversed by the attractor. This finding is consistent with previous assertions suggesting that the turbulent attractor is guided by ECSs (Hopf 1948; Cvitanović 2013; Budanur *et al.* 2017; Page *et al.* 2024).

We acknowledge that Linot & Graham (2023) presented DManD models for turbulent Couette flow with fewer than 18 degrees of freedom, which quantitatively capture key features of the dynamics, including the streak breakdown and regeneration cycle. At short times, these models track true trajectories for multiple Lyapunov times, while at long times they reproduce the Reynolds stresses, capture the energy balance and provide effective initial conditions for finding ECS. In the present study, we extend this methodology to pipe flow, which constitutes a more complex geometry and higher-dimensional dynamics; the DManD model identifies an embedding dimension of approximately 20. We demonstrate that DManD similarly captures the SSP, reproduces the LLE and provides effective initial conditions for ECSs in the pipe setting. Thus, while the essential capabilities of DManD are consistent across Couette and pipe flows, our results highlight its robustness and applicability to more complex wall-bounded shear flows, thereby broadening the scope of the approach.

Accurately modelling the turbulent dynamics with significantly fewer degrees of freedom than required for DNS, as demonstrated by the manifold dynamics models presented here, opens up exciting possibilities for dynamical-systems-type analyses. These models enable the calculation of local Lyapunov exponents in a computationally efficient manner. We also highlight that we have presented a global description of the manifold, but it would be possible to divide the global manifold topology into many local representations called charts (Floryan & Graham 2022; Fox *et al.* 2023). In experimental settings, temporal data are typically limited to partial observables (i.e. measurements at a few spatial locations) making full-state methods such as DManD difficult to implement. Nonetheless, data-driven models constructed from such partial measurements may still enable control-oriented strategies. Investigating this direction represents a promising avenue for future research.

Importantly, the models developed in this study are dependent on specific system parameters such as the Reynolds number. Therefore, when transitioning to different Reynolds numbers, it is necessary to obtain a new dataset to adjust the weights of the neural networks for the autoencoders and NODE. Therefore, a crucial direction for future research is to

develop models that capture this parameter dependence. The goal is to create robust low-dimensional models capable of transferring knowledge across different parameter regimes. Achieving this would enable broader applications and provide deeper insights into the turbulent flow dynamics. This represents a key direction for future research in this field.

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

Data availability. The code and data (in compressed format) used in the paper is available in the group GitHub repository. The complete dataset is available from the authors upon request.

We wish to thank A.P. Willis for making the ‘Openpipeflow’ code open source, and his help setting up the simulations and subsequent discussions. We gratefully acknowledge J. Buzhardt for helpful discussions. This work was supported by ONR N00014-18-1-2865 (Vannevar Bush Faculty Fellowship). The authors gratefully acknowledge the valuable feedback provided by the three anonymous reviewers during the revision process.

REFERENCES

- ABDI, H. & WILLIAMS, L.J. 2010 Principal component analysis. *Wiley Interdisciplin. Rev.: Comput. Stat.* **2** (4), 433–459.
- AVILA, M., BARKLEY, D. & HOF, B. 2023 Transition to turbulence in pipe flow. *Annu. Rev. Fluid Mech.* **55** (1), 575–602.
- BUDANUR, N.B., SHORT, K.Y., FARAZMAND, M., WILLIS, A.P. & CVITANOVIĆ, P. 2017 Relative periodic orbits form the backbone of turbulent pipe flow. *J. Fluid Mech.* **833**, 274–301.
- CHANDLER, G.J. & KERSWELL, R.R. 2013 Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow. *J. Fluid Mech.* **722**, 554–595.
- CHEN, R.T.Q., RUBANOVA, Y., BETTENCOURT, J. & DUVENAUD, D. 2019 Neural ordinary differential equations. [arXiv:1806.07366](https://arxiv.org/abs/1806.07366).
- CONSTANTE-AMORES, C.R., FOX, A.J., JESÚS, C.E.P.D. & GRAHAM, M.D. 2024a Data-driven Koopman operator predictions of turbulent dynamics in models of shear flows. [arXiv:2407.16542](https://arxiv.org/abs/2407.16542).
- CONSTANTE-AMORES, C.R. & GRAHAM, M.D. 2024 Data-driven state-space and Koopman operator models of coherent state dynamics on invariant manifolds. *J. Fluid Mech.* **984**, R9.
- CONSTANTE-AMORES, C.R., LINOT, A.J. & GRAHAM, M.D. 2024b Data-driven prediction of large-scale spatiotemporal chaos with distributed low-dimensional models. [arXiv:2410.01238](https://arxiv.org/abs/2410.01238).
- CVITANOVIC, P., ARTUSO, R., MAINIERI, R., TANNER, G., VATTAY, G., WHELAN, N. & WIRZBA, A. 2005 Chaos: classical and quantum. *ChaosBook. Org (Niels Bohr Institute, Copenhagen 2005)* **69**, 25.
- CVITANOVIĆ, P. 2013 Recurrent flows: the clockwork behind turbulence. *J. Fluid Mech.* **726**, 1–4.
- DARBYSHIRE, A.G. & MULLIN, T. 1995 Transition to turbulence in constant-mass-flux pipe flow. *J. Fluid Mech.* **289**, 83–114.
- DUGGLEBY, A., BALL, K.S., PAUL, M.R. & FISCHER, P.F. 2007 Dynamical eigenfunction decomposition of turbulent pipe flow. *J. Turbul.* **8**, N43.
- DUGUET, Y., WILLIS, A.P. & KERSWELL, R.R. 2008 Transition in pipe flow: the saddle structure on the boundary of turbulence. *J. Fluid Mech.* **613**, 255–274.
- ECKHARDT, B. 2009 Introduction. turbulence transition in pipe flow: 125th anniversary of the publication of Reynolds paper. *Phil. Trans. R. Soc. A* **367**, 449–455.
- ECKHARDT, B., SCHNEIDER, T.M., HOF, B. & WESTERWEEL, J. 2007 Turbulence transition in pipe flow. *Annu. Rev. Fluid Mech.* **39** (1), 447–468.
- FAISST, H. & ECKHARDT, B. 2003 Traveling waves in pipe flow. *Phys. Rev. Lett.* **91**, 224502.
- FLORYAN, D. & GRAHAM, M. 2022 Data-driven discovery of intrinsic dynamics. *Nat. Mach. Intell.* **4** (12), 1113–1120.
- FOIAS, C., MANLEY, O. & TEMAM, R. 1988 Modelling of the interaction of small and large eddies in two dimensional turbulent flows. *ESAIM: M2AN* **22** (1), 93–118.
- FOX, A.J., RICARDO CONSTANTE-AMORES, C. & GRAHAM, M.D. 2023 Predicting extreme events in a data-driven model of turbulent shear flow using an atlas of charts. *Phys. Rev. Fluids* **8**, 094401.
- FUKAMI, K., FUKAGATA, K. & TAIRA, K. 2019 Super-resolution reconstruction of turbulent flows with machine learning. *J. Fluid Mech.* **870**, 106–120.
- GIBSON, J.F. 2002 Dynamical systems models of wall-bounded, shear-flow turbulence. PhD thesis, Cornell University.
- GOODFELLOW, I., BENGIO, Y., COURVILLE, A. & BENGIO, Y. 2016 *Deep Learning*, vol. 1. MIT Press
- GRAHAM, M.D. & FLORYAN, D. 2021 Exact coherent states and the nonlinear dynamics of wall-bounded turbulent flows. *Annu. Rev. Fluid Mech.* **53** (1), 227–253.

- HAMILTON, J.M., KIM, J. & WALEFFE, F. 1995 Regeneration mechanisms of near-wall turbulence structures. *J. Fluid Mech.* **287**, 317–348.
- HINTON, G.E. & SALAKHUTDINOV, R.R. 2006 Reducing the dimensionality of data with neural networks. *Science* **313** (5786), 504–507.
- HOF, B., VAN DOORNE, C.W.H., WESTERWEEL, J., NIEUWSTADT, F.T.M., FAISST, H., ECKHARDT, B., WEDIN, H., KERSWELL, R.R. & WALEFFE, F. 2004 Experimental observation of nonlinear traveling waves in turbulent pipe flow. *Science* **305** (5690), 1594–1598.
- HOLMES, P., LUMLEY, J.L., BERKOOZ, G. & ROWLEY, C.W. 2012 Turbulence, coherent structures, dynamical systems and symmetry, 2nd edn. *Cambridge Monographs on Mechanics*. Cambridge University Press.
- HOPF, E. 1948 A mathematical example displaying features of turbulence. *Commun. Pure Appl. Maths* **1** (4), 303–322.
- JEONG, J. & HUSSAIN, F. 1995 On the identification of a vortex. *J. Fluid Mech.* **285**, 69–94.
- JIMÉNEZ, J. & MOIN, P. 1991 The minimal flow unit in near-wall turbulence. *J. Fluid Mech.* **225**, 213–240.
- JIMÉNEZ, K. 2013 Near-wall turbulence. *Phys. Fluids* **25** (10), 101302.
- JOLLIFFE, I.T. 1986 *Principal Component Analysis and Factor Analysis*, pp. 115–128. Springer New York.
- KASZÁS, B., CENEDESE, M. & HALLER, G. 2022 Dynamics-based machine learning of transitions in Couette flow. *Phys. Rev. Fluids* **7**, L082402.
- KASZÁS, B. & HALLER, G. 2024 Capturing the edge of chaos as a spectral submanifold in pipe flows. *J. Fluid Mech.* **979**, A48.
- KAWAHARA, G. & KIDA, S. 2001 Periodic motion embedded in plane Couette turbulence: regeneration cycle and burst. *J. Fluid Mech.* **449**, 291–300.
- KAWAHARA, G., UHLMANN, M. & VAN VEEN, L. 2012 The significance of simple invariant solutions in turbulent flows. *Annu. Rev. Fluid Mech.* **44** (2012), 203–225.
- KOGELBAUER, F. & HALLER, G. 2018 Rigorous model reduction for a damped-forced nonlinear beam model: an infinite-dimensional analysis. *J. Nonlinear Sci.* **28**, 1109–1150.
- KRAMER, M.A. 1991 Nonlinear principal component analysis using autoassociative neural networks. *AICHE J.* **37** (2), 233–243.
- KREILOS, T. & ECKHARDT, B. 2012 Periodic orbits near onset of chaos in plane Couette flow. *Chaos* **22** (4), 047505.
- KRYGIER, M.C., PUGHE-SANFORD, J.L. & GRIGORIEV, R.O. 2021 Exact coherent structures and shadowing in turbulent Taylor–Couette flow. *J. Fluid Mech.* **923**, A7.
- LAN, Y. & CVITANOVIĆ, P. 2004 Variational method for finding periodic orbits in a general flow. *Phys. Rev. E* **69**, 016217.
- LI, M., JAIN, S. & HALLER, G. 2022 Nonlinear analysis of forced mechanical systems with internal resonance using spectral submanifolds, Part I: Periodic response and forced response curve. *Nonlinear Dyn.* **110** (2), 1005–1043.
- LINOT, A.J., BURBY, J.W., TANG, Q., BALAPRAKASH, P., GRAHAM, M.D. & MAULIK, R. 2023 Stabilized neural ordinary differential equations for long-time forecasting of dynamical systems. *J. Comput. Phys.* **474**, 111838.
- LINOT, A.J. & GRAHAM, M.D. 2020 Deep learning to discover and predict dynamics on an inertial manifold. *Phys. Rev. E* **101**, 062209.
- LINOT, A.J. & GRAHAM, M.D. 2022 Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations. *Chaos* **32** (7), 073110.
- LINOT, A.J. & GRAHAM, M.D. 2023 Dynamics of a data-driven low-dimensional model of turbulent minimal Couette flow. *J. Fluid Mech.* **973**, A42.
- LIU, A., AXÁS, J. & HALLER, G. 2024 Data-driven modeling and forecasting of chaotic dynamics on inertial manifolds constructed as spectral submanifolds. *Chaos* **34** (3), 033140.
- DE LOZAR, A., MELLIBOVSKY, F., AVILA, M. & HOF, B. 2012 Edge state in pipe flow experiments. *Phys. Rev. Lett.* **108**, 214502.
- MILANO, M. & KOUMOUTSAKOS, P. 2002 Neural network modeling for near wall turbulent flow. *J. Comput. Phys.* **182** (1), 1–26.
- MULLIN, T. 2011 Experimental studies of transition to turbulence in a pipe. *Annu. Rev. Fluid Mech.* **43** (1), 1–24.
- NAGATA, M. 1990 Three-dimensional finite-amplitude solutions in plane Couette flow: bifurcation from infinity. *J. Fluid Mech.* **217**, 519–527.
- PAGE, J., BRENNER, M.P. & KERSWELL, R.R. 2021 Revealing the state space of turbulence using machine learning. *Phys. Rev. Fluids* **6**, 034402.

- PAGE, J. & KERSWELL, R.R. 2020 Searching turbulence for periodic orbits with dynamic mode decomposition. *J. Fluid Mech.* **886**, A28.
- PAGE, J., NORGAARD, P., BRENNER, M.P. & KERSWELL, R.R. 2024 Recurrent flow patterns as a basis for two-dimensional turbulence: predicting statistics from structures. *Proc. Natl Acad. Sci.* **121** (23), e2320007121.
- PARKER, J.P. & SCHNEIDER, T.M. 2022 Variational methods for finding periodic orbits in the incompressible Navier–Stokes equations. *J. Fluid Mech.* **941**, A17.
- PASZKE, A. 2019 Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* (ed. H. Wallach, A. Larochelle, F. Beygelzimer, E.F. d’Alché-Buc & P. Garnett), vol. 32. Curran Associates, Inc.
- PÉREZ-DE-JESÚS, C.E. & GRAHAM, M.D. 2023 Data-driven low-dimensional dynamic model of Kolmogorov flow. *Phys. Rev. Fluids* **8**, 044402.
- PEYRÉ, G. & CUTURI, M. 2020 Computational optimal transport. [arXiv:1803.00567](https://arxiv.org/abs/1803.00567)
- PRINGLE, C.C.T., DUGUET, Y. & KERSWELL, R.R. 2009 Highly symmetric travelling waves in pipe flow. *Phil. Trans. R. Soc. A: Math. Phys. Engng Sci.* **367** (1888), 457–472.
- PRINGLE, C.C.T. & KERSWELL, R.R. 2007 Asymmetric, helical, and mirror-symmetric traveling waves in pipe flow. *Phys. Rev. Lett.* **99**, 074502.
- REYNOLDS, O. 1883 An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Phil. Trans. R. Soc. A* **174**, 935–982.
- ROZBEDA, P. 2017 pylyapunov. <https://github.com/paulrozdeba/pylyapunov>
- SANDRI, M. 1996 Numerical calculation of Lyapunov exponents. *Math. J.* **6** (3), 78–84.
- SMITS, A.J., MCKEON, B.J. & MARUSIC, I. 2011 High–Reynolds number wall turbulence. *Annu. Rev. Fluid Mech.* **43** (1), 353–375.
- TEMAM, R. 1989 Do inertial manifolds apply to turbulence? *Phys. D: Nonlinear Phenom.* **37** (1), 146–152.
- VELA-MARTÍN, A. & AVILA, M. 2024 Large-scale patterns set the predictability limit of extreme events in Kolmogorov flow. *J. Fluid Mech.* **986**, A2.
- VISWANATH, D. 2007 Recurrent motions within plane Couette turbulence. *J. Fluid Mech.* **580**, 339–358.
- VISWANATH, D. & CVITANOVIĆ, P. 2009 Stable manifolds and the transition to turbulence in pipe flow. *J. Fluid Mech.* **627**, 215–233.
- WALEFFE, F. 2001 Exact coherent structures in channel flow. *J. Fluid Mech.* **435**, 93–102.
- WEDIN, H. & KERSWELL, R.R. 2004 Exact coherent structures in pipe flow: travelling wave solutions. *J. Fluid Mech.* **508**, 333–371.
- WILLIS, A.P. 2017 The openpipeflow Navier–Stokes solver. *SoftwareX* **6**, 124–127.
- WILLIS, A.P. 2019a Equilibria, periodic orbits and computing them. [arXiv:1908.06730](https://arxiv.org/abs/1908.06730)
- WILLIS, A.P. 2019b Equilibria, periodic orbits and computing them. [arXiv:1908.06730](https://arxiv.org/abs/1908.06730)
- WILLIS, A.P., CVITANOVIĆ, P. & AVILA, M. 2013 Revealing the state space of turbulent pipe flow by symmetry reduction. *J. Fluid Mech.* **721**, 514–540.
- WILLIS, A.P. & KERSWELL, R.R. 2008 Coherent structures in localized and global pipe turbulence. *Phys. Rev. Lett.* **100**, 124501.
- WILLIS, A.P., SHORT, K.Y. & CVITANOVIĆ, P. 2016 Symmetry reduction in high dimensions, illustrated in a turbulent pipe. *Phys. Rev. E* **93**, 022204.
- YASUDA, T. & LUCAS, D. 2024 Stabilising nonlinear travelling waves in pipe flow using time-delayed feedback. [arXiv:2406.15588](https://arxiv.org/abs/2406.15588)
- YOUNG, C.D., CORONA, P.T., DATTA, A., HELGESON, M.E. & GRAHAM, M.D. 2023 Scattering-Informed Microstructure Prediction during Lagrangian Evolution (SIMPLE)—a data-driven framework for modeling complex fluids in flow. *Rheol. Acta* **62** (10), 587–604.
- ZELIK, S. 2022 Attractors: then and now. [arXiv:2208.12101](https://arxiv.org/abs/2208.12101).