Autoencoder-based dimensionality reduction of turbulent channel flow under spanwise wall oscillations

Lou Guérin, Thomas Fisk, Laurent Cordier, Cédric Flageul, Lionel Agostini

Autoencoder-Based Dimensionality Reduction of Turbulent Channel Flow Under Spanwise Wall Oscillations

Lou Guérin, Thomas Fisk, Laurent Cordier, Cédric Flageul, Lionel Agostini

Pprime institute, Curiosity team, CNRS, Université de Poitiers, Poitiers, France.

Abstract

This study investigates the efficacy of AutoEncoder (AE) neural networks in reducing the dimensionality of a heated turbulent channel flow subjected to Spanwise Wall Oscillations (SWO) at $\mathrm{Re}_\tau = \mathbf{200}$. The dataset, generated using implicit Large Eddy Simulations (LES), comprises velocity and temperature fields at various oscillation amplitudes. Convolutional AutoEncoders (CAEs) are compared to the benchmark Proper Orthogonal Decomposition (POD) method to assess flow complexity and quantify improvements in reconstruction accuracy. $\boldsymbol{\beta}$-Variational AutoEncoders ($\boldsymbol{\beta}$-VAEs) are employed to enhance the orthogonality of the latent samples, aiming to approach the uncorrelated modes of POD. The results demonstrate that CAEs consistently outperform POD in identifying a more compact coordinate system, capturing smaller flow scales with greater accuracy, especially for the wall-normal velocity component. Despite significant pruning of the latent variables, the $\boldsymbol{\beta}$-VAE achieves a fine balance between the orthogonality of POD and the reconstruction capabilities of the CAE. Compared to POD, autoencoders provide superior separability of latent samples in two distinct aspects: (i) more effective amplitude separation, and (ii) more physically coherent phase separation. The CAE in particular leads to the best performance in both respects. This research highlights the potential of autoencoders in dimensionality reduction for complex turbulent flows and provides insights into the trade-off between reconstruction accuracy and latent-space interpretability.

Keywords: Dimensionality reduction, autoencoders, proper orthogonal decomposition, large eddy simulations, turbulent channel flow, spanwise wall oscillations.

1 Introduction

Societal Context

Wall-bounded turbulent flows are prevalent across engineering systems, from aerodynamic applications to heat exchangers, where their behaviour fundamentally determines system performance. In these flows, turbulence manifests through enhanced momentum mixing in the near-wall region, resulting in increased friction drag along solid surfaces. This drag force becomes more pronounced at higher Reynolds numbers, significantly impacting energy consumption in numerous applications [1]. The transportation sector, in particular, faces substantial efficiency losses due to friction drag, affecting both external aerodynamics and internal flows [2]. As environmental concerns intensify and regulations on transport-related emissions become more stringent, the mitigation of friction drag has emerged as a critical engineering challenge [3]. Parallel to drag reduction, the enhancement of heat transfer efficiency presents another crucial engineering objective, particularly in thermal management systems. While turbulent flows naturally promote heat transfer through increased near-wall mixing [4], the traditional coupling between momentum and thermal transport often results in a compromise: attempts to reduce drag typically lead to decreased heat transfer rates, and vice versa. This intricate relationship necessitates innovative control strategies that can effectively manage both phenomena.

Spanwise wall oscillations strategy

The near-wall region of turbulent flows exhibits distinct coherent structures that significantly influence momentum mixing and heat transfer processes. Understanding these structures and their interactions is fundamental for developing effective flow control strategies. Spanwise wall oscillations (SWO) have emerged as a promising control technique, capable of modifying these coherent structures [1]. By generating a spanwise strain layer, SWO disrupts the formation of near-wall streaks, attenuates quasi-streamwise vortices, and increases viscous sublayer thickness [5–7]. The effectiveness of this control approach depends on multiple parameters, including oscillation amplitude (W^+) and period (T^+). Numerical and experimental studies have demonstrated drag reduction of up to 40-50% at low Reynolds numbers [8, 9].

A study by Guérin et al. [10] has demonstrated that particular oscillation parameters ($T^+ = 500$, $W^+ = 30$) can produce heat-transfer enhancement at approximately double the rate of drag increase at $Pr = 1$, indicating a breakdown of the Reynolds analogy. The recent work of Rouhi et al. [11] systematically examined the influence of oscillation frequency and Prandtl number for both SWO and streamwise travelling waves, at fixed $Re_\tau = 590$ and $W^+ = 12$, with particular focus on simultaneous reduction of both heat transfer and drag. Their results reveal that for $Pr > 1$, heat transfer reduction can exceed drag reduction significantly, for example reaching a 10 % superior reduction at $Pr = 7.5$. The authors employ the Stokes layer protrusion height ($\ell^\star_{0.01}$), originally introduced by Rouhi et al. [12] as the elevation where $\overline{\widetilde{w}^2}^\star$ drops to about 1 % of the spanwise turbulent stress, rather than the conventional Stokes layer thickness [13]. This choice is fundamentally motivated by the protrusion height's near

proportionality with the maximum Stokes layer production - the sole external source term that directly injects energy into the turbulent stress budgets. As the oscillation frequency decreases, the Stokes layer becomes more protrusive. In this case, when $Pr > 1$, the conductive sublayer is thinner, thus exposing the energetic small scales of $\Theta' v'$ to larger Stokes layer production. Consequently, these temperature fluctuations gain more energy compared to their velocity counterparts, ultimately resulting in greater reduction of drag than heat transfer.

Reduced-order modelling

The investigation of turbulent flows often relies on Direct Numerical Simulation (DNS), which resolves all ranges of spatio-temporal structures populating the flow. However, for wall-bounded flows, DNS computational resources requirements scale with the Reynolds number Re as Re^3, making it prohibitively expensive, even untenable, for practical applications, particularly when exploring multiple control parameters in flow control strategies such as SWO [6]. This computational limitation has motivated the development of more efficient numerical approaches. Among them, Reduced-Order Models (ROMs), which address this challenge through a two-step process:

- The first one consists of identifying a more efficient coordinate system where the flow dynamics can be represented with fewer degrees of freedom. These purely *data-driven* methods have demonstrated remarkable efficiency in capturing essential dynamics within a low-dimensional manifold. This dimensionality reduction phase aims to capture the essential flow physics within a compact representation. Traditional approaches such as Proper Orthogonal Decomposition (POD) achieve this through linear projections [14], whereas neural network architectures like autoencoders can capture non-linear relationships, potentially leading to more compact representations [15].
- The second step leverages this compact representation to determine the underlying physics governing the flow evolution. This involves identifying relationships between elements in the reduced coordinate system, effectively establishing a simplified dynamical model.

ROMs have become particularly attractive for flow control applications, where computational efficiency is paramount for determining optimal control strategies. For instance, a data-driven ROM, called DManD ROM [16], employs hybrid autoencoders to find a more compact coordinate system for a turbulent minimal Couette flow, reducing the dimension from 10^5 to 18 DOFs. The combination of DManD ROM with the soft actor-critic reinforcement learning algorithm [17] has achieved significant speed-up compared to DNS while maintaining control strategy effectiveness. The distribution of parameters used to create a data-driven ROM plays an essential role in ensuring robustness, both in the dimensionality reduction phase and when predicting flow dynamics [18].

Recent Advances in Neural Network Architectures

Recent advances in neural-network architectures have significantly enhanced ROM capabilities. Traditionally, POD has been the primary technique for dimensionality reduction, generating compact coordinate systems through orthogonal modes that capture the most energetic flow structures [14]. While POD ensures interpretability by providing orthonormal spatial modes and temporal coefficients, its reliance on linear projections presents limitations when addressing complex, multi-scale, nonlinear turbulent systems.

Autoencoders have emerged as a promising alternative [19], with their non-linear activation functions proving particularly effective in capturing coherent structures in turbulent flows [20]. Convolutional autoencoders demonstrate notable effectiveness in detecting spatial patterns and correlations. Studies comparing POD and autoencoder performance [21, 22] consistently demonstrate that autoencoders typically require fewer modes to capture equivalent system energy, though their non-linear nature often results in non-orthogonal modes that reduce model interpretability.

β-Variational autoencoders present a balanced approach between the interpretability of POD and the compact representation of autoencoders [23]. Through the use of a probabilistic encoder and the incorporation of a Kullback-Leibler (KL) divergence term in the loss function, weighted by parameter β, these models encourage mode decorrelation by minimising the divergence between the latent distribution and a standard Gaussian [24]. The latent distribution is thus encouraged to remain close to a centered isotropic distribution, favouring a denser latent space, beneficial for coherent sample generation via latent interpolation [25]. Eivazi et al. [24], Kang et al. [25] each propose an algorithm that ranks latent variables with regards to their contribution to reconstruction, providing a reliable tool for enhancing model interpretability. More recently, there has been a strong focus on enhancing the physical coherence of AEs. For instance, Mohan et al. [26] incorporated the Helmholtz decomposition and fixed neural network weights to enforce incompressibility in the autoencoder reconstruction. Similarly, Fukami and Taira [27] introduced a lift-augmented autoencoder, where the latent space directly encodes lift-related information, achieved by integrating a multi-layer-perceptron that is simultaneously trained to predict lift from the latent representation.

The broader landscape of modern dimensionality reduction encompasses methods such as Koopman operator-based approaches, which seek to linearize nonlinear dynamics [28], as well as sparse regression techniques like the Sparse Identification of Nonlinear Dynamics (SINDy) [29] and its hybrid architecture, SINDy-SHRED [30], which derive interpretable, parsimonious governing equations directly from data.

Manuscript Structure

This study presents a comprehensive investigation of dimensionality reduction techniques, comparing traditional POD approaches with advanced neural network architectures. The analysis focuses on a heated turbulent channel flow described in Section

2. The dataset comprises two-dimensional slices of velocity components and temperature fields, with flow control implemented through spanwise wall oscillations at a period of $T^+ = 500$.

The investigation employs 21 numerical simulations with systematic variations in oscillation amplitude W^+, ensuring robust validation across a range of control parameters. The manuscript structure follows a systematic progression through methodology and results:

- Section 2 presents the numerical framework, detailing flow conditions, simulation parameters and the dataset derived from the simulations.
- Section 3 describes the implementation of POD and autoencoder variants. Particular attention is given to the non-linear capabilities of neural architectures in capturing coherent structures and their potential advantages and disadvantages over linear POD decomposition.
- In Section 4, the performance of CAEs is evaluated against POD, with a focus on reconstruction accuracy, the capacity to resolve different flow scales, and the minimal number of modes required to represent dominant flow structures.
- Section 5 investigates β-VAE architectures, focusing on their ability to provide latent variable pruning and to generate uncorrelated modes through KL-divergence regularisation.
- In Section 6, the aptitude of each AE variants and POD to effectively separate latent samples based on their oscillations amplitudes and phases is assessed.

The primary focus of this study is to establish optimal dimensionality reduction strategies rather than a dynamical model with temporal evolution prediction. This investigation aims to develop compact, physically meaningful representations of turbulent flows under controlled conditions, with particular emphasis on maintaining interpretability in the reduced-order space.

Notations

Table 1 Table of Notations (Part 1)

Symbol	Description
General Mathematical Notation	
$\langle \bullet \rangle$	Ensemble averaging operator
$\overline{\bullet}$	Time and periodic directions averaged quantity ($\overline{\bullet} = \langle \bullet \rangle_{t,x,z}$)
χ'	Fluctuating component of a variable χ ($\chi' = \chi - \overline{\chi}$)
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
Flow Variables & Parameters	
u, v, w	Streamwise, wall-normal, and spanwise velocity components
p	Pressure
Θ	Temperature
Re	Reynolds number (based on bulk velocity U_b)
Re_τ	Friction Reynolds number ($\mathrm{Re}_\tau = Hu_\tau/\nu$)
H	Channel half-height
u_τ	Friction velocity ($u_\tau = \sqrt{\frac{1}{\mathrm{Re}} \frac{\partial \overline{u_\mathrm{w}}}{\partial y}}$)
Θ_τ	Friction temperature ($\Theta_\tau = \frac{\alpha}{u_\tau} \frac{\partial \overline{\Theta}_\mathrm{w}}{\partial y}$)
ν, α	Kinematic viscosity, thermal diffusivity
Pr	Prandtl number
C_f	Friction coefficient ($C_f = \frac{9}{2\mathrm{Re}} \frac{\partial \langle u_\mathrm{w} \rangle_{x,z}}{\partial y}$)
Nu	Nusselt number ($\mathrm{Nu} = 2\frac{\partial \langle \Theta_\mathrm{w} \rangle_{x,z}}{\partial y}$)
$\Delta C_f, \Delta\mathrm{Nu}$	Percentage variation of C_f and Nu with actuation
W^+	Oscillation amplitude in wall units
T^+	Oscillation period in wall units
φ	Oscillation phase
χ_w	Quantity at the wall ($y = 0$, $y = 2H$)
χ^+	Quantity χ normalized in wall units
χ^+_nom	Quantity χ normalized in wall units using unactuated friction quantities
χ^+_ac	Quantity χ normalized in wall units using actuated friction quantities
Performance Metrics	
MSE	Mean Squared Error loss function (10)
$E_r(\chi)$	Energy reconstruction percentage for field χ (16)
$D_\mathrm{KL}[\bullet \,\|\|\, \bullet]$	Kullback-Leibler divergence over two probability distributions
M	Latent variable correlation matrix
$M_{m,n}$	Correlation coefficient between latent variables m and n (17)
Mean_c	Mean Correlation (average off-diagonal of $\lvert M \rvert$) (19)
$\mathrm{MeanMax}_c$	Mean-Max Correlation (20)
D_B	Davies-Bouldin clustering index (21)
$D_B(\mathcal{C}_n, \mathcal{C}_m)$	Davies-Bouldin clustering index between cluster $\mathcal{C}_n$ and $\mathcal{C}_m$ (22)
$C(\phi, LV(i))$	Pearson correlation between physical quantity ϕ and latent variable i (24)

Table 2 Table of Notations (Part 2)

Symbol	Description
Dataset & Indices	
$\chi^{(i)}_{j,k,l}$	Data point for field i, simulation j, snapshot k, grid point l
$i \in \{1,2,3,4\}$	Index for fields: u, v, w, Θ
$j \in \{0,\ldots,20\}$	Index for simulation / oscillation amplitude W^+
$k \in \{0,\ldots,7680\}$	Index for snapshots: subdomain (12) / actuation cycle (20) / actuation phase (32)
$l \in \{1,\ldots,2048\}$	Index for spatial grid point: $(n_x, n_z) = (64, 32)$
$\langle \chi^{(i)}_{j,k,l} \rangle_{k,l}$	Average of dataset samples with field i and simulation j over snapshots k, and grid points l
$\chi^{(i)}_{j,k,l}{}'$	Fluctuations of $\chi^{(i)}_{j,k,l}$ over snapshots k, and grid points l $\left(\chi^{(i)}_{j,k,l}{}' = \chi^{(i)}_{j,k,l} - \langle \chi^{(i)}_{j,k,l} \rangle_{k,l}\right)$
$\widetilde{\chi^{(i)}_{j,k,l}}$	Model reconstruction of $\chi^{(i)}_{j,k,l}$
μ_i, σ_i	Mean and standard deviation of field i, over simulations j, snapshots k, and gridpoints l $\left(\mu_i = \langle \chi^{(i)}_{j,k,l} \rangle_{j,k,l},\ \sigma_i = \sqrt{\langle (\chi^{(i)}_{j,k,l} - \mu_i)^2 \rangle_{j,k,l}}\right)$
$\chi^{(i)}_{\mathrm{std},j,k,l}$	Field wise standardized data (6)
$\chi^{(i)}_{\mathrm{norm},j,k,l}$	Final normalized version data (7)
Dimensionality Reduction Models	
L_d	Latent space dimension
$\boldsymbol{z}$	Latent space vector
$\boldsymbol{\mu}, \boldsymbol{\sigma}$	Latent mean and standard deviation vectors (VAE)
$z_{j,k}(n)$	n-th coordinate of the latent vector with simulation index j and snapshot index k
$\langle z_{j,k} \rangle_{j,k}$	Average of the latent vectors over simulations j and snapshots k
$z'_{j,k}$	Fluctuations of latent sample $z_{j,k}$ over simulations j and snapshots k $\left(z'_{j,k} = z_{j,k} - \langle z_{j,k} \rangle_{j,k}\right)$
U, Σ, V	SVD matrices (left singular vectors, singular values, right singular vectors)
$q_{\boldsymbol{\psi}}(\boldsymbol{z}\|\boldsymbol{x})$	Encoder's approximate posterior distribution (VAE)
$q_{\boldsymbol{\psi}}(\boldsymbol{z})$	Marginal latent distribution (VAE)
$p_{\boldsymbol{\theta}}(\boldsymbol{x}\|\boldsymbol{z})$	Decoder's likelihood distribution (VAE)
$p_{\boldsymbol{\theta}}(\boldsymbol{z})$	Prior distribution (standard Gaussian $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$)
$p_{\boldsymbol{\theta}}(\boldsymbol{x})$	Underlying distribution

2 Heated turbulent channel flow under spanwise wall oscillation

This section provides detailed information on the flow conditions and numerical methods used to generate the dataset. Furthermore, it outlines the pre-processing steps and explains the rationale behind the decisions that shaped the dataset structure.

2.1 Flow conditions

The flow conditions for the present study are comprehensively described in Guérin et al. [10, 31], and are briefly outlined here for convenience. Large Eddy Simulations of

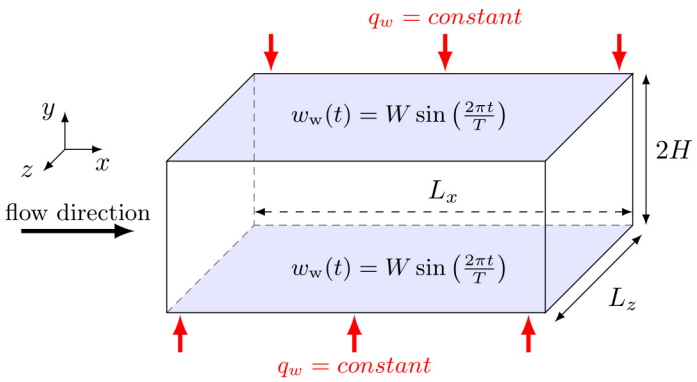


Fig. 1 Schematic of the flow configuration.

a canonical channel flow are carried out at $\mathrm{Re}_\tau = \frac{Hu_\tau}{\nu} \approx 200$ where u_τ is the friction velocity and ν the kinematic viscosity. The numerical domain is defined as:

$$\Omega = \left\{ \boldsymbol{x} = \begin{pmatrix} x & y & z \end{pmatrix}^{\mathsf{T}} \in \mathbb{R}^3 \;\middle|\; x \in [0, L_x],\ y \in [0, 2H],\ z \in [0, L_z] \right\},$$

where x, y, and z represent the streamwise, wall-normal, and spanwise directions, respectively (see Fig. 1). The velocity, pressure, and temperature fields are denoted by $\boldsymbol{u}(\boldsymbol{x}, t) = \begin{pmatrix} u(\boldsymbol{x}, t) & v(\boldsymbol{x}, t) & w(\boldsymbol{x}, t) \end{pmatrix}^{\mathsf{T}}$, $p(\boldsymbol{x}, t)$, and $\Theta(\boldsymbol{x}, t)$.

The evolution of the velocity $\boldsymbol{u}$ is governed by the incompressible Navier-Stokes equations, while the evolution of the temperature Θ follows the passive transport equation of a scalar by the velocity field $\boldsymbol{u}$. Assuming a flow density equal to 1 and adopting the non-dimensional framework of Guérin et al. [10], the following set of governing equations is obtained. All quantities are expressed in dimensionless form, and thus no physical units are associated with them.

$$\begin{cases} \nabla \cdot \boldsymbol{u} = 0 \\ \dfrac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p + \dfrac{1}{\mathrm{Re}} \Delta \boldsymbol{u} + \boldsymbol{f} \\ \dfrac{\partial \Theta}{\partial t} + \boldsymbol{u} \cdot \nabla \Theta + f_\Theta = \dfrac{1}{\mathrm{Re\ Pr}} \Delta \Theta, \end{cases} \tag{1}$$

where $\mathrm{Re} = \frac{HU_b}{\nu}$, with U_b with the bulk velocity . Pr denotes the Prandtl number, which quantifies the ratio of momentum to heat diffusion. To simplify the problem and promote initial similarity between temperature and streamwise velocity, the Prandtl number is set to 1. To further promote this unactuated similarity between thermal and velocity components, the mixed boundary conditions [32] are employed for the temperature. By definition, χ_w represents the boundary conditions imposed on any variable χ on the walls, *i.e.* $\chi_\mathrm{w} = \{\chi(\boldsymbol{x}) \mid x \in [0, L_x],\ z \in [0, L_z]$ and $y = 0$ or $y = 2H\}$. Periodicity is assumed in the x and z directions, while the top and bottom wall satisfy no-slip

boundary conditions. As mentioned in Section 1, in the actuated case, the spanwise velocity is set to an in-phase periodic oscillation, varying in time. This control action depends on two parameters, the oscillation period T, and the amplitude W:

$$\begin{aligned} \Theta_{\mathrm{w}} &= 0, \\ \mathbf{u}_{\mathrm{w}}(t) &= \left(0,\ 0,\ W \sin\left(\tfrac{2\pi t}{T}\right)\right)^{\mathsf{T}}. \end{aligned} \tag{2}$$

Variables denoted with a superscript $+$ indicates values scaled by wall units, *i.e.* non dimensionalized by the friction velocity $u_\tau = \sqrt{\frac{1}{\mathrm{Re}}\frac{\partial \overline{u_{\mathrm{w}}}}{\partial y}}$, where $\frac{\partial \overline{u_{\mathrm{w}}}}{\partial y}$ is the wall-normal derivative of the averaged streamwise velocity at the wall, and the friction temperature $\Theta_\tau = \frac{\alpha}{u_\tau}\frac{\partial \overline{\Theta_{\mathrm{w}}}}{\partial y}$, where α is the thermal diffusivity and $\frac{\partial \overline{\Theta_{\mathrm{w}}}}{\partial y}$ is the derivative of the averaged wall temperature. Due to the periodic time-dependence introduced by the wall oscillations, two distinct variants of wall units are employed for the non-dimensionalization of the actuated flow. These are distinguished by the friction velocity (or temperature) used for scaling: χ^+_{nom} uses the nominal friction velocity from the unactuated flow ($u_{\tau,\mathrm{nom}}$), providing a fixed reference, while χ^+_{ac} uses the instantaneous friction velocity of the actuated flow ($u_{\tau,\mathrm{ac}}$), which captures the modulated near-wall dynamics.

The friction coefficient C_f and Nusselt number Nu characterize drag and heat transfer, respectively. In this configuration, these quantities are defined by:

$$C_f(t) = \frac{9}{2\mathrm{Re}}\frac{\partial \langle u_{\mathrm{w}}\rangle_{x,z}}{\partial y}, \quad \mathrm{Nu}(t) = 2\frac{\partial \langle \Theta_{\mathrm{w}}\rangle_{x,z}}{\partial y}. \tag{3}$$

In Guérin et al. [10], the parameters $T^+ = 500$ and $W^+ = 30$ led to a preferential increase in Nu over C_f. The variation of these quantities induced by SWO can be expressed as a percentage in the following quantity:

$$\Delta C_f = 100 \times \left(\frac{\langle C_{f,\mathrm{ac}}\rangle_t}{\langle C_{f,\mathrm{unac}}\rangle_t} - 1\right), \tag{4}$$

where $C_{f,\mathrm{ac}}$ and $C_{f,\mathrm{unac}}$ account for the friction coefficient of the controlled and unactuated flow, respectively. ΔNu is defined analogously, using $\mathrm{Nu}_{\mathrm{ac}}$ and $\mathrm{Nu}_{\mathrm{unac}}$.

2.2 Numerical methods

In this study, the numerical simulations are carried out using the open-source Xcompact3d framework [33–35]. The spatial discretization employs sixth-order compact finite difference schemes, while the time integration is handled by a third-order explicit Runge-Kutta method. To ensure zero velocity divergence, a fractional step method is used, which solves a Poisson equation for the pressure gradient through 3D fast fourier transforms [33]. Given the extensive parameter space investigation required for W^+ variations in this study, implicit LES are employed instead of DNS to lower computational costs. This approach also effectively reduces the dimensionality of the original flow space representation. Details of the coarse mesh utilized in the LES can be found in Table 3, along with DNS comparison references [10, 36, 37].

Table 3 Simulations computational conditions, against references [10, 36, 37].

Cases	(Lx, Ly, Lz)	Δx^+	Δy^+_{min}	Δy^+_{max}	Δz^+	Δt^+	$\text{t}^+_{\text{collect}}$	Re_τ
[36]	$(12.8, 2, 6.4)$	3.0	0.20	5.93	3.0	NA	3960	180
[37]	$(6.4, 2, 3.2)$	1.1	0.05	0.97	1.1	NA	1677	180
[10]	$(24, 2, 6)$	10.68	0.43	6.16	5.34	$1.56\ 10^{-2}$	> 10000	179
S_1	$(8, 2, 6)$	11.0	2.0	15.7	11.0	$7.3\ 10^{-2}$	10000	196

The DNSs conducted by Guérin et al. [10] employ a significantly finer mesh compared to the present S_1 configuration. Specifically, the mesh resolution in the streamwise (Δx), spanwise (Δz), and the two wall-normal directions (Δy_{min} and Δy_{max}) is enhanced by factors of 1.03, 2.06, 4.65, and 2.55, respectively. Additionally, the computational domain extends three times further in the streamwise direction, providing a more detailed resolution of the turbulent structures. Consistent with the methodology of Flageul et al. [38], the S_1 mesh incorporates uniform numerical dissipation across all spatial directions. This dissipation is implemented using a fourth-order accurate spectral vanishing viscosity scheme [39]. The integration of added numerical dissipation, combined with the coarser mesh refinement of the S_1 configuration, classifies these simulations as implicit LES. As validated in Section 2.3, the S_1 mesh configuration leads to results comparable to those obtained from the higher-resolution grid employed in the reference DNS studies.

Each simulation starts with a warm-up period of five oscillation cycles, equivalent to 2500 nominal wall-unit time steps, ensuring that an equilibrium state is achieved. Following equilibrium, an additional 20 oscillation cycles are conducted to gather data snapshots. During the data collection process, 32 snapshots are recorded at regular intervals within each actuation cycle, capturing the flow dynamics comprehensively across 32 distinct phases (see Section 6 for phase separability analysis). All simulations are initiated from a common set of input snapshots, derived from the steady-state of an unactuated simulation using the S_1 configuration.

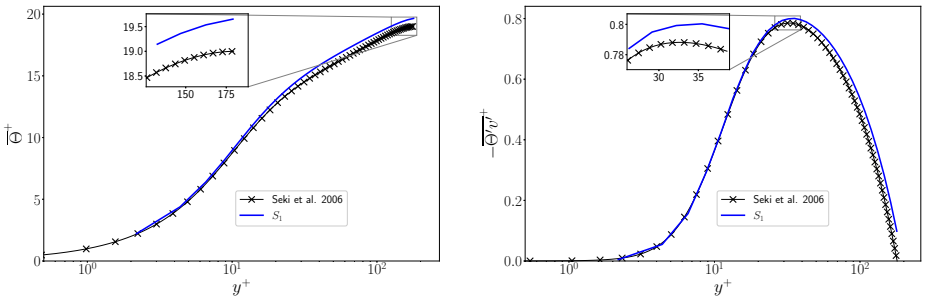
The current numerical strategy balances computational efficiency with the need to accurately capture essential flow features, particularly the interactions between turbulent structures and spanwise wall oscillations.

2.3 Validation

The validation of numerical simulations entails a systematic comparison of unactuated flow statistics against established reference data. Validation of the velocity statistics is already conducted in Guérin et al. [31], whereas the temperature statistics are evaluated against the numerical results of Seki et al. [37]. The assessment of controlled flow characteristics, specifically heat transfer and drag coefficients, is conducted through comparison with the data reported by Guérin et al. [10]. The validation procedure necessitates a precise definition of the averaging methodology and associated flow fluctuations. For this purpose, a generalised decomposition is applied to any flow variable χ (which may represent velocity components, temperature, or pressure) into its mean and fluctuating components. This decomposition is expressed mathematically as follows:

$$\chi(x, y, z, t) = \overline{\chi}(y) + \chi'(x, y, z, t), \tag{5}$$

where $\overline{\chi} = \langle \chi \rangle_{x,z,t}$ represents the time, streamwise and spanwise-averaged value at a specific wall-normal position y, and χ' denotes the fluctuations around this mean.



(a) (b)

Fig. 2 Wall-normal distribution of (a) mean temperature $\overline{\Theta}$, (b) turbulent heat flux $-\overline{\Theta'v'}$. Statistics obtained from simulations on S_1 are represented by the plain blue line. Reference results from Seki et al. [37] are depicted by the black line with crosses.

Fig. 2 presents a comparison between the unactuated simulation statistics and those reported by Seki et al. [37]. The results demonstrate a high level of agreement, with minor discrepancies primarily attributed to differences in Reynolds numbers. It is noteworthy that the discrepancies are more pronounced for the mean temperature $\overline{\Theta}$ compared to the mean streamwise velocity $\overline{u}$ [31], in the vicinity of its peak. However, the turbulent heat flux $-\overline{\Theta'v'}$, which is directly related to the Nusselt number Nu, exhibits minor discrepancies. Specifically, in the unactuated case, the DNS resolution from Guérin et al. [10] at $\mathrm{Re}_\tau = 200$ shows a 4.7% and 4% difference in C_f and Nu, respectively, when compared to the S_1 resolution. These results demonstrate sufficient accuracy, particularly considering the computational efficiency gained through the S_1 mesh configuration (see Guérin et al. 31), thus validating the robustness of the numerical approach.

Table 4 Comparison of the percentage variation in the Nusselt number (ΔNu) and friction coefficient (ΔC_f) between the S_1 mesh configuration and the reference DNS by Guérin et al. [10] for oscillation parameters $T^+ = 500$, $W^+ = 30$.

Configuration	ΔNu	ΔC_f
S_1	18.11%	9.79%
Guérin et al. [10]	14.93%	7.72%

Table 4 presents a comparative analysis of the variations in averaged Nusselt number (Nu) and friction coefficient (C_f) induced by SWO between the S_1 mesh configuration and the DNS results from Guérin et al. [10]. The comparison focuses on results for oscillation parameters given by $T^+ = 500$ and $W^+ = 30$. The S_1 configuration demonstrates strong correlation with the reference DNS data, albeit with marginally enhanced variations in both Nu and C_f. The equilibrium state and transitory characteristics maintain consistency with the reference study.

The LES accuracy was verified for two boundary conditions: the unactuated state ($W^+ = 0$) and maximum actuation amplitude ($W^+ = 30$), for which the Stokes layer protrusion height is at its peak, exerting a more pronounced influence on the flow dynamics, particularly for an actuation period of $T^+ = 500$. In this investigation, simulations will be performed using oscillation amplitudes within the range $W^+ \in [0, 30]$. The demonstrated agreement between LESs and DNSs solutions at these extrema substantiates the capability of the S_1 mesh configuration to accurately resolve the fundamental flow physics under both unactuated and controlled conditions, thereby ensuring reliable predictions of the wall-bounded turbulent flow behaviour throughout the parametric space.

2.4 Dataset

The present investigation utilizes a comprehensive dataset generated through 21 implicit LESs using the S_1 configuration detailed in Table 3. The oscillation period is fixed at $T^+ = 500$. The restriction to amplitude variations whilst maintaining a fixed oscillation period ($T^+ = 500$) represents a deliberate choice grounded in the underlying physics of the flow-control mechanism. As shown by Guérin et al. [10], the combination of $W^+ = 30$ and $T^+ = 500$ yielded a preferential heat-transfer enhancement nearly twice that of the corresponding drag increase. This configuration is not arbitrary but reflects the requisite spatial and temporal scales necessary for effective modulation of near-wall coherent structures.

The oscillation period T^+ governs two fundamental aspects of the control mechanism: the spatial penetration of the Stokes layer and the temporal synchronisation with the self-sustaining streak-formation process, populating the buffer layer. In contrast, the oscillation amplitude W^+ directly controls the intensity of the forcing, i.e. the energetic magnitude of the perturbation applied to the near-wall streaks. Variations in W^+ modulate the strength of streak disruption and reorganisation, which in turn governs the degree of momentum and thermal mixing enhancement. As the amplitude parameter represents a more straightforward control of the forcing intensity acting upon the physically targeted zone already optimised through the selection of $T^+ = 500$, the present parametric investigation is focused on W^+ variations within the range $[0, 30]$ (using a regular subdivision of this interval: $W^+ \in \{x = 1.5 \times i \mid i \in \{0, \ldots, 20\}\}$). This choice permits systematic examination of the relationship between forcing strength and flow modification whilst maintaining spatial and temporal targeting established by the fixed oscillation period.

Furthermore, the amplitude parameter exhibits more predictable and systematic behaviour than the period, rendering it a more suitable focus for a preliminary investigation into model robustness and interpolation capabilities. The decision to vary solely

the oscillation amplitude is thus rooted in the physics of the flow-control mechanism and represents a methodologically sound approach to exploring the parameter space.

Each simulation generated 640 snapshots comprising velocity components u, v, w, and temperature Θ. The data reduction methodology employs a systematic hierarchical approach to efficiently compress the high-dimensional flow field data. Initially, the three-dimensional computational snapshots, encompassing approximately $800,000$ grid points per field, undergo a primary dimensionality reduction by extracting two-dimensional slices at the wall-normal position $y^+_{\mathrm{ac}} \approx 15$ from both channel boundaries. This specific wall-normal location is selected because the small-scale coherent structures, which constitute the predominant turbulent features at the present Reynolds number, exhibit maximum intensity at this distance from the wall. The strategic selection of these wall-normal positions is fundamentally important for characterising the enhancement mechanisms of thermal and momentum transport induced by spanwise wall oscillations. At these locations, the manifestation of important turbulent phenomena—including streak intensity modulation, structural distortions, and alterations in sweep and ejection events—reaches peak intensity [10]. Consequently, predicated on the well-established premise that the two-dimensional slices at $y^+_{\mathrm{ac}} \approx 15$ encapsulate the essential physical information of the flow field, the compression methodology is specifically applied to these planar sections rather than the complete three-dimensional channel domain. This targeted approach confers substantial computational efficiency, significantly reducing the processing requirements associated with the compression of the full three-dimensional flow field. This initial dimensionality reduction diminishes the data volume to $128 \times 96 = 12,288$ grid points per field.

To further enhance computational tractability, a subsequent spatial decomposition partitions the computational domain into six distinct subdomains: two sections along the streamwise direction (x-axis) and three sections along the spanwise direction (z-axis). As evidenced in Fig. 3, this strategic partitioning preserves the integrity of coherent structures within each subdomain, thereby maintaining the essential physical characteristics of the flow. Following this decomposition, the resultant snapshots comprise $64 \times 32 = 2,048$ grid points per field. The four primary flow fields (u, v, w and Θ) are subsequently consolidated into a unified three-dimensional array with dimensions $64 \times 32 \times 4 = 8,192$ degrees of freedom, wherein the third dimension represents the stacked field variables. This integrated approach to dimensionality reduction demonstrates superior efficiency compared to individual field compression methodologies, as comprehensively documented in Section 4.

The complete dataset comprises $161,280$ samples, derived from $7,680$ samples per simulation. This 2D field extraction and domain decomposition strategy, combined with the implicit-LES approach, constitutes the initial phase of dimensionality reduction while maintaining sufficient physical information within the dataset.

The test subset consists of samples from simulations with $W^+ \in \{6, 12, 18, 24\}$, whereas the training subset encompasses the remaining samples, establishing an approximate 80/20 distribution. This partitioning strategy enables assessment of the model's interpolation capabilities across intermediate oscillation amplitudes, thus evaluating its robustness to variations in the control parameter. The dataset structure follows a four-dimensional indexing system where $\chi^{(i)}_{j,k,l}$ represents individual data

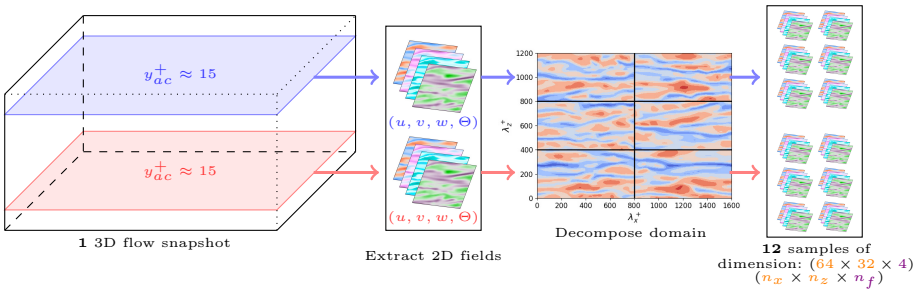


Fig. 3 Processing of 3D snapshots into 2D fields. 2D fields extraction: Slices are taken at $y_{\text{ac}}^{+} \approx 15$ and its upper channel symmetric. Domain decomposion: the x-axis is divided into two segments, and the z-axis is split into three parts. λ_x^+ and λ_z^+ here represent the length (in wall-units) of the domain in the streamwise and spanwise directions respectively.

points. The indices correspond to: field type $i = 1, \ldots, 4$ for u, v, w, Θ, simulation $j \in \{0, \ldots, 20\}$, snapshot number $k = 1, \ldots, n_{\text{samples}}$ (taking into account both time indexes and sub-domains), and spatial position $l = 1, \ldots, 2048$ within each snapshot. The preprocessing stage addresses the inherent scale disparities among the flow fields, as illustrated in Fig. 4(a). The distributions of u and Θ exhibit positive shifts relative to the zero-centred distributions of v and w. The v component displays a particularly concentrated distribution with minimal variance, necessitating careful consideration during the reconstruction process to maintain appropriate weighting among variables.

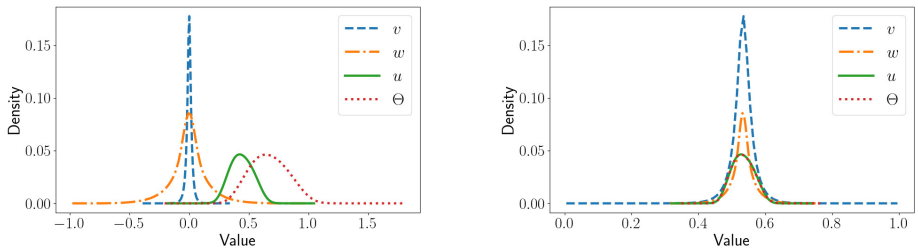


Fig. 4 Dataset Scaling: (a) Distributions Prior to Scaling, (b) Distributions Following Scaling. ——— : u distribution, - - - : v distribution, -.-.- : w distribution, ... : Θ distribution.

The data preprocessing implements a two-stage normalisation procedure. The first one applies field-specific standardisation:

$$\chi_{\text{std},j,k,l}^{(i)} = \frac{\chi_{j,k,l}^{(i)} - \mu_i}{\sigma_i}, \tag{6}$$

where $\mu_i = \langle \chi^{(i)}_{j,k,l} \rangle_{j,k,l}$ denotes the mean and $\sigma_i = \sqrt{\langle (\chi^{(i)}_{j,k,l} - \mu_i)^2 \rangle_{j,k,l}}$ represents the standard deviation for each field i, computed across all simulations from the training subset j, snapshots k and grid points l. The second stage implements a global MinMax normalisation:

$$\chi^{(i)}_{\text{norm},j,k,l} = \frac{\chi^{(i)}_{\text{std},j,k,l} - \min\limits_{i,j,k,l}(\chi^{(i)}_{\text{std},j,k,l})}{\max\limits_{i,j,k,l}(\chi^{(i)}_{\text{std},j,k,l}) - \min\limits_{i,j,k,l}(\chi^{(i)}_{\text{std},j,k,l})}. \tag{7}$$

The resultant distributions, illustrated in Fig. 4(b), exhibit centrality around 0.55 with harmonised variance and bounded values within $[0, 1]$. This normalisation strategy enhances the performance of both POD and AE methods, particularly improving the reconstruction accuracy of the wall-normal velocity component. The normalisation parameters derived from the training dataset are applied consistently across all model implementations. The inverse transformations restore the physical scaling when computing performance metrics. This methodology ensures robust preprocessing while preventing statistical contamination between training and test datasets.

3 Models

This section provides a comprehensive overview of the theoretical foundations and practical implementation details for all the models employed in this study, including POD and all AE variants.

3.1 Proper Orthogonal Decomposition

As outlined in Section 1, POD serves as a benchmark model for this study. POD is a technique used to extract dominant patterns from high-dimensional data by identifying a set of orthogonal modes ranked by their energy contribution [40]. The process involves computing the eigenvalue decomposition of mean-centered data's covariance matrix. Equivalently, the modes can be obtained by singular value decomposition of the data matrix. The singular values of this decomposition represent the energy associated with each mode, while the left singular vectors form spatial basis functions. These modes are ordered by decreasing singular value magnitude, ensuring that the most energetic structures are captured first.

The POD approximation of a function X depending on space x and time t is expressed as:

$$X(x,t) \approx \sum_{i=1}^{L_d} \sigma_i \alpha_i(t) \rho_i(x), \tag{8}$$

where the right hand side represents the L_d-mode approximation of the function X. Here, $\rho_i(x)$ are the spatial basis functions (orthonormal left singular vectors), σ_i are the singular values and $\alpha_i(t)$ are the corresponding temporal coefficients, which also form an orthonormal basis.

The orthogonality of both its spatial modes and temporal coefficients is a key feature of POD, giving a clear interpretation of their individual contributions to the

overall reconstruction. This property is a specific instance of a broader goal in dimensionality reduction: achieving a statistically independent latent space. In this context, a model's *latent samples orthogonality* refers to the orthogonality of the learned latent representations, which implies a statistically uncorrelated structure. A latent space where variables are uncorrelated is often more interpretable and efficient, as each dimension is encouraged to capture a distinct, independent factor of variation. For POD, this is guaranteed by construction through the orthogonality of the temporal coefficients $\alpha_i(t)$.

An other notable advantage of POD is the ranking of the modes with regards to their energy contributions using σ_i which yields additional interpretability. However, achieving convergence may require a large number of modes in scenarios involving turbulent flows or highly nonlinear governing equations. This is particularly relevant for the data considered in this study, where the fields u, v, w, and Θ exhibit complex spatial and temporal dynamics. To apply POD, each sample of shape $64 \times 32 \times 4$ is flattened into an 8192-dimensional vector, representing the spatial domain x in (8). To replicate the testing process used in the AE models, the extended POD approach [41] is employed in this study. This involves projecting the test data onto the POD spatial basis derived from the training data, *i.e.*

$$V_{\text{test}}^{\mathsf{T}} = \Sigma_{\text{train}}^{-1} U_{\text{train}}^{\mathsf{T}} X_{\text{test}}, \tag{9}$$

where V represents the right singular vectors (temporal coefficients), U the left singular vectors (spatial basis), Σ the matrix of singular values and X is the input data matrix. This process compromises the orthogonality of the test data temporal coefficients which no longer form a strict orthonormal basis, as attested in Section 5.

3.2 Convolutional AutoEncoder (CAE)

This section provides a concise overview of AutoEncoders (AEs) and Convolutional AutoEncoders (CAEs), introducing key concepts of convolutional layers, such as kernels and strides. The neural network architecture employed in this study is then detailed, and serves as a foundational framework for all autoencoder models discussed in this paper. The section also examines additional model hyperparameters, including activation functions, learning rate and batch size. An autoencoder is a neural network designed to compress high-dimensional data into a lower-dimensional representation (encoding) and reconstruct the original input from this compressed representation (decoding), see Fig. 5. The compressed representation, referred as the latent variable and denoted by the symbol $\boldsymbol{z}$, is the coordinate vector of the input data in the low-dimensional space. This representation is essential for the reconstruction of the input during the decoding process. During training, the encoding and decoding components are optimized collaboratively using backpropagation algorithms to identify the best low-dimensional representation. This is achieved by minimizing via gradient descent algorithms the difference between the input and its reconstruction, typically quantified using the Mean Squared Error (MSE) loss function.

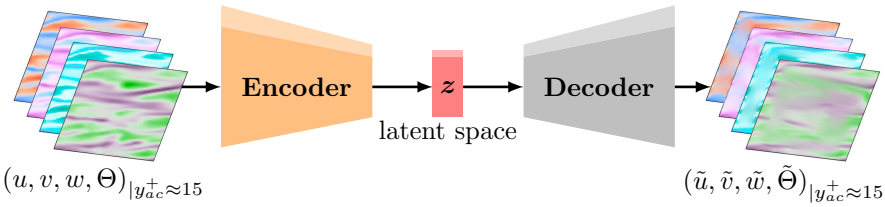


Fig. 5 Autoencoder representation for the reconstruction of the (u, v, w, Θ) 2D fields. z is the latent representation of (u, v, w, Θ), and ($\widetilde{u}$, $\widetilde{v}$, $\widetilde{w}$, $\widetilde{\Theta}$), its reconstruction through the autoencoder.

In this paper, the training process minimizes the MSE loss function. Adopting the same notation as in Section 2.4, where $\widetilde{\chi^{(i)}_{j,k,l}}$ represents the autoencoder reconstruction of grid point $\chi^{(i)}_{j,k,l}$, the MSE loss for a dataset χ is expressed as:

$$\mathrm{MSE}(\chi) = \sum_{i=1}^{4} \langle (\chi^{(i)}_{j,k,l} - \widetilde{\chi^{(i)}_{j,k,l}})^2 \rangle_{j,k,l}. \tag{10}$$

A convolutional autoencoder builds upon the traditional autoencoder architecture by using convolutional layers for encoding and transposed convolutional layers for decoding. In this study, 2D convolutions are applied, treating each input field as an individual channel. In the encoding phase, a 2D convolutional layer applies a small matrix called a kernel to a 2D input field using a sliding window approach. The kernel moves across the input in discrete steps, performing element-wise multiplication with the input region it overlaps, and then summing the results to produce a single value in the output. This operation detects local features such as edges, textures, or patterns in the input data. The kernel typically slides with a specified stride (the step size between applications) and may involve padding (extra values added around the input) to preserve the spatial dimensions of the data (see Fig. 6). The decoder uses transposed convolutional layers, which reverse the convolution operation. This process is often referred as up-sampling and allows the model to reconstruct the spatial details of the original input while maintaining the learned features. By using convolutional layers in both encoding and decoding, CAEs are particularly effective for processing spatial data, such as images or flow fields, e.g. velocity and temperature distributions [22, 42], as they capture hierarchical spatial patterns while efficiently reducing the dimensionality of the input data.

Using zero padding with a stride of 1 and a 3×3 kernel ensures that the center of the kernel passes over every element of the input field, resulting in an output field with the same dimensions as the input. This strategy is applied in layers 1, 2, and 3 of the present encoder architecture (see Fig. 7), where the field size remains fixed at 64×32, while the number of channels (or filters) increases from 4 in the input to 24 in the third layer. The stride of 1 in the first three convolutional layers is designed to capture the fundamental patterns of the flow, without reducing its dimensionality at this stage. In contrast, setting the stride to 2 causes the kernel to skip every other row and column, reducing the spatial dimensions of the output to one-fourth the size of the input. Fig. 6 illustrates this effect, where white blocks are skipped due to the

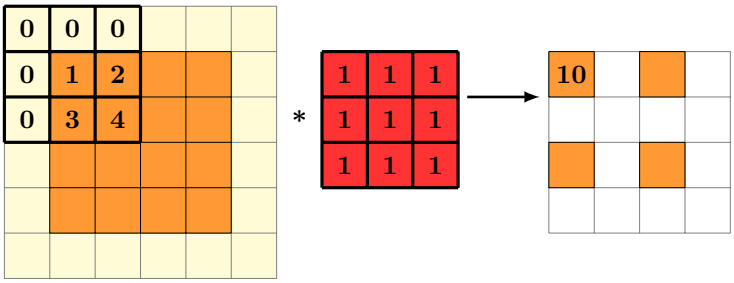


Fig. 6 Example of a convolutional layer with a 3×3 kernel (red matrix), with zero padding (yellow blocks) and stride $= 2$ on a 4×4 input matrix. Orange blocks on the output correspond to locations where the convolutions are performed on the input.

stride of 2, leading to a 2×2 output field. This downsampling strategy is used in layers 4, 5, and 6 of the current model to progressively reduce the spatial dimensions, while capturing the more complex features of the flow. After layer 6, the output tensor is flattened into a 4096-dimensional vector and mapped to a 500-dimensional latent vector via a fully connected layer (see Fig. 7). This aggressive dimensionality reduction accelerates convergence while preserving accuracy, outperforming architectures with intermediate dense layers.

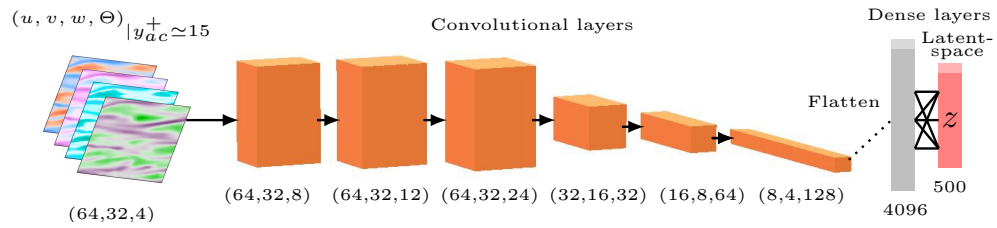


Fig. 7 Architecture of the encoder component of the CAE. Orange blocks represent convolutional layers, while solid arrows indicate batch normalization. Dotted lines represent the flattening layer, and solid lines denote the fully connected (dense) layers.

The decoder architecture implements a hierarchical expansion of the compressed latent representation through a sequence of precisely configured layers. The reconstruction process begins with a fully-connected layer that transforms the latent vector into a 4096-dimensional space, which is subsequently reshaped into a (8,4,128) tensor format. This initial transformation is followed by a series of transposed convolutional operations that progressively restore the spatial dimensions of the flow field through controlled upsampling procedures. The decoder structure exhibits partial symmetry with the encoder (Fig. 5), but introduces a modification in the filter distribution. The implemented architecture employs an asymmetric filter arrangement of $(128, 128, 64, 32, 12, 4)$ across its layers, diverging from the traditional symmetric configuration of $(128, 64, 32, 24, 12, 8, 4)$. This architectural adaptation has demonstrated slightly superior reconstruction capabilities, particularly for the wall-normal

velocity component. This architecture creates an information bottleneck at the latent space, compelling the encoder-decoder pair to operate synergistically in extracting and preserving the most salient flow features necessary for accurate reconstruction.

The neural network architecture incorporates exponential linear unit activation functions [43] throughout its structure, facilitating non-linear transformations within the normalised $[0, 1]$ range established in Section 2.4. The training protocol encompasses:

- 200 epochs with 64 samples per batch
- 3×3 convolution kernels
- Rectified adaptive moment estimation optimisation algorithm [see 44]
- Initial learning rate: 10^{-3}
- Adaptive rate reduction: 50% decrease after 10 epochs without improvement
- Minimum learning rate threshold: 10^{-6}

The latent dimension (L_d) is established at 500 components, ensuring a minimum 80% energy reconstruction threshold (as given by (16)) across all flow fields. This configuration emerges from systematic parameter exploration, demonstrating optimal performance metrics in the reconstruction task.

3.3 β-Variational AutoEncoder (VAE)

Although conventional autoencoders demonstrate robust capabilities in reconstructing complex fluid dynamics, their latent space representations often exhibit significant intermodal correlations, which impede the formation of interpretable basis. The traditional POD method maintains strict orthogonality, but its inherent linear framework limits its compression efficiency. The β-VAE architecture represents a sophisticated solution to the challenge of balancing these competing objectives. The fundamental distinction of VAEs lies in their probabilistic approach to latent space encoding. Rather than implementing deterministic mappings, VAEs encode input data into probability distributions from which latent variables (LVs) are sampled, as depicted in Fig. 8. This probabilistic framework introduces a regulated stochastic element to the encoding process. The present VAE implementation maintains architectural consonance with the previously described convolutional autoencoder (Fig. 7), preserving identical hyperparameter configurations. The distinguishing elements manifest in three critical aspects: (i) the probabilistic nature of the latent distribution, (ii) the stochastic sampling mechanism employed for latent variable generation, and (iii) the modified loss function formulation. This adaptation enables the VAE to maintain reconstruction fidelity whilst encouraging the emergence of more interpretable latent representations.

The encoder models the approximate posterior distribution $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x})$, where $\boldsymbol{x}$ is the input, as a multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma}^2)$. For each sample in the dataset, the encoder produces a mean vector $\boldsymbol{\mu}$ and a standard deviation vector $\boldsymbol{\sigma}$, representing the diagonal elements of the covariance matrix $\boldsymbol{\Sigma}$.

Directly sampling $\boldsymbol{z}$ from a Gaussian distribution would render both the latent variables and the encoder weights non-differentiable, preventing the use of backpropagation. To address this, the reparameterization trick is employed. Instead of sampling $\boldsymbol{z}$

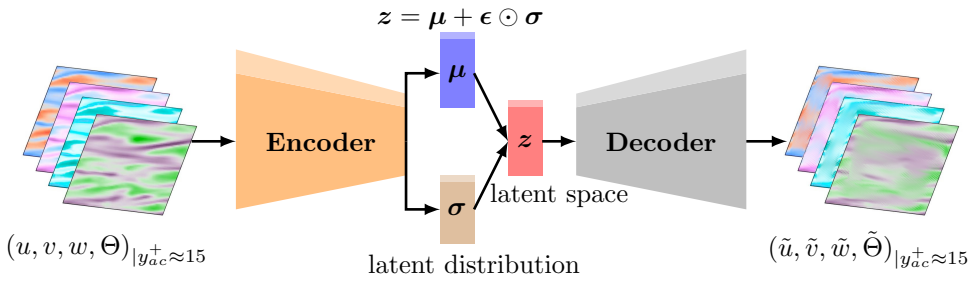


Fig. 8 VAE representation. $\boldsymbol{\epsilon}$ is sampled from a standard Gaussian distribution of zero mean and identity matrix for covariance. $\boldsymbol{z}$ is obtained via a linear combination of the predicted mean $\boldsymbol{\mu}$, and an element-wise product of the sampled $\boldsymbol{\epsilon}$ and the predicted standard deviation $\boldsymbol{\sigma}$.

directly, the encoder outputs the mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$, and $\boldsymbol{z}$ is computed as:

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \odot \boldsymbol{\sigma}, \tag{11}$$

where $\boldsymbol{\epsilon}$ is sampled from a standard normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. This approach shifts the randomness to $\boldsymbol{\epsilon}$, making the loss function differentiable with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, thus allowing effective backpropagation to optimize these parameters.

VAEs are based on the assumption that the input $\boldsymbol{x}$ is a sample from an unknown underlying distribution, denoted as $p_{\boldsymbol{\theta}}(\boldsymbol{x})$, which represents the probability of generating real samples $\boldsymbol{x}$. Introducing conditional probability, $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ is given by

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}) = \int p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}) p_{\boldsymbol{\theta}}(\boldsymbol{z})\, d\boldsymbol{z}, \tag{12}$$

where the prior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{z})$ represents the assumption of the model concerning the distribution of the latent space, while $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})$ describes the probabilistic decoder distribution that generates the data $\boldsymbol{x}$ given the latent variable $\boldsymbol{z}$. The objective is to maximize $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ or equivalently its log-marginal likelihood $\log(p_{\boldsymbol{\theta}}(\boldsymbol{x}))$.

Using Bayes' theorem and the logarithmic properties, the following inequality can be derived [45]:

$$\log(p_{\boldsymbol{\theta}}(\boldsymbol{x})) \geq -D_{\mathrm{KL}}[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x})||p_{\boldsymbol{\theta}}(\boldsymbol{z})] + \mathbb{E}_{q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x})}[\log(p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}))], \tag{13}$$

where D_{KL} denotes the KL divergence, a measure of dissimilarity between two probability distributions. This equation, known as the Evidence Lower Bound (ELBO), provides a computable lower bound on the log-marginal likelihood of the data. By maximizing the ELBO (the right hand side), the log-marginal likelihood of generating real samples, is indirectly maximized. To achieve this, one must minimize the KL divergence between the approximate posterior $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x})$ and the prior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{z})$, while simultaneously maximizing the expected log-likelihood of generating $\boldsymbol{x}$ given $\boldsymbol{z}$. The second term in the ELBO is equivalent to minimizing the reconstruction error (MSE) used in the CAE loss function (10).

The prior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{z})$, which regularizes the latent space and is used in the KL divergence term, is here set to a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Not only is this prior commonly used due to its simplicity and mathematical tractability, but it also enforces the latent space to be dense and orthogonal, thereby reducing correlations between the latent variables. Indeed, minimizing the KL divergence between $p_{\boldsymbol{\theta}}(\boldsymbol{z})$ and $q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x})$ indirectly minimizes the KL divergence between $p_{\boldsymbol{\theta}}(\boldsymbol{z})$ and the marginal latent distribution $q_{\boldsymbol{\psi}}(\boldsymbol{z})$ (see Hoffman and Johnson 46). Strong regularization thus ensures the latent distribution remains close to the isotropic standard Gaussian, encouraging each dimension to encode more independent factors of variation. This reduced correlation between latent variables thus enhances model interpretability. A dense latent space also facilitates smoother interpolations, which is particularly valuable for sample generation tasks. The combination of these characteristics makes the standard Gaussian prior especially suitable for VAEs in both representation learning and generative modeling.

Given the chosen distributions, the KL divergence term in the ELBO simplifies to:

$$D_{\mathrm{KL}}[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x})||p_{\boldsymbol{\theta}}(\boldsymbol{z})] = -\frac{1}{2}\left\langle \sum_{i=1}^{L_d} \left(\log(\sigma_{j,k}(i)^2) + 1 - \mu_{j,k}(i)^2 - \sigma_{j,k}(i)^2\right) \right\rangle_{j,k}, \quad (14)$$

where $\mu_{j,k}(i)$ denotes the mean of the i-th coordinate of the latent distribution, for simulation j and snapshot k. The same notations apply to the standard deviation $\sigma_{j,k}(i)$.

The β-VAE introduces a real number $\beta \in \mathbb{R}^+$, which acts as a parameter to balance the trade-off between orthogonality of the latent samples and reconstruction quality. Given the MSE loss defined in (10), the following loss function is obtained for β-VAEs:

$$\mathrm{Loss}_{\mathrm{VAE}}(\chi,\ \boldsymbol{\mu},\ \boldsymbol{\sigma},\ \beta) = \mathrm{MSE}(\chi) + \beta \times D_{\mathrm{KL}}[q_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{x})||p_{\boldsymbol{\theta}}(\boldsymbol{z})]. \quad (15)$$

This formulation allows for optimization of both the reconstruction accuracy and latent samples orthogonality.

Beyond their improved orthogonality and interpolation properties, β-VAEs often exhibit an automatic mode pruning effect [25]. The regularization term in (15) forces some of the latent distributions to collapse toward the standard Gaussian prior, effectively over-regularizing these variables and minimizing their contribution to reconstruction. Building on this observation, Kang et al. [25] developed a two-stage approach: (i) a ranking algorithm that assesses latent variable contribution through KL divergence analysis, and (ii) a pruning procedure that eliminates non-informative LVs while preserving reconstruction quality. This combined framework applied to β-VAEs can provide a method for determining optimal latent dimensionality, which is particularly valuable for dimension reduction tasks.

3.4 Metrics

To compare the performance of various models with different loss functions, it is necessary to define metrics that allow meaningful evaluations. These metrics can be

separated in three categories: (i) reconstruction of the four fields, (ii) latent samples orthogonality, and (iii) latent samples separability with regards to flow parameters.

For the first category, the metric of interest is the per-field energy reconstruction [23, 24], which measures the energy percentage captured by the model. It is particularly relevant to evaluate the reconstruction of each variable separately, as they are not all reconstructed with the same accuracy, a point that will be discussed in detail in Section 4.

Mathematically, the metric is defined as the squared difference between the fluctuations of input and reconstructed fields, normalized by the squared norm of the input fluctuations. This metric for a field $\chi^{(i)}$, using the same notation as in Section 2.4, is expressed as:

$$E_r(\chi^{(i)}) = 100 \times \left(1 - \left\langle \frac{\sum_{l=1}^{2048} [\chi^{(i)}_{j,k,l}{}' - \widetilde{\chi^{(i)}_{j,k,l}}']^2}{\sum_{l=1}^{2048} [\chi^{(i)}_{j,k,l}{}']^2} \right\rangle_{j,k} \right), \tag{16}$$

where $\chi^{(i)}_{j,k,l}{}' = \chi^{(i)}_{j,k,l} - \langle \chi^{(i)}_{j,k,l} \rangle_{k,l}$ and $\widetilde{\chi^{(i)}_{j,k,l}}$ represents the reconstruction of the grid point $\chi^{(i)}_{j,k,l}$ obtained by the model. The means utilized to calculate the fluctuations are therefore specific to each field i and simulation j.

To evaluate the orthogonality of the latent samples, a correlation matrix M is constructed. The correlation coefficient between two components m and n of the latent vector $\boldsymbol{z}$ is defined as:

$$M_{m,n} = \frac{C_{mn}}{\sqrt{C_{mm} C_{nn}}} \quad \text{for } 1 \leq n, m \leq L_d. \tag{17}$$

C_{mn}, the covariance between modes m and n, is given by:

$$C_{mn} = \left\langle z'_{j,k}(m) z'_{j,k}(n) \right\rangle_{j,k} , \tag{18}$$

where $z_{j,k}(m)$ denotes the m-th latent coordinate of simulation j and snapshot k, and $z'_{j,k} = z_{j,k} - \langle z_{j,k} \rangle_{j,k}$ represents the corresponding fluctuation, obtained by averaging over all dataset latent samples.

Technically, the correlation matrix contains all the information about the orthogonality (or correlation) of each mode with the others. However, extracting a scalar metric from this matrix to represent orthogonality is often more practical, especially for comparing different models. A common approach is to calculate the determinant of the correlation matrix [23, 24], which equals 1 in the case of perfect orthogonality, corresponding to an identity matrix. However, with high-dimensional latent spaces, this metric becomes less reliable, as even modes that are nearly orthogonal can result in a determinant close to zero. This limitation is evident in the current study, affecting all AE models as well as the extended POD method [41].

To address this, two alternative metrics are introduced: the mean correlation (Mean_c) and the mean-max correlation (MeanMax_c). These quantities are defined as follows.

1. Mean Correlation (Mean_c): The mean correlation measures the average off-diagonal coefficients of the correlation matrix in absolute value, providing an overall indication of how correlated the latent modes are with each other. Based on the definition of the covariance coefficients (17) taken in absolute value, it is reasonable to express this metric as a percentage:

$$\text{Mean}_c = 100 \times \langle |M_{i,j}| \rangle_{1 \leq i,j \leq L_d,\ i \neq j}. \tag{19}$$

While Mean_c provides a good overview of the orthogonality of the latent samples , it can sometimes dilute the impact of a few highly correlated modes due to averaging.

2. Mean-Max Correlation (MeanMax_c): To complement the Mean_c metric, the mean-max correlation is introduced. This metric calculates the average of the maximum correlation coefficient (in absolute value) for each mode, helping to assess whether modes are uncorrelated with all other modes in the system.

$$\text{MeanMax}_c = 100 \times \Big\langle \max_{1 \leq j \leq L_d,\ j \neq i} (|M_{i,j}|) \Big\rangle_{1 \leq i \leq L_d}. \tag{20}$$

In a well-structured and orthogonal latent space, both the Mean_c and MeanMax_c metrics should exhibit consistency, with both being close to 0. Together, these two metrics provide a comprehensive view of the degree of correlation in the latent space, ensuring that no significant dependencies among the latent modes are overlooked.

Finally, there remains to evaluate the separability of the latent samples with regards to flow parameters. This refers to the global structure of the latent set, namely, how well it can be partitioned into clusters that correspond to distinct generative factors. In simple flows, for which the data can be compressed into a space of dimension 2-3, visualizing the latent space allows for straightforward evaluation of this separability. However, with the present turbulent flow dataset requiring much larger latent dimensions, direct visualization is impossible, necessitating the use of quantitative metrics to assess this aspect effectively.

The chosen approach consists of selecting all samples corresponding to a specific amplitude (or phase) within the test dataset, and to create a cluster with these samples, defined by its centroid and contained data samples. From these clusters, one can easily apply any clustering metric, taking into account intra or inter cluster variances. A relevant metric that takes into account both of these aspects is the Davies-Bouldin indicator [47], which is calculated as the maximal ratio of the averaged distance between a point and its centroid and the distance between two centroids, averaged over all clusters, *i.e.*

$$D_B = \left\langle \max_{m \neq n} \left(\frac{\delta_n + \delta_m}{\text{dist}(\boldsymbol{c}_n, \boldsymbol{c}_m)} \right) \right\rangle_n. \tag{21}$$

Here, δ_n is the average distance of all the points contained in $\mathcal{C}_n$, cluster of index n, to the centroid $\boldsymbol{c}_n$. This cluster intra-variance is estimated as $\delta_n =$

$\frac{1}{|\mathcal{C}_n|}\sum_{\boldsymbol{z_{j,k}}\in\mathcal{C}_n}\text{dist}(\boldsymbol{z_{j,k}},\boldsymbol{c}_n)$ where $\boldsymbol{z_{j,k}}$ corresponds to a latent sample and $\text{dist}(\boldsymbol{x},\boldsymbol{y})$ measures the euclidean distance between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

The perfect Davies-Bouldin score is 0 which means clusters are well separated, clearly defined and dense. Higher scores mean clusters are poorly defined and overlap.

When evaluating the separability of only two clusters, the Davies-Bouldin indicator reduces to:

$$D_B(\mathcal{C}_n,\mathcal{C}_m)=\frac{\delta_n+\delta_m}{\text{dist}(\boldsymbol{c}_n,\boldsymbol{c}_m)}. \tag{22}$$

Using this metric, one can estimate a matrix D such that $D_{n,m}=D_B(\mathcal{C}_n,\mathcal{C}_m)$, describing how each cluster is separated with the others.

4 Comparing CAE with POD

This section conducts a comparative evaluation of CAE and POD methodologies, focusing on three core objectives: assessing their reconstruction performance, examining their interpolation capabilities at unseen oscillation amplitudes W^+, and investigating the advantages of compressing all fields simultaneously versus processing them individually. Unless otherwise noted, all results are derived from test dataset.

Table 5 Comparison of energy reconstruction for each field, where the AE is trained either to reconstruct all fields simultaneously with $L_d=500$ (Together) or to reconstruct them individually using 125 modes for each (Separately).

	Together (L_d=500)	Separately (L_d=125)
$E_r(u)$	92%	84%
$E_r(v)$	81%	65%
$E_r(w)$	92%	90%
$E_r(\Theta)$	90%	78%

Table 5 highlights the superior reconstruction accuracy achieved by the CAE when trained on all fields simultaneously within a 500-dimensional latent space, compared to independent field training using 125-dimensional latent spaces per field. This comparison is dimensionally consistent, as the cumulative latent dimensionality for individual field processing also totals 500. The simultaneous compression approach improves reconstruction accuracy for the u, v, and Θ fields by 8%, 16% and 12% respectively. This enhancement leverages the strong physical correlations between fields, particularly between the streamwise velocity (u) and temperature (Θ) fields, as expected from the Reynolds analogy. Notably, the wall-normal velocity component (v) consistently exhibits lower reconstruction accuracy across both approaches, a limitation that warrants further investigation. Considering these results, dimensionnality reduction models are systematically applied to the four fields together instead of separately for the rest of this study.

Fig. 9 compares the reconstructed-energy percentage of POD and CAE across various latent dimensions, for each field. The results show that the CAE consistently

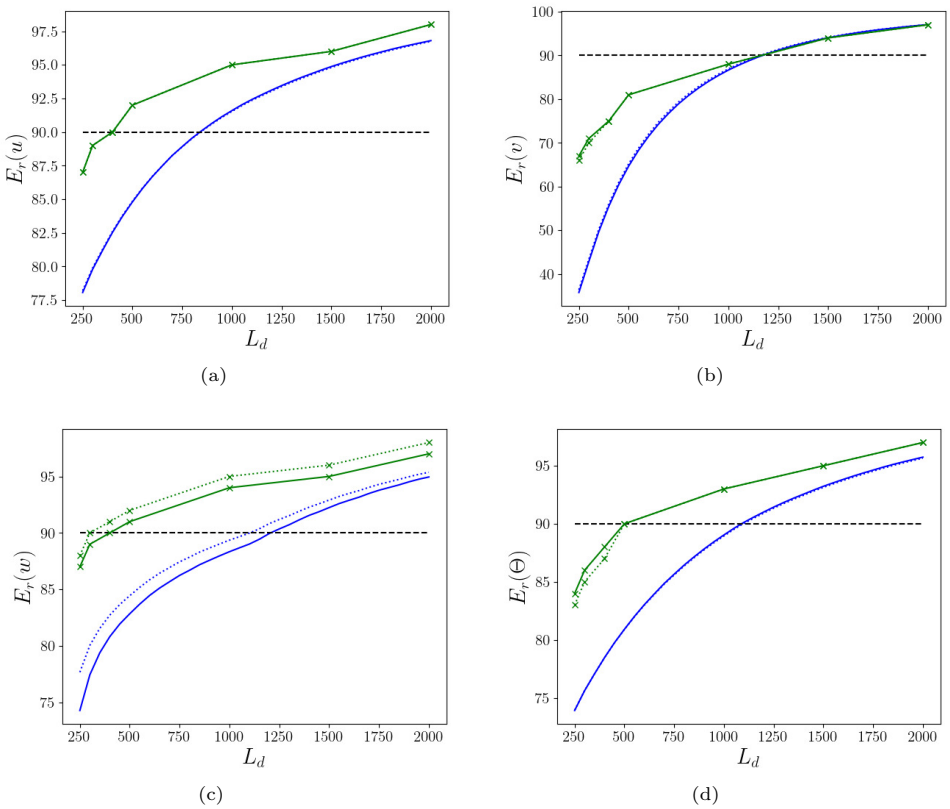


Fig. 9 Energy reconstruction scores (E_r) as a function of the latent dimension (L_d) for extended POD (blue) and CAE (green). Solid lines indicate the training data results, while dotted lines highlight scores on the test data. The fields correspond to the following: (a) Streamwise velocity u, (b) Wall-normal velocity v, (c) Spanwise velocity w, and (d) Temperature Θ.

outperforms extended POD for all L_d-values. Extended POD only matches the performance of CAE on v when $L_d \geq 1000$, with both methods performing similarly for larger dimensions. For each field, the performance gap narrows as L_d increases, with the greatest difference observed at the lowest value of L_d analysed, *i.e.* 250. Notably, this gap is most pronounced for v, where a difference of over 50% in E_r is observed at $L_d = 250$, highlighting the advantage of the non-linear CAE approach, particularly for smaller latent dimensions. Another major result is the number of modes for which each model reaches the 90% reconstruction accuracy benchmark. Excluding the v-field, where both models reach this target around $L_d = 1200$, extended POD requires 2 to 3 times more modes than the CAE to achieve 90% accuracy. This underscores the CAE's superior ability to compress information, as it reaches comparable or better accuracy with significantly fewer modes, making it better suited for this dataset.

As noted earlier, the v-field consistently achieves lower E_r scores than the other three fields. Fig. 9 shows that for $L_d = 250$, $E_r(u)$, $E_r(w)$, and $E_r(\Theta)$ all exceed 80% for the CAE, while $E_r(v)$ remains around 65%. As mentioned previously, this issue

is even more pronounced for POD, further highlighting the challenges specific to the v-field. This is at least partially attributed to the smaller turbulent scales contained in the v-field compared to the others, which require more modes to be captured. The v-component is also initially less energy-dense compared to the other fields, and the applied scaling may not have fully remedied this discrepancy. Additionally, the long tail observed in the scaled v-distribution (Fig. 4(b)) does not contribute significantly to this disparity, as a pre-scaling truncation of this tail did not improve the reconstruction accuracy. Another possible explanation is that SWO enhance sweeps and ejections in the wall-normal direction [10], further expanding the v-distribution in the actuated case.

Table 6 Energy reconstruction scores ($E_r(\chi)$ for $\chi = [u, v, w, \Theta]$), mean-correlation (Mean_c) and mean max-correlation (MeanMax_c) on the test dataset, with $L_d = 500$, on all models. Comparison of results for extended POD, CAE and β-VAE for $\beta = 5 \times 10^{-7}$ (see Section 5).

	CAE	β-VAE	POD
$E_r(\chi)$	[92,81,92,90]	[89,74,89,85]	[85,65,84,81]
Mean_c	2.1	1.1	0.7
MeanMax_c	9.7	6.9	4.1

The selection of latent dimension $L_d = 500$ as the principal case investigated is established to guarantee that the CAE maintains a minimum reconstruction accuracy of 80% across all four target fields when assessed against the testing dataset. The comparative analysis at $L_d = 500$ reveals the superior performance of the CAE relative to extended POD when utilizing an identical number of modes. Specifically, the CAE demonstrates enhanced reconstruction efficiency with improvements of 7%, 16%, 8%, and 9% in E_r for the u, v, w, and Θ fields, respectively (see Table 6). These quantitative results provide substantial evidence for the enhanced capabilities of the non-linear autoencoder methodology compared to its linear counterpart.

Both models demonstrate robust generalization capabilities, as evidenced by the comparable behaviour of training and testing curves across all flow fields in Fig. 9. This characteristic was specifically evaluated through the train-test partition detailed in Section 2.4. The observed higher test scores compared to training scores for the w field are likely attributable to the disparity in dataset sizes, with the training set containing simulations at 17 distinct W^+ values, while the test set comprises only 4 W^+ values. This imbalance in sampling density may lead to better generalization performance on the sparser test dataset.

Fig. 10 offers enhanced insight into the interpolation performance of both CAE and POD (with $L_d = 500$) by disaggregating the reconstruction errors across the four distinct oscillation amplitudes ($W^+ \in \{6, 12, 18, 24\}$) in the test dataset. Notably, reconstruction accuracy varies significantly with amplitude for both methods, with performance gaps in E_r reaching $\approx 7.5\%$ for the CAE compared to $\approx 12\%$ for POD, across all fields except w, where the observed gap is slightly lower. This smaller

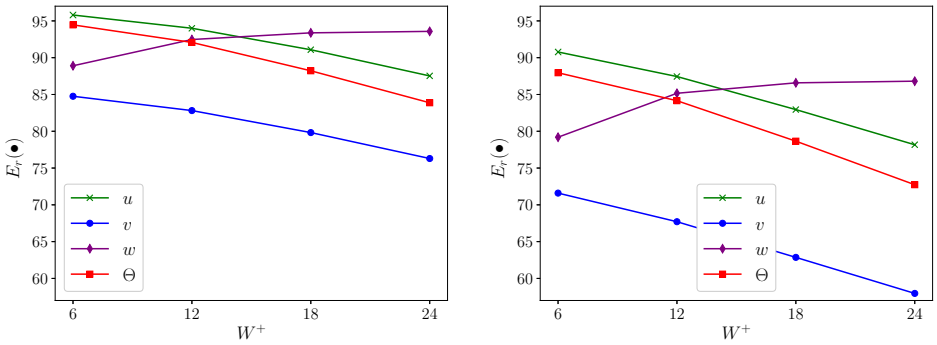


Fig. 10 Energy reconstruction for each W^+ case in the test dataset, with $L_d = 500$. (a): CAE results, (b): extended POD results.

performance variation demonstrates the CAE's superior robustness to changes in actuation amplitude. A clear difference is observed between the reconstructed fields: while reconstruction of u, v, and Θ progressively deteriorates with increasing W^+, w reconstruction conversely improves. This dichotomy aligns with physical expectations, as higher oscillation amplitudes transform near-wall w fluctuations into smoother periodic motions, whereas lower amplitudes retain more stochastic behavior characteristic of the unactuated case. The degradation in reconstructing other fields with increasing W^+ suggests that despite the more periodic nature, larger amplitudes introduce additional flow complexity, consistent with the rising Re_τ observed by Guérin et al. [10] at $W^+ = 30$.

The spatial reconstructions depicted in Fig. 11 present comparative visualisations across the complete LES domains, incorporating sliding averages to maintain subdomain continuity (see Fig. 3). The comparative analysis encompasses Ground Truth representations (left column), CAE reconstructions (centre column), and extended POD reconstructions (right column) for each flow field. These snapshots are picked from the simulation of the test dataset corresponding to the highest oscillation amplitude: $W^+ = 24$. This simulation is characterized by its important small-scale fluctuations and by the largest gap in performances when comparing CAE and POD (see Fig. 10). The Ground Truth data exhibits pronounced, spatially intermittent fluctuations of significant amplitude, characteristic of turbulent small-scale structures. These fine-scale turbulent manifestations present significant reconstruction challenges, necessitating the implementation of 500 modes to achieve satisfactory accuracy.

The v-field exhibits particularly pronounced small-scale structural elements, presenting substantial reconstruction challenges for both methodologies. Its distribution demonstrates a notable centralised tendency (see Fig. 4), which manifests in reduced reconstruction emphasis from both approaches. The extended POD reconstruction demonstrates adequate representation of large-scale flow features. However, the reconstructed flow structures exhibit considerable "smoothing" relative to the Ground Truth, with numerous small-scale features absent from the reconstruction.

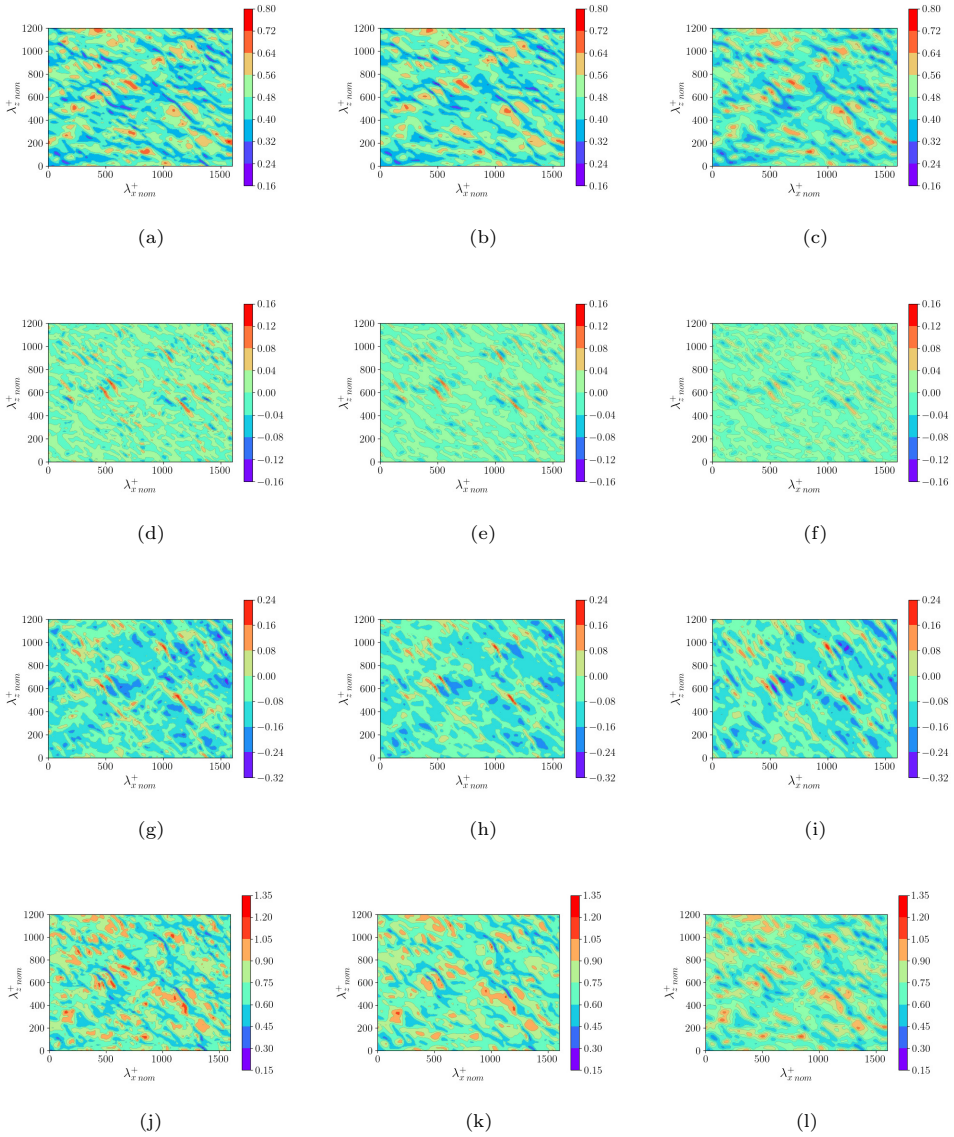


Fig. 11 Reconstruction of snapshots with $L_d = 500$ for the maximal amplitude simulation within the test dataset ($W^+ = 24$): (Left) Ground truth, (Middle) CAE reconstruction, and (Right) extended POD reconstruction. The four fields are displayed from top to bottom: u, v, w, and Θ. Each snapshot represents the full LES domain.

Visual evidence of the ongoing challenge faced by POD to adequately reconstruct small-scale fluctuations in the flow is provided by the Power Spectral Densities (PSDs) in the spanwise direction represented in Fig. 12 for all components. Pre-multiplied PSDs are compared for the ground truth data (black solid lines) of the test dataset

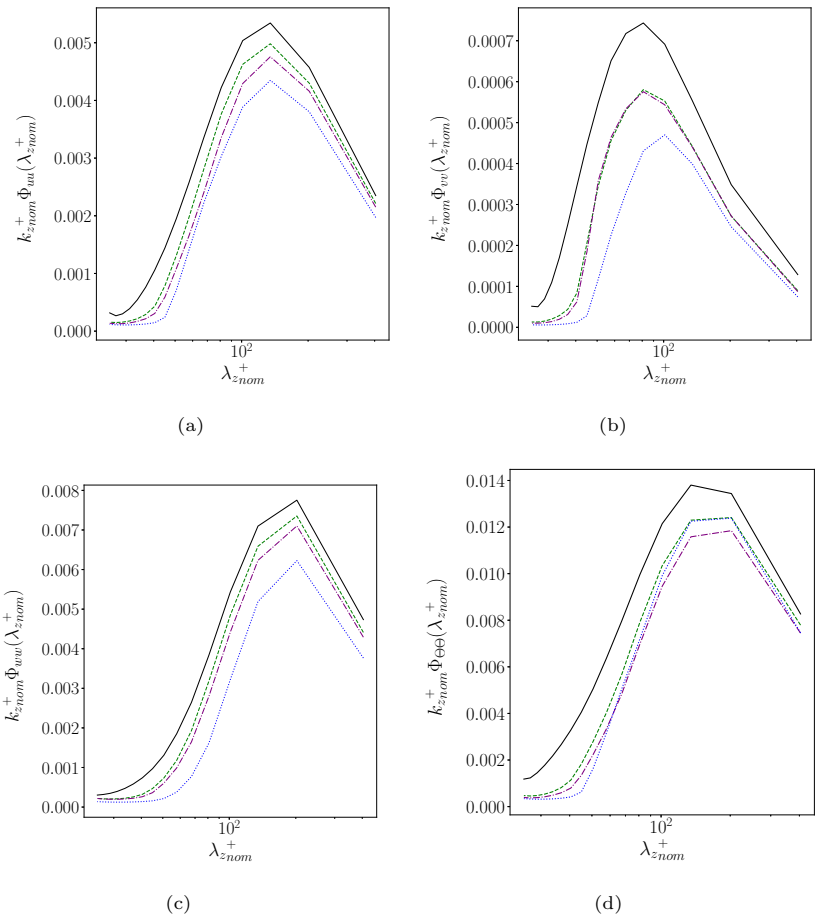


Fig. 12 Pre-multiplied Power spectral density $k_z \Phi_{\chi^{(i)}\chi^{(i)}}(\lambda_z)$ of (a) u, (b) v, (c) w, (d) Θ, for the maximal amplitude simulation within the test dataset ($W^+ = 24$). The spectras are determined on the subdomains of Fig. 3 and not on the full LES-domain. Black solid lines account for the Ground Truth data, green dashed lines for the CAE estimation, blue dotted lines for the extended POD predictions, and purple dash-dotted lines for the β-VAE predictions.

simulation with oscillation amplitude $W^+ = 24$, CAE reconstruction (green dashed lines) and extended POD reconstruction (blue dotted lines). For all fields, the blue curve remains beneath the black curve for all flow scales, with the gap narrowing for the larger scales of the flow, resulting in POD not capturing the total energy of the system. The model particularly fails to capture the smaller scales of the flow, which is further confirmed by the flat left tail of the blue spectras. As anticipated, this phenomenon is even more pronounced for the wall-normal velocity v (see Fig. 12(b)), which exhibits a significantly larger energy gap compared to the other fields. This behavior is expected, as a substantial portion of the energy in v is concentrated in its smaller-scale structures. This energy distribution likely explains the comparatively

lower reconstruction scores for the v-field, as accurately capturing these fine-scale features presents a greater challenge. This aligns with the theoretical foundation of POD, where truncating the modes inherently excludes the lower-energy modes associated with these fine scales. Furthermore, the reconstructed snapshots reveal that extended POD struggles to capture extreme values within flow structures, demonstrating a tendency towards mean-value approximation. These constraints highlight the inherent limitations of linear methods for representing turbulent flows, where small-scale dynamics are crucial for an accurate reconstruction.

The CAE reconstruction methodology demonstrates enhanced capability in capturing localised variations (see Fig. 11), preserving these features with increased fidelity. The reconstruction quality is substantial across the majority of fields, with superior representation of small-scale non-linear features compared to the POD approach. This difference in model performances is confirmed by the spectral analysis of Fig. 12, in which the CAE consistently reconstructs all the flow scales up to $\lambda_{z_{nom}}^{+} \approx 200$ more accurately than POD. Intriguingly, the POD performances are closer to those of the CAE on the temperature field, as they provide similar estimations for wavelengths $\lambda_{z_{nom}}^{+} > 50$. The CAE demonstrates significantly lower performance in the v field in comparison to the other components, primarily due to the inherent small-scale characteristics of this field. Nevertheless, the wall-normal velocity is characterised by the largest discrepancy between the POD and CAE spectra, further demonstrating the superiority of the latter approach in reconstructing turbulent flows. These improvements are attributed to the architectural implementation of convolutional layers (see Section 3.2) in conjunction with non-linear activation functions, specifically designed for the detection and preservation of localised, non-linear features in spatial data structures. Despite the generally satisfactory reconstruction performance, regions characterised by significant noise, continue to present reconstruction challenges for the CAE methodology.

The CAE demonstrates superior performance in reconstruction relative to extended POD. However, it exhibits suboptimal results in terms of latent samples orthogonality (see Table 6). Although extended POD is not strictly orthogonal (as explained in Section 3.1), it achieves mean correlation scores (Mean_c) that are three times lower than those of the CAE, and its modes are, at most, less than half as correlated as the CAE latent variables (MeanMax_c). An ideal model should achieve the high reconstruction accuracy of the CAE while producing correlation scores closer to those of extended POD, allowing for more interpretable modes. This balance will be explored further in the next section.

5 Balancing orthogonality of the latent samples and reconstruction accuracy with β-VAE

In this section, the orthogonality of the latent samples is examined by sampling latent variables in a manner that encourages independence and proximity to the standard Gaussian distribution, as opposed to using a deterministic, unstructured approach (see Section 3.3). More precisely, β-VAEs are employed to determine whether a suitable

compromise can be achieved between the orthogonality of extended POD and the reconstruction accuracy of the CAE.

A parametric study on β showed that the value 5×10^{-7} led to the best trade-off between reconstruction and orthogonality. This value was subsequently retained. Table 6 highlights the obtained trade-off. Even with a relatively small β value, the reconstruction accuracy across all fields experiences a slight decrease, underscoring the significant impact of the latent space structure on energy reconstruction and the difficulty of finding an appropriate balance. The orthogonality of the latent samples is largely enhanced comparatively to the CAE, with a decrease of 1% and 2.8% in Mean_c and MeanMax_c, respectively. Interestingly, this reduction of high-correlation relationships in the latent space impacts the reconstruction of v and Θ fields the most, as attested by Table 6. This suggests that the improvements in latent samples orthogonality may come at the expense of reconstruction accuracy for these fields, which is likely linked to the high correlation among all four fields.

Fig. 12 demonstrates that, like the CAE, the β-VAE (purple dash-dotted line) outperforms extended POD in reconstructing the flow, particularly for the smaller turbulent scales and at the energy peak. However, this is not the case for the temperature field, where extended POD achieves better accuracy at the energy peak, aligning with the previously noted decline in β-VAE performance for this field. While the β-VAE performs marginally worse than the CAE in flow scale reconstruction overall, it matches the CAE's performance in the v field for this specific simulation.

The mode ranking algorithm developed by Kang et al. [25] was applied following the completion of model training. As detailed in Section 3.3, latent variables are sorted according to their Kullback-Leibler divergence from the one-dimensional standard Gaussian prior, wherein higher divergence values are indicative of greater information content. Through the regularisation term in (15), latent variables with minimal KL divergence are forced towards the standard Gaussian distribution. As a direct consequence, these over-regularised variables are rendered functionally redundant, contributing negligibly to the reconstruction process. Dimensional reduction through mode pruning is thus mathematically justified, whereby a systematic excision of redundant modes is performed based on their information content. In this pruning procedure, the most informative latent variables (classified as physics-aware in Kang et al. 25) are preserved within the reduced-order representation, while the less informative components (physics-unaware) are systematically eliminated. This dimensional truncation is implemented by fixing the physics-unaware modes to their respective estimated mean values (23a), effectively neutralising their contribution to the flow field reconstruction.

The KL divergence is computed using (14) (applied individually to each latent variable) and the empirically estimated parameters $\widehat{\mu}_n$ and $\widehat{\sigma}_n$ determined as

$$\widehat{\mu}_n = \langle z_{j,k}(n) \rangle_{j,k}, \tag{23a}$$

$$\widehat{\sigma}_n = \sqrt{\langle (z'_{j,k}(n))^2 \rangle_{j,k}}. \tag{23b}$$

This procedure allows the mode ranking and pruning algorithm to be applied to the CAE, despite its deterministic nature.

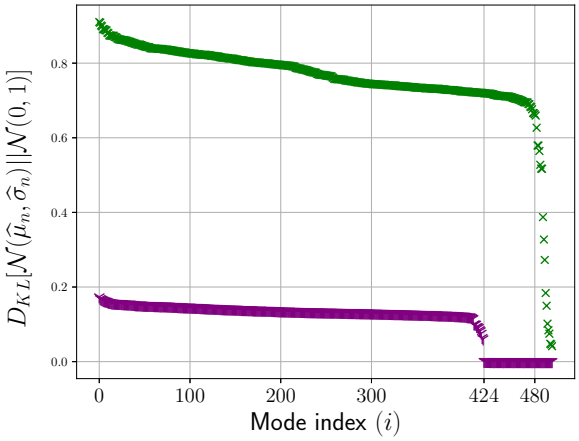


Fig. 13 KL divergence of each latent variable with the 1D standard Gaussian prior, for the CAE (green crosses) and β-VAE (purple Ys), with $L_d = 500$ and ranking of the modes with regards to their KL divergence score.

Fig. 13 displays the KL divergence of each latent variable, ranked in descending order, with CAE results marked by green crosses and β-VAE results by purple Ys. The CAE modes exhibit a KL divergence roughly four times higher than those of the β-VAE in average, highlighting the efficacy of the regularization term in (15). A sharp decline in KL divergence beyond latent variable 424 of the β-VAE separates the physics-aware modes from the physics-unaware ones, suggesting the presence of the pruning effect commonly observed in VAEs. Here, 76 latent distributions are effectively reduced to a standard Gaussian, contributing negligibly to reconstruction. A similar, though less pronounced, KL drop occurs for the CAE at mode 480. However, even the least informative CAE latent variables retain a KL divergence of approximately 0.04, preventing it from being classified as physics-unaware.

Following the methodology of [25], a physics-informed β-VAE can be derived by retaining only the 424 modes with the highest KL divergence. Table 7 confirms that the pruned model achieves reconstruction performance identical to the full model, validating the pruning effect induced by the β-VAE regularization term.

Table 7 Energy reconstruction scores ($E_r(\chi)$ for $\chi = [u, v, w, \Theta]$), mean-correlation (Mean_c) and mean max-correlation (MeanMax_c) on the test dataset, for the CAE with $L_d = 400$, and the pruned β-VAE, using 424 modes.

	CAE_{400}	$\beta\text{-VAE}_{\text{pruned}}$
$E_r(\chi)$	[90,75,91,87]	[89,74,89,85]
Mean_c	1.8	1.3
MeanMax_c	9.4	6.5

Before comparing the models in terms of latent samples orthogonality, it is important to note that the mean correlation (Mean_c) should be interpreted with caution, as it may underestimate the presence of highly correlated modes. Fig. 14(a) shows the correlation matrix of the CAE, with LVs ranked according to their KL divergence. The last 10 to 20 modes, corresponding to the KL divergence drop of Fig. 13, exhibit nearly perfect correlation with each other and are significantly correlated with the remaining modes compared to the average behavior. However, these LVs have a limited effect on the overall Mean_c score, which remains low (2.1%, as shown in Table 6). This highlights the potential of Mean_c to obscure redundancies within the latent space. By contrast, the mean-max correlation (MeanMax_c) offers a more sensitive perspective. This metric captures the maximum correlation for each mode, making it better suited to identify hidden redundancies. For the CAE, the MeanMax_c score (9.7%) reveals significant entanglement among modes that the Mean_c score fails to reflect. The CAE latent space demonstrates clear entanglement, especially for the least contributing LVs, a significant limitation of this architecture. This entanglement implies that features detected by its convolutional layers lack independence and interpretability. Consequently, the CAE risks encoding redundant features, where similar patterns are repeatedly captured across different modes.

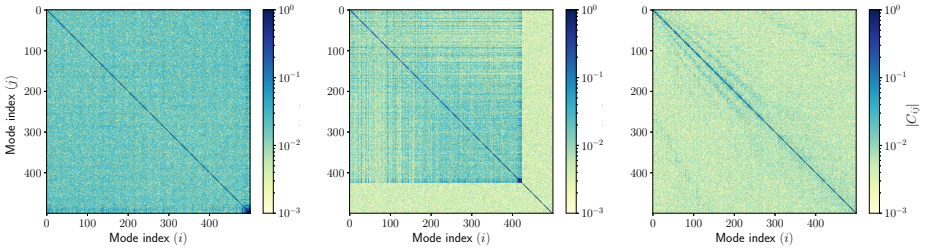
Although the extended POD spatial basis is orthogonal, when validating the results on the test data, the temporal coefficients do not necessarily form an orthogonal basis, as illustrated in Fig. 14(c), which is generated from the temporal coefficients of each sample. The ranking algorithm of Kang et al. [25] is not applied in this case, as POD already organizes its modes by decreasing energy contribution. This ranking leads to modes with similar energy levels being more likely to exhibit correlation due to their tendency to describe similar flow structures. This pattern is reflected in the correlation matrix, where coefficients closer to the diagonal show stronger correlations, while those further away tend to be less correlated, as indicated by the light green regions in the bottom-left and top-right corners. Interestingly, a few secondary blue diagonal lines, both near and distant from the matrix main diagonal, indicate higher than average correlations between modes with significantly different energy contributions. Unlike the CAE, the extended POD does not exhibit dark blue dots outside the diagonal, implying that no mode is highly correlated with another. This is consistent with the MeanMax_c score of 4.1% (see Table 6). As a result, the temporal coefficients of extended POD can be considered near-orthogonal.

The β-VAE correlation matrix (Fig. 14(b)) demonstrates a notable improvement over that of the CAE. Overall, the matrix is characterized by larger green regions, indicating that the correlation coefficients of these modes are significantly damped. This improvement is further validated in Table 6, where the mean correlation (Mean_c) drops from 2.1% for the CAE to 1.1% for the β-VAE. The 76 physics-unaware LVs are clearly visible in the correlation matrix, represented by light green lines that highlight their purely stochastic behavior. Similar to the CAE correlation matrix, the least contributing modes (those just above the pruned threshold) show high mutual correlation, though to a much lesser degree. This observation suggests that orthogonality improvement is not merely a global average effect but occurs at multiple levels.

Supporting this conclusion, the β-VAE's mean max correlation (MeanMax_c) drops to 6.9%, nearly 3 percentage points lower than the CAE's MeanMax_c of 9.7%.

However, one might question to what extent the β-VAE's superior orthogonality stems from its physics-unaware modes, exhibiting very low correlations (see Fig. 14(b)). To investigate this, Table 7 compares the pruned β-VAE with a CAE using $L_d = 400$ (a relevant comparison since the pruned β-VAE utilizes 424 latent variables). While both models achieve comparable reconstruction quality (with the CAE showing slightly better scores), the pruned β-VAE maintains significantly better orthogonality scores, particularly in terms of MeanMax_c. This demonstrates that the β-VAE's regularization achieves more than simple mode pruning, establishing an effective trade-off between reconstruction fidelity and latent samples orthogonality.

A comparison of the β-VAE correlation matrix with the extended POD matrix reveals that the latter model continues to demonstrate superior performance with respect to latent samples orthogonality. The results presented in Tables 6 and 7 substantiate this conclusion. As with the extended POD correlation matrix, discernible patterns are evident in the β-VAE matrix, albeit of a markedly different nature. Indeed, when excluding the physics-unaware LVs, the overall correlation of each remaining LV with other modes appears to decrease as its KL divergence increases.



(a) (b) (c)

Fig. 14 Correlation matrix obtained on the test dataset, with $L_d = 500$, for models: (a) CAE, (b) β-VAE, (c) extended POD. Absolute values are applied to all coefficients. Colors range from light green (uncorrelated) to dark blue (highly correlated).

The results demonstrate that the β-VAE achieves a balanced compromise between latent samples orthogonality, effective W^+ interpolation, and field reconstruction, while not outperforming either the CAE or extended POD in their specific strengths. By applying the mode ranking algorithm from [25], the β-VAE can be successfully reduced to a model with $L_d = 424$, while maintaining a fair trade-off between reconstruction quality and orthogonality. These results highlight the inherent challenges in simultaneously achieving both orthogonal latent variables and precise reconstruction accuracy for complex datasets requiring high latent dimensions.

6 Separability of generative factors within the latent space

The interpretability of the latent space extends beyond the orthogonality of the latent samples and the ranking of latent variables based on reconstruction contribution. An ideal latent space should exhibit effective separability between samples derived from different generative factors of the flow, a property particularly important for generating new data through latent interpolation. Clear distinctions in the latent space enable smoother and more meaningful transitions between interpolated values. In this section, the latent space organisation of each proposed dimensionality reduction model is analysed with regard to the separation of samples with different oscillation amplitudes W^+ or different oscillation phases φ (see Section 2.2 for their definitions).

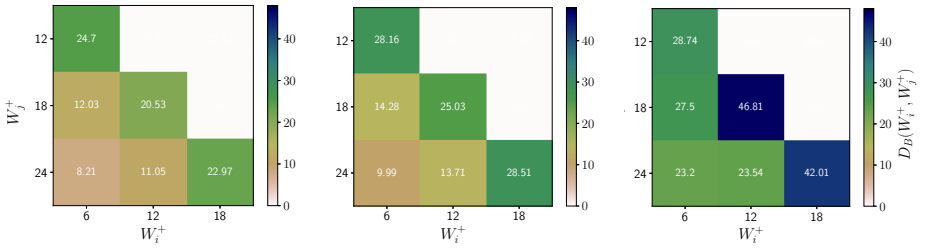
Table 8 Global Davies-Bouldin indicator as given by (21) applied to clusters of oscillation amplitudes ($D_B(W^+)$) and phases ($D_B(\varphi)$), using the test dataset, with $L_d = 500$, on all models. Comparison of results for extended POD, CAE and β-VAE for $\beta = 5 \times 10^{-7}$ (see Section 5).

	CAE	β-VAE	POD
$D_B(W^+)$	23.84	28.35	41.10
$D_B(\varphi)$	27.70	33.90	27.67

Table 8 presents the Davies-Bouldin indicator (21) for the latent spaces of each model, evaluated for both W^+ and φ. Based on this metric, the CAE achieves the best separability with respect to W^+, marginally outperforming the β-VAE, while extended POD produces the weakest clustering performance. For oscillation phases (φ), both extended POD and CAE exhibit better separability than the β-VAE. These results indicate that the latent spaces of each model are structured differently depending on these oscillation parameters, with the CAE demonstrating the most effective overall separation.

The cluster separation of oscillation amplitudes W^+ is investigated in Fig. 15 through the Davies-Bouldin indicator for two clusters W_i^+ and W_j^+ as given by (22).

Higher D_B values, indicating worse separability, are observed for clusters with similar W^+ values across all considered models. This effect is particularly pronounced in the CAE and β-VAE results, where the D_B score decreases dramatically when comparing cases with W^+ gaps of 12 and even more so for gaps of 18. For the autoencoders, the most challenging differentiations occur for $D_B(6, 12)$ and $D_B(18, 24)$, while distinguishing $W^+ = 12$ from $W^+ = 18$ proves relatively easier. This behaviour can be explained through physical interpretation: $W^+ = 12$ represents conditions closer to the unactuated case ($W^+ = 0$) characterised by stochastic fluctuations, while $W^+ = 18$ approaches the fully actuated case ($W^+ = 30$), characterised by smoother periodic fluctuations. Although both autoencoder approaches show similar trends, consistently

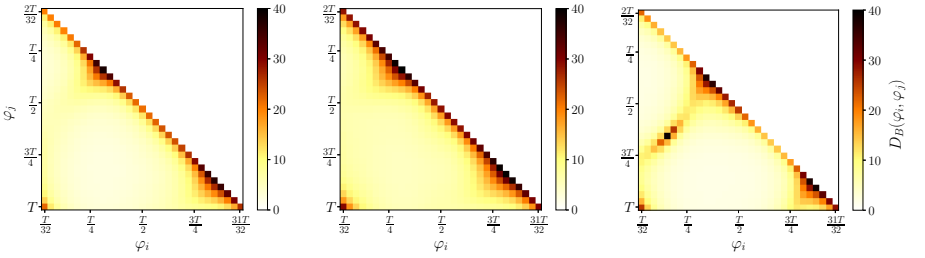


(a) (b) (c)

Fig. 15 Davies-Bouldin indicator for two clusters W_i^+ and W_j^+ ($D_B(W_i^+, W_j^+)$) for the test dataset, with $L_d = 500$, for models: (a) CAE, (b) β-VAE, (c) extended POD. Only the lower triangular part of the matrix is shown due to the symmetry of the operator.

lower coefficients are featured in the CAE matrix, indicating superior overall W^+ separability, a result consistent with Table 8.

A similar yet weaker dependence on W^+ gaps is revealed in the extended POD matrix. For instance, comparable magnitudes are demonstrated by $D_B(6, 12)$ and $D_B(6, 18)$. Different behaviour is shown by extended POD compared to the autoencoders, with $D_B(12, 18)$ proving the most difficult case to differentiate. This further confirms that the latent space is organised fundamentally differently by POD with respect to this generative factor. Overall, significantly inferior W^+ separability is demonstrated by POD compared to both autoencoder approaches.



(a) (b) (c)

Fig. 16 Davies-Bouldin indicator for two clusters φ_i and φ_j ($D_B(\varphi_i, \varphi_j)$) for the test dataset, with $L_d = 500$, for models: (a) CAE, (b) β-VAE, (c) extended POD. Only the lower triangular part of the matrix is shown due to the symmetry of the operator.

As detailed in Section 2.2, each actuation cycle is decomposed into 32 distinct phases in the dataset. All simulations in the test dataset feature actuation ($W^+ > 0$), exhibiting periodic behaviour that justifies phase-based analysis. The dependence on the phase of the oscillation cycle becomes more prominent with higher values of W^+. The separability analysis protocol is applied in Fig. 16 to evaluate phase separability across models.

For the autoencoder approaches, similar trends are exhibited by both models, with the CAE consistently achieving substantially enhanced separability. A pattern analogous to the W^+ case is observed, where poorer separability is demonstrated by temporally adjacent phases compared to distant phases. Notably, the highest intra-group D_B scores are exhibited by phases in the ranges $\varphi \in \left[\frac{T}{4} + \delta_\varphi, \frac{T}{2} - \delta_\varphi\right]$ and their symmetric counterparts $\varphi \in \left[\frac{3T}{4} + \delta_\varphi, T - \delta_\varphi\right]$, where δ_φ represents a small phase offset. These ranges, corresponding to the ascending drag regime (transitioning from $C_{f,\min}$ to $C_{f,\max}$), are distinguished by steeper temporal gradients in wall friction compared to other phases of the cycle, as identified in Guérin et al. [10]. Additionally, the phases $\varphi = \frac{T}{4}$ and $\varphi = \frac{3T}{4}$ correspond to the extremal values of the control signal $w_\mathrm{w}(t)$, where $w_\mathrm{w}\left(\frac{T}{4}\right) = W$ and $w_\mathrm{w}\left(\frac{3T}{4}\right) = -W$, which may explain the poor separability scores observed around these phases.

While the symmetric nature of friction variation arising from the sinusoidal wall oscillation [10] might suggest difficulty in distinguishing phases separated by $\frac{T}{2}$, efficient separation is actually demonstrated in the autoencoder latent spaces for these cases. This suggests that other physical distinctions, such as Stokes strain sign reversal or streak orientation (see Guérin et al. 10), are effectively captured by the autoencoders.

Distinct behaviour is exhibited by the extended POD approach, as evidenced by elevated D_B scores for phases φ near $\frac{T}{8}$ with their symmetric counterparts at $\varphi \approx \frac{5T}{8}$. This observation suggests that difficulties are encountered by the extended POD latent space in distinguishing these phase pairs, which correspond to similar drag and heat transfer characteristics but opposing streak orientations. More broadly, the presence of the corresponding non-diagonal line in Fig. 16(c) reveals that phases $\varphi_1 \in \left[0, \frac{T}{4}\right]$ are not well separated from phases $\varphi_2 \approx \frac{3T}{4} - \varphi_1$. This implies that the extended POD methodology struggles to differentiate between friction-descending phase pairs φ_1/φ_2 originating from opposite halves of the actuation cycle and displaying nearly identical distances from $C_{f,\min}/C_{f,\max}$ phases. These phase pairs are characterised by substantially different streak orientations and opposite Stokes strain directions in the actuated case, which confirms the difficulty of extended POD to effectively capture related flow features within its latent space.

Beyond these distinctions, phase separability trends similar to the autoencoders are demonstrated by extended POD, albeit with better performance for phase pairs outside the challenging regions with ascending drag.

Three primary results are obtained from the separability analysis of generative factors of the flow:

1. Superior performance in separating different W^+ amplitudes is consistently achieved by autoencoder architectures compared to extended POD, with the CAE achieving optimal capability. This suggests that amplitude-dependent flow features are better captured in the latent representations of autoencoders.
2. Similar phase separation characteristics are exhibited by both CAE and β-VAE, showing:
   - Reduced separability in ascending friction phases.
   - Superior overall performance achieved by the CAE.

- Effective discrimination of phases with different streak orientation or opposing Stokes strain direction.

3. The best global $D_B(\varphi)$ score is achieved by extended POD (see Table 8) and excellent separability is shown for certain phase combinations. However, weak separation of part of the descending friction phase-pairs is provided by this method, which differ significantly in terms of streak patterns and Stokes strain.

While extended POD is not matched by autoencoders in terms of latent samples orthogonality, more physically meaningful characterisations of the generative factors of the flow are provided by their learned representations, offering enhanced interpretability. The best performance overall was achieved by the CAE, especially with regard to oscillation phases. The approach to investigate these separabilities remains basic, considering one cluster for each phase or amplitude. Deeper insights into the separability of SWO amplitudes and phases, as well as mode correlations, may be provided by employing Cluster-based Reduced-Order Modelling (CROM) as proposed by Kaiser et al. [48]. Additionally, new information about the interactions between C_f and Nu, particularly concerning the breakdown of the Reynolds analogy, may be revealed. This extended cluster-based analysis remains a topic for future investigations.

7 Latent variable interpretability

This study has demonstrated that autoencoders outperform POD in both reconstruction quality and the separability of the flow's generative factors within the latent space. To further assess interpretability, a preliminary exploratory analysis is conducted wherein correlations are examined between the latent variables and four physical quantities: the friction coefficient C_f, the Nusselt number Nu, the oscillation phases φ and the amplitudes W^+. These quantities are of particular interest, as the present actuation strategy resulted in preferential heat-transfer enhancement at $W^+ = 30$ [10].

7.1 Methodological Limitations

It is essential to acknowledge a fundamental constraint of this analysis: the Pearson correlation coefficient employed herein measures exclusively linear relationships, whereas the principal advantage of autoencoders over POD resides in their capacity to capture non-linear projections. This limitation is particularly pronounced for the CAE, whose training objective (minimisation of reconstruction error) imposes no constraints favouring disentangled representations. Consequently, physical information is expected to be distributed across multiple latent variables in complex, non-linear patterns that linear correlation analysis is inherently unable to detect. The absence of strong linear correlations thus does not indicate the absence of encoded physical information, but rather reflects the inadequacy of the linear probing methodology for non-linear latent spaces.

To quantify linear relationships, the Pearson correlation coefficient is computed across all dataset samples. For a physical quantity ϕ (W^+, φ, C_f, or Nu), the correlation between ϕ and the i^{th} latent variable is given by:

$$C(\phi, LV(i)) = \frac{\left\langle z'_{j,k}(i)\phi'_{j,k} \right\rangle_{j,k}}{\sqrt{\left\langle z'_{j,k}(i)z'_{j,k}(i) \right\rangle_{j,k} \left\langle \phi'_{j,k}\phi'_{j,k} \right\rangle_{j,k}}}, \tag{24}$$

where the averaging and fluctuations over simulations j and snapshots k are defined as in Section 3.4.

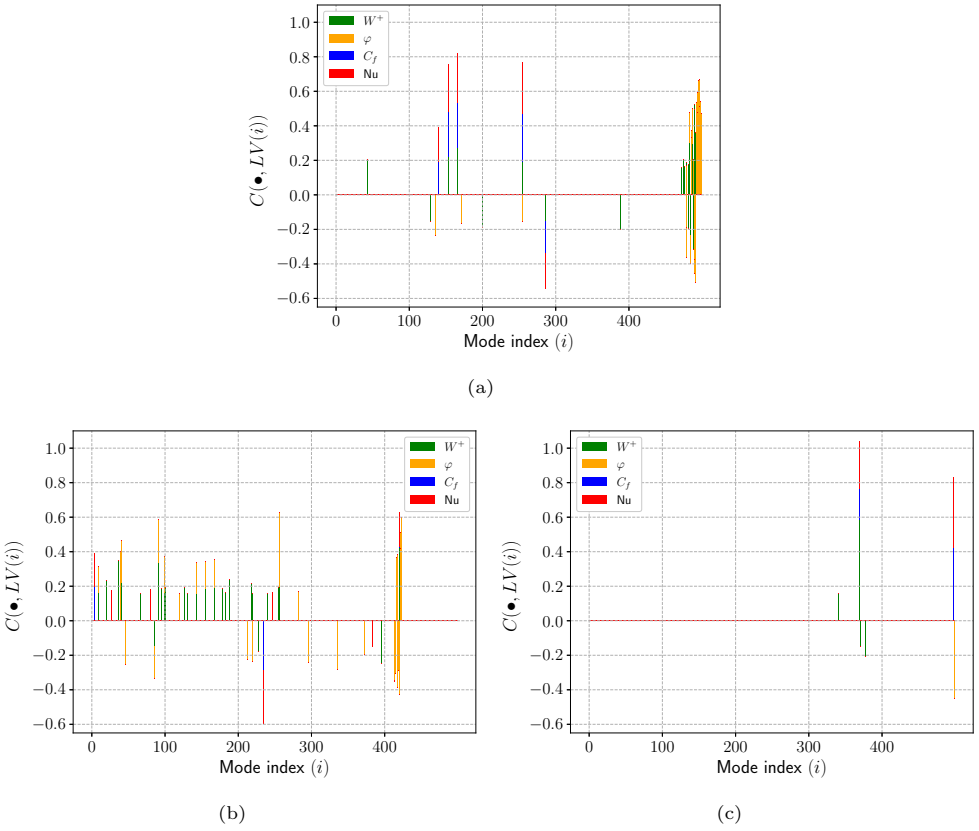


Fig. 17 Pearson correlation between latent variables and physical quantities: oscillation amplitudes W^+ in green, oscillation phases φ in orange, friction coefficient C_f in blue, and Nusselt number Nu in red. For each latent variable, correlations with the physical quantities are displayed as stacked bars, with each subsequent bar beginning at the endpoint of the previous one to enable direct comparison. Latent variables are ranked according to Section 5, and correlations with $|C(\bullet, LV(i))| < 0.15$ are omitted for clarity. Results are presented on the test data for (a): CAE, (b): β-VAE, and (c): extended POD.

7.2 Results and Interpretation

Fig. 17 presents the correlation results for the CAE, β-VAE, and extended-POD models, with latent variables ranked using the algorithm of Kang et al. [25] for autoencoders, whilst POD modes retain their inherent energy-based ordering. Correlations below 0.15 in absolute value are omitted. Across all models, no significant correlation (above 0.7) is observed, and the majority of latent variables exhibit correlations below the threshold, indicating predominantly weak or absent linear relationships.

The CAE results (Fig. 17(a)) demonstrate negligible correlations for the most reconstruction-contributing latent variables (first 100 LVs), consistent with the expectation that unconstrained non-linear optimisation produces highly entangled representations wherein physical information is encoded in complex, non-linear patterns inaccessible to linear probing. Intermediate modes (indices 100–300) exhibit modest correlations (0.2–0.3), whilst the least contributing modes display stronger correlations with W^+ and φ (up to 0.6).

The β-VAE exhibits a substantially greater number of detectable linear correlations (Fig. 17(b)), including amongst its most reconstruction-contributing variables. This enhancement is directly attributable to the regularisation term in equation (15), which penalises complex, entangled structures by minimising the KL divergence between the latent distribution and a standard Gaussian prior. This constraint forces simpler, more disentangled representations wherein relationships between latent variables and physical quantities manifest in more direct, linearly detectable forms. The decorrelation of latent variables imposed by the regularisation thus renders the encoded physical information more amenable to linear probing techniques, particularly amongst the most reconstruction-contributing modes. Importantly, this apparent improvement in "interpretability" represents a trade-off: the β-VAE sacrifices some reconstruction efficiency and optimal non-linear encoding capacity to achieve a simpler latent structure that is more accessible to linear analysis. The enhanced detectability of correlations thus reflects regularisation-imposed simplification rather than superior information encoding per se. Deep entanglement persists despite regularization, as numerous distinct latent variables demonstrate correlations to the same physical quantities.

Extended-POD demonstrates the weakest linear correlations, with only six modes exceeding the threshold (Fig. 17(c)), all corresponding to higher-order (low-energy) modes. This observation is consistent with POD's disentanglement (as quantified by the correlation metrics in Table 6), yet also reflects a fundamental limitation of the POD methodology: modes are ordered by energy contribution rather than by physical significance. Consequently, there is no *a priori* reason for specific modes (especially the most energetic ones) to encode particular physical quantities. This encoding strategy, whilst maintaining strong orthogonality properties, does not prioritise alignment with specific physical generative factors or quantities.

A common result across all models is that latent variables correlated with C_f also exhibit a similar correlation with Nu, consistent with the Reynolds analogy. For a limited number of these modes, a concurrent correlation with W^+ is observed, likely due to the gradual breakdown of the Reynolds analogy as W^+ increases. Furthermore, the least contributing modes in both AE models (excluding the pruned ones for the β-VAE) demonstrate strong correlations with φ or W^+. This highlights the significant

entanglement among these modes, as previously illustrated in Fig. 14, and suggests that they contribute to capturing the phase and amplitude behavior of the flow.

7.3 Connection to Latent Space Separability

The correlation patterns observed herein may be related to the separability analysis presented in Section 6, wherein the Davies-Bouldin index was employed to quantify cluster separation with respect to oscillation parameters. It is pertinent to recognise, however, that these two metrics probe fundamentally distinct aspects of latent space organisation. The Pearson correlation coefficient assesses local, linear relationships between individual latent variables and physical quantities, thereby measuring the extent to which a single mode varies linearly with a given parameter. In contrast, the Davies-Bouldin index evaluates global geometric structure, quantifying the separation between cluster centroids relative to intra-cluster variance across the entire latent space. As such, strong linear correlations are neither necessary nor sufficient for achieving superior cluster separability, as effective geometric organisation may be realised through complex, non-linear patterns distributed across multiple latent dimensions.

This conceptual distinction is demonstrated by the CAE results. The model achieves superior separability for amplitude-based clustering (lowest $D_B(W^+)$ score in Table 6) despite exhibiting weak linear correlations between individual latent variables and W^+. This apparent paradox is resolved by recognising that physical distinctions between amplitudes are embedded through complex, non-linear encodings distributed across multiple latent dimensions. Whilst individual modes exhibit negligible linear correlation with W^+, the collective geometric structure of the latent space permits effective amplitude-based clustering. The Davies-Bouldin metric, being sensitive to centroid separation in high-dimensional space, successfully detects this global organisation, whereas the Pearson coefficient, probing only linear relationships in individual dimensions, fails to capture the distributed, non-linear encoding mechanism.

The β-VAE demonstrates moderately degraded separability relative to the CAE (intermediate D_B scores in Table 6), yet exhibits substantially enhanced linear correlation, particularly amongst its most reconstruction-contributing modes. This inverse relationship between separability and linear detectability represents the direct consequence of the regularisation strategy: the KL divergence term in equation (15) constrains the latent distribution to approximate a standard Gaussian prior, thereby encouraging simpler, more linearly structured encodings. Whilst this regularisation enhances the linear detectability of physical relationships (rendering them accessible to correlation-based probing), it simultaneously constrains the model's capacity to construct optimally separated geometric clusters through complex, non-linear transformations. The trade-off is thus explicit: the β-VAE sacrifices some degree of optimal geometric organisation to achieve a more linearly interpretable latent structure.

Extended-POD exhibits particularly poor separability for amplitude-based clustering (highest $D_B(W^+)$ score in Table 6) alongside intermediate performance for phase-based organisation, despite maintaining superior orthogonality (as quantified in Table 6). This behaviour is directly attributable to POD's fundamental organisational principle: modes are ordered by energy contribution rather than by alignment with

physical generative factors. The orthogonal basis constructed through POD is optimised exclusively for capturing variance in descending order, with no mechanism to ensure that specific modes encode particular physical parameters such as W^+ or φ. Samples characterised by different amplitudes or phases are thus not systematically organised within distinct geometric regions of the latent space, resulting in poorly separated clusters and elevated Davies-Bouldin scores. The autoencoders, by contrast, learn latent representations through data-driven non-linear optimisation, thereby constructing structures that inherently reflect the underlying generative factors. This fundamental difference in organisational principle explains why the autoencoders achieve substantially superior physical separability despite the CAE's employment of highly non-linear, entangled encodings and despite POD's superior orthogonality.

8 Conclusions and perspectives

This study comprehensively evaluated the effectiveness of proper orthogonal decomposition and various autoencoder architectures in reducing the dimensionality of turbulent channel flow data at $\mathrm{Re}_\tau = 200$ and $\mathrm{Pr} = 1$. The dataset, comprising velocity and temperature fields, was generated using Large Eddy Simulations under spanwise wall oscillations with a fixed oscillation period and varying amplitudes. Specific attention was set towards small turbulent scales reconstruction, orthogonality of the latent samples, and latent separability of generative factors. The main results from the investigation are summarised as follows:

- CAEs consistently outperformed POD in terms of reconstruction accuracy across all flow fields, particularly at lower latent dimensions. The most significant improvement was observed in the wall-normal velocity component (v), highlighting the advantage of non-linearity offered by AE architectures in capturing complex turbulent dynamics. Furthermore, the CAE demonstrated superior generalization capability across the range of oscillation amplitudes.
- Compressing all four fields (u, v, w, Θ) simultaneously resulted in superior reconstruction performance compared to individual field compression. This enhancement leverages the inherent physical correlations between the variables, especially between the streamwise velocity (u) and temperature (Θ) fields.
- Both POD and AE methodologies encountered difficulties in accurately reconstructing small-scale turbulent structures, particularly within the v-field. POD, constrained to 500 modes, was limited to capturing only large-scale flow characteristics, and exhibited a tendency towards mean-value approximation.
- β-VAEs achieved a balance between the reconstruction performance of CAEs and latent samples orthogonality of extended POD. Although there was a decrease in accuracy for the v and Θ fields, correlation metrics indicated improved latent space structure compared to CAEs.
- The β-VAE maintains optimal performance with only 424 latent variables ($L_d = 424$), preserving an effective balance between reconstruction fidelity and orthogonality. Notably, in this pruned architecture, latent variables exhibiting greater KL divergence tend to show reduced inter-mode correlations.

- Both CAEs and β-VAEs demonstrate superior capability in distinguishing amplitude-related features (W^+), compared to extended POD. Autoencoders additionally incorporate key physical insights by differentiating phases (φ) with substantially different Stokes strain or streak orientation. Despite similar trends, the CAE consistently outperformed the β-VAE in terms of latent separability.
- The β-VAE sacrifices some degree of optimal geometric organisation to achieve a more linearly interpretable latent structure. The CAE's weak linear correlations for its dominant modes do not indicate absence of physical information, but rather its encoding in complex, non-linear patterns. Extended-POD's weak correlations and poor physical separability arise from its energy-based rather than physics-based mode ordering. For all models, latent variables correlated with C_f consistently exhibit similar correlations with Nu, reflecting the Reynolds analogy.

Based on the current results, future research directions are proposed to further enhance the understanding and application of dimensionality reduction techniques in turbulent flows or to provide a reduced order dynamical model for the present dataset:

- Employ CROM techniques [48] to elucidate the dynamics encoded within the latent spaces of each model. This approach, which is more complex than that of the present study, may reveal more meaningful information regarding how each model structures its latent space with regards to physical parameters.
- Investigate the application of three-dimensional convolutional layers to compress the entire 3D channel, as opposed to the current two-dimensional slice approach. This holistic compression strategy may capture spatial correlations more effectively, improving reconstruction accuracy and model generalisation. However, compressing the 3D space, characterized by a much larger dimension, may prove a challenging task.
- Given the inherent limitations of linear probing in non-linear latent spaces, future investigations should employ non-linear probing techniques (such as mutual information analysis, supervised classification tasks, or non-linear dimensionality reduction of the latent space) to thoroughly evaluate the physical information encoded within these high-dimensional representations.

A primary and promising direction is to extend the learned latent representations for temporal forecasting and control. This could be achieved by combining the AE-based dimensionality reduction with long short term memory networks or transformers [24] to develop an efficient predictive model of the flow, thereby facilitating real-time flow control and optimisation in engineering applications.

Appendix

Generalised Stokes Layer Solution

For an arbitrary periodic spanwise wall motion, the laminar Stokes layer solution is constructed through harmonic superposition [13]. Consider a wall velocity signal decomposed into its Fourier series:

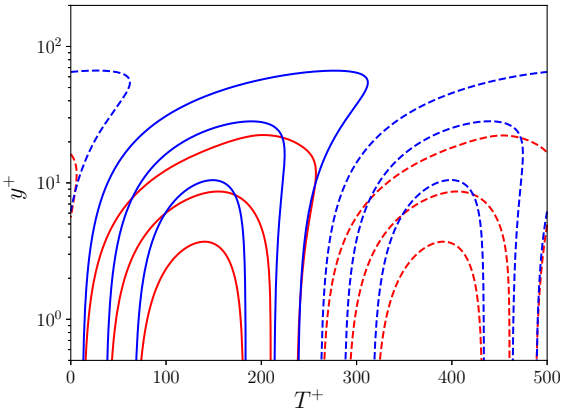


Fig. 18 Phase averaged wall-normal velocity $\widetilde{w}(y, \varphi)$ for oscillation parameters $T^+ = 500$, $W^+ = 30$, at $\mathrm{Re}_\tau = 180$. Blue contours indicate the laminar analytical case (27), while red contours correspond to the turbulent case. Solid contours correspond to positive values while dashed contours indicate negative values.

$$w_{\mathrm{w}}(t) = W_m \sum_{n=1}^{\infty} A_n \sin\left(\frac{2\pi n t}{T} + \phi_n\right). \tag{25}$$

where W_m represents the maximum wall velocity, A_n are the normalised harmonic amplitudes satisfying $\sum_{n=1}^{\infty} |A_n|^2 = 1$, and ϕ_n are the phase angles. The corresponding laminar Stokes layer velocity profile is given by:

$$w_{\mathrm{St}}(y,t) = W_m \sum_{n=1}^{\infty} A_n \exp\left(-\sqrt{n}\frac{y}{\delta_{s,1}}\right) \sin\left(\frac{2\pi n t}{T} + \phi_n - \sqrt{n}\frac{y}{\delta_{s,1}}\right), \tag{26}$$

where $\delta_{s,1} = \sqrt{2\nu T/(2\pi)}$ represents the characteristic Stokes layer thickness for the fundamental frequency.

Sinusoidal Specialisation:

For the specific case of sinusoidal wall motion $w_{\mathrm{w}}(t) = W \sin(2\pi t/T)$, we have $W_m = W$, $A_1 = 1$, $A_n = 0$ for $n > 1$, and $\phi_1 = 0$, reducing equation (26) to:

$$w_{\mathrm{St}}(y,t) = W \exp\left(-\frac{y}{\delta_s}\right) \sin\left(\frac{2\pi t}{T} - \frac{y}{\delta_s}\right), \tag{27}$$

where $\delta_s = \delta_{s,1}$ is the classical Stokes layer thickness. Fig. 18 illustrates the difference between the laminar and turbulent phase-averaged profiles, where the analytical solution results in a deeper penetration of the Stokes layer into the buffer layer.

Statements and Declarations

Conflict of interest

The authors report no conflict of interest.

Data availability statement

The dataset used in this study is publicly available on the following link: https://doi.org/10.57745/GU68MY.

References

[1] Ricco, P., Skote, M., Leschziner, M.A.: A review of turbulent skin-friction drag reduction by near-wall transverse forcing. Progress in Aerospace Sciences **123**, 100713 (2021). Publisher: Elsevier

[2] Abdulbari, H.A., Zulkifli, R., Salleh, M.Z., Yusoff, M.Z.: Skin-Friction Drag Reduction Using Riblets, Dimples, and Oscillating Wall. Journal of Mechanical Engineering and Sciences **4**, 452–464 (2013)

[3] Asidin, S., Kuntjoro, W., Salleh, M.Z.: A Review on Flow Control Methods in Turbine Internal Cooling. Applied Sciences **9**(16), 3233 (2019)

[4] Zhang, X., Li, T., Jiang, N.: Flow Control Strategies for Improving Performance of Turbomachinery: A Review. Progress in Aerospace Sciences **115**, 100595 (2020)

[5] Agostini, L., Touber, E., Leschziner, M.: The turbulence vorticity as a window to the physics of friction-drag reduction by oscillatory wall motion. International Journal of Heat and Fluid Flow **51**, 3–15 (2015)

[6] Quadrio, M., Ricco, P.: Critical assessment of turbulent drag reduction through spanwise wall oscillations. Journal of Fluid Mechanics **521**, 251–271 (2004)

[7] Choi, K., Clayton, B.R.: The mechanism of turbulent drag reduction with wall oscillation. International Journal of Heat and Fluid Flow **22**(1), 1–9 (2001) https://doi.org/10.1016/S0142-727X(00)00070-9

[8] Viotti, C., Quadrio, M., Luchini, P.: Streamwise oscillation of spanwise velocity at the wall of a channel for turbulent drag reduction. Physics of fluids **21**(11), 115109 (2009)

[9] Touber, E., Leschziner, M.: Near-wall streak modification by spanwise oscillatory wall motion and drag-reduction mechanisms. Journal of Fluid Mechanics **693**, 150–200 (2012)

[10] Guérin, L., Flageul, C., Cordier, L., Grieu, S., Agostini, L.: Preferential enhancement of convective heat transfer over drag via near-wall turbulence manipulation using spanwise wall oscillations. International Journal of Heat and Fluid Flow **110**, 109564 (2024) https://doi.org/10.1016/j.ijheatfluidflow.2024.109564

[11] Rouhi, A., Hultmark, M., Smits, A.J.: Spanwise wall forcing can reduce turbulent heat transfer more than drag. Journal of Fluid Mechanics **1010**, 59 (2025) https://doi.org/10.1017/jfm.2025.310

[12] Rouhi, A., Fu, M.K., Chandran, D., Zampiron, A., Smits, A.J., Marusic, I.: Turbulent drag reduction by spanwise wall forcing. Part 1. Large-eddy simulations. Journal of Fluid Mechanics **968**, 6 (2023) https://doi.org/10.1017/jfm.2023.499

[13] Cimarelli, A., Frohnapfel, B., Hasegawa, Y., Angelis, E.D., Quadrio, M.: Prediction of turbulence control for arbitrary periodic spanwise wall movement. Physics of Fluids **25**(7) (2013) https://doi.org/10.1063/1.4813807 . Publisher: AIP Publishing

[14] Berkooz, G., Holmes, P., Lumley, J.L.: The proper orthogonal decomposition in the analysis of turbulent flows. Annual review of fluid mechanics **25**(1), 539–575 (1993)

[15] Agostini, L.: Exploration and prediction of fluid dynamical systems using auto-encoder technology. Physics of Fluids **32**(6), 067103 (2020) https://doi.org/10.1063/5.0012906

[16] Linot, A.J., Graham, M.D.: Dynamics of a data-driven low-dimensional model of turbulent minimal Couette flow. Journal of Fluid Mechanics **973**, 42 (2023) https://doi.org/10.1017/jfm.2023.720

[17] Linot, A.J., Zeng, K., Graham, M.D.: Turbulence control in plane Couette flow using low-dimensional neural ODE-based models and deep reinforcement learning. International Journal of Heat and Fluid Flow **101**, 109139 (2023) https://doi.org/10.1016/j.ijheatfluidflow.2023.109139

[18] Gruber, A., Gunzburger, M., Ju, L., Wang, Z.: A comparison of neural network architectures for data-driven reduced-order modeling. Computer Methods in Applied Mechanics and Engineering **393**, 114764 (2022) https://doi.org/10.1016/j.cma.2022.114764

[19] Bank, D., Koenigstein, N., Giryes, R.: Autoencoders. arXiv. arXiv:2003.05991 [cs] (2021)

[20] Agostini, L., Leschziner, M.: Auto-encoder-assisted analysis of amplitude and wavelength modulation of near-wall turbulence by outer large-scale structures in channel flow at friction Reynolds number of 5200. Physics of Fluids **34**(11), 115142 (2022) https://doi.org/10.1063/5.0123119

[21] Du, X., Qu, X., He, Y., Guo, D.: Single Image Super-Resolution Based on Multi-Scale Competitive Convolutional Neural Network. Sensors **18**(3), 789 (2018) https://doi.org/10.3390/s18030789

[22] Csala, H., Dawson, S.T.M., Arzani, A.: Comparing different nonlinear dimensionality reduction techniques for data-driven unsteady fluid flow modeling. Physics of Fluids **34**(11) (2022) https://doi.org/10.1063/5.0127284 . Publisher: AIP Publishing

[23] Solera-Rico, A., Vila, C.S., Gómez-López, M., Wang, Y., Almashjary, A., Dawson, S.T.M., Vinuesa, R.: β-Variational autoencoders and transformers for reduced-order modelling of fluid flows. Nature Communications **15**(1), 1–15 (2024) https://doi.org/10.1038/s41467-024-45578-4 . Number: 1 Publisher: Nature Publishing Group

[24] Eivazi, H., Clainche, S.L., Hoyas, S., Vinuesa, R.: Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows. Expert Systems with Applications **202**, 117038 (2022). Publisher: Elsevier

[25] Kang, Y., Yang, S., Yee, K.: Physics-aware Reduced-order Modeling of Transonic Flow via β-Variational Autoencoder. Publisher: arXiv Version Number: 2 (2022). https://doi.org/10.48550/ARXIV.2205.00608

[26] Mohan, A.T., Lubbers, N., Chertkov, M., Livescu, D.: Embedding hard physical constraints in neural network coarse-graining of three-dimensional turbulence. Physical Review Fluids **8**(1), 014604 (2023) https://doi.org/10.1103/PhysRevFluids.8.014604 . Publisher: American Physical Society

[27] Fukami, K., Taira, K.: Grasping extreme aerodynamics on a low-dimensional manifold. Nature Communications **14**(1), 6480 (2023) https://doi.org/10.1038/s41467-023-42213-6

[28] Bevanda, P., Sosnowski, S., Hirche, S.: Koopman operator dynamical models: Learning, analysis and control. Annual Reviews in Control **52**, 197–212 (2021) https://doi.org/10.1016/j.arcontrol.2021.09.002 . Accessed 2025-10-19

[29] Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences **113**(15), 3932–3937 (2016) https://doi.org/10.1073/pnas.1517384113 . Accessed 2025-10-19

[30] Williams, J.P., Zahn, O., Kutz, J.N.: Sensing with shallow recurrent decoder networks. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **480**(2298), 20240054 (2024) https://doi.org/10.1098/rspa.2024.0054 . Accessed 2025-10-19

[31] Guérin, L., Cordier, L., Flageul, C., Grieu, S., Agostini, L.: Policy-based optimization for drag reduction via spanwise wall oscillations. Neural Computing and Applications (2025) https://doi.org/10.1007/s00521-025-11067-y

[32] Kasagi, N., Tomita, Y., Kuroda, A.: Direct Numerical Simulation of Passive Scalar Field in a Turbulent Channel Flow. Journal of Heat Transfer **114**(3), 598–606 (1992) https://doi.org/10.1115/1.2911323

[33] Laizet, S., Lamballais, E.: High-order compact schemes for incompressible flows:

A simple and efficient method with quasi-spectral accuracy. Journal of Computational Physics **228**(16), 5989–6015 (2009) https://doi.org/10.1016/j.jcp.2009.05.010

[34] Laizet, S., Li, N.: Incompact3d: A powerful tool to tackle turbulence problems with up to O (10^{5}) computational cores. International Journal for Numerical Methods in Fluids **67**(11), 1735–1757 (2011) https://doi.org/10.1002/fld.2480

[35] Bartholomew, P., Deskos, G., Frantz, R.A., Schuch, F.N., Lamballais, E., Laizet, S.: Xcompact3D: An open-source framework for solving turbulence problems on a Cartesian mesh. SoftwareX **12**, 100550 (2020) https://doi.org/10.1016/j.softx.2020.100550

[36] Abe, H., Antonia, R.A., Kawamura, H.: Correlation between small-scale velocity and scalar fluctuations in a turbulent channel flow. Journal of Fluid Mechanics **627**, 1–32 (2009) https://doi.org/10.1017/S0022112008005569

[37] Seki, Y., Iwamoto, K., Kawamura, H.: Prandtl number Effect on Turbulence Quantities through High Spatial Resolution DNS of Turbulent Heat Transfer in a Channel Flow. In: Proceedings of the International Symposium on Turbulence, Heat and Mass Transfer, pp. 301–304. Begellhouse, Dubrovnik, Croatia (2006). https://doi.org/10.1615/ICHMT.2006.TurbulHeatMassTransf.560

[38] Flageul, C., Benhamadouche, S., Lamballais, E., Laurence, D.: DNS of turbulent channel flow with conjugate heat transfer: Effect of thermal boundary conditions on the second moments and budgets. International Journal of Heat and Fluid Flow **55**, 34–44 (2015) https://doi.org/10.1016/j.ijheatfluidflow.2015.07.009

[39] Lamballais, E., Fortuné, V., Laizet, S.: Straightforward high-order numerical dissipation via the viscous term for direct and large eddy simulation. Journal of Computational Physics **230**(9), 3270–3275 (2011) https://doi.org/10.1016/j.jcp.2011.01.040

[40] Cordier, L., Bergmann, M.: Proper Orthogonal Decomposition: an overview. In: Lecture Series 2002-04, 2003-03 and 2008-01 on Post-processing of Experimental and Numerical Data, pp. 1–46. Von Kármán Institute for Fluid Dynamics, ??? (2008). ISBN 978-2-930389-80-X

[41] Borée, J.: Extended proper orthogonal decomposition: a tool to analyse correlated events in turbulent flows. Experiments in Fluids **35**(2), 188–192 (2003) https://doi.org/10.1007/s00348-003-0656-3 . Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 2 Publisher: Springer-Verlag

[42] Fukami, K., Nakamura, T., Fukagata, K.: Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data. Physics of Fluids **32**(9) (2020) https://doi.org/10.1063/5.0020721 . Publisher: AIP Publishing

[43] Rasamoelina, A.D., Adjailia, F., Sinčák, P.: A review of activation function for artificial neural network. In: 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 281–286 (2020). IEEE

[44] Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., Han, J.: On the variance of the adaptive learning rate and beyond. arXiv preprint arXiv:1908.03265 (2019)

[45] Kingma, D.P., Welling, M.: An Introduction to Variational Autoencoders. Foundations and Trends® in Machine Learning **12**(4), 307–392 (2019) https://doi.org/10.1561/2200000056

[46] Hoffman, M.D., Johnson, M.J.: Elbo surgery: yet another way to carve up the variational evidence lower bound. In: Workshop in Advances in Approximate Bayesian Inference, NIPS, vol. 1 (2016)

[47] Xu, D., Tian, Y.: A Comprehensive Survey of Clustering Algorithms. Annals of Data Science **2**(2), 165–193 (2015) https://doi.org/10.1007/s40745-015-0040-1

[48] Kaiser, E., Noack, B.R., Cordier, L., Spohn, A., Segond, M., Abel, M., Daviller, G., Östh, J., Krajnović, S., Niven, R.K.: Cluster-based reduced-order modelling of a mixing layer. Journal of Fluid Mechanics **754**, 365–414 (2014). Publisher: Cambridge University Press