



Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows

Hamidreza Eivazi^{a,*}, Soledad Le Clainche^b, Sergio Hoyas^c, Ricardo Vinuesa^{a,d,**}

^a FLOW, Engineering Mechanics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

^b School of Aerospace Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain

^c Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, 46022 Valencia, Spain

^d Swedish e-Science Research Centre (SeRC), Stockholm, Sweden

ARTICLE INFO

Keywords:

Non-linear mode decomposition
Turbulent flows
Variational autoencoders
Convolutional neural networks
Machine learning

ABSTRACT

Modal-decomposition techniques are computational frameworks based on data aimed at identifying a low-dimensional space for capturing dominant flow features: the so-called modes. We propose a deep probabilistic-neural-network architecture for learning a minimal and near-orthogonal set of non-linear modes from high-fidelity turbulent-flow data useful for flow analysis, reduced-order modeling and flow control. Our approach is based on β -variational autoencoders (β -VAEs) and convolutional neural networks (CNNs), which enable extracting non-linear modes from multi-scale turbulent flows while encouraging the learning of independent latent variables and penalizing the size of the latent vector. Moreover, we introduce an algorithm for ordering VAE-based modes with respect to their contribution to the reconstruction. We apply this method for non-linear mode decomposition of the turbulent flow through a simplified urban environment, where the flow-field data is obtained based on well-resolved large-eddy simulations (LESs). We demonstrate that by constraining the shape of the latent space, it is possible to motivate the orthogonality and extract a set of parsimonious modes sufficient for high-quality reconstruction. Our results show the excellent performance of the method in the reconstruction against linear-theory-based decompositions, where the energy percentage captured by the proposed method from five modes is equal to 87.36% against 32.41% of the POD. Moreover, we compare our method with available AE-based models. We show the ability of our approach in the extraction of near-orthogonal modes with the determinant of the correlation matrix equal to 0.99, which may lead to interpretability.

1. Introduction

Analysis of turbulent flows is challenging due to the presence of a vast range of spatio-temporal coherent structures in high-dimensional non-linear dynamics. However, the fact that common flow features appear across a wide range of fluid flows suggests that there are key dominant phenomena that serve as the foundation of many flows. Modal-decomposition techniques offer methods to identify a low-dimensional coordinate system for capturing dominant flow features (Taira et al., 2017, 2020) useful for developing reduced-order models, analyzing non-linear and chaotic dynamics, and designing efficient flow-control schemes. Proper-orthogonal decomposition (POD) (Lumley, 1967) and dynamic-mode decomposition (DMD) (Rowley et al., 2009; Schmid, 2010) are two mode-decomposition methods based on linear algebra that have been widely used to extract the dominant spatio-temporal

features in fluid flows. Balanced POD (BPOD) (Rowley, 2005), spectral POD (SPOD) (Towne et al., 2018), higher-order DMD (HODMD) (Le Clainche & Vega, 2017) and spatio-temporal Koopman decomposition (STKD) (Clainche & Vega, 2018) are several successful variants of POD and DMD for analysis of turbulent flows.

Recent developments in deep learning for engineering problems bring advanced and innovative approaches to improve the efficiency, flexibility, and accuracy of the predictive models. Some of the outstanding applications of deep neural networks (DNNs) in the domain of computational physics are solution of partial differential equations (PDEs) (Eivazi, Tahani, Schlatter, & Vinuesa, 2021; Jin et al., 2021; Raissi et al., 2019), operator learning (Gin et al., 2020; Li et al., 2021; Lu et al., 2021), linear embedding of non-linear dynamics (Lusch et al., 2018), and model reduction of dynamical systems (Lee & Carlberg,

* Corresponding author.

** Corresponding author at: FLOW, Engineering Mechanics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden.

E-mail addresses: hamidre@kth.se (H. Eivazi), soledad.leclainche@upm.es (S. Le Clainche), sergio.hoyas@mot.upv.es (S. Hoyas), rvinuesa@mech.kth.se (R. Vinuesa).

<https://doi.org/10.1016/j.eswa.2022.117038>

Received 31 October 2021; Received in revised form 18 February 2022; Accepted 27 March 2022

Available online 9 April 2022

0957-4174/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2020), besides many other successful applications of data-driven methods (Albu et al., 2019; Dhini et al., 2021; Hedrea et al., 2021; Kashani et al., 2021; Tseng et al., 2021). Fluid mechanics has been one of the active research topics for development of innovative DNN-based approaches (Brunton et al., 2020; Duraisamy et al., 2019; Kutz, 2017; Vinuesa & Brunton, 2021). Successful application of DNNs has been shown in the data-driven turbulence closure modeling (Jiang et al., 2021; Ling et al., 2016), prediction of temporal dynamics of a low-order model of turbulence (Eivazi, Guastoni, Schlatter, Azizpour, & Vinuesa, 2021; Srinivasan et al., 2019), extraction of turbulence theory for two-dimensional decaying isotropic turbulent (Jiménez, 2018), non-intrusive sensing in turbulent flows (Guastoni et al., 2021; Güemes et al., 2021), and active flow control through deep reinforcement learning (Tang et al., 2020).

Besides the aforementioned linear methods for modal decomposition of flow-field data, DNN-based models have shown promising performance in learning a compact latent representation of high-dimensional data by accounting for the non-linearity in the low-dimensional mapping using non-linear activation functions (Hinton & Salakhutdinov, 2006). In particular, unsupervised learning based on autoencoders (AEs) has been shown suitable for efficient mode decomposition and reduced-order modeling with superior performance in flow reconstruction over the linear POD (Eivazi et al., 2020; Milano & Koumoutsakos, 2002).

Moreover, convolutional neural networks (CNNs) (Lecun et al., 1998) and their ability in pattern recognition have received increasing attention by the fluid-mechanics community (Fukami et al., 2019; Guastoni et al., 2021; Kim et al., 2021; Kim & Lee, 2020; Lee & You, 2019). Murata et al. (2020) proposed a CNN-based autoencoder architecture for decomposition of flow-fields into non-linear low-dimensional modes and to visualize each mode. They applied this so-called mode-decomposing convolutional-neural-network autoencoder (MD-CNN-AE) to a relatively simple laminar flow around a circular cylinder at $Re_D = 100$ (where Re_D is the Reynolds number based on freestream velocity and cylinder diameter). Their results showed the superior performance of the CNN-based autoencoder over POD where the reconstruction of the flow from only two MD-CNN-AE modes contains also the higher-order POD modes. The architecture of AE-based methods allows a non-linear low-dimensional mapping leading to a superior performance against linear-theory-based methods. However, the AE-based methods do not benefit from the useful properties of the eigenvalue or singular-value-decomposition techniques, e.g., optimality and orthogonality. In contrast to the POD modes, which are an orthogonal set of basis vectors arranged in the order of their energy content, the AE-based modes are neither orthogonal nor ranked. This may lead to the lack of interpretability and robustness of the AE-based modes (Vinuesa & Sirmacek, 2021). In order to obtain ranked modes, Fukami et al. (2020) proposed a hierarchical CNN-AE architecture inspired by the concept of hierarchical autoencoder (AE) (Saegusa et al., 2004). The proposed method was first applied to a laminar cylinder wake and its transient process and further to an in-plane cross-sectional velocity field of turbulent channel flow at $Re_\tau = 180$ (note that Re_τ is the friction Reynolds number, based on channel half height and friction velocity). They showed that the hierarchical autoencoder (AE) can rank the AE modes following their contributions to the reconstructed field while achieving efficient order reduction. However, issues related to interpretability and non-uniqueness remained unanswered.

Finding an efficient set of reduced coordinates is an essential part of developing reduced-order models (ROMs) (Vinuesa & Brunton, 2021). AE-based models can be employed to learn a set of nonlinear manifold coordinates that can considerably improve the compression in the latent space (Lee & Carlberg, 2020). These coordinates typically describe the amplitudes of important flow structures. Once an efficient coordinate system is found, the dynamics of these coordinates may be modeled using many machine-learning approaches, including long short-term memory (LSTM) network (Abadía-Heredia et al., 2022;

Srinivasan et al., 2019), echo-state networks (Pathak et al., 2018), or through sparse identification of nonlinear dynamics (SINDy) (Brunton et al., 2016). ROMs provide a fast surrogate model for more expensive fluid flow simulations. This is extremely useful for optimization and active flow-control tasks that require many model iterations or fast and online predictions with increasing importance, both from a fundamental science perspective and for many industrial engineering applications.

In this paper, we propose a probabilistic method based on β -variational autoencoders (β -VAEs) (Higgins et al., 2017) and CNNs in order to extract a minimal (parsimonious) set of near-orthogonal non-linear modes from turbulent flows. We applied the proposed machine-learning method for the modal decomposition of high-fidelity turbulent-flow simulation data of a simplified urban environment. The flow simulation is carried through well-resolved large-eddy simulation (LES) by means of the spectral-element code Nek5000 (Fischer et al., 2008). The results from the proposed method are compared with the results from a conventional CNN-AE model, a hierarchical CNN-AE model, and POD. It should be noted that the complexities of the compared methods are not the same, and the goal of this comparison is to assess how different model architectures and training strategies affect the learned latent modes and the quality of the reconstructions. Through the training process of the available AE-based modal-decomposition methods, the objective is to only minimize the reconstruction loss. Therefore, the focus of the available AE-based methods for non-linear modal decomposition is only on the performance of the model in the reconstruction and not on the properties of the learned modes, such as orthogonality and minimality. In contrast, here we minimize the correlation between the latent variables and penalize the size of the latent vector in addition to the minimization of the reconstruction loss. By solving this multi-objective optimization problem using our CNN- β VAE approach, we seek a minimal set of uncorrelated non-linear modes that are able to accurately describe the turbulent flow-field data. The obtained modes are extremely useful for the development of compact reduced-order surrogate models or designing flow-control strategies. Moreover, understanding the physics of the turbulent flow through the extraction of uncorrelated non-linear mechanisms is another incentive for applying CNN- β VAEs for modal decomposition. In particular, the development of accurate predictive models and understanding flow structures in urban environments are of significant importance due to their impact on urban planning, air quality management, and pollutant dispersion (Vinuesa et al., 2015). The essential need to address sustainable development from an urban perspective is pronounced in the 2030 Agenda (UN General Assembly, 2015)¹ through the sustainable development goals (SDGs) 11 (on sustainable cities and communities) and 13 (on climate action). Air pollution is a major cause of premature death and disease and is the single largest environmental health risk in Europe. Heart disease and stroke are the most common reasons for premature deaths attributable to air pollution, followed by lung diseases and lung cancer (European Environment Agency et al., 2020). Predictive models for air-quality detection are in particular important to provide protection from excessive pollution concentrations. However, the available predictive models are unable to provide the spatial and temporal accuracy required for reproducing pollutant-dispersion patterns within urban environments. Therefore, there is a pressing need for improved prediction and assessment methods to tackle these challenges and enable urban sustainability in the near future (Vinuesa et al., 2020). Furthermore, it is necessary to gain further insight into the physics responsible for the pollutant-concentration and thermal distributions within cities (Torres et al., 2021).

This article is organized as follows: in Section 2 we provide an overview of the flow physics and discuss the theoretical background

¹ United Nations (UN).

relevant to linear-theory-based modal decomposition; in Section 3 we discuss three different CNN-AE-based methods from a methodological point of view; the performance of the methods and their characteristics are compared in Section 4; and finally, in Section 5 we provide a summary and the conclusions of the study.

2. Modal decomposition

Modal decomposition is a mathematical method to extract energetically- and dynamically-important features from fluid flows. The spatial features are represented by a set of modes ranked in terms of kinetic energy or their largest amplitude (connected with the norm of the mode). These modes are generally obtained via solving an eigenvalue problem. The obtained eigenvalues could represent the energy content of the modes or the growth rates and frequencies modeling the temporal dynamics of the flow motion. Modal-decomposition techniques produce a low-dimensional coordinate system for capturing dominant flow structures. These dominant flow structures are extremely useful not only for flow analysis but also for reduced-order modeling and flow control. For a detailed discussion of the modal-decomposition techniques, the readers are referred to the reviews on the topic (Taira et al., 2017, 2020). In the following, we discuss flow physics in a simplified urban environment and apply POD for modal decomposition.

2.1. Training data

We employ a database (Martínez-Sánchez et al., 2022; Torres et al., 2021) of the flow through a simplified urban environment to obtain the training data. This database was obtained through well-resolved large-eddy simulation (LES) using the open-source numerical code Nek5000 (Fischer et al., 2008), which is based on the spectral-element method (SEM), to solve the incompressible Navier–Stokes equations:

$$\nabla \cdot \mathbf{u} = 0, \quad (1)$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \nu \nabla^2 \mathbf{u},$$

where \mathbf{u} represents the velocity vector, t is the time, and ν is the kinematic viscosity. p is the pressure, including the constant density. Note that the SEM combines the geometrical flexibility required to discretize the urban geometry with the high-order accuracy of spectral methods. The geometry of the buildings is discretized using hexahedral elements and the scale disparity of the turbulent flows is resolved using the Gauss–Lobatto–Legendre (GLL) quadrature within each element. The flow case includes two wall-mounted obstacles with width-to-height ratio b/h and height-to-separation ratio h/ℓ of 0.5 and 1.25, respectively. Here x , y and z denote streamwise, wall-normal and spanwise directions, respectively. The x – z cross-sectional velocity fields at $y = 0.5h$ are extracted and used as the input data. As we focus on the flow around the obstacles, we extract the following region of the computational domain: $-1h \leq x \leq 5h$ and $-1.5h \leq z \leq 1.5h$. To facilitate the data-handling process, we performed spectral interpolation of the 1200 instantaneous fields considered in this work from the original SEM mesh to another one containing the following number of grid points in x and z : $(N_x, N_z) = (96, 192)$. We use the fluctuation component of the streamwise velocity u as the input of the models.

2.2. Proper-orthogonal decomposition (POD)

In this section, we apply POD for modal decomposition and discuss some of its useful properties, i.e., optimality and orthogonality. The POD technique (also known as the Karhunen–Loève (KL) procedure (Karhunen, 1946; Loève, 1955)) was first introduced to the fluid dynamics/turbulence community by Lumley (1967) as a mathematical algorithm to extract coherent structures from turbulent flows. POD extracts modes based on minimizing the mean-square error between the

signal and its reconstruction and also minimizing the number of modes required for such a reconstruction. This leads to a minimal number of basis functions or modes to capture as much energy as possible. Let us consider a vector field as: $q(\xi, t)$ representing, e.g., velocity with coordinates ξ , where t denotes time. Having the temporal mean $\bar{q}(\xi)$ subtracted, the unsteady component of the vector field can be decomposed as:

$$q(\xi, t) - \bar{q}(\xi) = \sum_j a_j(t) \phi(\xi), \quad (2)$$

where $\phi(\xi)$ and a_j represent the spatial modes and temporal (expansion) coefficients respectively. To this end, we first prepare snapshots of the flow-field as a collection of finite-dimensional data vectors:

$$\mathbf{x}(t) = q(\xi, t) - \bar{q}(\xi) \in \mathbb{R}^n, \quad t = t_1, t_2, \dots, t_m. \quad (3)$$

where $\mathbf{x}(t)$ represents the fluctuating component of the vector data, n is the number of grid points modeling the vector data and m is the number of snapshots selected to model the flow dynamics. We arrange the data into a matrix \mathbf{X} through the concatenation of m snapshots as follows:

$$\mathbf{X} = [\mathbf{x}(t_1) \ \mathbf{x}(t_2) \ \dots \ \mathbf{x}(t_m)] \in \mathbb{R}^{n \times m}. \quad (4)$$

The POD modes can be determined as the eigenvectors of the covariance matrix $\mathbf{R} = \mathbf{X} \mathbf{X}^T$:

$$\mathbf{R} \phi_j = \lambda_j \phi_j, \quad \phi_j \in \mathbb{R}^n, \quad \lambda_1 \geq \dots \lambda_n \geq 0. \quad (5)$$

The eigenvalues λ_j show how well each mode ϕ_j represents the reference data in the ℓ_2 sense. Considering the velocity vectors as $\mathbf{x}(t)$, each eigenvalue shows the kinetic energy captured by its corresponding mode.

Another approach is to apply singular-value decomposition (SVD) (Sirovich, 1987) directly on the matrix \mathbf{X} as:

$$\mathbf{X} = \Phi \Sigma \Psi^T, \quad (6)$$

where $\Phi \in \mathbb{R}^{n \times n}$ and $\Psi \in \mathbb{R}^{m \times m}$ are the left and right singular vectors of \mathbf{X} , respectively, and Σ is a diagonal matrix containing the singular values $(\sigma_1, \dots, \sigma_m)$. The singular vectors Φ and Ψ are identical to the eigenvectors of $\mathbf{X} \mathbf{X}^T$ and $\mathbf{X}^T \mathbf{X}$, respectively, and the singular values are related to the eigenvalues by $\sigma_j^2 = \lambda_j$. Two important properties of the POD are:

- **Optimality:** This indicates that the POD is the most efficient decomposition, in the sense that for a given number of modes k , the projection on the subspace spanned by the leading k modes contains the greatest possible energy on average among all linear decompositions.
- **Orthogonality:** This implies that the time series of the coefficients $a_j(t)$ are linearly uncorrelated, which is an attractive property for constructing reduced-order models.

It is important to note that the POD is a linear procedure. Linearity is the origin of the strength of the method, its applicability, but it is also its limitation. As stated above regarding the optimality of the POD results, it should be noted that optimality is implied only with respect to other linear representations (Holmes et al., 1996).

We applied POD using the SVD method on the urban-flow database discussed above. Fig. 1 shows the eigenvalues λ_i (left) and the cumulative eigenvalue spectrum $\sum_{j=1}^i \lambda_j$ (right) normalized with the cumulative sum of the eigenvalues $\sum_{j=1}^m \lambda_j$, where i indicates the number of modes. We observed that 247 modes are required to capture 99% of the energy as it is depicted by the vertical red line in Fig. 1 (right). This result implies that it is impractical to represent turbulent flows as a linear superposition of a few modal functions, and thus, more sophisticated algorithms enabling a non-linear modal decomposition are required.

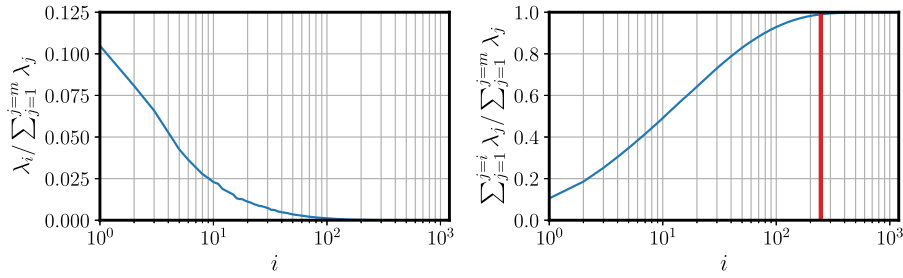


Fig. 1. Eigenvalues λ_i (left) and the cumulative eigenvalue spectrum $\sum_{j=1}^i \lambda_j$ (right) normalized with the cumulative sum of the eigenvalues $\sum_{j=1}^m \lambda_j$, where i indicates the number of modes. The solid red line shows the number of modes required to capture 99% of the energy.

3. CNN-based autoencoders for modal decomposition

An autoencoder is a deep neural network (DNN) with an architecture suitable for unsupervised feature extraction. The network comprises two parts: an encoder that maps the input data to a low-dimensional latent space $x \mapsto r$, and a decoder that projects the latent vector r back to the reference space $r \mapsto \tilde{x}$. We refer to the encoder and decoder parts as \mathcal{E} and \mathcal{D} , respectively. Through the model training, the autoencoder learns to extract the most important features in the data that are required for reconstruction by optimizing the model parameters w to minimize the reconstruction loss \mathcal{L}_{rec} :

$$F = \mathcal{D} \circ \mathcal{E}, \quad (7a)$$

$$\tilde{x} = F(x; w), \quad (7b)$$

$$\mathcal{L}_{\text{rec}} = \epsilon(x, \tilde{x}), \quad (7c)$$

where x denotes the input data, which in our case is the fluctuating component of the streamwise flow velocity; \tilde{x} is the reconstruction of the input data and ϵ represents the loss function. The autoencoder architecture is attractive for modal decomposition as it provides a framework that can incorporate non-linearity in the mappings through the use of non-linear activation functions.

Another challenge in the modal decomposition of turbulent flows is the process of information from input fields that contain multiscale coherent features. The presence of coherent features motivates the use of convolutional layers in autoencoder models to process the input information. Through a two-dimensional convolution layer and using a so-called filter with size $H \times H \times K$, the data is processed as:

$$z_{ijm}^{(l)} = \varphi \left(\sum_{k=1}^K \sum_{p=1}^H \sum_{q=1}^H z_{i+p,j+q,k}^{(l-1)} w_{pqkm}^{(l)} + b_{ijm}^{(l)} \right), \quad (8)$$

where $z^{(l-1)}$ and $z^{(l)}$ represent the input and output variables, respectively. Note that $w^{(l)}$ and $b^{(l)}$ denote the weights and the biases of layer l , and φ denotes the activation function. Furthermore, K is the number of the so-called kernels, which are the two-dimensional slices of the filter. Since $H \ll N_x, N_y$, the use of kernels greatly reduces the number of parameters that need to be learned during training. It is common to use convolution layers together with the pooling and upsampling layers. Through a pooling operation, the data are compressed by a factor of $(1/P)^2$ so that a region with size of $P \times P$ is represented by e.g. its maximum value; this is the so-called max-pooling operation. The upsampling operation is used to expand the data dimension by e.g. nearest-neighbor or bilinear interpolation. In the following, we discuss two NN-architectures based on CNNs and autoencoders, namely CNN-based autoencoders (CNN-AEs) and CNN-based hierarchical autoencoders (CNN-HAEs). Finally, we propose CNN-based β -variational autoencoders (CNN- β VAEs) for non-linear modal decomposition of turbulent flows. For all the models, we use mean-squared error as the loss function for reconstruction \mathcal{L}_{rec} and the Adam algorithm (Kingma & Ba, 2017) to optimize the model parameters w . All the models are

initialized using the Xavier scheme (Glorot & Bengio, 2010). We employ the early-stopping criterion and obtain the best model based on the validation loss to avoid overfitting, where 20% of the data is randomly selected for validation.

3.1. CNN-based autoencoders

Fig. 2 depicts a schematic representation of the CNN-AE model. For simplicity, consider the fluctuating component of the streamwise velocity u as the input/output of the model, but it is also possible to consider all the velocity components (u, v, w) as the input/output. The first convolution layer contains 16 filters with a size of (3×3) , and it is followed by a max-pooling layer with $P = 2$. At each convolution step, we double the number of feature maps to extract more information from the turbulent-flow data while at each downsampling step we reduce the dimension. This allows the next layer to combine the features individually identified in each feature map, enabling the extraction of larger and more complex features for progressively deeper convolutional networks from simple non-linear combinations of the previous ones. Therefore, convolutional layers can learn to recognize turbulent-flow patterns of various complexity and scales (Guastoni et al., 2021). After five steps of convolution and max pooling, the extracted features are flattened and fed to fully connected layers to reduce the dimension to the latent vector r with a size of d . The latent vector r is mapped back to the reference space through the consecutive upsampling and convolution operations using nearest-neighbor interpolation. Throughout the model, we use a filter size of (3×3) with the stride of one for convolution layers and (2×2) max pooling and upsampling operations with the same stride. We employ the hyperbolic-tangent (\tanh) function $\varphi(z) = (e^z - e^{-z}) / (e^z + e^{-z})$ as the non-linear activation function, since it led to the best performance in our study.

3.2. CNN-based hierarchical autoencoders

Based on the idea proposed by Fukami et al. (2020), Saegusa et al. (2004) proposed a CNN-based hierarchical autoencoder for modal decomposition of fluid flows in order to extract the modes ranked in terms of their contribution to the reconstruction while achieving more efficient data compression. To this end, the first subnetwork \mathcal{F}_1 is trained to map the high-dimensional data to a latent vector with size $d = 1$. The latent vector can be obtained using the encoder part of the first subnetwork as $r_1 = \mathcal{E}_1(x)$. The second subnetwork \mathcal{F}_2 is then trained to reconstruct the input data at the output from a two-dimensional latent vector comprising the first latent vector r_1 , which has already been obtained, and the second latent vector r_2 , being updated through the training of \mathcal{F}_2 , as $[r_1 \ r_2]$. The third network is trained in a similar way where the output is reconstructed from a latent vector as $[r_1 \ r_2 \ r_3]$. A schematic of the CNN-HAE architecture is illustrated in Fig. 3. In the present study, we employ the same hyperparameters as those of the CNN-AE model for the CNN-HAE.

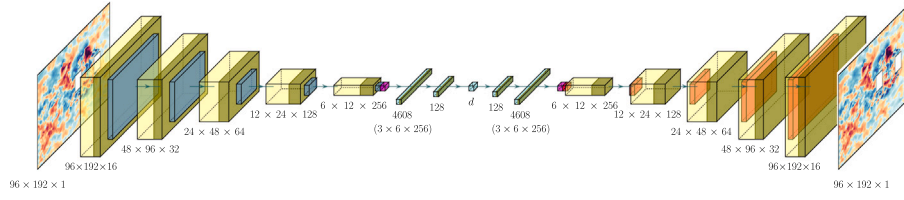


Fig. 2. Schematic view of the CNN-AE architecture. The color coding for each layer is: 2D-convolution (yellow), tanh activation (green), max pooling (blue), reshape (purple), fully-connected layer (cyan), upsampling (orange).

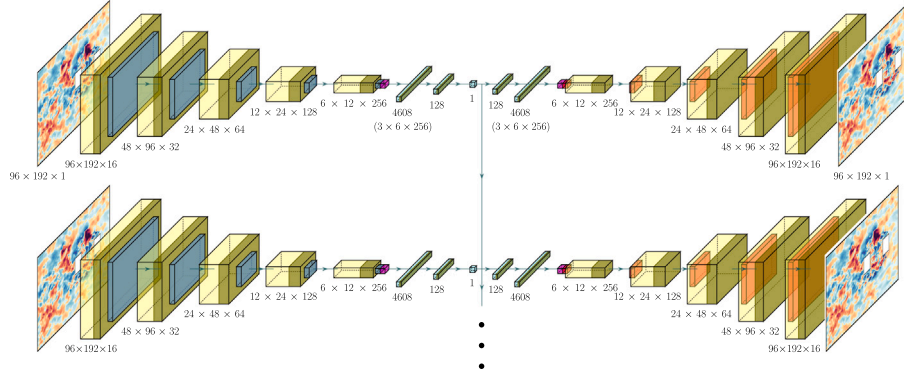


Fig. 3. Schematic view of the CNN-HAE architectures with the same color coding as in Fig. 2.

3.3. CNN-based β -variational autoencoders

In this section, we propose a modified version of the so-called variational autoencoders (VAEs) (Kingma & Welling, 2014; Rezende et al., 2014), a probabilistic generative neural architecture, for modal decomposition of turbulent flows. VAEs are powerful generative models emerging from the combination of statistics and information theory with the flexibility of DNNs to efficiently generate new high-dimensional data. The goal is to map the data to a latent distribution, from which new meaningful samples can be generated. VAEs have gained increasing attention in the scientific community (Iten et al., 2020; Maulik et al., 2020), both due to their strong probabilistic foundation and their precious application in the field of representation learning (Bengio et al., 2013).

3.3.1. Marginal likelihood

Let us consider a data sample \mathbf{x} in some high-dimensional space \mathcal{X} with the distribution $p(\mathbf{x})$. Let us assume that we have a vector of latent variables \mathbf{r} in a low-dimensional space \mathcal{R} with the probability density function (PDF) $p(\mathbf{r})$, from which we can easily sample new datapoints. Considering a family of deterministic functions $f(\mathbf{r}; \theta)$, parameterized by a vector θ in some space Θ in such a way that $f : \mathcal{R} \times \Theta \mapsto \mathcal{X}$, $f(\mathbf{r}; \theta)$ is a random variable in the space of \mathcal{X} if \mathbf{r} is random and θ is fixed. This implies that by optimizing θ to maximize the probability of each \mathbf{x} , we can sample \mathbf{r} from $p(\mathbf{r})$ and expect that $f(\mathbf{r}; \theta)$ resembles the samples \mathbf{x} in our dataset. Therefore, we aim to maximize:

$$p_{\theta}(\mathbf{x}) = \int p_{\theta}(\mathbf{x}|\mathbf{r})p(\mathbf{r})d\mathbf{r}, \quad (9)$$

where $f(\mathbf{r}; \theta)$ has been replaced by a distribution $p_{\theta}(\mathbf{x}|\mathbf{r})$ to show the dependence of \mathbf{x} on \mathbf{r} according to the law of total probability, and $p_{\theta}(\mathbf{x})$ is the so-called marginal likelihood, which is the approximation of $p(\mathbf{x})$ with parameters θ . This is the so-called maximum likelihood estimation (MLE), and the distribution $p(\mathbf{r})$ is often called the prior distribution. However, Eq. (9) is typically intractable due to the integral and it is usually computationally infeasible (Kingma & Welling, 2014).

3.3.2. Evidence lower bound (ELBO)

VAEs define another probability distribution $q_{\phi}(\mathbf{r}|\mathbf{x})$, a so-called probabilistic encoder (or recognition model). In a similar vein $p_{\theta}(\mathbf{x}|\mathbf{r})$ is referred to as a probabilistic decoder (or generative model). The marginal likelihood is obtained as a sum over the marginal likelihoods of individual datapoints:

$$\log p_{\theta}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}) = \sum_{i=1}^m \log p_{\theta}(\mathbf{x}^{(i)}), \quad (10)$$

and for each \mathbf{x} it can be defined as:

$$\log p_{\theta}(\mathbf{x}) = D_{\text{KL}}(q_{\phi}(\mathbf{r}|\mathbf{x}) \parallel p_{\theta}(\mathbf{r}|\mathbf{x})) + C(\theta, \phi; \mathbf{x}), \quad (11)$$

where the first right-hand-side (RHS) term is the Kullback–Leibler (KL) divergence D_{KL} between $q_{\phi}(\mathbf{r}|\mathbf{x})$ and $p_{\theta}(\mathbf{r}|\mathbf{x})$. The KL-divergence is non-negative, which indicates that the second RHS term $C(\theta, \phi; \mathbf{x})$ is a lower bound on the marginal likelihood, and can be written as:

$$\log p_{\theta}(\mathbf{x}) \geq C(\theta, \phi; \mathbf{x}) = -D_{\text{KL}}(q_{\phi}(\mathbf{r}|\mathbf{x}) \parallel p_{\theta}(\mathbf{r})) + \mathbb{E}_{q_{\phi}(\mathbf{r}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{r})]. \quad (12)$$

This term is usually called evidence lower bound (ELBO). Since the ELBO is more tractable than the MLE, it is used as the cost function for the training of NNs to optimize the unknown parameters θ and ϕ .

3.3.3. Reparameterization trick

Since the operation that samples a latent vector from $q_{\phi}(\mathbf{r}|\mathbf{x})$ is not differentiable, we need to perform a change of variable, the so-called reparameterization trick (Kingma & Welling, 2014), to differentiate ELBO with respect to both θ and ϕ . We assume $q_{\phi}(\mathbf{r}|\mathbf{x})$ to be a Gaussian distribution,

$$\log q_{\phi}(\mathbf{r}|\mathbf{x}) = \log \mathcal{N}(\mathbf{r}; \mu, \sigma^2 \mathbf{I}), \quad (13)$$

where the mean μ and the standard deviation σ are outputs of the encoder, and \mathbf{I} is the identity matrix. We sample from $q_{\phi}(\mathbf{r}|\mathbf{x})$ using $\mathbf{r} = \mu + \sigma \odot \epsilon$ where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is an auxiliary normally-distributed random number, and \odot indicates an element-wise product. Moreover, the term $\mathbb{E}_{q_{\phi}(\mathbf{r}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{r})]$, which is the so-called log-likelihood, encourages accurate reconstruction of the data and can be estimated as a negative

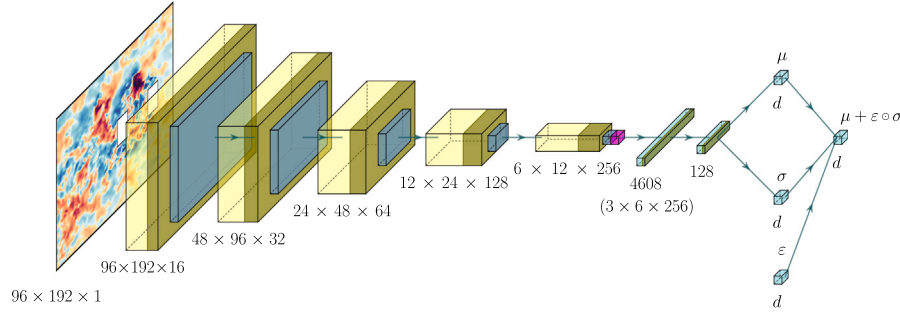


Fig. 4. Schematic view of the encoder part \mathcal{E} of the CNN- β VAE with the same color coding as in Fig. 2.

reconstruction error in an autoencoder setting (Kingma & Welling, 2014). This leads to the VAE cost function \mathcal{C} , and we can take the negative of it as a loss function \mathcal{L} for training the NNs:

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \mathcal{L}_{\text{rec}} - \frac{1}{2} \sum_{i=1}^d (1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2). \quad (14)$$

3.3.4. Disentangled representation

In the field of representation learning (Bengio et al., 2013), it is of interest to find a latent representation of the high-dimensional data as an uncorrelated representation with a minimal number of parameters (factors), the so-called disentangled representation, which can be useful for a large variety of tasks and domains. Higgins et al. (2017) proposed to augment the original VAE loss function with a single hyperparameter $\beta \geq 0$ that controls the extent of the learning constraints. The goal is to encourage learning of statistically-independent latent variables r_i and penalize the size of the latent vector r . This can be obtained by minimizing the distance $D_{\text{KL}}[p(r) \parallel \prod_i p(r_i)]$ between $p(r)$ and the product of its marginals. In practice, this is performed by upweighting the KL term in the ELBO, see Eq. (12), with a penalization factor β leading to the following loss function:

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \mathcal{L}_{\text{rec}} - \frac{\beta}{2} \sum_{i=1}^d (1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2) \quad (15)$$

for β -VAEs. A detailed discussion on disentangling in β -VAEs can be found in the works by Achille and Soatto (2018), Burgess et al. (2018), Higgins et al. (2017) and Locatello et al. (2019).

In the present work, we employ a CNN-based β -VAE architecture for modal decomposition of turbulent flows. Our goal is to minimize the correlation between the latent variables, motivating the network to extract a set of orthogonal modes, and also penalize the size of the latent vector d . This leads to an efficient representation of the high-dimensional data useful for flow analysis, reduced-order modeling, and flow control. It should be noted that for the testing steps the mean part of the encoder output μ of the CNN- β VAE is taken as the vector of latent variables r . Fig. 4 illustrates the schematic of the encoder part of the CNN- β VAE model. The decoder part is the same as the decoder part of the CNN-AE model depicted in Fig. 2.

4. Results and discussion

The key insight of the present study is to encourage independence of the latent variables r_1, \dots, r_d to extract near-orthogonal modes from turbulent flows using CNN- β VAEs. This is to motivate disentangled representations in the language of representation learning. We impose a limit on the capacity of the latent information and motivate learning statistically independent latent variables using the penalization factor β . The objective is to motivate orthogonality in the latent space to obtain modes that are useful for flow analysis, reduced-order modeling, and flow control. Algorithm 1 represents the training process of the CNN- β VAEs.

Algorithm 1: Training of the CNN- β VAE.

Model encoder \mathcal{E} and decoder \mathcal{D} of the CNN- β VAE. Note that θ and ϕ are the parameters of the encoder and the decoder, respectively.

Data $\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{val}}$.

$\theta, \phi \leftarrow$ random initialization using the Xavier scheme (Glorot & Bengio, 2010).

repeat

for \mathbf{x} in $\mathbf{X}_{\text{train}}$: # \mathbf{x} represents a batch of samples.

$\mu, \sigma, \epsilon \leftarrow \mathcal{E}(\mathbf{x})$

$r \leftarrow \mu + \sigma \odot \epsilon$

$\tilde{\mathbf{x}} \leftarrow \mathcal{D}(r)$

$\mathcal{L}_{\text{rec}} \leftarrow \text{MSE}(\mathbf{x}, \tilde{\mathbf{x}})$

$\mathcal{L} \leftarrow \text{mean}(\mathcal{L}_{\text{rec}} - \frac{1}{2} \sum_{i=1}^d (1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2))$

update θ, ϕ using the backpropagation algorithm.

$\mathcal{L}_{\text{val}} \leftarrow$ computed from \mathbf{X}_{val} .

if $\mathcal{L}_{\text{val}} < \min(\mathcal{L}_{\text{val}}, \text{history})$ then

 save model parameters θ and ϕ

until final epoch is reached.

The KL term in the β -VAE loss function, see Eq. (15), acts as a regularizer. This regularization may create a trade-off between reconstruction fidelity and the quality of learning independent representations. In this section, we investigate the effect of the penalization factor β and the size of the latent vector d on the performance of the CNN- β VAE models, both in terms of reconstruction accuracy and independence of the modes. Furthermore, we compare the performance of the CNN- β VAE in the modal decomposition of the turbulent flow through a simplified urban environment with that of the CNN-HAE, CNN-AE, and POD. To this end, we define two evaluation metrics to measure the quality of the reconstructions and the orthogonality (disentanglement) of the latent variables. For reconstruction quality, we evaluate the energy percentage E_k that is captured by the model reconstructions as:

$$E_k = \left(1 - \left\langle \frac{\sum_{i=1}^n (u - \tilde{u})^2}{\sum_{i=1}^n u^2} \right\rangle \right) \times 100, \quad (16)$$

where $\langle \cdot \rangle$ indicates ensemble averaging in time, u and \tilde{u} denote the reference value of the fluctuating component of streamwise velocity and its reconstruction, respectively, and n is the number of grid points. To measure the independency of the latent variables, we compute the determinant of the correlation matrix multiplied by 100 and refer to it as $\det_{\mathbf{R}}$, where $\mathbf{R} = (R_{ij})_{d \times d}$ is the correlation matrix defined by:

$$R_{ii} = 1 \text{ and } R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}}, \quad (17)$$

for all $1 \leq i \neq j \leq d$, and C_{ij} denotes the components i, j of the covariance matrix \mathbf{C} . Note that $\det_{\mathbf{R}}$ is 100 when all the variables are

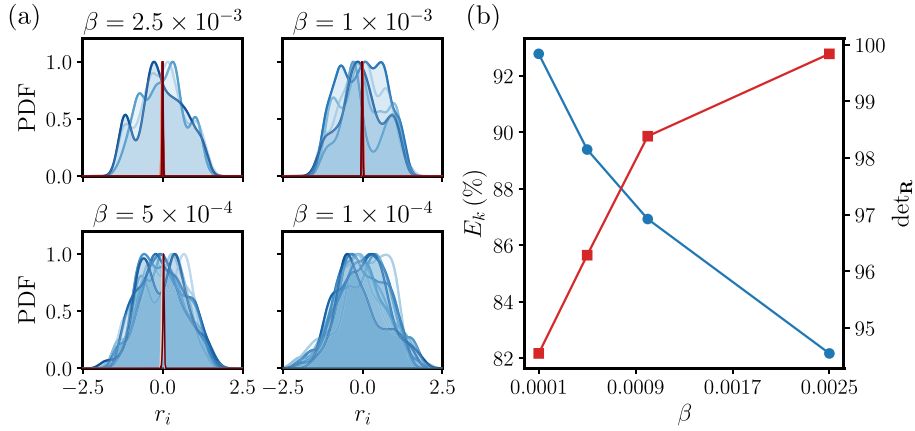


Fig. 5. Effects of the penalization factor β on the performance of the CNN- β VAE models. (a) Normalized probability density function (PDF) of the latent variables obtained from models with different values of β as indicated in each panel. The variables with considerably large values are colored by blue and the variables with very small deviations from zero are depicted by red. (b) Effect of β on E_k (blue) and \det_R (red).

completely uncorrelated ($R_{ij} = 0$) and zero when they are completely correlated ($R_{ij} = 1$). We report the value of \det_R as a metric for independency of the latent variables.

4.1. Learning orthogonal and parsimonious modes

The introduction of non-linearity in the process of modal decomposition using AEs through the use of non-linear activation functions leads to excellent performance in terms of reconstruction accuracy in comparison to the linear-theory-based decomposition methods such as POD. However, the AE-based modes lack the useful properties of the linear-theory-based modes such as orthogonality and ranking, which may lead to interpretability. We constrain the shape of the latent space to motivate orthogonality (disentanglement) of the features extracted by the CNN- β VAEs and obtain parsimonious (minimal) modes. The penalization factor β regulates the balance between the information preservation and the information capacity of the latent vector (Higgins et al., 2017).

The results obtained from CNN- β VAEs using four different values of β are reported in Fig. 5. For these tests we consider $d = 10$ and $\beta = 10^{-4}$, 5×10^{-4} , 10^{-3} , and 2.5×10^{-3} . Fig. 5(a) depicts the normalized probability density function (PDF) obtained from each variable. We note that using $\beta = 2.5 \times 10^{-3}$, only three latent variables have significant values and the others are comparatively very small. These variables have a very small deviation from zero, and thus can be regarded as noise. We also observe that the spatial modes corresponding to these variables are very similar. Therefore, by imposing a limit on the capacity of the latent information using $\beta = 2.5 \times 10^{-3}$ the network minimizes the latent-space dimension to three and leads to almost perfectly independent variables where $\det_R = 99.84$. This test also leads to a very good reconstruction of the turbulent flow with $E_k = 82.17\%$ as it is shown in Fig. 5(b). By reducing the penalization factor β to 10^{-3} , the latent-space dimension increases to five as a consequence of a lighter constrain on the capacity of the latent information, which also leads to a reduction of \det_R to 98.38. However, the reconstruction performance is improved leading to $E_k = 86.93\%$. By further reduction of β , the dimension of the latent space is increased where all the 10 latent variables have considerable values using $\beta = 10^{-4}$. Moreover, it can be observed in Fig. 5(b) that by reducing β , the independence of the variables diminishes while the reconstruction accuracy is improved. We obtain $\det_R = 94.55$ and $E_k = 92.78\%$ using $\beta = 10^{-4}$.

Next, we investigate the effect of the size of the latent vector d on the performance of the CNN- β VAEs. For these tests, we consider $\beta = 10^{-3}$ and $d = 5, 10, 20$. We observe that in all three tests, only

five latent variables have considerable values and the rest are just noise. Our results show that although we increase d , the penalization using $\beta = 10^{-3}$ encourages the network to learn a minimal number of near-orthogonal modes (only five) for accurate reconstruction of the turbulent flow. Since the network maps the information into only five variables, increasing d to more than five does not improve the performance of the model. We obtain \det_R equal to 99.20, 98.38, and 99.22 and E_k equal to 87.36%, 86.93%, and 87.15% for the CNN- β VAEs with d equal to 5, 10, and 20, respectively.

4.2. Ranking the CNN- β VAE modes

As we discussed in Section 2.2, POD modes are sorted in terms of their energy content. This property is extremely useful for understanding and analyzing the dominant patterns in complex flows. Fukami et al. (2020) implemented hierarchical autoencoders to extract AE-based modes in the order of their contribution in the reconstruction, which requires training multiple NNs and might be cumbersome especially for the extraction of higher-order modes. Here, we propose a strategy for ranking the CNN- β VAE modes, which is represented in Algorithm 2. Later, we show that CNN- β VAEs are able to extract near-orthogonal and parsimonious modes from turbulent flows. These properties allow us to rank these modes after the training process and based on their contribution to the reconstruction. In particular, we rank CNN- β VAE modes based on the maximum E_k that can be obtained from q modes, where q represents the rank. To this end, after the training process, we first use the encoder to map high-dimensional data to the latent vector $\mathcal{E} : \mathbf{x} \mapsto \mathbf{r}$. We zero out all the latent variables except the i th variable, which leads to a latent vector $\hat{\mathbf{r}}_i$. Then, we employ the decoder part of the CNN- β VAE to send this latent vector to the original space $\mathcal{D} : \hat{\mathbf{r}}_i \mapsto \hat{\mathbf{x}}_i$. This procedure is performed for all the time steps. The energy percentage that is captured by only considering the i th mode is evaluated as E_k^i . The first mode is selected as the mode leading to the maximum value of E_k^i . For the second mode, we perform the same procedure while we preserve the first mode and look for the mode, which in combination with the first mode, leads to the maximum value of E_k^i . In a similar way, the third mode is selected as the mode which gives the maximum E_k^i in combination with the first and second modes. We continue this procedure to rank all the modes. Fig. 6 illustrates the ranked modes obtained from a CNN- β VAE model with $d = 5$ and $\beta = 10^{-3}$, together with the modes obtained from the POD, CNN-AE, and CNN-HAE methods with $d = 5$. The POD and the CNN-HAE modes are already ranked and we also perform the ordering procedure for the CNN-AE modes. A clear resemblance can be

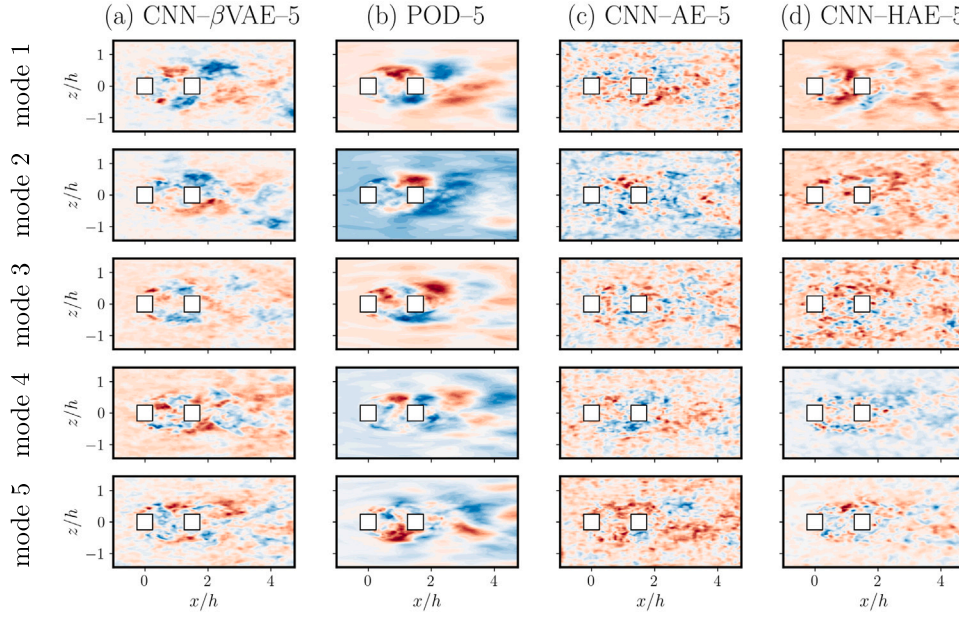


Fig. 6. The ranked spatial modes obtained from the CNN- β VAE (a), POD (b), CNN-AE (c), and the CNN-HAE (d). The size of the latent vector d is equal to 5.

observed between the first two modes of the CNN- β VAE, see Fig. 6(a), and those from POD, as shown in Fig. 6(b), indicating the ability of CNN- β VAEs in the extraction of interpretable modes from turbulent flows. These modes correspond to the large-scale vortex shedding from around the obstacles into the wake region. Moreover, these results show that using the ranking procedure it is possible to sort the CNN- β VAE modes based on their importance for reconstruction. It also can be seen in Fig. 6(c) that it is extremely difficult to relate CNN-AE modes to physical processes, a fact that is referred to as the lack of interpretability. We observe that although the CNN-HAE model is able to extract large-scale features first, the obtained modes may not be physically interpretable, as shown in Fig. 6(d).

Algorithm 2: Ranking the CNN- β VAE modes.

Model trained encoder \mathcal{E} and decoder \mathcal{D} of the CNN- β VAE.
Data \mathbf{x}
Initialize J vector containing indices of the ranked modes.
 $d \leftarrow$ vector containing indices of the modes.
 $\mu, \sigma, \epsilon \leftarrow \mathcal{E}(\mathbf{x})$
 $\mathbf{r} \leftarrow \mu$
for j **in** d :
 Initialize E
 for i **in** $d[\sim J]$:
 $I \leftarrow [i, J]$ # concatenation
 $\hat{\mathbf{r}}_i \leftarrow$ zero out all the latent variables except members of I
 $\tilde{\mathbf{x}}_i \leftarrow \mathcal{D}(\hat{\mathbf{r}}_i)$
 $E \leftarrow E_k(\mathbf{x}, \tilde{\mathbf{x}}_i)$ # append
 $J \leftarrow d[\sim J][\text{argmax}(E)]$ # append
Output: J

4.3. Comparison of CNN-AE-based models

In previous sections, we proposed CNN- β VAEs for modal decomposition of turbulent flows and showed their ability in the extraction

of near-orthogonal and parsimonious modes. Here, we compare the performance of CNN- β VAEs with that of CNN-HAEs, CNN-AEs and POD in terms of reconstruction accuracy and orthogonality of the latent variables. We select $d = 5$ as the size of the latent vector for all models. For the CNN- β VAE we use $\beta = 10^{-3}$. Fig. 7 shows the reconstruction of the first time step in the dataset obtained from different methods in comparison with the reference data. It can be observed that a more accurate reconstruction can be obtained using NN-based methods in comparison with 5 POD modes due to the introduction of non-linearity in the algorithm. The CNN-AE model leads to the best reconstructions with $E_k = 94.22\%$ while 32.41% of the energy is captured by the 5 POD modes. Both CNN-HAE and CNN- β VAE models also lead to excellent reconstructions with E_k of 91.84% and 87.36%, respectively, which are slightly lower than that of the CNN-AE. For the CNN- β VAE, it is due to the fact that the regularization with the KL term in the β -VAE loss function, Eq. (15), induces a trade-off between the reconstruction quality and learning independent representations. However, our results show that excellent reconstructions can be obtained using all the three CNN-AE-based models leading to E_k of about 90% using only 5 modes.

The reconstruction results are also reported in a detailed area between the two obstacles in Fig. 8 to provide a clear insight into the fidelity of the reconstructions. It can be seen that although some small-scale features are lost, all three CNN-AE models are able to preserve the dominant structures of the turbulent flow. However, POD cannot reconstruct the turbulent flow properly from 5 modes.

Next, we compare the independence of the latent variables obtained from different methods. As we discussed in Sections 2.2 and 4.1, orthogonality of the modes is a useful property for flow analysis, reduced-order modeling and flow control. Moreover, motivating the orthogonality of the modes may lead to interpretability. Results are depicted in Fig. 9 as the absolute value of the correlation matrix \mathbf{R} corresponding to the latent variables from the CNN-HAE, CNN-AE, CNN- β VAE with $d = 5$, and also POD with 5 modes. It can be seen that although the CNN-HAE model extract modes in the order of their contribution in the reconstruction, the latent variables are correlated leading to the lowest value for $\det \mathbf{R}$ among all methods. As mentioned above, it is possible to motivate the disentanglement or independence

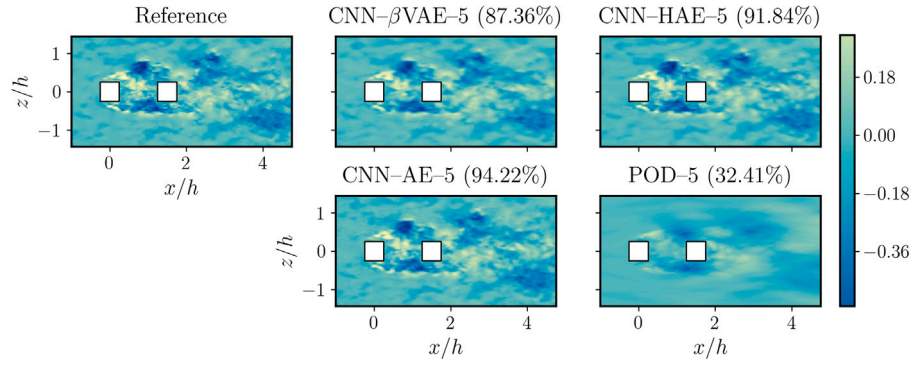


Fig. 7. Reconstruction of the fluctuating component of the streamwise velocity obtained from different methods, as indicated in each panel, in comparison with the reference data. The value in brackets on each panel indicates the obtained E_k .

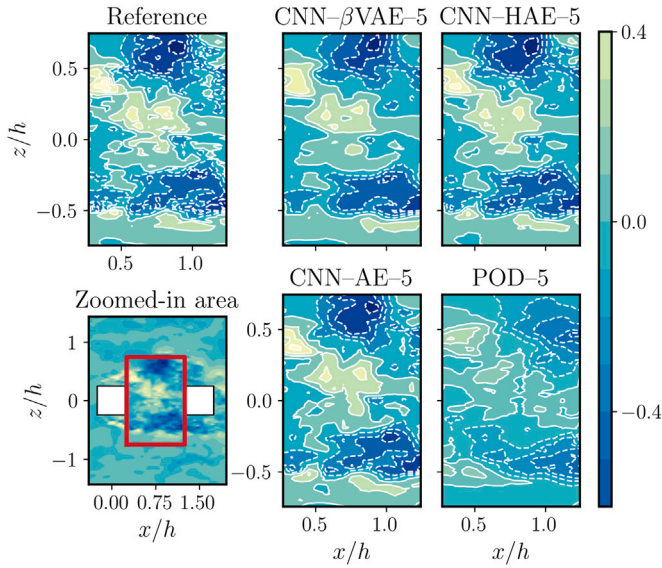


Fig. 8. Reconstruction of the fluctuating component of the streamwise velocity obtained from different methods, as indicated on each panel, in comparison with the reference data for the zoomed-in area between the two obstacles marked by the red rectangle.

of the latent variables using CNN- β VAEs and obtain near-orthogonal modes. It can be observed that the correlation between the latent variables is reduced for the CNN- β VAE in comparison to that of the CNN-AE, where $\det_{\mathbf{R}}$ is equal to 99.20 for the CNN- β VAE method and 87.59 for the CNN-AE technique.

5. Summary and conclusions

In this study, we propose a probabilistic deep-neural-network architecture based on β -VAEs and CNNs for non-linear mode decomposition of turbulent flows. The objective is to learn a compact, near-orthogonal and parsimonious latent representation of high-dimensional data by introducing non-linearity in the process of dimension reduction and also minimizing the correlation between the latent variables, as well as penalizing the size of the latent vector. This may lead to a set of interpretable modes useful for flow analysis, reduced-order modeling and flow control. Since the correlations among the learned latent variables are minimized, we proposed an algorithm to rank the VAE-based modes based on their contribution to the reconstruction. We applied the proposed CNN- β VAE architecture for modal decomposition of the

turbulent flow through a simplified urban environment. Furthermore, we compared the performance of the CNN- β VEs in terms of the quality of the reconstructions and orthogonality of the extracted modes with that of the CNN-AEs and CNN-HAEs. The flow-field data were obtained from a well-resolved large-eddy simulation (LESs) database (Torres, 2021; Torres et al., 2021). Our results from modal decomposition using POD indicates that 247 modes are required to obtain 99% of the energy from the reconstruction. This indicates that it is challenging to represent turbulent flows as a linear superposition of a few POD modes. Our proposed CNN- β VAE model with a latent vector size $d = 5$ and a penalization factor of $\beta = 10^{-3}$ leads to E_k of 87.36% against 32.41% obtained from POD, which shows the excellent performance of the CNN- β VAE in the reconstruction of the turbulent flow from only five modes. This model also leads to near-orthogonal modes, where $\det_{\mathbf{R}}$ is equal to 99.20. We showed that by constraining the shape of the latent space and motivating orthogonality of the modes, we can extract meaningful non-linear features where the first mode of this CNN- β VAE model represents the large-scale vortex shedding from around the obstacles into the wake. Moreover, we investigated the effects of the size of the latent vector d and the penalization factor β on the performance of the model. Our results indicate the ability of the CNN- β VAEs in learning a minimal (parsimonious) set of near-orthogonal modes which are required for reconstruction, where the penalization encourages the model to minimize the number of learned latent variables and it does not change by increasing d . Moreover, we observed that there is a trade-off between the reconstruction accuracy and the quality of learning independent representations. Our results showed that a lighter constrain on the capacity of the latent information leads to a slightly better reconstruction while allowing more correlations among the latent variables. Our comparison between the CNN- β VAE, CNN-HAE, and CNN-AE models indicates that although motivating orthogonality of the modes decreases the reconstruction accuracy, very good reconstructions can be obtained from five modes using the CNN- β VAE leading to E_k of 87.36% against 93.93% and 91.84% of the CNN-AE and the CNN-HAE, respectively. The CNN- β VAE model leads to a set of near-orthogonal modes with the highest $\det_{\mathbf{R}}$ among the AE-based models.

The proposed CNN- β VAE architecture can be extended in future works for the development of reduced-order surrogate models or advanced flow-control methods, among others. In particular, the proposed method can be employed for non-linear modal decomposition of the data obtained from thermal-imaging cameras in complex urban environments, with extensive application in the context of urban air-quality forecast and control.

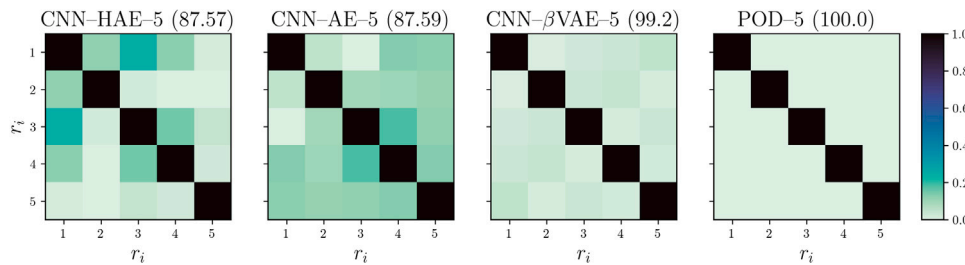


Fig. 9. Correlation matrix \mathbf{R} for the latent variables obtained from different models as indicated on the panels. The value in brackets indicates the corresponding $\det_{\mathbf{R}}$ for each case.

CRedit authorship contribution statement

Hamidreza Eivazi: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Visualization.
Soledad Le Clainche: Methodology, Writing – review & editing.
Sergio Hoyas: Methodology, Writing – review & editing.
Ricardo Vinuesa: Conceptualization, Methodology, Data processing, Writing – review & editing, Supervision, Funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge Álvaro Martínez for his contributions to this work. RV acknowledges the Göran Gustafsson foundation, Sweden for the financial support of this research. SH has been supported by project RTI2018-102256-B-I00 of Mineco/FEDER. SLC acknowledges the support of the Spanish Ministry of Science and Innovation under the grant PID2020-114173RB-I00.

References

- Abadía-Heredia, R., López-Martín, M., Carro, B., Arribas, J. I., Pérez, J. M., & Le Clainche, S. (2022). A predictive hybrid reduced order model based on proper orthogonal decomposition combined with deep learning architectures. *Expert Systems with Applications*, 187, Article 115910. <http://dx.doi.org/10.1016/j.eswa.2021.115910>.
- Achille, A., & Soatto, S. (2018). Information dropout: Learning optimal representations through noisy computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40, 2897–2905. <http://dx.doi.org/10.1109/TPAMI.2017.2784440>.
- Albu, A., Precup, R. -E., & Teban, T. -A. (2019). Results and challenges of artificial neural networks used for decision-making and control in medical applications. *Facta Universitatis, Series: Mechanical Engineering*, 17, 285–308. <http://dx.doi.org/10.22190/FUME190327035A>.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 1798–1828. <http://dx.doi.org/10.1109/TPAMI.2013.50>.
- Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52, 477–508.
- Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *PANS*, 113, 3932–3937.
- Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., & Lerchner, A. (2018). Understanding disentangling in β -VAE. [arXiv:1804.03599](https://arxiv.org/abs/1804.03599).
- Clainche, S. L., & Vega, J. M. (2018). Spatio-temporal Koopman decomposition. *Journal of Nonlinear Science*, 28, 1793–1842. <http://dx.doi.org/10.1007/s00332-018-9464-z>.
- Dhini, A., Surjandari, I., Kusumoputro, B., & Kusiak, A. (2021). Extreme learning machine – radial basis function (ELM-RBF) networks for diagnosing faults in a steam turbine. *Journal of Industrial and Production Engineering*, 1–9. <http://dx.doi.org/10.1080/21681015.2021.1887948>.
- Duraisamy, K., Iaccarino, G., & Xiao, H. (2019). Turbulence modeling in the age of data. *Annual Review of Fluid Mechanics*, 51, 357–377.

- Eivazi, H., Guastoni, L., Schlatter, P., Azizpour, H., & Vinuesa, R. (2021). Recurrent neural networks and koopman-based frameworks for temporal predictions in a low-order model of turbulence. *International Journal of Heat and Fluid Flow*, 90, Article 108816. <http://dx.doi.org/10.1016/j.ijheatfluidflow.2021.108816>.
- Eivazi, H., Tahani, M., Schlatter, P., & Vinuesa, R. (2021). Physics-informed neural networks for solving Reynolds-averaged Navier–Stokes equations. In *Proc. 13th ERCOFTAC symp. on engineering turbulence modelling and measurements (etmm13)*. Rhodes, Greece, September 16–17: [arXiv:1804.03599](https://arxiv.org/abs/1804.03599).
- Eivazi, H., Veisi, H., Naderi, M. H., & Eshfahanian, V. (2020). Deep neural networks for nonlinear model order reduction of unsteady flows. *Physics of Fluids*, 32, Article 105104. <http://dx.doi.org/10.1063/5.0020526>.
- European Environment Agency, Soares, J., Guerreiro, C., & González Ortiz, A. (2020). *Air quality in Europe: 2020 report*. LU: Publications Office of the European Union, <http://dx.doi.org/10.2800/786656>.
- Fischer, P., Lottes, J., & Kerkemeier, S. (2008). Nek5000 web page. <http://nek5000.mcs.anl.gov>.
- Fukami, K., Fukagata, K., & Taira, K. (2019). Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics*, 870, 106–120. <http://dx.doi.org/10.1017/jfm.2019.238>.
- Fukami, K., Nakamura, T., & Fukagata, K. (2020). Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data. *Physics of Fluids*, 32, Article 095110. <http://dx.doi.org/10.1063/5.0020721>.
- Gin, C. R., Shea, D. E., Brunton, S. L., & Kutz, J. N. (2020). DeepGreen: Deep learning of Green's functions for nonlinear boundary value problems. [arXiv:2101.07206](https://arxiv.org/abs/2101.07206).
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feed-forward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256). Chia Laguna Resort, Sardinia, Italy: URL: <https://proceedings.mlr.press/v9/glorot10a.html>.
- Guastoni, L., Güemes, A., Ianiro, A., Discetti, S., Schlatter, P., Azizpour, H., & Vinuesa, R. (2021). Convolutional-network models to predict wall-bounded turbulence from wall quantities. *Journal of Fluid Mechanics*, 928, A27. <http://dx.doi.org/10.1017/jfm.2021.812>.
- Güemes, A., Discetti, S., Ianiro, A., Sirmacek, B., Azizpour, H., & Vinuesa, R. (2021). From coarse wall measurements to turbulent velocity fields with deep learning. *Physics of Fluids*, 33, Article 075121.
- Hedrea, E. -L., Precup, R. -E., Roman, R. -C., & Petriu, E. M. (2021). Tensor product-based model transformation approach to tower crane systems modeling. *Asian Journal of Control*, 23, 1313–1323. <http://dx.doi.org/10.1002/asjc.2494>.
- Higgins, I., Matthey, L., Pal, A., Burgess, C. P., Glorot, X., Botvinick, M., Mohamed, S., & Lerchner, A. (2017). beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313, 504–507. <http://dx.doi.org/10.1126/science.1127647>.
- Holmes, P., Lumley, J. L., & Berkooz, G. (1996). *Cambridge monographs on mechanics, Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge University Press, <http://dx.doi.org/10.1017/CBO9780511622700>.
- Iten, R., Metger, T., Wilming, H., del Rio, L., & Renner, R. (2020). Discovering physical concepts with neural networks. *Physical Review Letters*, 124, Article 010508. <http://dx.doi.org/10.1103/PhysRevLett.124.010508>.
- Jiang, C., Vinuesa, R., Chen, R., Mi, J., Laima, S., & Li, H. (2021). An interpretable framework of data-driven turbulence modeling using deep neural networks. *Physics of Fluids*, 33, Article 055133.
- Jiménez, J. (2018). Machine-aided turbulence theory. *Journal of Fluid Mechanics*, 854, R1. <http://dx.doi.org/10.1017/jfm.2018.660>.
- Jin, X., Cai, S., Li, H., & Karniadakis, G. E. (2021). NSFnets (Navier–Stokes flow nets): Physics-informed neural networks for the incompressible Navier–Stokes equations. *Journal of Computational Physics*, 426, Article 109951. <http://dx.doi.org/10.1016/j.jcp.2020.109951>.
- Karhunen, K. (1946). Zur spektraltheorie stochastischer prozesse. *Annales Academiae Scientiarum Fennicae, Series AI (Mathematica)*, 34.
- Kashani, M., Arashi, M., Rabiei, M. R., D'Urso, P., & De Giovanni, L. (2021). A fuzzy penalized regression model with variable selection. *Expert Systems with Applications*, 175, Article 114696. <http://dx.doi.org/10.1016/j.eswa.2021.114696>.

- Kim, H., Kim, J., Won, S., & Lee, C. (2021). Unsupervised deep learning for super-resolution reconstruction of turbulence. *Journal of Fluid Mechanics*, 910, A29. <http://dx.doi.org/10.1017/jfm.2020.1028>.
- Kim, J., & Lee, C. (2020). Prediction of turbulent heat transfer using convolutional neural networks. *Journal of Fluid Mechanics*, 882, A18. <http://dx.doi.org/10.1017/jfm.2019.814>.
- Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv:1412.6980*.
- Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. *arXiv:1312.6114*.
- Kutz, J. N. (2017). Deep learning in fluid dynamics. *Journal of Fluid Mechanics*, 814, 1–4.
- Le Clainche, S., & Vega, J. M. (2017). Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16, 882–925. <http://dx.doi.org/10.1137/15M1054924>.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 2278–2324. <http://dx.doi.org/10.1109/5.726791>.
- Lee, K., & Carlberg, K. T. (2020). Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404, Article 108973. <http://dx.doi.org/10.1016/j.jcp.2019.108973>.
- Lee, S., & You, D. (2019). Data-driven prediction of unsteady flow over a circular cylinder using deep learning. *Journal of Fluid Mechanics*, 879, 217–254. <http://dx.doi.org/10.1017/jfm.2019.700>.
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021). Fourier neural operator for parametric partial differential equations. *arXiv:2102.08895*.
- Ling, J., Kurzwski, A., & Templeton, J. (2016). Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, 807, 155–166. <http://dx.doi.org/10.1017/jfm.2016.615>.
- Locatello, F., Bauer, S., Lucic, M., Ratsch, G., Gelly, S., Schölkopf, B., & Bachem, O. (2019). Challenging common assumptions in the unsupervised learning of disentangled representations. *arXiv:1811.12359*.
- Loève, M. (1955). *Probability theory*. Princeton, New York.
- Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3, 218–229. <http://dx.doi.org/10.1038/s42256-021-00302-5>.
- Lumley, J. L. (1967). The structure of inhomogeneous turbulent flows. In A. M. Yaglom, & V. I. Tatarski (Eds.), *Atmospheric turbulence and radio wave propagation* (pp. 166–178). Moscow: Nauka.
- Lusch, B., Kutz, J. N., & Brunton, S. L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9, 4950. <http://dx.doi.org/10.1038/s41467-018-07210-0>.
- Martínez-Sánchez, A., Lazpita, E., Corrochano, A., Clainche, S. L., Hoyas, S., & Vinuesa, R. (2022). Data-driven assessment of arch vortices in urban flows. *arXiv:2202.01667*.
- Maulik, R., Fukami, K., Ramachandra, N., Fukagata, K., & Taira, K. (2020). Probabilistic neural networks for fluid flow surrogate modeling and data recovery. *Physical Review Fluids*, 5, Article 104401. <http://dx.doi.org/10.1103/PhysRevFluids.5.104401>.
- Milano, M., & Koumoutsakos, P. (2002). Neural network modeling for near wall turbulent flow. *Journal of Computational Physics*, 182, 1–26. <http://dx.doi.org/10.1006/jcph.2002.7146>.
- Murata, T., Fukami, K., & Fukagata, K. (2020). Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *Journal of Fluid Mechanics*, 882, A13. <http://dx.doi.org/10.1017/jfm.2019.822>.
- Pathak, J., Hunt, B., Girvan, M., Lu, Z., & Ott, E. (2018). Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical Review Letters*, 120, Article 024102. <http://dx.doi.org/10.1103/PhysRevLett.120.024102>.
- Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707. <http://dx.doi.org/10.1016/j.jcp.2018.10.045>.
- Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. *arXiv:1401.4082*.
- Rowley, C. W. (2005). Model reduction for fluids using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, 15, 997–1013. <http://dx.doi.org/10.1142/S0218127405012429>.
- Rowley, C. W., Mezić, I., Bagheri, S., Schlatter, P., & Henningson, D. S. (2009). Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641, 115–127. <http://dx.doi.org/10.1017/S0022112009992059>.
- Saegusa, R., Sakano, H., & Hashimoto, S. (2004). Nonlinear principal component analysis to preserve the order of principal components. *Neurocomputing*, 61, 57–70. <http://dx.doi.org/10.1016/j.neucom.2004.03.004>.
- Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656, 5–28.
- Sirovich, L. (1987). Turbulence and the dynamics of coherent structures, part I: Coherent structures. *Quarterly of Applied Mathematics*, 45, 561–571.
- Srinivasan, P. A., Guastoni, L., Azizpour, H., Schlatter, P., & Vinuesa, R. (2019). Predictions of turbulent shear flows using deep neural networks. *Physical Review Fluids*, 4, Article 054603.
- Taira, K., Brunton, S. L., Dawson, S. T. M., Rowley, C. W., Colonius, T., McKeon, B. J., Schmidt, O. T., Gordeyev, S., Theofilis, V., & Ukeiley, L. S. (2017). Modal analysis of fluid flows: An overview. *AIAA J.*, 55, 4013–4041. <http://dx.doi.org/10.2514/1.J056060>.
- Taira, K., Hemati, M. S., Brunton, S. L., Sun, Y., Duraisamy, K., Bagheri, S., Dawson, S. T. M., & Yeh, C. -A. (2020). Modal analysis of fluid flows: Applications and outlook. *AIAA J.*, 58, 998–1022. <http://dx.doi.org/10.2514/1.J058462>.
- Tang, H., Rabault, J., Kuhnle, A., Wang, Y., & Wang, T. (2020). Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning. *Physics of Fluids*, 32, Article 053605. <http://dx.doi.org/10.1063/5.0006492>.
- Torres, P. (2021). *High-order spectral simulations of the flow in a simplified urban environment: A study of flow statistics* (Master thesis), Spain: Polytechnic University of Valencia.
- Torres, P., Le Clainche, S., & Vinuesa, R. (2021). On the experimental, numerical and data-driven methods to study urban flows. *Energies*, 14, 1310.
- Towne, A., Schmidt, O. T., & Colonius, T. (2018). Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics*, 847, 821–867. <http://dx.doi.org/10.1017/jfm.2018.283>.
- Tseng, M. -L., Tran, T. P. T., Ha, H. M., Bui, T. -D., & Lim, M. K. (2021). Sustainable industrial and operation engineering trends and challenges Toward Industry 4.0: A data driven analysis. *Journal of Industrial and Production Engineering*, 38, 581–598. <http://dx.doi.org/10.1080/21681015.2021.1950227>.
- UN General Assembly (2015). Transforming our world: The 2030 agenda for sustainable development. Resolut. A/RES/70/1, 25, (pp. 1–35). URL: <https://undocs.org/en/A/RES/70/1>.
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11, 233.
- Vinuesa, R., & Brunton, S. L. (2021). The potential of machine learning to enhance computational fluid dynamics. *arXiv:2110.02085*.
- Vinuesa, R., Schlatter, P., Malm, J., Mavriplis, C., & Henningson, D. S. (2015). Direct numerical simulation of the flow around a wall-mounted square cylinder under various inflow conditions. *Journal of Turbulence*, 16, 555–587. <http://dx.doi.org/10.1080/14685248.2014.989232>.
- Vinuesa, R., & Sirmacek, B. (2021). Interpretable deep-learning models to help achieve the sustainable development goals. *Nature Machine Intelligence*, 3, 926. <http://dx.doi.org/10.1038/s42256-021-00414-y>.