

# ModelFLOWs-app: data-driven post-processing and reduced order modelling tools

Ashton Hetherington<sup>a</sup>, Adrián Corrochano<sup>a</sup>, Rodrigo Abadía-Heredia<sup>a</sup>, Eneko Lazpita<sup>a</sup>, Eva Muñoz<sup>b,c</sup>, Paula Díaz<sup>a</sup>, Egoitz Maiora<sup>a</sup>, Manuel López-Martín<sup>a</sup>, Soledad Le Clainche<sup>1a</sup>

<sup>a</sup>*ETSI Aeronáutica y del Espacio, Universidad Politécnica de Madrid, Plaza Cardenal Cisneros, 3, Madrid, 28040, Spain*

<sup>b</sup>*Aero-Thermo-Mechanics Department, Université Libre de Bruxelles, Brussels, 1000, Belgium*

<sup>c</sup>*Brussels Institute for Thermal-fluid systems and clean Energy (BRITE), Université Libre de Bruxelles and Vrije Universiteit Brussel, Brussels, 1000, Belgium*

---

## Abstract

This article presents an innovative open-source software named ModelFLOWs-app<sup>1</sup>, written in Python, which has been created and tested to generate precise and robust hybrid reduced order models (ROMs) fully data-driven. By integrating modal decomposition and deep learning methods in diverse ways, the software uncovers the fundamental patterns in dynamic systems. This acquired knowledge is then employed to enrich the comprehension of the underlying physics, reconstruct databases from limited measurements, and forecast the progression of system dynamics. These hybrid models combine experimental and numerical database, and serve as accurate alternatives to numerical simulations, effectively diminishing computational expenses, and also as tools for optimization and control. The ModelFLOWs-app software has demonstrated in a wide range of applications its great capability to develop reliable data-driven hybrid ROMs, highlighting its potential in understanding complex non-linear dynamical systems and offering valuable insights into various applications. This article presents the mathematical background, review some examples of applications and introduces a short tutorial of ModelFLOWs-app.

---

<sup>1</sup>Correspondence to: soledad.leclainche@upm.es

<sup>1</sup>The website of the software is available at <https://modelflows.github.io/modelflowsapp/>

*Keywords:* open-source software, reduced order models, data analysis, deep-learning, patterns identification, data-driven methods

---

## 1. Introduction

The availability of high-quality data has sparked a revolution in machine learning and reduced order modeling. Data-driven equation-free models offer a promising approach to understanding complex non-linear dynamical systems, even without prior knowledge of the governing equations. Machine learning tools allow us to extract knowledge directly from the data, rather than relying solely on theoretical principles. This shift encourages the use of data to uncover new hypothesis and models. While these techniques present challenges, they also provide significant opportunities for advancing our understanding of such systems, with applications in various industries including aerospace, automotive, construction, pharmaceuticals, chemicals, manufacturing, and more.

There are two primary approaches to data-driven modeling. The first approach involves data forecasting models, which focus on predicting future data using machine learning techniques like deep neural networks. These models do not incorporate explicit physical insights into their construction. The second approach is represented by reduced order models (ROMs) with physical insights. These hybrid models incorporate physical understanding and utilize pattern identification techniques such as proper orthogonal decomposition [1] or dynamic mode decomposition [2] to extract relevant spatio-temporal information from the data.

By employing hybrid data-driven ROMs in non-linear dynamical systems, several advantages emerge over relying solely on deep neural networks. These ROMs enable the identification of key instabilities and mechanisms within the studied database by capturing important information about the underlying physics. Furthermore, they facilitate the development of powerful tools for optimization and control. Having a deeper physical understanding of the problem allows for enhanced system phase prediction, adoption of more controlled and robust strategies, reduction in computational costs for numerical simulations, and streamlined information collection in experiments.

This article introduces the methodology and some relevant results obtained with a totally novel software, named as ModelFLOWS-app, which has been tested suitable to develop accurate and robust fully data-driven hybrid

ROMs. Modal decomposition and deep learning strategies are combined in several ways to reveal the main patterns in dynamical systems, and uses this information to understand the physics of the problem under study, to reconstruct databases from sparse measurements or to predict the evolution of the system dynamics, providing accurate alternatives to numerical simulations, reducing in this way related computational cost.

The origin of this tool was for the analysis of complex flows, where ModelFLOWs-app has been tested in a wide range of applications, for instance, (i) for temporal forecasting in reactive flows [3], wind turbines [4], compressible flows with buffeting [5], etc., (ii) for identification of flow instabilities and patterns in wall bounded turbulent flows [6, 7], synthetic jets [8], urban flows [9], etc., (iii) for data reconstruction in experiments modelling turbulent or complex flows [10, 11, 12], etc. Additionally, due to the robustness and exceptional properties presented by the models and algorithms composing ModelFLOWs-app, the software has been tested and extended to a wide range of industrial applications solving problems in complex non-linear dynamical systems. Some examples of the multiple applications of this tool include, patterns identification and reconstruction in medical imaging [13, 14], wind velocity predictions with LiDAR experimental measurements [15], identification of cross-flow instabilities [4], prediction of flutter instability in flight test experiments [16, 17], to name a few. We invite the readers to review carefully the applications presented in this article and to complement their knowledge with the tutorials and videos presented in ModelFLOWs-app website [18].

The article is organized as follows. A general description of the methodological framework behind ModelFLOWs-app is introduced in §2, and the methodology connected to the two big modules forming this software is shown in §3 and §4. A review of some of the main results of ModelFLOWs-app is presented in §5. Finally, §6 explain the main conclusions.

## 2. Methodology

ModelFLOWs-app methodological framework is formed by two big modules: Module 1 uses modal decomposition methods, and Module 2 is formed by hybrid machine learning tools, which combine modal decomposition with deep learning architectures. Each module solves three different applications: (1) patterns identification, suitable to study the physics behind the data analysed; (2) data reconstruction, capable to reconstruct two- or three- di-

dimensional databases from a set of selected points, using data from sensors, or repairing missing data; (3) data forecasting, which builds reduced order models (ROMs) to predict the spatio-temporal evolution of the signal analysed. Figure 1 shows an sketch with the general organization of the software.

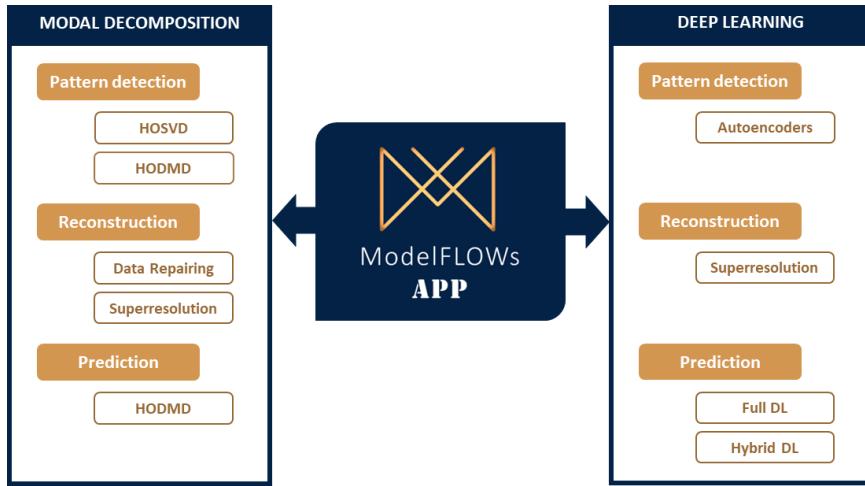


Figure 1: General distribution of ModelFLOWS-app software.

This section introduces the mathematical principles behind each Module, and also provides some examples of application. A detailed guide on how to treat and analyse the selected databases to exploit the properties of this software is presented in the ModelFLOWS-app website [18]. The software is presented in two different modalities. The first one is the application-web, which is mounted on an interface over *streamlit* [19]. This modality is suitable to understand what are the possibilities of ModelFLOWS-app, what are the setting parameters to run each application, what are the input and output files and to understand the capabilities of the software. The second one is formed by the main code, written in Python [20], which calls the modules and applications. This modality is suitable to be used in complex problems and big databases. The codes are open to the general public, so it can be locally modified and adapted to the user needs.

### 2.1. Data organization

ModelFLOWS-app is a fully data-driven software, so the user should first provide the desired databases to analyze and next select the application to

solve.

In matrix form, a database is formed by a set of  $K$  snapshots  $\mathbf{v}_k = \mathbf{v}(t_k)$ , where  $t_k$  is the time measured at instant  $k$ , that for convenience, are collected in the following *snapshot matrix*

$$\mathbf{V}_1^K = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_{K-1}, \mathbf{v}_K]. \quad (2.1)$$

For some applications, it is more convenient to re-organize the database into tensor form, called as *snapshot tensor*. In that case, the various components that constitute the database are separated and re-organized into different tensor components, and similarly, the different spatial coordinates are also separated too. Generally, the database components are formed by velocity components (especially in fluid dynamics applications), although it is possible to consider any type of variable depending on the application solved. For instance, in combustion databases the components are formed by the different species [3], to identify flutter instability in flight test, the components are formed by the signal given by an array of accelerometers [17], in atmospheric boundary layer flows, in addition to the velocity components, the pressure is included [12], etcetera.

In the snapshot tensor, the matrix snapshots fit a multidimensional array, which depends on more than two indexes. The *fibers* of the tensor are formed by the corresponding matrix columns and rows. Fig. 2 shows an example of a third order tensor, where the tensor fibers are identified.

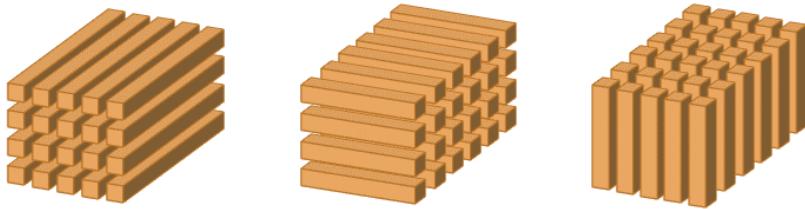


Figure 2: The fibers of a third order tensor.

Generally, the algorithms presented in ModelFLOWS-app consider fourth and fifth order tensors for the analysis of two- and three-dimensional databases, respectively. For instance, let us consider a two-dimensional database (plane) formed by two velocity components (although, as mentioned before, the type and number of components are dependent on the database studied), streamwise and normal velocity components  $v_x$  and  $v_y$ , of the in-plane velocity  $\mathbf{v}$  in

a Cartesian coordinate system with dimension  $J_2 \times J_3$ , as

$$\mathbf{v}(x_{j_2}, y_{j_3}, t_k) \quad \text{for } j_2 = 1, \dots, J_2, \quad j_3 = 1, \dots, J_3, \quad k = 1, \dots, K. \quad (2.2)$$

The snapshots can be re-organized in a fourth-order  $J_1 \times J_2 \times J_3 \times K$ -tensor  $\mathbf{V}$ , whose components  $V_{j_1 j_2 j_3 k}$  are defined as

$$V_{1j_2j_3k} = v_x(x_{j_2}, y_{j_3}, t_k), \quad V_{2j_2j_3k} = v_y(x_{j_2}, y_{j_3}, t_k). \quad (2.3)$$

The indexes  $j_1$ ,  $j_2$ ,  $j_3$  and  $k$  labels the velocity components (in this particular case  $j_1 = 1, 2$ , where  $J_1 = 2$ ), the discrete values of the two spatial coordinates, and the values of time. Note that, although we present a particular case for simplicity, the snapshot tensor  $V_{j_1 j_2 j_3 k}$ , can be used with different number of components  $J_1$ , satisfying the need of the problem under study.

For three-dimensional time-dependent databases, the database is organized in a fifth-order  $J_1 \times J_2 \times J_3 \times J_4 \times K$ -tensor  $\mathbf{V}$ , whose components  $V_{j_1 j_2 j_3 j_4 k}$  are defined as

$$\begin{aligned} V_{1j_2j_3k} &= v_1(x_{j_2}, y_{j_3}, z_{j_4}, t_k), \\ V_{2j_2j_3k} &= v_2(x_{j_2}, y_{j_3}, z_{j_4}, t_k), \\ &\dots \\ V_{j_1 j_2 j_3 k} &= v_2(x_{j_2}, y_{j_3}, z_{j_4}, t_k), \\ &\dots \\ V_{J_1 j_2 j_3 k} &= v_{J_1}(x_{j_2}, y_{j_3}, z_{j_4}, t_k). \end{aligned} \quad (2.4)$$

$J_1$  is the number of selected components forming the database (i.e., three velocity components), the indexes  $j_2$ ,  $j_3$  and  $j_4$  correspond to the discrete values of the three spatial coordinates,  $x$ ,  $y$  and  $z$ , which are streamwise, normal and spanwise components, and  $k$  is the index representing the time instant.

Let us note that the simple connection between the snapshot matrix, presented in eq. (2.1), and the present snapshot tensor formulation. In the snapshot matrix, the tensor indexes  $j_1$ ,  $j_2$ ,  $j_3$  (and  $j_4$  for three-dimensional databases) are folded together into a single index  $j$ . So,  $\mathbf{V}_1^K \in J \times K$ , where  $J = J_1 \times J_2 \times J_3 (\times J_4)$ . Hence, data can be easily transformed from matrix to tensor and tensor to matrix just by re-shaping the database, using the *reshape* function found in Numpy in Python[20], as it is done in ModelFLOWs-app.

## 2.2. Relative root mean square error (RRMSE)

The relative root mean square error (RRMSE) is computed to measure the quality of the ROMs developed in each one of the modules as

$$RRMSE = \sqrt{\frac{\sum_{k=1}^K \|\mathbf{v}_k - \mathbf{v}_k^{approx.}\|^2}{\sum_{k=1}^K \|\mathbf{v}_k\|^2}}, \quad (2.5)$$

where  $\|\cdot\|$  is the usual Euclidean norm and vectors  $\mathbf{v}_k$  and  $\mathbf{v}_k^{approx.}$  correspond to the real and approximated solution.

## 3. Module 1: modal decomposition

### 3.1. Singular value decomposition (SVD) and proper orthogonal decomposition (POD)

Lumley[21] introduced Proper Orthogonal Decomposition (POD) as a mathematical approach for extracting coherent structures from turbulent flows. The primary objective of POD is to decompose data into modes that optimize the mean square of a field variable under analysis. The classical method to calculate POD modal expansion is based on the covariance of a state vector that changes over time, with the size of the state vector being based on the spatial degrees of freedom of the data. This method becomes extremely computationally expensive for large two-dimensional or three-dimensional problems. In such cases, a different technique, singular value decomposition (SVD) is used to obtain the POD modes, introduced by Sirovich [1].

In fluid dynamics, specifically in the study of turbulent flows, SVD is a widely used factorization technique. SVD captures the directions of a matrix which represent the vectors that can either shrink or grow. These directions are determined by the eigenvalues and eigenvectors of a rectangular matrix. In addition to fluid dynamics, SVD has found a wide range of applications, particularly in low-rank matrix approximations. This approach is beneficial because it reduces the size of the data being analyzed, removes noise and filters spatial redundancies [6].

It is remarkable that the literature often uses the terms SVD and POD interchangeably, but SVD is one of the two possible techniques that can applied to obtain a POD decomposition.

SVD (or POD) decomposes a collection of spatio-temporal data  $\mathbf{v}(x, y, z, t)$  into a set of proper orthogonal spatial modes, also known as SVD or POD

modes, represented by  $\Phi_n(x, y, z)$ , which are weighted by the temporal coefficients  $\mathbf{c}_n(t)$ , as

$$\mathbf{v}(x, y, z, t) \simeq \sum_{n=1}^N \mathbf{c}_n(t) \Phi_n(x, y, z). \quad (3.1)$$

SVD algorithm factorizes the snapshot matrix  $\mathbf{V}_1^k$ , eq. (2.1), which is decomposed into the spatial orthogonal modes  $\mathbf{W}$ , the SVD or POD modes, temporal modes  $\mathbf{T}$  and singular values  $\Sigma$  as

$$\mathbf{V}_1^K \simeq \mathbf{W} \Sigma \mathbf{T}^\top. \quad (3.2)$$

where  $(\cdot)^\top$  denotes the matrix transpose. The diagonal of matrix  $\Sigma$  contains the singular values  $\sigma_1, \dots, \sigma_K$ ,  $\mathbf{W}^\top \mathbf{W} = \mathbf{T}^\top \mathbf{T} = N \times N$ -unit matrix, being  $N$  the number of SVD modes retained. This parameter is also called *spatial complexity*, and will be referred in the following sections, when DMD-based methods will be introduced. It is remarkable the difference between  $J$ , the *spatial dimension* of the database, and the spatial complexity  $N$ , where  $N \leq \min(J, K)$ .

SVD modes are ranked in descending order based on their singular values. Typically, the modes with the highest singular values embody the system's general dynamics, representing coherent structures or patterns, meanwhile, the modes with the smallest singular values may be omitted from the approximation, assuming a certain level of error. These modes could be related to noise, especially in the case of experimental databases, with spatial redundancies or, in fluid dynamics, they are generally connected (although not always) to small size flow scales.

The number of  $N$  SVD modes to be retained in the approximation, to construct the expansion eq. (3.1), can be determined using different criteria, as discussed in Ref. [22] (in this context, the POD or SVD approaches are also referred to as *Principal Component Analysis - PCA*). In the present context, the number of retained modes are estimated for a certain tolerance (tunable),  $\varepsilon_{svd}$ , which could be comparable to the level of noise (in the case of experimental results), could be connected to the size of the flow structures (in turbulent flows), et cetera, defined as

$$\sigma_{N+1}/\sigma_1 \leq \varepsilon_{svd}. \quad (3.3)$$

### 3.2. High order singular value decomposition (HOSVD)

Introduced by Tucker in 1966 [23], the HOSVD algorithm has gained popularity in recent years, particularly due to its implementation by de Lath-

auwer et al. [24, 25]. This algorithm has proven to be effective in diverse fields such as aeronautic database generation [26], database compression [27], conceptual aeronautic design [28], and real-time control of automotive engines [29].

HOSVD decomposes databases organized in tensor form, where SVD is applied to each one of the fibers of the tensor. For instance, HOSVD of the fifth order tensor defined in eq. (2.4) is presented as

$$V_{j_1 j_2 j_3 j_4 k} \simeq \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} \sum_{p_3=1}^{P_3} \sum_{p_4=1}^{P_4} \sum_{n=1}^N S_{p_1 p_2 p_3 p_4 n} W_{j_1 p_1}^{(1)} W_{j_2 p_2}^{(2)} W_{j_3 p_3}^{(3)} W_{j_4 p_4}^{(4)} T_{k n}, \quad (3.4)$$

where  $S_{p_1 p_2 p_3 p_4 n}$  is the *core tensor*, another fifth-order tensor, and the columns of the matrices  $\mathbf{W}^{(1)}$ ,  $\mathbf{W}^{(2)}$ ,  $\mathbf{W}^{(3)}$ ,  $\mathbf{W}^{(4)}$  are  $\mathbf{T}$  are known as the *modes* of the decomposition.

The first set of modes (i.e, the columns of the matrices  $\mathbf{W}^{(l)}$  for  $l = 1, 2, 3$  and 4) correspond to the number of components of the database and the spatial variables, so they are known the spatial HOSVD modes, while the columns of the matrix  $\mathbf{T}$  correspond to the time variable, these are the temporal HOSVD modes.

The singular values of the decomposition is now formed by five sets of values,

$$\sigma_{p_1}^{(1)}, \quad \sigma_{p_2}^{(2)}, \quad \sigma_{p_3}^{(3)}, \quad \sigma_{p_4}^{(4)}, \quad \text{and } \sigma_n^t, \quad (3.5)$$

which are also sorted in decreasing order.

Similarly to SVD, without truncation the HOSVD (3.4) is exact. Nevertheless, truncation is advised to filtering noise, spurious artifacts, or reducing the data-dimensionality depending on the need of our application. As in SVD, the number of modes retained for each case generally depends on a tolerance (tunable) as

$$\begin{aligned} \sigma_{P_1+1}/\sigma_1 &\leq \varepsilon_{svd_1}, \\ \sigma_{P_2+1}/\sigma_1 &\leq \varepsilon_{svd_2}, \\ \sigma_{P_3+1}/\sigma_1 &\leq \varepsilon_{svd_3}, \\ \sigma_{P_4+1}/\sigma_1 &\leq \varepsilon_{svd_4}, \\ \sigma_{N+1}/\sigma_1 &\leq \varepsilon_{svd_5}. \end{aligned} \quad (3.6)$$

Generally, the tolerance is set the same for all the cases, so  $\varepsilon_{svd_l} = \varepsilon_{svd}$ , for  $l = 1, 2, 3, 4, 5$ . Although, for highly complex dynamics, these tolerances

should be set different, depending on the database studied. This is one of the main advantages of HOSVD algorithm compared to SVD, where the dimensionality of all the directions and components of the database is reduced similarly. Using HOSVD, it is possible to distinguish between different noise levels or components magnitude.

After truncation, HOSVD (3.4) is written as

$$V_{j_1 j_2 j_3 j_4 k} \simeq \sum_{n=1}^N W_{j_1 j_2 j_3 j_4 n} \hat{V}_{kn}, \quad (3.7)$$

where  $W_{j_1 j_2 j_3 j_4 n}$  and  $V_{kn}$  are the spatial and temporal modes, and  $N$  is the spatial complexity defined above. The spatial modes are defined as

$$W_{j_1 j_2 j_3 J_4 n} = \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} \sum_{p_3=1}^{P_3} S_{p_1 p_2 p_3 p_4 n} W_{j_1 p_1}^{(1)} W_{j_2 p_2}^{(2)} W_{j_3 p_3}^{(3)} W_{j_4 p_4}^{(4)} / \sigma_r^t, \quad \hat{V}_{kn} = \sigma_r^t T_{kn}. \quad (3.8)$$

### 3.3. Gappy SVD and gappy HOSVD for data repairing and resolution enhancement

Repairing databases with corrupted or incomplete information, enlarging the dimension of the original database, and increasing its spatial resolution are some of the most relevant applications of SVD. In addition to the classical applications of SVD, patterns identification and reducing data dimensionality, when used properly, SVD is also a very useful tool for the post-processing and treatment of databases. The algorithm for this application is very simple and is based on the properties of the decomposition, which re-organizes the SVD modes as function of their contribution into the reconstruction of the original database.

Starting from the snapshot matrix eq. (2.1), with dimension  $J \times K$ , the algorithms for data repairing or resolution enhancement are presented below. For simplicity, the SVD-based algorithms introduced are particularized for a single snapshot, two-dimensional, organized in matrix form, while for three-dimensional databases and/or temporal variations, the algorithms uses HOSVD and the snapshot tensor (2.4), or its corresponding version adapted to the number of components composing the tensor.

### 3.3.1. Gappy SVD

Gappy SVD, also known as Gappy POD, uses SVD iteratively to repair and reconstruct incomplete or corrupted databases. The database analysed is particularized for a single snapshot vector  $\mathbf{v}_k$  with dimension  $J \times 1$ , which is re-organized in matrix form, named as  $\widehat{\mathbf{V}}^0$  with dimension  $\widehat{N}_1 \times \widehat{N}_2 = J$ , being  $\widehat{N}_1$  and  $\widehat{N}_2$  the dimensions associated to the streamwise and normal directions. The initial database contains some information that is corrupted, for instance, given by *Nan* (*Not a Number*) information. To repair the database, Gappy SVD algorithm is as follows:

- Step 1.** Initialize the database,  $\widehat{\mathbf{V}}^0$ , giving the points *Nan* an initial value, which can be zero or can be calculated as the mean or as the linear or non-linear interpolation between the surrounded points.
- Step 2.** Apply SVD to the previous matrix  $\widehat{\mathbf{V}}^i = \mathbf{W}^i \Sigma^i (\mathbf{T}^i)^T$ , for  $i = 0$  in the initial iteration, and reduce the previous matrix dimensions by retaining  $P'$  singular values (tunable).

- Step 3.** Reconstruct the new reduced snapshot matrix

$$\widehat{\mathbf{V}}^{i+1} = \mathbf{W}_{P'}^i \Sigma_{P'}^i (\mathbf{T}_{P'}^i)^T, \quad (3.9)$$

with  $\mathbf{W}_{P'}^i \in \mathbb{R}^{\widehat{N}_1 \times P'}$ ,  $\Sigma_{P'}^i \in \mathbb{R}^{P' \times P'}$ , and  $\mathbf{T}_{P'}^i \in \mathbb{R}^{\widehat{N}_2 \times P'}$ .

- Step 4.** Update the gaps with the values of  $\widehat{\mathbf{V}}_k^{i+1}$ .

- Step 5.** Calculate the MSE at the gaps between iterations as

$$MSE_{gaps} = \frac{1}{N_{gaps}} \sqrt{\sum_{n=1}^{N_{gaps}} |\widehat{\mathbf{V}}^i - \widehat{\mathbf{V}}^{i-1}|}, \quad (3.10)$$

where  $N_{gaps}$  is the number of gaps in the database. While  $MSE_{gaps} < 10^{-6}$ , update  $i = i + 1$  and repeat steps 2–4 in the new matrix from Step 4.

The final reconstructed matrix  $\widehat{\mathbf{V}}_k^{i+1}$  (for iteration  $i$ ) is the new repaired matrix. Figure 3 shows an sketch representing the general methodology.

For three-dimensional data, the algorithm uses databases organized in tensor form, which particularized for a single snapshot has dimension  $\widehat{N}_1 \times$

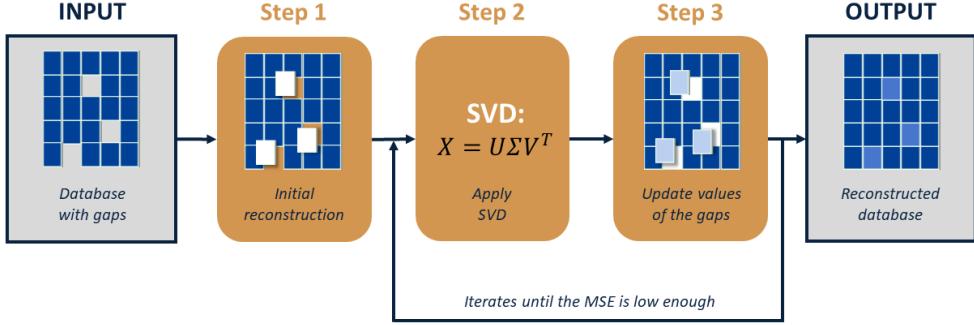


Figure 3: Gappy SVD: sketch summarizing the methodology.

$\widehat{N}_2 \times \widehat{N}_3$ , with  $\widehat{N}_3$  as the spanwise component. The previous algorithm is then the same, but SVD is replaced by HOSVD to properly repair all the components of the database. This algorithm is also valid for larger dimension databases, or it can be connected to other components of the database, different to the spatial dimension. See more details of the algorithm in Refs. [30, 31, 32].

### 3.3.2. Increasing resolution using modal decomposition

SVD can be used to increase the resolution of the database. The method uses SVD iteratively, and it also makes profit of the properties of SVD to re-organize SVD modes as function of their contribution into the original database. The algorithm is particularized for a single snapshot  $\mathbf{v}_k$ , which is organized in matrix form  $\mathbf{V}^{DS,i}$  ( $i$  is the iteration number of the algorithm) with (down-sampled) dimension  $N_1 \times N_2 < J$ , being  $N_1$  and  $N_2$  the dimensions associated to the streamwise and normal directions. The main goal of this algorithm is to get a new database better resolved in space, with dimension  $\widehat{N}_1 \times \widehat{N}_2 = J$ . The algorithm is as follows:

- Step 1.** Apply SVD to the initial under-resolved or downsampled database, and set the number of singular values to  $P'$  (tunable), as

$$\mathbf{V}^{DS,i} \simeq \mathbf{W}_{P'}^{DS,i} \Sigma_{P'}^{DS,i} (\mathbf{T}_{P'}^{DS,i})^T, \quad (3.11)$$

with  $\mathbf{W}_{P'}^{DS,i} \in \mathbb{R}^{N_1 \times P'}$ ,  $\Sigma_{P'}^{DS,i} \in \mathbb{R}^{P' \times P'}$  and  $\mathbf{T}_{P'}^{DS,i} \in \mathbb{R}^{N_2 \times P'}$ , for  $i = 0$  in the first iteration.

- Step 2.** Enlarge the dimension of the matrices from the previous decomposition using a linear (or non-linear) interpolation as:  $\mathbf{W}_{P'}^{DS,i+1} \in \mathbb{R}^{2N_1 \times P'}$  and

$\mathbf{T}_{P'}^{DS,i+1} \in \mathbb{R}^{2N_2 \times P'}$  interpolating between two points, and reconstruct the new enlarged matrix

$$\mathbf{V}^{DS,i+1} \simeq \mathbf{W}_{P'}^{DS,i+1} \Sigma_{P'}^{DS,i} (\mathbf{T}_{P'}^{DS,i+1})^T, \quad (3.12)$$

**Step 3.** Update the iteration number as  $i = i + 1$ .

**Step 4.** Repeat steps 1 – 3  $i = s$  times, until  $2^s \times N_1 = \hat{N}_1$  and  $2^s \times N_2 = \hat{N}_2$ .

The new matrix  $\mathbf{V}^{DS,s+1}$  is the matrix with enhanced resolution.

When working with three-dimensional data, the algorithm employs tensor databases that are organized in a specific way. For a single snapshot, the database dimension is  $\hat{N}_1 \times \hat{N}_2 \times \hat{N}_3$ , where  $\hat{N}_3$  represents the spanwise component. To refine this process further, the previous algorithm is modified by replacing SVD with HOSVD. Also, when dealing with temporal components, HOSVD is used to enhance the resolution of the databases, and the associated temporal information (tunable) can either remain constant withing the algorithm, only enhancing the resolution of the spatial components, or may also be enlarged, interpolating to time instants that were not included in the original dataset.

See more details regarding this algorithm and some applications in Refs. [33, 14]

### 3.4. Higher order dynamic mode decomposition (HODMD)

Higher order dynamic mode decomposition (HODMD) [34] is an extension of dynamic mode decomposition (DMD) [2] introduced for the analysis of complex flows and highly non-linear dynamical systems [35].

HODMD decomposes spatio-temporal data  $\mathbf{v}_k$ , as an expansion of  $M$  DMD modes  $\mathbf{u}_m$ , weighted by their corresponding amplitude  $a_m$  as

$$\mathbf{v}(x, y, z, t_k) \simeq \sum_{m=1}^M a_m \mathbf{u}_m(x, y, z) e^{(\delta_m + i\omega_m)t_k}, \quad (3.13)$$

for  $k = 1, \dots, K$ , where  $\omega_m$  is the oscillation frequency and  $\delta_m$  corresponds to the growth rate, representing the mode temporal growing, decaying or showing when the mode remains neutral in time.

HODMD algorithm is briefly introduced here, where the method is summarized in two main steps. A detailed description of the algorithm can be found in Ref. [34]. We recommend the book by Vega & Le Clainche [35],

where it is possible to find a wide range of examples and applications, as well as the implementation of the algorithms in Matlab [36].

The algorithm is presented as follows.

- **Step 1: Dimension reduction via SVD.** SVD is applied to the snapshot matrix (2.1) to remove noise or spatial redundancies and reduce data dimensionality from the spatial dimension  $J$  to the spatial complexity  $N$  (number of SVD modes retained). At this step, tolerance  $\varepsilon_{svd}$  is fixed to select  $N$ .

Starting from eq. (3.2), it is possible to define the *reduced snapshot matrix* as

$$\widehat{\mathbf{V}}_1^K = \Sigma \mathbf{T}^T, \quad (3.14)$$

with  $\mathbf{V}_1^K = \mathbf{W} \widehat{\mathbf{V}}_1^K$  and  $\widehat{\mathbf{V}}_1^K$  with dimension  $N \times K$ .

- **Step 2: The DMD-d algorithm.** The *high order Koopman assumption* is applied to the reduced snapshot matrix as

$$\widehat{\mathbf{V}}_{d+1}^K \simeq \widehat{\mathbf{R}}_1 \widehat{\mathbf{V}}_1^{K-d} + \widehat{\mathbf{R}}_2 \widehat{\mathbf{V}}_2^{K-d+1} + \dots + \widehat{\mathbf{R}}_d \widehat{\mathbf{V}}_d^{K-1}, \quad (3.15)$$

where  $\widehat{\mathbf{R}}_k = \mathbf{W}^T \mathbf{R}_k \mathbf{W}$  for  $k = 1, \dots, d$  are the Koopman operators, linear, which contain the system dynamics. The snapshot matrix eq. (2.1) is then divided into  $d$  blocks formed by  $K - d$  time-delayed snapshots each. For  $d = 1$ , the method is similar as standard DMD [2].

The previous equation is reorganized in the following way

$$\begin{bmatrix} \widehat{\mathbf{V}}_2^{K-d+1} \\ \vdots \\ \widehat{\mathbf{V}}_d^{K-1} \\ \widehat{\mathbf{V}}_{d+1}^K \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \widehat{\mathbf{V}}_1^{K-d} \\ \widehat{\mathbf{V}}_2^{K-d+1} \\ \vdots \\ \widehat{\mathbf{V}}_d^{K-1} \end{bmatrix}. \quad (3.16)$$

The previous expression can also be represented as

$$\tilde{\mathbf{V}}_2^{K-d+1} = \tilde{\mathbf{R}} \tilde{\mathbf{V}}_1^{K-d}, \quad (3.17)$$

where  $\tilde{\mathbf{V}}_1^{K-d+1}$  and  $\tilde{\mathbf{R}}$  correspond to the *modified snapshot matrix* and *modified Koopman matrix*, respectively. SVD is again applied to this

matrix to remove expected spatial redundancies, where the tolerance  $\varepsilon_{svd}$  is again used to retain  $N' > N$  SVD modes, as

$$\tilde{\sigma}_{N'+1}/\tilde{\sigma}_1 < \varepsilon_{svd}. \quad (3.18)$$

It is remarkable that the key of the good performance of HODMD algorithm lies in matrix  $\tilde{\mathbf{R}}$ , which increases the spatial complexity of the data from  $N$  to  $N'$ .

The next step calculates the DMD modes to form the DMD expansion eq. (3.13). These modes are obtained via solving an eigenvalue problem in the new reduced version of the modified Koopman matrix, whose eigenvalues and eigenvectors are adapted to represent the DMD frequencies and growth rates, and the DMD modes, respectively. Finally, the mode amplitudes are calculated via least-squares fitting, following the optimized DMD method [37]. It is remarkable that depending on the way the amplitudes are calculated, the mode weight can be different. See for instance HODMD with criterion (HODMDc) [5], where the modes also consider the contribution of the growth rates in the amplitude calculations. A detailed description on different ways to calculate the DMD mode amplitudes can be found in the review paper Ref. [38].

Finally, sorting the mode amplitudes in decreasing order, the DMD modes are reordered. The  $M \leq N'$  retained DMD modes are calculated based on a second tolerance  $\varepsilon_a$  (tunable) as

$$a_{M+1}/a_1 < \varepsilon_a. \quad (3.19)$$

$M = \min(K, N')$  is known as the *spectral complexity*, different from the *spectral dimension*  $K$ . For a sufficiently large number of snapshots  $K$  (which is the common case), in cases in which the spatial complexity  $N$  is smaller than the spectral complexity  $M$  (for  $K > N$ ), the high order Koopman assumption completes the lack of spatial information (reduced to  $N$ ) and ensures the good performance of the DMD method. When using standard DMD, it is possible to find some cases in which  $N < M$ , hence the algorithm fails. HODMD in contrast, overcome such limitation, extending the range of application of the standard algorithm. HODMD method has shown potential in the study of various flow types such as transition to turbulent [39, 40] and turbulent flows

[6, 7], identification of flow features from experimental data with noise [39, 41, 16], in the analysis of data obtained from limited spatial locations like field measurements [16, 42], or even in medical imaging [13].

### 3.5. Multi-dimensional iterative higher order dynamic mode decomposition (mdHODMD-it)

The multi-dimensional HODMD (mdHODMD) algorithm uses the snapshot tensor given in equation (2.4). So, instead of SVD, HOSVD is applied as first step to reduce data-dimensionality. This method considers the variables and components of the database separately, so different modes can be selected to better represent each fiber forming the tensor.

The algorithm is summarized in two steps:

- **Step 1:** Application of HOSVD.
  - 1.1 Perform HOSVD on the snapshot tensor  $\mathbf{V}$ , eq. (2.4), without truncation. The ranks of the matrices, whose columns are the fibers, determine the number of modes in the tensor  $\mathbf{V}$ . These ranks are  $P_1 = \min J_1, J_2 J_3 J_4 K$ ,  $P_2 = \min J_2, J_1 J_3 J_4 K$ ,  $P_3 = \min J_3, J_1 J_2 J_4 K$ , and  $N = \min K, J_1 J_2 J_3 J_4$ . The resulting singular values are selected as shown in equation (3.5).
  - 1.2 Choose spatial and temporal tolerances, denoted as  $\varepsilon_{svd}$  and  $\varepsilon_a$  respectively, to determine the number of modes retained in each direction. These numbers are denoted by  $P_1, P_2, P_3, P_4$  and  $N$ . The smallest values of  $P_1, P_2, P_3, P_4$  and  $N$  that satisfy the condition given by equation (3.6) are chosen.
  - 1.3 Perform truncated HOSVD with the numbers of modes determined in the previous step. This gives the core tensor and the modes that define the truncated HOSVD on the right-hand side of equation (3.4).
  - 1.4 Compute the spatial and temporal modes as defined in equation (3.8).
- **Step 2:** Calculation of the multidimensional DMD expansion. Step 2 in the HODMD algorithm described in §3.4 is applied to the reduced snapshot matrix defined by the temporal modes  $\hat{V}_{kr}$  calculated in the previous step item 1.4. The mode amplitudes, frequencies, growth rates and DMD modes are calculated at this step.

When the data are too noisy or too complex, mdHODMD can be applied iteratively, and the algorithm is called as multi-dimensional iterative HODMD (mdHODMD-it). More specifically, as first step the mdHODMD technique is reapplied to the reconstructed snapshots with the same tolerances as in the first application, resulting in cleaner data. Next, the algorithm is applied to the newly reconstructed snapshots, and this process repeats until the number of HOSVD modes is maintained after two consecutive iterations. With each iteration, the algorithm recalculates and orders both HOSVD and DMD modes based on their new corresponding amplitudes, ultimately improving the quality of the DMD reconstructions. The primary benefit of the iterative approach is the removal of irrelevant or inconsistent HOSVD modes based on the tolerance  $\varepsilon_{svd}$ .

#### 4. Module 2: hybrid machine learning tools

Module 2 is formed by a group of deep learning algorithms that are combined with the modal decomposition methods presented in Module 1. Modal decomposition are suitable to identify patterns that contain relevant information about the physics of the dynamical system. As previously mentioned, these techniques have excellent properties to reduce the data-dimensionality, which is very convenient in fluid dynamics, other complex problems and industrial applications. The dimensionality of the original database is reduced from hundred thousands or even millions degrees of freedom to a few modes, POD or DMD, generally varying from dozens to hundred, depending on the problem under study. These modes represent the dynamics of the system, which is represented by modal decomposition with physical interpretability. As second step, the reduced database is combined with different algorithms of deep learning. ModelFLOWS-app in particular offers two possibilities, convolutional neural networks (CNNs) and recurrent neural networks (RNNs), although the algorithm is opened to be combined with other types of architectures. This so-called hybrid ROM, can be used for several applications, including data repairing and reconstruction (Module 2 - application 2) and temporal forecasting (Module 2 - application 3), which is used to predict the non-linear dynamics in complex systems. This application is able to predict saturated (converged or statistically steady) solutions from transient stages in numerical simulations, which generally results in a notorious reduction of computational times to generate new databases. In the examples presented,

the computational cost is reduced from hundred or even thousands of computational hours to a few seconds [43].

Module 2 - application 1 also explores the possibility of using deep learning architectures to identify patterns in the dynamical system under study. For such aim, several architectures of autoencoders have been tested, and compared with results obtained with modal decompositions. This application completes the Deep Learning Module, which is complementary to the modal decomposition module. Module 1 and Module 2 solves similar applications using classical (modal decomposition) and more modern (neural networks) tools, respectively.

#### 4.1. Dimensionality reduction and via autoencoders (AEs)

Autoencoders (AEs) are unsupervised neural networks designed to learn a compressed representation of a database. The encoder is responsible for projecting the data onto a low-dimensional nonlinear manifold,  $\mathbf{v} \mapsto \mathbf{r}$ , while the decoder reconstructs the data from the latent space back to the reference space and reduces the reconstruction error, See Fig. 4

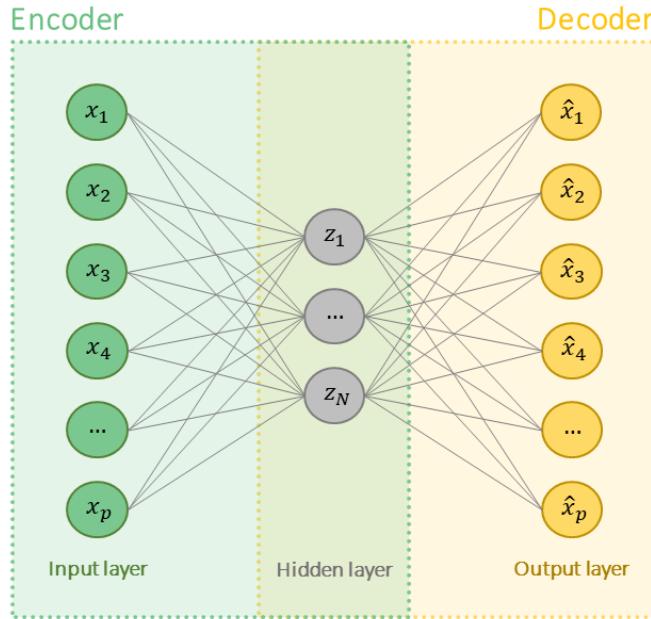


Figure 4: Basic architecture of autoencoders.

By training the autoencoder model, it acquires the ability to identify the crucial features in the data necessary for reconstruction. This is accomplished

by optimizing the model parameters  $\mathbf{w}$  to minimize the reconstruction loss  $\mathcal{L}_{\text{rec}}$ . For  $\mathcal{E}$  and  $\mathcal{D}$  being the encoder and decoder respectively, this can be expressed as,

$$\mathcal{F} = \mathcal{D} \circ \mathcal{E}, \quad (4.1a)$$

$$\hat{\mathbf{v}} = \mathcal{F}(\mathbf{v}; \mathbf{w}), \quad (4.1b)$$

$$\mathcal{L}_{\text{rec}} = \epsilon(\mathbf{v}, \hat{\mathbf{v}}), \quad (4.1c)$$

being  $\epsilon$ ,  $\mathbf{v}$  and  $\hat{\mathbf{v}}$  the loss function, the input data, and the reconstruction of the input data, respectively.

The simplest form of AEs is formed by non-recurrent, feed-forward neural networks that comprise an input layer, one or more hidden layers, and an output layer with the same number of neurons as the input layer. Their primary objective is to reconstruct the input data, minimizing the difference between input and output through a loss function that can accommodate regularization and sparsity terms. ModelFLOWs-app uses this type of AEs to identify patterns in the data analysed. Different types of activation functions, number of layers, or compression rates, among other parameters, can be selected. Snapshot matrix eq. (2.1) is the input of the AE architecture used, and the output provides the reconstructed matrix using a specific (tunable) number of AEs, as well as the patterns collected in each AE. These AE are re-organized in decreasing order as function of their highest influence in the reconstructed field, minimising the RRMSE reconstruction error eq. (2.5).

It is important to note that a shallow AE with linear activation functions is equivalent to principal component analysis (PCA), an extension of POD. Hence, in this case, similarities are found between AE modes and POD (or SVD) modes. Nevertheless, in the context of modal decomposition, the autoencoder architecture offers an appealing framework that can effectively incorporate non-linearity in mappings by utilizing non-linear activation functions [44].

#### 4.2. Hybrid reduced order models

Similarly to HODMD algorithm section 3.4, hybrid ROMs are defined by two main steps. The first step starts from eq. (3.2), for simplicity repeated here as

$$\mathbf{V}_1^K \simeq \mathbf{W} \Sigma \mathbf{T}^\top, \quad (4.2)$$

to define the *reduced snapshot matrix* (with dimension  $N \times K$ ) as

$$\widehat{\mathbf{V}}_1^K = \Sigma \mathbf{T}^T, \quad (4.3)$$

with  $\mathbf{V}_1^K = \mathbf{W}\hat{\mathbf{V}}_1^K$ . The second steps applies deep learning architectures into this reduced matrix. Depending on the type of architecture, it is possible to predict temporal evolution of a signal or to reconstruct two- or even three-dimensional databases from sensors, as presented in sections 4.2.1 and 4.2.2, respectively. Figure 5 shows an sketch summarizing the hybrid ROM methodology for temporal forecasting, where  $p$  snapshots are predicted. So, initial snapshot matrix  $\mathbf{V}_1^K$  eq. (2.1) is then transformed to the new snapshot matrix containing  $p$  additional snapshots  $\mathbf{V}_1^{K+p}$ .

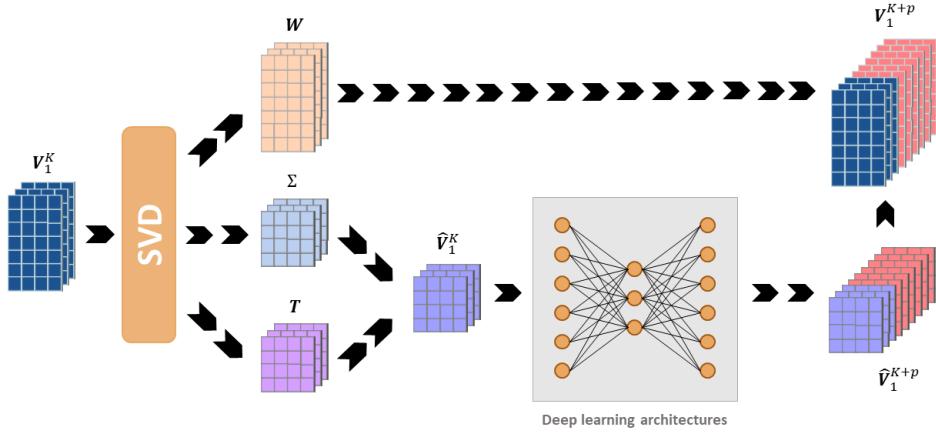


Figure 5: Steps involved in the architecture of the hybrid reduced order models.

It is remarkable that using this method, one-dimensional architectures are required, since they are directly applied to the reduced snapshot matrix. In contrast, two- or even three-dimensional architectures should be applied to the original snapshot matrix, which also contains relevant spatial information of the database. Hence, the reduction in computational time and memory of this hybrid methodology is one of the key points of the algorithms presented.

The performance of the architectures presented below is optimal for standard fluid dynamics problems. However, when the complexity of the database increases, it is necessary to take into account the magnitude and variance of the variables analysed. For instance, in reactive flows, where a large number of variables are involved representing the multiple species of the flow (more than 80), it is necessary to combine the previous methodology with other pre-processing techniques, such as centering and scaling. These two methods are applied to each one of the multiple variables of the database before forming the snapshot matrix  $\mathbf{V}_1^K$  eq. (2.1). Centering removes the temporal

mean in each variable, in this way the analysis only considers the fluctuation field. Scaling normalizes all the variables so they can be compared on the same bases. There are several possibilities to scale the data. The following equation represents centering and scaling of the data,

$$\tilde{\mathbf{v}}_j(t_k) = \frac{\mathbf{v}_j(t_k) - \bar{v}_j}{c_j}, \quad (4.4)$$

where  $\mathbf{v}_j$  is the  $j$ -th variable,  $\bar{v}_j$  is the mean averaged in time,  $c_j$  is the scaling factor used and  $\tilde{\mathbf{v}}_j$  is the scaled variable. Three scaling techniques have been implemented in ModelFLOWS-app:

1. *Range* scaling: The difference between the maximum and minimum value of each variable is used as the scaling factor.
2. *Auto* scaling: the standard deviation of the  $j$ -th variable ( $\sigma_j$ ) is used as the scaling factor.
3. *Pareto* scaling: the square root of the standard deviation of the  $j$ -th variable ( $\sqrt{\sigma_j}$ ) is used as the scaling factor.

Several studies in the literature [22, 45, 46] have examined the impact of centering and scaling on modal decomposition techniques like PCA and HODMD in the context of combustion applications.

#### 4.2.1. Time-series forecasting models using convolutional and recurrent neural networks (CNN and RNN)

The proposed architectures have been chosen to best approximate temporal evolution of databases with high-dimensional nature of time series driven by strongly sequential dynamics. The architectures are fit to predict the longest data sequence as possible, intending to minimize the quantity of training data. In this way, the predictive hybrid ROM can be used to predict new databases in numerical simulations, generally associated to large computational costs. Also, this strategy avoids overfitting.

Fig. 6 shows an sketch showing this predictive deep learning models. Each column of matrix  $\widehat{\mathbf{V}}_{k+1}$  represents the temporal dynamics of each snapshot. The model predicts the snapshot  $k + 1$ , given at time  $t_{k+1}$  in the reduced matrix,  $\widehat{\mathbf{V}}_{k+1}$ , by utilizing data from the  $q$  previous snapshots, defined as  $\widehat{\mathbf{V}}_k$ ,  $\widehat{\mathbf{V}}_{k-1}$ ,  $\dots$ ,  $\widehat{\mathbf{V}}_{k-q+1}$ .

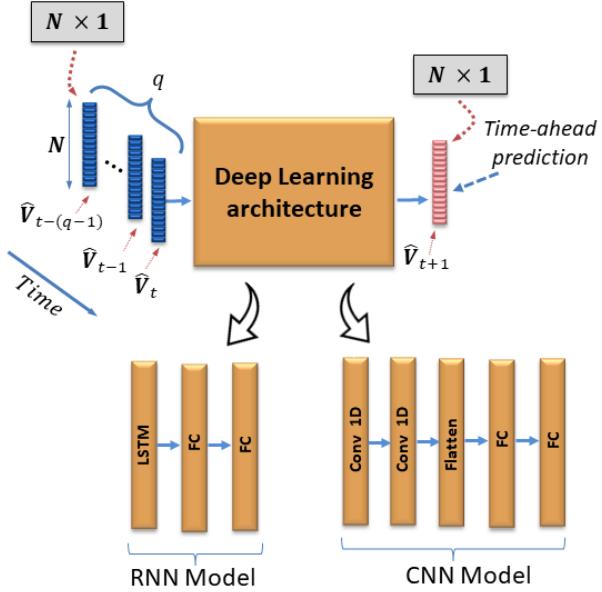


Figure 6: Hybrid predictive ROM. Recurrent and convolutional neural networks to predict a time-ahead snapshot,  $\hat{V}_{k+1}$ , based on the previous  $q$  snapshots.  $N$  is the number of POD modes.

The model using CNN architectures is composed by one-dimensional convolutional (Conv 1D) layers followed by FC layers. Conv 1D applies a one dimensional kernel, with no padding and a stride of 1 [47]. Flatten function is included in thee intermediate layers that bridge the gap between convolutional and FC layers, to transform the matrix structure employed by Conv 1D into a vector structure compatible with the FC layers.

The RNN model is formed by short-term memory (LSTM) layers and FC layers [48]. Different dimensions of the output space (number of units) for the LSTM layers have been considered to form this model. Although, in ModelFLOWS-app, different architecture details can be considered for this application as function of the problem analysed. Details about the number and type of layers, activation function, layer dimensions, ..., are presented in some examples in section 5.6. Hyper-parameter approach to automatically set these parameters to get the best predictions have also been implemented in ModelFLOWS-app. This function also set the optimization [49] and learning rate parameters, batch size and number of epoch, used with early stopping method.

Before entering the deep learning models, the reduced snapshot matrix  $\hat{\mathbf{V}}_1^K$  can be scaled to improve the performance of the neural network. Some of them have been already mentioned in the previous section. Apart from them, a new scaling method (*Max per Mode (MpM)*) has been implemented, in which each column of  $\hat{\mathbf{V}}_1^K$  ( $\mathbf{v}_j$ ) is scaled with the sum of the maximum values of all columns, as

$$\hat{\mathbf{v}}_j = \frac{\mathbf{v}_j}{\sum_{j=1}^N \max|\mathbf{v}_j|}. \quad (4.5)$$

This scaling method has been proved suitable for the prediction of non-periodic temporal modes [50, 3].

For the training and validation, the columns  $K$  of matrix  $\hat{\mathbf{V}}$  are separated in three blocks with dimension  $N \times K_{\text{training}}$ ,  $N \times K_{\text{validation}}$  and  $N \times K_{\text{test}}$ , for the training, validation and test set, respectively, where  $K = K_{\text{training}} + K_{\text{validation}} + K_{\text{test}}$ . An sketch is presented in Fig. 7.

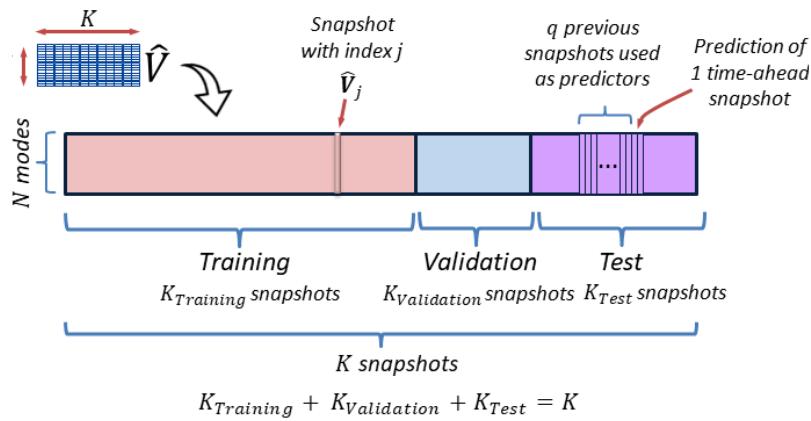


Figure 7: Sketch with the training, validation and test set distribution for the deep learning module.

Similar to HODMD algorithm (section 3.4), a rolling-window method is used. The method generates one output, the model prediction, for  $q$  inputs (model information input) organized in data batches. As seen in the sketch of Fig. 8, 1 is the offset considered between the successive rolling windows.

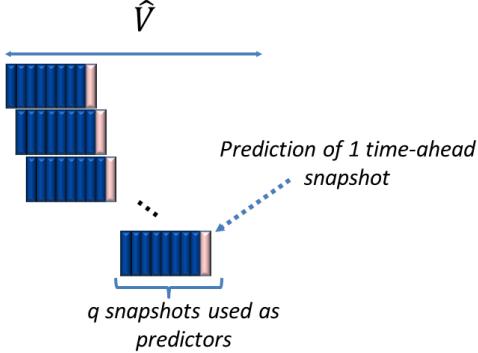


Figure 8: Rolling window method calculating 1 output from  $q$  inputs.

Mean Squared Error Loss ( $MSE_{Loss}$ ) comparing the real and predicted databases ( $\hat{\mathbf{V}}_t^{real}$ ,  $\hat{\mathbf{V}}_t^{predicted}$ ) is minimized using batch stochastic gradient descent algorithm. The global loss ( $MSE_{Loss}$ ) is based on averaging in time the local loss, which is calculated for each prediction as

$$MSE_{Loss}(t) = \frac{1}{N} \|\hat{\mathbf{V}}_t^{predicted} - \hat{\mathbf{V}}_t^{real}\|^2, \quad (4.6)$$

being  $N$  the number of singular values. The local error is calculated at the end of each epoch, and early stopping method is used in the validation set to get the best network parameters. See more details of this architecture in Ref. [43].

Finally, the RRMSE eq. (2.5) is also computed comparing the predictions with the real solution in the neural networks.

#### 4.2.2. Increasing resolution using hybrid deep learning models

This application reconstructs two- or three-dimensional databases using the information contained in a few sensors. As previously mentioned, the first step of the algorithm applies SVD to the snapshot matrix as defined in eq. (4.2). In this case, the snapshot matrix  $\mathbf{V}_1^K$  eq. (2.1), is now called as  $\mathbf{V}^{D_S}$ , as it represents a database under-resolved, only formed by a few points where temporal information about velocity vector, pressure field or other quantities are collected in each one of such points. For each temporal snapshot, the dimensions of the *down-sampled* snapshot matrix  $\mathbf{V}^{D_S}$  are  $N_1 \times N_2$ , being  $N_1 \times N_2 < J$ . For two-dimensional databases,  $N_1$  and  $N_2$  correspond to the streamwise and normal components, respectively, while

for three-dimensional databases,  $N_2$  collects the information of the normal and spanwise components (although the three spatial components could be re-organized differently into  $N_1$  and  $N_2$  based on our needs of filling gaps into the databases). Eq. (4.2) particularized for a single snapshot, re-organizes the information into the down-sample matrix and can be re-written as

$$\mathbf{V}^{DS} \simeq \mathbf{W}^{DS} \Sigma^{DS} (\mathbf{T}^{DS})^\top. \quad (4.7)$$

The number of singular values retained is  $P'$ , hence  $\Sigma^{DS} \in \mathbb{R}^{[P', P']}$ . Matrices  $\mathbf{W}^{DS} \in \mathbb{R}^{[N_1, P']}$  and  $\mathbf{T}^{DS} \in \mathbb{R}^{[N_2, P']}$ , are introduced into a deep learning architecture to enlarge the dimension of  $N_1$  and  $N_2$  as in the original database, to  $\widehat{N}_1$  and  $\widehat{N}_2$ , with  $\widehat{N}_1 \times \widehat{N}_2 = J$ , being the dimension of these new enlarged matrices as  $\mathbf{W} \in \mathbb{R}^{\widehat{N}_1, P'}$  and  $\mathbf{T} \in \mathbb{R}^{\widehat{N}_2, P'}$ . Once these new matrices have been modelled, they are again combined to re-construct the database with spatial dimension  $J$  for each snapshot  $K$ , as

$$\mathbf{V}_k \simeq \mathbf{W} \Sigma^{DS} (\mathbf{T})^\top, \quad (4.8)$$

where  $\mathbf{V}_k \in \mathbb{R}^{[\widehat{N}_1, \widehat{N}_2]}$ . The deep learning architectures gives this reconstructed matrix as an output. This architecture is formed by two groups of neural networks, defined as decoders (the decoding part of an autoencoder) working in parallel that meets in the output layer, where the reconstructed solution is calculated as in eq. (4.8). Hence, the reconstruction error (RRMSE, see eq. (2.5)) used to improve the weights is calculated by comparing the reconstructed database with the original solution. Fig. 9 shows an sketch of the present architecture, formed by 5 layers, although this is a tunable parameter of ModelFLOWS-app.

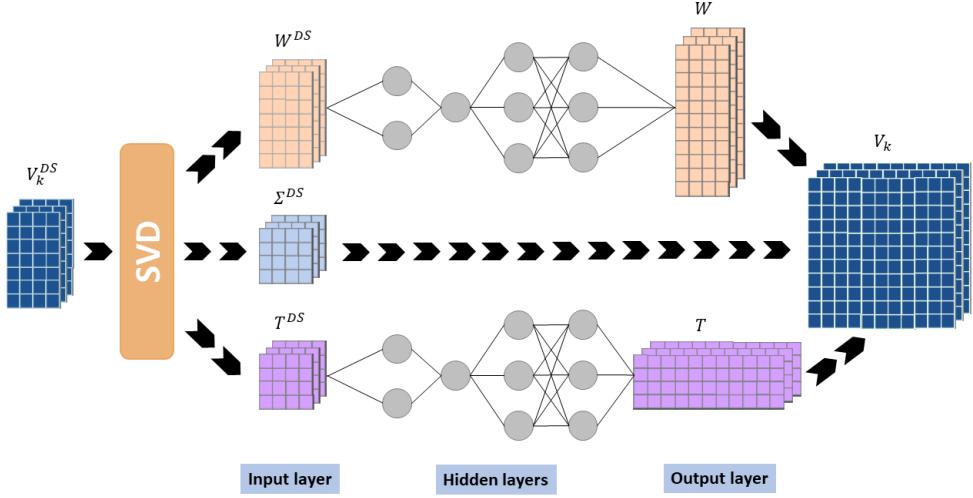


Figure 9: Reconstruction of databases combining SVD and deep learning. Sketch of the methodology.

To organize the database, the strategy followed divides the total number of available snapshots into training, validation and test, as in Fig. 7. It is also notorious that this architecture is able to predict in time the reconstruction of databases.

See more details about this architecture and the model previously described in Ref. [12].

## 5. Results

### 5.1. Module 1 - application 1: patterns identification

HODMD algorithm (Sec. 3.4) has been applied to identify temporal patterns in complex flows. The good performance of HODMD and mdHODMD-it algorithms depends on the selection of the parameter  $d$ , which defines the number and sizes of the window selecting sub-areas of the analyzed matrix, and the tolerances  $\varepsilon_{svd}$  and  $\varepsilon_a$ . As explained in Ref. [34], as a reference  $d$  should be set in the interval  $K/10 < d < K/2$ . There is not only a single optimal value for  $d$ . Generally, robust results are obtained for several values. Also, in saturated flows or converged signals, robust results should be obtained using similar values of  $d$  applying the algorithm in different time intervals. The optimal values of  $d$  also varies with the tolerances  $\varepsilon_{svd}$  and

$\varepsilon_a$ , the time interval  $\Delta t$  and the number of snapshots  $K$ . Also  $d$  scales with the total number of snapshots, so if the  $K$  is doubled,  $d$  should also be multiplied by 2. Finally, the tolerances  $\varepsilon_{svd}$  and  $\varepsilon_a$  should be comparable to the expected uncertainty of the measurements. If the level of noise of the database is known,  $\varepsilon_{svd}$  should be set similar or slightly smaller (when the iterative method is applied) to such level. More details on the calibration of the algorithm can be found in Refs. [39, 35, 6].

This first example reviews an application of patterns identification using HODMD that has been published in Ref. [9]. The problem presented studies a turbulent flow in a simplified urban environment consisting of two buildings, modelled by wall-mounted obstacles, separated by a certain distance. The main goal of this study was to identify the main patterns connected to high concentration of pollution levels. More specifically, it is known that the arch vortex (see Fig. 10 a), a pattern characteristic of this type of flows which forms downstream buildings, plays a crucial role in the dispersion of pollutants in urban areas. HODMD was applied to analyze the flow to identify the main patterns and frequencies leading flow dynamics connected to the presence of the arch vortex. The findings presented in this study are of utmost significance for urban sustainability, as they shed light on the critical factors that contribute to the concentration of pollutants in urban environments. So, we encourage to read the original article Lazpita *et al.* [9] and other related articles for more information [51, 52, 53].

The database for this problem was obtained through numerical simulation applied to a simplified urban environment consisting of two wall-mounted obstacles, where the distance between buildings was varied to obtain different regimes. According to Oke [54], these regimes are called skimming flow, wake interference, and isolated roughness, ordered from shortest to longest distance between buildings. In the case of skimming flow, the spatial dimension of the computational domain is defined for the streamwise, normal and spanwise direction in the interval  $x \in [-1, 5]$ ,  $y \in [0, 2]$  and  $z \in [-1.5, 1.5]$ , respectively. The temporal interval between snapshots is  $\Delta t = 0.35$ . The database used to obtain the results is composed of a spatial grid formed by  $100 \times 125 \times 50$  points (streamwise, normal and spanwise directions, respectively), and 224 snapshots. Therefore, the spatial dimension of the problem is  $J = 1,875,000$ , and the temporal dimension is  $K = 224$ .

HODMD was applied to analyse this database using the parameters listed in Table 1, selected based on the criteria discussed earlier.

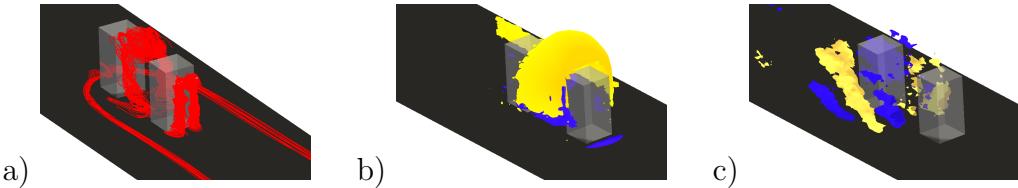


Figure 10: Flow patterns in skimming flow regime in a simplified urban environment. (a) Velocity streamlines representing the arch vortex, (b) and (c) real part of the streamwise velocity of DMD modes: Iso-surfaces normalized with the  $L_\infty$ -norm for (b) the generator mode ( $\omega_m = 0.11$ ), and (c) the breaker mode ( $\omega_m = 1.1$ ). The values used are given by  $c_{max} \max(U)$  (yellow) and  $c_{min} \min(U)$  (blue)

Parameter	Symbol	Value
Number of windows	$d$	50
SVD tolerance	$\varepsilon_{svd}$	$10^{-3}$
DMD tolerance	$\varepsilon_a$	$10^{-3}$

Table 1: Values of the parameters to choose on the HODMD algorithm for the database presented in this section.

From this analysis, two types of modes were identified: the so-called *generator modes* associated with low frequencies in the computed DMD mode spectrum, which were connected to the presence of the main vortical structures found in the flow (i.e., arch vortex [55, 56]), and the so-called *breaker modes* that were high frequency modes and were connected to the wake formed behind the buildings, connected to flow dispersion. These two types of modes are presented in Fig. 10 b and c, showing the streamwise velocity component. The remaining modes forming the spectrum are combination (formed by non-linear interaction) of these two types of modes.

### 5.2. Module 1 - application 2: data reconstruction

SVD algorithm has been applied to repair and enhance the resolution of images complex flows (Sec. 3.3). These two applications are detailed below.

#### 5.2.1. Application 2.1: data repairing

Gappy SVD (and HOSVD for databases in tensor form), as introduced in Section 3.3.1, depends on two main parameters: the type of initial reconstruction of the database to fill the gaps generally identified with *NaN* values (Step 1 of the algorithm), and the number of  $P'$  modes that are retained after

applying SVD to the reconstructed database (Step 2). On the one hand, in the initial reconstruction, the gaps can be replaced by (i) 0 values, (ii) the mean value between consecutive points or (iii) using a linear (or non-linear) interpolation exploiting the information of the surrounding points. On the other hand, regarding the number of SVD modes  $P'$ , which are used in the reconstruction of the database (Step 3), using a large number could increase the reconstruction error, since some of these modes could contain information related to noise or spatial redundancies. On the contrary, using a small number of them could imply losing relevant information related to the dynamics of the system. Hence, calibration is crucial to ensure the proper performance of the method.

Gappy SVD has been applied to repair a numerical database modelling the two-dimensional wake past a circular cylinder at Reynolds number (computed with the cylinder diameter) 100. Details about the numerical simulations and the generation of this database can be found in Ref. [35]. The spatial dimension of the computational domain is defined for the streamwise direction in the interval  $x \in [-1, 8]d$  and for the normal direction in the interval  $y \in [-2, 2]d$ , being  $d$  the diameter of the cylinder. The dataset analysed is composed by the two velocity components, 449 points in the streamwise direction and 199 in the normal direction and 150 snapshots, equi-distant with time interval  $\Delta t = 0.2$ . Therefore, the data can be organized into a four-order tensor with dimension  $2 \times 449 \times 199 \times 100$ . From this database, the  $\sim 63\%$  of the values are selected randomly and are removed to obtain gaps.

Gappy SVD is applied using 0 as initial reconstruction value (replacing Nan), and the number of retained SVD modes selected is 10. Figure 11 shows the initial database, the initial reconstruction carried out by the algorithm (iteration  $i=1$ ) and compares the real solution with the final reconstruction. The RRMSE eq. (2.5) of the reconstruction is smaller than 2%.

For reconstruction of datasets considering the temporal component, similar results are obtained, using the algorithm Gappy HOSVD instead. See details in Ref. [18].

### 5.2.2. Application 2.2: superresolution

This second type of application of SVD algorithm is tested to enhance the resolution of a database. The only parameter necessary to set is the final resolution that is desired to obtain in the super-resolved database (see details in Section 3.3.2). As described in the methodology, the resolution

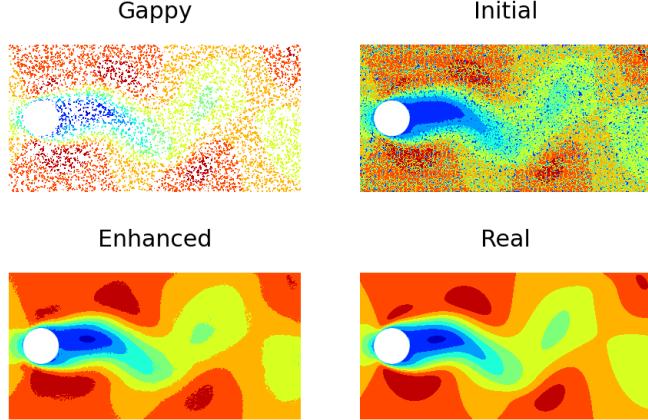


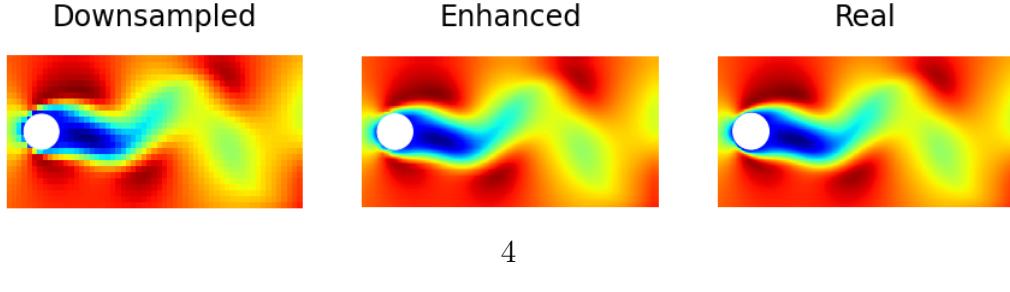
Figure 11: Gappy SVD algorithm to reconstruct the two-dimensional wake behind a circular cylinder. NaN is found in the 63% of the points forming the initial database. From left to right and top to bottom: the corrupted snapshot, initial reconstruction, final reconstruction and real database.

of the database is doubled in each iteration the algorithm is applied, so the final resolution will be a number that is calculated as the power of 2 times the initial resolution.

As in the previous section, the method is tested to enhance the resolution of a database modelling the two-dimensional wake behind a circular cylinder at Reynolds number 100. The dimension of the database, organized in tensor form, is  $2 \times 449 \times 199 \times 150$ , corresponding to the velocity components, grid points along the streamwise and normal directions, respectively, and the number of snapshots. For this application, the dimension of the database is reduced to  $2 \times 63 \times 63 \times 100$ , so the spatial dimensions have been downsampled by a factor of  $2^3 = 8$  (only 1 every 8 points are retained in the reduced database). In ModelFLOWS-app then it would be necessary to include the downsampled database and the parameter 3, to enhance the resolution to the original dimension.

Figure 12 shows the downsampled database and compares the resolution enhanced database with the original one, where the RRMSE eq. (2.5) computed is smaller than 3%.

For reconstruction of datasets considering the temporal component, similar results are obtained, using the algorithm HOSVD instead. See details in Ref. [18].



4

Figure 12: Application of the superresolution algorithm to a database extracted from the flow past a cylinder. From left to right: the downsampled snapshot snapshot, the solution gave by the algorithm and the real snapshot extracted from the simulation.

### 5.3. Module 1 - application 3: predictive models

Let us examine a specific data-oriented ROM developed using HODMD (Sec. 3.4) for temporal prediction. This example is presented in detail in Refs. [57, 35]. The data-driven predictive ROM is developed to expedite the computational efficiency of Nek5000 [58], an open source spectral element code employed for solving the incompressible continuity and Navier-Stokes equations [59], thereby enabling the computation of the system’s final attractor for temporal forecasting. Our focus lies on the three-dimensional wake of a circular cylinder in the context of incompressible fluid dynamics, a well-established benchmark problem [60, 61]. For this problem, we define the Reynolds number (with the cylinder diameter) as 210. The spanwise length of the computational geometry studied corresponds to a spanwise wavenumber of 4. Descriptions of the computational domain, the mesh, and the analysis of the accuracy of the numerical simulations can be found in [57]. To construct the data-driven ROM, we first perform numerical simulations and then we use a collection of 500 snapshots, equi-distant with time interval  $\Delta t = 1$ , obtained from the saturated regime of such simulations, specifically within the time interval  $575 \leq t \leq 825$ . These snapshots serve as the basis for developing the data-driven model.

Parameter	Symbol	Value
Number of windows	$d$	250
SVD tolerance	$\varepsilon_{svd}$	$10^{-4}$
DMD tolerance	$\varepsilon_a$	$3 \cdot 10^{-3}$

Table 2: Values of the parameters to choose on the predictive HODMD algorithm for the database presented in this section.

HODMD is then applied to analyse the previous dataset, where the calibration of the method involved a careful selection of values, presented in Tab. 2, ensuring consistency and robustness in the obtained results [39]. Within our analysis, we have distinguished two distinct types of DMD modes: transient modes characterized by  $\delta < 0$  and permanent modes with  $\delta \simeq 0$  (see the DMD expansion eq. (3.13)). Visual representations of the relationships between damping rates, mode amplitudes, and retained frequencies can be found in Fig. 13.

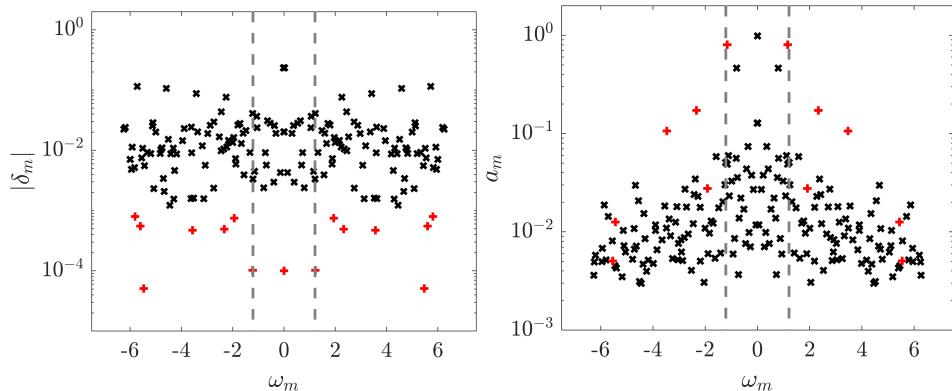


Figure 13: Frequencies vs. damping rates (left) and mode amplitudes (right) of the DMD modes calculated in the transient of a numerical simulation modelling the three-dimensional wake behind a circular cylinder. Red: modes with smaller damping rates, retained to form the ROM. Black: remaining modes. Fundamental frequencies are indicated by vertical lines.

To construct the ROM, permanent modes are selected based on a condition for their growth rate:  $|\delta_m| < \varepsilon = 10^{-3}$ . The HODMD method is then employed to extrapolate the results, excluding transient modes, to the attractor for  $t \geq 1900$ . For such application, ModelFLOWS-app follows the next steps: (i) the original dataset is reconstructed utilizing the DMD expansion eq. (3.13) only using the selected DMD modes, (ii) the growth rate of the selected DMD modes is set to 0 (although it is also possible not to change the growth rate of the modes, as function of the needs of the ROM), (iii) the temporal term of the DMD expansion is set as  $t \geq 1900$ . The ROM predicts then the temporal evolution of the flow for temporal instants  $\geq 1900$ .

The predicted solution exhibits a RRMSE error eq. (2.5) of  $\sim 6\%$  when compared to the real solution. Figure 14 shows a representative snapshot from the attractor, comparing the outcomes of the data-driven ROM with the original data. Notably, the ROM approximation performs worse in the

spanwise velocity component compared to the other two components, mainly due to its significantly smaller magnitude. However, it can be inferred that the spatio-temporal symmetry in the spanwise velocity component is essentially preserved. The steep-up factor of this ROM compared to the numerical simulations is larger than 100.

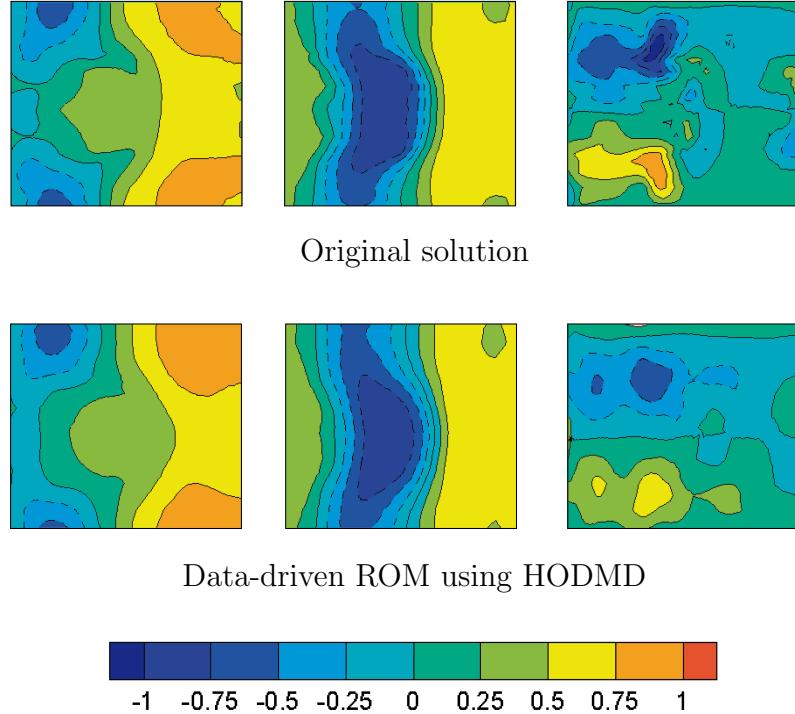


Figure 14: Predictive ROM using HODMD tested in the three-dimensional way of a circular cylinder. Streamwise (left), normal (center), and spanwise (right) velocity components in the mid  $x - z$  plane for a representative snapshot of the attractor at  $t = 2900$ . The data has been normalized using the maximum velocity value from the original data.

#### 5.4. Module 2 - application 1: patterns identification

Autoencoders algorithm (Sec. 4.1) has been applied to identify temporal patterns in complex flows. More specifically, we applied this technique on the flow generated by two planar synthetic jets. These devices are characterized by the periodic movement of a membrane or piston inside a cavity. Each periodic cycle consists of an injection and suction phases in which the fluid is ejected and reintroduced into the cavity through a jet nozzle, respectively.

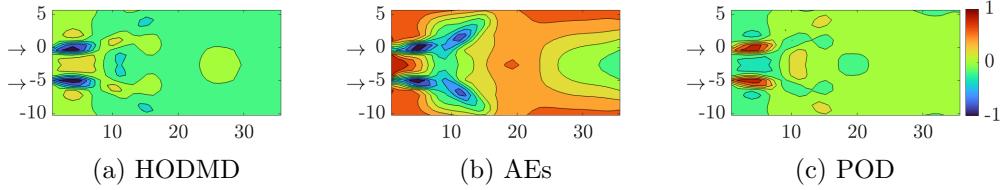


Figure 15: a) AEs, b) HODMD and c) POD first contributing mode on two synthetic jets. Arrows represent the jets' input. Legend from c) applies to all pictures.

The database analysed model two synthetic jets working synchronously at Strouhal number (defined with the jet diameter) 0.03 and Reynolds number (defined with the jet diameter) 100 (from Ref. [62]). The spatial dimension of the computational domain is defined for the streamwise and normal direction in the interval  $x \in [0.9, 35.6]D$  and  $y \in [-10, 5.7]D$ , respectively, where  $D$  refers to the diameter of the nozzle and the axis center is the center of upper jet exit. The temporal interval between snapshots is  $\Delta t = 5.34\text{e-}02 U/D$ , where  $U$  is the characteristic velocity of the flow.

The main patters of the flow are analysed using POD and HODMD methods (both from ModelFLOWS-app module 1), and this section shows a new application where also AEs are used to extract the main patterns connected to the flow dynamics. More details about the results presented in this section can be found in Ref. [63].

Figure 15 compares the dominant mode identified by HODMD, AEs and POD methods. In previous research [62], HODMD relates this mode with the oscillation frequency driving the synthetic jet, Strouhal number 0.03. In the results obtained, these three modes present some differences, as expected since we use three different methodologies. Nevertheless, the high intensity areas found in the three modes are similar: the three modes recognize two high intensity regions in form of two ovals at the jets exit. A region with still high (but less than before) intensity extends further downstream the two jet exits. This region is more strengthned in AEs. The difference found in the shape of the modes could be connected to the non-orthogonality behind AEs calculations. More results and a detailed explanation can be found in Ref. [63]. Also Ref. [44] shows a detailed comparison of differerent type of AEs and SVD modes, as well as their properties for reduced order modelling in turbulent urban flows.

The calibration of AEs depends on the characteristics of the database

analyzed and its spatial and temporal dimensions,  $J$  and  $K$ , respectively. The parameters to choose in ModelFLOWS-app are the training percentage,  $\%_{train}$ , that splits the data into the training and test size ( $K_{training} = \%_{train}K$ ,  $K_{test} = (1 - \%_{train})K$ , respectively) and it is recommended to be smaller or equal to 80%; the batch size,  $N_{batch} \in [1, K_{training}]$  is the size of the packages used for the training, and it is recommended to be a power of 2, being usually 32 and 64 the best values[64]; the number of AE modes retained,  $M < K_{training}$ , also called as encoding dimension; and the maximum number of epochs,  $N_{epochs}$ , is the maximum number of passes the algorithm takes around the training data, and possible examples of its value are 100, 200, 500, . . . . We propose high numbers for this parameter because we have implemented early stopping in the code. Therefore, if  $N_{epochs}$  is high, the early stopping will stop the training when the convergence is achieved, and it is too low, the training could stop before without assuring convergence. Finally, it is notorious that the number of AE modes retained is the parameter that will most affect the results. If the main application of AEs is to observe the main dynamics of the flow, we advise retaining a small number (i.e., 5, 10, 20), but to obtain small reconstruction errors, the number of modes retained should be larger. Based on these recommendations, the parameters chosen in the database analysed in this section, with  $J= 1980$  and  $K=4369$ , are stored in Tab. 3.

Parameter	Symbol	Value
Training percentage	$\%_{train}$	80 %
Batch size	$N_{batch}$	32
# modes (autoencoder)	$M$	10
Epoch number	$N_{epochs}$	200

Table 3: Calibration in AEs algorithm for patters indentification in two plana synthetic jets. Database with dimension  $J= 1980$  and  $K=4369$ .

### 5.5. Module 2 - application 2: data reconstruction

Hybrid ROMs combining SVD with deep learning architectures (Sec. 4.2.2) has been applied to reconstruct downsampled databases in complex flows. This application is explained in detail in Ref. [12], where the authors propose a hybrid model, combination of SVD with neural networks. The neural network considered consists on combining two autoencoders that work in parallel and are joined into the last layer. The architecture has been proved

to be robust and generalizable, which allows reconstructing databases with a low reconstruction error, from databases containing information from a few sensors (as it is the case in experimental measurements).

As in the example presented in module 1 (Sec. 5.2.1), the case analysed is the saturated regime modelling the two-dimensional wake behind a circular cylinder at Reynolds number (defined with the cylinder diameter) 100. The database analysed has been obtained numerically, where details about the performance of the numerical simulation can be found in Ref. [35]. The spatial dimension of the computational domain is defined for the streamwise direction in the interval  $x \in [-1, 8]d$  and for the normal direction in the interval  $y \in [-2, 2]d$ , being  $d$  the diameter of the cylinder. The database analysed is formed by 3 variables, the two velocity components (streamwise and normal) and the spanwise vorticity component, 449 spatial points along the streamwise direction and 199 points along the normal direction, and 150 snapshots equispaced in time with an interval of 0.2. From this database, a downsampled tensor is created, taking one every 30 points in both spatial directions, resulting in a tensor with dimensions  $3 \times 15 \times 7 \times 150$ .

Figure 16 shows the downsampled database and compares the reconstruction of a representative snapshot of the database, carried out using the present methodology, with the original solution. The RRMSE eq. (2.5) in the reconstruction is smaller than 8% in the whole database analysed. The algorithm has also been successfully tested with initial databases composed by a very small number of grid points (i.e. 8 in each spatial direction). More details can be found in Ref. [12].

The calibration of the method depends on several parameters, which are listed in Tab. 4. If selected, the database can be pre-processed with the parameter  $1^{st} Scaling$ , as well as the matrices that will enter the deep learning model, in that case the parameter is called as  $2^{nd} Scaling$ . The parameter  $\%_{train}$  refers to the data used in the training process, which is recommended to be less than the 80% of the samples. The batch size  $N_{batch}$  can be also selected, as in the previous section, as well as the number of epoch ( $N_{epoch}$ ).

For the deep learning model, ModelFLOWS-app offers the possibility to selecting and using the optimal model hyperparameters of the deep learning architecture: we use *RandomSearch Keras* tuner. Additionally, these parameters can be directly selected by the user. These are: (i) the number of neurons  $N_{neurons}$ , the activation function  $AF$ , the loss function  $l_f$  and the learning rate  $l_r$ .

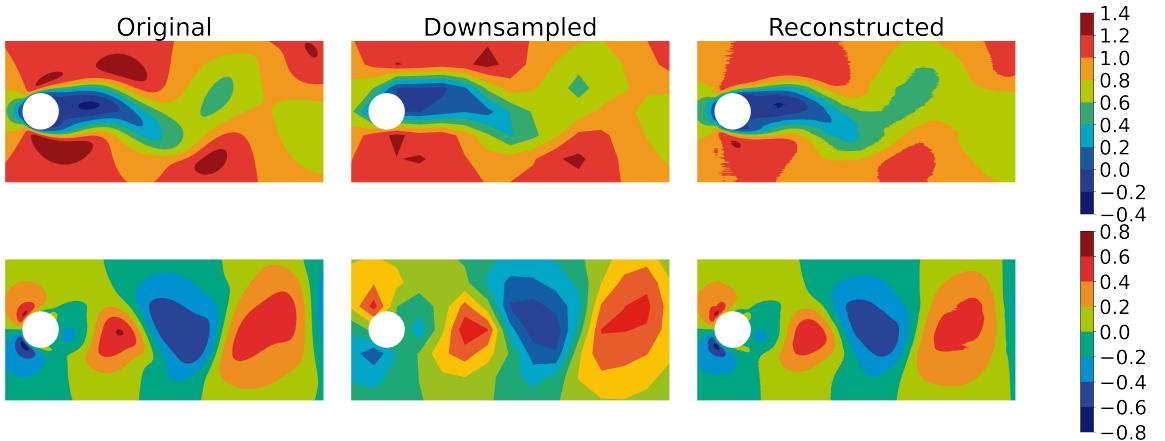


Figure 16: Reconstruction of a representative snapshot of the flow past a cylinder using ModelFLOWs-app. From left to right: original snapshot, the downsampled snapshot (input of the neural network) and the reconstructed snapshot (output of the algorithm). From top to bottom: Streamwise and normal velocity components.

Parameter	Symbol	Value
First Scaling	$1^{st} Scaling$	No
Second Scaling	$2^{nd} Scaling$	No
Training size	$\%_{test}$	0.8
Batch size	$N_{batch}$	23
Epoch number	$N_{epochs}$	500
Hyper parameterization	$Hyper$	No
Activation functions (only for $Hyper = No$ )	$AF$	ReLU
# Neurons (only for $Hyper = No$ )	$N_{neurons}$	13
Learning rate (only for $Hyper = No$ )	$l_r$	0.002
Loss function (only for $Hyper = No$ )	$l_f$	$mse$

Table 4: Parameters to choose on the data reconstruction algorithm and the values selected for the database presented in this section.

### 5.6. Module 2 - application 3: time-series forecasting models

Time-series forecasting models based on deep learning (Sec. 4.2.1) have been applied to two different problems modelling different types of complex flows. The main goal of this application is to extract data from the transient region of a numerical simulation and to predict the evolution of the simulation, so these ROMs have been developed to speed-up numerical simulations.

To show the good properties and robustness of the models proposed, two different methodologies have been followed. We first apply the hybrid framework (HybridDL) presented in Refs. [43, 3], where the proposed models are a combination of SVD and deep learning architectures. Secondly we follow a fully deep leaning framework (FullyDL), as in Ref. [65], where the proposed models are fully based on deep learning architectures. It is worth to mention that for this second application, in Ref. [65], dimensionality reduction was also carried out by HODMD, instead of SVD. Nevertheless, the present article only shows the application of the deep learning architecture itself. As seen in this section, the two methodologies available in ModelFLOWS-app to develop predictive ROMs can be applied to different kind of problems.

As mentioned before, HybridDL models first apply SVD to the database. This pre-processing allows to reduce the number of trainable parameters required by the models, simplifying the training. But, at the cost of loosing part of the spatial information, since the spatial dimension tensor is flattened into a single vector before applying SVD, forming the snapshot matrix eq. (2.1). This methodology shows good performance in several examples: the three-dimensional wake of a circular cylinder, synthetic jets [43], and reactive flows (axisymmetric, time varying, non-premixed, co-flow nitrogen-diluted methane flame [3]). Results about the latter case are shown. More information about the numerical simulacion of the laminar flame and the numerical set-up can be found in Refs. [66, 67]. From this detailed simulation, a database is extracted, composed by 10 variables (the temperature and 9 chemical species). The spatial dimension of the computational domain is defined for the axial direction: 100 points in the interval  $z \in [0, 0.12]$ ; and for the radial direction: 75 points in the interval  $r \in [0, 0.02]$ , respectively. The number of snapshots extracted is  $K = 999$ , equidistant in time with  $\Delta t = 2.5 \times 10^{-4}$ . Figure 17 shows the reactive flow test case. It compares the real solution with the predictions carried out using RNN and CNN architectures. The number of SVD modes retained in the first step of the model is 18. The figure shows a representative snapshot of the  $CO_2$  mass fraction and the evolution in time of temperature field extracted at two representative points from the computational domain. The RRMSE eq. (2.5) in the predictions is smaller than 4% and 3% for the CNN and RNN architectures, respectively, and the speed-up factor in the numerical simulations is higher than 100 in both cases. See more details about this example in Ref. [3].

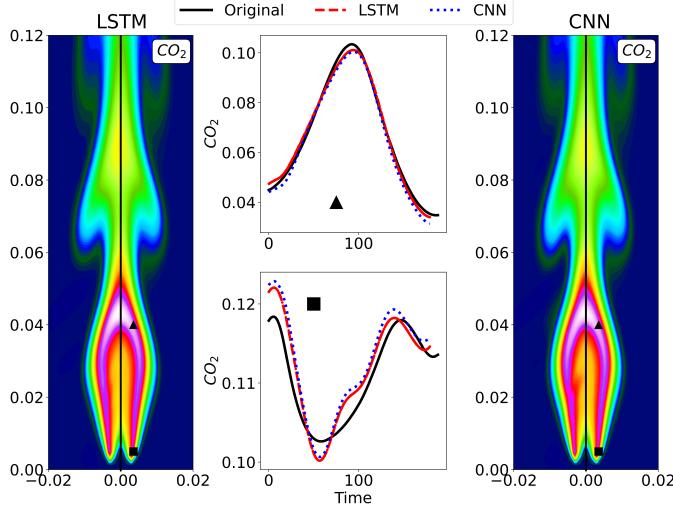


Figure 17: Predictive hybrid ROM to predict reactive flows. Representative snapshot showing concentration of  $CO_2$ . Predictions carried out using RNN (left) and CNN (right) architectures. Right and left parts of each contour show the original snapshot the prediction. Mid part of the figure: temporal evolution of temperature at two characteristic points of the computational domain. Points extracted where the triangle and square are located in the figure contour.

The second example uses the methodology FullyDL, which in contrast to HybridDL, any dimensionality reduction is carried out in the original database. This increases the number of trainable parameters and, therefore, the complexity of the neural network training. However, we keep the spatial information contained in the snapshots, i.e., the flow structures that define the flow dynamics. This methodology has been successfully tested in several problems modelling complex flows, as explained in Refs. [68, 65]. This section presents the results of applying these models to predict the evolution of a two-phase flow concentric jet. The flow comprises two liquid jets, consisting of two incompressible, viscous, and immiscible fluids, with Reynolds numbers of  $Re_1 = 30$  and  $Re_2 = 200$  for each phase, respectively, and a Weber number of  $We = 80$ . Both jets arise from two nozzles separated by a gap, whose length can be close to zero. The spatial dimension of the computational domain is defined for the streamwise direction in the interval  $x \in [0, 16]$  and  $y \in [0, 8]$ , respectively. The temporal interval between snapshots is  $\Delta t = 0.005$ . Details about the numerical simulations carried out to solve this problem can be found in Ref. [65]. The database used to obtain the results in Figure 18 is composed of a spatial grid formed by  $100 \times 100$  points

for the streamwise velocity, and 301 snapshots. Therefore, in this case the spatial dimension of the problem is  $J = 100 \times 100$  (in this framework we do not flatten the spatial dimension), and the temporal dimension is  $K = 301$ . Figure 18 compares two representative snapshot of the predictions carried out using the CNN and RNN models. The RRMSE eq. (2.5) computed in these predictions is 0.044 and 0.1 for the CNN and RNN architectures, respectively. The speed-up in the numerical simulations is 9.88 and 9.03 for the CNN and RNN architectures, respectively.

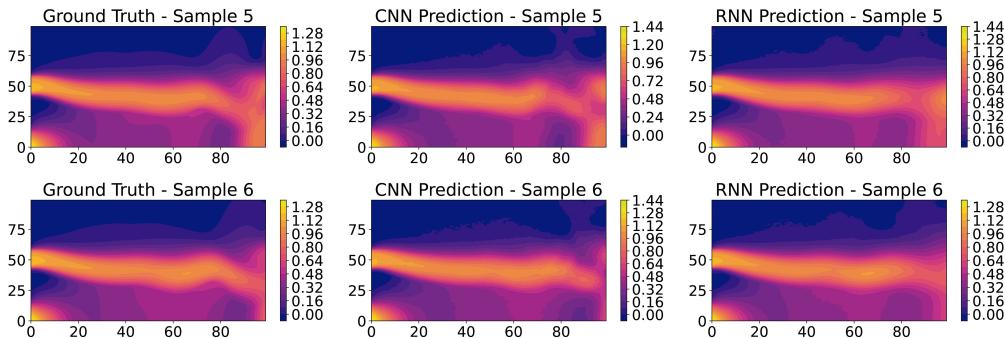


Figure 18: Predictions carried out using models inside FullyDL framework. Streamwise velocity in the original solution (left), and in the predictions using CNN (middle) and RNN (right) architectures.

In each methodology, we developed two different models: CNN and RNN, which is based on the Long Short-Term Memory (LSTM) architecture [69]. It is important to note that the LSTM architecture only uses a vector structure, which is not a problem in HybridDL because the spatial dimension was already flattened to apply SVD. However, the FullyDL framework takes the original snapshots as input, which means that the spatial dimension needs to be flattened. Due to this, the RNN model exhibits significantly better performance in the HybridDL framework compared to FullyDL. On the other hand, the CNN model shows better performance in the FullyDL framework than in HybridDL. This is because Convolutional Neural Networks were developed for analyzing snapshots, particularly in the field of computer vision[70]. Therefore, this architecture is capable of better capturing the spatial structures present in the snapshots.

As seen in Tabs. 5, 6 the calibration of the RNN and CNN architectures differ depending on the framework: HybridDL or FullyDL. Details are

presented below.

### 5.6.1. Calibration of models in HybridDL framework

Parameter	Symbol	Value
# modes (SVD)	$N$	18
First scaling	$1^{st}Scaling$	Auto
Second scaling	$2^{nd}Scaling$	MpM
Testing size	$\%_{test}$	0.2
Validation size	$\%_{val}$	0.15
Batch size	$N_{batch}$	12
# epochs	$N_{epochs}$	400
# input samples	$k$	10
# time ahead predictions	$p$	6
Hyper parameterization	$Hyper$	No
Hidden activation functions (only for $Hyper = \text{No}$ )	$HiddenAF$	ELU
Output activation functions (only for $Hyper = \text{No}$ )	$OutAF$	Tanh
# Neurons (only for $Hyper = \text{No}$ )	$N_{neurons}$	100
Shared dimension (only for $Hyper = \text{No}$ )	$SharedDim$	80
Learning rate (only for $Hyper = \text{No}$ )	$l_r$	0.005
Loss function (only for $Hyper = \text{No}$ )	$l_f$	custom_loss

Table 5: Hyperparameters of models belonging to the HybridDL framework. The symbol  $\#$  means number and the Value column shows the configuration used in Ref. [3].

Both RNN and CNN models share the same hyperparamenters for calibration, listed in Table 5. Note this calibration depends on the database. From the mentioned case, the reactive flow, presented in Ref. [3], the database analysed is formed by: the temperature and the most important 9 chemical species, 100 points on the streamwise direction, 75 on the normal direction and 999 snapshots. This gives a fourth-order tensor of dimensions  $10 \times 100 \times 75 \times 999$ .

The hyperparameter  $1^{st}Scaling$  indicates the general scaling to be applied to the database. This step is necessary in combustion databases, as the variables can differ significantly in magnitude among them. In this case, *autoscaling* has been used, which utilizes the standard deviation of each variable, ensuring that all variables have equal importance [22]. After this scaling step, the tensor is reshaped into a matrix, grouping all dimensions

except the temporal one into a single dimension. This results in a matrix with dimensions of  $75000 \times 999$ . Subsequently, SVD is applied, and the first  $N$  modes are retained, which contain the large scales.

The term *2<sup>nd</sup>Scaling* refers to the scaling applied to the temporal coefficients of the SVD, specifically for the reduced database ( $\hat{\mathbf{V}}_1^K$ ) used for training the models. In our case, *MpM* scaling has been employed, as it has been demonstrated to be suitable when dealing with non-periodic databases [3].

The hyperparameters  $\%_{val}$  and  $\%_{test}$  refer to the percentage of the total number of samples ( $K$ ) that is used respectively for validation and testing, the remaining are used for training.  $N_{batch}$  is used to select the batch size and  $N_{epochs}$  to select the number of epochs to use for training, i.e., the training length. When using the rolling window technique (previous Fig. 8), it is necessary to specify the number of input samples for each window  $k$  and the number of target samples (time ahead predictions)  $p$ .

Both the RNN and CNN models have a feature called *Hyper* that allows for automatic hyperparameterization of the remaining features. The *HiddenAF* parameter sets the activation function of the hidden layers, while *OutAF* sets the activation function of the output layer. In the case of the RNN model, the *N\_neurons* parameter specifies the number of neurons to use in the LSTM layer. Additionally, the *SharedDim* hyperparameter modifies the number of neurons in the models. The *lr* parameter specifies the learning rate, and *lf* determines the loss function used for training the models.

For the *lf* parameter, a new physics-aware loss function called *PA-MSE* has been implemented specifically for reacting flows databases. This loss function aims to achieve a good reconstruction while ensuring that the sum of the species is equal to 1. In ModelFLOWS-app, this loss function can be selected as *customLoss*.

If the *Hyper* parameter is set to *Yes*, these last six hyperparameters can be automated. Otherwise, they must be explicitly specified.

See more details about this application in Ref. [3].

### 5.6.2. Calibration of models in framework FullydDL

Parameter	Symbol	Value
Training size	$\%_{train}$	0.8
Validation size	$\%_{val}$	0.2
Batch size	$N_{batch}$	5
Type of model	$Model_T$	rnn
# epochs	$N_{epochs}$	140
# input samples	$k$	10
# time ahead predictions	$p$	2
Hyper parameterization	$Hyper$	No
Hidden activation functions (only for $Hyper = \text{No}$ )	$HiddenAF$	Dense
Output activation functions (only for $Hyper = \text{No}$ )	$OutAF$	None
# Neurons (only for RNN and when $Hyper = \text{No}$ )	$N_{neurons}$	400
Shared dimension (only for $Hyper = \text{No}$ )	$SharedDim$	80
Learning rate (only for $Hyper = \text{No}$ )	$lr$	0.005

Table 6: Hyperparameters of models belonging to the FullyDL framework. The symbol # means number and the Value column shows the configuration used in Ref. [65].

Similar to the previous subsection, in this one we summarize the hyperparameters available, listed on Tab. 6, for both models, RNN and CNN, in the FullyDL framework. Some of them share the same meaning as in previous subsection, these are  $N_{batch}$ ,  $N_{epochs}$ ,  $k$ ,  $p$ ,  $Hyper$ ,  $HiddenAF$ ,  $OutAF$ ,  $SharedDim$  and  $lr$ . Hyperparameters  $\%_{train}$  and  $\%_{val}$  specify which percentage of samples will be used for training and validation, respectively. While the entire database is used for testing. Also, in this framework the hyperparameter  $Model_T$  specifies which model to use: RNN or CNN. Finally,  $N_{neurons}$  shows the number of neurons used in the LSTM layer, inside the RNN model. Similar to HybridDL framework, the four bottom hyperparameters in Tab. 6 can be automated if  $Hyper = \text{Yes}$ , otherwise they must be specified.

## 6. Conclusions

This article introduces the ModelFLOWS-app, a novel software tool that has proven to be effective in developing accurate and robust fully data-driven hybrid reduced order models. The software combines modal decomposition and deep learning strategies to uncover key patterns in dynamical systems,

enabling a deeper understanding of the underlying physics. It offers valuable capabilities such as database reconstruction from sparse measurements and accurate prediction of system dynamics, providing an alternative to computationally expensive numerical simulations.

Although initially developed for analyzing turbulent flows, ModelFLOWs-app has demonstrated its versatility and exceptional properties in a wide range of industrial applications involving complex non-linear dynamical systems. Examples include various applications in fluid dynamics such as temporal forecasting in wind turbines, predictions in reactive flows and future projections in compressible flows with buffeting, identification of flow instabilities in turbulent flows, medical imaging pattern identification and reconstruction, wind velocity prediction using lidar measurements, identification of flutter instability in flight tests, et cetera.

Readers are encouraged to thoroughly explore the applications presented in this article and further enhance their knowledge by accessing the tutorials and videos available on the ModelFLOWs-app website [18].

## 7. Acknowledgements

The authors would like to acknowledge the collaboration of the following researchers, who have contributed by sharing databases, writing articles, providing new ideas, and engaging in fruitful discussions. The collective work carried out with these researchers over the past years has greatly enhanced the robustness of the current codes. These researchers are: Prof. J.M. Vega (UPM), Dr. R. Vinuesa (KTH), Prof. A. Parente (ULB), Prof. L. Brant (KTH), Dr. M. Rosti (OIST), Prof. O. Tammisola (KTH), and Prof. J. Soria (Monash Uni.). The authors would also like to express their gratitude to the research group ModelFLOWs for their valuable discussions, assistance in generating new databases, and for their support in testing some of the developed tools. S.L.C., A.C. and S.R.A. acknowledge the grant PID2020-114173RB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and the support of Comunidad de Madrid through the call Research Grants for Young Investigators from Universidad Politécnica de Madrid. A.C. also acknowledges the support of Universidad Politécnica de Madrid, under the programme ‘Programa Propio’. E.L. and S.L.C. acknowledge the support provided by Grant No. TED2021-129774B-C21 and by Grant No. PLEC2022-009235, funded by MCIN/AEI/10.13039/501100011033 and by the European Union “NextGenerationEU”/PRTR.

## References

- [1] L. Sirovich, Turbulence and the dynamic of coherent structures, parts i–iii, *Q. Appl. Math.* 45(3) (1987) 561.
- [2] P. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (2010) 5–28.
- [3] A. Corrochano, R. Freitas, A. Parente, S. Le Clainche, A predictive physics-aware hybrid reduced order model for reacting flows, arXiv:2301.09860 (2023).
- [4] S. Le Clainche, E. Ferrer, A reduced order model to predict transient flows around straight bladed vertical axis wind turbines, *Energies* 11 (2018) 566–578.
- [5] J. Kou, S. Le Clainche, W. Zhang, A reduced-order model for compressible flows with buffeting condition using higher order dynamic mode decomposition with a mode selection criterion, *Phys. Fluids* 30 (2018) 016103.
- [6] S. Le Clainche, D. Izvassarov, M. Rosti, L. Brandt, O. Tammisola, Coherent structures in the turbulent flow of an elastoviscoplastic fluid, doi:10.1017/jfm.2020.31, *J. Fluid Mech.* (2020) A5.
- [7] S. Le Clainche, M. Rosti, L. Brandt, A data-driven model based on modal decomposition: application to the turbulent channel flow over an anisotropic porous wall, *J. Fluid Mech.* 939 (2022).
- [8] S. Le Clainche, Prediction of the optimal vortex in synthetic jets, *Energies* 12 (2019) 1635–1655.
- [9] E. Lazpita, Á. Martínez-Sánchez, A. Corrochano, S. Hoyas, S. Le Clainche, R. Vinuesa, On the generation and destruction mechanisms of arch vortices in urban fluid flows, *Physics of Fluids* 34 (2022) 051702.
- [10] J. Perez, F. Sastre, S. Le Clainche, A. Velazquez, J. Vega, Three-dimensional flow field reconstruction in the wake of a confined square cylinder using planar piv data, *Exp. Therm. Fluid Sci.* 133 (2022) 110523.

- [11] F. Sastre, S. Le Clainche, A. Velazquez, J. Vega, Higher order dynamic mode decomposition applied to post-process a limited amount of noisy piv data, 47th AIAA Fluid Dynamics Conference, AIAA paper 3304 (2017).
- [12] P. Díaz, A. Corrochano, M. López-Martín, S. L. Clainche, Deep learning combined with singular value decomposition to reconstruct databases in fluid dynamics, 2023. [arXiv:2305.08832](https://arxiv.org/abs/2305.08832).
- [13] N. Groun, M. Villalba-Orero, E. Lara-Pezzi, E. Valero, J. Garicano-Mena, S. Le Clainche, Higher order dynamic mode decomposition: From fluid dynamics to heart disease analysis, Comp. Biolog. Medic. 144 (2022) 105384.
- [14] N. Groun, M. Villalba-Orero, E. Lara-Pezzi, E. Valero, J. Garicano-Mena, S. Le Clainche, A novel data-driven method for the analysis and reconstruction of cardiac cine mri, Comp. Biolog. Medic. 151 (2022) 106317.
- [15] S. Le Clainche, L. Lorente, J. Vega, Wind predictions upstream wind turbines from a lidar database, Energies 11 (2018) 543.
- [16] S. Le Clainche, R. Moreno-Ramos, P. Taylor, J. M. Vega, A new robust method to study flight flutter testing, J. Aircraft DOI: 10.2514/1.C034863 (2018).
- [17] C. Méndez, S. Le Clainche, R. Moreno, J. Vega, A new method to predict flutter, Aerosp. Sci. Tech. 114 (2021) 106749.
- [18] D. by ModelFLOWS research group, Modelflows-app, The website of the software is available at <https://modelflows.github.io/modelflowsapp/> (2023).
- [19] streamlit, Available at <https://streamlit.io/> (????).
- [20] Python, Available at <https://www.python.org/downloads/> (????).
- [21] J. L. Lumley, The structure of inhomogeneous turbulent flows, In: Yaglam, A.M., Tatarsky, V.I. (eds.) Proceedings of the International Colloquium on the Fine Scale Structure of the Atmosphere and Its Influence on Radio Wave Propagation. Doklady Akademii Nauk SSSR, Nauka, Moscow (1967).

- [22] A. Parente, J. C. Sutherland, Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity, *Combustion and flame* 160 (2013) 340–350.
- [23] L. R. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika* 16 (1966) 279–311.
- [24] L. De Lathawer, B. De Moor, J. Vandewalle, On the best rank-1 and rank- $(r_1, r_2, \dots, r_n)$  approximation of higher-order tensors, *SIAM J. Matrix. Anal. Appl.* (2000) 1324–1342.
- [25] L. De Lathawer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, *SIAM J. Matrix. Anal. Appl.* (2000) 1253–1278.
- [26] L. Lorente, J. Vega, A. Velazquez, Generation of aerodynamics databases using high-order singular value decomposition, *J. Aircraft* (2008) 1779–1788.
- [27] L. Lorente, J. Vega, A. Velazquez, Compression of aerodynamic databases using high-order singular value decomposition, *Aeros. Sci. Techn.* (2010) 168–177.
- [28] S. de Lucas, J. Vega, A. Velazquez, Aeronautic conceptual design optimization method based on higher order singular value decomposition, *AIAA Journal* 49 (2011) 2713–2725.
- [29] N. Benito, J. Arias, A. Velazquez, J. Vega, Real time performance improvement of engineering control units via higher order singular value decomposition: Application to a si engine, *Control Eng. Prac* 11 (2011) 1315–1327.
- [30] A. Moreno, A. Jarzabek, J. Perales, J. Vega, Aerodynamic database reconstruction via gappy high order singular value decomposition, *Aerosp. Sci. Tech.* 52 (2016) 115–128.
- [31] D. Venturi, G. Karniadakis, Gappy data and reconstruction procedures for flow past a cylinder, *J. Fluid. Mech.* 519 (2004) 315–336.
- [32] J. Beckers, M. Rixen, Eof calculations and data filling from incomplete oceanographic datasets, *J. Atmos. and Ocean. Tech.* 20 (2003) 1839–1856.

- [33] L. Lorente, J. Vega, A. Velazquez, Efficient computation of the pod manifold containing the information required to generate a multi-parameter aerodynamic database, *Aerosp. Sci. Tech.* 25 (2013) 152–160.
- [34] S. Le Clainche, J. M. Vega, Higher order dynamic mode decomposition, *SIAM J. Appl. Dyn. Sys.* 16 (2) (2017) 882–925.
- [35] J. Vega, S. Le Clainche, Higher order dynamic mode decomposition and its applications, BOOK: Elsevier (2020).
- [36] Matlab, mathworks, Available at <https://it.mathworks.com/products/matlab.html> (????).
- [37] K. Chen, J. Tu, C. Rowley, Variants of dynamic mode decomposition: boundary condition, koopman and fourier analyses, *TJ. Nonlinear Sci.* 22 (2012) 8871—887.
- [38] B. Begiashvili, N. Groun, J. Garicano-Mena, S. Le Clainche, E. Valero, Data-driven modal decomposition methods as feature detection techniques for flow problems: a critical assessment, *Phys. of Fluids* (accepted) (2023).
- [39] S. Le Clainche, J. M. Vega, J. Soria, Higher order dynamic mode decomposition of noisy experimental data: the flow structure of a zero-net-mass-flux jet, *Exp. Therm. Fluid Sci.* 88 (2017) 336–353.
- [40] S. Le Clainche, Prediction of the optimal vortex in synthetic jets, *Energies* 12 (9) (2019).
- [41] S. Le Clainche, F. Sastre, A. Velazquez, J. M. Vega, Higher order dynamic mode decomposition applied to study flow structures in noisy piv experimental data, AIAA 2017-3304 . Proceedings of 47th AIAA Fluid Dynamics Conference, 5 – 9 June, Denver, CO, USA (2017).
- [42] S. Le Clainche, L. Lorente, J. Vega, Wind predictions upstream wind turbines from a lidar database, *Energies* 11 (3) (2018) 543.
- [43] R. Abadía-Heredia, M. López-Martín, B. Carro, J. Arribas, J. Perez, S. Le Clainche, A predictive hybrid reduced order model based on proper orthogonal decomposition combined with deep learning architectures, *Expert Systems with Applications* 187 (2022) 115910. doi:<https://doi.org/10.1016/j.eswa.2021.115910>.

- [44] H. Eivazi, S. Le Clainche, S. Hoyas, R. Vinuesa, Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows, *Expert Syst. Appl.* 202 (2022) 117038.
- [45] G. D'Alessio, A. Attili, A. Cuoci, H. Pitsch, A. Parente, Analysis of turbulent reacting jets via principal component analysis, in: *Data Analysis for Direct Numerical Simulations of Turbulent Combustion*, Springer, 2020, pp. 233–251.
- [46] A. Corrochano, G. D'Alessio, A. Parente, S. Le Clainche, Higher order dynamic mode decomposition to model reacting flows, *International Journal of Mechanical Sciences* 249 (2023) 108219. URL: <https://www.sciencedirect.com/science/article/pii/S0020740323001212>. doi:<https://doi.org/10.1016/j.ijmecsci.2023.108219>.
- [47] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D. J. Inman, 1d convolutional neural networks and applications: A survey, *Mech. Syst. Signal Process.* 151 (2021) 107398.
- [48] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: Lstm cells and network architectures, *Neural Comput.* 31(7) (2019) 1235—1270.
- [49] D. P. Kingma, J. L. Ba, Adam: A method for stochastic optimization, Dec. 2015, Accessed: Sep. 21, 2020. [Online]. Available: <https://arxiv.org/abs/1412.6980v9> (2020).
- [50] A. Kaptanoglu, K. Morgan, C. Hansen, S. Brunton, Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches, *Physical Review E* 104 (2021). doi:[10.1103/PhysRevE.104.015206](https://doi.org/10.1103/PhysRevE.104.015206).
- [51] M. Atzori, P. Torres, A. Vidal, S. L. Clainche, S. Hoyas, R. Vinuesa, High-resolution large-eddy simulations of simplified urban flows, arXiv preprint arXiv:2207.07210 (2022).
- [52] Á. Martínez-Sánchez, E. Lazpita, A. Corrochano, S. Le Clainche, S. Hoyas, R. Vinuesa, Data-driven assessment of arch vortices in simplified urban flows, *International Journal of Heat and Fluid Flow* 100 (2023) 109101.

- [53] Á. Martínez-Sánchez, E. López, S. L. Clainche, A. Lozano-Durán, A. Srivastava, R. Vinuesa, Causality analysis of large-scale structures in the flow around a wall-mounted square cylinder, arXiv preprint arXiv:2209.15356 (2022).
- [54] T. R. Oke, Street design and urban canopy layer climate, *Energy and buildings* 11 (1988) 103–113.
- [55] B. Monnier, S. A. Goudarzi, R. Vinuesa, C. Wark, Turbulent structure of a simplified urban fluid flow studied through stereoscopic particle image velocimetry, *Boundary-layer meteorology* 166 (2018) 239–268.
- [56] B. M. Sumer, N. Christiansen, J. Fredsøe, The horseshoe vortex and vortex shedding around a vertical wall-mounted cylinder exposed to waves, *Journal of Fluid Mechanics* 332 (1997) 41–70.
- [57] S. Le Clainche, J. M. Vega, Higher order dynamic mode decomposition to identify and extrapolate flow patterns, *Phys. Fluids* 29 (8) (2017) 084102.
- [58] Nek5000, ???? URL: <https://nek5000.mcs.anl.gov/>.
- [59] C. K. Batchelor, G. K. Batchelor, An introduction to fluid dynamics, Cambridge university press, 1967.
- [60] C. H. Williamson, Oblique and parallel modes of vortex shedding in the wake of a circular cylinder at low reynolds numbers, *Journal of Fluid Mechanics* 206 (1989) 579–627.
- [61] C. H. Williamson, Vortex dynamics in the cylinder wake, *Annual review of fluid mechanics* 28 (1996) 477–539.
- [62] E. Muñoz, S. Le Clainche, On the topology patterns and symmetry breaking in two planar synthetic jets, *Physics of Fluids* 34 (2022). doi:[10.1063/5.0080834](https://doi.org/10.1063/5.0080834).
- [63] E. Muñoz, H. Dave, G. D'Alessio, A. Parente, S. Le Clainche, Extraction and analysis of flow features in planar synthetic jets using different machine learning techniques, *SSRN Preprint* (2023). doi:<https://dx.doi.org/10.2139/ssrn.4429450>.

- [64] Y. Bengio, Practical recommendations for gradient-based training of deep architectures, 2012. [arXiv:1206.5533](https://arxiv.org/abs/1206.5533).
- [65] L. Mata, R. Abadía-Heredia, M. Lopez-Martin, J. M. Pérez, S. L. Clainche, Forecasting through deep learning and modal decomposition in two-phase concentric jets (2023). [arXiv:2212.12731](https://arxiv.org/abs/2212.12731).
- [66] G. D'Alessio, A. Cuoci, G. Aversano, M. Bracconi, A. Stagni, A. Parente, Impact of the partitioning method on multidimensional adaptive-chemistry simulations, *Energies* 13 (2020) 2567.
- [67] G. D'Alessio, A. Parente, A. Stagni, A. Cuoci, Adaptive chemistry via pre-partitioning of composition space and mechanism reduction, *Combustion and Flame* 211 (2020) 68–82.
- [68] M. Lopez-Martin, S. Le Clainche, B. Carro, Model-free short-term fluid dynamics estimator with a deep 3d-convolutional neural network, *Exp. Syst. Appl.* 177 (2020) 114924.
- [69] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, *Neural Computation* 9 (1997) 1735–1780. URL: <https://doi.org/10.1162/neco.1997.9.8.1735>. doi:[10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [70] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86 (1998) 2278–2324. doi:[10.1109/5.726791](https://doi.org/10.1109/5.726791).