
SPARSE SURFACE PRESSURE-BASED RECONSTRUCTION OF THE FLOW AROUND A THICK AIRFOIL OVER A RANGE OF ANGLES OF ATTACK

Quentin Bucquet

EM2C, CentraleSupélec,
Université Paris-Saclay,
France

quentin.bucquet@centralesupelec.fr

Bérengère Podvin

EM2C, CentraleSupélec,
Université Paris-Saclay,
France

Caroline Braud

LHEEA, CNRS, Centrale Nantes,
Nantes-Université,
France

Emmanuel Guilmineau

LHEEA, CNRS, Centrale Nantes,
Nantes-Université,
France

October 27, 2025

ABSTRACT

We present an efficient neural-based approach to estimate the instantaneous flow field around an airfoil from limited surface pressure measurements. The model, denoted SNN-POD, relies on two independent shallow neural networks to predict the instantaneous flow over a wide range of angles of attack $[10^\circ, 20^\circ]$. At all angles the global model correctly recovers the average characteristics of the flow from single-time sensor data, thus allowing combination with local, angle-dependent models. The method is applied to 2D URANS simulations of a thick airfoil at a Reynolds number of $Re = 4.5 \times 10^6$. The training set consists of snapshots obtained from a coarse sampling ($1 - 2^\circ$) of the angle of attack range. A variance-based criterion is used to determine the number and positions of sensors. Tests are carried out for unseen snapshots at angles of attack within the set (sampled angles) as well as outside the set (interpolated angles). The maximum MSE error of attack for sampled and interpolated angles is respectively 2.9% and 6.6%. This makes it possible to develop adaptive strategies to improve the estimation if necessary.

Keywords Sparse Flow Reconstruction; Data-driven; Reduced-Order Modeling; Wind Energy

1 Introduction

In the context of wind energy, accurate prediction of mean and instantaneous wind turbine aerodynamic loads is essential for ensuring the structural integrity and efficiency of wind turbines, particularly during gusts or dynamic stall. Originally introduced by Glauert (1935), the Blade Element Momentum (BEM) method has long been the workhorse for aerodynamic load prediction. Its appeal lies in its simplicity and computational efficiency, assuming stationary and uniform inflow. Over the decades, BEM has been refined with numerous corrections to account for three-dimensional effects such as tip losses (Prandtl and Betz, 1927), root effects (Chaviaropoulos and Hansen, 2000), yawed inflow (Madsen et al., 2020), and dynamic inflow conditions (Snel and Schepers, 1995). Based on comparisons with large-scale field test campaigns, Potentier et al. (2021) identified the most impactful corrections and demonstrated that, when supplemented with unsteady aerodynamic models, BEM can reliably capture mean load trends in most conditions. However capturing instantaneous loads remains an important challenge, as it requires knowledge of the surrounding flow field—information that is challenging to access in practical or in-situ settings where only sparse, low-resolution wall-pressure measurements are typically available.

The field of sparse flow reconstruction, where a reliable estimator of the flow is built using only limited or noisy measurements, has been widely studied over the last decades in many engineering applications, ranging from medical diagnostics to atmospheric measurements. In the medical field, the spatially and temporally limited acquisition of the four-dimensional flow magnetic resonance imaging (4D MRI) is complemented by flow reconstruction techniques to determine blood flow during cerebral aneurysms (Zhang et al., 2022). Reconstruction of turbulent flows from restricted observations is also crucial for climatology (Schneider et al., 2017), weather forecast (Kalnay, 2002), atmospheric (Kaimal and Finnigan, 1994) and urban boundary layers studies (Blackman and Perret, 2016; Lu and Papadakis, 2023), where metrology techniques (meteorological masts, LIDAR, pressure balloons) give access to only a small fraction of the atmospheric volume of interest. The same problem applies in wind energy when estimating the inflow wind or the downstream wake for both wind turbines and wind farms using high-frequency SCADA data on the turbines (Rott et al., 2020).

Several challenges for flow reconstruction need to be addressed. Firstly, the target flow field to estimate is often turbulent and therefore is characterized by high-dimensional data, which generally needs to be reduced down to a more compact representation. Secondly, physical and economic constraints can limit the number, position and acquisition rate of the sensors so that optimization strategies may be needed, (Hollenbeck et al., 2024; Bukowski et al., 2025). Thirdly, the mapping between the observations and the target field may be nonlinear, as for instance in the reconstruction of a velocity field from pressure measurements (Graziani et al., 2018). Finally, an adequate compromise should be found between the cost and the accuracy of the estimation technique.

Comprehensive reviews of sparse reconstruction methodologies can be found in Callaham et al. (2019); Nair and Goza (2020); Manohar et al. (2018), but a few key principles are worth recalling here. The majority of approaches rely on Reduced-Order Models (ROMs) to project the high-dimensional flow dynamics onto a low-order manifold. The reconstruction task then reduces to estimating the reduced state, which can subsequently be mapped back to the full-order field. The projection of the high-dimensional dynamics of the complex flow field can be carried out using a linear decomposition basis that can be data-agnostic, such as Fourier modes or wavelets, or data-driven, such as Proper Orthogonal Decomposition aka POD (Lumley, 1967), or Dynamic Mode Decomposition aka DMD (Schmid, 2010). A comprehensive review of modal decompositions can be found in Taira et al. (2017). However, linear reconstructions may be limited in their ability to capture nonlinear flow features. This has motivated the development of more sophisticated methods such as sparse regression techniques to address these issues (Loiseau et al., 2018).

Generally speaking, recent advances in machine learning have opened new perspectives for flow reconstruction : neural-network-based approaches can learn nonlinear mappings between sparse measurements and the full-order flow, although linear expansions can still be used to reduce the dimensionality of the data (Swischuk et al., 2019). Among the various neural network architectures that have attracted interest, autoencoders have shown promise for discovering nonlinear manifolds with improved reconstruction accuracy compared to linear subspaces (Lee and Carlberg, 2019; Xu and Duraisamy, 2020; Tan, 2023) and can provide causality insights (Fukami and Taira, 2025). Graph neural networks (GNNs) have also been shown to be a versatile and powerful tool for estimation purposes, as was recently evidenced by Duthé et al. (2023), who developed a message-passing GNN to predict steady two-dimensional pressure and velocity fields from surface pressure distributions around airfoils of arbitrary shape.

Possible drawbacks of data-driven neural frameworks include their lack of interpretability and their disconnection from physics-based knowledge. These limitations are addressed by Physics-Informed Neural Networks (PINNs), introduced by Raissi et al. (2019), which are trained with a loss that seeks to both minimize the discrepancy with the observations and to enforce physics conservation laws. Implementations included velocity reconstruction in a cylinder wake from sparse and incomplete velocity measurements (Xu et al., 2023), bluff-body wake and Kolmogorov flow reconstruction with a Convolution Neural Network trained for various physics-based loss formulations (Mo and Magri, 2025), and reconstruction of three-dimensional turbulent wall-bounded flows from two-dimensional planar measurements with a physics-constrained variational autoencoder (Hora et al., 2024). However, while PINNs show great potential, they can induce high computational costs, and can struggle on complex flows, such as multi-scale dynamic systems (Li et al., 2024).

In contrast, a low-cost alternative that has shown substantial promise is the shallow neural network model introduced by Erichson et al. (2020) (referred therein as "shallow decoder", denoted as SNN by Carter et al., 2021), which relies on a simple neural architecture to map sparse sensor inputs to high-dimensional flow fields. The shallow neural network model has demonstrated superior performance compared to POD-based linear reconstructions in test cases such as flow past a cylinder, sea-surface temperature dynamics, and synthetic turbulence (Erichson et al., 2020). Shallow neural networks have further been used as nonlinear refinements of linear state estimation (Carter et al., 2021) or as ROM-based models for reconstructing vorticity fields from vorticity- and surface stress-measuring sensors on the body of the flat plate at high angle of attack (Nair and Goza, 2020). The framework of shallow neural networks has also been extended to taken into account temporal information. Williams et al. (2024) successfully tested a shallow recurrent

decoder (SHRED) that relies on sensor past history for reconstruction and forecasting tasks. This approach can be complemented by a ROM (SHRED-ROM, Kutz et al., 2024; Tomasetto et al., 2025).

In the present work, a hybrid approach blending POD and shallow neural networks is used to reconstruct the flow velocity in the wake of a thick airfoil using sparse, instantaneous surface pressure measurements. The dataset consists of unsteady two-dimensional URANS simulations at high Reynolds number obtained over various angles of attack. The model output or target is constituted by the instantaneous POD amplitudes of the flow velocity, which are split into a time-dependent and a time-independent parts, each of which is learnt independently by a shallow neural network.

The remainder of this paper is structured as follows. Section 2 describes the URANS simulations from which the training and test datasets are constructed, with a focus on the surface pressure signal (input) and the streamwise velocity field (final output). The reconstruction procedure is detailed in Section 3. Evaluation of the approach and of its robustness is carried out in Section 4. A conclusion and discussion of future developments are given in Section 5.

2 Dataset presentation

2.1 Simulation set up

The present dataset consists in two-dimensional simulations of a wind turbine airfoil for angles of attack ranging from 10° to 20° . It is generated with the ISIS-CFD flow solver (Visonneau et al., 2014), developed by Centrale Nantes and CNRS and part of the FINETM/Marine computing suite, using an incompressible Unsteady Reynolds-Averaged Navier Stokes (URANS) approach. The $k - \omega$ SST model is used for turbulence modeling. A finite volume method spatially discretizes the transport equations on an unstructured mesh which is automatically prolonged (refined) or restricted (derefined) using Adaptive Mesh Refinement (AMR) to accurately capture local gradients without leading to excessive computational costs. Flow variables are stored at the center of the arbitrary shaped cells. Volume and surface integrals are computed using second-order approximations, whereas fluxes at the mesh faces are reconstructed by linear extrapolation of the integrand from the neighboring cell centers. A second-order backward difference scheme is used for time discretization. A constant time step computed as $\Delta t U_\infty / c = 0.01$ convective time units, where U_∞ is the inlet velocity and c the blade chord, is applied for all the simulations, and snapshots are stored every 5 time steps.

The airfoil geometry is extracted from a 3D scan of a 2MW wind turbine blade at 80% of the rotor diameter, featuring a chord length $c = 1.25$ m and a thickness of 20%. The chord-based Reynolds number is $Re = 4.5 \times 10^6$, which reproduces the full-scale geometry in the CSTB climatic wind tunnel (Neunaber et al., 2022; Braud et al., 2024). A comparison of the experimental and numerical global loads was carried out in Neunaber et al. (2022) and showed a good agreement.

For each angle of attack, the simulation was run for at least 20000 time steps, corresponding to approximately 200 convective time units, and required approximately 300 core hours of computational time. While AMR is used when running the simulations, the post-processing and training / testing of our flow reconstruction methodology are performed on a fixed grid, taken as the mesh of the last iteration of the 14° simulation, consisting of $N_s = 91\,912$ grid points.

2.2 Angle of attack discretization

The angles of attack (AoA) featured in the training set form a set $I_{train} = \{10^\circ, 12^\circ, 14^\circ, 15^\circ, 16^\circ, 18^\circ, 20^\circ\}$. This range is selected to cover various flow regimes: cases with fully attached flows, transitional cases where the boundary layer starts to separate from the trailing edge (around 14° in the simulations, 12° in the experiments Braud et al., 2024), and cases where the boundary layer is separated, with the separation point moving upstream toward the leading edge. This diversity of AoA reflects the natural variability of the angle of attack for a wind turbine due to changes in rotation speed and pitch in normal operating conditions. Additionally, the turbine and the airfoil are immersed in a complex turbulent atmospheric environment which induces strong local fluctuations of the angle of incidence due to wind gusts, wind ramps (i.e. rapid changes in incoming wind speed) or yaw misalignment, for instance. The angle of attack increments were of 1° in the range $[14^\circ, 16^\circ]$, which corresponds to the onset of separation, and 2° otherwise. This coarse discretization over the range of AoA was chosen in order to allow fast training and to evaluate the generalization capabilities of the model.

2.3 Wall pressure

This section describes the pressure measurements on the airfoil, a subset of which will constitute the model inputs.

2.3.1 Wall pressure distribution

The distribution of the time-averaged wall pressure coefficient $C_p = \frac{\Delta p}{q_0}$, where Δp is the differential between the wall pressure and the reference static pressure taken at freestream, and q_0 is the dynamic pressure, is shown in Fig. 1(a) for $\text{AoA} \in I_{\text{train}}$. For all angles of attack, the distributions exhibit similar characteristics : a strong suction peak at the leading edge that increases with AoA, a high suction region in the first half of the blade, followed by a slow decrease of the suction down to the trailing edge. It should be noted that some asperities in the profile are due to discontinuities in the geometry, which was directly obtained from a coarse 3D scan of a real-life blade. Beyond separation, for $\text{AoA} \geq 14^\circ$, the progressive flattening of the curves with AoA near the trailing edge indicates a shift of the flow separation point towards the leading edge. This evolution is essentially similar to that described in Neunaber et al. (2022) and Braud et al. (2024) for the corresponding experimental configuration (however, in the experiment, separation was observed around $\text{AoA} \geq 12^\circ$ which can be attributed to slight differences in the ambient environment that is more noisy experimentally).

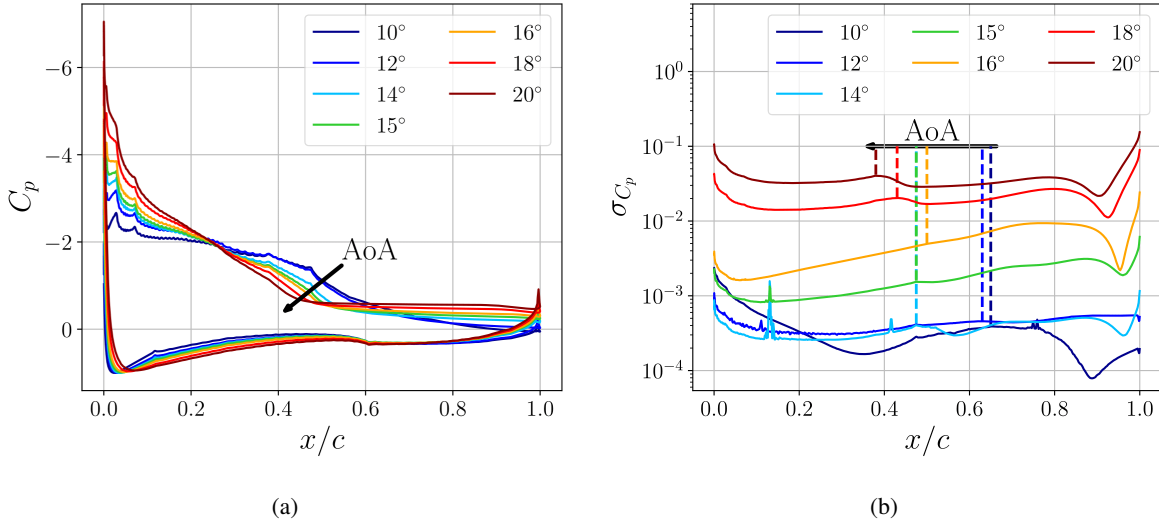


Figure 1: a) Time-averaged pressure coefficient C_p distribution around the airfoil b) Standard deviation σ_{C_p} of C_p (suction side only) in chordwise direction for angles of attack from 10° to 20° . Vertical dashed lines correspond to the chordwise location of the Intermittent Separation Point (ISP), which is defined as the local maximum of σ_{C_p} on the mid-chord pressure suction side

Figure 1(b) shows the distribution of the standard deviation of the wall pressure coefficient, σ_{C_p} . The level of pressure fluctuations remains very low for angles of attack where the flow remains attached or starts to separate (10° , 12° , 14°). For higher angles of attack associated with separated flow regimes, σ_{C_p} increases progressively. At high Reynolds number, this monotonic increase with the angle of the attack is consistent with experimental observations (Broeren and Bragg, 2001, B. Hanna, private communication). For all angles of attack, a distinct Leading Edge Maximum, associated with fluctuations near the suction peak, is observed. A local maximum of σ_{C_p} was also consistently observed across all angles of attack in the chordwise region $x/c \in [0.3; 0.7]$. It shifts gradually toward the leading edge as the angle of attack increases, with an approximately constant rate of displacement. The location of the fluctuation maximum was called the Intermittent Separation Point (ISP) in Braud et al. (2024) as it characterizes the region where intermittent separation point takes place. Another peak—referred to as the Trailing Edge Maximum—appears around $x/c \approx 0.8$, linked to the unsteady interaction between the separated shear layer and the airfoil surface in the trailing-edge region. These three regions of large fluctuations were also identified in the experiment (Braud et al., 2024) for individual chords. However some differences were noted with the experiment: fluctuation levels associated with the ISP were found to be higher in the experiment, and the additional peak found at the trailing edge in the simulation above separation was not observed there. These differences could be respectively explained by the 2D character of the simulation and the relatively low resolution of the experiment.

2.3.2 Wall pressure fluctuations

In this section and in the remainder of the work, the frequency auto-spectrum of the variable x is identified as $S_x(f)$, with f denoting the frequency. Fig. 2 shows the pre-multiplied frequency spectrograms of the pressure signal fS_{C_p}

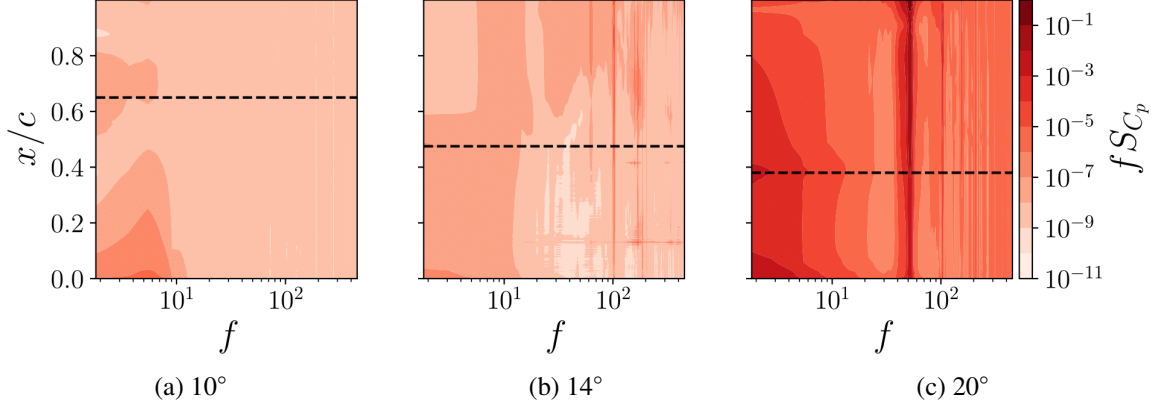


Figure 2: Pre-multiplied time spectrograms of C_p for AoA = 10° (a), AoA = 14° (b), AoA = 20° (c). Black dashed lines corresponds to the location of the ISP

for the suction side of the blade at three different angles of attack corresponding to attached flow (10°), the onset of separation (14°) and full separation (20°). As expected from figure 1(b), the energy levels increase with AoA. Below separation, low-frequency components ($f \leq 10$ Hz), typically associated with the strong suction peak dynamics, dominate near the leading edge. At the onset of separation, a sharp frequency band centered at 100 Hz emerges from the spectrum. The frequency band width increases and its center increases with the angle of attack, as it reaches around 50 Hz at 20° . A Strouhal number based on the freestream velocity U_∞ and the shear layer height h determined from the chord separation point and AoA yields a nearly constant value of $S_t = \frac{fh}{U_\infty} \sim 0.26$, which is consistent with the literature on 2D bluff bodies (Williamson, 1996).

2.4 Wake characterization

This section describes the main features of the velocity field, referred to hereafter as the "ground truth". Throughout the paper, the focus of the reconstruction will be the instantaneous streamwise velocity component, although a similar approach could be carried out for the cross-stream component. Time-averaged then turbulent features are presented below.

2.4.1 Time-averaged wake

To characterize the mean features of the wake velocity deficit, defined as $\Delta u(x, y) = (U_\infty - \bar{u}(x, y))$, where $\bar{\cdot}$ represents a time average, the following two quantities are introduced:

- The wake width δ is defined as the spanwise distance at which the time-averaged velocity deficit $\Delta u(x, y)$ reaches a fraction θ of the local maximum velocity deficit, such that $\Delta u(x, y = \delta) = \theta \Delta u_{\max}(x)$. Although a common criterion in wind turbine wake modeling is to use $\theta = 0.1$, a more conservative threshold $\theta = 0.5$ was used as it was found to provide a more robust estimate with respect to small-scale perturbations in the surrounding flow (Duda et al., 2021).
- The wake deflection angle α is defined as the angle between the line corresponding to the locations of the maximum velocity deficit and the chord. It is computed as $\alpha = \arctan(\frac{y_{\max} - y_0}{x_{\max} - x_0})$, where $(x_{\max}, y_{\max})$ are the coordinates of the maximum velocity deficit and (x_0, y_0) the coordinates of the trailing edge.

Fig. 3 depicts the streamwise evolution of the wake width and the wake deflection angle for all angles of attack. Concerning the wake width, a general trend observed is that both δ and $d\delta/dx$ increase with the angle of attack. For separated flow regimes, a specific trailing edge effect can be identified in the near wake ($x/c \in [1; 2]$), as the width $\delta(x)$ first goes through a local minimum then a local maximum, and $d\delta/dx$ jumps to higher values. At each angle of attack, the far wake evolution in the streamwise direction ($x > 4c$) can be characterized by a nearly constant growth rate, which increases with the angle of attack. Figure 3 also shows that the deflection angle α increases with the angle of attack, with large variations in the near wake, and a quasi-linear increase in the far wake (it should be noted that the small oscillations observed there are an artefact of the computational mesh). The existence of a monotonic relationship thus makes it possible to infer the angle of attack from the deflection angle α . The noisy behavior observed in the far wake profiles of both δ and α can be attributed to the mesh coarsening away from the airfoil.

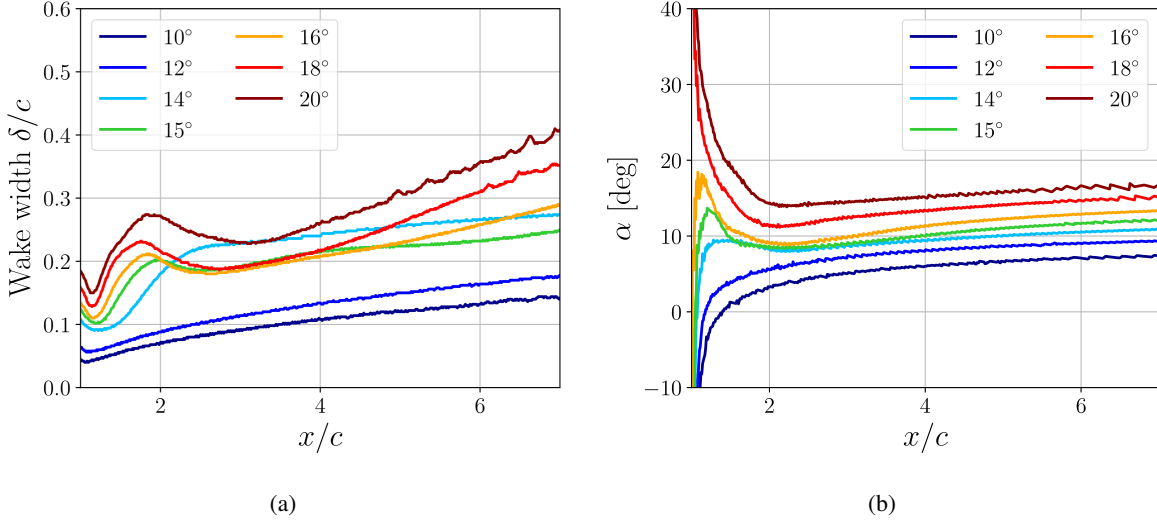


Figure 3: Streamwise evolution of the time-averaged wake width $\delta(x)/c$ (a) and streamwise evolution of the time-averaged wake deflection angle α (b). The profiles are shown for $x/c \in [1; 7]$, where $x/c = 1$ corresponds to the streamwise location of the trailing edge

2.4.2 Wake flow fluctuations

Since the reconstruction method aims to recover instantaneous velocity fields, attention is now turned to the time-dependent features of the wake. Snapshots of the instantaneous streamwise velocity field at an arbitrary time $t = 5s$ are shown in the top row of Fig. 4 at the same angles of attack as those in Fig. 2. It is recalled that all fields shown in the paper are rotated by placing the airfoil at a 0° incidence and aligning the flow at the correct angle with respect to the AoA). The vertical white dashed line at $x/c = 2$ (near wake) corresponds to the section where pre-multiplied velocity spectra are extracted and shown on the bottom row of Fig. 4.

As observed for the pressure field, the total energy of the fluctuations increase with the angle of attack. The dominant frequencies observed in the pressure spectra in Fig. 2 are also present in the pre-multiplied velocity spectra fS_u (100 and 50 Hz for AoA = 14° and 20°) and can be found to correspond to vortex shedding dynamics. However several differences between pressure and velocity spectra can be noted, thus reflecting the nonlinear relationship between the pressure and velocity that the model needs to handle.

In this section, we have characterized the inputs and outputs of the model to be described later, namely the surface pressure and the streamwise velocity field. Both time-averaged and time-dependent flow properties, such as the mean wake and its spectral signature, will be used to assess the model's performance.

3 Methodology

We now detail the estimation method for reconstructing the streamwise velocity field from a limited number of pressure measurements. We first describe how the sensor number and location are defined.

3.1 Sensor placement and number

Owing to cost and practical constraints, the number of sensors needs to be kept low in real-life applications. The sensor location is then critical because the information content is not uniformly distributed over the blade surface : regions such as the leading edge, trailing edge, or the intermittent separation zone carry substantially more information about the surrounding flow.

Much effort has been devoted to sensor placement optimization (see for instance Manohar et al., 2018; Hollenbeck et al., 2024; Bukowski et al., 2025; Carter et al., 2021). The approach of Manohar et al. (2018) was followed here. It consists in placing the sensors in order to maximize their contribution to the variance. However, the variations of the

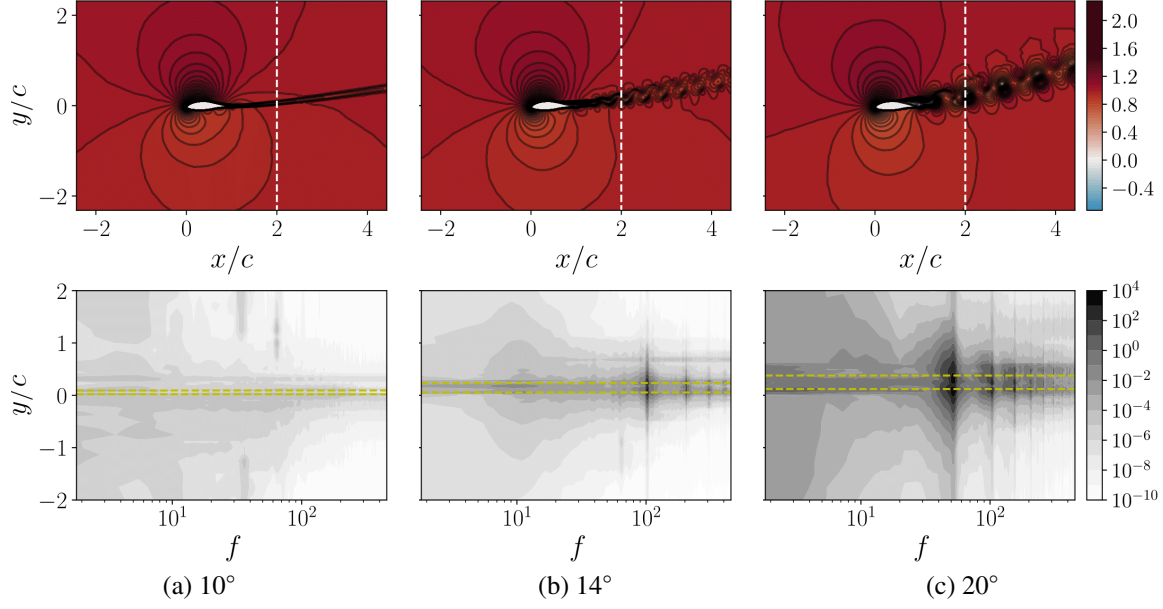


Figure 4: Snapshots of the normalized instantaneous streamwise velocity field u/U_∞ (top) and spectrograms of the pre-multiplied velocity spectra fS_u for the cases $\text{AoA} = 10^\circ$ (a), $\text{AoA} = 14^\circ$ (b), and $\text{AoA} = 20^\circ$ (c). The vertical white dashed lines in the snapshots denote the $x/c = 2$ section where the spectrograms are computed. The horizontal yellow dashed lines in the spectrograms indicate the wake width

variance with the angle of attack need to be taken into account, so that the information content associated with low fluctuations at the onset of separation does not get drowned out by the energetic fluctuations at higher angles of attack.

Algorithm Maximum variance approach

- 1: cluster blade points in n_c clusters based their positions.
 - 2: define a variance threshold σ_{thres}^2 to be represented by sensors.
 - 3: **for** each AoA_i **do**
 - 4: compute the total variance of C_p for each cluster
 - 5: keep first p_i most energetic clusters
 - 6: choose locations of variance maximum within these p_i clusters as sensor locations
 - 7: **if** selected locations are too close **do**
 - 8: move point of lesser variance up to a minimum distance $d_{\min}$
 - 9: **end if**
 - 10: **end for**
 - 11: concatenate all selected sensors
 - 12: **if** concatenated sensors are too close **do**
 - 13: keep point of higher variance, remove the other
 - 14: **end if**
-

Table 1: Methodology for selecting the wall pressure sensors based on maximizing the represented variance

The first step consists in clustering blade surface points into n_c clusters based on their positions. The k-means clustering method is used (MacQueen, 1967) and the number of clusters $n_c = 25$ is determined using the elbow method (Thorndike, 1953). In the second step the optimal sensor locations are determined for each angle of attack : the wall-pressure variance is computed within each cluster, and the clusters are ranked according to their relative contribution to the total variance at this angle of attack. Given a variance threshold σ_{thres}^2 (expressed as a percentage of the total variance) to be captured, the minimal number of clusters p_i necessary to capture this fraction is determined, and the corresponding clusters selected. In each of these p_i clusters, one point only is selected (the point of maximum variance) as sensor location, with the assumption that the fluctuations of the point are representative of the fluctuations of the cluster. If two candidate sensors are found to be separated by less than $d_{\min}$, the one associated with the highest variance is kept in

place, whereas the other is moved within its cluster at a distance d_{min} . The minimum distance d_{min} is taken to be $0.01c$ and roughly corresponds to the pressure sensor spacing of Neunaber et al. (2022)'s experiment. The third step consists in concatenating the lists of sensor positions obtained for the total number of angles of attack N_{AoA} . At this last stage, if two sensors are found to be at a distance less than d_{min} from each other, only the sensor associated with the highest variance - regardless of the angle of attack - is retained. The procedure yields a total number of $p \leq p_i N_{AoA}$ sensors. The sensor selection process is summarized in table 1.

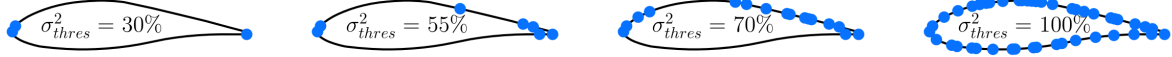


Figure 5: Sensor layouts selected with the variance-maximization strategy for different threshold levels

Table 2 reports the correspondence between the target variance threshold σ_{thres}^2 to be captured and the resulting number of sensors p required. The relationship between the variance threshold and the number of selected sensors is strongly nonlinear. At low threshold levels ($\leq 30\%$), the number of sensors remains nearly constant or increases only marginally. This can be attributed to the dominance of the Leading Edge Maximum and Trailing Edge Maximum, whose strong fluctuations alone account for up to $\sigma_{thres}^2 = 30\%$ of the total variance. As the threshold increases, additional sensor positions are created — first in the leading and trailing edge regions, then in the separation region, associated with the ISP, and finally covering the surface of the airfoil, as shown in Fig. 5. Based on the trade-off between reconstruction accuracy and the number of sensors, an optimal number of $p = 7$ sensors (hence representing 55% of wall pressure variance) was identified and will be used for most of the results presented below. This choice will be further detailed in Sect. 4.2.1.

σ_{thres}^2 (%)	10	20	30	40	50	60	70	80	90	100
p	2	2	3	4	6	10	16	23	31	44

Table 2: Number of sensors p as a function of the variance fraction σ_{thres}^2

3.2 The SNN-POD model

The SNN-POD model combines Proper Orthogonal Decomposition and the shallow decoder framework of Erichson et al. (2020) to reconstruct the full velocity field from sparse wall pressure measurements taken over a range of angles of attack. The model is trained with snapshots corresponding to the set I_{train} of angles of attack of size $N_{AoA} = 7$. For each angle of attack, a sequence of N_t snapshots is selected to constitute the training set, which therefore consists of $N_t \times N_{AoA}$ input/output couples, where each input is a pressure measurement vector of size p , and each output is constituted by the snapshot of the velocity field of size N_s with only one velocity component (streamwise). Since the m -th snapshot, where $1 \leq m \leq N_t N_{AoA}$, can be associated to a time t_k and an angle of attack AoA_i , it will be referred to indifferently with the index m or the tuple (k, i) . The acquisition time resolution between snapshots was taken to be 0.05 convective time units (about 1.1 ms). We used $N_t = 150$ snapshots, so that the length of the training set is about 0.16s, which corresponds to 8 – 16 vortex shedding cycles (defined using the vortex shedding frequencies observed in Fig. 4), depending on the angle of attack.

The approach can be summarized as follows (see Fig. 7) : POD is first used to compress information to a low-order manifold, a dimensionality reduction process which is presented in Sect. 3.2.1. A model based on shallow neural networks is then trained to learn the nonlinear mapping between the pressure measurements and the latent subspace (POD velocity amplitudes), the details of which are given in Sect. 3.2.2. The instantaneous flow field is then reconstructed from the temporal amplitudes estimated by the model and the spatial statistics of the training set.

3.2.1 Dimensionality reduction

Due to the high dimension of the full-order state to be reconstructed (see Sec. 2), each velocity field in the training set is first projected onto a low-dimensional linear subspace using Proper Orthogonal Decomposition (Lumley, 1967). Since $N_t N_{AoA} < N_s$ the method of snapshots (Sirovich, 1987) is applied to the velocity fluctuation defined with respect to a mean $\langle u \rangle$, where $\langle . \rangle$ refers to an average taken over all times and angles of attack. The method relies on the construction of a snapshot matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{u}^1 & \mathbf{u}^2 & \dots & \mathbf{u}^{N_t N_{AoA}} \\ | & | & & | \end{bmatrix}$$

where each column of $\mathbf{X} \in \mathbb{R}^{N_s \times N_t N_{AoA}}$ represents a velocity snapshot. It is to be noted that $\mathbf{u}^1$ and $u(\mathbf{x}, t_1, AoA_1)$ will be considered as equivalent notations.

Singular value decomposition (SVD) of the snapshot matrix $\mathbf{X}$ yields :

$$\mathbf{X} = \Phi \mathbf{S} \mathbf{A}^T$$

where

$$\Phi = \begin{bmatrix} \phi_1 & \dots & \phi_{N_t N_{AoA}} \end{bmatrix} \in \mathbb{R}^{N_s \times N_t N_{AoA}}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{N_t N_{AoA}} \end{bmatrix} \in \mathbb{R}^{N_t N_{AoA} \times N_t N_{AoA}},$$

$$\mathbf{S} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_{N_t N_{AoA}}^{1/2}) \in \mathbb{R}^{N_t N_{AoA} \times N_t N_{AoA}},$$

with $\mathbf{S}$ the diagonal matrix of singular values ranked by magnitude, Φ the matrix of linear POD modes and $\mathbf{A}$ the matrix of POD time coefficients. Each POD mode $\phi_n(\mathbf{x})$, $n = 1, \dots, N_t N_{AoA}$, satisfies the orthonormality condition as well each amplitude vector $\mathbf{a}_n$. The value λ_n represents the contribution of the n -th mode to the variance of the velocity field. Each streamwise velocity field of the training set can thus be expressed at any angle of attack AoA_i with a unique POD basis:

$$u(\mathbf{x}, t_k, AoA_i) = \langle u(\mathbf{x}) \rangle + u'(\mathbf{x}, t_k, AoA_i) \text{ where } u'(\mathbf{x}, t_k, AoA_i) = \sum_{n=1}^{N_t N_{AoA}} \lambda_n^{1/2} a_n^{k,i} \phi_n(\mathbf{x}) \quad (1)$$

Figure 6 shows the first six most energetic POD spatial modes $\phi_n(\mathbf{x})$, $n = 1, \dots, 6$ and their amplitudes a_n which are plotted separately for each angle of attack. The six spatial modes can be split into two groups : on the one hand, modes 1, 2, 3 and 5 are important in the shear layer and separation zone while on the other hand modes 4 and 6 are dominant in the wake. The energy level of the latter modes is similar, and their shape is consistent with a pair of vortex shedding modes. It can be further observed that the amplitudes of the modes in the first group are characterized by non-zero time-averaged values which are distinct for each angle of attack, which indicates that they correspond to time-independent field deformation due to the angle of attack. In contrast the time-average of the amplitudes in the second group is zero and the amplitudes are characterized by large oscillations which increase with separation. A criterion can thus be defined to determine whether a POD mode a_n is *steady* or *unsteady* by comparing its time average for each angle of attack $\bar{a}_n(AoA)$ with the amplitude of its variations. Specifically, a mode will be considered to be steady (resp. unsteady) if :

$$\max_{AoA} |\max_{N_t} \mathbf{a}_n - \min_{N_t} \mathbf{a}_n| > \max_{AoA} |\bar{\mathbf{a}}_n|. \quad (2)$$

This makes it possible to split POD modes into steady and unsteady modes, which will be indexed respectively with the suffixes *st* and *un*. A discussion on the relevance and robustness of this splitting is provided in Sec. 4.2.2.

3.2.2 Construction of the model

Given an input vector of pressure measurements of size p , the model learns the projection onto the POD basis by estimating the r most energetic POD coefficients $\hat{a}_i$, which are then combined to reconstruct the flow field of size N_s using :

$$\hat{u}(\mathbf{x}) = \langle u(\mathbf{x}) \rangle + \sum_{1 \leq n \leq r} \lambda_n^{1/2} \hat{a}_n \phi_n(\mathbf{x}) \quad (3)$$

where $\langle u(\mathbf{x}) \rangle$, λ_n and ϕ_n are determined from the training set. A preliminary analysis conducted on a single AoA configuration (not shown) indicated that increasing the number of POD modes generally enhances the reconstruction accuracy; however, the improvement becomes marginal for $r \geq 50$ retained modes. For this reason, the value of $r = 50$ POD modes, corresponding to a compression ratio of 96.75%, was chosen for the present study.

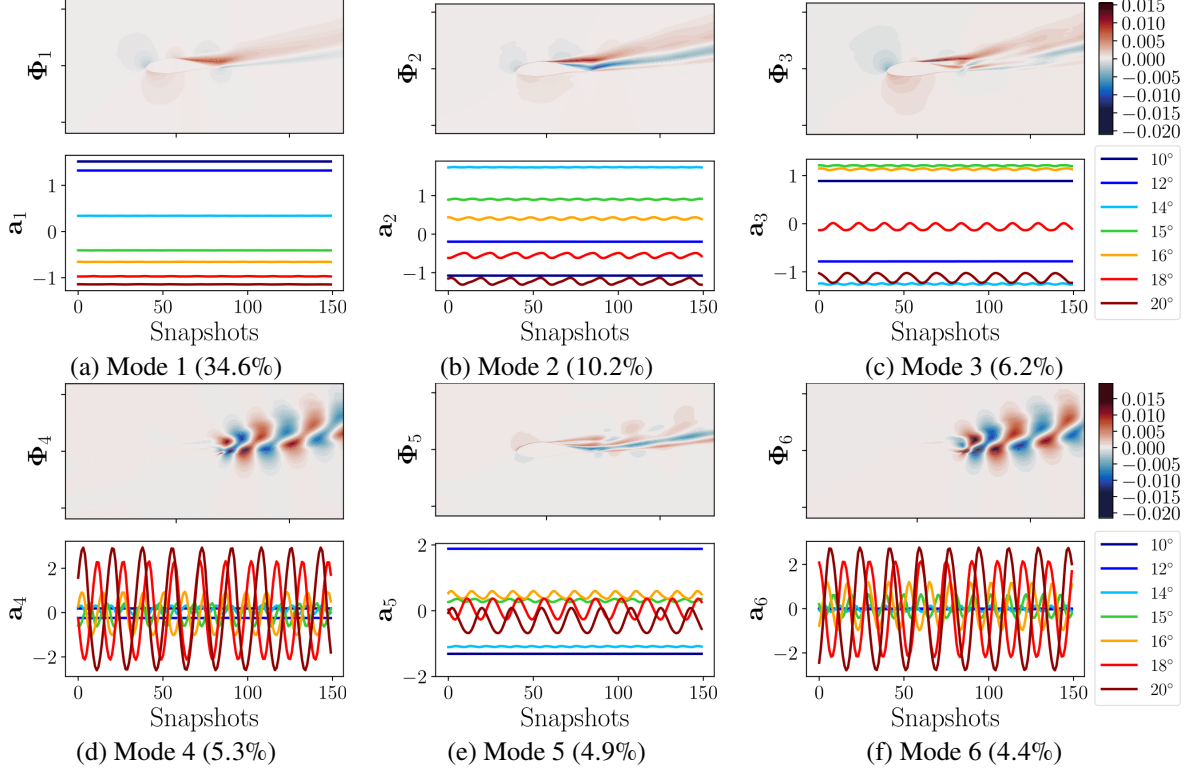


Figure 6: Sets of POD modes Φ_i (top) and their associated POD coefficients $\mathbf{a}_i$ (bottom) for $i = 1, \dots, 6$ (a,...,f). Their contribution to the total variance is denoted in parenthesis in captions

The amplitudes to be learnt $\hat{a}_n$ are split into two output vectors consisting of steady and unsteady modes—according to previous section—which are independently learnt by two networks based on the shallow decoder architecture outlined in Erichson et al. (2020), which we first briefly review. A Shallow Neural Network (SNN) is a neural network consisting of L fully-connected hidden layers (with $L \leq 3$ in general). Given an input $\mathbf{s}$, the output of the model is defined as

$$\mathcal{F}(\mathbf{s}; \mathbf{w}) = \mathbf{w}^L g(\mathbf{w}^{L-1} \dots g(\mathbf{w}^1 \mathbf{s})),$$

where g is a non-linear activation function and $\mathbf{w}$ the set of weights matrices. Here $L = 3$ and g is the ReLu function. The model used in the present study relies on two SNNs with identical architectures. The robustness of each network is improved by the addition of a dropout with a drop ratio of 10%. An Adaptive Moment Estimation (ADAM) optimizer is used along with a regularization method (weight decay) to prevent overfitting. Training batches are shuffled and mini-batch size is set to 128. Both networks are trained over 4000 epochs but training may be shortened by early stopping that acts as a regularizer to avoid overfitting (Goodfellow et al., 2016). In practice early stopping is applied if the error on the testing dataset does not reduce over 100 epochs. The choice of the optimal set of hyperparameters (i.e. parameters external to the learning process which are directly learnt within estimators) is determined by using Optuna hyperparameter optimization software framework based on Bayesian optimization (Akiba et al., 2019). The list of optimized hyperparameters and their associated values is given in Table 3.

Learning rate	Learning rate change	Weight decay	Weight decay change	Dropout probability
1.5×10^{-2}	7×10^{-5}	0.8	0.7	0.1
Input layer size	Hidden layer size 1	Hidden layer size 2	Output layer size	
p	r	$1.4r$	r	

Table 3: List of the tuned hyperparameters and their optimal values

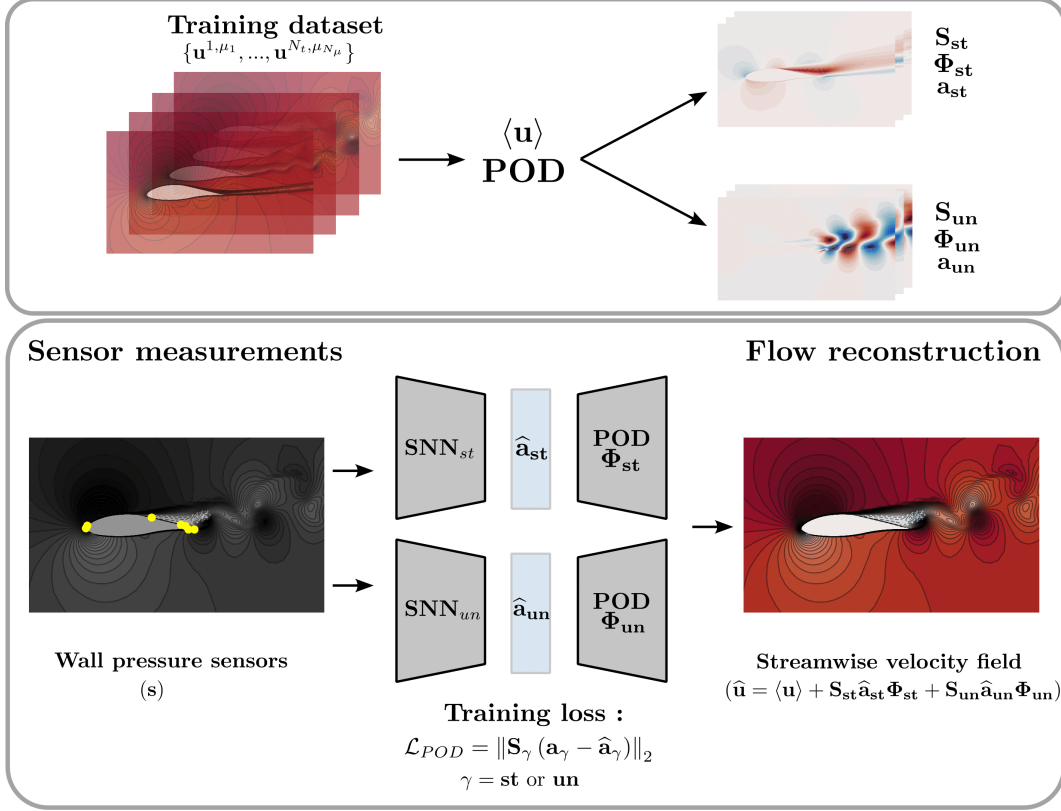


Figure 7: Schematic representation of the **SNN-POD** reconstruction framework. Steady (denoted with the *st* subscript) and unsteady (denoted with the *un* subscript) modes resulting from the Proper Orthogonal Decomposition of the training dataset are treated separately by the **SNN-POD**. Two Shallow Neural Networks (SNN) learn the mapping between the sparse wall pressure measurements and the steady and unsteady POD coefficients using a energy-weighted loss. The high-dimensional flow field is then recovered from the POD training basis Φ and estimated model amplitudes

Each SNN is trained to minimize the L_2 -norm of the error between the reduced state and its estimation, weighted by the associated eigenvalue (contribution of the mode to the variance), so that the loss is given by :

$$\mathcal{L}_{POD} = \|\mathbf{S} [\mathbf{A}^k - \mathcal{F}(\mathbf{s}^k; \mathbf{w})]\|_2 = \sum_{1 \leq n \leq r} \lambda_n \sum_{1 \leq m \leq N_t N_{AoA}} \|a_n^m - \hat{a}_n^m\|^2 \quad (4)$$

The singular values $\mathbf{S}$ included in the loss function are not learned by the model, but are directly extracted from the POD decomposition of the training dataset. In a sense, this formulation can be interpreted as a physics-based loss where the contribution of each coefficient to the loss represents its contribution to the total variance.

The test set consists of unseen snapshots for angles of attack both within and outside I_{train} . Specifically, three new angles were considered: 15.5° , 17° , 19° so the total number of angles in the test set is $N_{AoA}^{test} = 10$. For each $AoA \in I_{test}$ a sequence of $N_t^{test} = 50$ snapshots was selected to constitute the test set presented in the paper. Results did not change significantly for larger numbers or random selections of snapshots.

To evaluate the model accuracy on the test set, the following metrics were used:

- the normalized mean-square error (MSE) E_{test} , given by :

$$E_{test} = \frac{1}{N_{AoA}^{test} N_t^{test}} \sum_{\substack{1 \leq i \leq N_{AoA}^{test} \\ 1 \leq k \leq N_t^{test}}} \frac{\|\mathbf{u}^{k,i} - \hat{\mathbf{u}}^{k,i}\|_2}{\|\mathbf{u}^{k,i}\|_2} \quad (5)$$

- the fluctuation error ε_{test} , based on the fluctuating part of the velocity field u' (with respect to the average taken over the training set), defined as

$$\varepsilon_{test} = \frac{1}{N_{AoA}^{test} N_t^{test}} \sum_{\substack{1 \leq i \leq N_{AoA}^{test} \\ 1 \leq k \leq N_t^{test}}} \frac{\|\mathbf{u}'^{k,i} - \widehat{\mathbf{u}}'^{k,i}\|_2}{\|\mathbf{u}'^{k,i}\|_2} \quad (6)$$

By construction, the fluctuation error is larger than the MSE. It is also independent from the fluctuation to mean ratio, which can bias the MSE.

- the projection error ε_{test}^Φ measures the relative difference between the reconstruction and the projection of the ground truth velocity field onto the POD basis of r modes:

$$\varepsilon_{test}^\Phi = \frac{1}{N_{AoA}^{test} N_t^{test}} \sum_{\substack{1 \leq i \leq N_{AoA}^{test} \\ 1 \leq k \leq N_t^{test}}} \frac{\|\mathbf{u}'_\Phi^{k,i} - \widehat{\mathbf{u}}'^{k,i}\|_2}{\|\mathbf{u}'_\Phi^{k,i}\|_2}. \quad (7)$$

where $\mathbf{u}'_\Phi^{k,i} = \sum_{1 \leq n \leq r} a_n^{k,i} \Phi_n$ is the projection of the ground truth field onto the POD basis of the training set, which represents the best approximation achievable by the SNN-POD. Comparing the fluctuation and the projection error therefore provides a measure of the suitability of the POD training basis.

4 Results

In this section, the performance of the SNN-POD model trained on the full range of angles of attack is evaluated on the test set. The robustness of the model is then examined through a sensitivity analysis (Sec. 4.2.1) of the sensor placement strategy described in Sec. 3.1, followed by a study of the influence of the modal splitting approach (Sec. 4.2.2) on the model performance. Finally, Sec. 4.3 describes how combining SNN-POD models trained on different subsets of angles of attack can locally improve the reconstruction.

4.1 Evaluation of SNN-POD

This section evaluates the ability of the SNN-POD model to estimate both the POD coefficients and the instantaneous velocity field u from the URANS simulations, denoted thereafter as the “ground truth”. As mentioned earlier, the test set includes angles sampled during training as well as angles which are not part of the original discretization (interpolated angles). Figs. 8-11 compare characteristic features of the true and reconstructed flow fields for three sampled (10° , 15° , 20°) and three interpolated angles (15.5° , 17° , 19°). Each column corresponds to an angle of attack. Figs. 8 and 10 provide a comparison of a ground truth snapshot (top row) with its reconstruction (bottom row) for sampled and interpolated angles, respectively. In all cases a very good agreement is observed for the main features of the flow, although some slight differences are noticeable for the interpolated angles, particularly in the far wake.

A more quantitative assessment can be made by examining the top rows of Figs. 9 and 11, which show the average wake width and deflection angles. The wake width is well recovered for the sampled angles (fig. 9). For the interpolated angles (fig. 11), the wake width is characterized by small oscillations for $x/c > 1.5$ and tends to be overpredicted for the two higher angles, although the growth rate $d\delta/dx$ generally appears to be correct. In contrast, the deflection angle is well recovered for both sampled and interpolated angles, which shows that the model is able to identify the angle of attack of the configuration from a single set of pressure measurements. This result is particularly important in the context of wind energy, where the estimation of instantaneous wind loads is directly linked to the determination of the AoA.

The bottom lines of Figs. 9 and 11 compare the time spectra of the true and estimated velocity fields for sampled and interpolated angles, respectively. For all AoAs, the energy levels are accurately predicted, and the characteristic frequencies associated with flow separation are also well captured. The largest discrepancies in the spectrograms are observed for the attached flow case (10°), where wake turbulence is low in energy. The model is able to predict the frequency shift observed for both sampled and interpolated angles and reproduces the complexity of the spectra, although unsurprisingly the agreement is closer for sampled angles.

A more direct evaluation of the model performance is provided by figure 12, which compares the estimated POD coefficients with their true values at all angles. Again, snapshots have been organized by increasing angle of attack and in chronological order so that the dynamics of the reconstruction can be evaluated, although no time dependence

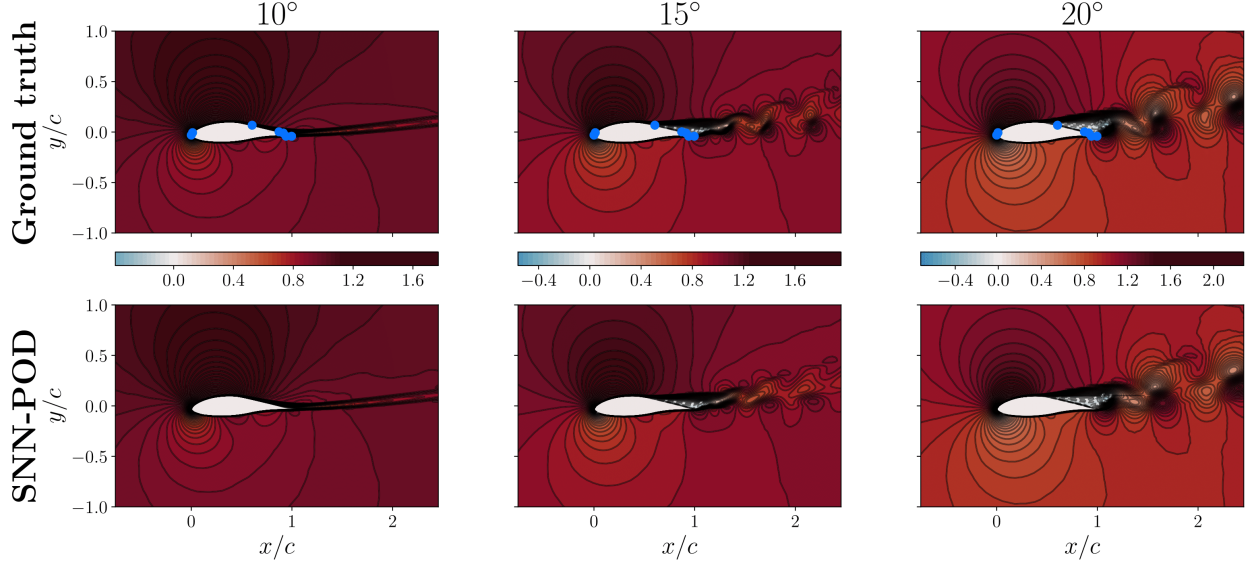


Figure 8: Comparison of streamwise velocity $u(\mathbf{x}, t_s)/U_\infty$ contours at snapshot $t_s = 150s$ of the ground truth obtained from URANS (left) and its SNN-POD estimation (right) for three AoA : 10° (left), 15° (center) and 20° (right). The locations of the $p = 7$ pressure sensors on the blade are indicated with blue markers

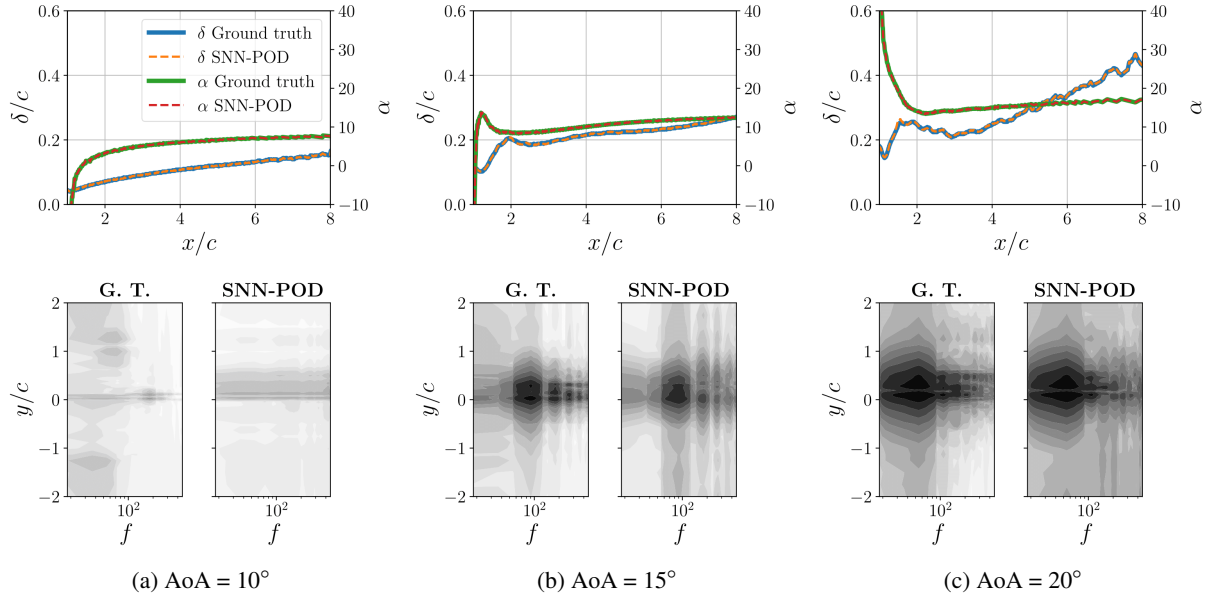


Figure 9: Top plots denote the wake widths and wake deflection angles comparison between the ground truth and the estimation for three AoA : 10° (left), 15° (center) and 20° (right). Bottom plots correspond to the spectrograms of the pre-multiplied velocity spectra fS_u of the ground truth (G.T., left) and estimation (SNN-POD, right) at the $x/c = 2$ section. The same logarithmic colormap is applied for both spectrograms

is accounted in the model. Interpolated angles are indicated by green shaded regions. As noted in Fig. 6, strong non-monotonic variations are observed in the amplitudes of the first three modes around AoA = 14° , while the evolution appears to be more gradual for AoA $\geq 15^\circ$. The time-averaged values of the amplitudes appear to be correctly predicted for all angles of attack, with the largest discrepancy observed for mode 3 at AoA = 15.5° (note that the true value does not correspond to a direct interpolation of the angles AoA = 15° and AoA = 16° , a trend which the model captures but overpredicts).

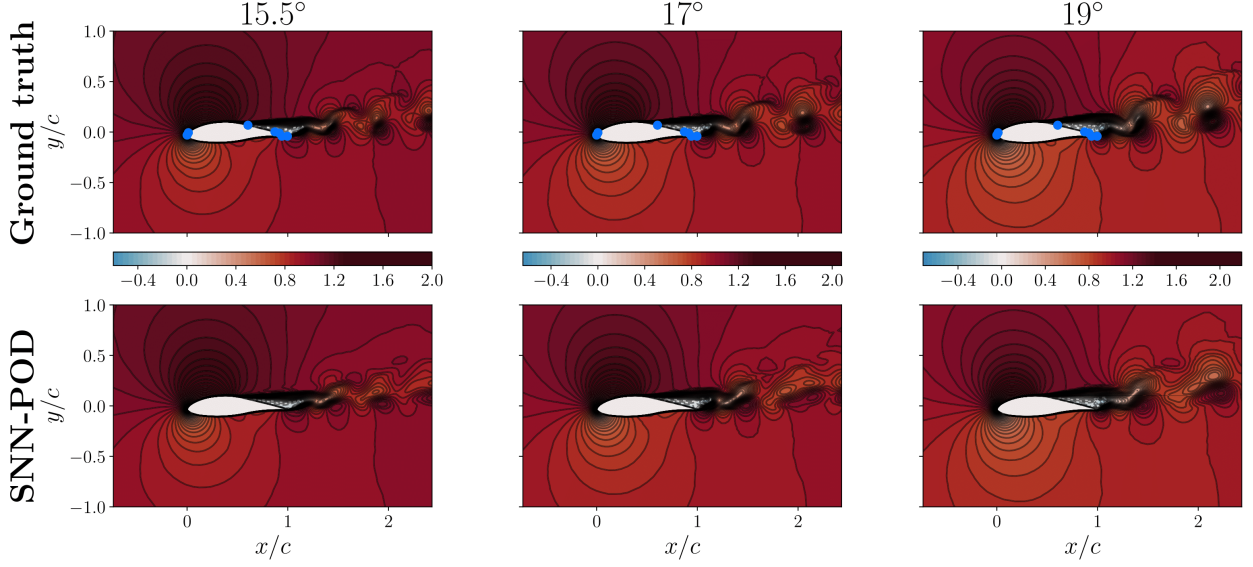


Figure 10: Comparison of streamwise velocity $u(\mathbf{x}, t_s)$ contours at snapshot $t_s = 150s$ of the ground truth obtained from URANS (first plot from left to right) and its SNN-POD estimation (second plot) for three interpolated AoA : 15.5° (left), 17° (center) and 19° (right). The locations of the $p = 7$ pressure sensors on the blade are indicated with blue markers

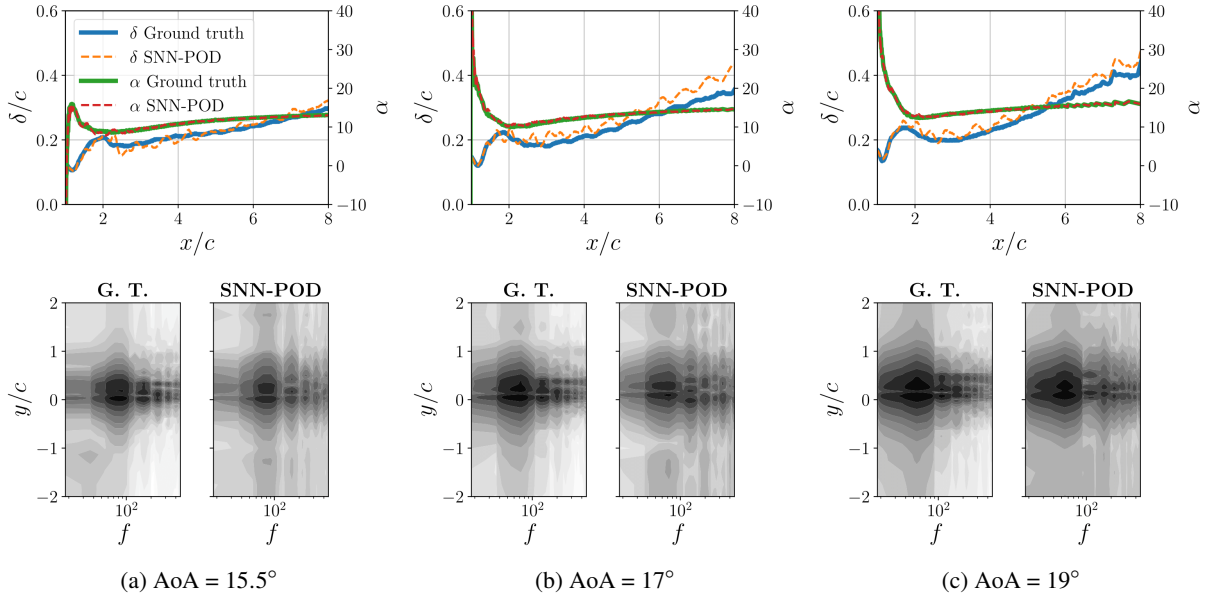


Figure 11: Wake widths and wake deflection angles comparison between the ground truth and the estimation (top) for three unseen AoA : 15.5° (left), 17° (center) and 19° (right). Bottom plots correspond to the spectrograms of the pre-multiplied velocity spectra fS_u of the ground truth (G.T., left) and estimation (SNN-POD, right) at the $x/c = 2$ section. The same logarithmic colormap is applied for both spectrograms

The mean, fluctuation and projected fluctuation errors defined in section 3.2.2 are computed individually for each angle of attack of the test dataset and shown in Fig. 13. As expected, errors are higher for interpolated AoA than for sampled AoA. Nonetheless, the mean reconstruction error remains below 6.6% (maximum observed for 17°), while the largest fluctuation errors $\varepsilon_{\text{test}}$ are about 30% for 15.5° and 17° , and reach about 23% for 19° . The projection error is essentially equal to the fluctuation error for sampled angles but about 25% lower for interpolated angles, a fraction that tends to decrease as the AoA increases (25% for 15.5° and 17° , and under 20% for 19°). Overall, this points to a moderate influence of the POD basis on the model error. It can be noted that the projection error is largest for 15.5° , despite the

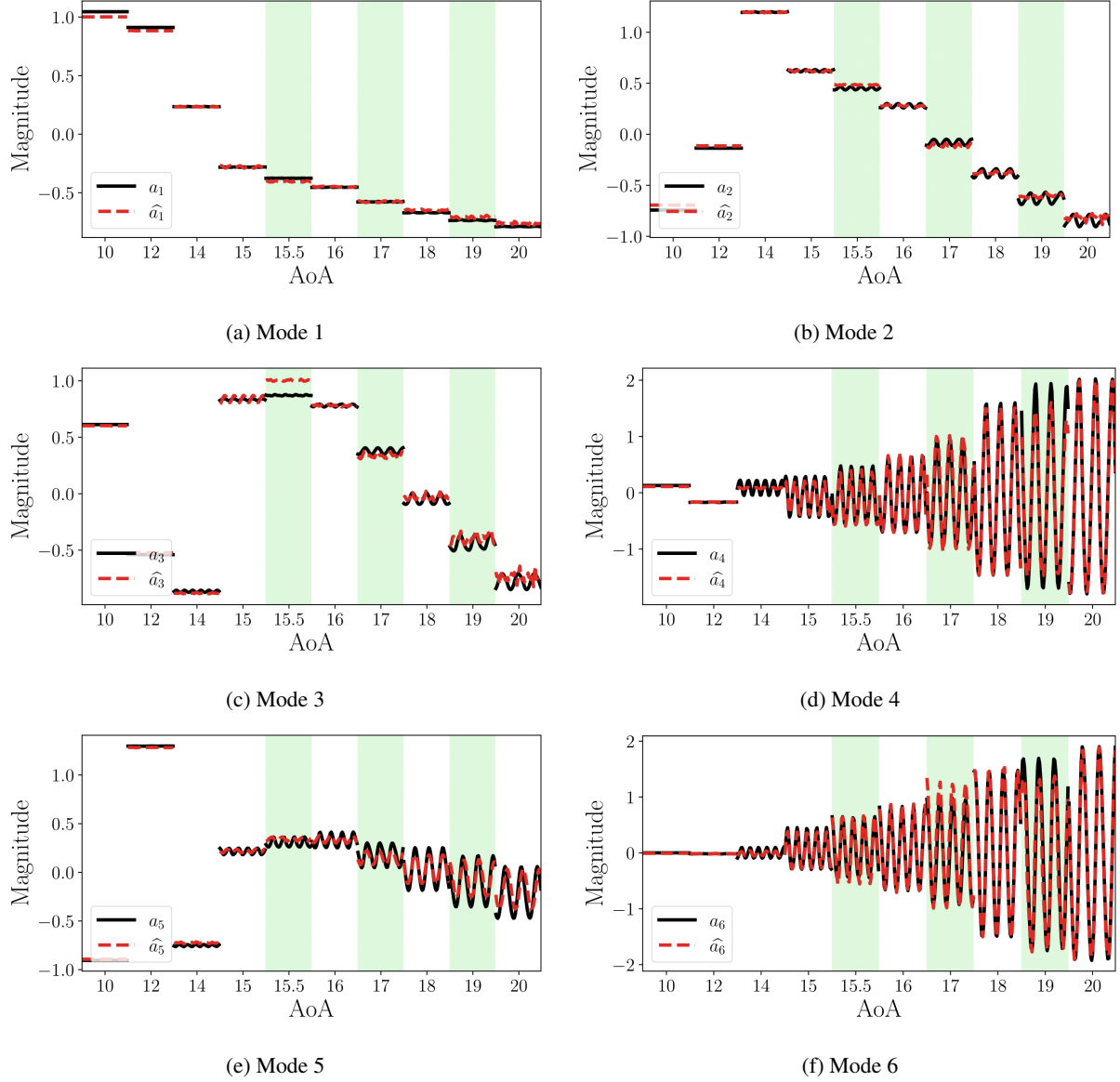


Figure 12: Comparison between the true POD amplitudes a_i (black lines) and the model estimates $\hat{a}_i$ (SNN-POD estimation red dashed lines) on the test dataset, with $i = 1, 2, 3, 4, 5, 6$ (a,b,c,d,e,f respectively). The test snapshots are organized into chronological time and sorted by angles of attack-angles of attack seen during training are displayed on white background, those not seen are highlighted with light green backgrounds

finer AoA resolution in that range. This suggests that significant changes occur in this range of angle of attack, which is also consistent with the higher errors observed at the angles 14° , 15° , and to a lesser extent 16° . This point will be further discussed in Sec. 4.3.

4.2 Robustness of the model

4.2.1 Evaluation of the sensor placement strategy

In order to assess the influence of sensor placement, a comparative analysis involving three sensor selection approaches was carried out: (i) the variance-maximization strategy described in Sec. 3.1 (ii) a fully random placement where all sensors are placed randomly, (iii) a clustered random placement where preliminary clustering is first performed and then one sensor location is drawn with a uniform probability from each cluster, which ensures a more uniform

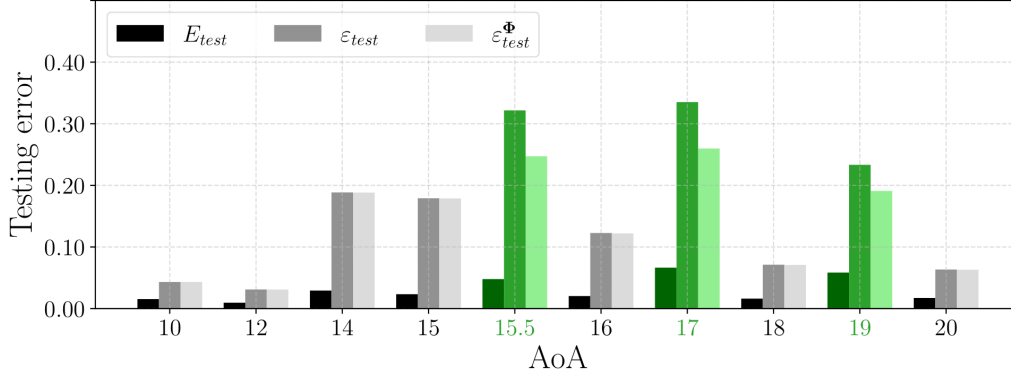


Figure 13: Testing mean error E_{test} , fluctuating error ε_{test} and projected fluctuating error $\varepsilon_{test}^{\Phi}$ on testing datasets for each AoA configuration. The three interpolated AoA $\{15.5^\circ; 17^\circ; 19^\circ\}$ are highlighted in green

sensor distribution than the full random approach. Evaluation on the testing set defined above was performed over 30 realizations of the model, which were trained on the same training set with different initial conditions. In all cases a minimum distance d_{min} was imposed between sensors ($d_{min} = c/100 = 1.25 \times 10^{-2}$ m as used in the maximum variance approach). Fig. 14 shows the evolution of the fluctuation error with the number of sensors for the three approaches. Unsurprisingly, as the number of sensors p increase, the reconstruction error decreases towards a value that is the same for all strategies (within statistical uncertainty). However, when the number of sensors is small ($3 \leq p \leq 15$), the reconstruction error is lower for the maximum variance strategy, with significantly less variability. This validates the choice of $p = 7$ sensors made in the study.

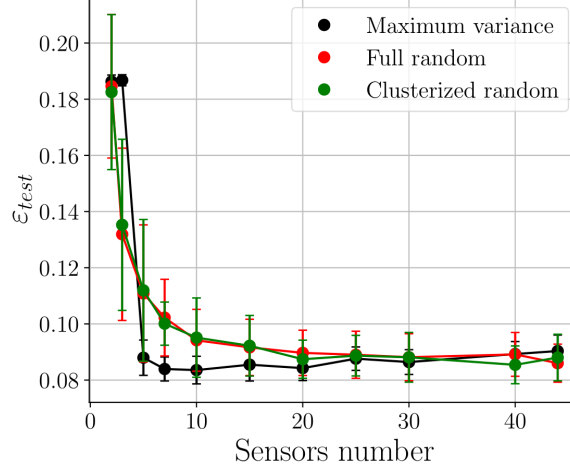


Figure 14: Sensitivity study of the pressure sensors number

4.2.2 Modal splitting sensitivity study

In this section the robustness of the model with respect to the splitting strategy is evaluated. Different architectures based on the shallow neural network framework were compared in terms of performance and of training cost. For each architecture, 30 models were trained and evaluated, and average values are reported in Table 4.

The following models were considered :

- the SNN-POD model described Sec. 3, based on two independent networks for the steady and unsteady component of the POD amplitudes

- the SNN model which directly learns the mapping between the pressure measurements and the velocity field (no latent space reduction)
- the SNN-POD-0 model which learns the mapping in POD space with a single network (all POD amplitudes are mixed together)
- the SNN-POD-50 model which learns the mapping in POD space with a network for each POD mode.

The SNN has both a high training cost and a high reconstruction error, which illustrates the importance of learning latent features. The SNN-POD-50 has the lowest reconstruction error but also the highest training cost, which is due to the high number of networks. Unsurprisingly, the lowest training cost is observed for the SNN-POD-0, but it also has the highest reconstruction error.

	SNN-POD	SNN	SNN-POD-0	SNN-POD-50
Number of SNNs trained	2	1	1	50
E_{test}	2.0×10^{-2}	3.2×10^{-2}	3.9×10^{-2}	1.9×10^{-2}
ε_{test}	8.5×10^{-2}	1.4×10^{-1}	1.6×10^{-1}	8.0×10^{-2}
CPU training time (s)	330	3530	220	6000

Table 4: Average training cost and MSE for different SNN-based architectures. See text for details

The sensitivity of the model to the modal splitting was next examined. Different choices for the two-mode split were evaluated, with comparison to the no-split case SNN-POD-0, denoted as $\{\}$ in Fig. 15. Six mode-splitting configurations were investigated, where one to ten modes are selected to be reconstructed by one shallow neural network, while the remainder of the 50 POD modes is estimated by another. In the first four configurations, ($\{1\}$, $\{1; 2\}$, $\{1; 2; 3\}$, $\{1; 2; 3; 5\}$), the output of the first network consisted of an increasing number of modes that are all considered steady according to criterion Eq.(2). In contrast, for the other two configurations $\{1 - 6\}$ and $\{1 - 10\}$, the output of the first network respectively consisted of the first six and ten most energetic modes (energy-based splitting). This means that the first network reconstructs both steady and unsteady modes.

Fig. 15 represents the fluctuation error for each configuration on the portion of the test dataset described in section 4 limited to sampled angles only. It is observed that the strategy used to separate steady and unsteady modes has a significant impact on the global reconstruction error, with a drop of nearly 40% if at least one steady mode is trained independently from the others. For these configurations, the error varies between 8.5% and 9.1%, so that it depends only weakly on the exact splitting. In contrast, the performance drops to a level comparable to that of the no-split model (SNN-POD-0, $\{\}$) when training SNN_{st} on unsteady modes, as it is the case for $\{1 - 6\}$ and $\{1 - 10\}$. These results illustrate both the importance and the robustness of the mode-splitting strategy.

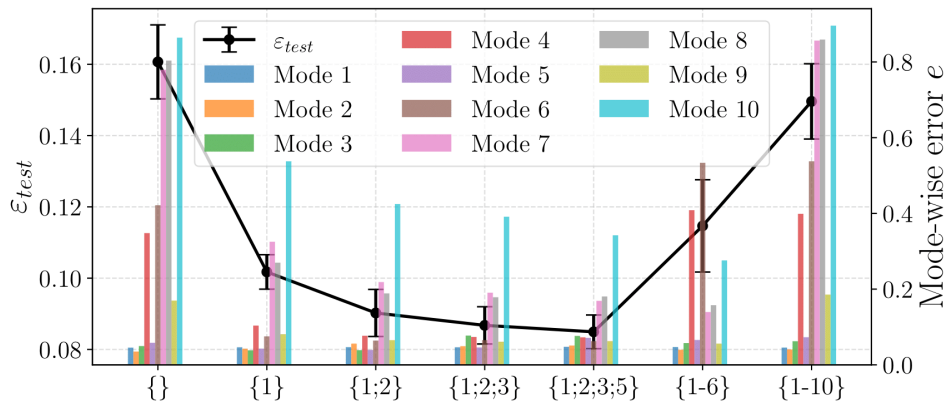


Figure 15: Reconstruction error ε_{test} and mode-wise reconstruction errors e of the first 10 POD modes for different scale separations. Statistics are computed over 30 independent training runs

Fig. 15 also shows the mode reconstruction error for each of the first ten POD modes, where the reconstruction error e_i for the i -th POD mode is defined as :

$$e_i = \frac{\|a_i - \hat{a}_i\|_2}{\|a_i\|_2} \quad (8)$$

Examination of the errors e_i reveals that the steady modes (1,2,3,5) are well recovered in all cases. In view of the loss formulation in Eq. (4), it is expected that the error increases with the POD mode index. Including mode 4 (which is unsteady, as can be seen in figure 6) with the steady modes significantly degrades its reconstruction, particularly evident in the $\{1 - 6\}$ configuration where the first six modes are learnt together. This further confirms the relevance of the steadiness-based splitting.

4.3 Combining SNN-POD models

As reported in section 4.1, the model globally trained over the full range of angles of attack is generally able to recover the time-averaged features of the flow such as the wake deflection angle (and therefore the angle of attack), while flow reconstruction of the time-dependent dynamics can be more challenging, particularly at the onset of separation where fluctuations are relatively small. A solution to improve the reconstruction could therefore be to combine the global model with local models that would be trained for specific flow regimes. The global model would first be used to infer the steady part of the flow and in particular the configuration angle of attack, based on which a model specifically trained for this flow regime could be selected.

Combined model (SNN-POD-C)

- 1: from pressure input evaluate $\hat{\mathbf{a}}_{\text{st}}$ with global model (steady global network)
 - 2: compute steady flow field $\hat{\mathbf{u}}_{\text{st}} = \hat{\mathbf{a}}_{\text{st}} \Phi_{\text{st}}$
 - 3: compute deflection angle α for each reconstructed steady flow field
 - 4: **if** $\alpha(x/c = 4) \in [\alpha_1; \alpha_2]$ (or other suitable criterion)
 - 5: evaluate $\hat{\mathbf{a}}_{\text{st}}^{\text{I}}$ and $\hat{\mathbf{a}}_{\text{un}}^{\text{I}}$ with local model
 - 6: reconstruct $\hat{\mathbf{u}} = \langle \mathbf{u} \rangle^{\text{I}} + \mathbf{S}_{\text{st}}^{\text{I}} \hat{\mathbf{a}}_{\text{st}}^{\text{I}} \Phi_{\text{st}}^{\text{I}} + \mathbf{S}_{\text{un}}^{\text{I}} \hat{\mathbf{a}}_{\text{un}}^{\text{I}} \Phi_{\text{un}}^{\text{I}}$
 - 7: **else** evaluate $\hat{\mathbf{a}}_{\text{un}}$ with global model (unsteady global network)
 - 8: reconstruct $\hat{\mathbf{u}} = \langle \mathbf{u} \rangle + \mathbf{S}_{\text{st}} \hat{\mathbf{a}}_{\text{st}} \Phi_{\text{st}} + \mathbf{S}_{\text{un}} \hat{\mathbf{a}}_{\text{un}} \Phi_{\text{un}}$
 - 8 : **end if**
-

Table 5: Combined model SNN-POD-C based on global and local models

To test this idea, a local model was developed by training a SNN-POD model on the angles $14^\circ, 15^\circ, 16^\circ$. It is emphasized that the POD basis and mean field are generally different for the global and local models since they are associated with different training sets - the suffix **I** will be used for the local model. The approach is summarized in Tab. 5. The combination of the global and the local models, referred to as SNN-POD-C, was then evaluated for the testing dataset in the range $[14^\circ, 16^\circ]$.

Figure 16 compares the SNN-POD and SNN-POD-C reconstructions with the ground truth at the challenging angle of AoA = 14° , corresponding to the onset of separation. Comparing the two left columns on each row shows that the instantaneous features of the wake are more accurately captured by the combined model. The mean characteristics of the wake (width and deflection angle), shown in the middle column, are equally well recovered by both models. Comparison of the spectrograms (two right-most columns) confirms that the SNN-POD-C model provides a more accurate estimate of the fluctuation level, which is underestimated by the SNN-POD model

Fig. 17 presents the three types of reconstruction errors for the standard SNN-POD (shaded areas) and the combined SNN-POD-C models (denoted by red hatched bars) for the range $[14^\circ, 16^\circ]$. All three errors are reduced by a factor of 2 for the angles $14^\circ, 15^\circ, 16^\circ$, which are seen during training. Interestingly, no reduction is observed for the interpolated angle of 15.5° , which shows that the gain associated with specific training does not necessarily lead to better generalization properties.

The main outcome of this coupling framework is that the global model demonstrates good performance in reconstructing instantaneous fields for fully attached and separated regimes, while its training over a wide AoA coverage enables interpolation to unseen configurations. Moreover, thanks to the mode-splitting, the global model can infer the AoA regime, allowing it to be coupled with local models that achieve higher accuracy on the transitional AoA regime.

5 Conclusion

This work describes an efficient data-driven methodology for the one-shot reconstruction of instantaneous velocity fields around a thick 2D airfoil from sparse wall-pressure measurements. The dataset is generated from 2D high-

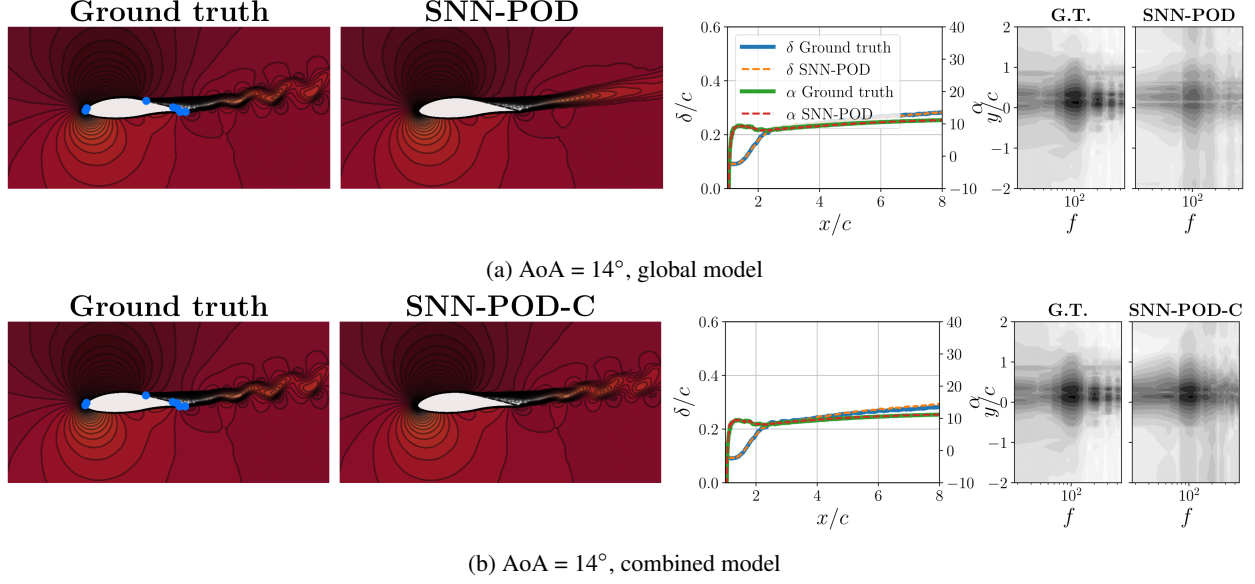


Figure 16: Side-by-side comparison of streamwise velocity $u(\mathbf{x}, t_s)$ contours at snapshot $t_s = 150$ s of the ground truth obtained from URANS (first plot from left to right) and its SNN-POD / SNN-POD-C estimation (second plot). The wake widths and wake deflection angles comparison between the ground truth and the estimation are plotted on the third subfigure. The fourth and fifth plots correspond to the spectrograms of the pre-multiplied velocity spectra $f S_u$ of the ground truth (G.T., left) and estimation (SNN-POD / SNN-POD-C, right) at the $x/c = 2$ section. The same logarithmic colormap is applied for both spectrograms

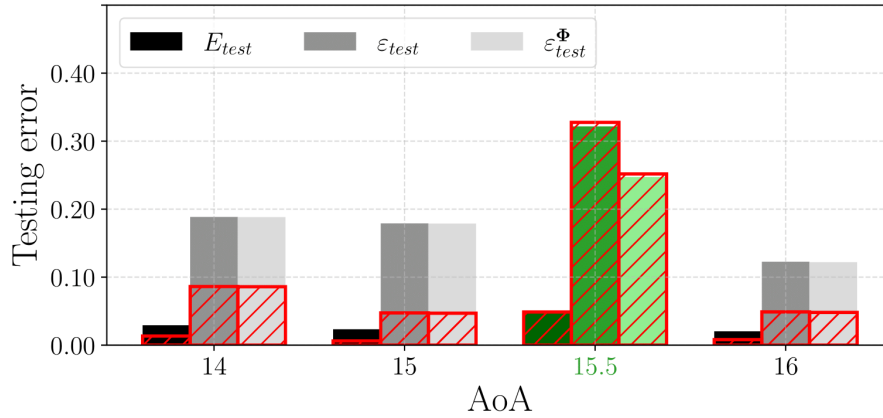


Figure 17: Mean error E_{test} , fluctuating error ϵ_{test} and projected fluctuating error ϵ_{test}^{Φ} for transitional AoA without (colored bars) and with (red hatched bars) the use of a local model. The interpolated AoA = 15.5° is highlighted in green

Reynolds-number URANS simulations over a range of angle of attacks between 10° and 20°, spanning a variety of flow regimes ranging from attached flow to full separation. Proper Orthogonal Decomposition provides a low-dimensional representation of the velocity field that can be split into steady modes, the amplitudes of which depend mostly on the angle of attack, and a remainder of unsteady modes. The model is constituted by two shallow neural networks that independently predict the unsteady and steady parts of the POD amplitudes from limited surface pressure measurements. A sensor placement strategy based on a maximization variance criterion was found to achieve better accuracy than randomized placements when the number of sensors is kept low. The model is trained on a coarse discretization of the angle of attack range (1 – 2°) in order to test its generalization abilities. The maximum reconstruction error (MSE) was found to be 2.9% for sampled angles and 6.6% for interpolated angles. In all cases the time-averaged flow features at a given angle of attack were well recovered from a single-time set of instantaneous pressure measurements.

Reconstruction of the time-dependent features could be improved by combining the model trained over the full range of angles of attack with one trained over a more local range. It is emphasized that this proposed strategy merely aims to highlight one of the multiple possibilities offered by the SNN-POD framework.

Several perspectives for extending the study can be identified. Firstly, in the present manuscript, only variations in AoA were considered to define different flow configurations; however, it would be relevant to investigate the performance of SNN-POD under changes in other parameters such as turbulence intensity or Reynolds number. Secondly, the current framework is limited to training on 2D URANS simulations, which cannot capture three-dimensional effects, which are particularly important in near flow separation conditions for thick airfoils. A natural extension is thus the application of SNN-POD to higher-fidelity 3D datasets for the reconstruction of fully unsteady 3D flow fields, ultimately making it a valuable tool for experimentalists. Lastly, the reconstruction methodology has so far been assessed only on static simulations with constant AoA, thereby neglecting the additional nonlinearities and hysteresis effects characteristic of dynamic stall. Extending the model to dynamic AoA variations would constitute a next step, enabling the evaluation of the model under more complex and unsteady operating conditions. To this end, more sophisticated approaches may be required, such as incorporating temporal history to capture coherent turbulent structures within the model.

Acknowledgments This work was granted access to the HPC resources of IDRIS under the allocation 2025-AD012A16542 made by GENCI.

Funding Statement This research was performed within the French-Swiss project MISTERY funded by the French National Research Agency (ANR PRCI grant no. 266157) and the Swiss National Science Foundation (grant no. 200021L 21271). The authors are thankful to the anonymous reviewers for their helpful suggestions.

Competing Interests The authors declare no competing interests exist.

Author Contributions Conceptualization: Q.B; B.P; C.B; E.G. Methodology: Q.B; B.P. Data curation: Q.B; C.B; E.G. Data visualization: Q.B. Writing - original draft: Q.B. Writing - review & editing: Q.B; B.P; C.B; E.G. All authors approved the final submitted draft.

Supplementary Material State whether any supplementary material intended for publication has been provided with the submission.

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Blackman, K. and Perret, L. (2016). Non-linear interactions in a boundary layer developing over an array of cubes using stochastic estimation. *Physics of Fluids*, 28(9):095108.
- Braud, C., Podvin, B., and Deparday, J. (2024). Study of the wall pressure variations on the stall inception of a thick cambered profile at high Reynolds number. *Physical Review Fluids*, 9(1):014605.
- Broeren, A. and Bragg, M. (2001). Spanwise variation in the unsteady stalling flowfields of two-dimensional airfoil models. In *39th Aerospace Sciences Meeting and Exhibit*.
- Bukowski, T. J., Rennie, M. R., and Gordeyev, S. (2025). Sparse sensing and surface pressure field reconstruction for flow over a hemisphere. In *AIAA SCITECH 2025 Forum*, page 1884.
- Callaham, J. L., Maeda, K., and Brunton, S. L. (2019). Robust flow reconstruction from limited measurements via sparse representation. *Physical Review Fluids*, 4(10):103907.
- Carter, D. W., De Voogt, F., Soares, R., and Ganapathisubramani, B. (2021). Data-driven sparse reconstruction of flow over a stalled aerofoil using experimental data. *Data-Centric Engineering*, 2:e5.
- Chaviaropoulos, P. and Hansen, M. (2000). Investigating Three-Dimensional and Rotational Effects on Wind Turbine Blades by Means of a Quasi3D Navier-Stokes Solver. *Journal of Fluids Engineering-transactions of The Asme - J FLUID ENG*, 122.
- Duda, D., Uruba, V., and Yanovych, V. (2021). Wake Width: Discussion of Several Methods How to Estimate It by Using Measured Experimental Data. *Energies*, 14(15):4712.
- Duthé, G., Abdallah, I., Barber, S., and Chatzi, E. (2023). Graph Neural Networks for Aerodynamic Flow Reconstruction from Sparse Sensing.

- Erichson, N. B., Mathelin, L., Yao, Z., Brunton, S. L., Mahoney, M. W., and Kutz, J. N. (2020). Shallow Neural Networks for Fluid Flow Reconstruction with Limited Sensors. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2238):20200097.
- Fukami, K. and Taira, K. (2025). Observable-augmented manifold learning for multi-source turbulent flow data. *Journal of Fluid Mechanics*, 1010:R4.
- Glauert, H. (1935). *Airplane Propellers*, pages 169–360. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- Graziani, A., Kerhervé, F., Martinuzzi, R., and Keirsbulck, L. (2018). Dynamics of the recirculating areas of a forward-facing step. *Exp. Fluids*, 59:154.
- Hollenbeck, A. C., Grandhi, R., Hansen, J. H., and Pankonien, A. M. (2024). Sparse flow sensor placement optimization for flight-by-feel control of 2d airfoils. *AIAA Journal*, 62(10):3803–3812.
- Hora, G. S., Gentine, P., Momen, M., and Giometto, M. G. (2024). Physics-informed data-driven reconstruction of turbulent wall-bounded flows from planar measurements. *Physics of Fluids*, 36(11):115191.
- Kaimal, J. C. and Finnigan, J. J. (1994). *Atmospheric Boundary Layer Flows: Their Structure and Measurement*. Oxford University Press.
- Kalnay, E. (2002). Atmospheric Modeling, Data Assimilation and Predictability.
- Kutz, J. N., Reza, M., Faraji, F., and Knoll, A. (2024). Shallow Recurrent Decoder for Reduced Order Modeling of Plasma Dynamics.
- Lee, K. and Carlberg, K. (2019). Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders.
- Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., Azizzadenesheli, K., and Anandkumar, A. (2024). Physics-Informed Neural Operator for Learning Partial Differential Equations. *ACM / IMS J. Data Sci.*, 1(3):9:1–9:27.
- Loiseau, J. C., Noack, B. R., and Brunton, S. L. (2018). Sparse reduced-order modelling: sensor-based dynamics to full-state estimation. *J. Fluid Mech.*, 844:459–490.
- Lu, S. and Papadakis, G. (2023). Flow Reconstruction Around a Surface-Mounted Prism from Sparse Velocity and/or Scalar Measurements Using a Combination of POD and a Data-Driven Estimator. *Flow, Turbulence and Combustion*, 110(4):1059–1090.
- Lumley, J. (1967). The structure of inhomogeneous turbulent flows. In Iaglom, A. and Tatarski, V., editors, *Atmospheric Turbulence and Radio Wave Propagation*, pages 221–227. Nauka, Moscow.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In of California Press, U., editor, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, page pp. 281–297.
- Madsen, H. A., Larsen, T. J., Pirrung, G. R., Li, A., and Zahle, F. (2020). Implementation of the blade element momentum model on a polar grid and its aeroelastic load impact. *Wind Energy Science*, 5(1):1–27.
- Manohar, K., Brunton, B. W., Kutz, J. N., and Brunton, S. L. (2018). Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns. *IEEE Control Systems Magazine*, 38(3):63–86.
- Mo, Y. and Magri, L. (2025). Reconstructing unsteady flows from sparse, noisy measurements with a physics-constrained convolutional neural network. *Physical Review Fluids*, 10(3):034901. Publisher: American Physical Society.
- Nair, N. J. and Goza, A. (2020). Leveraging reduced-order models for state estimation using deep learning. *Journal of Fluid Mechanics*, 897:R1.
- Neunaber, I., Danbon, F., Soulier, A., Voisin, D., Guilmineau, E., Delpech, P., Courtine, S., Taymans, C., and Braud, C. (2022). Wind tunnel study on natural instability of the normal force on a full-scale wind turbine blade section at Reynolds number $4.7 \cdot 10^6$. *Wind Energy*, 25(8):1332–1342.
- Potentier, T., Braud, C., Guilmineau, E., Finez, A., and Le Bourdat, C. (2021). Analysis of the DANAERO wind turbine field database to assess the importance of different state-of-the-art blade element momentum (BEM) correction models. *Energy Science and Engineering*, 9(9):1477–1500.
- Prandtl, L. and Betz, A. (1927). *Vier Abhandlungen zur Hydrodynamik und Aerodynamik: (Flüssigkeit mit kleiner Reibung; Tragflügeltheorie, I. und II. Mitteilung; Schraubenpropeller mit geringstem Energieverlust)*. Selbstverlag des Kaiser Wilhelm Instituts für Strömungsforschung.

- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707.
- Rott, A., Petrović, V., and Kühn, M. (2020). Wind farm flow reconstruction and prediction from high frequency SCADA Data. *Journal of Physics: Conference Series*, 1618(6):062067.
- Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28.
- Schneider, T., Lan, S., Stuart, A., and Teixeira, J. (2017). Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophysical Research Letters*, 44(24).
- Sirovich, L. (1987). Turbulence and the dynamics of coherent structures part i: Coherent structures. *Quart. Appl. Math.*, 45(3):561–571.
- Snel, H. and Schepers, J. G. (1995). Joint investigation of dynamic inflow effects and implementation of an engineering method. *Stichting Energieonderzoek Centrum Nederland, Netherlands Energy Research Foundation ECN*.
- Swischuk, R., Mainini, L., Peherstorfer, B., and Willcox, K. (2019). Projection-based model reduction: Formulations for physics-based machine learning. *Computers and Fluids*, 179:704–717.
- Taira, K., Brunton, S. L., Dawson, S. T. M., Rowley, C. W., Colonius, T., McKeon, B. J., Schmidt, O. T., Gordeyev, S., Theofilis, V., and Ukeiley, L. S. (2017). Modal analysis of fluid flows: An overview. *AIAA Journal*, 55(12):4013–4041.
- Tan, A. C. C. (2023). Flow reconstruction from sparse sensors based on reduced-order autoencoder state estimation. *Physics of Fluids*, 35(7):075127.
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4):267–276.
- Tomasetto, M., Williams, J. P., Braghin, F., Manzoni, A., and Kutz, J. N. (2025). Reduced Order Modeling with Shallow Recurrent Decoder Networks.
- Visonneau, M., Queutey, P., Deng, G., Guilmineau, E., Leroyer, A., and Wackers, J. (2014). Logiciel de simulation ISIS-CFD, version 3.0-1.
- Williams, J. P., Zahn, O., and Kutz, J. N. (2024). Sensing with shallow recurrent decoder networks.
- Williamson, C. H. K. (1996). Vortex dynamics in the cylinder wake. *Annual Review of Fluid Mechanics*, 28(Volume 28, 1996):477–539.
- Xu, J. and Duraisamy, K. (2020). Multi-level convolutional autoencoder networks for parametric prediction of spatio-temporal dynamics. *Computer Methods in Applied Mechanics and Engineering*, 372:113379.
- Xu, S., Sun, Z., Huang, R., Guo, D., Yang, G., and Ju, S. (2023). A practical approach to flow field reconstruction with sparse or incomplete data through physics informed neural network. *Acta Mechanica Sinica*, 39(3):322302.
- Zhang, J., Brindise, M. C., Rothenberger, S. M., Markl, M., Rayz, V. L., and Vlachos, P. P. (2022). A multi-modality approach for enhancing 4D flow magnetic resonance imaging via sparse representation. *Journal of The Royal Society Interface*, 19(186):20210751.