

RESEARCH ARTICLE

# Data-driven sparse reconstruction of flow over a stalled aerofoil using experimental data

Douglas W. Carter<sup>1</sup>, Francis De Voogt<sup>1</sup>, Renan Soares<sup>1</sup> and Bharathram Ganapathisubramani<sup>1</sup><sup>\*</sup>

<sup>1</sup>Department of Aeronautical & Astronautical Engineering, University of Southampton, Southampton SO17 1BJ, United Kingdom

<sup>\*</sup>Corresponding author. E-mail: [g.bharath@soton.ac.uk](mailto:g.bharath@soton.ac.uk)

**Received:** -; **Revised:** -; **Accepted:** -

**Keywords:** Sparse reconstruction; Extended proper orthogonal decomposition; Shallow neural network; Separated aerofoil

## Abstract

Recent work has demonstrated the use of sparse sensors in combination with the proper orthogonal decomposition (POD) to produce data-driven reconstructions of the full velocity fields in a variety of flows. The present work investigates the fidelity of such techniques applied to a stalled NACA 0012 aerofoil at  $Re_c = 75000$  at an angle of attack  $\alpha = 12^\circ$  as measured experimentally using planar time-resolved Particle Image Velocimetry (PIV). In contrast to many previous studies, the flow is absent of any dominant shedding frequency and exhibits a broad range of singular values due to the turbulence in the separated region. Several reconstruction methodologies for linear state estimation based on classical compressed sensing and extended POD methodologies are presented as well as non-linear refinement through the use of a shallow neural network (SNN). It is found that the linear reconstructions inspired by the extended POD are inferior to the compressed sensing approach provided that the sparse sensors avoid regions of the flow with small variance across the global POD basis. Regardless of the linear method used, the non-linear SNN gives strikingly similar performance in its refinement of the reconstructions. The capability of sparse sensors to reconstruct separated turbulent flow measurements is further discussed and directions for future work suggested.

## Impact Statement

Sparse reconstruction of full-field information using a limited subset of data is widely relevant to data-centric engineering applications; from reconstructing human faces with limited pixels to predicting laminar and turbulent flow fields from limited sensors. The focus of the present study is of the latter example with high relevance to active flow control in aerospace and related industry. There are multiple data-driven methodologies for obtaining flow field reconstructions from sparse measurements ranging from the linear unsupervised proper orthogonal decomposition to the use of non-linear supervised neural networks. The feasibility of such methods to flow fields that are highly turbulent as well as obtained via experiment remains an open area of research. The present study reveals the capability of these techniques to create a time-invariant library that can predict instantaneous states of the flow from sparse measurements alone (provided that these states are within the bounds of the applied training data). The proposed linear methods, as well as the neural network architecture, provide well-characterized frameworks for future efforts in sparse sensing and state estimation applications; particularly for highly non-linear underlying systems such as turbulent flow.

## 1. Introduction

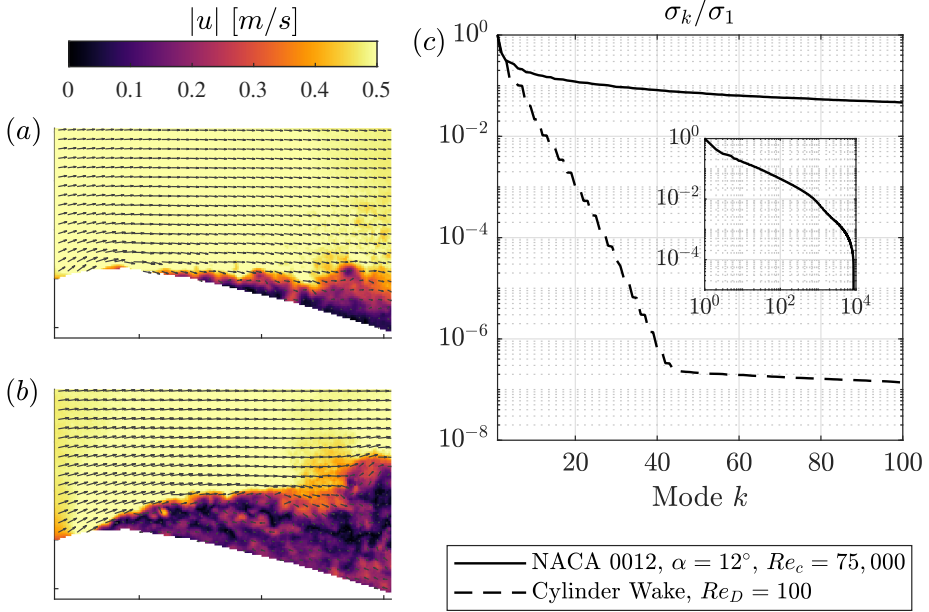
Sparse reconstruction is a technique used to obtain accurate details about the full scale features of a system using a sparse subset of information (for example, a few pixels or measurements within the system) and has been the subject of interest for some decades (Donoho, 2006; Candès et al., 2006). Applications for such state estimation problems range from reconstructing faces from limited or corrupted data (Wright et al., 2008) to deblurring and improving image resolution (Dong et al., 2011) to estimating global sea surface temperatures (Manohar et al., 2018; Callaham et al., 2019). The literature concerning state estimation and sparse reconstruction is rapidly developing. In the following we motivate and present a brief review of the literature as it relates to the present study. For a more comprehensive review, we refer the interested reader to Manohar et al. (2018) and Nair and Goza (2019).

In general, the requirements for accurate sparse reconstructions from limited data are that (i) the basis underlying the data exhibits sparsity (as will be discussed) and (ii) that full-state information for the system can be obtained or approximated *a-priori* to generate a global basis applicable to any sample or instant (Brunton and Kutz, 2019). For example, suppose a set of images of weathered ancient hieroglyphics are only partially discernible. If a library of images of known undamaged hieroglyphics is tabulated *a-priori*, the principle of sparsity can be used to estimate the weathered hieroglyphics before they were damaged (Roman-Rangel et al., 2012). Further examples of sparse reconstruction exist over a diverse range of engineering disciplines. Liu et al. (2017) demonstrated the use of sparsity to monitor and detect faults in an industrial Tennessee Eastman Process using a novel variation of principle component analysis. Bao et al. (2017) showed the use of sparse strain sensors to estimate the full stress state of a structure using a Fourier basis. Iyer et al. (2020) used recurrent neural networks and sparse observations for reconstructing and forecasting road traffic.

Sparse reconstruction may also be extended to data-assimilation applications. Brajard et al. (2020) demonstrated assimilating data from noisy sparse observations for predicting a chaotic 40-variables Lorenz system using a convolutional neural network. Further applications of data-assimilation using sparsity remain to be explored. For example, if the entire flow field in the neighborhood of a gas turbine blade is simulated and a global basis tabulated, the flow field in the neighborhood of a real turbine blade fitted with flow sensors can be estimated. Such a reconstruction technique is widely applicable to any system exhibiting many degrees of freedom. Therefore, the number of possible applications for sparse reconstructions in science and engineering is large.

There remains much to be done in order to test the limitations of sparse reconstructions numerically, experimentally, and via data-assimilation for a variety of engineering systems. For example: how do variations of non-dimensional parameters that characterize a system impact the reconstruction? How many known full-state snapshots are required to generate a global basis? How many sparse sensors or probes are needed to achieve a desired reconstruction accuracy? Where should the probes be placed? These are some of the underlying questions that motivate the current study seeking to expand the application of sparse sensing to engineering problems.

The key underlying principle that allows for full reconstructions using only limited measurements is the sparsity of the representative basis. For the application of interest, a suitable basis must be chosen onto which to project the sparse signal (the limited measurements) for the full reconstruction. The basis of choice depends on the data in question but is typically taken to be either a Fourier transform (Candès and Wakin, 2008) or a data-driven basis such as the proper orthogonal decomposition (POD: Sirovich, 1987; Berkooz et al., 1993). If the basis contains many entries that are small (near-zero) then the system is said to exhibit sparsity. For example, a Fourier transform may indicate only a few frequencies have significant amplitudes. Sparse reconstruction takes advantage of the sparsity of the basis functions to produce full reconstructions (Brunton and Kutz, 2019) that can yield surprisingly accurate full state estimations. For example, using POD as a representative basis Manohar et al. (2018) showed that using



**Figure 1.** Instantaneous velocity fields (a,b) from the PIV data (every fifth vector shown for clarity) presented in this study at two separate instants; highlighting the variation in the size of the separated wake. The singular values are shown in (c) normalized by the first singular value (inset: up to 9000 modes). This is also shown for the laminar cylinder wake (dashed) of diameter  $D$  at  $Re_D = 100$  from the DNS of Brunton and Kutz (2019) for comparison.

only 7 probes they could reconstruct the full vorticity field of laminar flow over a cylinder exceeding 90% accuracy.

POD is a data-driven method commonly used as a basis for reconstructions due to its attractive properties of being energy optimal and having time-independent spatial modes, though other choices are possible (Bai et al., 2015; Jayaraman et al., 2019). Manohar et al. (2018) demonstrated that the reconstruction accuracy is significantly improved if the choice of probe locations is made carefully such that it takes better advantage of the sparse basis. They demonstrated that the use of greedy algorithms to intelligently place probes within the flow improved the reconstructions greatly compared to random probe placement. However, this also depends on the complexity of the underlying system. When the analysis was applied to sea surface temperatures whose POD basis requires many more modes, the reconstruction accuracy relied more heavily on the optimal probe placement; pointing to the difficulty of reconstructing systems with a large range of spatio-temporal scales.

Several alternative methods originating in the fluid mechanics literature have been proposed for sensing turbulent flow structures and obtaining reconstructions, such as linear stochastic estimation (LSE: Adrian and Moin, 1988). In LSE, the state of the flow field is tabulated and conditioned upon the state of an event, leading to a linear map through an  $l_2$  minimization procedure. Several studies have demonstrated the fidelity of reconstructions based on the concept of LSE (Picard and Delville, 2000; Lasagna et al., 2013). Borée (2003) showed that LSE is equivalent to the so-called *extended* POD, where the temporal POD modes of one quantity are projected onto by the state of the flow field. This produces a set of extended modes revealing the spatial structures of the flow that are correlated with the quantity of interest. The extended POD framework is a straightforward and flexible approach offering alternative means of obtaining reconstructions using the temporal modes of the sparse information, as demonstrated by Hosseini et al. (2015). Extended POD was pushed further by Discetti et al. (2018) to

obtain up-sampled time-resolved velocity fields from sparse sensors; demonstrating the flexibility of the framework.

Although the aforementioned techniques produce sparse reconstructions that are obtained entirely linearly, supervised machine learning provides a non-linear framework with additional flexibility. In fact, POD itself is a form of unsupervised machine learning (Brunton et al., 2020) and one may interpret the POD as an unsupervised neural network with a single layer and linear activation function (Milano and Koumoutsakos, 2002). The use of neural networks (NNs) to obtain sparse reconstructions has seen recent attention in the literature. Nair and Goza (2019) demonstrated the ability of NNs to outperform the linear counterparts using a network of three layers for low Reynolds number flow over a flat plate at high angles of attack. Similarly, Erichson et al. (2019) introduced a “shallow” neural network (SNN) consisting of two layers for reconstructing laminar flow over a cylinder with high fidelity. They further applied this method to a comparatively more challenging numerical simulation of isotropic turbulence and found difficulty in predicting future states. They discussed the use of regularization of the loss function to avoid overfitting; a problem that can easily arise with limited training data. Intelligent choices for loss functions offers promising potential for improved neural network performance and generalization. For example, imposing physical constraints on the loss function has been demonstrated as a successful approach for a variety of systems governed by partial differential equations (Raissi et al., 2019; Sun and Wang, 2020). More recently, convolutional neural networks have been demonstrated for flow reconstruction from wall shear measurements in a turbulent channel flow (Guastoni et al., 2020), however the inherent spatial dependence renders convolution-based approaches unattractive for the particular case of spatially sparse reconstruction. This does however represent progress in the direction of reconstructing turbulent flows, as their inherent range of spatial and temporal scales makes them significantly more challenging than the laminar case.

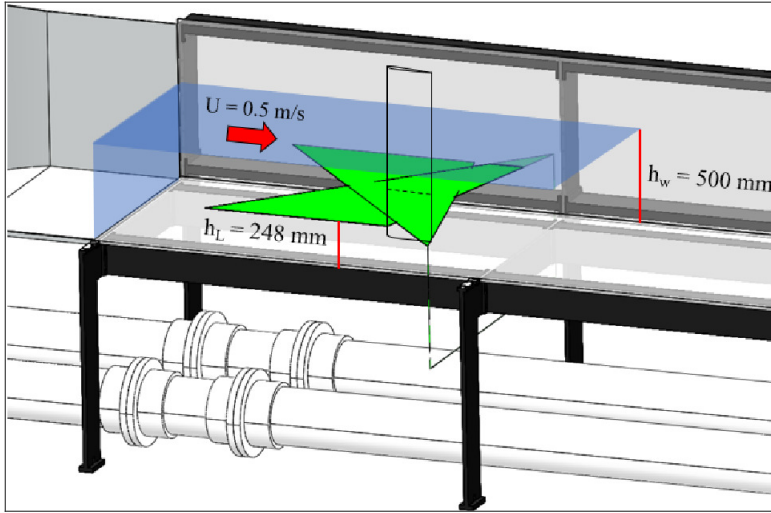
The present study focuses on the problem of sparse sensing motivated by the need to detect flow structures in aerospace applications. For example: to detect anomalous structures, the onset of stall over an aerofoil, or to inform control systems designed to reduce drag. Here the focus is on the situation of an aerofoil in stalled conditions. We seek to explore how a global basis may be used to predict the unseen flow about a stalled aerofoil using limited single-point sensors placed in the flow.

As of yet, data-driven sparse sensing has not been widely applied to advective turbulent flows such as the case of a stalled aerofoil due to the inherently large range of scales and difficulty in capturing a low-order representation of the dynamics. As a result, the flow requires many modes compared to e.g. the case of a laminar cylinder (figure 1) to capture the dynamics. This fact motivates exploration of reconstructions obtained outside the classic compressed sensing approach, which relies more heavily on optimal placement for systems with many modes. The present study seeks to explore how effectively the various linear and non-linear approaches for sparse reconstruction perform for (i) a situation which is highly turbulent and (ii) data generated via experiment and therefore contaminated by experimental noise.

The present study will be structured as follows. The experiments performed to obtain the two-dimensional time-resolved flow fields of the separated aerofoil are outlined in section 2. The methodology used to select probe placements and obtain the sparse reconstructions with and without neural network refinement is presented in section 3. The results are presented in section 4. Conclusions and outlook for future work are finally presented in section 5.

## 2. Experimental Method

Particle image velocimetry (PIV) has seen widespread use in experimental fluid mechanics over the past decades due to its ability to obtain highly spatially-resolved instantaneous planar two-component velocity fields (Adrian et al., 2011). Combined with time-resolved capabilities of high-speed cameras, it is the measurement method of choice for the present data-driven analysis due to the ability to generate large spatial domains (and resulting spatial modes) as well as time information at each point in the



**Figure 2.** Illustration of the experimental setup focusing on the test section of the water flume flow facility at the University of Southampton. The NACA 0012 aerofoil is illuminated from both sides, however for this study a field of view focusing on the suction side of the aerofoil is used to capture the separation of the wake. The water level  $h_w$  and laser sheet level  $h_L$  are indicated.

flow; enabling the sparse reconstruction investigation via a pseudo-probe analysis. The details of the data collection and post-processing are outlined in the following sections.

### 2.1. Water Flume Experiment

To obtain planar time-resolved velocity fields on the suction side of a separated NACA 0012 aerofoil, a PIV campaign was performed in the water flume flow facility located at the University of Southampton as illustrated in figure 2. A NACA 0012 aerofoil model of chord length  $c = 15$  cm and span  $s = 70$  cm was fixed vertically in the center of the span of the flume immediately following the contraction into the test section. The aerofoil was fixed at angle of attack  $\alpha = 12^\circ$  using a stepper motor attached to an overhead carriage system with precise control over the stepper motor angle. A 4 Megapixel Phantom v641 camera mounting a 105 mm Ex Sigma lens ( $f = 5.6$ ) was directed upward from underneath the flume at a standoff distance 88 cm, resulting in a stream-wise wall-parallel field of view approximately 16 cm in the stream-wise direction  $x$  and 10 cm in the stream-normal wall-parallel direction  $y$ . The field of view was illuminated via a set of sheet-forming optics directing laser light from a Litron 527 nm Nd:YLF high-speed laser into the test section from the side of the facility at a wall-normal height of  $h_L = 248$  mm from the bottom of the flume. The flume was filled until the water reached a height  $h_w = 500$  mm for which the maximum frequency of the flume pumps yielded a free stream velocity  $U_\infty = 0.5$  m/s corresponding to a Reynolds number based on the chord length  $Re_c = \frac{U_\infty c}{\nu} = 75,000$ , where  $\nu$  is the kinematic viscosity.

To collect the images, Davis 8.3.1 PIV software was used with a LaVision high-speed controller to ensure synchronous timing of laser and camera. The flow was seeded with Vestosint 2157 polyamide particles of nominal diameter  $55 \mu\text{m}$  verified to behave as faithful flow tracers. The seeding density was iteratively adjusted until a satisfactory number of particle reflections across the field of view were observed. The images were captured at full resolution ( $2560 \times 1600$  pixels) at a frequency of 750 Hz. In total, 9000 separate snapshots were collected for data training purposes and one fully time-resolved run corresponding to 5468 sequential snapshots across 7.3 seconds of continuous measurement



was collected to test sparse reconstruction methodologies. At an angle of attack  $\alpha = 12^\circ$ , substantial separation from the surface of the aerofoil was observed (figure 1).

## 2.2. Data Post-Processing

The images were background-subtracted and low frequency noise was attenuated using a convolution with a Gaussian high-pass filter of standard deviation of 5 pixels. A mask corresponding to the location of the aerofoil in the images was stored and applied during PIV processing. The images were processed using a verified in-house PIV code based in MATLAB with iterative interrogation window stepping from  $64 \times 64$  pixels to  $32 \times 32$  pixels to  $24 \times 24$  pixels with 50% overlap. Subpixel displacements were obtained via a Gaussian 3-point fit, and detected outliers were flagged and replaced via local interpolation. Interrogation windows found to overlap with the image mask by greater than 25% were discarded.

For each component of the velocity field, the velocity was decomposed into a mean and fluctuating component  $u = U + u'$  and  $v = V + v'$  where the capitals here denote the mean velocity field and the prime denotes the fluctuating components of the horizontal (stream-wise)  $u$  and vertical (stream normal) fields  $v$ . To reduce the random noise of the PIV fields, the filtering method based on POD described in Raiola et al. (2015) was used (see section 3.1) with the suggested value of the parameter  $F = 0.999$  (the ratio of forward to backward residuals of the rank-restricted POD reconstruction). The location of the rank-order filtering truncation (see 3.1) was found to correspond to only 5% of the energy of the velocity fluctuations. To improve the accuracy of the velocity fields, particularly in the separated region, gappy POD was employed using iterative replacement of detected outliers (Gunes et al., 2006) until satisfactory convergence was achieved. It was verified that the number of POD modes used to reduce random noise and improve the accuracy of the velocity fields was greater than the number of modes used to test the sparse reconstructions.

## 3. Sparse Reconstruction Methodology

### 3.1. Proper Orthogonal Decomposition (POD)

For the present study, the proper orthogonal decomposition (POD: Sirovich, 1987; Berkooz et al., 1993) is used as a basis for the reconstruction due to its unique property as an energy-optimal basis. POD is commonly utilized as a basis for sparse reconstruction (Candès et al., 2006), though other choices (such as a Fourier basis) are possible (Jayaraman et al., 2019). In the present notation the fluctuating velocity fields are rearranged into a matrix  $U$  of size  $n_t \times 2n_x$ , where  $n_t$  is the number of snapshots and  $n_x$  is the total number of spatial points (PIV vectors). At each instant the fluctuating horizontal velocity field  $u'$  is appended row-wise with the vertical  $v'$  resulting in twice the number of spatial points in the rows of  $U$  (see e.g. Taira et al., 2017). Henceforth, capital  $U$  will denote the appended fluctuating velocity fields (not to be confused with the mean velocity fields). The POD is then calculated as:

$$U = \Psi \Sigma \Phi, \quad (1)$$

where  $\Psi$  is an orthogonal matrix of temporal modes of size  $n_t \times n_t$  satisfying  $\Psi^T \Psi = I$  with corresponding columns  $\psi_k$  containing the  $k$ -th temporal mode.  $\Sigma$  is a diagonal matrix of singular values (the square root of the eigenvalues of the velocity autocorrelation) of size  $n_t \times n_t$  containing the relative contribution of each mode where  $\sigma_k$  is the  $k$ -th diagonal entry. Finally,  $\Phi$  is an orthogonal matrix of spatial modes of size  $n_t \times 2n_x$  satisfying  $\Phi^T \Phi = I$  with corresponding rows  $\phi_k$  containing the  $k$ -th spatial mode. POD is an exact decomposition of the velocity fields, however a low-order representation of the fields is easily obtained by imposing a rank-order truncation for  $k$  desired modes. POD is a data-driven decomposition in the sense that the number of available modes depends on the number snapshots used to perform the POD. Often in the literature the temporal modes and the singular values are combined into a single matrix  $A = \Psi \Sigma$  referred to as the POD coefficients with columns  $a_k$  corresponding to the

$k$ -th mode coefficients across all samples and  $a(t)_k$  the  $k$ -th mode coefficient for a particular sample at time  $t$ . A truncated reconstruction corresponding to  $k$  modes is then obtained using the coefficients at a particular instant as

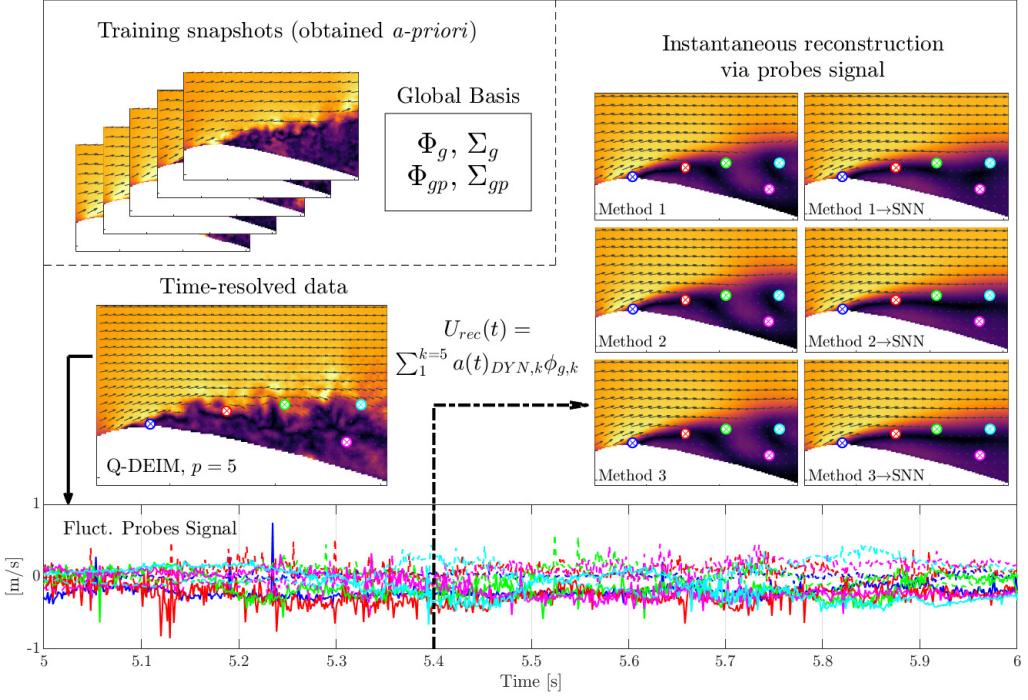
$$\mathbf{U}(t) = \sum_1^k a(t)_k \boldsymbol{\phi}_k \quad (2)$$

where here the explicit time dependence in  $\mathbf{U}(t)$  is retained to emphasize it is a vector of fluctuating velocities of length  $2n_x$  at time instant  $t$ . A key feature of POD that allows for sparse reconstruction is in the time independence of the spatial modes  $\boldsymbol{\phi}_k$ . With enough samples the most energetic modes (associated to the underlying physical mechanisms) converge (i.e. the modes do not change with additional samples). For these time-independent energetic modes, the reconstruction becomes a matter of approximating  $\mathbf{A}$  using a limited number of sensors. In practice, approximating the coefficients can be achieved in several ways including the use of LSE (Lasagna et al., 2013) or through the use of a Kalman filter (Kalman, 1960; Troshin and Seifert, 2019). For the present study we investigate three distinct methods of obtaining a reconstruction from the sparse probes. The underlying motivation for applying these separate approaches is detailed in section 3.3. Psuedo-probes (hereafter referred to as just *probes*) taken at specific locations in the velocity fields are utilized to explore the sparse reconstruction.

### 3.2. Probe Placement

The optimal placement of probes for sparse reconstruction is a challenging and ongoing subject of research (Manohar et al., 2018). As the problem involves both the number of sensors as well as the number of modes used for reconstruction, the optimization is combinatorial in nature; making it intractable for even a modest number of possible sensor locations. The objective is to maximize the signal to noise ratio of the reconstruction by minimizing the condition number of the sparse basis (Manohar et al., 2018; Jayaraman et al., 2019). A body of literature has been reported exploring heuristic greedy algorithms known as empirical interpolation methods (EIM: Barrault et al., 2004; Willcox, 2006; Yildirim et al., 2009) or discrete EIMs (DEIM: Chaturantabut and Sorensen, 2010; Drmac and Gugercin, 2016) to identify optimal locations for the probes. In addition, optimal design literature provide methods of identifying probe locations based on the moment matrix of the basis (e.g. A, C, D, E-optimal design, see Atkinson and Donev, 1992; Cox and Reid, 2000). Considerations for non-linear placement have also been explored as summarized in the recent work of Otto and Rowley (2021) and references therein.

The work of Manohar et al. (2018) shows that the optimal choice of placement (for the linear reconstruction) corresponds to the locations that contribute the maximum variance ( $l_2$ -norm) across the spatial basis. The QR decomposition with pivoting (Van Loan and Golub, 1983) is a commonly utilised heuristic for identifying such locations, resulting in an upper-triangular matrix  $\mathbf{R}$  with entries ordered accordingly. The resulting pivot locations approximate the best sensor locations. When the number of modes used for the reconstruction is equal to the number of probes, the QR with column pivoting is applied to the spatial basis and is known as the Q-DEIM (Drmac and Gugercin, 2016). When the number of probes exceeds the number of modes used for reconstruction the problem is over defined and the QR-based method requires additional treatment (Manohar et al., 2018). As the present study will investigate reconstructions using non-linear methods (in addition to linear), the optimal placement is not trivial. The scope of the present study will be limited to considering probes placed randomly (the sub-optimal case) and using the Q-DEIM with pivoting (to approximate the optimal placement). For the case of randomly placed probes, one set of random locations is used (as opposed to testing multiple sets of random locations). This is due to the difficulty in generating and storing a large number of random global probe libraries, the need for which is presented in section 3.3.2.



**Figure 3.** Conceptual illustration of the instantaneous reconstruction methodology using  $p = 5$  probes and Q-DEIM placement. The global basis is obtained *a-priori* and the real-time probe signals are used to approximate the instantaneous fields. The probe signals are shown with a solid line for  $u'$  and dashed for  $v'$  and color-coded according to their indicated locations. The total velocity shown in the plots are calculated by summing the mean and fluctuating fields. See sections 3.3 and 3.4 for details of the methods.

### 3.3. Reconstruction

The accuracy of the reconstruction is a matter of approximating the real POD coefficients of the full velocity fields using a set of “dynamic” coefficients  $\mathbf{A}_{DYN}$  that are estimated via the sparse probe signals  $\mathbf{U}_p$ . For the present analysis, the probe signals are obtained via  $\mathbf{U}_p = \mathbf{U}\mathbf{C}$ , where  $\mathbf{C}$  is a Boolean matrix known as the *sparse matrix* of size  $2n_x \times p$ , where  $p$  is the number of probes. Each column of the sparse matrix contains a single entry equal to one and zero elsewhere; providing a map that subsamples from all spatial locations  $\mathbf{x}$  to the sparse locations of the probes  $\mathbf{x}_p$  (Jayaraman et al., 2019). With the POD coefficients estimated, the reconstruction is obtained via the global POD basis

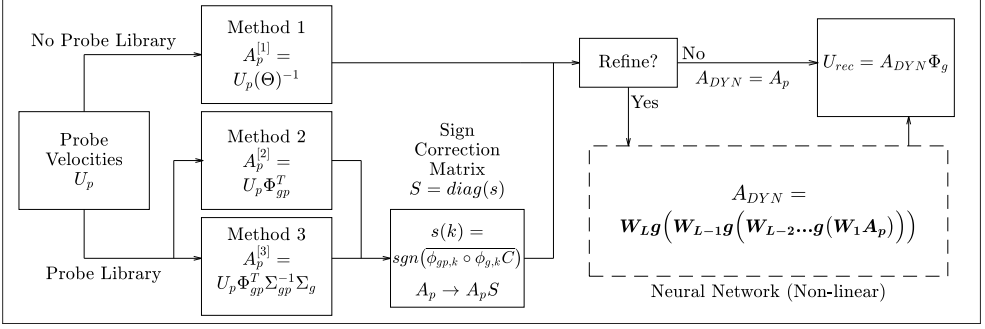
$$\mathbf{U}_{rec} = \mathbf{A}_{DYN} \mathbf{\Phi}_g, \quad (3)$$

or, more explicitly, for a up to a specified number of modes at a particular instant:

$$\mathbf{U}_{rec}(t) = \sum_{k=1}^k a(t)_{DYN,k} \phi_{g,k} \quad (4)$$

where  $\mathbf{\Phi}_g$  are the *global* spatial modes from the POD performed on a set of training data  $\mathbf{U}_g$ . There lies an implicit assumption in the application of equation 4 such that the external state of the system is unchanged during the reconstruction compared to the conditions under which the global basis was tabulated. The instantaneous state must remain within the bounds of the variation captured by the training data. In other words, using the aerofoil as an example, variations in the fixed free stream





**Figure 4.** Block diagram of the reconstruction starting from the probe signal  $U_p$ .

velocity or angle of attack are not accounted for. As such, the scope of the methodology is limited to testing the efficacy of the sparse reconstructions without external changes to the state of the system.

For the present study, the training data consists of 9000 samples. The training data is used to generate the time-independent global basis  $\Phi_g$  in order to predict the complete spatio-temporal evolution of the flow based on sparse time-resolved measurements via equation 4. The large training data set is needed due to the high spatio-temporal variability of the turbulent separated flow, leading to a slowly decreasing set of singular values  $\Sigma_g$  compared to e.g. a laminar flow (figure 1) and therefore many modes are required within the global basis. The global POD was calculated accordingly using  $n_t = 9000$  and  $2n_x = 50,952$ , resulting in a computation time of 418 seconds (7 minutes) on a single desktop computer (double precision with 3.6GHz CPU and 16GB RAM). It was determined that a larger set of training data was not required as the global POD library was found to be satisfactorily converged up to the maximum number of modes used for reconstructions at  $k = 500$ . As such the accuracy of the reconstructions depends only on how closely the dynamic coefficients  $A_{DYN}$  approximate the real ones up to the number of modes used for the reconstructions. The time-resolved reconstruction was then tested on all 5468 samples of the time-resolved data set, resulting in a training-to-validation ratio of approximately 1.65.

There are several methods used in order to approximate the coefficients that will be outlined in the following subsections. For all of the methods, provided that the global POD and optimal placement calculations are performed using a training data set *a-priori*, the calculations are possible to be performed in real-time. This is conceptually illustrated in figure 3. These methods are therefore highly relevant to flow-sensing and control applications. The methods for approximating the coefficients originating from the probes  $A_p$  can either be used for reconstruction immediately, or improved using a shallow neural network (figure 4).

### 3.3.1. Method 1: Sparse recovery reconstruction

The first method used is the most common implementation of sparse recovery originating from the compressed sensing literature (Donoho, 2006; Candès and Wakin, 2008; Callaham et al., 2019). This method allows to calculate a set of coefficients that approximate the real coefficients by taking advantage of the fact that many entries of the global basis are negligibly small (c.f. Jayaraman et al., 2019). With this method, the POD coefficients are estimated by projecting the probes signal into a sparse basis corresponding to the global POD modes evaluated at the locations of the probes

$$A_p^{[1]} = U_p \Theta^{-1} \quad (5)$$

where the bracketed superscript denotes the method used. Here,  $\Theta$  is the sparse basis formed by the product of the global basis with the sparse matrix  $\Theta = \Phi_g C$ . The number of rows in  $A_p$  corresponds to the number of samples (rows) of  $U_p$ , and the number of columns is the number of probe modes. As we

center our analysis around the Q-DEIM placement method described in section 3.2, the sparse basis  $\Theta$  is a square matrix with the number of probes  $p$  equal to the number of reconstruction modes  $k$ .

As mentioned in section 3.2, the optimal locations of the probes are the locations within  $\Phi_g$  that minimize the matrix condition number of  $\Theta$ . In other words, locations within the global basis with very little variance across the modes will lead to error propagation and low signal to noise ratio (SNR) upon the inversion in equation (5). Therefore, the best locations will correspond to the locations with the largest variance across the modes of the global basis.

### 3.3.2. Method 2: Extended POD reconstruction

Motivated by studies that seek to approximate the POD coefficients using either a different variable or a separate measurement, extended POD (Borée, 2003) offers a flexible framework with which to approximate the coefficients. Borée (2003) showed that by projecting the temporal modes of one quantity into the measurements of another, one obtains a set of extended spatial modes that can be used as a basis to reconstruct the part of the measurements that is correlated to the quantity of interest. When all modes are included, this is equivalent to the well known LSE (LSE: Adrian and Moin, 1988; Lasagna et al., 2013; Hosseini et al., 2015).

For the present study, a framework inspired by extended POD is implemented treating the probes as a quantity to be correlated to the full flow field. If the assumption is made that the coefficients of the probes are highly correlated with the coefficients of the flow field, one can use the same training data used to calculate the global POD basis  $\Phi_g$  to produce a global *probe* POD basis  $\Phi_{gp}$ . The coefficients are then obtained as

$$A_p^{[2]} = U_p \Phi_{gp}^T \quad (6)$$

where the transpose is used on the global probe basis as it is orthogonal by construction. The global probe basis is generated for all tested probe numbers and positions across all 9000 training samples used to construct the global libraries. For the sake of memory allocation, only one set of random probe locations is tested to avoid calculating and storing hundreds of additional global probe libraries. It was confirmed after testing two other sets of random placements that the results presented hereafter were qualitatively unaffected.

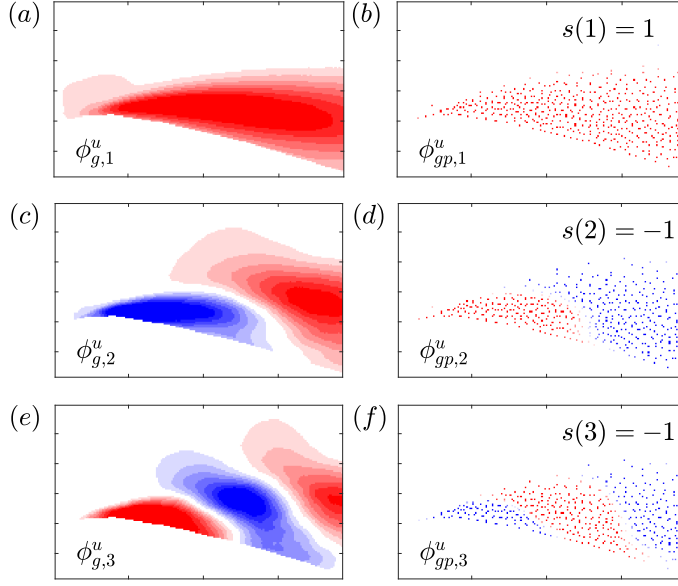
### 3.3.3. Method 3: Quasi-Orthogonal Extended POD reconstruction

Building off of the methodology of the previous subsection, the quasi-orthogonal extended POD reconstruction utilizes all of the available information from the global POD and global probe POD libraries to produce the reconstruction. Instead of simply assuming that the coefficients produced from the global probe modes will approximate the real coefficients, instead a quasi-orthogonal basis  $\Psi_p$  is calculated using the pseudoinverse of the singular values of the global probe modes  $\Sigma_{gp}^{-1}$ . The term “quasi” is used to emphasize that there is no guarantee that the resulting basis will be perfectly orthogonal as it is constructed using the training data and subsequently projected into by independent time-resolved testing data. The coefficients can then be approximated using the global singular values to re-scale the quasi-orthogonal basis as

$$A_p^{[3]} = U_p \overbrace{\Phi_{gp}^T \Sigma_{gp}^{-1}}^{\Psi_p} \Sigma_g. \quad (7)$$

A similar principle was used by Discetti et al. (2018) to up-sample time-resolved fields using simultaneous probe signals at a higher sampling frequency.

We remark that the extended-POD methodologies presented here represent a novel departure from previous studies, e.g. Hosseini et al. (2015). In their case, the temporal information contained within the pressure probes were used to reconstruct the velocity fields using a basis determined by the extended spatial velocity modes. The present work, by contrast, uses the global POD basis for the reconstruction and estimates the coefficients using a separate POD basis tailored for the probes. The POD basis tailored



**Figure 5.** Global spatial modes (a,c,e) and corresponding global spatial probe modes (b,d,f) of  $u'$  mapped to the locations of the full field for 500 probes placed using the Q-DEIM. The colorbars range across  $\pm 3\sigma$  of the corresponding global modes  $\phi_{g,k}^u$  from blue to red. Modes with spatial locations that are in phase give a sign correction  $s = 1$  (a-b) and out of phase  $s = -1$  (c-f).

for the specific sets of probes  $\Phi_{gp}$  is the main novelty of the present approach. As the probes themselves are velocity measurements, the underlying assumption here is in the inherent correlation between the probe velocities, their coefficients, and the full-field velocities.

### 3.3.4. Sign Correction Matrix

For the extended POD based methods of the two previous subsections, a new global probe basis is used to obtain an approximation of the coefficients. The velocity autocorrelation of the probes (the underlying physical quantity used to calculate the POD) is missing information due to the sparsity of the probe signals. This may necessarily lead to spatial probe modes  $\Phi_{gp}$  that, when mapped to the locations of the full field modes  $\Phi_g$ , have opposite sign. Especially for modes corresponding to the largest singular values, this can lead to reconstructions that are anti-correlated with the true underlying velocity fields. This effect is illustrated in figure 5, showing the first three modes for the fluctuating horizontal velocity component from the global library and the global probe library side by side for the case of 500 probes placed using Q-DEIM.

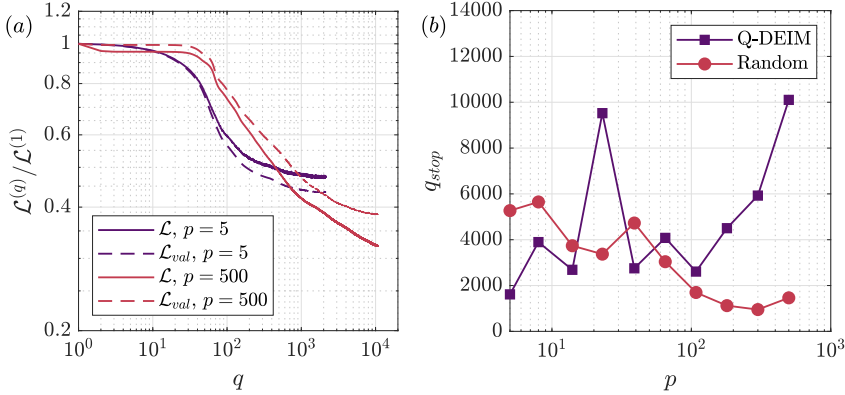
To account for this effect, a sign correction matrix calculation is proposed to be calculated *a-priori* to ensure that the two spatial bases  $\Phi_{gp}$  and  $\Phi_g$  are not anticorrelated (or in other words, as correlated as possible):

$$s(k) = \text{sgn}(\overline{(\phi_{gp,k} \circ \phi_{g,k} C)}), \quad (8)$$

where the overline denotes spatial averaging across the locations of the probes and the  $\circ$  symbol denotes the Hadamard (element-wise) product. The full diagonal (and orthogonal) sign correction matrix is then

$$S = \text{diag}(s) \quad (9)$$

where the diag function outputs a square matrix of zeros with the entries of the vector argument  $s$  along the diagonal. The coefficients are then updated through multiplication with  $S$  as shown in figure 4.



**Figure 6.** Normalized training and validation loss  $\mathcal{L}^{(q)} / \mathcal{L}^{(1)}$  within the SNN for method 1 using 5 and 500 probes and Q-DEIM placement vs number of epochs  $q$  (a) and number of epochs before stopping  $q_{stop}$  vs number of probes for method 1 for each placement (b).

### 3.4. Shallow Neural Network Refinement

As outlined in figure 4, the option to apply non-linear refinement via a neural network is explored. Here we briefly review literature concerning neural network size to frame the specific implementation utilised in this study. Early on, Cybenko (1989) provided proof for the universal approximation theorem via neural networks of arbitrary width with a single hidden layer and a sigmoidal activation function. A remaining issue however was in the capability to train networks with a large number of nodes in a single layer, thereby ceding computational advantage to unsupervised methods such as POD. Further generalization of the universal approximation theorem was later shown for different activation functions and multiple layers (Hornik, 1991). In combination with new training methods, the differences between networks of a single layer with many nodes or multiple layers with less nodes has since faded (Hinton et al., 2006). Deep neural networks with multiple layers remain commonplace, however fewer layers can provide similar results taking advantage of modern hardware and training methods to significantly reduce the required training. Although single layer neural networks are easier to interpret, it has been shown that networks with two layers provide better generalization capabilities (Thomas et al., 2017). Networks with few layers are typically less sensitive to the specific choices of hyperparameters than their deep counterparts. As such, a shallow neural network (SNN) is adopted for the present study.

We apply an SNN architecture as outlined in the work of Erichson et al. (2019). In general, we seek a function  $\mathcal{F}$  that consists of multiple fully connected or convolutional layers with associated scalar non-linear activation functions  $g$  and weights  $\mathbf{W}$  applied to input  $\mathbf{A}$

$$\mathcal{F}(\mathbf{A}; \mathbf{W}) := \mathbf{W}_L g(\mathbf{W}_{L-1} g(\mathbf{W}_{L-2} \cdots g(\mathbf{W}_1 \mathbf{A}))) \quad (10)$$

where a shallow network has a small number of layers  $L$ . For the present study  $L = 3$  is used, where two layers are hidden and followed by an output layer. Due to the spatial sparsity of the inputs in the present study, we favor the fully connected layers over the convolutional layers in line with similar investigations (Erichson et al., 2019; Nair and Goza, 2019).

For the SNN architecture, shown in table 1, we construct a first “imagination” layer with  $1.5p$  nodes followed by a  $1.2p$  “refinement” layer with a 5% dropout layer in between to improve the generalization (Erichson et al., 2019). As the network scales with the input size  $p$ , the amount of trainable parameters varies significantly for different networks considering varying amounts of probes. The training data however has a consistent amount of samples. This requires the larger networks to utilize additional generalization. This is achieved by the 5% dropout layer (and leaves the small networks largely unaffected). The final linear layer contains the same number of inputs (number of probes/modes) as outputs. For the

activation functions the rectified linear unit function (ReLU) is used for the hidden layers and a linear function for the output layer. The optimization is performed using the ADAM adaptive moment optimization (Kingma and Ba, 2014). The learning rate is set to 0.001, with the exponential decay rate for the first moment estimates equal to 0.9 and the exponential decay rate for the second moment estimates equal to 0.999. For numerical stability, the recommended  $\hat{\epsilon} = 10^{-7}$  value is used.

**Table 1.** The neural network architecture. The dropout layer is only active during training.

Layer	Layer type	Number of nodes	Activation
input	input	$p$	linear
hidden 1	dense	$1.5p$	ReLU
(hidden)	dropout	$0.05 \cdot 1.5p$	linear
hidden 2	dense	$1.2p$	ReLU
output	dense	$p$	linear

The same training data used to obtain the POD basis  $\Phi_g$  is used to train each neural network  $\mathcal{F}$  using 89% of the samples  $n = 8000$ , reserving  $m = 1000$  random samples for iterative learning validation. As demonstrated by Guastoni et al. (2020), the neural network for the present study is trained to recover the POD coefficients. The global coefficients for training the outputs are obtained as  $A_g = U_g \Phi_g^T$  up to  $k = p$  modes (where  $p$  is the number of probes), and the sparse probe coefficients  $A_{gp}^{[i]}$  using the  $i$ -th linear method applied to the training data  $U_g C$  as the inputs. For each set of probe locations and for each linear method, a separate network and corresponding set of weights is trained to minimize the loss

$$W = \arg \min_{\tilde{W}} \mathcal{L}(\tilde{W}) \quad (11)$$

where the loss is defined for each epoch  $q$  as

$$\mathcal{L}^{(q)}(\tilde{W}) = \sum_{j=1}^n \left\| A_g(t_j) - \tilde{\mathcal{F}}(A_{gp}^{[i]}(t_j); \tilde{W}) \right\|_1 \quad (12)$$

where  $A(t_j)$  indicates the vector of coefficients corresponding to the  $j$ -th training sample and the  $\tilde{\cdot}$  indicates a dummy function or variable used for training. The  $l_1$  norm is chosen in this study for evaluating the loss as it was found to outperform the commonly used  $l_2$  norm. The use of the  $l_1$  norm (in combination with unscaled coefficients, as will be discussed) aids the network in prioritizing an error reduction on the coefficients corresponding to the modes with larger singular values. In an effort to maintain the singular value scaling of the corresponding temporal modes, the coefficient inputs and outputs are not scaled to a zero-to-one range as is often done to speed up learning. This appears to have a minor influence on the training; only effecting initial optimization epochs where the optimizer must adjust before reducing the loss (figure 6). The use of batch normalization for generalization was omitted and instead large batches of 3000 samples were used. These have the added benefit of speeding up training. Other methods such as layer activation regularization or layer weight regularization were explored but were found to produce less consistent results over the various neural networks without repetitive hyperparameter optimization.

It is possible to include additional constraints on the loss function in equation 11 in order to keep the network more generalized (Erichson et al., 2019; Sun and Wang, 2020). Instead for this study early stopping was opted for to avoid overfitting. The trained network was evaluated on the validation subset every epoch and for every iteration where the loss decreased the weights were stored. If the network did not experience a new minimum in the validation loss within 500 epochs, the training was halted and the last stored weights adopted. The value of 500 was chosen ad-hoc for this particular study as it was found to strike a balance between “kick-starting” the network with enough initial epochs but was large enough to revert any apparent overfitting. This is shown in figure 6a for  $p = 5$  and 500 probes via input



coefficients from method 1. With this architecture and early stopping criterion, the number of epochs was on the order of a few thousand, increasing with increasing probes to a maximum of approximately 10 thousand epochs (figure 6b).

We remark that the reconstructions use the same number of modes as probes, therefore for the present study the more probes used for the inputs the greater the performance ceiling for the SNN (up to the limit of the real coefficients truncated at  $p$  modes).

### 3.5. Quantification

Two metrics are utilized to quantify the reconstruction accuracy in the present study. To quantify the phase properties of the reconstruction, the normalized correlation is defined as

$$\rho_{u'} = \frac{\langle u'(\mathbf{x}, t) u'_{rec}(\mathbf{x}, t) \rangle}{\langle u'(\mathbf{x}, t)^2 \rangle^{1/2} \langle u'_{rec}(\mathbf{x}, t)^2 \rangle^{1/2}} \quad (13)$$

where the angled brackets  $\langle \cdot \rangle$  indicated averaging over all space and time and here  $u'_{rec}$  is the reconstruction of the horizontal fluctuating velocity. The correlation takes a value of 1 for a reconstruction that is completely in phase, -1 when it is out of phase, and 0 when it is uncorrelated. To evaluate the overall difference between the original and reconstructed fields, the root-mean-square error is defined as

$$e_{u'} = \frac{\langle (u'(\mathbf{x}, t) - u'_{rec}(\mathbf{x}, t))^2 \rangle^{1/2}}{\langle u'(\mathbf{x}, t)^2 \rangle^{1/2}}. \quad (14)$$

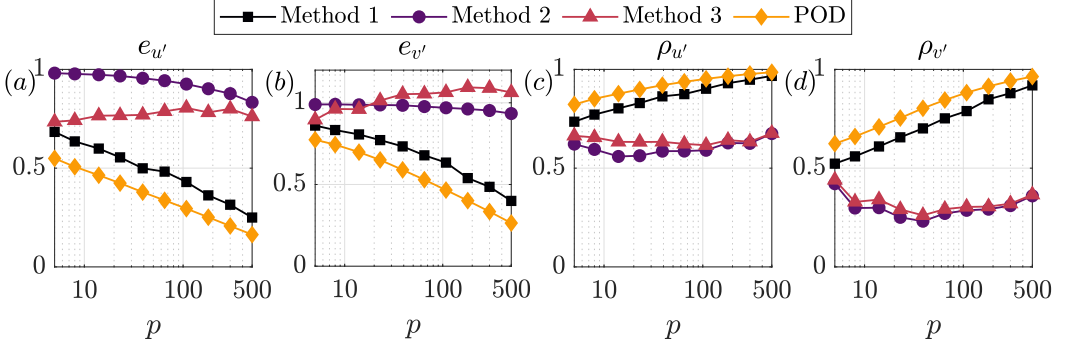
The root-mean-square error describes the fraction by which the reconstruction differs from the original fields, with a perfect reconstruction at  $e = 0$ . Equations 13 and 14 are analogously defined for the vertical fluctuations  $v'$ .

## 4. Results

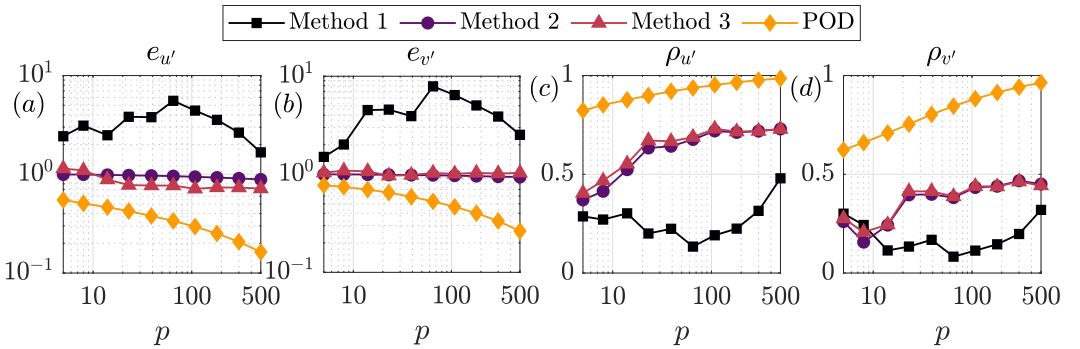
The results are presented first using the linear methods alone, followed by with non-linear SNN refinement as outlined in section 3 and figure 4. The reconstructions evaluate the ability of the linear and non-linear methods to predict instantaneous flow fields using the testing data consisting of 5468 samples. For all results, the number of modes  $k$  used in the reconstructions is equal to the number of probes  $p$ . This was chosen based on the underlying principles of the calculation for the optimal placement using the Q-DEIM (see section 3.2); revealing optimal probe locations for reconstruction method 1 specifically (section 3.3.1). These were found to correspond to locations within the separated region of the flow (figure 5). We remark, however, that the other linear methods (methods 2 and 3) need not necessarily use all  $k = p$  modes in their reconstructions. For the present study we opt to present  $k = p$  for all methods for consistency of comparison.

### 4.1. Linear Reconstruction

The normalized root-mean-square error and correlations for the three linear methods are presented for each component of the fluctuating velocity in figures 7 and 8 for Q-DEIM and random placement, respectively. The best possible performance is indicated in both figures by the reconstruction calculated using the rank-truncated POD via equation 2. It is immediately apparent that the maximum possible performance with the maximum number of probes  $p = 500$  corresponds to a normalized root-mean-square error of 16% and 24% and correlations of 0.99 and 0.96 for  $u'$  and  $v'$ , respectively. This highlights the challenge of applying a reduced order model such as POD to a turbulent flow as in the present case; resulting in many required modes to capture the fluctuations.

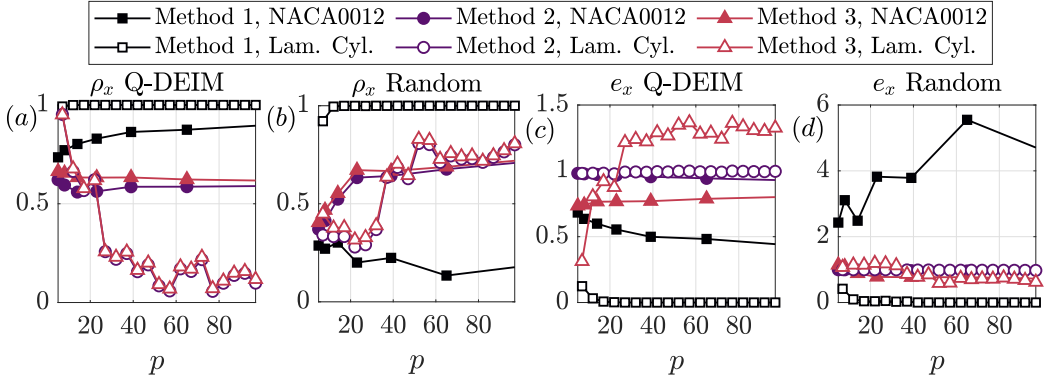


**Figure 7.** Normalized root mean square error (a,b) and correlations (c,d) vs number of probes for  $u'$  (a,c) and  $v'$  (b,d) using the Q-DEIM for probe placement applied to the testing data via method 1 (squares), method 2 (circles), method 3 (triangles) and POD (diamonds). The number of reconstruction modes  $k$  is equal to the number of probes used  $p$ . The POD-based reconstructions are obtained via equation 2 using the coefficients from projecting the full velocity fields into the global basis.



**Figure 8.** Normalized root mean square error (a,b) and correlations (c,d) vs number of probes for  $u'$  (a,c) and  $v'$  (b,d) using random probe placement applied to the testing data from via method 1 (squares), method 2 (circles), method 3 (triangles) and POD (diamonds). The number of reconstruction modes  $k$  is equal to the number of probes used  $p$ . The POD-based reconstructions are obtained via equation 2 using the coefficients from projecting the full velocity fields into the global basis.

The results in figure 7 highlight the differences in the three proposed methods when the probes are placed using Q-DEIM. For this placement, the sparse recovery via method 1 significantly outperforms the other methods. At 500 probes, method 1 was able to recover the spatio-temporal fluctuations to within 25% and 40% for  $u'$  and  $v'$ , respectively, with correlations exceeding 90%. The relative superiority of method 1 is enforced by construction, as the Q-DEIM ensures that method 1 produces coefficients that most efficiently sample the POD basis. Methods 2 and 3 show mixed results. These approaches were only able to recover the fluctuations with approximately 75% error for  $u'$  and even exceeding 100% for  $v'$  with correlations in the approximate range of 30-60%. Both methods produce reconstructions with very similar correlations. This is unsurprising, as both methods use the same global probe basis. Method 3, on the other hand, effectively contains a re-scaling of the coefficients by a factor of  $\Sigma_p^{-1}\Sigma_g$ , resulting in notable differences in the root-mean-square error. For  $u'$ , this leads to comparatively lower error using method 3. However for  $v'$ , method 3 quickly compares worse to method 2 beyond 14 probes.

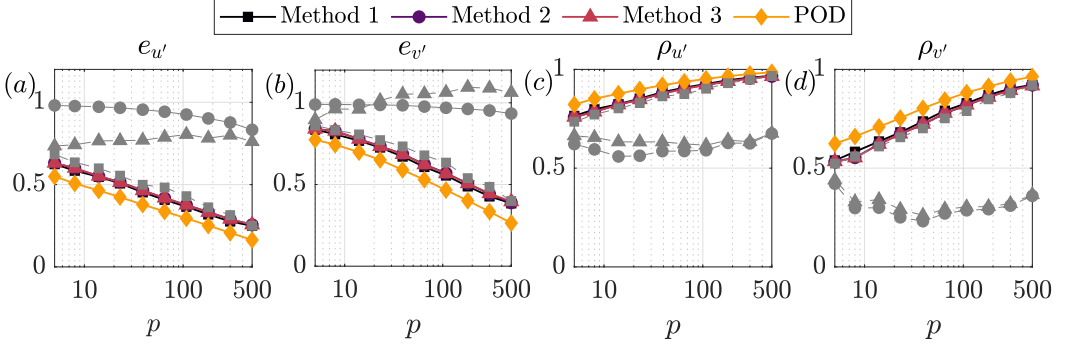


**Figure 9.** Comparison of normalized reconstruction correlation (a,b) and root mean square error (c,d) for the reconstruction of  $u'$  in the present case (filled symbols,  $x = u'$ ) to the laminar cylinder at  $Re_D = 100$  from the DNS of Brunton and Kutz (2019) (unfilled symbols,  $x = \omega$ ) vs number of probes using Q-DEIM (a,c) and random placement (b,d) via method 1 (squares), method 2 (circles), and method 3 (triangles).

We remark that although the experimental data presented is inherently limited in its accuracy and contains noise, no systematic variation of noise on the fidelity of the reconstructions is presented. This is due to the underlying assumption that the measurement noise is largely confined to low-energy high-rank structure that was explicitly treated for in the data post-processing (section 2.2). The cut-off mode number for treating the random noise was found to be higher than the maximum number of modes tested for the sparse reconstructions. Nevertheless, the limited sub-pixel accuracy of the PIV (Adrian et al., 2011) should be considered in the interpretation of the results. As can be seen in figure 8, the vertical velocity fluctuation  $v'$  was found to be consistently more difficult to reconstruct across methods. This is likely due to the reduced signal-to-noise ratio of  $v'$  in the measurements, for which the vertical fluctuations span a smaller range of pixel values (average deviation  $\sigma_{std,v} = 1.2$  pixels) than the horizontal ( $\sigma_{std,u} = 1.8$  pixels) and are therefore more greatly effected by sub-pixel accuracy in the PIV.

An interesting shift in the results occurs for when the probes are placed randomly as shown in figure 8. The sparse recovery approach of method 1 now performs significantly worse than methods 2 and 3. This is due to the fact that the random probes have some locations outside of the shear layer. These locations contribute very little variance to the global POD basis, and when they are inverted (equation 5) the singular values become large and amplify the noise as a result. This points to the necessity that reconstructions obtained via method 1 must avoid probe locations with very little variance across the global basis. For the case of a separated aerofoil, this means avoiding probes outside of the separated region. On the other hand, the random placement does not greatly impact the extended-POD based methods 2 and 3. Both methods perform similarly between the random and Q-DEIM placements; suggesting they may be more robust to arbitrary placement. Methods 2 and 3 do not suffer from the same effect as method 1 because they use an independent global probe basis that is tailored to the sets of probe locations using individual PODs obtained *a-priori*.

The extended-POD methods appear more robust for the present PIV data of a separated aerofoil at high Reynolds number, but a natural question to ask is whether such robustness extends to the laminar case. To assess this, we use the DNS data of a laminar cylinder in crossflow at  $Re_D = 100$  provided by Brunton and Kutz (2019) and perform sparse reconstructions of the vorticity as described by Manohar et al. (2018). A brief outline of the analysis is as follows. The laminar data consists of 151 snapshots for which the mean is subtracted to isolate the fluctuations. The first 100 snapshots are used to generate the global basis and the global probe basis using POD for both Q-DEIM and random placements (see



**Figure 10.** Normalized root mean square error (a,b) and correlations (c,d) vs number of probes for  $u'$  (a,c) and  $v'$  (b,d) using SNN refinement and the Q-DEIM for probe placement applied to the testing data from via method 1 (squares), method 2 (circles), method 3 (triangles) and POD (diamonds). The linear results are shown in gray dashed lines with corresponding symbols. The number of reconstruction modes  $k$  is equal to the number of probes used  $p$ . The POD-based reconstructions are obtained via equation 2 using the coefficients from projecting the full velocity fields into the global basis.

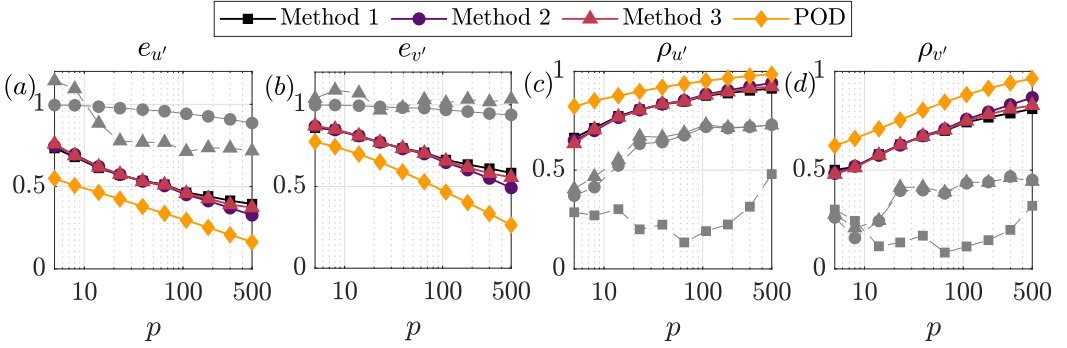
figure 1c). The remaining 51 snapshots are used for assessing the reconstructions quantified using the same metrics as the present study (section 3.5).

The comparison to the laminar cylinder is presented in figure 9. The results from the laminar cylinder are shown with unfilled symbols and from the present study for  $u'$  with filled symbols using Q-DEIM and random placement. Method 1 outperforms methods 2 and 3 for reconstructing the laminar cylinder for all probe numbers and locations. Interestingly, for very few probes ( $p = 7$ ) methods 2 and 3 correlate reasonably well for the Q-DEIM placement around the cylinder. However, as the number of probes increases the correlation rapidly drops off. It is observed that as the number of probes increases, the optimal locations for the probes shift from within the wake region to concentrating near the leading edge of the cylinder. This appears to negatively impact the ability for extended POD to capture the correlation between probes and flow. The random placement however appears to give improved correlations for methods 2 and 3 compared to the Q-DEIM; indicating that the extended POD benefits from the dispersed probe locations.

The comparison between the reconstructions in the present study and with the laminar cylinder indicates that the compressed sensing methodology is superior for systems captured by relatively few modes. This is because the compressed sensing methodology does not rely on underlying correlations between probes, but rather how their location contributes to the global basis and its corresponding sparse inversion. By contrast, the improvement in the extended POD based methods between Q-DEIM and randomly placed probes suggests that the extended POD relies more heavily on the underlying spatial correlations captured by the POD of the global probe basis. Indeed, it was confirmed (not shown) that spatially concentrated probes in the wake of the cylinder yielded significantly poorer reconstructions than spatially dispersed configurations. This effect may explain why the extended POD methods are more robust for the separated aerofoil; the spatially dispersed probes capture a portion of the underlying spatial correlations whilst the compressed sensing approach of method 1 may be sensitive to specific regions with low variance across the global basis.

#### 4.2. Non-Linear Reconstruction

The root-mean-square error and correlations using the non-linear SNN for refinement are presented in figures 10 and 11 for the Q-DEIM and random placements, respectively. In both figures, the gray



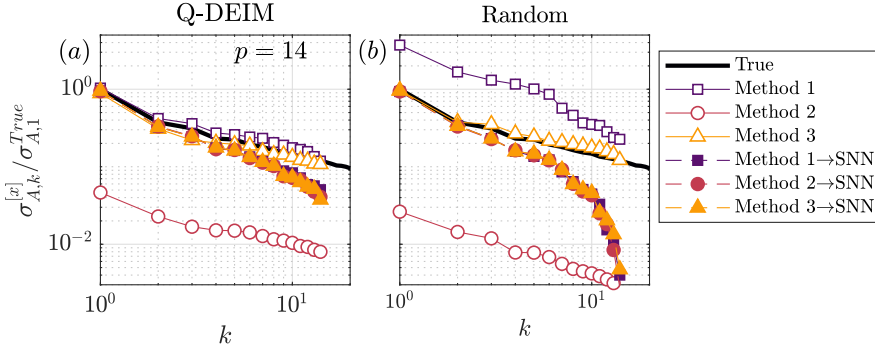
**Figure 11.** Normalized root mean square error (a,b) and correlations (c,d) vs number of probes for  $u'$  (a,c) and  $v'$  (b,d) using SNN refinement and random probe placement applied to the testing data from via method 1 (squares), method 2 (circles), method 3 (triangles) and POD (diamonds). The linear results are shown in gray dashed lines with corresponding symbols. The number of reconstruction modes  $k$  is equal to the number of probes used  $p$ . The POD-based reconstructions are obtained via equation 2 using the coefficients from projecting the full velocity fields into the global basis.

dashed lines indicate the reconstructions without applying the SNN. Regardless of which linear method is used to supply the input for the SNN, the output coefficients perform nearly identically for each placement. The Q-DEIM placement appears to slightly outperform the random placement; pointing to the importance of the underlying correlations between the input and output coefficients on their performance of the SNN. Only for very large numbers of probes using the random placement can some small differences arise between input methods (figure 11). In any case, the SNN outperforms all of the linear methods both in terms of correlation and root-mean-square-error. The SNNs are trained using a loss function that is designed to minimize the difference between the known POD coefficients and the output of the SNN. It is therefore unsurprising that it outperforms the linear counterparts, although it does not fully recover the true underlying POD coefficients for either placement.

In order to understand more about how the SNN is performing compared to the linear methods, the singular values from the estimated coefficients  $\mathbf{A}_{DYN}$  are extracted. This separates the orthogonal temporal part of the coefficients in order to reveal the *effective* singular values used for the full reconstruction. This is presented for the linear methods and with SNN refinement for the case  $p = 14$  in figure 12 for Q-DEIM and random placement. The results are seemingly counter-intuitive. For both placements, the linear method 3 appears to most closely approximate the true singular values of the validation run; but this does not mean that it will necessarily lead to a better reconstruction. In fact, this result is unsurprising as method 3 is constructed to have a quasi-orthogonal set of temporal probe modes that are re-scaled using the global ones, and therefore will most closely resemble the true singular values upon extraction from the coefficients. The other two methods on the other hand do nothing explicit to regulate the orthogonal basis of the coefficients, and this is manifested by singular values in figure 12 that depart from the true ones.

Regardless of the linear method used, the singular values from the output of the SNN collapse together and consistently underestimate the true ones. It is important to note that no orthogonal regularization was imposed on the SNN in its construction. Instead it simply attempts to minimize the loss function between the training inputs and the training data by adjusting the weights. As the root-mean-square error and correlations unilaterally improve via the SNN as shown in figures 10 and 11, the results of figure 12 imply that recovering the underlying temporal correlation (the phase information of the temporal modes) is inherently more important than recovering the precise singular values (at least





**Figure 12.** Singular values  $\sigma_{A,k}$  extracted from the estimated coefficients  $A_{DYN}$  normalized by the leading order singular value of the true coefficients for  $p = 14$  probes using linear methods (unfilled symbols) and SNN refinement (filled symbols) via Q-DEIM (a) and random placement (b).

for minimizing the imposed loss function). This becomes increasingly less important as the mode number increases, as evidenced by the sharp decrease in singular values with increasing  $k$ . This is shown for the case of  $p = 14$ , however similar results were found for all probe numbers tested.

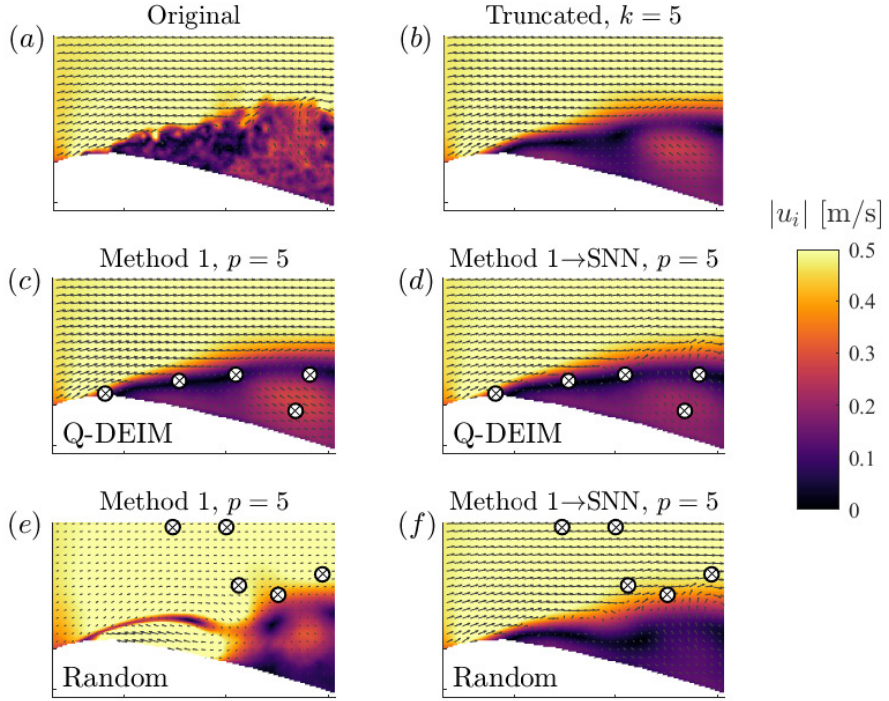
## 5. Conclusions

We have presented three distinct linear reconstruction methodologies and a shallow neural network (SNN) for non-linear refinement to obtain state estimations of separated turbulent flow over a NACA 0012 aerofoil at  $\alpha = 12^\circ$  and  $Re_c = 75000$  using time-resolved PIV from limited probes. Each methodology was trained using 9000 samples of PIV training data and instantaneous sample prediction was tested using all 5468 samples of an independent time-resolved data set. This is the first systematic investigation of sparse reconstructions for a moderately high Reynolds number advective turbulent flow obtained via experiment.

The reconstructions were first tested through the use of linear methods alone. It was found that the performance of the linear methods depended largely on the choice of placement. When the placement was chosen carefully using the Q-DEIM, the compressed-sensing approach of method 1 greatly outperformed the extended POD based approaches of methods 2 and 3 (figure 7). When the probes were placed randomly (figure 8), it was found that the extended POD methods performed more favorably than the compressed sensing approach of method 1. This was due to locations within the global POD basis outside of the shear layer with small variance causing errors in the inversion of the sparse basis for method 1. This is illustrated clearly in figure 13c and e, showing how the reconstruction is affected when probe locations reside in the free stream.

In order to investigate how the methods compare for the laminar case, the extended POD methods were applied to the DNS of laminar flow over a cylinder at  $Re_D = 100$  of Brunton and Kutz (2019) (figure 9). It was found that the extended POD based methods rely on the underlying correlations of their global probe POD basis. When the probes were grouped closely together, the reconstructions performed more poorly. When the probes were more dispersed, the autocorrelation underlying the global probe POD basis was able to capture more features. Despite this, the compressed sensing approach of method 1 outperformed the extended POD approaches for both Q-DEIM and random placement in the laminar case. As the laminar data did not contain measurement noise and was captured by relatively few modes, small variances within the global basis were comparatively less problematic.

Non-linear refinement of the estimated coefficients was tested through the application of an SNN using the various linear methodologies as inputs. The performance of the SNN was found to be nearly identical regardless of the linear method used and slightly more favorable for Q-DEIM placement



**Figure 13.** Original (a), POD-based (b), linear via method 1 (c,e), and non-linear (d,f) reconstructions of the total velocity at one arbitrary instant using 5 probes and Q-DEIM (c,d) and random (e,f) placement with every fifth velocity vector shown for clarity. Vectors are scaled automatically with respect to their individual fields and not across panels.

(figures 10 and 11). In all cases, the SNN improved the root-mean-square error and correlations of the reconstructions. Interestingly, it was found that the output coefficients of the SNN did not contain singular values that most closely matched the true singular values (figure 12). Instead, the SNN naturally placed more emphasis on the phase information of the underlying temporal correlations; leading to improved reconstruction performance.

The implications for sparse reconstructions in a rapidly evolving turbulent flow (the most common type of flow in engineering applications) based on the findings of the present study can be summarized as follows. Firstly, several hundred modes (and therefore probes) are required to capture the velocity fluctuations using POD as the global basis. This implies an extensive full-field data set must be tabulated *a-priori*, in this case 18 times more samples than the maximum modes used for reconstruction (9000 samples and 500 modes). Many fewer probes may be used for reconstructions depending on the desired accuracy for the application of interest (figure 7). Second, the placement of the probes depends on the reconstruction method of choice. The compressed sensing approach based on sparsity (method 1) is ideal but relies on chosen locations within the flow with high variability. The extended-POD approaches (methods 2 and 3) are less optimal however are comparatively robust and instead depend on being spatially dispersed. Finally, the non-linear refinement via the SNN appears to equalize the linear approaches, with only a small improvement when placed favorably. This result begs the question: why should one bother with linear methods at all? The answer again depends on the application of interest. If the objective is to reconstruct the full-field as accurately as possible our results indicate one should employ as many sensors as possible and use any linear method for the inputs to the SNN. This requires training the network with the entirety of the full-field data obtained *a-priori*. If the application instead

does not require very high accuracy, the expensive task of training a neural network may be avoided in which case the differences between the linear methods become relevant once again.

This study has demonstrated the capability of linear and non-linear methods to recover instantaneous reconstructions of the full experimental velocity fields using sparse measurements. This is visualized in figure 13. Avenues for future work include sparse reconstructions with flow control within the sensing region (which would require a distinct global basis depending on if the flow control is active or not). A further exploration of sensitivity for variations in the system e.g. changes in the freestream velocity  $U_\infty$  or angle of attack  $\alpha$  is warranted, as the cost of training separate global bases for variations in external conditions is undesirably high (for the present study, this implies separate training data sets on the order of thousands of samples each for each  $\alpha$ ). Future investigations may also explore variations of the SNN architecture, which might benefit from e.g. splitting the input coefficients into singular values and orthogonal modes and training the modes with orthogonal regularization. One challenge with promising potential is to explore obtaining a greater number of outputs (modes) from the SNN than inputs; potentially achieving reconstructions that are more accurate than the rank-truncated POD itself.

**Acknowledgments.** The authors wish to thank Dr. John Lawson for his technical assistance with the PIV code.

**Funding Statement.** We gratefully acknowledge financial support from EPSRC (Grant ref No: EP/R010900/1) and EU H2020 project HOMER (Grant Ref No: 769237).

**Competing Interests.** None

**Data Availability Statement.** Code containing the interactive plots and the presented methodologies will be made available upon publication at <https://doi.org/10.5258/SOTON/D1752>.

**Ethical Standards.** The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

**Author Contributions.** Conceptualization: B.G.; D.W.C. Methodology: D.W.C.; F.D.V. Data curation: R.S.; D.W.C. Data visualisation: D.W.C. Writing original draft: D.W.C; B.G. All authors approved the final submitted draft.

## References

- Adrian, L., Adrian, R. J., and Westerweel, J. (2011). *Particle image velocimetry*. Number 30. Cambridge university press.
- Adrian, R. J. and Moin, P. (1988). Stochastic estimation of organized turbulent structure: homogeneous shear flow. *Journal of Fluid Mechanics*, 190:531–559.
- Atkinson, A. C. and Donev, A. N. (1992). *Optimum experimental designs*.
- Bai, Z., Wimalajeewa, T., Berger, Z., Wang, G., Glauser, M., and Varshney, P. K. (2015). Low-dimensional approach for reconstruction of airfoil data via compressive sensing. *AIAA journal*, 53(4):920–933.
- Bao, Y., Shi, Z., Wang, X., and Li, H. (2017). Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring. *Structural Health Monitoring*, 17(4):823–836.
- Barrault, M., Maday, Y., Nguyen, N. C., and Patera, A. T. (2004). An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique*, 339(9):667–672.
- Berkooz, G., Holmes, P., and Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575.
- Bořée, J. (2003). Extended proper orthogonal decomposition: a tool to analyse correlated events in turbulent flows. *Experiments in fluids*, 35(2):188–192.
- Brajard, J., Carassi, A., Bocquet, M., and Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the lorenz 96 model. *arXiv preprint arXiv:2001.01520*.
- Brunton, S. L. and Kutz, J. N. (2019). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- Brunton, S. L., Noack, B. R., and Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52:477–508.
- Callahan, J. L., Maeda, K., and Brunton, S. L. (2019). Robust flow reconstruction from limited measurements via sparse representation. *Physical Review Fluids*, 4(10):103907.
- Candès, E. J. et al. (2006). Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain.
- Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30.
- Chaturantabut, S. and Sorensen, D. C. (2010). Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764.

- Cox, D. R. and Reid, N. (2000). *The theory of the design of experiments*. CRC Press.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314.
- Discetti, S., Raiola, M., and Ianiro, A. (2018). Estimation of time-resolved turbulent fields through correlation of non-time-resolved field measurements and time-resolved point measurements. *Experimental Thermal and Fluid Science*, 93:119–130.
- Dong, W., Zhang, L., Shi, G., and Wu, X. (2011). Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on image processing*, 20(7):1838–1857.
- Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306.
- Drmac, Z. and Gugercin, S. (2016). A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions. *SIAM Journal on Scientific Computing*, 38(2):A631–A648.
- Erichson, N. B., Mathelin, L., Yao, Z., Brunton, S. L., Mahoney, M. W., and Kutz, J. N. (2019). Shallow learning for fluid flow reconstruction with limited sensors and limited data. *arXiv preprint arXiv:1902.07358*.
- Guastoni, L., Güemes, A., Ianiro, A., Discetti, S., Schlatter, P., Azizpour, H., and Vinuesa, R. (2020). Convolutional-network models to predict wall-bounded turbulence from wall quantities. *arXiv preprint arXiv:2006.12483*.
- Gunes, H., Sirisup, S., and Karniadakis, G. E. (2006). Gappy data: To krig or not to krig? *Journal of Computational Physics*, 212(1):358–382.
- Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554.
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257.
- Hosseini, Z., Martinuzzi, R. J., and Noack, B. R. (2015). Sensor-based estimation of the velocity in the wake of a low-aspect-ratio pyramid. *Experiments in Fluids*, 56(1):13.
- Iyer, S. R., An, U., and Subramanian, L. (2020). Forecasting sparse traffic congestion patterns using message-passing rnns. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3772–3776.
- Jayaraman, B., Al Mamun, S., and Lu, C. (2019). Interplay of sensor quantity, placement and system dimension in pod-based sparse reconstruction of fluid flows. *Fluids*, 4(2):109.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lasagna, D., Orazi, M., and Iuso, G. (2013). Multi-time delay, multi-point linear stochastic estimation of a cavity shear layer velocity from wall-pressure measurements. *Physics of Fluids*, 25(1):017101.
- Liu, Y., Zhang, G., and Xu, B. (2017). Compressive sparse principal component analysis for process supervisory monitoring and fault detection. *Journal of Process Control*, 50:1–10.
- 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.
- Milano, M. and Koumoutsakos, P. (2002). Neural network modeling for near wall turbulent flow. *Journal of Computational Physics*, 182(1):1–26.
- Nair, N. J. and Goza, A. (2019). Leveraging reduced-order models for state estimation using deep learning. *arXiv preprint arXiv:1912.10553*.
- Otto, S. E. and Rowley, C. W. (2021). Inadequacy of linear methods for minimal sensor placement and feature selection in nonlinear systems: a new approach using secants. *arXiv preprint arXiv:2101.11162*.
- Picard, C. and Delville, J. (2000). Pressure velocity coupling in a subsonic round jet. *International Journal of Heat and Fluid Flow*, 21(3):359–364.
- Raiola, M., Discetti, S., and Ianiro, A. (2015). On piv random error minimization with optimal pod-based low-order reconstruction. *Experiments in Fluids*, 56(4):75.
- 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.
- Roman-Rangel, E., Odobez, J.-M., and Gatica-Perez, D. (2012). Assessing sparse coding methods for contextual shape indexing of maya hieroglyphs. *Journal of Multimedia*, 7(ARTICLE):179–192.
- Sirovich, L. (1987). Turbulence and the dynamics of coherent structures. iii. dynamics and scaling. *Quart. Appl. Math.*, 45:583–590.
- Sun, L. and Wang, J.-X. (2020). Physics-constrained bayesian neural network for fluid flow reconstruction with sparse and noisy data. *arXiv preprint arXiv:2001.05542*.
- Taira, K., Brunton, S. L., Dawson, S. T., 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.
- Thomas, A., Petridis, M., Walters, S., Gheytaasi, S., and Morgan, R. (2017). Two hidden layers are usually better than one. *Communications in Computer and Information Science: Engineering Applications of Neural Networks*, 744.
- Troshin, V. and Seifert, A. (2019). Data-driven modeling of pitching and plunging tandem wings in hovering flight. *Experiments in Fluids*, 60(3):39.
- Van Loan, C. F. and Golub, G. H. (1983). *Matrix computations*. Johns Hopkins University Press Baltimore.
- Willcox, K. (2006). Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & fluids*, 35(2):208–226.

- Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2008). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227.
- Yildirim, B., Chrysostomidis, C., and Karniadakis, G. (2009). Efficient sensor placement for ocean measurements using low-dimensional concepts. *Ocean Modelling*, 27(3-4):160–173.