


**Sparse sensor-based cylinder flow estimation using artificial neural networks**Kevin H. Manohar <sup>\*</sup>, Chris Morton,<sup>†</sup> and Paul Ziadé <sup>‡</sup>*Department of Mechanical and Manufacturing Engineering University of Calgary 2500  
University Drive NW Calgary, Alberta, Canada T2N 1N4* (Received 27 December 2020; accepted 25 January 2022; published 17 February 2022)

Advancements in the estimation of fluid flows from remote sensors have rapidly progressed through the emerging use of artificial neural networks (ANNs). However, further development of the robustness of estimation strategies warrants their validation on more complex benchmark cases. This work presents a family of challenge flow cases for the advancement of sparse sensor-based flow estimation techniques. The flows under investigation are the laminar wake development and wake interactions of two cylinders in proximity. Vortex shedding occurs from both the cylinders, and the flow regime is defined based on the nature of the interaction of the two wakes. Two flow regimes are studied: one where the wake interaction yields a low-frequency beating behavior, known as modulated-periodic (MP), and another where the wake is characterized by highly nonlinear interactions that result in irregular flow behavior, called flip-flop (FF). The estimation of the velocity flow fields is performed in a proper orthogonal decomposition (POD)-based low-dimensional subspace using sparsely placed pressure sensors. The estimation performance is assessed for modified linear and quadratic stochastic estimators (LSE/QSE-POD), shallow feedforward neural networks (FFNN), and long short-term memory (LSTM) networks. For the MP regime, the results show that both ANNs significantly outperform the LSE/QSE-POD approaches both instantaneously and in an average least-square sense. For the FF regime, the LSTM-based estimator significantly outperforms the other approaches in both the time and frequency domains. For both regimes, it is found that the trained LSTM-based estimators are far more robust to sensor noise and are relatively insensitive to sensor placements considered in this study. Unlike the single cylinder in cross-flow, the wake interactions in the two-cylinder flow regimes investigated are characterized by multiscale spatiotemporal phenomena, offering ample complexity for systematic development of techniques in sensor-based estimation, scale-extraction, and model prediction.

DOI: [10.1103/PhysRevFluids.7.024707](https://doi.org/10.1103/PhysRevFluids.7.024707)**I. INTRODUCTION**

The estimation of fluid flow fields from remote sensors is a tool with important practical utility in fluid mechanics. It has applications in flow control and sensing in marine, aerospace, and biomedical sectors [1]. One major challenge in this area is the ability to sufficiently estimate the flow fields from sensors that are sparsely placed either within or removed from the spatial domain of interest. Instead of attempting to estimate the entire high-dimensional state space, a more efficient method is to exploit the large-scale coherent structures that dominate mixing behavior, fluctuations in body forces, and represent a substantial portion of the fluctuating kinetic energy in the flow. The reduction

---

<sup>\*</sup>Corresponding author: [kevin.manohar@ucalgary.ca](mailto:kevin.manohar@ucalgary.ca)

<sup>†</sup>[chris.morton@ucalgary.ca](mailto:chris.morton@ucalgary.ca)

<sup>‡</sup>[paul.ziade@ucalgary.ca](mailto:paul.ziade@ucalgary.ca)

of the flow state to a low-dimensional subspace significantly reduces computational effort [2], and more importantly enhances reconstruction quality by effectively filtering incoherent, and typically small-scale dynamics.

Stochastic estimation methods since the pioneering work of Adrian and Moin [3] have now matured and have become the standard toolkit for the estimation of coherent flow dynamics in free and wall-bounded flows. A review of these methods can be found in Clark *et al.* [4]. The inclusion of a low-dimensional subspace into the stochastic estimation models using the proper orthogonal decomposition (POD) modes was first introduced by Bonnet *et al.* [5]. Artificial neural networks (ANNs) have been previously used to estimate low-dimensional states using the POD [6,7]. Cohen *et al.* [6] found that a standard feedforward neural network (FFNN) was able to outperform the linear and quadratic stochastic estimators (LSE/QSE-POD) with multitime delays for a four-sensor configuration.

Approaches to improve estimators by incorporating time history via multi-time-delay techniques as proposed by Durgesh and Naughton [8] and expanded on by Lasagna *et al.* [7] include more expressive models that are capable of capturing high degrees of nonlinearity. Lasagna *et al.* [7] demonstrated that nonlinear techniques better reconstruct complex spectral flow features compared with linear approaches, and that the coupling with time-delayed sensor signals improves estimation performance in some cases.

More recently, the rapid advancements in regression techniques emerging from the machine learning community has accelerated progress in fluid dynamics, among many other prominent fields where function approximators are frequently used. The emergence of long short-term memory (LSTM) networks [9] in the fluid dynamics community is duly noted. The LSTM network is a form of a recurrent neural network (RNN) with a predictive power that is conditional on the context and history of input data sequences. RNNs suffer from vanishing (or exploding) gradients for long-term dependencies [10]. The LSTM circumvents this issue through the use of gates, facilitating its success in accomplishing complex tasks such as sequential modeling [11], handwriting recognition [12], anomaly detection [13], and machine translation [14]. They have been recently used to construct reduced-order models in the context of dynamical systems [15,16] as well as in fluid mechanics [17–22].

ANNs have been used to predict flow fields based on sparsely placed sensors [23–26]. Deng *et al.* [24] used the LSTM-POD to make sensor-based time-resolved velocity field predictions from non-time-resolved particle image velocimetry measurements of a turbulent inverted flag flow exhibiting strong flapping periodicity at a Reynolds number of  $Re = 6200$ . The results suggested that their LSTM network was able to predict the reduced-order flow fields with remarkable accuracy. After training on non-time-resolved data sampled at 5 Hz, they were able to produce a time-resolved reconstruction at 2000 Hz. The study of Deng *et al.* [24] showed great promise in the application of the LSTM-POD architecture for the prediction of flow fields. This has motivated the present work, where additional underlying issues remain to be addressed, such as (i) the more practical use of pressure sensors within the domain rather than (virtual) velocity sensors extracted directly from the PIV flow field, (ii) the remote placement of sensors far from the energetic fluctuations to address situations with limited access to install sensors, and (iii) exploring and assessing the expressivity of the LSTM architecture to predict complex dynamics of high-dimensional systems, i.e., training on more than 50 modes within the POD subspace using a single LSTM model.

Typically, a low-order representation (LOR) of the flow can be obtained by truncating the POD series to capture a specified amount of energy. While considerable success has been shown with this energy-based mode selection criterion, the exclusion of less-energetic modes can result in insufficient reconstruction of dynamically relevant flow features. The incorporation of these high-order modes can be critical to reconstruct the dynamics of flows exhibiting complex multiscale phenomena, which form the majority of flows of practical use. However, high-order modes can also have an element of stochasticity, which can lead to problems such as overfitting. Erichson *et al.* [23] suggest a method to reduce the low-order reconstruction error by obtaining an estimation basis

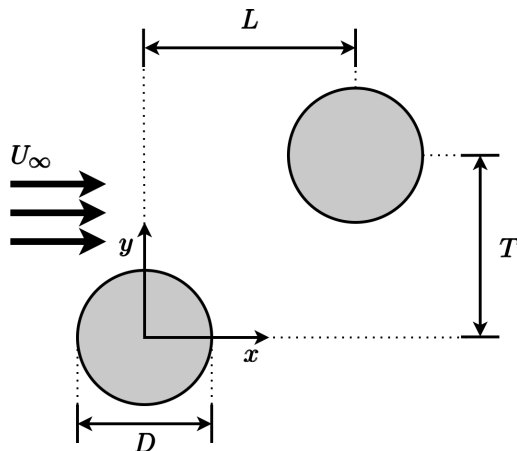


FIG. 1. Geometrical parameters defining the two-cylinder configuration in the present work.

that is informed by both the snapshot flow and the sensor data. It is claimed that their method may potentially improve classical POD in terms of obtaining physically interpretable modes.

#### A. Flow cases

Of interest in the present work is to introduce a challenging data set for the purpose of accelerating the advancement of sensor-based estimation techniques. As noted by Kutz [27], challenge data sets are an important factor in ANN development and are lacking in the fluid mechanics community. The challenge flow case selected in the present work is that of a two-cylinder geometry, which provides a variety of complex unsteady flow phenomena [28]. A significant amount of literature [28–39] exists for verification and validation purposes, thus serving as an ideal test case for the development of sensor-based flow estimation strategies. Herein, a brief overview of this geometry and the flow development is provided, while additional details can be found in Sumner [38].

Cylindrical structures in a cross-flow are found in a wide range of engineering applications such as bridges, antennae, risers, heat exchangers, power lines, and chimneys. In many of these applications, multiple cylindrical structures are placed in proximity, leading to a much more complex flow compared to the single-cylinder case. This flow is often unsteady, involving the periodic formation and shedding of vortices from the structure(s), which induces significant fluctuating loading. Unwanted flow-induced vibrations of the structure may occur due to this shedding phenomenon or the interactions between vortices shed from multiple bodies in proximity to one another. Thus, it is critical for engineers and designers to have a detailed understanding of the geometrical and flow parameters that influence the vortex shedding process.

For a single isolated cylinder, substantial research efforts [40] have enabled a detailed classification of the boundary layer, shear layer, and wake development across a wide range of Reynolds numbers. For two cylinders in proximity, the additional geometrical parameters  $T/D$  and  $L/D$  (see Fig. 1) defining the relative position of the cylinders in the flow domain give rise to a greater number of possible flow regimes for a given Reynolds number. This presents a significant challenge to classify the flow regime due to a coupling of Reynolds number and geometrical effects.

Sumner [38] describes several approaches to understand the flow behavior. One intuitive approach is to classify the flow based on the type of interference occurring between the upstream and downstream cylinders [28]:

(i) wake interference, when the downstream cylinder is immersed in the wake of the upstream cylinder,

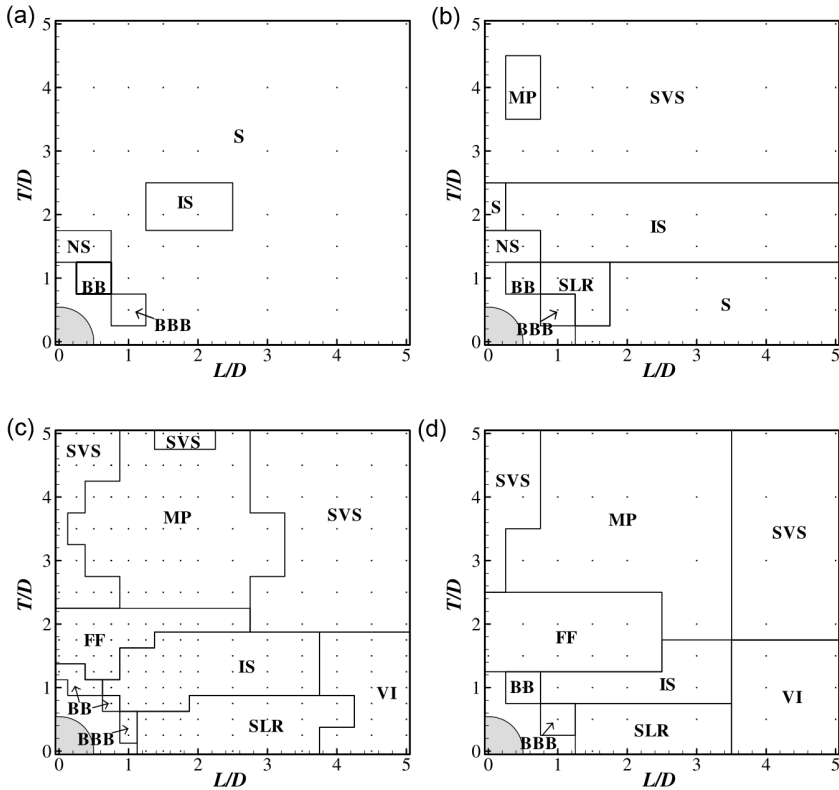


FIG. 2. Flow patterns for the flow past two circular cylinders, (a)  $Re_D = 40$ , (b)  $Re_D = 50$ , (c)  $Re_D = 100$ , (d)  $Re_D = 160$  (reprinted from Ref. [37] with permission from Elsevier). A full description of the various flow regimes can be found in Lee and Yang [37].

(ii) proximity interference, when the two cylinders are located close to one another but there is no immersion within the wake of the upstream cylinder,

(iii) no interference, where the flow development over each cylinder is approximately unaffected from that of a single isolated cylinder.

Further refinement of this approach would enable each regime (e.g., wake interference) to be specified further based on the nature of the interference effects. This has been achieved through direct interpretation of experimental and numerical visualizations and measurement data characterizing the vortex shedding frequencies, mean and fluctuating drag, and mean and fluctuating lift forces.

To provide a greater sense of the challenge in developing a complete understanding of the flow development for two cylinders in proximity, Lee and Yang [37] have shown that at low Reynolds numbers ( $Re_D < 160$ ), up to ten distinct flow regimes can be identified through visualization of their simulation results. To make matters more difficult, the geometrical ( $T/D$  and  $L/D$ ) boundaries between the flow regimes change significantly with relatively small changes in Reynolds number, as seen, for example, in the flow regime maps provided in Fig. 2. More recent experimental and numerical studies as well as comprehensive reviews (e.g., Zhou and Alam [39]) continue to indicate that the multicylinder problem is a major challenge not only owing to the complexity and number of flow regimes, but also due to the appearance of bistable, tristable, and even quadristable flow states.

Coupling the complexity of the two-cylinder flow regimes with the simplicity of the geometry presents a platform to systematically test, develop and improve estimation algorithms. We demonstrate the scalability of the flow complexity by applying estimation techniques to two flow cases: modulated-periodic (MP) and flip-flop (FF). The distinct and complex dynamics that characterize these regimes are discussed, which form the crux of the estimation challenge.

## B. Objectives

In review of the relevant flow estimation, ANNs, and two-cylinder flow development literature, the objectives of the present study are as follows:

(i) Present a comprehensive description of both two-cylinder flow cases studied, modulated periodic (MP) and flip-flop (FF), and analyze the wake interactions based on the dominant POD modes.

(ii) Provide a robust methodology for the incorporation of a POD subspace within an FFNN or LSTM network for the purposes of sparse pressure sensor-based estimation of velocity fields.

(iii) Compare estimator performance for LSTM, FFNN, QSE-POD, and LSE-POD approaches.

The main contributions of this work are summarized below. The present study is the first to:

(i) Study the application of POD to two-cylinder flows, which includes an analysis of the modes and dynamics characterizing the wake development and wake interaction. It is shown that POD can be used to separate the flow into a reduced set of coherent energetic motions, thus lowering the dimension of the estimation problem and significantly simplifying the estimation architecture;

(ii) Employ the LSTM-POD architecture to establish correlations between pressure sensors and (low-order) velocity fields. The expressivity of the architecture is assessed by studying its ability to reconstruct a large number of modes that govern the complex wake interactions. Robustness of the LSTM-based estimator to sensor noise and various sensor placement configurations is also assessed—both of which are important topics and have not been addressed in the fluid dynamics literature in the context of LSTM-based estimation.

## II. METHODOLOGY

Section II A presents the physical geometry and flow simulation setup, followed by an outline of the proper orthogonal decomposition. Section II B presents the flow estimation methodology, including details on the types of estimators used in this work as well as efforts to address model generalizability.

### A. Flow setup and analysis

#### 1. Flow simulation

In this study, two cases of the two-cylinder configuration are considered. The Reynolds number is held constant at  $Re_D = 160$ . The first case consists of a geometric configuration with  $T/D = 2$  and  $L/D = 2$ , and the second case corresponds to  $T/D = 2$  and  $L/D = 0$ . According to Lee and Yang [37], these two cases would exhibit modulated-periodic (MP) and flip-flopping (FF) behavior, respectively (see Fig. 2).

Two-dimensional flow field data are produced using highly resolved incompressible fluid flow simulations using the open-source software, OpenFOAM v8 [41]. The time-resolved flow is computed using a cell-centered finite volume method. A second-order total variation diminishing scheme and a second-order backward implicit time-stepping scheme are used for the convective and temporal terms, respectively. The pressure-momentum coupling is addressed using the pressure-implicit with splitting of operators algorithm. The system of equations is solved using a geometric-algebraic multigrid procedure. A no-slip boundary condition is applied at the cylinder surfaces. The inlet is assigned a laminar inflow with prescribed velocity. A convective outflow is defined at the outlet. All boundary conditions can be found in Table I.

TABLE I. Computational boundary conditions.

Boundary	Velocity condition	Pressure condition
Inlet	Laminar inflow	Zero gradient
Outlet	Convective	Zero gradient
Front/back	Empty	Empty
Top/bottom	Slip	Slip
Cylinder surface	No-slip	Zero gradient

Results presented in this study are obtained from highly resolved simulations with approximately 700 000 hexahedral elements (see inset in Fig. 3). The domain has a total length of  $38D$  in the direction of the flow and  $20D$  in the spanwise direction (see Fig. 3). Agreement with previous studies is addressed in Sec. III. Full field snapshots are output at a rate of approximately 20 snapshots per shedding cycle. A fixed time step  $\Delta t = 10^{-5}$  s is used, resulting in an average Courant number below 0.3.

The time-averaged vorticity fields for the MP and FF regimes are provided along with the pressure sensor locations in Figs. 4(a) and 4(b), respectively. Six (virtual) pressure sensors ( $N_p = 6$ ), whose values are output at every time step, are sparsely placed along the outer edges of the flow domain for each case. The streamwise spacing of the sensors are chosen such that they are not integer multiples of an integral length scale (such as the cylinder diameter  $D$ ). Their spanwise positions ensure that they are sufficiently removed from the region of high velocity magnitude, thus mimicking remoteness. These settings provide additional challenges for the estimators. The effects of sensor placement along the outer edges of the spatial domain and the number of sensors on estimation performance are briefly discussed in Sec. III B 3 b for select configurations.

The velocity field, defined as  $\mathbf{U}(\mathbf{x}, t)$  with its two components  $U(\mathbf{x}, t)$  and  $V(\mathbf{x}, t)$ , represents the velocity components in the  $x$  and  $y$  directions in the Cartesian coordinate system defined in Fig. 1, respectively (note, this can be extended to three-dimensional flow fields and the figure maintains its generality). After obtaining the spatiotemporal flow snapshots of interest, the time-averaged (mean) field  $\bar{U}(\mathbf{x})$  is subtracted from each of the snapshots at times  $t_k = k \Delta t$  via the Reynolds decomposition to yield the fluctuating velocity fields  $\mathbf{u}(\mathbf{x}, t_k) = \mathbf{U}(\mathbf{x}, t_k) - \bar{U}(\mathbf{x})$ . The corresponding fluctuating velocity components are defined as  $u(\mathbf{x}, t)$  and  $v(\mathbf{x}, t)$ .

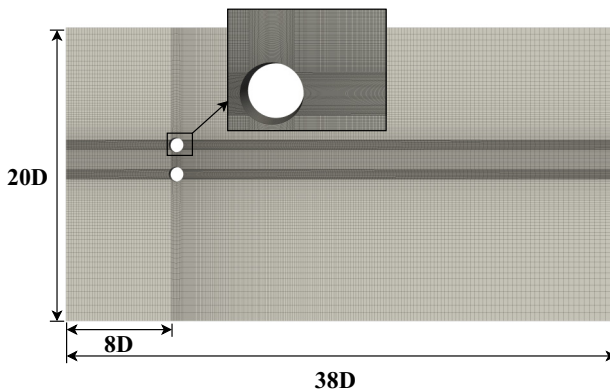


FIG. 3. Computational domain geometry and dimensions for  $L/D = 0$ ,  $T/D = 2$  (every fourth point shown to improve figure clarity). Inset: Closeup of mesh refinement near solid cylinder surface.

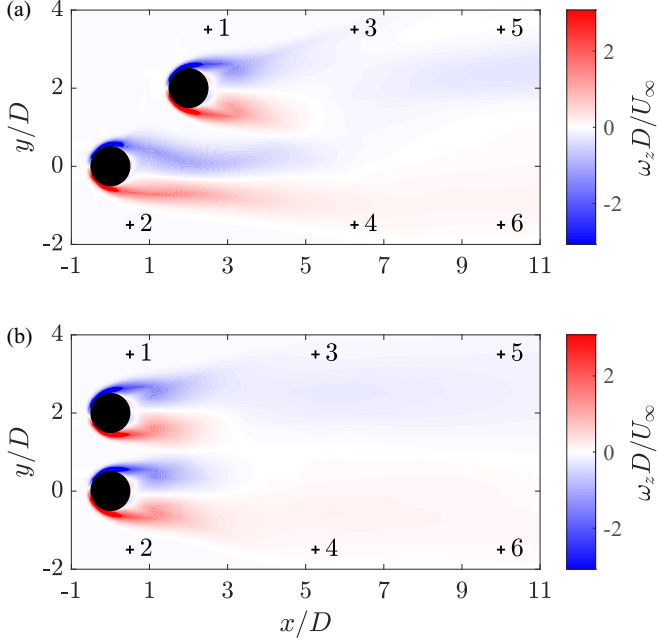


FIG. 4. Time-averaged nondimensional vorticity field of the (a) MP and (b) FF flows with  $N_p = 6$  sensors. The nondimensional coordinates  $(x/D, y/D)$  of the sensor locations (labeled 1 through 6) indicated by (+) marks for the MP case are  $(2.5, 3.5)$ ,  $(6.25, 3.5)$ ,  $(10, 3.5)$ ,  $(0.5, -1.5)$ ,  $(6.25, -1.5)$ , and  $(10, -1.5)$ . Similarly, the sensor locations for the FF case are  $(0.5, 3.5)$ ,  $(5.25, 3.5)$ ,  $(10, 3.5)$ ,  $(0.5, -1.5)$ ,  $(5.25, -1.5)$ , and  $(10, -1.5)$ .

## 2. Proper orthogonal decomposition

The important flow structures are extracted from the high-dimensional flow using the proper orthogonal decomposition (POD). Also known as the Karhunen-Loève decomposition, the concept was first introduced to the study of fluid mechanics by Lumley [42]. In this work, we use the method of snapshots [43] to compute the POD, which is outlined by Tropea *et al.* [44] and extended to matrix space.

Given the Reynolds decomposition of the velocity field, we can decompose the fluctuating ensemble  $\mathbf{u}(\mathbf{x}, t)$  of  $N$  two-dimensional snapshots into a finite sum of (eigen-)basis functions  $\Phi_k(\mathbf{x})$  and corresponding time-dependent coefficients  $a(t_k)$ :

$$\mathbf{u}(\mathbf{x}, t) = \sum_{k=1}^N \Phi_k(\mathbf{x}) a(t_k). \quad (1)$$

Additionally, the energy captured by the first  $r$  modes defined as

$$E_r \equiv \sum_{k=1}^r \frac{1}{2} \overline{a(t_k)^2} \quad (2)$$

is maximized. That is, the basis functions generated using POD are such that the first  $r$  modes contain more energy than would be generated by any other orthogonal decomposition. Maximization of  $\overline{a(t_k)^2}$  results in the basis  $\Phi_k(\mathbf{x})$  satisfying the orthonormality condition in  $\mathcal{L}^2$ -Hilbert space, and the time-dependent coefficients can thus be obtained by orthogonal projection. The corresponding

coefficients additionally satisfy orthogonality via the scaling relationship defined in Tropea *et al.* [44]:

$$(a_i(t), a_j(t)) = \sum_{k=1}^N a_i(t_k) \cdot a_j(t_k) = N\lambda_i\delta_{ij}, \quad (3)$$

where  $a_i(t)$  and  $a_j(t)$  are the time-varying coefficients spanning  $N$  snapshots that correspond to  $\Phi_i(\mathbf{x})$  and  $\Phi_j(\mathbf{x})$ , respectively;  $\{\lambda_i\}$  are the POD modal energies, which are equivalently known as the POD eigenvalues. Lastly, the modes are rank-ordered by their energy content— $\lambda_1 > \lambda_2 > \dots > \lambda_N$ . Thus, a truncated set of eigenfunctions forms an energy-optimal basis in which the corresponding coefficients evolve in time.

The two-dimensional computation of the turbulent kinetic energy (TKE),  $\overline{q^2}$ , is given by

$$\overline{q^2} = \frac{1}{2}(\overline{u^2} + \overline{v^2}). \quad (4)$$

The TKE in the flow domain  $\Omega$  and its relationship with the POD eigenvalues is:

$$k = \iint_{\Omega} \overline{q^2} d\Omega = \frac{1}{2} \sum_{n=1}^N \lambda_n. \quad (5)$$

Fuller treatments of this topic can be found in Lumley [42], Pope [45] and Holmes *et al.* [46].

## B. Estimation

### 1. Overview of the estimation process

A general POD-based flow estimation procedure is summarized in Fig. 5. For the entirety of the estimation, we are targeting the fluctuating velocity fields. Stage I of the estimation involves the extraction of POD modes through the flow field decomposition. Following the acquisition of flow snapshots, an appropriate training data size is defined based on a convergence criterion for the modes. Then, a mode selection procedure is employed to obtain the modes that are representative of the underlying flow dynamics. The POD is then computed over the training ensemble, thus providing (i) a spatial basis  $\Phi^c(\mathbf{x})$ , labeled as “POD basis,” for the low-dimensional subspace and (ii) the time-varying signals labeled “POD coefficient training data,”  $a^c(t_k)$ ,  $\forall t_k \in \mathcal{T}_{\text{train}}$  where  $\mathcal{T}_{\text{train}}$  is the interval spanning the training data and  $k \in [1, N_{\text{train}}]$  such that  $N_{\text{train}}$  is the number of training snapshots. The coefficients describe the temporal evolution of the basis vectors for the training duration. A detailed description of the mode selection process used in this work is given in Appendix A 1. The obtained basis is momentarily set aside while the temporal coefficients serve as training data for Stage II of the process—the estimator training stage.

To begin the estimation, we obtain synchronized measurement training data obtained from the sparsely placed pressure sensors,  $P(\mathbf{x}_s, t_k)$ , in the spatial domain,  $\Omega_s$ . The fluctuating pressure sensor data,  $p(\mathbf{x}_s, t_k) = P(\mathbf{x}_s, t_k) - \overline{P(\mathbf{x}_s, t_k)}$ , are used to train the estimator.

During the training stage, a mapping  $\mathcal{F}$  is approximated by correlating the instantaneous sensor states, defined as  $p(\mathbf{x}_s, t_k)$  to the coefficients  $a^c(t_k)$  representing the state of the POD modes at time  $t_k$ , for  $k \in [1, N_{\text{train}}]$ . That is,

$$\mathcal{F} : p(\mathbf{x}_s, t_k) \mapsto a^c(t_k). \quad (6)$$

The training process is briefly summarized in Fig. 6. First, the sensor training data and the POD coefficient training data are scaled using the  $z$ -score normalization through their individual mean and standard deviation. Second, an estimator is employed to approximate a function that represents the mapping  $\mathcal{F}$ . In this work, four types of estimators are considered:

- (1) LSE-POD,
- (2) QSE-POD,



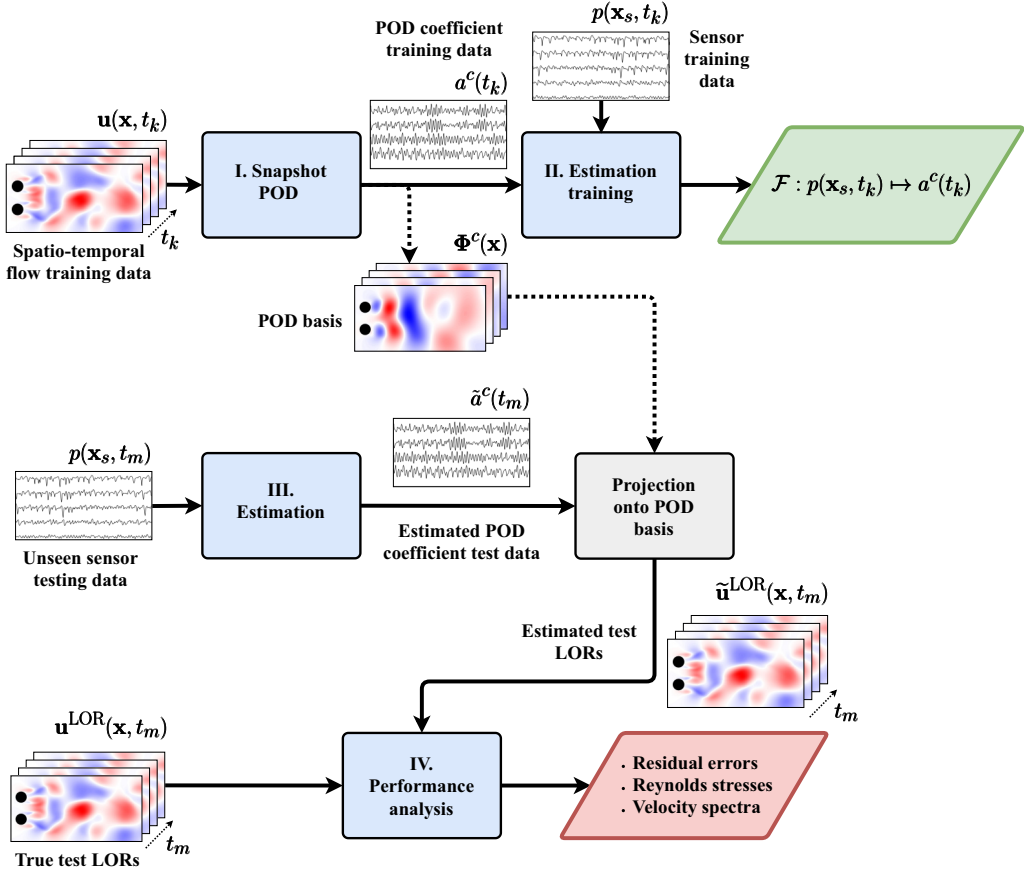


FIG. 5. Block diagram of a generic POD-based flow estimation process including a performance assessment. Note that the time interval in Stages I and II are spanned by the training samples (i.e.,  $t_k \in \mathcal{T}_{\text{train}}$ ) and the time interval in Stages III and IV are spanned by unseen testing samples (i.e.,  $t_m \in \mathcal{T}_{\text{test}}$ , where  $\mathcal{T}_{\text{test}} \cap \mathcal{T}_{\text{train}} = \emptyset$ )

- (3) shallow<sup>1</sup> feedforward artificial neural networks (FFNN), and
- (4) single-layer LSTM networks.

Note that the orange block in Fig. 6 represents the general estimation training steps involved for an FFNN with two fully connected hidden layers, which is later detailed in Sec. II B 2 b. In the case of the LSE/QSE-POD, this block can be replaced by a least-square regression approach to obtain the LSE/QSE coefficients (see Sec. II B 2 a). For the FFNN, the layers are linked through linear combinations of weights and biases (note that nonlinear activation functions are also used between the layers, but are not explicitly marked in Fig. 6). The input pressure sensor data are passed through the hidden layers to provide an estimate of the output POD coefficients. The deviation of these estimates is then quantified using a mean-square-error (MSE) loss function,  $\mathcal{L}$ . This loss is then used to perform gradient descent (via backpropagation) to tune the weights and biases of the

<sup>1</sup>In this study, the term “shallow” network is used to qualitatively characterize a network with three or fewer hidden layers.

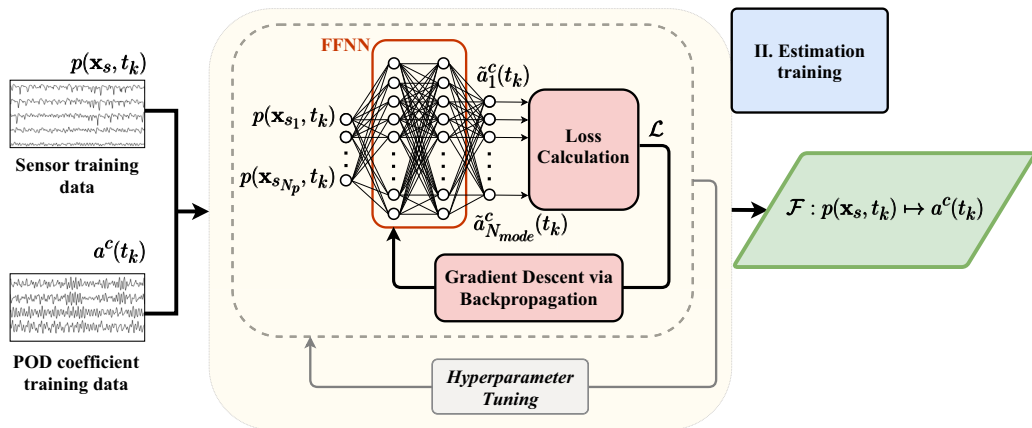


FIG. 6. Block diagram describing Stage II of the estimation process: Estimation training for a sample ANN with two fully connected hidden layers. Input pressure data at times  $t_k \in \mathcal{T}_{\text{train}}$  are passed through the network to estimate the velocity POD coefficients through the mapping  $\mathcal{F}$ . The network is trained to minimize the loss function denoted by  $\mathcal{L}$ . Additional details are found in Sec. II B 2 b.

network. The process is repeated until a stopping criterion is met such that the loss is sufficiently low and/or the model does not overfit the training data. Optionally, to improve performance, the hyperparameters of the ANN (number of hidden layers and units, parameters of the gradient descent optimizer, etc.) can be tuned. The trained network is then an approximation of the desired mapping  $\mathcal{F}$ . Additional details are provided in Sec. II B 2 b.

Stage III is the estimation, where unseen sensor data, denoted by  $p(\mathbf{x}_s, t_m)$ , are mapped using  $\mathcal{F}$  to estimate the POD coefficients, for  $t_m \in \mathcal{T}_{\text{test}}$  where  $\mathcal{T}_{\text{test}}$  is the time interval spanned by the unseen test data and  $\mathcal{T}_{\text{test}} \cap \mathcal{T}_{\text{train}} = \emptyset$ , where  $\emptyset$  is the empty set. To recover the flow fields, the sensor-based estimation of the flow fields is performed using the obtained basis  $\Phi^c(\mathbf{x})$  from Stage I by projecting the estimated test coefficients  $\tilde{a}^c(t_m)$  onto the basis. This yields the estimated snapshots  $\tilde{\mathbf{u}}^{\text{LOR}}(\mathbf{x}, t_m)$ .

Stage IV involves the performance analysis of each estimator. A true assessment is only possible if new, unseen data are available. The estimated LORs  $[\tilde{\mathbf{u}}^{\text{LOR}}(\mathbf{x}, t_m)]$  are compared with the true test LORs  $[\mathbf{u}^{\text{LOR}}(\mathbf{x}, t_m)]$ . We assess the performance of the estimators by studying three metrics:

- (1) Residual error time-series and the time-averaged values of the estimated (unseen) test snapshots,  $\mathbf{u}_k = (u, v)_k$ , computed via their  $\ell_2$ -norms over the flow-dependent spatial domain  $\Omega$ ,
- (2) Estimated time-averaged (Reynolds) shear stress fields,
- (3) Estimated power spectral density functions (PSDFs) of the spanwise velocity at two locations: (1) the location of maximum  $v$ -RMS (root-mean-square) and (2) a secondary location where a different flow behavior is expected, depending on the flow case considered.

The first metric is a standard statistical quantity used to assess the performance of a regressor. The second metric is physics-based and can provide insight into the estimated physics from a qualitative perspective. The third metric, a form of frequency content analysis within different locations in the flow field, is required to characterize the ability of the estimator to retain the important flow dynamics.

## 2. Flow estimators

*a. Linear/quadratic stochastic estimation.* The modified linear and quadratic stochastic estimators (LSE/QSE) are used in the context of POD coefficients as a point of comparison for the ANN estimators. See Guezennec [47] for a thorough review of the implementation and derivation of the stochastic estimation methods for coherent flow structures. The LSE-POD implementation

of Bonnet *et al.* [5] is used to compute the LSE coefficients from the pressure sensor and POD coefficient training data. The algorithm outlined in Murray and Ukeiley [48] for the modified QSE-POD approach is used in this work to compute the QSE coefficients. A brief overview of stochastic estimation is presented below.

In the context of the present work, stochastic estimation methods, which were first presented by Adrian and Moin [3], are formulated as a conditional average of the velocity POD coefficients  $\tilde{a}^c(t)$  given the event of sparse pressure measurements  $p(\mathbf{x}_s, t)$ . That is, the conditional average of the temporal coefficients of some mode  $i$  is

$$\tilde{a}_i^c(t) = \mathbb{E}\{a_i^c(t)|p(\mathbf{x}_s, t)\}, \quad (7)$$

where  $\mathbb{E}\{\cdot\}$  is the expectation operator.

As shown by Guezennec [47], this conditional average can be estimated by a power series:

$$\tilde{a}_i^c(t) = A_{ij}p(\mathbf{x}_{s_j}, t) + B_{ijk}p(\mathbf{x}_{s_j}, t)p(\mathbf{x}_{s_k}, t) + C_{ijkl}p(\mathbf{x}_{s_j}, t)p(\mathbf{x}_{s_k}, t)p(\mathbf{x}_{s_l}, t) + \dots, \quad (8)$$

where  $p(\mathbf{x}_{s_j}, t)$  is the fluctuating pressure time-series measured at spatial location  $\mathbf{x}_{s_j} \in \Omega_s$ , and  $A_{ij}$ ,  $B_{ijk}$  and  $C_{ijkl}$  are the coefficients to be solved.

Truncating Eq. (8) to the first and second terms yields the linear and quadratic stochastic estimates, respectively. The linear estimate is thus expressed as

$$\tilde{a}_i^c(t) = A_{ij}p(\mathbf{x}_{s_j}, t), \quad (9)$$

and the quadratic estimate is expressed as

$$\tilde{a}_i^c(t) = A_{ij}p(\mathbf{x}_{s_j}, t) + B_{ijk}p(\mathbf{x}_{s_j}, t)p(\mathbf{x}_{s_k}, t). \quad (10)$$

The coefficients  $A_{ij}$  and  $B_{ijk}$  are obtained by minimizing the MSE  $[\tilde{a}_i^c(t) - a_i^c(t)]^2$  with respect to the coefficients. In particular, the minimization problem is constrained by Eqs. (9) and (10) for the LSE and QSE problems, respectively. Differentiating the MSE with respect to the coefficients and equating to zero yields a linear system of equations, which can be solved to find  $A_{ij}$  and  $B_{ijk}$ .

*b. Feedforward and long short-term memory neural networks.* The feedforward ANN (FFNN) can be formulated as a nonlinear function approximator that maps input states to output states. A detailed overview of the feedforward architecture can be found in Goodfellow *et al.* [49]. If the input layer of pressure sensor measurements at  $t_k$  is defined as the vector  $\mathbf{p} = [p(\mathbf{x}_{s_1}, t_k), \dots, p(\mathbf{x}_{s_{N_p}}, t_k)]^T \in \mathbb{R}^{N_p}$ , and the output layer of network-predicted POD coefficients is defined as  $\tilde{\mathbf{a}} = [\tilde{a}_1^c(t_k), \dots, \tilde{a}_{N_{\text{mode}}}^c(t_k)]^T \in \mathbb{R}^{N_{\text{mode}}}$ , then  $N_p$  and  $N_{\text{mode}}$  denote the input and output state dimensions, respectively.<sup>2</sup> The network can have many hidden layers between the input and output layers, whose (hidden) units are interconnected through linear combinations of weights and biases. The goal is to obtain appropriate values for these weights and biases to fit a function that maps the input state to the output state.

A high-level schematic of the FFNN architecture used in this work is provided in Fig. 6. Mathematically, if  $d_i$  denotes the dimension of the  $i$ th layer, the function  $G_i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d_{i+1}}$  then describes the mapping between the input state of the  $i$ th layer, denoted as  $\mathbf{g}_i \in \mathbb{R}^{d_i}$ , to that of the  $(i+1)$ th layer,  $\mathbf{g}_{i+1} \in \mathbb{R}^{d_{i+1}}$ . In particular,  $G_i = \xi_i(\Theta_i)$ . Here,  $\Theta_i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d_{i+1}}$  is an affine transformation of the layer input that takes the form  $\Theta_i(\mathbf{g}_i) = \mathbf{W}_i^T \mathbf{g}_i + \mathbf{b}_i$ , where the matrix  $\mathbf{W}_i \in \mathbb{R}^{d_i \times d_{i+1}}$  contains the weights, the vector  $\mathbf{b}_i \in \mathbb{R}^{d_{i+1}}$  contains the biases, and  $\xi_i : \mathbb{R}^{d_{i+1}} \rightarrow \mathbb{R}^{d_{i+1}}$  is a Hadamard (elementwise) evaluation of a nonlinear activation function. An FFNN with  $l$  fully connected layers is expressed for all  $k$  by

$$\tilde{\mathbf{a}} = G_l \circ G_{l-1} \circ \dots \circ G_2 \circ G_1, \quad (11)$$

where  $\circ$  is the composition operator, and  $G_1 = \xi_1(\Theta_1) = \xi_1(\mathbf{W}_1^T \mathbf{p} + \mathbf{b}_1)$ .

<sup>2</sup>The superscript “ $T$ ” denotes the matrix transpose.

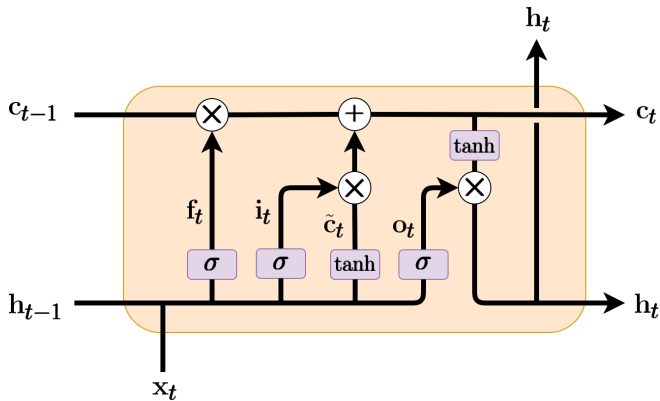


FIG. 7. LSTM Cell.

Dropout layers were first applied to recurrent neural networks (RNNs) with LSTM cells in the context of handwriting recognition [50], which showed significant performance improvement over LSTMs that did not have dropout layers. These layers act as regularizers and are thus incorporated after each hidden layer in the ANNs used in the present work. The loss function used for the FFNN at some time instant  $t_k$  is the half-MSE, not normalized by  $N_{\text{mode}}$ , computed via

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^{N_{\text{mode}}} [\tilde{a}_i^c(t_k) - a_i^c(t_k)]^2. \quad (12)$$

Note that  $L_2$ -regularization is also used during training. Using this loss, the weights and biases are tuned via stochastic gradient descent using the adaptive moment estimation algorithm (Adam) [51]. The gradient is computed using backpropagation.

The LSTM network is a gated RNN whose main benefit is the presence of paths through which the gradient can flow for long durations without vanishing or exploding [10]. This results in the LSTM being able to detect temporal dependence, so that the predictions are conditional on the context and history of the input sequence. Herein, a brief description of the LSTM architecture is presented. For a more detailed review, see the original work by Hochreiter and Schmidhuber [9].

A diagram of an LSTM cell at time  $t = t_k$  is shown in Fig. 7. It contains three gates: an input gate, a forget gate, and an output gate. For a cell containing  $d$  hidden units, the cell state,  $\mathbf{c}_t \in \mathbb{R}^d$ , contains the long-term information the network propagates in time. The forget gate, denoted by  $\mathbf{f}_t \in \mathbb{R}^d$ , decides which information to keep and which to discard. This is the leftmost gate shown in the diagram, and is represented mathematically as

$$\mathbf{f}_t = \sigma(\mathbf{W}_f^T \mathbf{x}_t + \mathbf{U}_f^T \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (13)$$

where  $\sigma$  is the sigmoid unit, which scales the output between 0 and 1,  $\mathbf{x}_t \in \mathbb{R}^{N_p}$  is the input vector,  $\mathbf{W}_f \in \mathbb{R}^{N_p \times d}$  is the input weight,  $\mathbf{U}_f \in \mathbb{R}^{d \times d}$  is the recurrent weight,  $\mathbf{b}_f \in \mathbb{R}^d$  is the bias, and  $\mathbf{h}_{t-1} \in \mathbb{R}^d$  is the (previous) hidden state vector.

The incoming cell state,  $\mathbf{c}_{t-1} \in \mathbb{R}^d$ , is updated through the input gate (denoted by  $\mathbf{i}_t \in \mathbb{R}^d$ ). The input gate activation vector is defined as

$$\mathbf{i}_t = \sigma(\mathbf{W}_i^T \mathbf{x}_t + \mathbf{U}_i^T \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (14)$$

where  $\mathbf{W}_i \in \mathbb{R}^{N_p \times d}$ ,  $\mathbf{U}_i \in \mathbb{R}^{d \times d}$  and  $\mathbf{b}_i \in \mathbb{R}^d$  are the corresponding (input) weight, recurrent weight and bias, respectively. The input activation vector is multiplied with an intermediate cell

state  $\tilde{\mathbf{c}}_t \in \mathbb{R}^d$ :

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c^T \mathbf{x}_t + \mathbf{U}_c^T \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (15)$$

where  $\mathbf{W}_c \in \mathbb{R}^{N_p \times d}$ ,  $\mathbf{U}_c \in \mathbb{R}^{d \times d}$ , and  $\mathbf{b}_c \in \mathbb{R}^d$  are the corresponding (intermediate) weight, recurrent weight, and bias, respectively. The new cell state,  $\mathbf{c}_t \in \mathbb{R}^d$ , is obtained as follows:

$$\mathbf{c}_t = \mathbf{c}_{t-1} \odot \mathbf{f}_t + \tilde{\mathbf{c}}_t \odot \mathbf{i}_t, \quad (16)$$

where  $\odot$  denotes the Hadamard product, which is a componentwise multiplication of two vectors. The output of the cell,  $\mathbf{h}_t \in \mathbb{R}^d$ , is dependent on the current cell state,  $\mathbf{c}_t$ , and the new input into the cell. The output gate activation vector is defined as

$$\mathbf{o}_t = \sigma(\mathbf{W}_o^T \mathbf{x}_t + \mathbf{U}_o^T \mathbf{h}_{t-1} + \mathbf{b}_o), \quad (17)$$

where  $\mathbf{W}_o \in \mathbb{R}^{N_p \times d}$ ,  $\mathbf{U}_o \in \mathbb{R}^{d \times d}$ , and  $\mathbf{b}_o \in \mathbb{R}^d$  are the output weight, recurrent weight, and bias, respectively. It is multiplied by the current cell state, yielding the cell output

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \quad (18)$$

The fully connected final output layer at time  $t$  is defined through the network-predicted estimate,  $\tilde{\mathbf{a}}$ . Using Eqs. (13)–(18), the LSTM cell is delayed (or is looped back in time) by  $S$  time steps, indicating that the hidden state at time  $t$ ,  $\mathbf{h}_t$ , is conditional on  $\mathbf{h}_{t-S:t-1}$ . We refer to the variable  $S$  as the sequence length. The same loss function,  $\mathcal{L}$ , is used to train the LSTM networks. Note that  $L_2$ -regularization is also used during training. The weights and biases are computed using the Adam [51] optimizer, and the gradient is computed via the backpropagation through time algorithm [52].

The number of hidden units in the LSTM cell,  $d$ , controls the expressivity of the network to encode time history. A large size can significantly increase computational time but can also increase expressive power while risking overfitting of the training samples. While a small value of  $d$  might not have the same expressivity as a large layer size, the risk of overfitting and the computational effort decreases, as with any ANN architecture. Thus, the layer size must be tuned. The final LSTM network architecture consists of the input layer of  $N_p$  units, followed by an LSTM layer of  $d$  units, a dropout layer with a prescribed (dropout) probability, and finally a fully connected regression layer followed by the output layer each containing  $N_{\text{mode}}$  units.

Complex network architectures can be difficult to tune when there are several hyperparameters. The black-box nature of these estimators makes the process especially difficult when no known direct correlations exist between the underlying system and network parameters. Fortunately, these hyperparameters can be tuned in several ways to save computation time and effort. In this work, some hyperparameters are tuned using Bayesian optimization via Gaussian process regression [53]. The objective function of the tuning algorithm is the  $\ell_2$ -norm of the estimated validation snapshots. This approach was similarly employed by Lui and Wolf [54].

*c. Addressing model generalizability.* Due to the interpolative nature of neural networks, the model function, or mapping  $\mathcal{F}$  spans the probability distribution of the training data [2]. Thus, to obtain a generalizable architecture, an effort must be made to ensure that the trained model does not overfit the training data. During training, the network is cross-validated using validation data that are disjoint from the training samples. Following best practices from Goodfellow *et al.* [49], the training set of  $N_{\text{train}}$  samples is split into two disjoint subsets with a ratio of 80:20, where the larger of the two subsets is used for training and the other is used for validation. A well-accepted metric to evaluate the model generalizability is  $1 - \epsilon_{\text{train}}/\epsilon_{\text{val}}$ , where  $\epsilon_{\text{train}}$  and  $\epsilon_{\text{val}}$  are the training and validation errors, respectively. These errors are computed using the average  $\ell_2$ -norm of the residual error between the true and network-predicted values. Based on our results for both the MP and FF flows cases, the metrics are less than 1/500. Early stopping [49] is employed by terminating the training phase when the validation error does not reduce over 50 epochs.

As mentioned earlier, dropout layers are used as a form of regularization after each hidden layer for FFNNs, and after an LSTM cell for the LSTM estimators. To improve generalizability, white

Gaussian noise is added to the training (and validation) pressure sensor data during training. It is well-known that training ANNs using input data that are injected with random noise is equivalent to regularizing the networks [49,55]. The noise levels added to the training data are tuned based on the test errors obtained when the trained models are tested with sensors containing various noise levels. As a result, there is a slight sacrifice in performance with clean (no added noise) test data in favor of lower test errors in the presence of noisy test data. This discussion is revisited in Sec. III B 3.

### III. RESULTS AND DISCUSSION

The wake development and wake interaction in the modulated-periodic (MP) and flip-flop (FF) flow regimes at  $Re_D = 160$  are discussed in Secs. III A 1 and III A 2, respectively. The results of the POD mode selection method are presented in Appendix A 2. The performance analysis of the estimators is finally provided for the MP flow regime in Sec. III B 1 and for the FF regime in Sec. III B 2.

#### A. Wake development and proper orthogonal decomposition-based reconstruction

##### 1. Flow development in the modulated-periodic regime

A sequence of instantaneous vorticity contours for the MP regime is shown in Fig. 8. Qualitatively, the image sequence indicates that unsteady vortical structures are shed periodically from the top and bottom cylinders. This vortex shedding process is most clearly identifiable downstream of the top cylinder, with the alternating appearance of concentrations of spanwise vorticity in a typical von Kármán vortex street that is symmetric with respect to the  $y/D = 2$  axis. However, the bottom cylinder vortices appear to be weaker and significantly modified in both their trajectory and shape due to the presence of the top cylinder. In particular, the bottom cylinder vortices form in the near wake and upon shedding from the body are deflected in the  $-y$  direction, resulting in significant asymmetry relative to the  $y/D = 0$  axis. These patterns and results are consistent with those reported in flow visualizations [29,31–33] and numerical simulations [35]. Additional qualitative features observed include the nature of the vortex interactions between vortices shed from both cylinders in the wake region ( $x/D > 4$ ). For example, there is evidence of vortex merging at approximately  $x/D = 9.5$  and  $y/D = 0.25$ , see Fig. 8(a). Conversely, cross-annihilation of vorticity can be seen with the vortex located at  $x/D = 4.75$ – $5.5$  and  $y/D = 0$  as it is surrounded by shear of opposite sign in Figs. 8(a)–8(c). These observations are of nonlinear processes [29] and give rise to a range of spatial scales and velocity fluctuations within the domain of interest shown in Fig. 8.

The quantitative characterization of the wake is provided through the Reynolds stress fields and the spectral analysis of the velocity and pressure fluctuations, as shown in Fig. 9. Since the unsteady dynamics are the result of an interaction between laminar wakes, the Reynolds stress fields are directly representative of the velocity fluctuations induced by the formation and advection of these vortices. The stress fields are strongest in the wake of the top cylinder, where the periodic vorticity concentrations are evident—consistent with the results in Fig. 8. In fact, the top cylinder wake Reynolds stresses are similar to that of a stationary cylinder in an open free stream for  $x/D < 5$  (see Dong *et al.* [36]). Downstream of this location, vortex interactions take place and result in asymmetric distortion of the Reynolds stress fields [see Figs. 9(a)–9(c) for about  $x/D > 5$ ]. The wake of the bottom cylinder contains substantially weaker velocity fluctuations, resulting in peak Reynolds stresses which are less than 50% of that found in the top cylinder wake.

The frequency content of the wake is investigated through PSD analysis of the velocity and pressure field at points (i)–(vi), shown in Figs. 9(a) and 9(b). In particular, points (i)–(iii) are the locations of maximum RMS values of  $u$ ,  $v$ , and  $p$ , respectively, in the wake of both cylinders. Points (iv)–(vi) roughly correspond to the locations of maximum RMS of these quantities in the wake of the bottom cylinder.

As shown in Figs. 9(d) and 9(e), a multitude of frequencies are present in the wake, in contrast with the well-behaved single-cylinder in cross-flow [40]. Analysis of instantaneous sequences

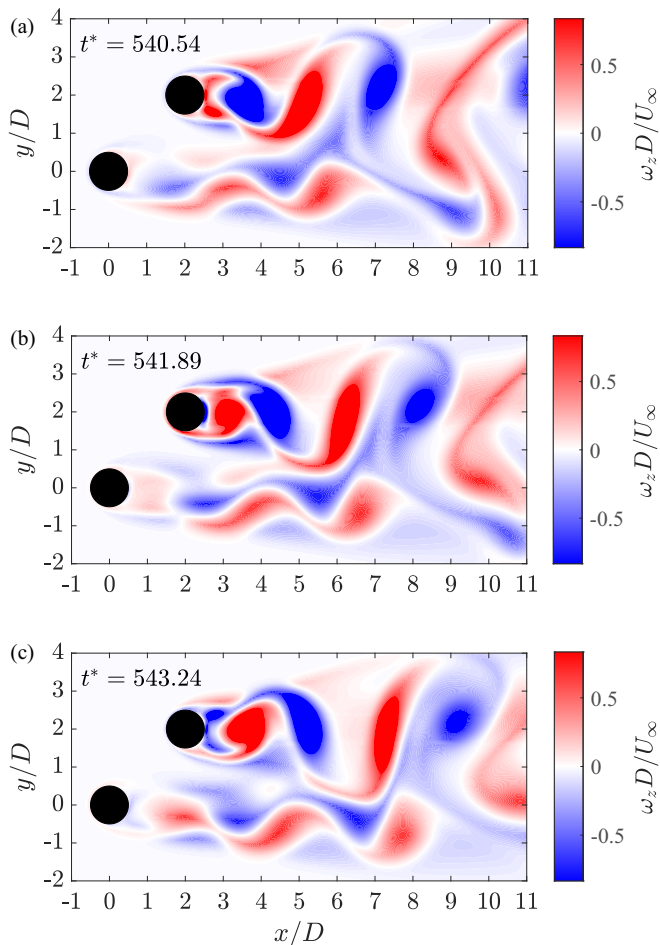


FIG. 8. Sequence of nondimensionalized vorticity contours for the MP case with nondimensional time stamps presented in the top-left corner of each snapshot.

allowed for identification of the vortex shedding frequencies of the top ( $f_{sh,t}^* = 0.19$ ) and bottom ( $f_{sh,b}^* = 0.22$ ) cylinders. These are labeled with vertical dashed lines in Figs. 9(d) and 9(e), and are nondimensionalized as  $f_{sh,t}^* = f_{sh,t} D / U_\infty$ . In particular, we observe that  $f_{sh,b}^* > f_{sh,t}^*$ , which is in agreement with previous studies [28]. Additional relevant frequencies  $f_A^* = 0.0083$ ,  $f_B^* = 0.024$  and  $f_{beat}^* = 0.032$  are also identified in Figs. 9(d) and 9(e). The beat frequency,  $f_{beat}^*$ , corresponds to the difference in the shedding frequencies of the top and bottom cylinders,  $f_{beat}^* = f_{sh,b}^* - f_{sh,t}^*$ . This frequency is associated with the change in relative phase alignment of the wake vortices shed from the top and bottom cylinders.

The frequencies  $f_A^*$  and  $f_B^*$  are most clearly identified at locations (ii), (iii) and (vi). The physical significance of  $f_A^*$  requires consideration that the vortices are in fact shed discretely from the body, as described in detail by McClure *et al.* [56] for a dual-step cylinder in the laminar shedding regime. It is also straightforward to recognize that the dynamics of the wake are related to the relative phase alignment of the vortices at formation. Thus, while the beat frequency,  $f_{beat}^*$ , describes the rate at which the vortices move in and out of phase, a much lower fundamental frequency  $f_A^*$  describes the rate at which the vortices move in phase and are also at the same physical state

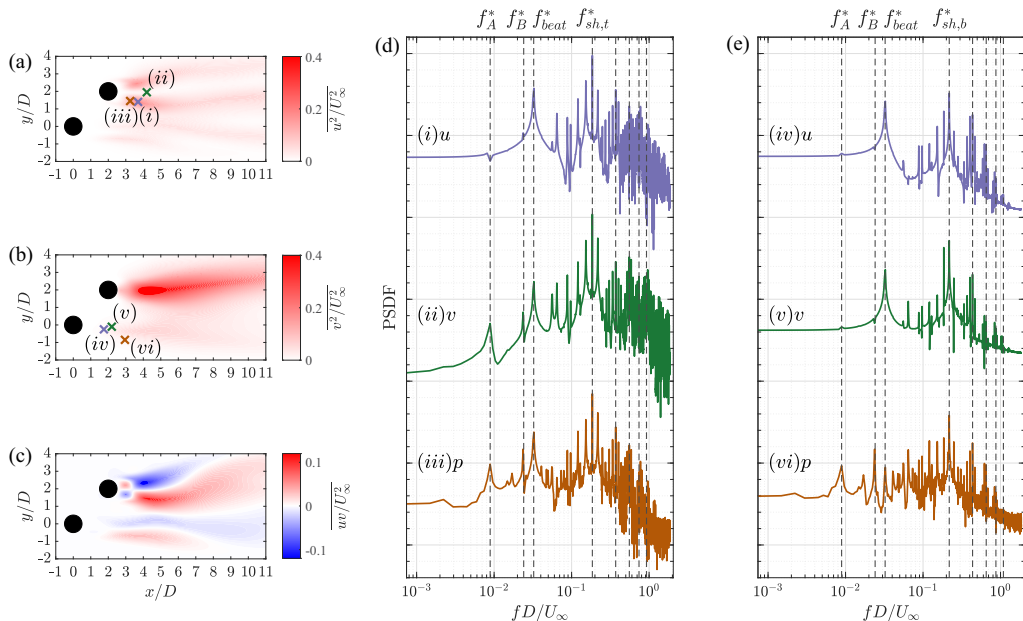


FIG. 9. Reynolds stresses (a)–(c) and PSDFs (d)–(e) of  $u$  and  $v$  velocities and pressures at their location of maximum RMS. The locations are indicated by (X) markings in (a), (b), respectively. The  $(x/D, y/D)$  coordinates are (i) (3.7,1.4), (ii) (4.2,1.95), (iii) (3.25,1.45), (iv) (1.75,−0.25), (v) (2.2,−0.1), and (vi) (2.95,−0.85).

in the vortex formation process. This fundamental frequency is approximately determined by the smallest integer ratio between the vortex shedding frequency of the top and bottom cylinders,  $m/n = f_{sh,t}^*/f_{sh,b}^*$ . Then,  $f_A^*$  is obtained as  $f_A D/U_\infty \approx f_{sh,t} D/mU_\infty \approx f_{sh,b} D/nU_\infty$ . In the MP case of the present study, it is found that  $23/20 \approx m/n \approx f_{sh,t}^*/f_{sh,b}^* \approx 1.15$ . This suggests that the fundamental phase dislocation process spans 23 shedding cycles of the top cylinder, or equivalently 20 shedding cycles of the bottom cylinder. To the best knowledge of the authors, the present study is the first to report and identify the phase dislocation phenomenon in a two-cylinder geometry. However, frequency  $f_B^*$  can be expressed as the difference between the beat frequency and the phase dislocation frequency—i.e.,  $f_B^* = f_{beat}^* - f_A^*$ . Hence,  $f_B^*$  manifests alongside  $f_A^*$  as a consequence of the nonlinear interaction in the wake with  $f_{beat}^*$  [see locations (ii), (iii), and (vi) in Figs. 9(d) and 9(e)].

Proper orthogonal decomposition analysis of the wake is a critical component of the flow estimation procedure and enables a deeper discussion of the most energetic structures. The first six POD modes of the MP flow case are presented along with the PSD functions of the corresponding temporal coefficients (see Fig. 10). It is observed that the first four modes form two mode pairs, as shown by similar energy and frequency content present in both modes of each pair. The first two modes are associated with the dominant vortex shedding process imparted by the top cylinder and the associated coherent structures. Modes 3 and 4 characterize the vortex shedding of the bottom cylinder. Modes 5 and 6 characterize the low-frequency modulation which is caused by the interaction of the two wakes.

The temporal evolutions of modes 1, 3, and 6 are presented through their coefficients in Fig. 11, which describe the shedding and modulation phenomena. Figure 11(a) shows the weakly modulated shedding of the top cylinder at  $f_{sh,t}^*$ ; Fig. 11(b) shows the modulated shedding of the bottom cylinder at  $f_{sh,b}^*$ , where the modulation is more apparent, and; Fig. 11(c) displays the low-frequency



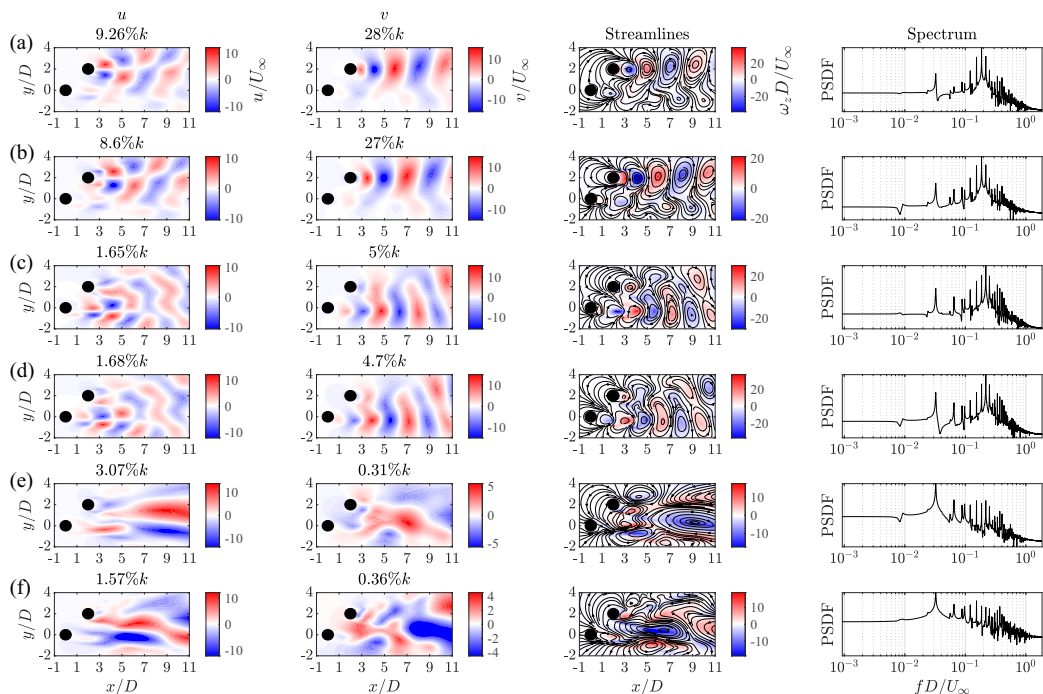


FIG. 10. The first six POD modes of MP flow regime computed using  $N_{\text{train}} = 4000$  snapshots (roughly 200 shedding cycles at  $f_{\text{sh},t}^*$ ). Rows (a)–(f) represent modes 1 through 6, respectively. The first and second columns show the nondimensionalized  $u$ - and  $v$ -velocity components of the spatial modes, respectively; the third column shows the corresponding streamlines of the spatial modes, with the nondimensionalized vorticity component  $\omega_z D/U_\infty$  as the colormap; the fourth column displays the PSDF of the corresponding temporal coefficients on a log-log axis scale. The modal energies contained within each velocity component is shown above each flow field in the first and second columns. The total energy contained in each mode (from mode 1 to 6) is: 37.11%  $k$ , 35.36%  $k$ , 6.65%  $k$ , 6.40%  $k$ , 3.38%  $k$ , and 1.93%  $k$ .

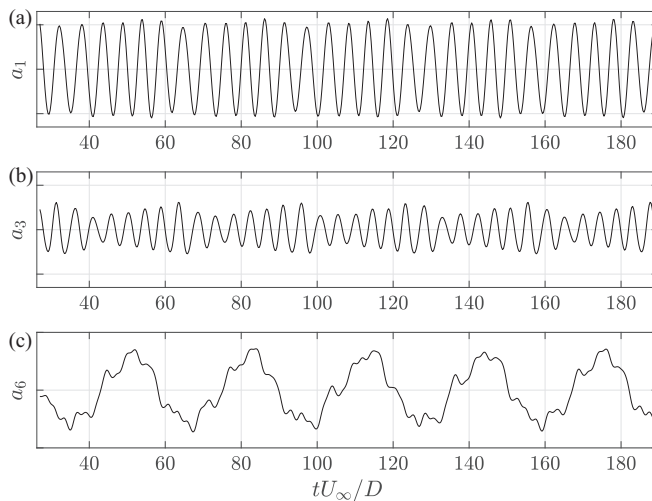


FIG. 11. Temporal coefficients  $a_1$ ,  $a_3$ , and  $a_6$  for MP regime.

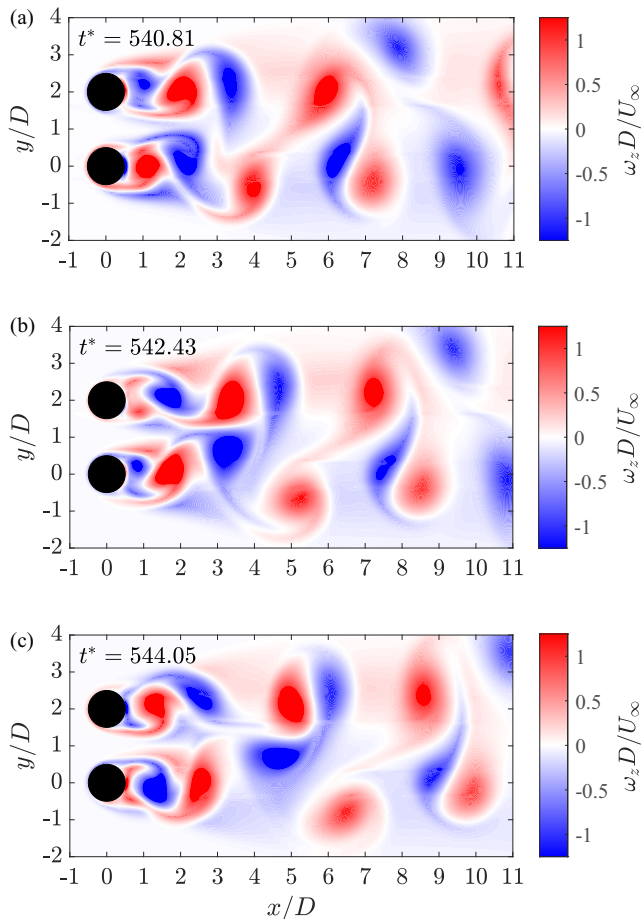


FIG. 12. Sequence of nondimensionalized vorticity contours for the FF case with nondimensional time stamps presented in the top-left corner of each snapshot.

modulation or the beat frequency  $f_{\text{beat}}^*$ . In the context of flow estimation, the clear separation of the shedding processes and their modulation into separate POD modes is promising for estimator development.

## 2. Flow development in the flip-flop regime

A sequence of vorticity field snapshots illustrating the flow development is shown in Fig. 12. The image sequence indicates that the vortex shedding from the top and bottom cylinders is symmetric about the  $y/D = 2$  and  $y/D = 0$  axes, respectively, for about  $x/D = 1-3$ . The shedding topology resembles the typical von Kármán vortex street in this region, with two pairs of counter-rotating vortices imparted by the two cylinders—as observed by Lee and Yang [37]. The cylinder wakes begin to interact at roughly  $x/D > 3$ . In this region of the flow, the vortices tend to either cluster in pairs of opposite signs or remain solitary, though no repeating pattern of these configurations is observed. This is in stark contrast to the observations in the MP regime, where the large-scale dynamics are characterized by specific low-frequency interactions. These observations are consistent with those reported by Lee and Yang [37] and Kang [35] in the FF regime. It is observed that the interaction

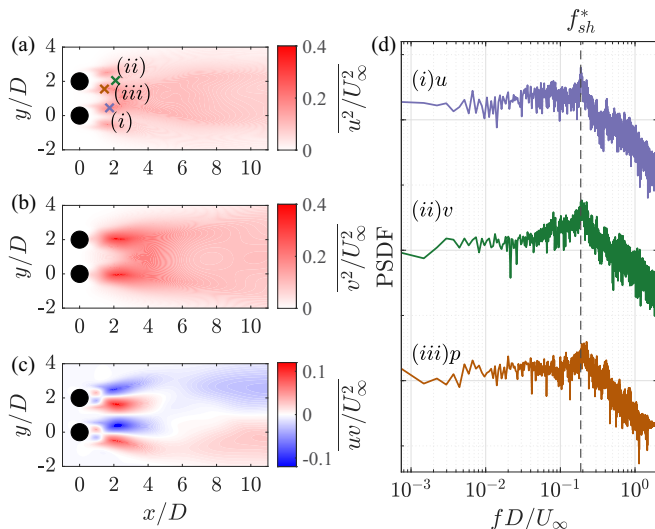


FIG. 13. Reynolds stresses (a)–(c) and PSDF (d) of  $u$  and  $v$  velocities and pressure values at locations where their RMS is maximum. The locations are indicated by (X) markings in panel (a). The  $(x/D, y/D)$  coordinates are (i) (1.7, 0.4), (ii) (2.15, 2.05), and (iii) (1.55, 1.55).

results in significant distortion to the flow symmetry relative to the  $y = 1$  axis (see  $x/D > 3$ ). The nonlinear wake interactions give rise to a large bandwidth of spatial scales and velocity fluctuations.

A quantitative characterization of the wake development is provided through the presentation of the Reynolds stresses and a PSD analysis, as shown in Fig. 13. Unlike the stresses observed in the MP case, the FF stresses generally appear symmetric. Both top and bottom cylinder wake Reynolds stresses are similar to that of a stationary cylinder in an open free stream for about  $x/D < 4$  [36]. The wake interactions downstream of this location generally result in symmetric principal stress fields [see Figs. 13(a) and 13(b)].

The frequency content is investigated by conducting a PSD analysis at points (i)–(iii), as shown in Fig. 13(d). These points are locations where the RMS of  $u$ ,  $v$ , and  $p$  quantities is maximum [see Fig. 13(a)]. The primary vortex shedding frequency is labeled  $f_{sh}^* = 0.19$ . Since there are many frequencies with similar spectral densities, it is difficult to identify the key frequencies other than the shedding frequency. The scattering of high spectral densities across the frequency domain indicates the presence of tightly coupled nonlinear multiscale interactions in the flow, as was also observed by Kang [35] for a study conducted at a slightly lower Reynolds number at  $Re_D = 100$ , and other experimental studies [30,32] conducted at higher Reynolds numbers.

The first six POD modes along with the PSD functions of the corresponding temporal coefficients are shown in Fig. 14. The time-varying coefficients of modes 1, 3, and 6 are plotted in Fig. 15. Three mode pairs characterize the majority of the fluctuating kinetic energy (amounting to 61%  $k$ ) of the FF flow. In particular, the first two spatial modes are symmetric with respect to the  $y/D = 1$  axis and are characterized by short and long wavelengths in the near- and far-wake, respectively. The energetic shorter wavelengths ( $x/D < 3$ ) correspond to the peak frequency at  $f_{sh}^*$  in the corresponding coefficient spectrum, and the longer wavelengths ( $x/D > 3$ ) correspond to the peak at  $f^* \approx 0.08$ . The anti-symmetric spatial patterns in modes 3 and 4 resemble classical von Kármán vortex shedding modes centered at  $f_{sh}^*$ . The time signal of  $a_3$  suggests that the shedding at  $f_{sh}^*$  in modes 3 and 4 is coupled with irregular modulations, likely due to a combination with frequency-centered phenomena at  $f^* \approx 0.08$  and  $f^* \approx 0.40$  [see Fig. 15(b)]. Modes 5 and 6 show similar characteristics as modes 1 and 2. Moreover, the mode energies decay slower compared

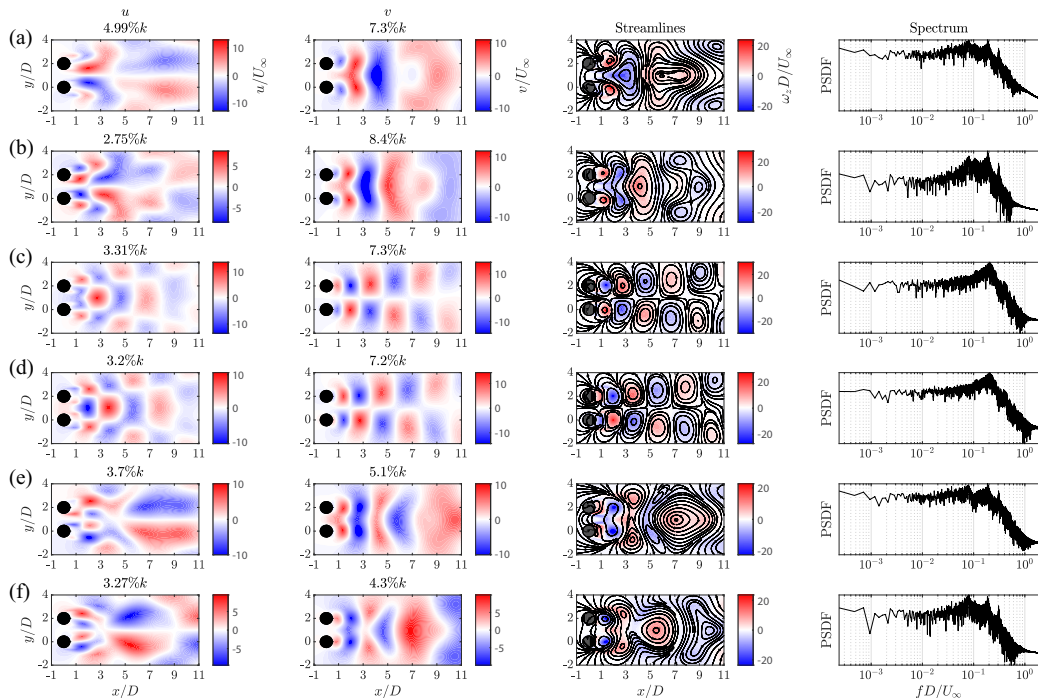


FIG. 14. The first six POD modes of FF flow case computed using  $N_{\text{train}} = 15\,671$  snapshots (roughly 784 shedding cycles at  $f_{\text{sh}}^*$ ), similarly represented as for the MP case in Fig. 10. The total energy in each mode (from mode 1 to 6) is: 12.29%  $k$ , 11.15%  $k$ , 10.61%  $k$ , 10.4%  $k$ , 8.8%  $k$ , and 7.57%  $k$ .

with the MP regime, which suggests that more modes are needed to reconstruct the energetically dominant flow features.

Qualitative assessments from previous studies [38] of the FF regime have noted a gap flow between the cylinders, along with an irregular flow bias either towards the top or bottom boundaries of the domain (hence the name “flip-flop”). The present study is the first to show that the seemingly “chaotic” wake flow in the FF regime is more simply described in a POD subspace. In particular, while the second mode pair describes the unbiased gap flow with vortex shedding from the cylinders, the first and third mode pairs characterize the dynamical coupling of the two wakes that gives rise to a multitude of spatiotemporal scales.

While the frequency scatter in the FF regime was observed in previous works [35,37], the POD analysis of the present study more clearly discerns five distinct frequencies. The shedding frequency is identified to be  $f_{\text{sh}}^* = 0.19$ , and four additional (and somewhat broadband) peaks are observed in the dominant POD modes:  $f^* \approx 0.08, 0.11, 0.31$ , and  $0.40$ . Besides the shedding frequency, the other frequencies have so far not been reported in the literature. These additional frequencies correspond to nonlinear secondary wake interactions, which are major contributors to the Reynolds stresses due to their presence in the most energetic modes. The flow physics that result in these frequencies requires further investigation.

In contrast to the MP temporal coefficients, the FF coefficient signals belonging to the most energetic modes do not exhibit behavior that can be easily predicted by standard statistical techniques. In the context of flow estimation, the irregular patterns arising from the wake interaction can pose significant challenges for standard estimators such as LSE/QSE-POD. This lack of mode separation

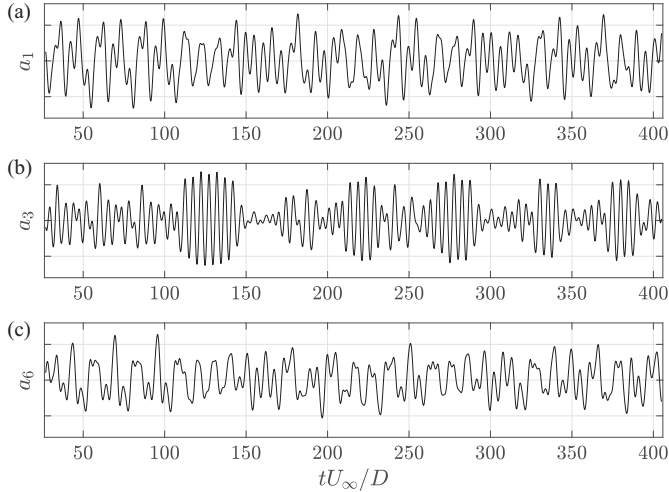


FIG. 15. Temporal coefficients  $a_1$ ,  $a_3$ , and  $a_6$  for FF regime.

in the frequency domain also makes the flow case an excellent testbed for developing more efficient scale-extraction methods.

## B. Flow estimation performance

Sections III B 1 and III B 2 present an analysis of the estimator performance in the MP and FF flow regimes, respectively. Four estimators are studied: LSE-POD, QSE-POD, FFNN, and LSTM networks. Three quantitative metrics are used to assess performance: (i) residual errors, (ii) Reynolds shear stresses, and (iii) velocity spectra. The number of modes to estimate is chosen based on the POD mode selection results provided in Appendix A 2. Section III B 3 provides a discussion on the robustness of the estimators.

### 1. Estimation performance in the modulated-periodic regime

For the MP case, the training data set contains  $N_{\text{train}} = 4000$  snapshots, corresponding to roughly 200 shedding cycles at  $f_{\text{sh},t}^*$ . From this set, 3000 snapshots are used for training and the remaining 1000 are used for cross-validation during the training phase. For the FFNN, the training and validation snapshots are randomly selected from the 4000 total snapshots. For the LSTM, the snapshots in the training and validation data sets are ordered in time due to the dependence of the LSTM network architecture on historical input sequences. For both estimators, the unseen test data set contains  $N_{\text{test}} = 2600$  snapshots, or roughly 130 shedding cycles at  $f_{\text{sh},t}^*$ . Both ANNs are trained using the Adam [51] optimizer, with all its parameters initialized to 0. The input weights are initialized using the method of Xavier [57].  $L_2$ -regularization is set to  $10^{-4}$ .

*FFNN training characteristics.* For the MP case,  $N_p = 6$  pressure sensors are used to estimate the first  $N_{\text{mode}} = 50$  modes. A three-layer fully connected network that consists of two hidden layers is used for training. It was observed that adding more layers did not improve performance. The input layer is  $N_p = 6$  units in length; the first hidden layer of  $d_1 = 80$  units is followed by a dropout layer with probability of 0.05; the second hidden layer of  $d_2 = 160$  units is followed by a dropout layer with 0.1 probability, and the final fully connected regression layer has  $N_{\text{mode}} = 50$  units. The Rectified Linear unit (ReLU) is used after both hidden layers as the activation function. Several other combinations of activation functions were tested with similar results. Training is stopped when the validation error does not reduce after 50 epochs. The initial learning rate of the Adam optimizer is

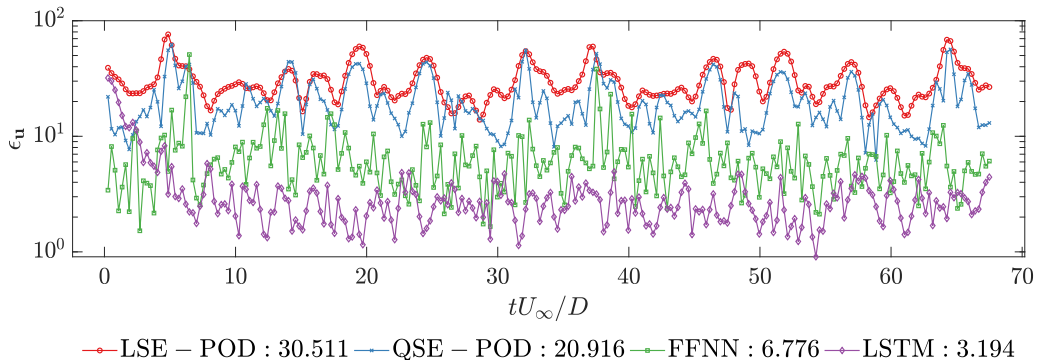


FIG. 16. Residual errors of select estimated test snapshots for  $N_p = 6$  sensors and  $N_{\text{mode}} = 50$  for the MP flow case. The mean residual error,  $\bar{\epsilon}_u$ , is shown in the corresponding legend for each estimator.

0.002, and is dropped by 5% every 50 epochs. The input (sensor) training data are injected with white Gaussian noise with a signal-to-noise ratio of 15 dB. This is later addressed in Sec. III B 3.

*LSTM training characteristics.* A single-layer LSTM network with  $d = 152$  hidden LSTM units is considered, with some parameters manually tuned and others determined through Bayesian optimization. The recurrent weights are initialized as an orthogonal matrix [58]. A dropout layer with 0.061 probability is added after the LSTM layer. The sequence length,  $S$ , and the gradient threshold,  $G_{\text{max}}$ , are manually tuned. It is observed that a longer sequence length resulted in better performance within the span of training samples used in this flow case, which is also accompanied by longer training times. To balance the tradeoff with increasing training times for improved model performance, a sequence length of  $S^* = SU_{\infty}/D = 324$  nondimensional time units (corresponding to roughly 60 shedding cycles at  $f_{\text{sh},t}^*$ ), is chosen along with a  $G_{\text{max}}$  value of 25. The following hyperparameters are tuned via Bayesian optimization: (i) initial learning rate of the Adam optimizer, (ii) dropout probability, and (iii) number of units in the LSTM layer,  $d$ . The initial learning rate of the Adam optimizer is 0.0965, and is dropped by 5% every 50 epochs. The input sensor training data are injected with white Gaussian noise with a signal-to-noise ratio of 10 dB.

First, the residual errors of the estimated snapshots are used to assess the performance of each estimator. The residual is computed as the  $\ell_2$ -norm ( $\epsilon_u$ ) over the flow domain between the estimated snapshots and the corresponding LOR. The results of the error time series are provided in Fig. 16.

The results indicate that the FFNN and LSTM estimators outperform the LSE/QSE-POD estimators on average by an order of magnitude. We observe that the QSE-POD outperforms the LSE-POD, the weakest performing estimator considered, by approximately 50% on average. This suggests that even a second order estimator is insufficient to capture the flow behavior present in the MP flow regime. On average, the LSTM performs over two times better than the FFNN. An initial decrease in the residual error curves for the LSTM is observed. This is because the LSTM is informed by historical data in the sequence (i.e., memory). Thus, the network error should not be evaluated during this initial transient phase (spanning approximately six convective timescales).

Assessing the residual errors in the frequency domain can provide insight into the frequencies (or length scales [7]) where the errors persist. Following the work of Lasagna *et al.* [7], the ratios of the PSDFs of the residuals plotted in Fig. 16 are computed to see how well the ANNs perform relative to the LSE-POD, the weakest performing estimator considered. The ratio is computed according to Eq. (19):

$$\eta_{\epsilon_u\text{-PSDF}} = \frac{\mathcal{S}[\epsilon_u^{(*)}]}{\mathcal{S}[\epsilon_u^{(\text{LSE-POD})}]}, \quad (19)$$

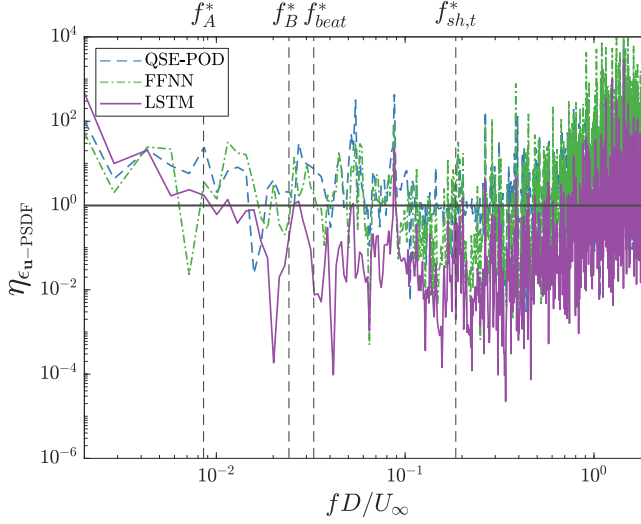


FIG. 17. Ratios of the PSDFs of the residual errors of QSE-POD, FFNN, and LSTM estimations relative to those of the LSE-POD. Results are for  $N_p = 6$ ,  $N_{\text{mode}} = 50$  for the MP flow. Values  $\eta_{\epsilon_u-\text{PSDF}}(f^*) < 1$  indicate better performance of the estimator compared to the QSE-POD at  $f^*$ .

where  $\mathcal{S}[\epsilon_{\mathbf{u}}^{(*)}]$  is the PSDF of the residual errors for either the QSE-POD, FFNN, or the LSTM network;  $\mathcal{S}[\epsilon_{\mathbf{u}}^{(\text{LSE-POD})}]$  is the PSDF of the residual errors for the LSE-POD estimator (see Fig. 17).

Values closer to  $\eta_{\epsilon_u-\text{PSDF}}(f^*) = 1$  indicate no significant difference between the LSE-POD and the other estimators. Values below unity indicate superior reconstruction performance compared to the LSE-POD. A lower value of  $\eta_{\epsilon_u-\text{PSDF}}$  is desirable to ensure a lack of periodicity in the error time series.

It is observed from Fig. 17 that the LSTM estimator performs better than the LSE-POD across most of the frequency domain, while the QSE-POD and FFNN do not significantly outperform the LSE-POD. Additionally, the LSTM curve is seen to fall below the FFNN and QSE-POD curves past  $f^* \approx f_A^*$ , suggesting that the former ANN is better at resolving frequencies  $f^* > f_A^*$ . Past  $f_{\text{beat}}^*$ , the amplitude of the LSTM  $\eta_{\epsilon_u-\text{PSDF}}$  curve is slightly lower than that of the FFNN. The accumulation of estimation errors for both the ANNs is most prominent at higher frequencies past the fundamental shedding frequency  $f_{sh,t}^*$  and at certain frequencies within the band  $f_{\text{beat}}^* < f^* < f_{sh,t}^*$ . The increase in the ratios with higher frequencies in the region  $f^* > 4f_{sh,t}^*$  indicates that all estimators show equally poor performance at these frequencies. Finally, none of the estimators is able to resolve the high frequencies of  $f^* > 1$ .

The second metric to assess the performance of the estimators is their ability to reconstruct the Reynolds shear stress field,  $\overline{uv}$ , presented in Fig. 18. The left column presents the estimated shear stress fields, while the right column shows the absolute residual errors between the corresponding field on the left, and its target. The first row shows the shear stress for the LOR, accompanied by its residual error with the original (nondecomposed) stress field. For the remaining rows, the residual fields on the right correspond to the absolute residual field created between the estimator when compared with the LOR in Fig. 18(a).

The 50-mode LOR is able to accurately reconstruct the Reynolds shear stresses [see Figs. 18(a) and 18(b)]. The residual errors for the LSE-POD are seen to have the same order of magnitude as the LSE-POD estimated shear stress field itself, indicating highly unsatisfactory estimation results [see Figs. 18(c) and 18(d)]. Notably, a trail of high stress magnitudes separated by approximately 2D in the wake of estimated LSE and QSE-POD stress fields is observed [see Figs. 18(c) and

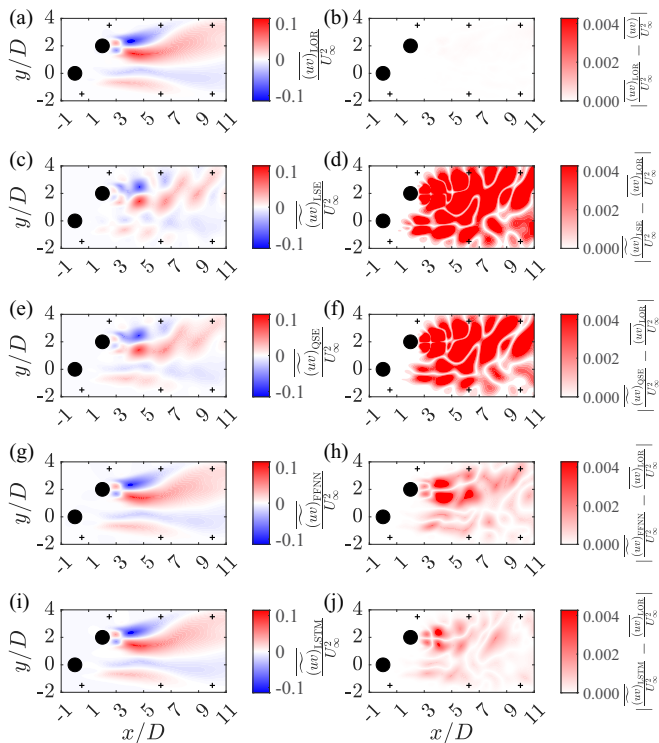


FIG. 18. Reynolds shear stresses,  $\overline{uv}$ , of the MP regime for (a) the LOR, (c) LSE-POD, (e) QSE-POD, (g) FFNN, and (i) LSTM estimates; and their associated absolute residual error fields with row-wise respect to the original (nondecomposed) shear stress field [for panel (b)], and to the LOR [for panels (d), (f), (h), (j)].

18(e)]. This separation length corresponds to the length scales that were not successfully resolved by the LSE/QSE-POD. Moreover, this trail is also apparent in the FFNN and LSTM residual fields, although the error magnitude shown in the contour plots in the right column are two orders of magnitude lower than their estimates in the left column [see Figs. 18(g), 18(h) and 18(i), 18(j)]. The Reynolds shear stress field is adequately reconstructed by both the ANNs for the MP regime, with maximum error in the wake of the top cylinder [see Figs. 18(h) and 18(j)].

The extent to which the estimators are able to reconstruct the individual POD coefficients is assessed. In particular, the ability of each estimator to retain the 50 modal energies is analyzed. The retention is given by the ratio  $\eta_{\lambda_i} = \lambda_i^{\text{est}} / \lambda_i^{\text{LOR}}$ , where  $\lambda_i^{\text{est}}$  is the eigenvalue of mode  $i$  of the estimated test data computed via  $\sum_{k=1}^{N_{\text{test}}} \frac{1}{2} \overline{\tilde{a}_i^2}(t_k)$ , which is similarly computed for the eigenvalues of the LOR, denoted  $\lambda_i^{\text{LOR}}$ . Thus, a ratio closer to unity indicates a more accurate reconstruction of mode energy. The results are given in Fig. 19. The total energy retention given by the sum of these ratios for all 50 modes is given in the legend.

Poor reconstructions of the high-order modes may indicate overfitting of the model. In particular, the rate of decay in energy retention with increasing mode order may indicate the degree of overfitting. Our results suggest—as expected—that the LSE-POD significantly overfits the training data with no clear-cut-off mode past which the retention ratio monotonically decreases. We see that its estimation completely fails past mode 15 with the exception of a few modes that are inadequately captured (i.e.,  $\eta_{\lambda_i} < 0.1$ ). We also note that the LSE-POD disproportionately retains the energies of modes 2 and 4, which are linked to the shedding phenomena as seen previously in Fig. 10. However, the QSE-POD rapidly decays past mode 10, and generally performs better than the LSE-POD by



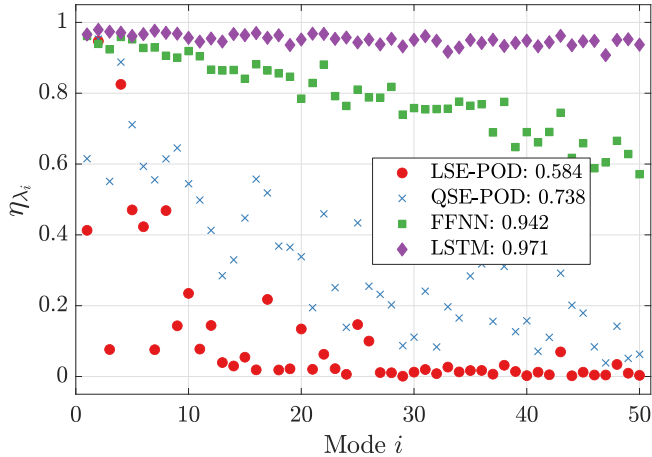


FIG. 19. Retention of MP mode energies captured by the different estimators. Ratio of the total energy captured by each estimator relative to the LOR is given in the legend. Ratios closer to unity imply higher retention of energy.

being able to capture in total 15% more energy. Conversely, the LSTM and FFNN significantly outperform the former two approaches. For the LSTM, we note minor energy deficits in some high-order mode estimates (see modes 19, 33–35, and 47). The FFNN exhibits a more linear decay of energy retention with mode number while being able to reconstruct 94% of the energy in the LOR. The LSTM can reconstruct 97% of the energy in the 50-mode LOR, which is remarkable given that only six sensors are used for training. Finer tuning of the model hyperparameters is expected to reduce overfitting. Finally, we note that while such an energy-based method is a coarse performance metric, it proves to be a quick and an efficient method to analyze the ability of estimators to predict certain (modal) dynamics.

Finally, the ability of the estimators to predict the important flow dynamics is assessed by studying the  $v$ -PSDFs from the LOR and the estimated ensemble spanning the test data at two distinct locations. Location (1) is where the RMS of the  $v$ -component of the velocity is maximum [see location (ii) in Fig. 9(a)]. Location (2) is dominated by the secondary shedding process [see point (iv) in Fig. 9(b)]. The results of the  $v$ -PSDFs are provided in Fig. 20.

We observe that the LSE/QSE-POD estimators inaccurately predict the dynamics of the major flow phenomena at location (1). For instance, they over-estimate the spectrum amplitude at  $f_{\text{beat}}^*$ , ranking it with the dominant shedding at  $f_{\text{sh},t}^*$  although the original and LOR spectra suggest that this location is dominated by the primary vortex shedding. The ANNs are better at predicting the spectrum compared to the LSE/QSE-POD estimators at this location. We notice some nonphysical predictions by the FFNN in the range  $f_A^* < f^* < f_{\text{beat}}^*$ , and some over-predicted frequency components past this region. However, the LSTM can faithfully predict the spectrum at this location without significant scatter in the low-frequency region. Qualitatively, the LSTM can predict the peak amplitudes up to the fourth shedding harmonic, with decreasing accuracy past this point.

At location (2), all estimators over-predict the phase dislocation process at  $f_A^*$ , and generally perform worse compared to location (1). The LSE/QSE-POD spectra at the second location are relatively more scattered compared to those of the ANNs, introducing nonphysical dynamics that are not observed in the original flow (e.g., at  $f_A^* < f^* < f_{\text{sh},b}^*$ ). We observe some over-prediction with the ANNs in the range  $f^* < f_{\text{beat}}^*$ . Moreover, both ANNs struggle to predict the high-frequency amplitude peaks past  $f^* \approx 2.5f_{\text{sh},b}^*$ .

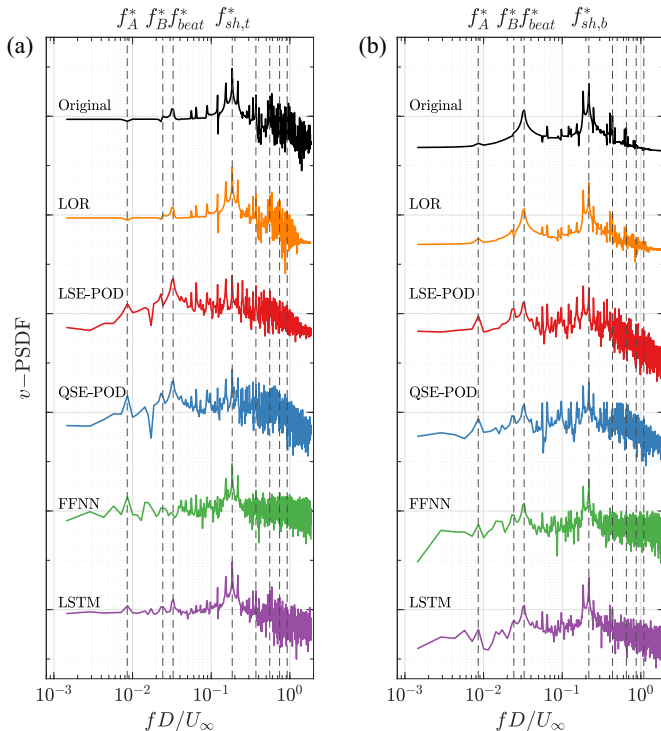


FIG. 20.  $v$ -PSDFs at (a) location (1) at (4.2, 1.95) and (b) location (2) at (1.75,  $-0.25$ ) for  $N_p = 6$  sensors and  $N_{\text{mode}} = 50$  for the MP flow. Unlabeled dotted lines represent the second, third, fourth, and fifth harmonics of  $f_{sh,t}^*$  in panel (a) and  $f_{sh,b}^*$  in panel (b), respectively.

To quantify how well the estimators reconstruct the amplitudes of the power spectra at these two locations, we present  $\eta_{v\text{-PSDF}}$  curves, which are ratios of the estimator  $v$ -PSDFs to those of the LOR (see Fig. 21). A value closer to unity indicates closer agreement with the LOR. At both locations, it is observed that the ratios corresponding to the LSTM are closer to unity for frequencies lower than the shedding frequencies (i.e.,  $f_{sh,t}^*$ ,  $f_{sh,b}^*$ ). At location (1), the curve corresponding to the LSTM confirms that the estimator can faithfully predict frequencies over the dynamical range of interest. The  $\eta_{v\text{-PSDF}}$  curves for the LSTM are marginally lower than those corresponding to the FFNN for regions past  $f^* \approx 0.4$ . Based on the curves at location (2), the LSTM performs slightly better than the FFNN, and both ANNs outperform the LSE/QSE-POD methods. These results agree with the qualitative observations from Fig. 20. All estimators considered are unable to accurately reconstruct the amplitude decay occurring at high frequencies past the fourth harmonics of  $f_{sh,t}^*$  and  $f_{sh,b}^*$  [see Figs. 21(a) and 21(b)]. Overall, the LSTM is the better estimator in the frequency domain.

## 2. Estimation performance in the flip-flop regime

The training data set for the FF flow contains  $N_{\text{train}} = 15671$  snapshots, corresponding to roughly 784 shedding cycles. From this set, 12537 snapshots are used for training and the remaining 3134 are used for cross-validation. The unseen test data set contains  $N_{\text{test}} = 6171$  snapshots or roughly 309 shedding cycles at  $f_{sh}^*$ . A similar approach to the MP flow is taken to train the ANNs.

*FFNN training characteristics.* The first  $N_{\text{mode}} = 65$  FF modes are estimated. All the training characteristics are identical to the MP case except for the number and size of the hidden layers, and the output layer size. The input layer is  $N_p = 6$  units in length; hidden layer of  $d_1 = 50$  units is

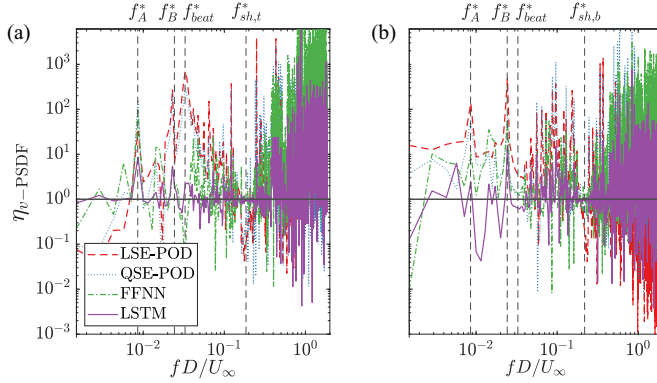


FIG. 21. Ratios of the  $v$ -PSDFs of QSE-POD, FFNN, LSTM estimates to that of the LOR in Fig. 20 at (a) location (1) at (4.2, 1.95) and (b) location (2) at (1.75, -0.25) for the MP case. Values closer to unity indicate better performance.

followed by a dropout layer with probability of 0.01; the second hidden layer of  $d_2 = 160$  units is followed by a dropout layer with 0.05 probability; and the final fully connected regression layer is  $N_{\text{mode}} = 65$  units in length. The corresponding activation function for both hidden layers is the hyperbolic tangent. The initial learning rate of the Adam optimizer is 0.008 and is dropped by 20% every 20 epochs. The input training data are injected with white Gaussian noise with a signal-to-noise ratio of 15 dB.

*LSTM training characteristics.* Identical steps to the MP case are taken to tune the LSTM architecture. The number of LSTM hidden units is  $d = 180$ , followed by a dropout layer with 0.075 (dropout) probability, and the sequence length is  $S^* = 270$  convective timescales (roughly 50 shedding cycles at  $f_{\text{sh}}^*$ ). The initial learning rate of the Adam optimizer is 0.0098 and is dropped by 20% every 20 epochs. The input training data are superimposed with white Gaussian noise with a signal-to-noise ratio of 15 dB.

The residual errors of the reconstructed snapshots are computed over the unseen test data and are presented in Fig. 22. On average, the QSE-POD performs 12% better than the LSE-POD, bridging the performance gap between the two by nearly four times compared with the MP case. Although the four-layer FFNN is expected to be expressive enough to capture strong nonlinear behavior, compared with the LSE/QSE-POD there is only a 30–35% improvement in an average  $\ell_2$  sense. However, the LSTM outperforms the LSE/QSE-POD on average by roughly six times, and the FFNN by over four times.

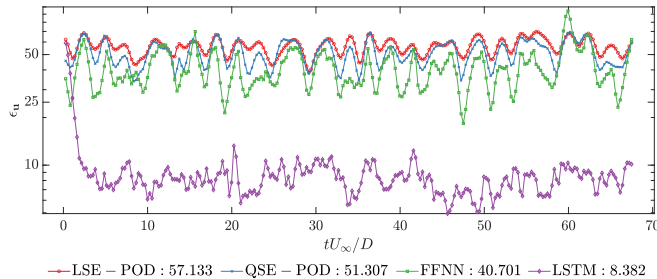


FIG. 22. Residual errors for the estimated test snapshots for  $N_p = 6$  and  $N_{\text{mode}} = 65$  in the FF flow.

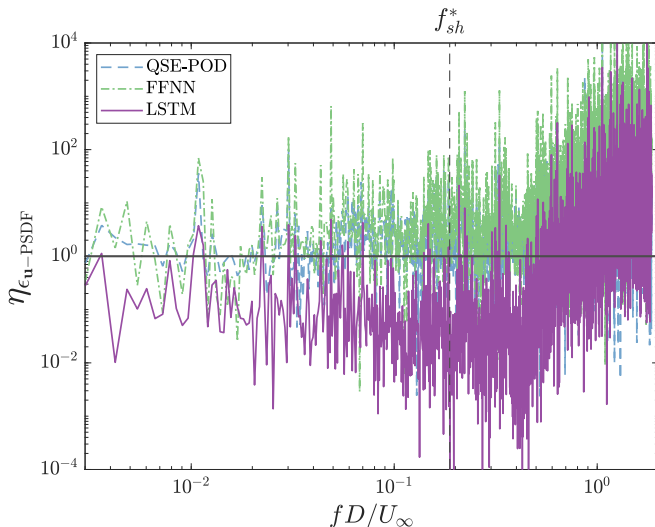


FIG. 23. Ratios of the PSDFs of the residual errors of QSE-POD, FFNN, and LSTM estimations relative to those of the LSE-POD. Results are shown for  $N_p = 6$ ,  $N_{\text{mode}} = 65$  for the FF flow. Values  $\eta_{\epsilon_u\text{-PSDF}}(f^*) < 1$  indicate better performance of the estimator than the QSE-POD.

Following the same approach taken in the previous section, the ratios of the PSDFs between the ANNs to those of the LSE-POD are computed (see Fig. 23). The results show that the LSTM consistently performs better than the LSE-POD. The FFNN and QSE-POD are seen to perform no better than the LSE-POD throughout the frequency domain. The ratios steadily increase for all estimators past  $4f_{sh}^*$ , suggesting that none of them adequately resolves the high frequencies past this point.

The Reynolds shear stresses and the associated residual fields are presented in Fig. 24. We observe that the LSTM faithfully reconstructs the shear stresses. The predictions of the other estimators show errors that are at least an order of magnitude larger compared to the LSTM (see the right column of Fig. 24). The LSE-POD is the weakest estimator, which fails to predict the shear stress across most of the flow domain.

The amount of energy retained in the estimated modes is quantified using the methods described in the previous section with MP case using the mode energy retention ratio  $\eta_{\lambda_i}$  (see Fig. 25).

We observe that the LSE-POD retains only 28% of the energy in the LOR by mainly predicting the first 16 modes. The predictions completely fail for the higher-order modes. The QSE-POD captures 12% more energy than the former method. Both LSE/QSE-POD approaches exhibit a rapid decay in energy retention. The FFNN shows a slightly more linear decay until mode 30 while being able to capture 51% of the energy in the LOR. Finally, the LSTM is able to retain up to 94% of this energy, with a gentle (linear) decay. Given the complex irregularities in the FF POD coefficients, our results demonstrate the expressivity of the LSTM architecture.

Continuing the performance assessment into the frequency domain,  $v$ -PSDFs of the original, LOR and estimated test data are computed at location (1)  $(x/D, y/D) = (2.1, 2.05)$ , position of maximum  $v$ -RMS [see point (ii) in Fig. 13] and location (2)  $(x/D, y/D) = (5, 1)$ , where the wake interactions persist. The results are provided in Figs. 26(a) and 26(b), respectively.

At location (1), we see that the LOR can generally resolve the major spectral content found in the original spectrum, with worsening performance past  $3f_{sh}^*$ . The LSE/QSE-POD estimators struggle to predict the major peak amplitudes and the general shape of the velocity spectrum. Qualitatively, we see that they also fail to reconstruct the general behavior of the amplitude decay. The FFNN

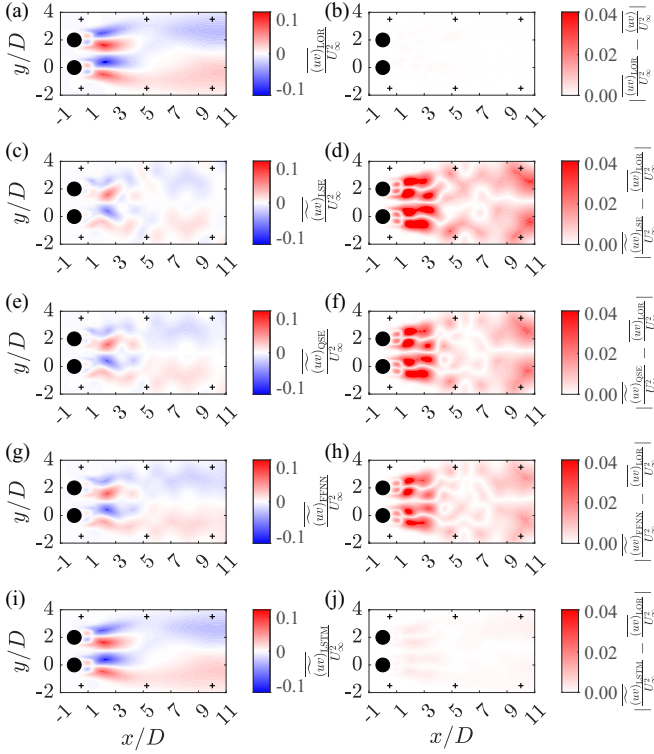


FIG. 24. Estimated Reynolds shear stresses  $\overline{uv}$  for  $N_p = 6$  and  $N_{\text{mode}} = 65$  for the FF flow. Data are organized similar to Fig. 18.

correctly identifies the peak at  $f_{sh}^*$  but over-predicts the lower frequencies in the range  $f^* < f_{sh}^*$  and fails to reconstruct the decay behavior at the higher frequencies. However, the LSTM successfully reconstructs both the major peak amplitudes and the general decay behavior.

At location (2), the LOR is once again able to sufficiently reconstruct the original spectrum, with worsening performance past  $3f_{sh}^*$ . Qualitatively, all estimators appear to resolve the major peak amplitudes, with the LSE/QSE-POD and the FFNN introducing nonphysical components (e.g., see FFNN at  $f^* \approx 10^{-2}$ ). The LSTM is seen to be the best at predicting the spectrum.

We provide a quantitative assessment by following the procedure outlined previously for the MP flow, and compute the ratios of the estimated  $v$ -PSDFs to those of the LOR at locations (1) and (2). The results are presented in Figs. 27(a) and 27(b), respectively.

At location (1), it is observed that the velocity spectrum of the LSTM estimate starts diverging slightly past the fundamental shedding frequency ( $f_{sh}^*$ ), and then steadily decreases. We also verify the superior performance of the LSTM in the low frequency range  $f^* < f_{sh}^*$ . All the estimators perform better at reconstructing the spectrum in this range at location (2), with the LSTM showing the best performance. Similar to location (1), the amplitude predictions at location (2) worsen at high frequencies.

By design, the ANN architectures enable learning of more complex nonlinear correlations as opposed to LSE/QSE-POD approaches, which are restricted by a polynomial structure. In principle, the QSE-POD should be sufficient to learn the necessary correlations given the existence of a quadratic relation between the pressure and velocity fields in incompressible fluid flows. We

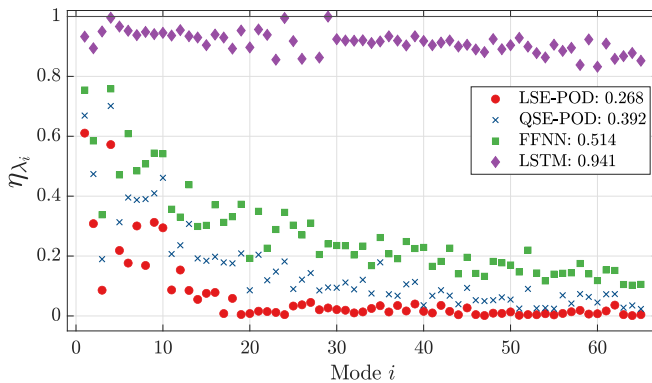


FIG. 25. Retention of FF mode energies captured by the different estimators. Ratio of the total energy captured by each estimator relative to the LOR is given in the legend. Ratios closer to unity imply higher retention.

hypothesize that the reason why the ANN approaches outperform the QSE-POD is due to the nature of the input pressure information, which is a highly sparse representation of the high-dimensional pressure field (i.e., sparse sensors).

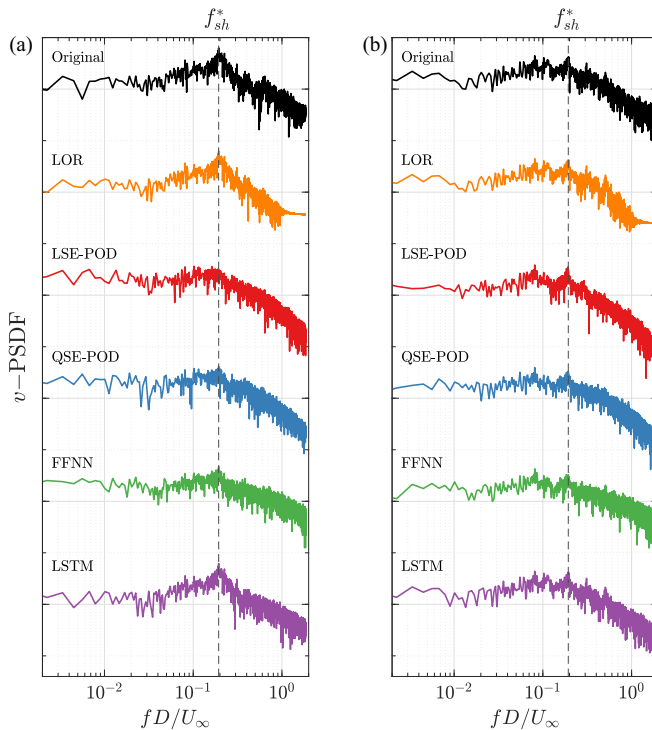


FIG. 26.  $v$ -PSDFs at locations (1)  $(x/D, y/D) = (2.1, 2.05)$  and (2)  $(x/D, y/D) = (5, 1)$  for the FF flow. Labeled in both subfigures is the fundamental shedding frequency,  $f_{sh}^*$ .

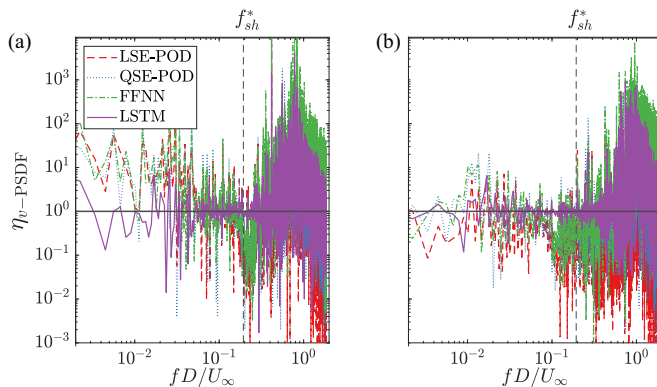


FIG. 27. Ratio of the  $v$ -PSDFs of the estimates of QSE-POD, FFNN, LSTM to the LOR in Fig. 26 at locations (1)  $(x/D, y/D) = (2.1, 2.05)$  and (2)  $(x/D, y/D) = (5, 1)$  for the FF flow case.

From our results, it is hypothesized that the ANN architecture overcomes this sparsity limitation to learn the essential information encoded in the pressure-velocity mapping to predict the low-order velocity field. In particular, the nonlinear activations, which are absent in the LSE/QSE-POD approaches, are likely required to overcome this limitation and are thus major contributors to the superior predictive capabilities of the ANNs. Rigorous studies are required to test this hypothesis. Hence, we expect the LSE/QSE-POD estimators to perform significantly better when several sensors spanning a larger portion of the domain are used since the (internal) estimates of the high-dimensional pressure field are expected to be more accurate.

Moreover, we see through the superior LSTM performance in the FF regime the importance of memory and time history in the input data. This has also been noted by Durgesh and Naughton [8] and Lasagna *et al.* [7] through the use of time-delayed sensor signals.

### 3. Robustness of the estimators

To demonstrate the practical potential of the estimators, robustness of the models must be assessed. Two different types of assessments are performed. First, the predictive capability of the trained estimators in the presence of sensor noise is evaluated in Sec. III B 3 a. This allows for the study of estimator ability to uncover the deterministic component of the sensor signals, which is also useful because instrumentation noise and error is inevitable in a practical setting. Other classes of noise and error sources are those that can be found in a physical experimental setup, such as instrumentation error and noise, inflow nonuniformity, flow three-dimensionality, mounting tolerances of the sensors, and so on. These are not easily replicable in the present computational study. Moreover, vortex flows are sensitive to perturbations [59] as they change the flow response, and thus the flow state. Thus, attempting to perturb the (laminar) two-cylinder flow cases using simulations is expected to change the flow regime, yielding an unknown flow response; determining this response is beyond the scope of this study. Therefore, we assess model robustness through studying estimator performance in the presence of Gaussian sensor noise.

Second, the performance sensitivity of the different estimators to the placement and number of sensors is briefly discussed in Sec. III B 3 b. This enables the identification of estimators that are relatively insensitive to the placement and number of sensors used, making them excellent candidates for systems with limited accessible regions to install sensors. Due to the poor performance of the LSE/QSE-POD estimators, the following discussion on robustness is restricted to the ANNs.

*a. Sensor noise.* The trained models are tested on pressure sensor data that are superimposed with random white Gaussian noise. The signal-to-noise ratio (SNR) is defined on a logarithmic decibel

(dB) scale. SNRs of 25 dB, 15 dB, 10 dB, and 5 dB are tested, and the corresponding flow field predictions are analyzed and compared to the case where no noise is added (defined  $\text{SNR} = \infty$ ).

Compared to training without noise, it is found that training on noisy sensor data results in significantly lower test errors (and thus a more robust model) when the trained model is tested with noisy sensors. In particular, the amount of noise added during training is tuned based on model performance in the presence of noisy unseen test sensors with SNRs of 5 and 10 dB. For the MP case, the LSTM model is trained on noisy sensors with  $\text{SNR} = 10$  dB and the FFNN with  $\text{SNR} = 15$  dB. For the FF case, both the LSTM and FFNN models are trained on noisy sensors with  $\text{SNR} = 15$  dB. No significant variation is observed in the validation and test errors when the above-mentioned SNR values (used for training) are slightly changed.

Four types of errors are used as performance metrics:

- (i) time-averaged velocity field residual errors,  $\overline{\epsilon_{\mathbf{u}}}$ ,
- (ii) Reynolds shear stress relative errors,

$$\epsilon_{\overline{uv}} = \frac{(\overline{uv} - \overline{uv}_{\text{LOR}})_F}{\overline{uv}_F} \times 100\%,^3 \quad (20)$$

(iii) relative PSD error in the energetic fluctuations in the near-wake at  $(x/D, y/D) = (4.2, 1.95)$  for the MP regime and at  $(2.1, 2.05)$  for FF,

$$\epsilon_{S^{(1)}} = \frac{\|\tilde{\mathcal{S}}\|_2 - \|\mathcal{S}_{\text{LOR}}\|_2}{\|\mathcal{S}_{\text{LOR}}\|_2} \times 100\%, \quad (21)$$

where  $\|\mathcal{S}_{\text{LOR}}\|_2$  is the  $\ell_2$ -norm of the PSDF from the LOR (computed as the magnitude using  $u$ - and  $v$ -PSDF components), and  $\|\tilde{\mathcal{S}}\|_2$  is the estimated quantity, and

(iv) relative PSD error in the far-wake, similarly defined at  $(1.75, -0.25)$  for MP and at  $(5, 1)$  for FF,  $\epsilon_{S^{(2)}}$ .

The results for the  $N_p = 6$  case are shown in Fig. 28.

While overall reconstruction errors are higher in the FF regime compared to the MP case, the LSTM is very robust against sensor noise. This is contrasted by the FFNN, which exhibits poor performance: (i) in the MP case, the FFNN performs relatively poorly in the presence of  $\text{SNR} \leq 15$  dB sensor noise, which suggests that it is more prone to overfitting compared with the LSTM, and (ii) in the FF case, the FFNN consistently performs poorly even in the absence of added sensor noise (see right column of Fig. 28).

For all SNRs considered, the LSTM can achieve reconstruction errors consistently less than 10% for the second-order moments (shear stresses) and spectral content for both flow regimes. This performance is remarkable given that the (noisy) sensors are placed along the outer edges of the flow domain, where there is limited information about the energetic motions of the flow.

*b. Effect of sensor placement.* As seen from the wake development Sec. III A, the wake dynamics greatly depends on the location within the flow domain. Thus, it is expected that the correlation information that underpins the mapping  $\mathcal{F}$  is sensitive to the placement of sensors. In the present work, the effect of sensor placement is investigated by considering seven unique configurations that are summarized in Table II in Appendix B. We consider  $N_p = 6$  (configuration A), 4 (B and C), 3 (D and E), and 2 (F and G) sensors that are placed closer to (B and F) and farther downstream (C and G) from the cylinders, and near the top (D) and bottom (E) boundaries of the flow domain. These configurations enable analysis of estimator performance based on the differing extent of physical information encoded inside the input state. The important hyperparameters used to train the different configurations for both flow regimes are summarized in Table III in Appendix B.

The average velocity reconstruction errors ( $\overline{\epsilon_{\mathbf{u}}}$ ) for configurations B–G are shown in Fig. 29. The FFNN is more sensitive to the number of sensors and sensor location compared to the LSTM.

<sup>3</sup>The subscript “F” denotes the Frobenius norm.



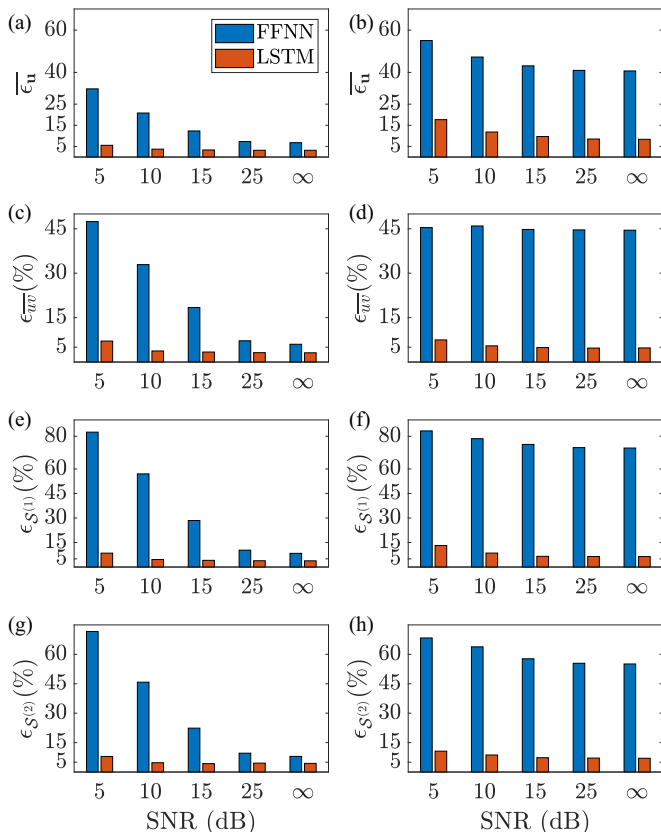


FIG. 28. Effect of increasing noise levels (SNR =  $\infty$  denotes the case where no noise is added) on reconstruction errors for the  $N_p = 6$  case in MP (left) and FF (right) regimes. Row 1 (a), (b) shows mean velocity residuals,  $\bar{\epsilon}_u$ ; Row 2 (c), (d) shows the residual in the Reynolds shear stresses computed via Eq. (20); Row 3 (e), (f) shows the residual in the PSDF in the near-wake and row 4 (g), (h) in the far-wake, computed via Eq. (21) at discrete locations.

In particular, for the FFNN, the errors increase as the number of sensors used decreases, and for configurations B, C and F, G, the performance is worse when the sensors are placed in the right-half of the domain, away from the cylinders. For the LSTM, configurations D and E yield roughly on average 10% higher residuals compared to configuration F, which has fewer sensors than the former two cases. This result suggests that the LSTM performance improves when sensors are placed near both the top and bottom boundaries of the domain (A–C, F, and G) and closer to the cylinders, rather than the top alone (D), or the bottom alone (E).

Both ANNs struggle to predict the flow with configuration G ( $N_p = 2$  sensors in the far-right of the domain). This is likely because the complex wake interactions may have limited memory and a damped signature of the near-wake flow by the time the flow reaches the sensors placed far downstream.

Overall, the LSTM is relatively insensitive to sensor location compared to the FFNN. Even with noisy  $N_p = 2$  sensors in configurations F/G, the LSTM yields acceptable reconstruction errors. Given its robustness to noise and sensor placements considered in this study, the LSTM estimator is an excellent choice when noisy sensors are inevitable and/or when there are limited options to install sensors.

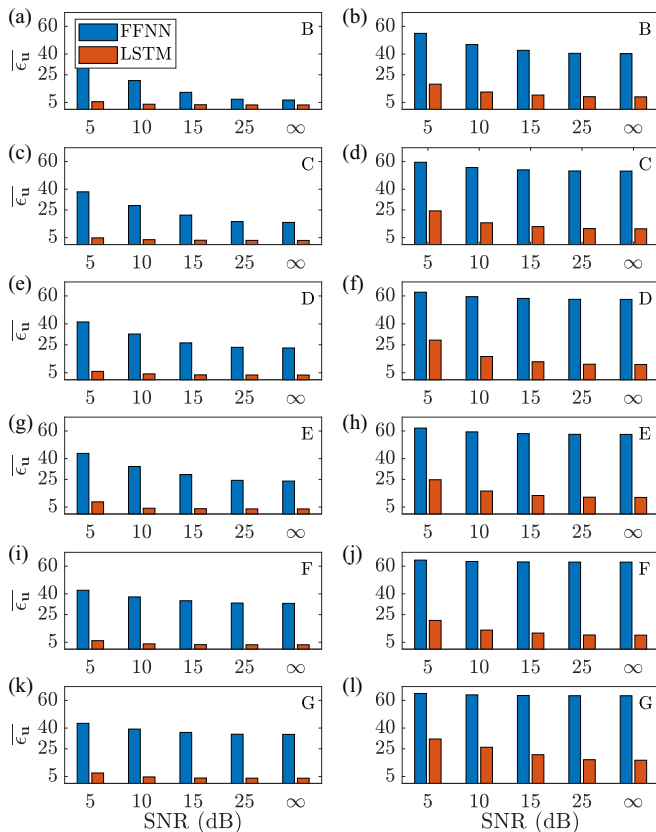


FIG. 29. Mean velocity reconstruction errors ( $\overline{\epsilon_u}$ ) for various sensor placement configurations in MP (left) and FF (right) regimes. The sensor configuration is identified in the top-right corner of each figure.

#### IV. CONCLUSIONS

In this study, sparse pressure sensor-based estimation of unsteady two-dimensional velocity flow fields is performed using modified linear and quadratic stochastic estimators (LSE/QSE-POD), shallow feedforward neural networks (FFNN), and long short-term memory (LSTM) networks in the context of low-dimensional representations based on the proper orthogonal decomposition (POD). Despite the rising use of powerful machine learning algorithms for flow estimation problems, a standard set of flows from which estimators can be developed, compared, and assessed is currently lacking within the fluid dynamics community. It was foreseen that a more organized development of these algorithms would make them more likely to be used in industrial applications, where their reliability can be assessed. Therefore, an intended outcome of this study is to support the accelerated advancement of sensor-based estimation techniques. The flow under investigation is the laminar ( $Re_D = 160$ ) wake development and wake interactions of two cylinders in proximity.

A detailed description of the flow physics is provided for two flow regimes: modulated-periodic (MP) and flip-flop (FF). Application of the POD to the MP and FF cases enables a clearer distinction of the dominant frequencies present in the wake and their associated dynamics. For both regimes, the first three mode pairs describe the majority of the fluctuating kinetic energy associated with the vortex shedding from the cylinders. We find that the MP regime, which is known to be characterized by a low-frequency phase modulation, also exhibits phase dislocation behavior that manifests at an

even lower frequency in the velocity and pressure spectra. For the FF flow, five distinct frequencies are identified, which correspond to the vortex shedding behavior and nonlinear secondary wake interactions.

The performances of the estimators are measured against low-order representations of the flow. Each estimator is assessed in terms of its ability to resolve the Reynolds shear stresses and pertinent spectral content in the wake. For the MP regime, the ANNs—particularly the LSTM—significantly outperform the LSE and QSE-POD approaches both instantaneously and in an average least-square sense. For the FF case, the LSTM network outperforms all other estimators on average by an order of magnitude in the time domain. Moreover, the LSTM-based estimator shows remarkable reconstruction performance in the frequency domain. Robustness of the estimators is demonstrated by considering pressure sensors with varying noise levels, revealing that the LSTM is far more robust than the FFNN in both flow regimes. Moreover, various sensor placement configurations are tested where combinations of two, three and four sensors with embedded noise are placed in different regions of the flow domain. Our study shows that the LSTM-based estimator is relatively insensitive to these sensor configurations compared to the FFNN. Based on the comprehensive characterization of the two flow cases and the detailed performance assessment of a variety of estimation methods on these data sets, the two-cylinder flows form an excellent suite of test cases to further develop and improve estimation algorithms.

## APPENDIX A: PROPER ORTHOGONAL DECOMPOSITION MODE SELECTION

### 1. Methodology for mode selection

Typically, the POD series is truncated to a user-defined rank number  $r$  to reconstruct a specified amount of energy, and the estimation algorithm is then trained to estimate this set of modes (e.g., Refs. [17,25]). In addition to energy, sufficient reconstruction of the spectral information is essential to recover the important unsteady dynamics that characterize the flow development. In the context of POD-based flow estimation, a robust methodology to incorporate the modes is required to (i) ensure statistical convergence of the modes as well as to (ii) obtain a reliable basis containing the important flow information. The convergence of mode  $i$  is computed as follows:

$$\varepsilon_{\lambda_{N_{\text{trial}}}}^i = \left| \frac{\lambda_{N_{\text{trial}}}^i - \lambda_{N_{\text{trial}}-m}^i}{\lambda_{N_{\text{trial}}-m}^i} \right| \times 100\%. \quad (\text{A1})$$

The convergence metric  $\varepsilon_{\lambda_{N_{\text{trial}}}}^i$  is defined as the relative % difference in eigenvalue of mode  $i$  computed from a data set containing  $N_{\text{trial}}$  snapshots relative to its eigenvalue when computed from a data set containing  $N_{\text{trial}} - m$  snapshots. Note that  $m$  is an arbitrary user-defined interval used to quantify the relative change of a POD eigenvalue after the addition of  $m$  snapshots to the data set. A value of  $m = 100$  is chosen to obtain a resolved trend of the mode energy convergence while considering computation time. It is expected for any given mode that this metric generally decreases with increasing number of snapshots. Therefore, an appropriate size for the training data set can be accordingly determined. The choice of a physics-informed threshold is still under investigation, and thus a threshold value of  $\varepsilon = 1\%$  is deemed sufficient to ensure statistical convergence in the POD energies of the flows studied in this work.

Successive spatial modes and the corresponding temporal coefficients derived from a converged training data set are then combined using Eq. (1) to form LORs. The spectral characteristics of the LORs are analyzed and compared to the original (nondecomposed) ensemble. Modes are added until the major dynamical phenomena are adequately captured. This is done by comparing the power spectral densities (PSDs) at two locations in the flow domain taken from the LORs and the original flow ensemble. The selected set of POD modes is then used for the estimation procedure.

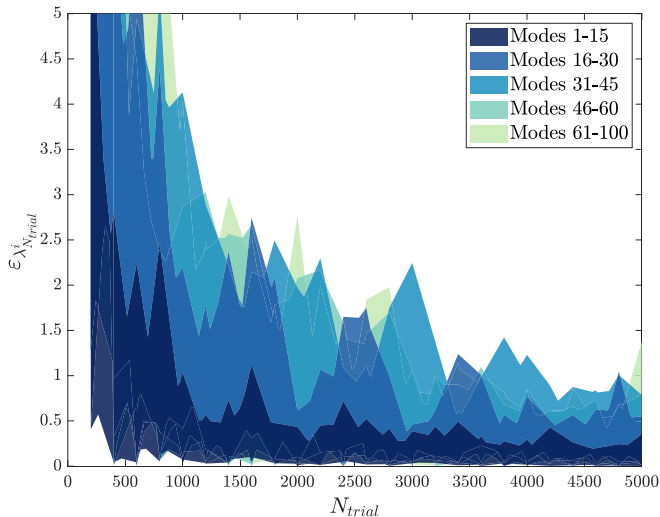


FIG. 30. Statistical convergence of the first 100 POD eigenvalues with increasing data set size for the MP flow calculated using Eq. (A1). Note that for each  $N_{\text{trial}}$  data set, the minimum and maximum  $\varepsilon$  values of the modes in that bin are plotted.

## 2. Results of selected modes for the modulated-periodic regime

The results of the POD mode selection procedure as outlined in Appendix A 1 are presented. For brevity, the results are only shown for the MP flow case. We present the results of the statistical convergence of mode energies to determine the size of the training data set, followed by results of the mode selection method.

An assessment of the statistical convergence of the first 100 POD modes is performed using Eq. (A1) over 49 data sets each containing  $N_{\text{trial}} = 200, 300, 400, \dots, 5000$  snapshots. To highlight the differing extents of convergence with respect to mode number, the convergence behaviors of modes 1–15, 16–30, 31–45, 46–60, and 61–100 are grouped into separate bins. The results are plotted in Fig. 30.

The results indicate that low-order modes converge faster than the high-order modes. This is expected due to the correlation-based approach taken by the snapshot POD algorithm. Modes 1–15 converge to  $<0.5\%$  by 3000 snapshots, while the higher-order modes generally require more snapshots to converge. The convergence in the higher-order modes appears to be irregular, without a monotonic decrease in relative energy difference. Following the 1% threshold criterion and accounting for high order modes, we estimate  $N_{\text{train}} = 4000$  snapshots to be sufficient to ensure statistical convergence.

The snapshot POD is computed over the training data set of  $N_{\text{train}} = 4000$  snapshots. We present  $v$ -PSDFs obtained from the (i) original data and (ii) the LOR constructed from the first 50 modes (99.8%  $k$ ), including (iii) the ratio of the latter to the former  $v$ -PSDFs in Fig. 31. An amplitude ratio that is closer to unity indicates better reconstruction of the dynamics at that frequency by the LOR. Figures 31(a) and 31(b) show results taken from regions dominated by the top and bottom cylinder wake dynamics, respectively. It is found that the first 50 modes are sufficient to reconstruct the important peak amplitudes in the power spectra. For the former location, we see that the 50-mode reconstruction can adequately capture frequency content up to  $f^* \approx 0.25$ , after which the spectral amplitudes significantly differ from those of the original. We observe similar behavior for the latter location except for few frequencies in the range  $0.023 < f^* < 0.13$ . Qualitatively, however, the LOR can capture the major peaks until  $f^* \approx 0.7$  at both locations. Note that over 200 modes are

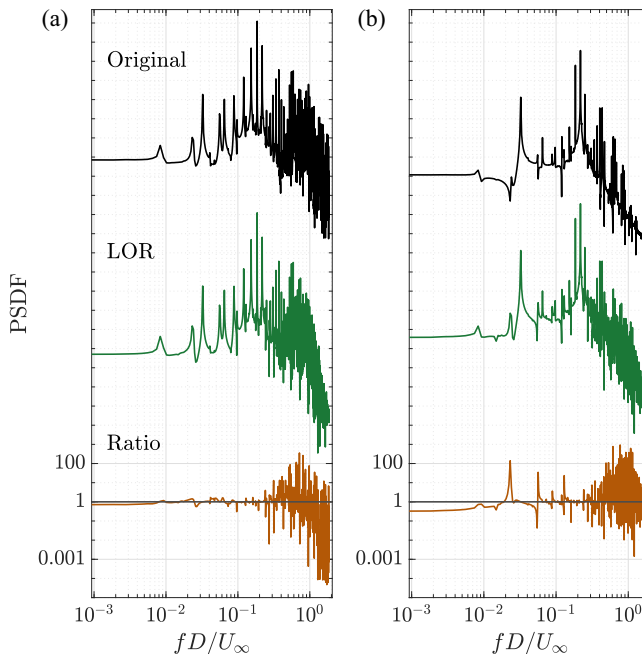


FIG. 31.  $v$ -PSDFs of the MP regime taken at  $(x/D, y/D) =$  (a)  $(4.2, 1.95)$  and (b)  $(1.75, -0.25)$  from the original ensemble, the 50-mode LOR and the ratio of the latter to the former PSDs. The axis at a ratio of unity is marked by a bolded solid line.

required to adequately resolve all the frequency components. Since a fixed six-sensor configuration is used, we refrain from using a large number of modes in this introductory work to account for computation time. The procedure is repeated for the FF regime and the first  $N_{\text{mode}} = 65$  FF modes (98.3%  $k$ ) are used for estimation.

## APPENDIX B: SENSOR PLACEMENT

The various sensor placement configurations used are provided in Table II. The mean velocity reconstruction errors for these configurations are shown in Fig. 29. A summary of all ANN hyperparameters used during training is provided in Table III.

TABLE II. Sensor placement configurations for both MP and FF regimes and sensors used based on labeling in Fig. 4.

Configuration	Sensors
A	1–6
B	1–4
C	3–6
D	1, 3, 5
E	2, 4, 6
F	1, 2
G	5, 6

TABLE III. Summary of pertinent ANN hyperparameters used during training for all sensor placement configurations A–G listed for both MP and FF regimes. Any hyperparameter not listed in this table uses the same value for the  $N_p = 6$  case (configuration A).

Configuration	A		B		C		D		E		F		G	
	MP	FF	MP	FF	MP	FF	MP	FF	MP	FF	MP	FF	MP	FF
Flow regime														
Number of Hidden Units, $d_1$	80	50	80	50	80	50	80	50	80	50	80	50	80	50
Activation Function, $\xi_1$	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh
Dropout Probability of Layer 1	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
Number of Hidden Units, $d_2$	160	160	160	160	160	140	160	120	160	120	160	100	160	80
Activation Function, $\xi_2$	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh	ReLU	tanh
Dropout Probability of Layer 2	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
Initial Learning Rate	0.002	0.008	0.002	0.008	0.002	0.05	0.002	0.1	0.002	0.1	0.002	0.01	0.002	0.1
Number of Hidden Units, $d$	152	180	152	180	152	180	152	180	152	180	152	180	152	180
Dropout Probability	0.061	0.075	0.061	0.075	0.061	0.075	0.061	0.075	0.061	0.075	0.061	0.075	0.061	0.075
Sequence Length, $S^*$	324	270	270	270	324	324	324	324	378	324	324	324	378	324
Initial Learning Rate	0.0965	0.0098	0.0965	0.0098	0.0965	0.05	0.0965	0.05	0.0965	0.05	0.0965	0.05	0.0965	0.05

- [1] S. L. Brunton and B. R. Noack, Closed-loop turbulence control: Progress and challenges, *Appl. Mech. Rev.* **67**, (2015).
- [2] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid mechanics, *Annu. Rev. Fluid Mech.* **52**, 477 (2020).
- [3] R. J. Adrian and P. Moin, Stochastic estimation of organized turbulent structure: Homogeneous shear flow, *J. Fluid Mech.* **190**, 531 (1988).
- [4] H. Clark, A. Naghib-Lahouti, and P. Lavoie, General perspectives on model construction and evaluation for stochastic estimation, with application to a blunt trailing edge wake, *Exp. Fluids* **55**, 1756 (2014).
- [5] J. P. Bonnet, D. R. Cole, J. Delville, M. N. Glauser, and L. S. Ukeiley, Stochastic estimation and proper orthogonal decomposition: Complementary techniques for identifying structure, *Exp. Fluids* **17**, 307 (1994).
- [6] K. Cohen, S. Siegel, J. Seidel, S. Aradag, and T. McLaughlin, Nonlinear estimation of transient flow field low dimensional states using artificial neural nets, *Expert Syst. Appl.* **39**, 1264 (2012).
- [7] D. Lasagna, L. Fronges, M. Orazi, and G. Iuso, Nonlinear multi-time-delay stochastic estimation: Application to cavity flow and turbulent channel flow, *AIAA J.* **53**, 2920 (2015).
- [8] V. Durgesh and J. Naughton, Multi-time-delay lse-pod complementary approach applied to unsteady high-Reynolds-number near wake flow, *Exp. Fluids* **49**, 571 (2010).
- [9] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.* **9**, 1735 (1997).
- [10] S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, *Int. J. Unc. Fuzz. Knowl. Based Syst.* **6**, 107 (1998).
- [11] J. Schmidhuber, D. Wierstra, and F. Gomez, Evolino: Hybrid neuroevolution/optimal linear search for sequence learning, in *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05)* (Morgan Kaufmann Publishers, San Francisco, CA, 2005), pp. 853-858
- [12] A. Graves, M. Liwicki, H. Bunke, J. Schmidhuber, and S. Fernández, Unconstrained on-line handwriting recognition with recurrent neural networks, in *Advances in Neural Information Processing Systems*, edited by J. Platt, D. Koller, Y. Singer, and S. Roweis, Vol. 20 (Curran Associates, Inc., 2008), pp. 577–584.
- [13] W. Luo, W. Liu, and S. Gao, Remembering history with convolutional LSTM for anomaly detection, in *2017 IEEE International Conference on Multimedia and Expo (ICME)* (IEEE, Piscataway, NJ, 2017), pp. 439–444.
- [14] C. Su, H. Huang, S. Shi, P. Jian, and X. Shi, Neural machine translation with Gumbel Tree-LSTM based encoder, *J. Visual Commun. Image Represent* **71**, 102811 (2020).
- [15] Z. Y. Wan, P. Vlachas, P. Koumoutsakos, and T. Sapsis, Data-assisted reduced-order modeling of extreme events in complex dynamical systems, *PLoS One* **13**, e0197704 (2018).
- [16] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Koumoutsakos, Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks, *Proc. R. Soc. A* **474**, 20170844 (2018).
- [17] A. T. Mohan, D. Tretiak, M. Chertkov, and D. Livescu, Spatio-temporal deep learning models of 3D turbulence with physics informed diagnostics, *J. Turbulence* **21**, 484 (2020).
- [18] S. Pawar, S. E. Ahmed, O. San, and A. Rasheed, Data-driven recovery of hidden physics in reduced order modeling of fluid flows, *Phys. Fluids* **32**, 036602 (2020).
- [19] H. Eivazi, H. Veisi, M. H. Naderi, and V. Esfahanian, Deep neural networks for nonlinear model order reduction of unsteady flows, *Phys. Fluids* **32**, 105104 (2020).
- [20] R. Maulik, R. Egele, B. Lusch, and P. Balaprakash, Recurrent neural network architecture search for geophysical emulation, in *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE, Piscataway, NJ, 2020), pp. 1–14.
- [21] R. Maulik, A. Mohan, B. Lusch, S. Madireddy, P. Balaprakash, and D. Livescu, Time-series learning of latent-space dynamics for reduced-order model closure, *Physica D* **405**, 132368 (2020).
- [22] S. R. Bukka, R. Gupta, A. R. Magee, and R. K. Jaiman, Assessment of unsteady flow predictions using hybrid deep learning based reduced-order models, *Phys. Fluids* **33**, 013601 (2021).
- [23] N. B. Erichson, L. Mathelin, Z. Yao, S. L. Brunton, M. W. Mahoney, and J. N. Kutz, Shallow learning for fluid flow reconstruction with limited sensors and limited data, *Proc. R. Soc. A* **476**, 20200097 (2020).

- [24] Z. Deng, Y. Chen, Y. Liu, and K. C. Kim, Time-resolved turbulent velocity field reconstruction using a long short-term memory (LSTM)-based artificial intelligence framework, *Phys. Fluids* **31**, 075108 (2019).
- [25] N. J. Nair and A. Goza, Leveraging reduced-order models for state estimation using deep learning, *J. Fluid Mech.* **897**, R1 (2020).
- [26] M. Le Provost, W. Hou, and J. Eldredge, Deep learning and data assimilation approaches to sensor reduction in estimation of disturbed separated flows, in *Proceedings of the AIAA Scitech 2020 Forum* (American Institute of Aeronautics and Astronautics, Orlando, Florida, 2020), p. 0799.
- [27] J. N. Kutz, Deep learning in fluid dynamics, *J. Fluid Mech.* **814**, 1 (2017).
- [28] M. M. Zdravkovich, The effects of interference between circular cylinders in cross flow, *J. Fluids Struct.* **1**, 239 (1987).
- [29] C. H. K. Williamson, Evolution of a single wake behind a pair of bluff bodies, *J. Fluid Mech.* **159**, 1 (1985).
- [30] H.-J. Kim and P. Durbin, Investigation of the flow between a pair of circular cylinders in the flopping regime, *J. Fluid Mech.* **196**, 431 (1988).
- [31] D. Sumner, S. Wong, S. J. Price, and M. P. Paidoussis, Two and three side-by-side circular cylinders in steady cross-flow, in *Proceedings of the 16th Canadian Congress of Applied Mechanics*, Vol. 1 (Quebec City, Quebec, 1997), p. 273.
- [32] D. Sumner, S. Wong, S. Price, and M. Paidoussis, Fluid behaviour of side-by-side circular cylinders in steady cross-flow, *J. Fluids Struct.* **13**, 309 (1999).
- [33] D. Sumner, S. Price, and M. Paidoussis, Flow-pattern identification for two staggered circular cylinders in cross-flow, *J. Fluid Mech.* **411**, 263 (2000).
- [34] J. Meneghini jr, F. Saltara, C. Siqueira, and J. Ferrari, Numerical simulation of flow interference between two circular cylinders in tandem and side-by-side arrangements, *J. Fluids Struct.* **15**, 327 (2001).
- [35] S. Kang, Characteristics of flow over two circular cylinders in a side-by-side arrangement at low Reynolds numbers, *Phys. Fluids* **15**, 2486 (2003).
- [36] S. Dong, G. E. Karniadakis, A. Ekmekci, and D. Rockwell, A combined direct numerical simulation-particle image velocimetry study of the turbulent near wake, *J. Fluid Mech.* **569**, 185 (2006).
- [37] K. Lee and K. S. Yang, Flow patterns past two circular cylinders in proximity, *Comput. Fluids* **38**, 778 (2009).
- [38] D. Sumner, Two circular cylinders in cross-flow: A review, *J. Fluids Struct.* **26**, 849 (2010).
- [39] Y. Zhou and M. Mahbub Alam, Wake of two interacting circular cylinders: A review, *Int. J. Heat Fluid Flow* **62**, 510 (2016).
- [40] C. H. K. Williamson, Vortex dynamics in the cylinder wake, *Annu. Rev. Fluid Mech.* **28**, 477 (1996).
- [41] H. G. Weller, G. Tabor, H. Jasak, and C. Fureby, A tensorial approach to computational continuum mechanics using object-oriented techniques, *Comput. Phys.* **12**, 620 (1998).
- [42] J. Lumley, The structure of inhomogeneous turbulence, in *Atmospheric Turbulence and Wave Propagation*, edited by A. M. Yaglom and V. I. Tatarski (Nauka, Moscow, 1967), pp. 166–177.
- [43] L. Sirovich, Turbulence and the dynamics of coherent structures: I. Coherent structures, II. Symmetries and transformations, III. Dynamics and scaling, *Q. Appl. Math.* **45**, 573 (1987).
- [44] C. Tropea and A. L. Yarin, *Springer Handbook of Experimental Fluid Mechanics* (Springer Science & Business Media, Berlin, 2007).
- [45] S. B. Pope, *Turbulent Flows* (IOP Publishing, Bristol, UK, 2001).
- [46] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems, and Symmetry* (Cambridge University Press, Cambridge, UK, 2012).
- [47] Y. Guezennec, Stochastic estimation of coherent structures in turbulent boundary layers, *Phys. Fluids* **1**, 1054 (1989).
- [48] N. E. Murray and L. S. Ukeiley, Modified quadratic stochastic estimation of resonating subsonic cavity flow, *J. Turbul.* **8**, N53 (2007).
- [49] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
- [50] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour, Dropout improves recurrent neural networks for handwriting recognition, in *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition* (IEEE, Piscataway, NJ, 2014), pp. 285-290.



- [51] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *Proceedings of the 3rd International Conference on Learning Representations (ICLR'15), San Diego, CA*, edited by Y. Bengio and Y. LeCun (ICLR, 2015).
- [52] P. J. Werbos, Backpropagation through time: What it does and how to do it, *Proc. IEEE* **78**, 1550 (1990).
- [53] J. Snoek, H. Larochelle, and R. P. Adams, Practical Bayesian optimization of machine learning algorithms, in *Advances in Neural Information Processing Systems*, Vol. 25 (NIPS, 2012).
- [54] H. F. S. Lui and W. R. Wolf, Construction of reduced-order models for fluid flows using deep feedforward neural networks, *J. Fluid Mech.* **872**, 963 (2019).
- [55] C. M. Bishop, Training with noise is equivalent to Tikhonov regularization, *Neural Comput.* **7**, 108 (1995).
- [56] J. McClure, C. Morton, and S. Yarusevych, Flow development and structural loading on dual step cylinders in laminar shedding regime, *Phys. Fluids* **27**, 063602 (2015).
- [57] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, edited by Y. W. Teh and M. Titterton, Proceedings of Machine Learning Research, Vol. 9 (PMLR, Sardinia, Italy, 2010), pp. 249–256.
- [58] A. M. Saxe, J. L. McClelland, and S. Ganguli, Dynamics of learning in deep linear neural networks, *NIPS Workshop on Deep Learning* (NIPS, 2013).
- [59] O. M. Griffin and M. S. Hall, Review—Vortex shedding lock-on and flow control in bluff body wakes, *J. Fluids Eng.* **113**, 526 (1991).