

Reconstruction of large-scale flow structures in a stirred tank from limited sensor data

Kirill Mikhaylov¹  | Stelios Rigopoulos²  | George Papadakis¹ 

¹Department of Aeronautics, Imperial College London, London, UK

²Department of Mechanical Engineering, Imperial College London, London, UK

Correspondence

George Papadakis, Department of Aeronautics, Imperial College London, London, SW7 2AZ, UK.

Email: g.papadakis@imperial.ac.uk

Funding information

Engineering and Physical Sciences Research Council, Grant/Award Number: Centre for Doctoral Training for Fluid Dynamics (EP/L016230/1) (Kirill Mikhaylov); UK Turbulence Consortium (EP/R029326/1), the UK Consortium for Turbulent Reacting Flows (EP/R029369/1), and the UK Materials and Molecular Modelling Hub (EP/P020194/1) (Computational Resources)

Abstract

We combine reduced order modeling and system identification to reconstruct the temporal evolution of large-scale vortical structures behind the blades of a Rushton impeller. We performed direct numerical simulations at Reynolds number 600 and employed proper orthogonal decomposition (POD) to extract the dominant modes and their temporal coefficients. We then applied the identification algorithm, N4SID, to construct an estimator that captures the relation between the velocity signals at sensor points (input) and the POD coefficients (output). We show that the first pair of modes can be very well reconstructed using the velocity time signal from even a single sensor point. A larger number of points improves accuracy and robustness and also leads to better reconstruction for the second pair of POD modes. Application of the estimator derived at $Re = 600$ to the flows at $Re = 500$ and 700 shows that it is robust with respect to changes in operating conditions.

KEYWORDS

computational fluid dynamics (CFD), fluid mechanics, turbulence

1 | INTRODUCTION

Stirred tanks are used widely in the chemical and process industries for mixing operations. The flow inside stirred tanks has been extensively studied experimentally and computationally. Although numerical results provide flow information within the whole vessel, experimental data are collected in planes if particle image velocimetry (PIV) is used, or at discrete probe points if laser doppler anemometry (LDA) is employed.

Both approaches have revealed the presence of large-scale vortical structures emanating from the rotating impeller blades. These have been extensively characterized by applying proper orthogonal decomposition (POD) mainly to two-dimensional PIV data captured on horizontal or vertical planes.^{1–8} When high sampling rates are used, these studies have found that the first two dominant POD modes are paired and correspond to the “trailing vortices” generated by the passage of the impeller blades through the fluid. The latter have long been

observed as prominent features of flows in stirred tanks agitated by a Rushton turbine and their location and trajectory has been extensively investigated.^{9–11} A second pair of POD modes with half the length scale and twice the frequency has also been found^{1–3}; these are formed by nonlinear interaction between the first pair of modes. The energy exchange between modes has been analyzed in Reference 4. A single (i.e., unpaired) mode with a much lower frequency and comparable magnitude to the second set of paired modes has also been observed in References 2, 3. In this case, the setup comprises a double Rushton turbine and the single mode arises from the interaction between the trailing vortices generated by the two turbines. All higher order modes typically represent small-scale turbulence. POD studies with long sampling time and low sampling frequency have been employed to investigate the macro-instabilities^{5,12} that have been observed in stirred tank flows.¹³ This flow feature has frequency of the order of $(0.02–0.1)F_N$, where F_N is the frequency of impeller rotation.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *AIChE Journal* published by Wiley Periodicals LLC on behalf of American Institute of Chemical Engineers.

To the best of our knowledge, only one study has applied POD to the three-dimensional (3D) flow field in a stirred tank.¹⁴ The author used Large Eddy Simulation (LES) to simulate the flow agitated by a three-blade pitched turbine using a combined rotating/static mesh approach. In total, 1500 snapshots of the 3D velocity field were stored over four rotations. The POD was carried out separately for the inner (rotating) domain surrounding the impeller and the outer (static) domain containing the rest of the flow.

The previous computational and experimental studies have clearly demonstrated that there are large-scale structures around the impeller that can be captured with a small number of dominant POD modes. Linear superimposition of the flow fields of the individual modes, scaled by the corresponding temporal coefficients, provides a reconstruction of the evolution of velocity fluctuations around a time-average flow. This approach provides a reduced order model (ROM) of the fluctuating flow. The necessary ingredients are the spatial modes and their time-varying coefficients; both are obtained from the stored velocity snapshots.

In the present study, we exploit this compact description of the flow, but we go a step further. In particular, we used system identification to construct a linear estimator of the temporal coefficients of the leading order modes using as inputs the velocity signals at one or more sensor points in the flow. The estimator has the form of a multiple-input multiple-output linear, time-invariant system. The system matrices and the order of the system are determined optimally using the N4SID algorithm.¹⁵ The latter belongs to the family of subspace identification methods; the reader may refer to Reference 16 for a detailed description and to Reference 17 for an overview.

Although system identification has been applied to many different areas (see examples in Reference 16), its application to fluid flows is relatively recent. It was first applied by Guzmán-Iñigo et al.¹⁸ to derive an estimator for the dynamics of perturbations in a laminar boundary layer flow. The perturbations were assumed to have very small amplitude, so their evolution was governed by the Navier–Stokes equations, linearized around the Blasius velocity profile. The assumption of small perturbations was relaxed in Reference 19, where the method was applied to reconstruct nonlinear flows exhibiting limit-cycle oscillations. In particular, the strongly periodic flow around a 2D aerofoil was considered. POD was used to decompose the perturbations around the time-average flow, and the temporal coefficients of the two leading modes were estimated very accurately (with fit values above 90%) using the signals of the two velocity components at a single probe point in the wake.

In both references cited earlier, the flow was 2D and laminar. In the present study, we apply system identification to the complex 3D flow inside an unbaffled stirred tank operating at the transitional Reynolds number of 600; this setup was also examined by Tamburini et al.²⁰ Although the majority of studies on stirred tanks have focused on baffled configurations due to their enhanced mixing properties, unbaffled tanks are widely used in industry due to the ease of cleaning and lack of dead zones behind the baffles when run at lower Reynolds numbers. We chose a Reynolds number of 600 for two reasons. First, flows in stirred tanks at low-to-moderate Reynolds number

have important industrial applications, for example, in bioreactors where the shear rate has to be kept low to avoid cell damage. Second, at this Reynolds number, the transitional flow can be modeled accurately using Direct Numerical Simulation (DNS). This is not critical for the application of the system identification algorithm; since we are interested in the dominant pair of modes, LES would have also been appropriate. This is however the first time the method is applied to complex transitional 3D flows, and we decided to employ an accurate simulation method. As will be seen later, this also allows us to investigate the asymptotic behavior of higher order POD modes.

The ability to estimate the temporal coefficients of the dominant POD modes from velocity signals at a very small number of sensor points opens new opportunities. For example, one can construct the estimator at nominal operating conditions and then apply it at off-design conditions. In this case, the approach has predictive power, as only the input velocity signals are necessary in order to reconstruct the entire instantaneous velocity field of the large-scale structures at the new conditions. We explore the accuracy of the reconstructed flow at *Re* numbers 500 and 700.

A word of caution regarding the reconstructed velocity perturbations for transitional and/or turbulent flows from limited velocity measurements is warranted here. Such flows exhibit a wide range of spatial scales, the largest of which are encapsulated in the dominant POD modes. The reconstruction framework is therefore expected to work better for large scales that have a strong footprint at the probe locations. Smaller scales have a more narrow extent and are more difficult (or even impossible) to recover. For this reason, in this article, we aim at reconstructing the large-scale 3D vortical structures around the impeller blades.

We close this section by stressing that system identification can be used also for process optimization, such as mixing enhancement. For example, if we know that a control variable (such as injection velocity) affects the mixing performance in a stirred tank, then we can use this as an input variable and the POD coefficients of the scalar concentration as the output, and construct a linear, time-invariant system, which can be used for mixing optimization. It is very difficult to solve this optimization problem directly using the 3D unsteady velocity and scalar fields, but system identification offers a more tractable approach.

This article is organized as follows: the flow conditions, numerical methodology, and extraction of POD modes are detailed in Sections 2 and 3, respectively. In Section 4, we show that the form of the identified multi-input, multi-output system arises naturally from the flow equations and estimation theory. In the same section, we also present the steps for the application of N4SID algorithm. Results at design and off-design conditions are reported in Section 5, and we conclude in Section 6.

2 | FLOW CONDITIONS AND COMPUTATIONAL DETAILS

The flow configuration considered is that of an unbaffled top-covered stirred tank, agitated by a six-blade Rushton impeller. A schematic is

FIGURE 1 Sketch of the stirred tank and the impeller geometry. The axes of the Cartesian coordinate system are shown in blue

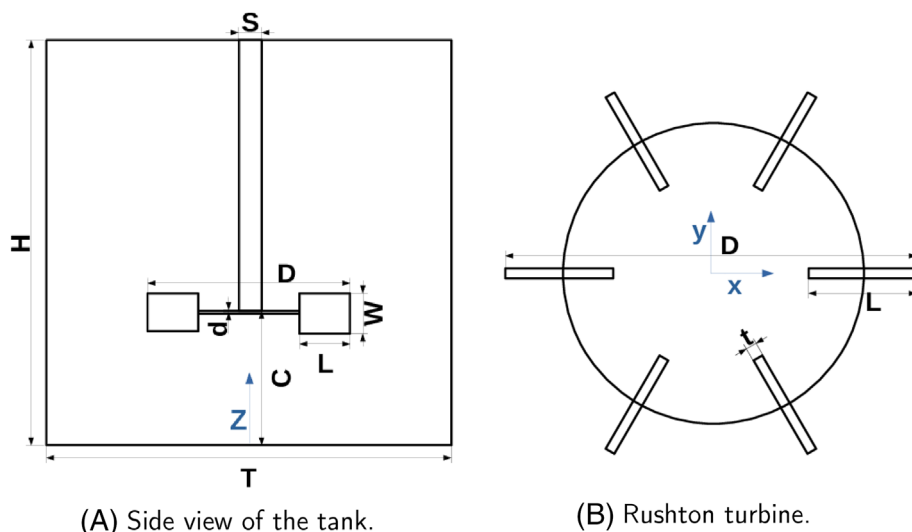


TABLE 1 Dimensions (mm) of the stirred tank considered

T	D	S	H	C	L	W	d	t
190	95	8.1	190	63.3	23.75	19	2.4	1.7

shown in Figure 1. The fluid is assumed to be incompressible and the Reynolds number, defined in the usual way as $Re = \rho ND^2/\mu$, where ρ is the density, N the impeller speed, and D the impeller diameter, is 600. The same geometry has been simulated with DNS by Tamburini et al.²⁰ and investigated experimentally by Scargiali et al.²¹; the physical dimensions are shown in Table 1. Simulations for two additional cases, with $Re = 500$ and 700 , were conducted in order to assess the ability of the reduced order model that was identified for $Re = 600$ to reconstruct the flow structures at off-design conditions.

Due to the absence of baffles, the geometry has rotational symmetry and it is more convenient to express and solve the flow equations in a coordinate system that rotates with the blades. In this system, shown in Figure 1, the impeller is held at a fixed position and the governing equations take the form

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} - [\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{x}) + 2\boldsymbol{\omega} \times \mathbf{u}], \quad (1a)$$

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

where $\mathbf{u}$ is the velocity vector, t the time, p the static pressure, ν the kinematic viscosity, $\mathbf{x} = (x, y, z)$ the position vector, and finally $\boldsymbol{\omega}$ is the angular velocity vector of the inertial coordinates. In the present case, this vector has a nonzero component only in the axial z direction, that is, $\boldsymbol{\omega} = (0, 0, \omega)$, which indicates that the rotation takes place in the (x, y) plane and ω is the angular velocity of the impeller. The two terms within square brackets in the momentum Equation (1a) represent the centrifugal and Coriolis acceleration terms. A Dirichlet boundary condition, $\mathbf{u}_w = -\boldsymbol{\omega} \times \mathbf{x}$, is imposed at the

top, bottom, and peripheral walls. This is known as the “Frozen Rotor” approach²² and is widely used when the geometry has rotational symmetry.

The above set of equations is solved using our in-house code “Pantarhei.” This is a finite volume, unstructured grid solver, which has been used extensively in simulating transitional and turbulent flows in various configurations, including stirred vessels.^{23–28} The convection terms are discretized in space using a second-order central scheme. The orthogonal viscous terms are treated implicitly, while the convection, nonorthogonal viscous, Coriolis, and centrifugal terms are all treated explicitly in time using quadratic extrapolation. A third-order accurate scheme is used to discretize the transient term. The pressure–velocity coupling is treated using the fractional step method. The code is parallelized with the aid of PETSc²⁹ and HYPRE libraries.³⁰

The computational domain was discretized using 100 cells in the radial direction, 516 in the tangential direction, and 149 in the vertical (axial) direction; the total number of cells was $\sim 7.6 \times 10^6$. This is between the third and fourth finest grids considered in Reference 20, with the authors choosing the third finest to perform their simulations. The time step for the $Re = 600$ case corresponds to 5621 time steps per impeller revolution and ensures a good temporal resolution of the flow. The corresponding CFL number was ~ 0.2 , significantly lower than that of the maximum value of ~ 0.6 required for the numerical scheme to be inviscidly stable. For the $Re = 500$ and $Re = 700$ cases, the time step was scaled to maintain approximately the same CFL number.

To check the grid independence of the results, we also performed simulations with a finer grid, consisting of 120, 636, and 192 cells in the radial, tangential, and vertical directions respectively, resulting in a total number of $\sim 19.6 \times 10^6$ cells. In Figure 2, we compare the profiles of time-averaged radial, tangential, and axial velocities along a line bisecting two blades at a height of 0.067 m, a few mm above the impeller disk, for the two grids. As can be seen, the matching of the time-averaged profiles is very good.

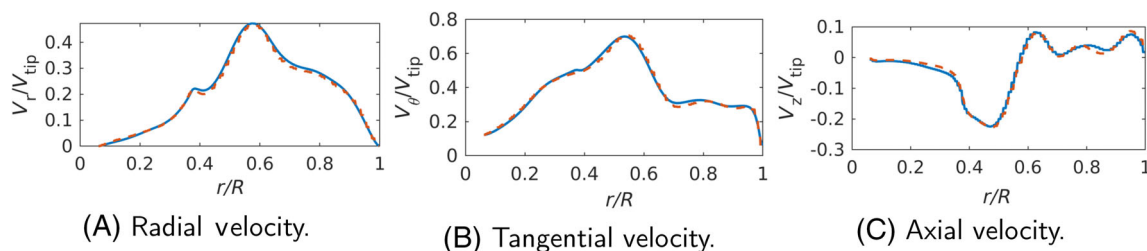


FIGURE 2 Time-averaged velocities along a line bisecting two blades at a distance of 67 mm from the tank bottom. Dashed line (—) high-resolution grid. Solid line (—) low-resolution grid

We also checked the resolution with respect to the Kolmogorov length scale, $\eta = (\nu^3/\epsilon)^{1/4}$, which is the smallest scale of the flow.³¹ The local ratio of the grid spacing to η was estimated from $V_{\text{cell}}^{1/3}/\eta_{\text{cell}}$, where V_{cell} is the cell volume. For the lower grid resolution, this ratio was found to be less than 2.0 everywhere, while the average value in the domain was 0.49. These metrics satisfy fully the resolution requirements for DNS³¹ and are comparable to those of other DNS studies in stirred tanks.^{20,32}

As a final check, we computed the power number, $N_p = \frac{P}{\rho N^3 D^5}$, where P is the time-average of total energy dissipation (i.e., due to mean and fluctuating fields) integrated over the tank volume. The value was found to be 2.1, which is close to 2.2 reported in References 20,21. Slight differences in geometry (e.g., the blade thickness) are likely to account for the small discrepancy. All the abovementioned tests confirm that the lower resolution grid is deemed sufficient, and this grid was used to obtain all the results presented henceforth.

3 | REDUCED ORDER MODEL OF THE FLUCTUATIONS AROUND THE TIME-AVERAGE FLOW

We use POD to construct a reduced order model of the velocity fluctuations around the mean flow. POD is a modal decomposition technique that yields a set of orthogonal modes that optimally represent the kinetic energy of the perturbation field. Below we provide only a brief summary; for more details refer to the book by Holmes et al.³³ and the review by Taira et al.³⁴

Using the Reynolds decomposition, the velocity vector $\mathbf{u}(\mathbf{x}, t) = (u, v, w)$ can be written as $\mathbf{u}(\mathbf{x}, t) = \bar{\mathbf{u}}(\mathbf{x}) + \mathbf{u}'(\mathbf{x}, t)$, where $\bar{\mathbf{u}}(\mathbf{x})$ is the time-average and $\mathbf{u}'(\mathbf{x}, t)$ the fluctuation. For later reference, the governing equations of the fluctuations, obtained by subtracting the time-averaged RANS equations from set (1), are provided as follows:

$$\frac{\partial \mathbf{u}'}{\partial t} + (\bar{\mathbf{u}} \cdot \nabla) \mathbf{u}' + (\mathbf{u}' \cdot \nabla) \bar{\mathbf{u}} + \frac{1}{\rho} \nabla p' - \nu \nabla^2 \mathbf{u}' + 2\boldsymbol{\omega} \times \mathbf{u}' = \overline{(\mathbf{u}' \cdot \nabla) \mathbf{u}'} - (\mathbf{u}' \cdot \nabla) \mathbf{u}', \quad (2a)$$

$$\nabla \cdot \mathbf{u}' = 0. \quad (2b)$$

We have arranged the equations so that all terms linear in the fluctuating variables are in the left-hand side. The centrifugal term cancels out, while the Coriolis term is linear and thus appears in the left-hand side. The velocity fluctuations can be expanded as

$$\mathbf{u}'(\mathbf{x}, t) = \sum_{i=1}^{\infty} a_i(t) \boldsymbol{\phi}_i(\mathbf{x}) \approx \sum_{i=1}^n a_i(t) \boldsymbol{\phi}_i(\mathbf{x}), \quad (3)$$

where $a_i(t)$ is the temporal coefficient, $\boldsymbol{\phi}_i(\mathbf{x})$ is the spatial distribution of mode i , and n is the order of the truncated expansion. We define the inner product between two fluctuating vectors $\mathbf{u}'_i$ and $\mathbf{u}'_j$ as

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \int_V \mathbf{u}_i^T(\mathbf{x}) \mathbf{u}_j(\mathbf{x}) dV \quad (4)$$

where T denotes the transpose operation and V the tank volume. If $\mathbf{u}'_i = \mathbf{u}'_j$, the norm is equal to twice the instantaneous turbulent kinetic energy integrated in the domain. The modes $\boldsymbol{\phi}_i(\mathbf{x})$ are obtained by maximizing the average projection of $\mathbf{u}'(\mathbf{x}, t)$ to $\boldsymbol{\phi}_i(\mathbf{x})$; the projection is defined using the inner product (Equation 4). It can be shown³³ that this optimization problem can be reduced to an eigenvalue problem, which we solve using the method of snapshots. More specifically, we store the velocity fluctuations around the time-average $\mathbf{u}' = \mathbf{u} - \bar{\mathbf{u}}$, $\mathbf{v}' = \mathbf{v} - \bar{\mathbf{v}}$, and $\mathbf{w}' = \mathbf{w} - \bar{\mathbf{w}}$ at each cell centroid $\mathbf{x}_k$ ($k = 1 \dots m$) at time t_j to form the vector

$$\mathbf{y}(t_j) = (u'(\mathbf{x}_1, t_j), \dots, u'(\mathbf{x}_m, t_j), v'(\mathbf{x}_1, t_j), \dots, v'(\mathbf{x}_m, t_j), w'(\mathbf{x}_1, t_j), \dots, w'(\mathbf{x}_m, t_j))^T, \quad (5)$$

and then stack K of these vectors column-by-column to form the snapshot matrix $\mathbf{Y}$

$$\mathbf{Y} = ((\mathbf{y}(t_1), \mathbf{y}(t_2), \dots, \mathbf{y}(t_K))) \quad (6)$$

It is shown in References 33,34 that the POD modes can be obtained from the singular value decomposition (SVD) of the matrix $\mathbf{V}^{1/2} \mathbf{Y}$, that is,

$$\mathbf{V}^{1/2} \mathbf{Y} = \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Psi}^T. \quad (7)$$

where $\mathbf{V}$ is a $3m \times 3m$ weighting matrix that accounts for the grid nonuniformity. For a second-order finite volume discretization, $\mathbf{V}$ is a diagonal matrix with the cell volume in the main diagonal. In Equation (7), $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values σ_i , while the left and right singular vectors are stored in matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$, respectively. The eigenvalues λ_i of the POD problem are $\lambda_i = \sigma_i^2/K$ and represent the energy content of mode i , the spatial modes $\phi_i(\mathbf{x})$ are the columns of $\mathbf{V}^{-1/2}\mathbf{\Phi}$, and the temporal coefficients $a_i(t)$ are obtained from the projection $\mathbf{Y}^T\mathbf{V}^{1/2}\mathbf{\Phi}$.

The process to extract the POD modes is the following. For all Reynolds numbers, the simulations first run for 40 impeller revolutions to allow the flow to reach a statistically steady state (10 probe points located in several areas of the tank were used to verify that this is the case). For the main $Re = 600$ case, the flow was simulated for another 120 rotations. During the simulation, we saved approximately 100 snapshots of the velocity field in the whole domain per rotation (i.e., one every 57 time steps). This provided a dataset of 12,000 snapshots from which the spatial distribution of the POD modes $\phi_i(\mathbf{x})$ ($i = 1 \dots n$) was computed using SLEPC,^{35,36} a library that provides parallel subroutines for solving SVD problems efficiently. Details about the temporal and spatial mode distributions are provided in Section 5.1.

From the above, it is clear that the same grid used for the flow simulation was also used for the POD mode extraction, that is, we did not down-sample. The size of the dataset is determined by two factors. First, it must be longer (by at least an order of magnitude) than the period of the flow features being examined. Second, there must be enough data to both train and validate the estimator. As such, a dataset of 20–40 rotations would have been sufficient for extracting the leading order modes and model training. We would then need additional 20–40 revolutions for model validation. In the present study, we used a conservatively long dataset (120 rotations) to ensure that we capture slowly varying features also (although these are not elaborated on in this study).

We now define the state vector of the temporal POD coefficients $\mathbf{a}(t) = (a_1(t), \dots, a_n(t))^T$ and obtain its evolution equation by performing Galerkin projection, that is, by substituting the expansion (3) into the fluctuating equation set (2) and taking into account that the basis is divergence-free and orthonormal. The resulting system takes the form

$$\frac{d\mathbf{a}}{dt} = \mathcal{A}\mathbf{a}(t) + \mathbf{F}(t) \quad (8)$$

where matrix $\mathcal{A}$ arises from the projection of the linear terms on ϕ_i while the forcing term $\mathbf{F}(t)$ arises from the nonlinear terms that appear in the right-hand side of (Equation 2a) as well as the fact that we retain only a finite number of POD modes (for more details on the derivation of Equation (8) see Reference 19). This equation represents a reduced-order model of the fluctuations. In the next section, we describe how to obtain an estimator that approximates the evolution of $\mathbf{a}(t)$ using information from a few sensor points in the flow. Note that we never compute the matrix $\mathcal{A}$ explicitly, all we need to know is that this matrix exists.

4 | CONSTRUCTION OF A LINEAR ESTIMATOR USING SYSTEM IDENTIFICATION

4.1 | Form of a linear dynamic estimator

We seek an estimate $\mathbf{a}_e(t)$ of the state vector $\mathbf{a}(t)$ from a set of velocity measurements $\mathbf{s}(t) = (s_1(t), \dots, s_p(t))^T$ at p sensor points located at $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ within the stirred tank. From expansion (3), we can write $\mathbf{s}(t) = \mathbf{S}\mathbf{a}(t) + \mathbf{g}(t)$, where matrix $\mathbf{S}$ depends on the sensor locations and the velocity components being recorded, while $\mathbf{g}(t)$ is noise arising from the fact that only a finite number of POD modes are retained in the state vector $\mathbf{a}(t)$. From linear estimation theory,³⁷ it is known that an optimal estimation $\mathbf{a}_e(t)$ can be obtained from

$$\frac{d\mathbf{a}_e}{dt} = \mathcal{A}\mathbf{a}_e(t) + \mathcal{L}(\mathbf{s}(t) - \mathbf{S}\mathbf{a}_e(t)) \quad (9)$$

where matrix $\mathcal{L}$, known as Kalman filter gain, is obtained from the solution of a Riccati equation. The forcing term in this equation involves the difference between the actual measurements, $\mathbf{s}(t)$, and their estimations, $\mathbf{S}\mathbf{a}_e(t)$. We can rearrange Equation (9) to get

$$\frac{d\mathbf{a}_e}{dt} = (\mathcal{A} - \mathcal{L}\mathbf{S})\mathbf{a}_e(t) + \mathcal{L}\mathbf{s}(t) \quad (10)$$

The above equation represents a linear state-space system with input $\mathbf{s}(t)$ and output $\mathbf{a}_e(t)$. Note that $\mathbf{S}$ can be computed from ϕ_i , but the matrices $\mathcal{A}, \mathcal{L}$ are not explicitly known. Equation (10) motivates the more general form of the estimator

$$\frac{d\mathbf{X}_e}{dt} = \mathbf{A}\mathbf{X}_e(t) + \mathbf{B}\mathbf{s}(t), \quad (11a)$$

$$\mathbf{a}_e(t) = \mathbf{H}\mathbf{X}_e(t), \quad (11b)$$

where we have introduced the state vector $\mathbf{X}_e(t)$. This is again a state-space system with input $\mathbf{s}(t)$ and output $\mathbf{a}_e(t)$ and becomes identical to (Equation 10) if $\mathbf{H} = \mathbf{I}$ (where $\mathbf{I}$ is the identity matrix), $\mathbf{A} = \mathcal{A} - \mathcal{L}\mathbf{S}$ and $\mathbf{B} = \mathcal{L}$. Note that the order of system (11), that is the size of the vector $\mathbf{X}_e(t)$, is not necessarily equal to the size n of $\mathbf{a}_e(t)$, which gives additional flexibility (in this case $\mathbf{H}$ is not a square matrix). Again matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{H}$ are not explicitly known and we seek to compute these matrices from the discrete values of the POD coefficients $\mathbf{a}(t_j)$ and the sensor data $\mathbf{s}(t_j)$ using system identification; this process is explained later.

4.2 | System identification from DNS data

Since the available data $\mathbf{a}(t_j)$ and $\mathbf{s}(t_j)$ from the DNS simulations are known at the discrete time instants t_j , we write system (11) in discrete form as

$$\mathbf{X}_e(t_{j+1}) = \mathbf{A}'\mathbf{X}_e(t_j) + \mathbf{B}'\mathbf{s}(t_j), \quad (12a)$$

$$\mathbf{a}_e(t_{j+1}) = \mathbf{H}'\mathbf{X}_e(t_j) \quad (12b)$$

where the primes indicate evolution matrices from t_j to t_{j+1} , for example, $\mathbf{A}' = e^{\mathbf{A}\delta t}$, where $\delta t = t_{j+1} - t_j$. Subspace system identification methods extract the matrices $\mathbf{A}'$, $\mathbf{B}'$, and $\mathbf{H}'$ from the input, $\mathbf{s}(t_j)$, and output, $\mathbf{a}(t_j)$, data such as the estimation $\mathbf{a}_e(t_j)$ optimally approximates $\mathbf{a}(t_j)$ over the length of the training dataset (see below). For system (12), this is known as deterministic identification. For stochastic systems, as in our case, noise terms $\mathbf{w}(t_j)$ and $\mathbf{v}(t_j)$ are included, that is,

$$\mathbf{X}_e(t_{j+1}) = \mathbf{A}'\mathbf{X}_e(t_j) + \mathbf{B}'\mathbf{s}(t_j) + \mathbf{w}(t_j), \quad (13a)$$

$$\mathbf{a}_e(t_{j+1}) = \mathbf{H}'\mathbf{X}_e(t_j) + \mathbf{v}(t_j) \quad (13b)$$

and stochastic system identification methods can also provide the covariance matrices, $\mathbf{Q}'$, $\mathbf{S}'$, and $\mathbf{R}'$, defined from

$$\mathbf{E} \left[\begin{pmatrix} \mathbf{w}(t_p) \\ \mathbf{v}(t_p) \end{pmatrix} \begin{pmatrix} \mathbf{w}(t_q)^T & \mathbf{v}(t_q)^T \end{pmatrix} \right] = \begin{pmatrix} \mathbf{Q}' & \mathbf{S}' \\ \mathbf{S}'^T & \mathbf{R}' \end{pmatrix} \delta_{pq}, \quad (14)$$

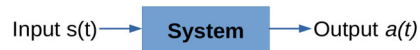
where $\mathbf{E}$ denotes the expectation operator and δ_{pq} is the Kronecker delta. This equation indicates that $\mathbf{w}(t_p)$ and $\mathbf{v}(t_p)$ are assumed to be white noise sequences.

The identification process consists of three procedural steps, which are shown schematically in Figure 3. In the first step, the DNS simulations are performed to generate the input and output data, $\mathbf{s}(t_j)$ and $\mathbf{a}(t_j)$, respectively. In the second step, the linear model matrices are computed by maximizing the fit between the output of the system and the model for part of the original data, known as training dataset. In the third step, the performance of the generated model is assessed on a validation dataset, which is different to the training dataset.

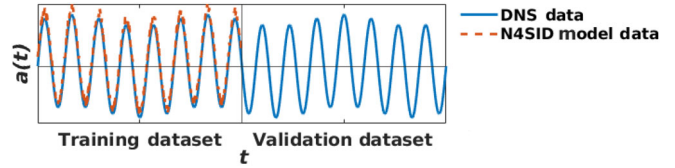
At the heart of the second step lies the singular value decomposition of a weighted matrix constructed using only the input and output data. The order of the estimator is obtained by the number of the retained singular values. In many cases, there is only a small number of large singular values and the rest can be discarded; this yields an optimal estimator order. From the corresponding left and right singular vectors, the states $\mathbf{X}_e$ are first obtained and then the matrices $\mathbf{A}'$, $\mathbf{B}'$, $\mathbf{H}'$, $\mathbf{Q}'$, $\mathbf{S}'$, and $\mathbf{R}'$ are computed using least squares; for more details, refer to References 16,38.

In total, 40 rotations, corresponding to 4000 data points, were used as the training dataset. A second nonoverlapping set of 40 rotations were used as the validation dataset. The output signals were taken to be the temporal components of the largest two POD mode pairs (more details are provided in the next section). Due to the transitional nature of the flow, the noise terms were retained. The N4SID function from the MATLAB toolbox³⁹ was used for the computation of the system matrices.

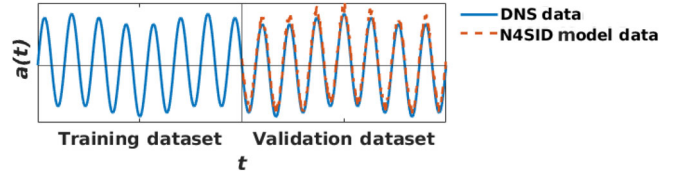
The location of the selected sensor points at which velocity fluctuations were recorded is important for the performance of the



(A) Step 1: Data generation.



(B) Step 2: Extraction of system matrices using a training dataset.



(C) Step 3: Model assessment using a validation dataset.

FIGURE 3 Procedural steps for system identification

algorithm. These points were selected based on the topological features of trailing vortices behind the blades; we elaborate on this later.

Finally, the percentage fit between the modeled and the true output for the i th mode was evaluated from

$$\text{FIT}_i[\%] = 100 \times \left(1 - \frac{\|a_i(t_k) - a_{i,e}(t_k)\|}{\|a_i(t_k) - a_i(t_k)\|} \right), \quad (15)$$

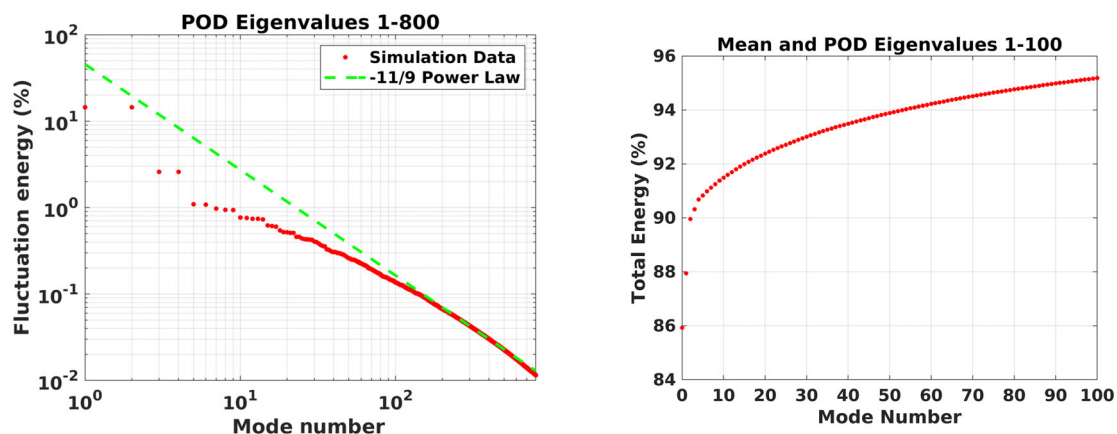
where $\|\cdot\|$ is the 2-norm of a vector, the subscript e denotes the estimated output, and the overbar denotes time-average (here equal to 0). A large value of FIT_i (max 100) indicates good matching between $a_i(t)$ and $a_{i,e}(t)$.

5 | RESULTS AND DISCUSSION

5.1 | Characterization of POD modes

Figure 4A shows the energy content of each POD mode as a fraction (percentage) of the turbulent kinetic energy integrated in the vessel. Modes 1 and 2 have identical contributions, equal to 14.32% each. Modes 3 and 4 also have nearly identical energy, but much smaller contributions, equal to 2.57% each. Modes 5 and above each contribute 1% or less to the fluctuation energy. It is clear also that the decay in the energy content is slow, with the first 100 modes capturing about 65% of the fluctuation energy. In Figure 4B, we plot the cumulative contribution, but we have included the time-average flow (Mode 0) as well. The mean flow contains a large portion, equal to 85.92%, of the total kinetic energy of the flow, and together with the first 100 modes capture more than 95% of the energy.

These results are very similar to those obtained for both a single Rushton turbine^{1,7} and a pair of Rushton turbines.^{2,3} It is interesting



(A) Percentage of fluctuation energy for modes 1 to 800.

(B) Cumulative percentage of total kinetic energy contained in the mean and first 100 modes.

FIGURE 4 Energy content of proper orthogonal decomposition modes

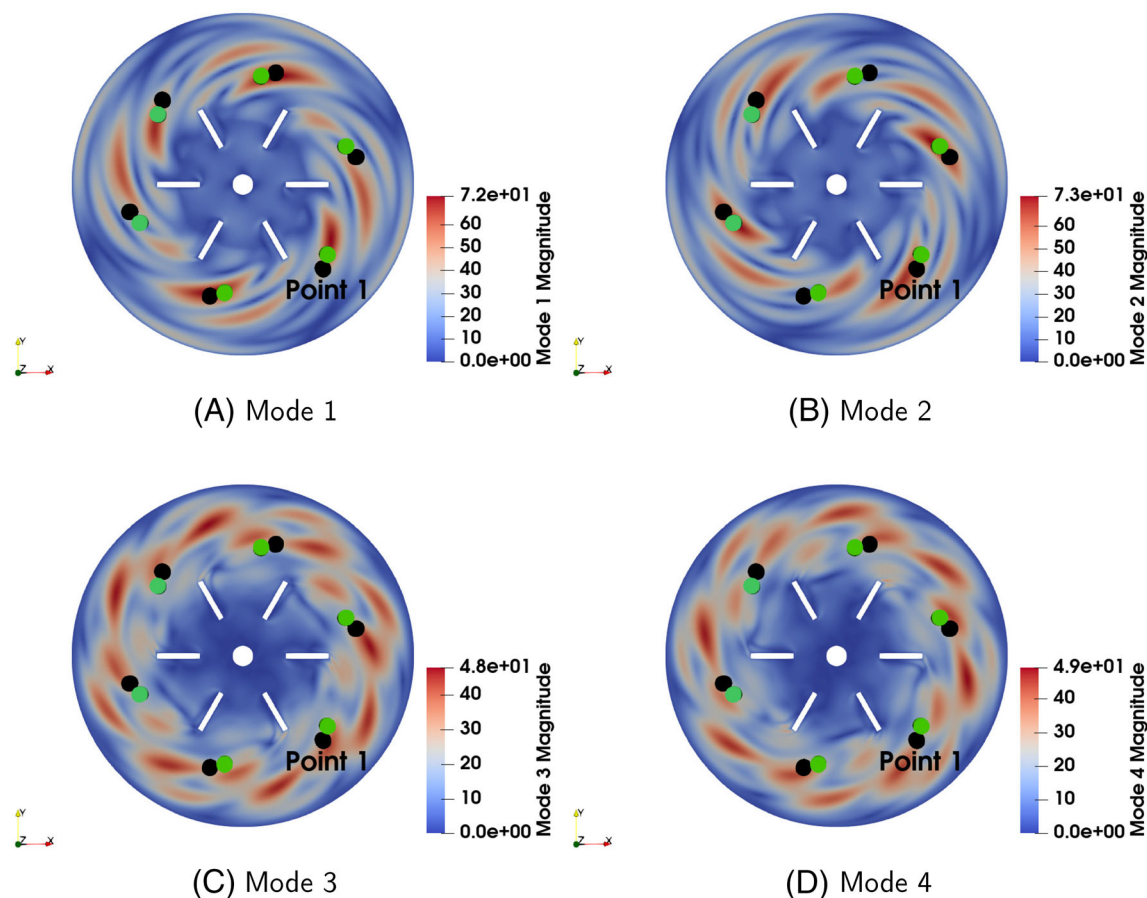


FIGURE 5 Contours of the velocity magnitudes of the first four proper orthogonal decomposition modes on a z plane just above the impeller disk at the height of the maximum RMS of velocity fluctuations (67 mm from the tank bottom). Black and green points denote the locations of maxima of the RMS of the velocity fluctuations and the RMS of the sum of Modes 1 and 2, respectively

to see that the slow decay of the energy content in Figure 4A follows a $-11/9$ power law for mode numbers between ≈ 130 and 600 (shown as a dashed green line). This power law for POD modes was first

derived theoretically in Reference 40 from the well-known $\kappa^{-5/3}$ law for the energy density with respect to wavenumber κ for homogeneous isotropic turbulence. This law has also been observed

experimentally in References 1,7 using 2D PIV for similar Reynolds numbers as in our case. The fact that we also capture it indicates that the complex nature of the examined transitional flow is accurately resolved using DNS. The nearly identical energies of the first and second, and of the third and fourth modes suggest that these modes form two pairs; this has also been observed in the literature.

Contour plots of the velocity magnitudes of Modes 1–4 in a horizontal slice just above the impeller disk (67 mm from the bottom of the tank) are shown in Figure 5. Both modes have rotational symmetry of 180° and show vortex structures that have also been observed in other studies^{1,2,4} and what is considered a mesoscale length scale.⁴¹ An animation of the velocity field $\mathbf{u}'_{1\&2}(\mathbf{x},t) = a_1(t)\boldsymbol{\phi}_1(\mathbf{x}) + a_2(t)\boldsymbol{\phi}_2(\mathbf{x})$ shows the periodic formation and shedding of vortices in the wake of each blade; once released, these vortices rotate around the vessel (the animation is provided as Video S1). The length scale of Modes 3 and 4 is approximately half that of the first two modes, a feature also noted in Reference 1. The locations of the maxima of the root mean square (RMS) velocities $\sqrt{\mathbf{u}^2}$ and $\sqrt{\mathbf{u}'_{1\&2}^2}$ are marked with black and green dots, respectively. Due to rotational symmetry, the dots are placed every 60° around the circumference, and the two sets are close to each other because Modes 1 and 2 together account for about 30% of the turbulent kinetic energy.

Contour plots of the velocity magnitudes in a vertical plane in line with the blades for Modes 1–4 are shown in Figure 6. The modes are symmetric with respect to the z -axis, as expected, and this indicates that a large enough number of samples were used. As can be seen, the modes impinge on the wall and are deflected in the vertical direction both upwards and downwards, thereby occupying a relative large fraction of the vessel. Of course, their presence is strongly close to the impeller. The z location of the maximum $\sqrt{\mathbf{u}'_{1\&2}^2}$ is indicated with a horizontal green line; this is at $z = 67$ mm, that is, the plane shown in Figure 5. The presented mode plots are very similar to the turbulent kinetic energy contour plots of Reference 42, which are phase-resolved at specific angles behind the impeller.

The first six modes come in pairs. The time signals of the temporal coefficients $a_i(t)$ for each pair are offset by a fixed time delay (plots not shown due to lack of space) and have identical spectra. In Figure 7, we plot the power spectral density of $a_i(t)$ for modes $i = 1, 3, 61$, and 251 . Modes 1 and 2 both have a very clear dominant frequency of $1.5F_N$, where F_N is the blade rotation frequency, while the frequency for Modes 3 and 4 is the first harmonic, that is, $3F_N$.

The spectra of $a_5(t)$ and $a_6(t)$ peak at $0.70F_N$ (close to the subharmonic of the dominant $1.5F_N$), but the spectral content starts to appear in other frequencies as well. Modes 1–6 all show clear

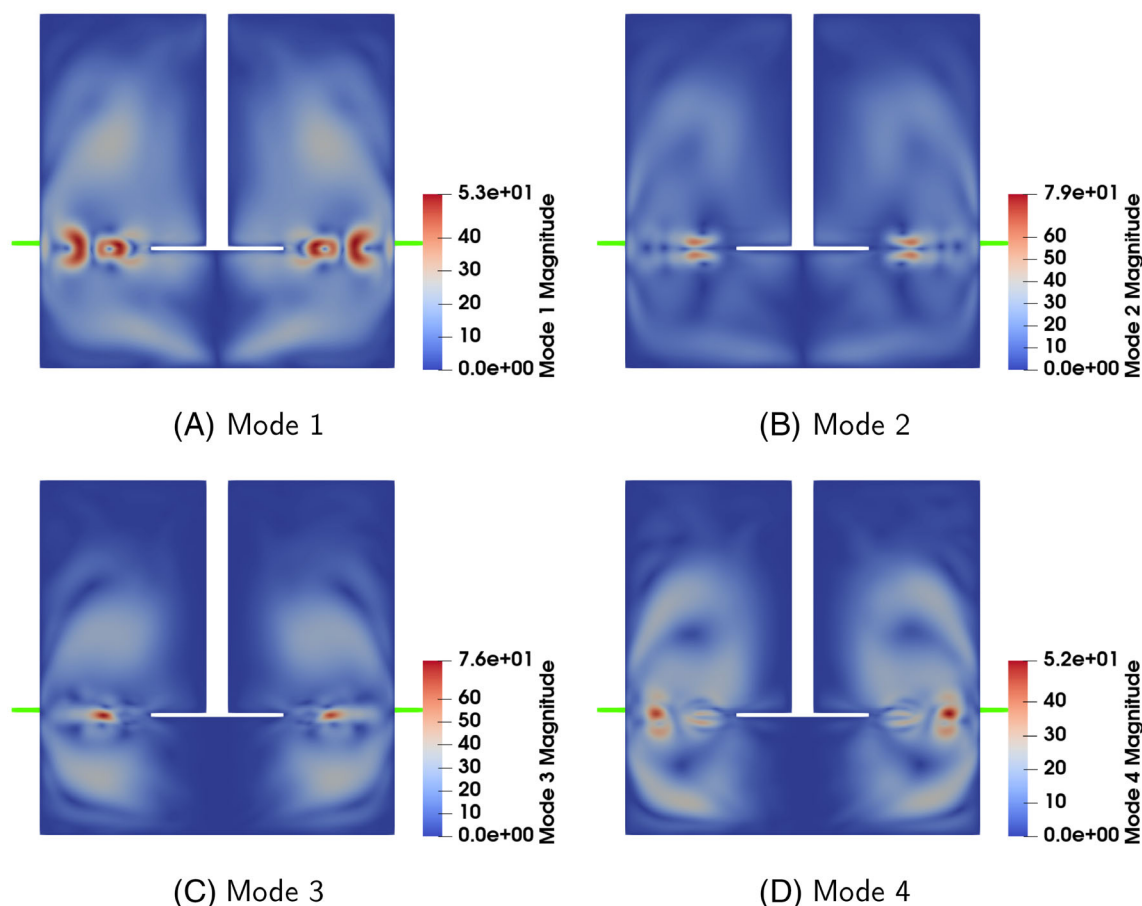


FIGURE 6 Contours of the velocity magnitudes of the first four proper orthogonal decomposition modes on a vertical plane bisecting the blades. The horizontal green lines indicate the height of maximum $\sqrt{\mathbf{u}^2}$ and $\sqrt{\mathbf{u}'_{1\&2}^2}$ values

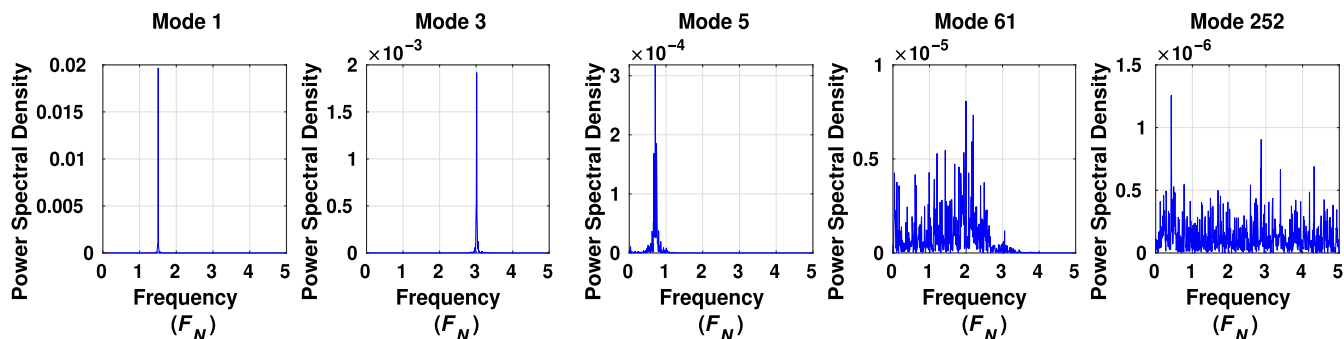


FIGURE 7 Power spectral density of proper orthogonal decomposition modes 1, 3, 5, 61, and 252

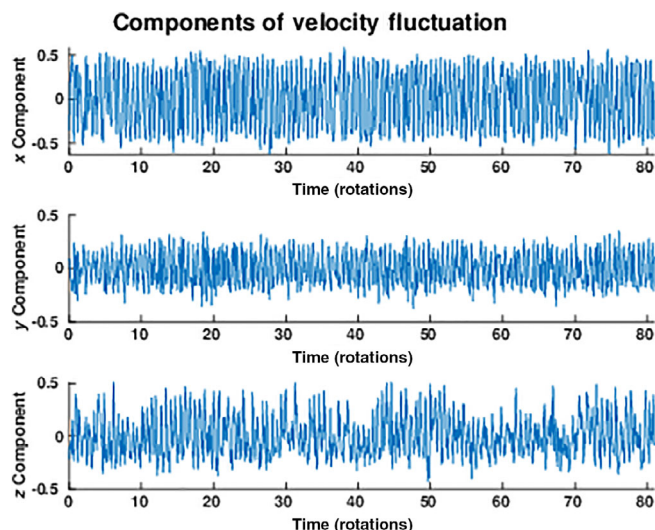


FIGURE 8 Velocity fluctuations at the point of maximum turbulent kinetic energy

evidence of structure and thus are far from the previously mentioned $-11/9$ power law, which is valid for homogeneous isotropic turbulence (refer to Figure 4A). As the mode number increases, the power spectrum becomes increasingly more noisy, with smaller energy content and less clear dominant frequencies, as evident for Mode 61. For higher modes, such as 252, the whole spectrum is filled and there is no dominant frequency, suggesting that the mode represents homogeneous isotropic turbulence and hence obeys the $-11/9$ power law. We examined also the probability density function plots of $a_i(t)$ (not shown) and found that these morph smoothly to Gaussian distribution with increasing mode number, as expected.

5.2 | Reconstruction of the first and second modes from point data

In this section, we consider the quality of reconstruction of the large-scale structures behind the impeller blades that correspond to Modes

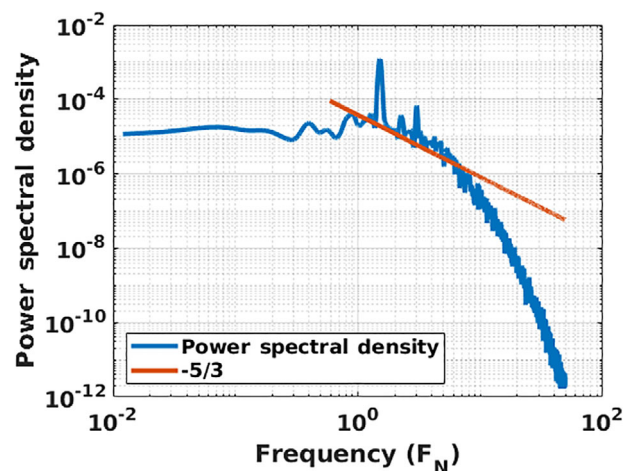


FIGURE 9 Power spectrum of the x velocity component at probe point 1 for $Re = 600$

1 and 2, that is, the output vector in Equation (12) which consists of the temporal coefficients $\mathbf{a}_e(t) = [a_{1,e}(t), a_{2,e}(t)]^T$. We start by employing as input the velocity signals from a single probe point, so $\mathbf{s}(t) = [u_1(t), v_1(t), w_1(t)]^T$, and proceed to explore the effect of including information from more points. More specifically, we consider between two to six points; in the last case, there is one point in the wake of each blade, that is, $\mathbf{s}(t) = [u_1(t), v_1(t), w_1(t), \dots, u_6(t), v_6(t), w_6(t)]^T$.

5.2.1 | Reconstruction using the velocity signals from a single sensor point

The location of the probe point is important for the quality of reconstruction.¹⁹ In the present work, this point was chosen to be at the position of maximum turbulent kinetic energy and is indicated with a black dot (Point 1) in Figure 5. The time signals of the velocity fluctuations for 80 revolutions at this point are shown in Figure 8 and the corresponding power spectrum for the x component in Figure 9 (the spectra for the other two components are very similar and not shown for brevity). The time signals are chaotic due to the transitional nature

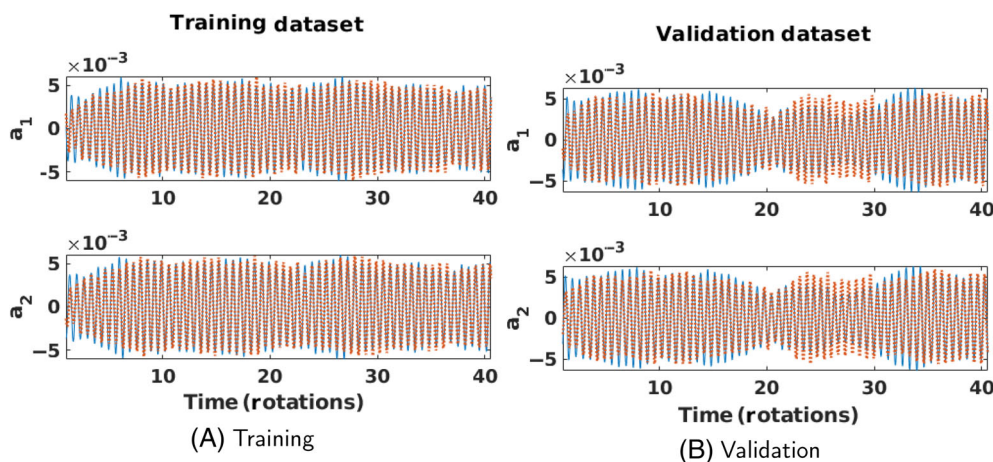


FIGURE 10 System identification using velocity data from probe point 1 (sixth-order ROM): DNS (blue) and N4SID (orange)

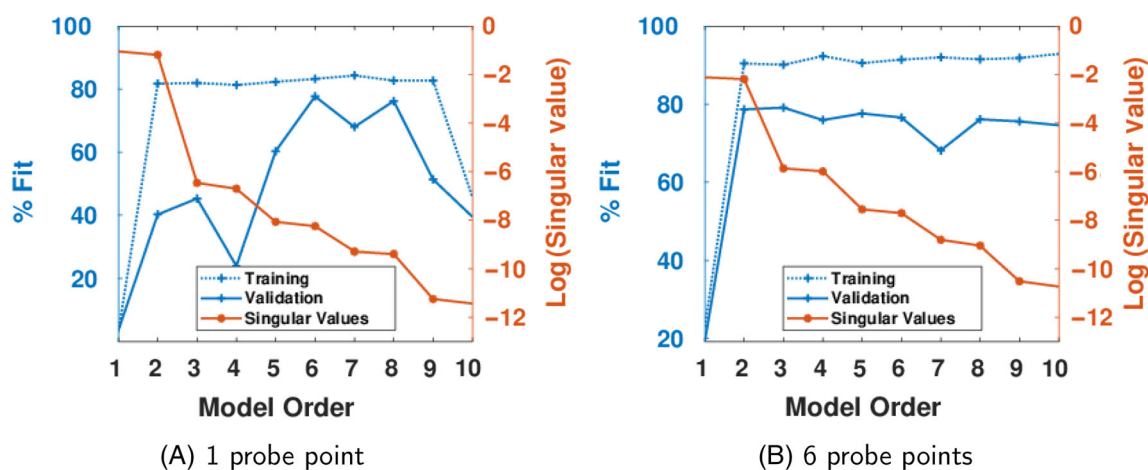


FIGURE 11 Effect of model order on the accuracy of reconstruction for the training and validation datasets using 1 (A) and 6 (B) probe points

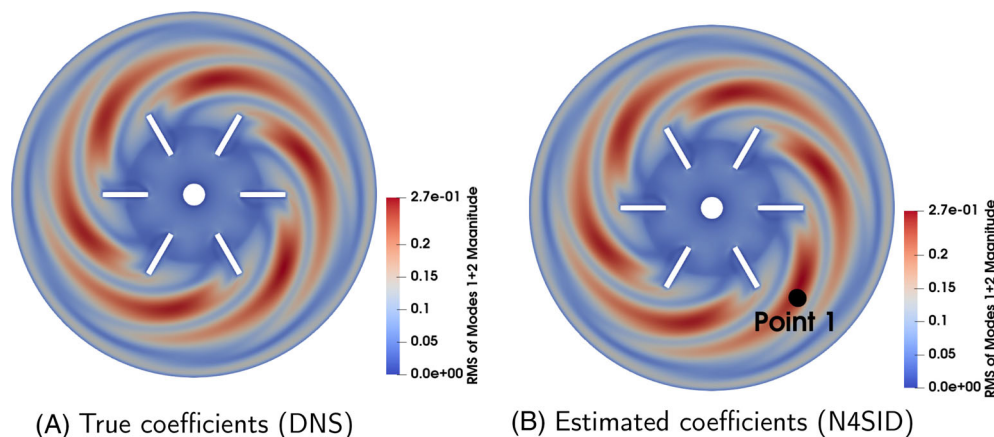
of the flow at the examined Re number. The spectral peaks at frequencies $1.5F_N$ and $3F_N$ are clearly visible in Figure 9. Moreover, a short range of frequencies (half a decade) follows the $-5/3$ power law due to the moderate value of the Reynolds number.

As mentioned in Section 4.2, velocity data from 40 revolutions were used for training and extraction of the matrices $\mathbf{A}'$, $\mathbf{B}'$, $\mathbf{H}'$, $\mathbf{Q}'$, $\mathbf{S}'$, and $\mathbf{R}'$. In Figure 10, we plot the variation of the temporal coefficients $a_1(t)$ and $a_2(t)$ as obtained from DNS (gray line) and as predicted by the estimator (using a sixth-order model). It can be seen that for the training dataset (Figure 10A), the predictions for both coefficients follow closely the DNS signals, indicating that the frequency and the time delay are correctly captured. This indicates that although the input signal contains a multitude of frequencies, as evidenced by the spectra in Figure 9, the matrices of the estimator are computed in a way that extracts the relevant frequency and phase difference while at the same time suppressing the other frequencies. This is a difficult task, but the algorithm performs very well. For model validation, we use the input signal at the same point from the following 40 revolutions (from 40–80), and employing the matrices obtained in the training phase, we can estimate the evolution of the two temporal

coefficients $a_1(t)$ and $a_2(t)$. Comparison with DNS is shown in Figure 10B. For the initialization of system (Equation 12), we use $\mathbf{a}_e(0) = \mathbf{0}$, but within a few revolutions the estimated values approach and capture very satisfactorily the DNS values. There are localized differences at several time instances but generally the agreement is quite good, especially bearing in mind that information is provided by a single sensor point only.

Next, we explore the effect of the order of the estimator (Equation 12), that is, the number of states of the vector $\mathbf{X}_e$, on the accuracy of reconstruction as determined by the FIT[%] value, shown in Equation (15). The results are shown in Figure 11A for both the training and the validation datasets (in computing the fit for the latter, we have excluded the first 500 data points to allow the model to “spin-up” from the initial zero state, see Figure 10). It can be seen that, for the training dataset, the fit for model orders between 2 and 9 is consistently over 80%. However, the performance of the generated models when tested on the validation dataset shows significant variations. The optimal fit values, 77% and 76%, are obtained for model orders 6 and 8, respectively. MATLAB identifies the eighth-order model as optimal. This is determined by examining the variation

FIGURE 12 Contour plots of $\sqrt{u_{1&2}^2}$ computed using (A) the true and (B) the estimated temporal coefficients $a_1(t)$ and $a_2(t)$. The black dot indicates the single sensor used



of the singular values, which are drawn with purple line in Figure 11A. It can be seen that the singular values drop by two orders of magnitude between the eighth and ninth order, indicating that orders higher than 8 do not offer significant advantage.

In Figure 12, we plot the spatial distribution of the velocity RMS due to the first two modes, $\sqrt{u_{1&2}^2}$, computed using the true (left panel) and the estimated (right panel) temporal coefficients $a_1(t)$ and $a_2(t)$. It can be seen that although there are slight deviations between the true and estimated coefficients over time, as evidenced in Figure 10, the time-average distributions are almost indistinguishable.

5.2.2 | Effect of the number of sensor points on the quality of reconstruction

We now investigate the quality of reconstruction as a function of additional probe points. To this end, we exploit the 60° rotational symmetry of the system and consider 2, 3, 4, 5 and 6 points placed at the locations of maximum turbulent kinetic energy. The points are located at the same radial distance from the impeller axis, but the azimuthal angle of each sensor location increases by 60° in the anticlockwise direction with respect to the preceding location; all probe points are indicated by black dots in Figure 5. The fit for the training dataset is very high for all cases considered. For example, using six probes, the fit is almost 90% as shown in Figure 11B. For the verification dataset, the fit reaches almost 80% and remains almost unaffected by the model order. Note also that a high fit value can be obtained even with very small models that contain only two states. This is in contrast to the performance with a single probe, which requires a model of higher order, as shown in Figure 11B. In general, as the number of sensor points increases, the optimal model order decreases; at the same time, the fit increases and the estimators become more robust.

5.3 | Reconstruction of the third and fourth modes

The ability of the identification algorithm to estimate the third- and fourth-order modes was also examined. For the training step, the

same velocity signals from the same six probe points as in the previous section were employed as inputs and the temporal coefficients $a_3(t)$ and $a_4(t)$ were used as outputs. For the training datasets, the model fits were good, with values $\approx 85\%$. However, the performance on the validation dataset was lower, with fits typically around 50%. The frequency of oscillation of the temporal coefficients, equal to $3F_N$ (refer to Figure 7), was predicted very well, with the loss of accuracy coming from the error in the prediction of the local amplitude. The results for the training and validation datasets for the optimal eighth-order model are shown in Figure 13.

The poor fit in the validation dataset is probably related to the fact that the selected sensor points do not correspond to the locations of the largest amplitude of Modes 3 and 4, as can be clearly seen from Figure 5C,D. Additionally, the energy content of these modes is substantially smaller (about 5 times) than that of the first pair, refer to Figure 4A (this can also be noticed by comparing the peaks at frequencies F_N and $3F_N$ in the velocity spectrum plot shown in Figure 9).

The velocity RMS due to Modes 3 and 4, $\sqrt{u_{3&4}^2}$, computed using the true and the estimated temporal coefficients, $a_3(t)$ and $a_4(t)$, are shown in the left and right panels, respectively, of Figure 14. As can be seen, the RMS of the velocity field corresponding to the second mode pair is predicted quite accurately; the error in the maximum value is less than 10%.

In Figure 15, we compare the turbulent kinetic energy obtained from the DNS simulations (left panel) and the energy from the superposition of the contributions of both mode pairs using the estimated coefficients $a_{i,e}(t)$, $i = 1...4$ (right panel). The plots are qualitatively identical, but quantitatively the estimated peak value is about 55% of the DNS value. This is expected since only four modes were used for reconstruction. It is interesting to notice that although in the full tank these four modes account for $\approx 34\%$ of the total kinetic energy (Figure 4A), the peak value at the plane considered is captured with significantly better accuracy. We have tried to estimate higher modes using data from the same sensor points, but the reconstruction becomes more erratic and the fit values drop. This is due to the fact that higher order modes have smaller energy content spread over a broader range of frequencies (Figure 7), thereby representing random, rather than organized, motions. The contribution of the white noise

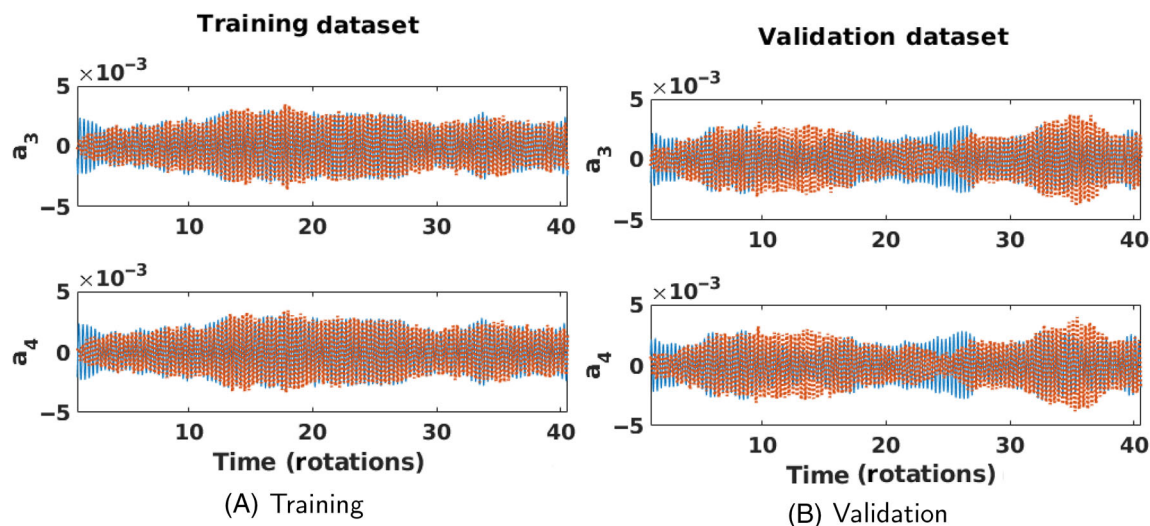


FIGURE 13 Comparison of temporal coefficients for the third and fourth proper orthogonal decomposition modes using an eighth-order model with six-point input: DNS (blue) and N4SID (orange)

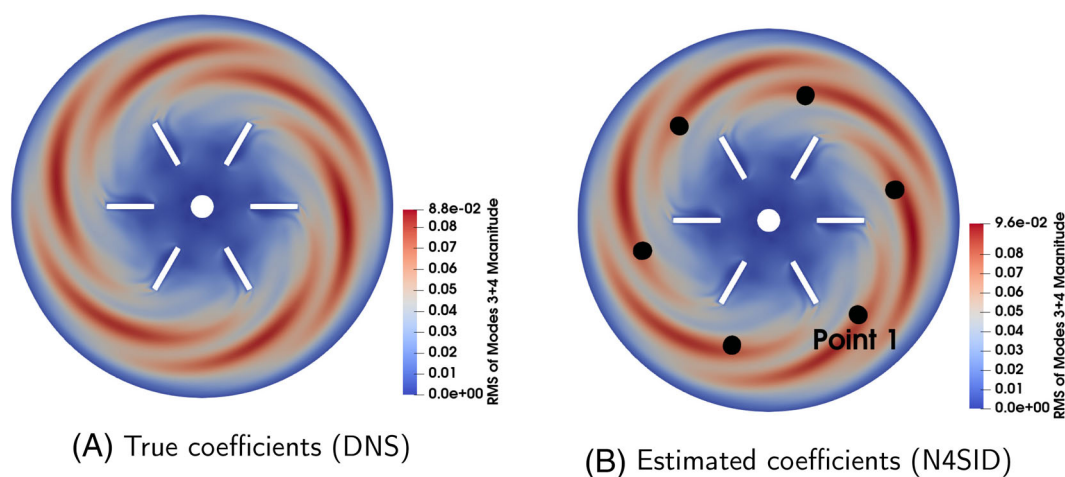


FIGURE 14 Contour plots of $\sqrt{u_{3&4}^2}$ computed using (A) the true and (B) the estimated temporal coefficients $a_3(t)$ and $a_4(t)$. The black dots indicate the sensor points used

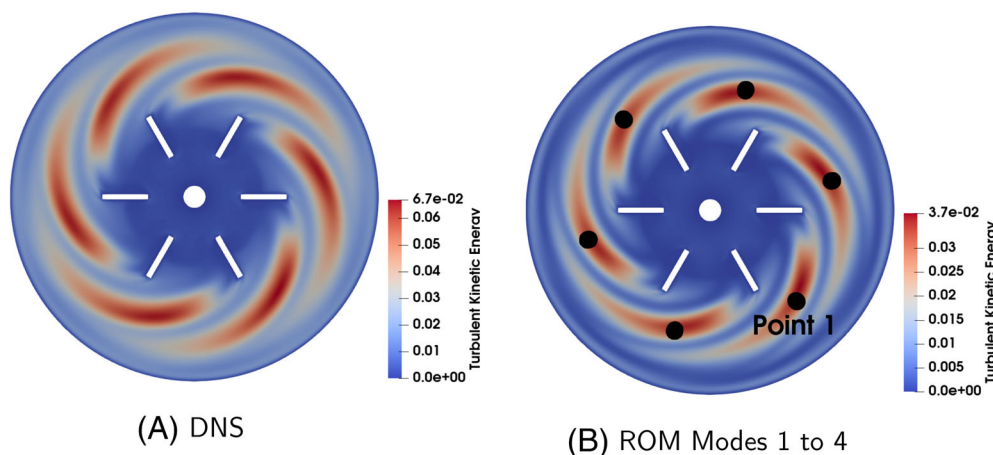


FIGURE 15 Contour plots of turbulent kinetic energy as (A) computed from the DNS simulation and (B) estimated by superposition of Modes 1–4 (for Modes 1 and 2, the input is only from sensor point 1, for Modes 3 and 4 from all six points)

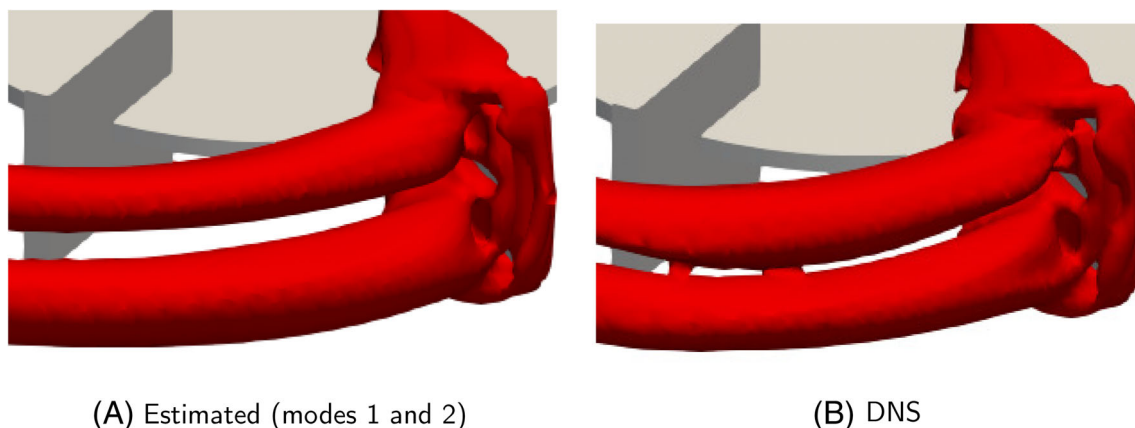
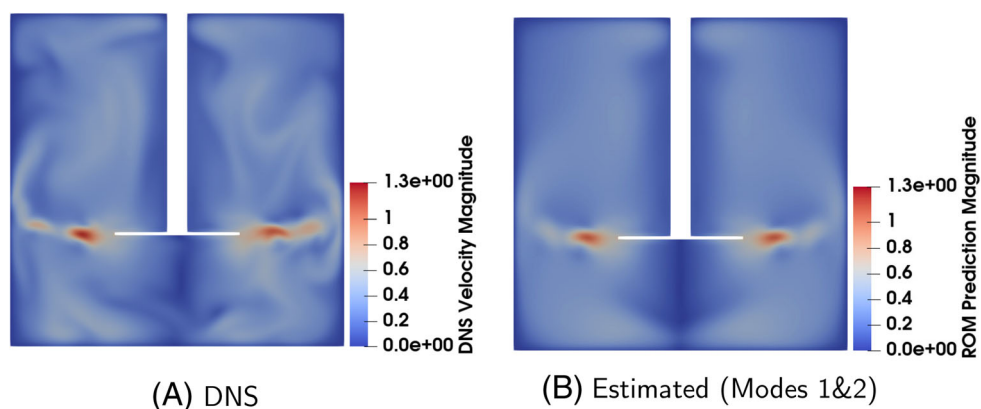


FIGURE 16 Isosurface of $\lambda_2/\omega^2 < -1.5$ in the vicinity of a blade at 20 rotations into validation dataset

FIGURE 17 True (DNS) and estimated velocity magnitudes at 20 rotations into validation dataset



terms $\mathbf{w}(t_i)$ and $\mathbf{v}(t_i)$ in model (13) becomes dominant, preventing meaningful estimation.

5.4 | Toward instantaneous 3D velocity field reconstruction

Combining the estimated coefficients $a_{1,e}(t)$ and $a_{2,e}(t)$ using one probe point, we now attempt to reconstruct the instantaneous 3D flow field in the whole vessel from $\mathbf{u}_e(\mathbf{x}, t) \approx \bar{\mathbf{u}}(\mathbf{x}) + a_{1,e}(t)\phi_1(\mathbf{x}) + a_{2,e}(t)\phi_2(\mathbf{x})$. Note that the time-average together with these two modes account for more than 90% of the total kinetic energy of the flow, as shown in Figure 4B.

We first consider the 3D flow close to the impeller and compute the λ_2 value for the instantaneous true and estimated flow fields at 20 rotations into the validation dataset. Recall that λ_2 is the second largest eigenvalue of the $\mathbf{S}^2 + \mathbf{\Omega}^2$, where $\mathbf{S}$ and $\mathbf{\Omega}$ are the strain and rotation rate tensors, respectively.⁴³ Isosurface plots of negative λ_2 (that visualize vortex structures) are shown in Figure 16. The flow close to the impeller is dominated by the large vortex structures captured by Modes 1 and 2 and, since the temporal coefficients are captured well (Figure 11A), the instantaneous flow is also rather well reconstructed. There are small-scale discrepancies close to the blade surface, but these are due to the higher modes.

Capturing the instantaneous flow away from the impeller is much more challenging. Contours of the true and estimated velocity magnitude on a vertical slice midway between the impeller blades at the same time instant are shown in Figure 17. Although at the height of the impeller the reconstruction is reasonably good, away from it the flow contains detailed features that are not reproduced. This makes physical sense. Recall that the Modes 1 and 2 are predominantly present around the impeller and have also a smooth but weak footprint in the rest of the vessel (Figure 6A,B). For this reason, the estimated velocity field in Figure 17B looks smooth away from the impeller, while the real one contains random, small scale, features. Such features in the bulk of the vessel cannot be captured by sensor points located close to the impeller because they have a narrow spatial extent (coherence).

5.5 | Performance at off-design conditions

The robustness and accuracy of reconstruction was also tested at off-design conditions, specifically at Reynolds numbers 500 and 700, obtained by adjusting the rotational velocity of the blades. The reconstruction process at the new conditions was the following: we performed DNS simulations for the aforementioned

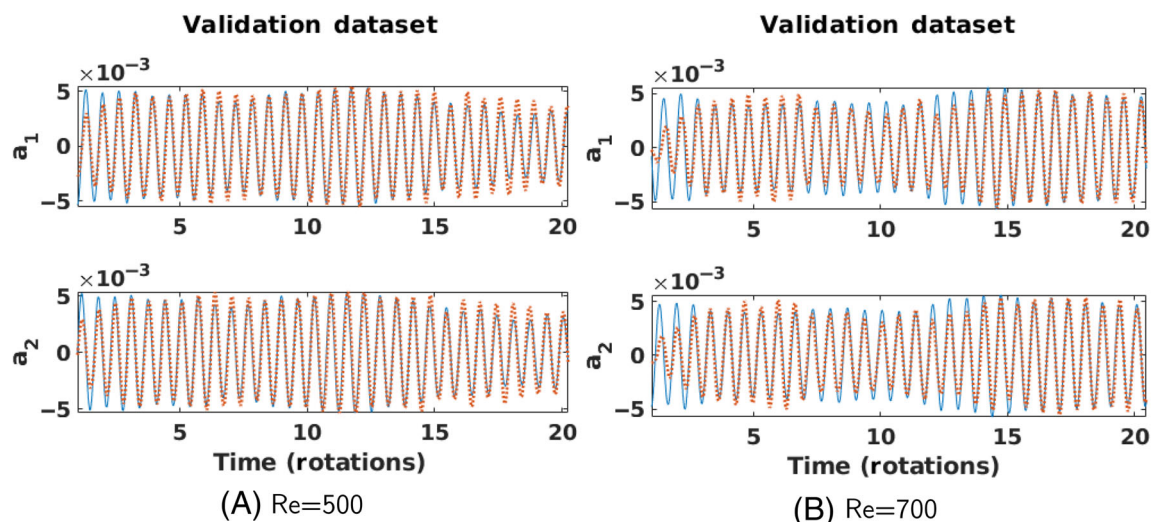


FIGURE 18 Temporal coefficients for Modes 1 and 2 at two off-design conditions using a single probe point and a sixth-order model; DNS (blue) and N4SID (orange). The temporal coefficients are normalized by the respective V_{tip} for each case

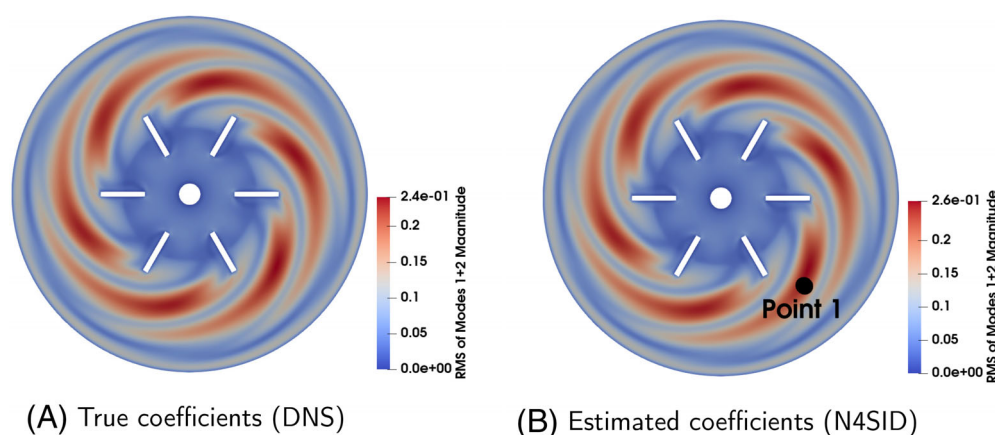


FIGURE 19 Contour plots of $\sqrt{u'_{1&2}{}^2}$ computed using (A) the true and (B) the estimated temporal coefficients $a_1(t)$ and $a_2(t)$ for the off-design condition of $Re = 500$. The black dot indicates the single sensor point used

Reynolds numbers and recorded the values of velocity at the same monitoring points. We then computed the temporal coefficients $a_1(t)$ and $a_2(t)$ by projecting the fluctuating velocity data obtained by DNS to the POD modes computed at the reference conditions, that is, $Re = 600$. Finally, we employed the model described by (12), which was trained in the reference case, to reconstruct the temporal coefficients using the input data for the new Reynolds numbers. It is important to note that this approach has truly predictive power, because we have trained a model using reference data (at $Re = 600$) and then use it to estimate the flow at off-design conditions ($Re = 500$ and 700). The only input required is information from sensor points at the new operating conditions.

The results using a single probe point and a sixth-order model are shown in Figure 18. The fits for modes (1 and 2) are (74.67%, 74.49%) for $Re = 500$ and (76.71%, 76.86%) for $Re = 700$. These values are marginally lower than those in the reference case of $Re = 600$. The

good quality of fit suggests that the spatial distribution of the leading pair of POD modes for $Re = 500$ and 700 is similar to that for $Re = 600$. As for the $Re = 600$ case, increasing the number of probe points to 6 leads to a third-order model with improved fits, equal to (82.77%, 82.5%) and (83.74, 83.21%) for Modes 1 and 2 for $Re = 500$ and 700 , respectively.

Contour plots of the velocity RMS due to Modes 1 and 2, $\sqrt{u'_{1&2}{}^2}$, computed using the true and the estimated temporal coefficients $a_1(t)$ and $a_2(t)$ for the off-design condition of $Re = 500$ are shown in Figure 19. Note the weak effect of the Reynolds number; the maximum RMS drops from 0.27 to 0.24, as can be seen by comparing Figures 12A and 19A. It is also clear that the employed algorithm has good predictive capability. The true and estimated peak RMS values are very close (the error is less than 10%), and this has been achieved using just one sensor point at the new Re number. Similar results are obtained for the other off-design value $Re = 700$ (not shown for brevity).

6 | CONCLUSIONS

In the present study, the POD technique was applied to reconstruct from limited data the large-scale vortical structures behind the blades of a Rushton impeller rotating inside an unbaffled stirred tank at the transitional Reynolds number of 600. About 34% of the turbulent kinetic energy was accounted for by two pairs of POD modes, with sharp spectral peaks at $1.5F_N$ and $3.0F_N$, respectively. Higher order modes were found to contain less than 1% of the total energy and have broader spectra. The $-11/9$ power law, corresponding to homogeneous isotropic turbulence, was found to hold for modes with orders greater than ≈ 120 .

The N4SID identification algorithm was then used to construct a data-driven estimator, that is, a linear, time-invariant system that reproduces the observed input-output relation between the velocity time signals at discrete sensor points and the temporal coefficients of each mode pair. Data from 1 to 6 sensor points were used. The points were placed at the location of maximum turbulent kinetic energy at the wake of each blade. For the first pair of modes, the quality of reconstruction was very good, even with a single probe point, but the order of the estimator was high, between 6 and 8. Increasing the number of probe points to 6 increased the fit and resulted in lower order and more robust estimators. Using a single input point to predict the second pair of POD modes was not successful, but a six-point input predicted well the frequency of the mode and yielded a satisfactory reproduction of the time trace of the temporal coefficients. The prediction of the RMS velocity for Modes 1 and 2 was excellent, while for Modes 3 and 4 very good (less than 10% error). The reconstruction of the entire 3D velocity field using the leading order modes was found to be reasonably good at the impeller height. However, further away from the impeller, where the footprint of the leading order modes was weaker, the detailed flow features were not reproduced.

The robustness of the estimator at off-design conditions, $Re = 500$ and 700 , was then examined. The estimator constructed for the nominal operating condition, $Re = 600$, was employed to reconstruct the temporal coefficients at the new conditions. The time signals of the Modes 1 and 2 were again very well reproduced, with small performance degradation, even using data from a single probe point.

We have demonstrated that system identification is capable of reconstructing the instantaneous evolution of 3D large-scale structures inside a stirred vessel using information from very few data points. Many questions remain to be answered. What is the performance of the algorithm at higher Reynolds numbers? How can we represent stochastically the truncated modes? What is the quality of reconstruction of other fields, such as Reynolds stresses? How accurate is the estimation of scalar mixing time using the reconstructed velocity field? These are topics of future research.

ACKNOWLEDGMENTS

The authors are grateful for computational resources made available through the UK Turbulence Consortium (EP/R029326/1), the UK Consortium for Turbulent Reacting Flows (EP/R029369/1), and the

UK Materials and Molecular Modelling Hub (EP/P020194/1). The first author, KM, would like to acknowledge funding from the EPSRC CDT for Fluid Dynamics Across Scales (EP/L016230/1). The authors would like to thank Profs. G. Micale and M. Ciofalo from the University of Palermo, Italy, for their responses to queries regarding Reference 20.

AUTHOR CONTRIBUTIONS

George Papadakis: Conceptualization; investigation; methodology; supervision; writing-review & editing. **Kirill Mikhaylov:** Investigation; software; validation; writing-original draft. **Stylianios Rigopoulos:** Supervision; writing-review & editing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Kirill Mikhaylov  <https://orcid.org/0000-0002-8451-9285>

Stelios Rigopoulos  <https://orcid.org/0000-0002-0311-2070>

George Papadakis  <https://orcid.org/0000-0003-0594-3107>

REFERENCES

1. Liné A, Gabelle J-C, Mochain J, Anne-Archard D, Augier F. On POD analysis of PIV measurements applied to mixing in a stirred vessel with a shear thinning fluid. *Chem Eng Res Design*. 2013;91(11):2073-2083.
2. de Lamotte A, Delafosse A, Calvo S, Toye D. Analysis of PIV measurements using modal decomposition techniques, POD and DMD, to study flow structures and their dynamics within a stirred-tank reactor. *Chem Eng Sci*. 2018;178:348-366.
3. de Lamotte A, Delafosse A, Calvo S, Toye D. Identifying dominant spatial and time characteristics of flow dynamics within free-surface baffled stirred-tanks from CFD simulations. *Chem Eng Sci*. 2018;192:128-142.
4. Gabelle J-C, Mochain J, Liné A. Kinetic energy transfer between first proper orthogonal decomposition modes in a mixing tank. *Chem Eng Technol*. 2017;40(5):927-937.
5. Ducci A, Doulgerakis Z, Yianneskis M. Decomposition of flow structures in stirred reactors and implications for mixing enhancement. *Ind Eng Chem Res*. 2008;47(10):3664-3676.
6. Doulgerakis Z, Yianneskis M, Ducci A. On the interaction of trailing and macro-instability vortices in a stirred vessel-enhanced energy levels and improved mixing potential. *Chem Eng Res Design*. 2009;87(4):412-420.
7. Gabelle J-C, Mochain J, Anne-Archard D, Augier F, Liné A. Experimental determination of the shear rate in a stirred tank with a non-Newtonian fluid: Carbopol. *AIChE J*. 2013;59(6):2251-2266.
8. Moreau J, Liné A. Proper orthogonal decomposition for the study of hydrodynamics in a mixing tank. *AIChE J*. 2006;52(7):2651-2655.
9. Riet KV, Smith JM. The trailing vortex system produced by Rushton turbine agitators. *Chem Eng Sci*. 1975;30(9):1093-1105.
10. Escudé R, Bouyer D, Liné A. Characterization of trailing vortices generated by a Rushton turbine. *AIChE J*. 2004;50(1):75-86.
11. Escudé R, Liné A. A simplified procedure to identify trailing vortices generated by a Rushton turbine. *AIChE J*. 2007;53(2):523-526.
12. Hasal P, Fort I, Kratena J. Force effects of the macro-instability of flow pattern on radial baffles in a stirred vessel with pitched-blade and Rushton turbine impellers. *Chem Eng Res Design*. 2004;82(9):1268-1281.

13. Nikiforaki L, Montante G, Lee K, Yianneskis M. On the origin, frequency and magnitude of macro-instabilities of the flows in stirred vessels. *Chem Eng Sci*. 2003;58(13):2937-2949.
14. Janiga G. Large-eddy simulation and 3D proper orthogonal decomposition of the hydrodynamics in a stirred tank. *Chem Eng Sci*. 2019;201:132-144.
15. Van Overschee P, De Moor B. N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*. 1994;30(1):75-93.
16. Van Overschee P, De Moor BL. *Subspace Identification for Linear Systems: Theory-Implementation-Applications*. Boston, MA, USA; London, UK; Dordrecht, The Netherlands: Kluwer Academic Publishers; 1996.
17. Qin S. An overview of subspace identification. *Comp Chem Eng*. 2006;30(10):1502-1513.
18. Guzmán-Iñigo J, Sipp D, Schmid PJ. A dynamic observer to capture and control perturbation energy in noise amplifiers. *J Fluid Mech*. 2014;758:728-753.
19. Guzmán-Iñigo J, Sodar MA, Papadakis G. Data-based, reduced-order, dynamic estimator for reconstruction of nonlinear flows exhibiting limit-cycle oscillations. *Phys. Rev. Fluids*. 2019;4:114703.
20. Tamburini A, Gagliano G, Micale G, Brucato A, Scargiali F, Ciofalo M. Direct numerical simulations of creeping to early turbulent flow in unbaffled and baffled stirred tanks. *Chem Eng Sci*. 2018;192:161-175.
21. Scargiali F, Tamburini A, Caputo G, Micale G. On the assessment of power consumption and critical impeller speed in vortexing unbaffled stirred tanks. *Chem Eng Res Design*. 2017;123:99-110.
22. Joshi JB, Nere NK, Rane CV, et al. CFD simulation of stirred tanks: comparison of turbulence models. Part i: radial flow impellers. *Can J Chem Eng*. 2011;89(1):23-82.
23. Yeoh S, Papadakis G, Yianneskis M. Determination of mixing time and degree of homogeneity in stirred vessels with large Eddy simulation. *Chem Eng Sci*. 2005;60(8-9):2293-2302.
24. Yeoh S, Papadakis G, Yianneskis M. Numerical simulation of turbulent flow characteristics in a stirred vessel using the LES and RANS approaches with the sliding/deforming mesh methodology. *Chem Eng Res Design*. 2004;82(7):834-848.
25. Başbuğ S, Papadakis G, Vassilicos JC. Reduced mixing time in stirred vessels by means of irregular impellers. *Phys Rev Fluids*. 2018;3:084502.
26. Başbuğ S, Papadakis G, Vassilicos JC. Reduced power consumption in stirred vessels by means of fractal impellers. *AIChE J*. 2018;64(4):1485-1499.
27. Yao H, Alves-Portela F, Papadakis G. Evolution of conditionally averaged second-order structure functions in a transitional boundary layer. *Phys. Rev. Fluids*. 2020;5:093902.
28. Thomareis N, Papadakis G. Resolvent analysis of separated and attached flows around an airfoil at transitional Reynolds number. *Phys. Rev. Fluids*. 2018;3:073901.
29. S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, D. A. May, L. C. McInnes, R. T. Mills, T. Munson, K. Rupp, P. Sanan, B. F. Smith, S. Zampini, H. Zhang, and H. Zhang, "PETSc Users Manual," Tech. Rep. ANL-95/11—Revision 3.9, Argonne National Laboratory, 2018.
30. Falgout RD, Yang UM. Hypre: a library of high performance preconditioners. In: Soot PMA, Hoekstra AG, Tan CJK, Dongarra JJ, eds. *Computational Science—ICCS*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2002:632-641.
31. Pope SB. *Turbulent Flows*. Cambridge, United Kingdom: Cambridge University Press; 2000.
32. Gillissen JJ, Van den Akker HEA. Direct numerical simulation of the turbulent flow in a baffled tank driven by a Rushton turbine. *AIChE J*. 2012;58(12):3878-3890.
33. Holmes P, Lumley JL, Berkooz G, Rowley CW. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge Monographs on Mechanics*. 2nd ed. Cambridge, United Kingdom: Cambridge University Press; 2012.
34. Taira K, Brunton SL, Dawson ST, et al. Modal analysis of fluid flows: an overview. *AIAA J*. 2017;55:4013-4041.
35. Hernandez V, Roman JE, Vidal V. SLEPC: a scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans Math Softw*. 2005;31(3):351-362.
36. Hernández V, Román JE, Tomás A. A robust and efficient parallel SVD solver based on restarted Lanczos bidiagonalization. *Electron Trans Numer Anal*. 2008;31:68-85.
37. Kailath T, Hassibi B, Sayed AH. *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall International; 2000.
38. Van Overschee P, De Moor B. A unifying theorem for three subspace system identification algorithms. *Automatica*. 1995;31(12):1853-1864.
39. Ljung L. *System Identification Toolbox: User's Guide*. Natick, MA, USA: The MathWorks, Inc; 2020.
40. Knight B, Sirovich L. Kolmogorov inertial range for inhomogeneous turbulent flows. *Phys Rev Lett*. 1990;65(11):1356.
41. Baldyga J, Bourne J. Interactions between mixing on various scales in stirred tank reactors. *Chem Eng Sci*. 1992;47(8):1839-1848.
42. Hartmann H, Derksen J, Montavon C, Pearson J, Hamill I, van den Akker H. Assessment of large eddy and rans stirred tank simulations by means of Ida. *Chem Eng Sci*. 2004;59:2419-2432.
43. Jeong J, Hussain F. On the identification of a vortex. *J Fluid Mech*. 1995;285:69-94.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Mikhaylov K, Rigopoulos S, Papadakis G. Reconstruction of large-scale flow structures in a stirred tank from limited sensor data. *AIChE J*. 2021;67(10): e17348. <https://doi.org/10.1002/aic.17348>