

**Decomposition and Reconstruction of the
Flow Field in a Stirred Tank Using
Data-Driven Methods and Machine
Learning**

Kirill Mikhaylov

Department of Aeronautics,
Imperial College London

This thesis is submitted for the degree of Doctor of
Philosophy

December 2022

Abstract

The aim of this thesis is to use data-driven and machine learning methods to identify, extract and reconstruct the large scale flow structures in an unbaffled stirred tank. Large scale flow structures have been observed behind the impeller blades and throughout the stirred tank. The origins and characteristics of these are not fully understood, but they have been found to play an important role in the rate of mixing and on products from chemical reactions in stirred tanks.

The computational cost of performing simulations with fine resolution to capture these is often prohibitive. On the other hand, the lack of spatial resolution from experimental measurements means that the latter can also only provide limited information. Data-driven and machine learning methods can be used to combine the strengths of the computational and experimental approaches with the aid of reduced order models.

The derived models can be used at both design and off-design conditions to provide quick estimates of the 3D flow field that can be used both in combination with chemical reaction simulation and for optimisation of mixing. For example, the approaches developed in the present thesis open the possibility to synchronise the reactant injection timing with local instantaneous flow features for more precise control of mixing. However, prior to this work, these models were largely limited to 2D flow problems.

An initial investigation was performed on a stirred tank in transitional flow conditions with $Re = 600$ simulated using Direct Numerical Simulations (DNS). Proper Orthogonal Decomposition (POD) was used to decompose the flow into modes. The first 4 POD modes were found to come in two pairs, accounting for 14%, and 2.6% per mode of the mean Turbulent Kinetic Energy (TKE) respectively. The first pair of modes was found to have a frequency of $1.5F_N$ and the second $3.0F_N$ (in the rotating reference frame) where F_N is the impeller rotation frequency. These modes were then reconstructed using the velocity input data from 1 and 6 input points using a system identification algorithm. Good fit values were obtained for both datasets for the first pair of modes for both the base design conditions and the off-design condi-

tions of $Re = 500$ and $Re = 700$. The second pair of modes were also reconstructed reasonably well with 6 input points.

Replacing the velocity input data with pressure data taken at the impeller and outer wall surfaces also allowed good prediction of the temporal evolution of the first two modes. Using 1 or 6 input points on the impeller surface, or 12 points (6 on the impeller surface and 6 on the outer wall) the temporal POD coefficients were reconstructed reasonably well for the base case. The successful application of the estimator to the off-design conditions for the 6 and 12 input point cases showed the robustness of the proposed approach to varying flow conditions.

The velocity POD modes were also used to decompose the pressure field into a time averaged component, and components that are linear and quadratic in terms of the temporal coefficient of the velocity modes. The spectrum of the power number of a single blade was found to have peaks at $1.5F_N$ and $3.0F_N$, corresponding to the first two POD mode pairs. However, when considering the power number computed from 3 adjacent blades or the total power number, these peaks vanished, suggesting that these modes have no effect on the total power number fluctuations. The power number spectrum was noisy but had noticeable peaks at $0.04F_N$ and $0.43F_N$. The former corresponded to the frequency of the 9th POD mode that represented fluctuations in the radial jet emanating from the impeller. The latter peak was not concentrated in a particular mode but was reproduced in the power number computed from the first 300 linear pressure modes.

The impeller speed was then increased to consider fully turbulent fluid flow conditions with a base case $Re = 30000$ and off-design conditions of $Re = 20000$ and $Re = 40000$ modelled using Large Eddy Simulations (LES). The first four modes were again found to come in two pairs with 5.5% and 1% per mode of the TKE respectively and frequencies of $1.4F_N$ and $0.8F_N$ in the rotating reference frame. In the inertial frame these were found to correspond to frequencies of $0.6F_N$ and $0.2F_N$ respectively. Investigation of their spatial components suggests that these modes are precessing Macro-Instabilities (MIs). In total, 4 algorithms (N4SID, N4SID with a preprocessing POD step on the probe data, and a 1 and 2 hidden layer LSTM Neural Network) were used to reconstruct the first pair of POD modes using velocity

data from 1, 2 and 6 input points. The modes were reconstructed well for both the design and off-design conditions by all algorithms with similar fit values. The reconstruction of the second pair of modes using a second set of 6 input points was less successful, with only the N4SID with the preprocessing algorithm giving qualitatively reasonable results.

Author's Declaration

I hereby declare that this thesis is based on my research work only. No parts of this thesis have been submitted for another degree. All references and other sources of information have been referenced and acknowledged.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution Non-Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

I would like to thank my supervisors Prof. George Papadakis and Dr Stelios Rigopoulos for their guidance and supervision provided over the course of this project.

I would also like to acknowledge funding from the EPSRC CDT for Fluid Dynamics Across Scales (EP/L016230/1) and the computational resources made available through the UK Turbulence Consortium (EP/R029326/1) and the UK Materials and Molecular Modelling Hub (EP/P020194/1).

Finally, I would like to thank my family, friends and everyone at ICXC, ICSWP and ICTri who has supported me over the past few years and made my time at Imperial a great one!

Dedication

In loving memory of my grandparents.

Contents

Abstract	1
Author's Declaration	4
Copyright Declaration	5
Acknowledgements	6
1 Introduction	34
1.1 Motivation	34
1.2 Literature Review	35
1.2.1 Flows in Stirred Tanks	35
1.2.2 Data-Driven and Machine Learning Methods	41
1.3 Objectives	48
1.4 Thesis Outline	49
2 CFD Methodology and Validation	50
2.1 Introduction	50

2.2	DNS Simulations	51
2.3	LES Simulations	54
3	Data-Driven and Machine learning Methods	59
3.1	Introduction	59
3.2	Proper Orthogonal Decomposition	59
3.3	Flow Field Reconstruction	63
3.3.1	Linear Estimator Using System Identification	63
3.3.2	Point-POD+N4SID	67
3.3.3	LSTM Neural Network	68
3.3.4	Fit Quality	72
3.3.5	Velocity Field Reconstruction	72
4	Reconstruction of Large Scale Flow Structures in a Stirred Tank from Sparse Sensor Data	73
4.1	Introduction	73
4.2	Characterisation of POD Modes	74
4.3	Reconstruction of the 1st and 2nd Modes from Point Data	78
4.3.1	Reconstruction Using the Velocity Signals from a Single Sen- sor Point	79
4.3.2	Effect of the Number of Sensor Points on the Quality of Re- construction	83
4.4	Reconstruction of the 3rd and 4th Modes	84

4.5	Towards Instantaneous 3D Velocity Field Reconstruction	88
4.6	Performance at Off-Design Conditions	89
4.7	Conclusions	91
5	Decomposition of Power Number and Reconstruction of Flow Structures from Surface Pressure Measurements	93
5.1	Introduction	93
5.2	Methodology	94
5.2.1	Governing Equations for the Pressure Field	94
5.3	Proper Orthogonal Decomposition	97
5.3.1	Velocity POD Modes	97
5.3.2	Governing Equations for the Partial Pressures	98
5.3.3	Boundary Conditions	100
5.4	Construction of a Data-Driven Linear Flow Estimator from Pressure Measurements	101
5.5	Power Number Decomposition	102
5.6	Characterisation of Partial Pressures and Power Number Decomposition	105
5.7	Analysis of Power Number Spectrum Using Partial Pressure Modes .	108
5.8	Reconstruction of Large Scale Vortical Structures from Surface Pressure Measurements	114
5.8.1	1 Point on the Blade Surface at Maximum Pressure RMS . . .	117
5.8.2	6 Points on Blade Surface Based on the Maximum of the Pressure RMS	120

5.8.3	12 Points based on the Maximum of the Pressure RMS on the Blades and on the Outer Wall	120
5.9	Reconstruction of Instantaneous Pressure	123
5.10	Conclusion	124
6	Identification and Reconstruction of Macro-Instabilities in a Turbu- lent Stirred Tank Using Data-Driven and Machine Learning Meth- ods	127
6.1	Introduction	127
6.2	Results and Discussion	128
6.2.1	Mean Flow	128
6.2.2	POD Mode Characterisation	129
6.2.3	Reconstruction of POD Modes 1 and 2	140
6.2.4	Reconstruction of Modes 3 and 4	148
6.3	Conclusions	150
7	Conclusions and Future Work	153
7.1	Summary of Main Achievements	153
7.1.1	Reconstruction of Large Scale Flow Structures in a Stirred Tank from Sparse Sensor Data	153
7.1.2	Decomposition of the Pressure Field and Power Number, and Reconstruction of Large Scale Flow Structures in a Stirred Tank from Blade Surface Pressure Measurements	154

7.1.3	Identification and Reconstruction of MIs in a Turbulent Stirred Tank Using Data-Driven and Machine Learning Methods . . .	156
7.2	Future Work	157
Bibliography		159

List of Tables

2.1	Normalised Dimensions (T) of the stirred tank considered.	51
6.1	Training and Validation Fits for the Temporal Coefficients of the First 2 POD Modes.	143
6.2	Fits for Off-design Conditions for the Temporal Coefficients of the First 2 POD Modes.	143

List of Figures

1.1	Measurements of the power number for an unbaffled stirred tank [1] and a baffled stirred tank [2]. The approximate transition points from laminar to transitional flow at $Re = 100$ and from transitional to turbulent flow at $Re = 10000$ are denoted by the vertical lines. . .	37
1.2	Schematics of a Trailing Vortex based upon [3] and precessing Macro Instabilities based upon [4]	39
1.2a	Trailing Vortex.	39
1.2b	Macro-Instabilities.	39
2.1	Sketch of the stirred tank and the impeller geometry. The axes of the Cartesian coordinate system are shown in blue.	51
2.1a	Side view of the tank.	51
2.1b	Schematic of Rushton turbine.	51
2.2	Time-averaged velocities along a line bisecting two blades at a distance of $0.353T$ from the tank bottom. Dashed line (- -) high resolution grid. Solid line (—) low resolution grid.	53
2.2a	Radial Velocity.	53
2.2b	Tangential velocity.	53

2.2c	Axial velocity.	53
2.3	Turbulent kinetic energy and dissipation rate of kinetic energy along a line bisecting two blades at a distance of $0.353T$ from the tank bottom. Dashed line (- -) high resolution grid. Solid line (-) low resolution grid.	53
2.3a	TKE.	53
2.3b	Dissipation.	53
2.4	Time and tangentially averaged velocity at $z/T = 0.211$ and trailing vortex path comparison for different meshes against data from other studies.	57
2.4a	Tangential velocity.	57
2.4b	Radial velocity	57
2.4c	Axial velocity.	57
2.4d	Trailing vortex path.	57
3.1	Procedural steps for system identification.	64
3.1a	Step 1: Data generation.	64
3.1b	Step 2: Extraction of system matrices using a training dataset.	64
3.1c	Step 3: Model assessment using a validation dataset.	64
3.2	Schematic of the Point-POD+N4SID.	68
3.3	Schematic of a LSTM with two hidden layers.	70
3.4	Schematic of a LSTM Cell.	70
4.1	Energy content of POD modes.	75

4.1a	Percentage of fluctuation energy for modes 1 to 800.	75
4.1b	Cumulative percentage of total kinetic energy contained in the mean and first 100 modes.	75
4.2	Contours of the velocity magnitudes of the first 4 POD modes on a z plane just above the impeller disk at the height of the maximum RMS of velocity fluctuations (0.353T 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 re- spectively.	75
4.2a	Mode 1	75
4.2b	Mode 2	75
4.2c	Mode 3	75
4.2d	Mode 4	75
4.3	Contours of the velocity magnitudes of the first 4 POD 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.	76
4.3a	Mode 1	76
4.3b	Mode 2	76
4.3c	Mode 3	76
4.3d	Mode 4	76
4.4	Power spectral density of POD modes 1, 3, 5, 61 and 252.	77
4.5	Probability density function plots (blue) and fitted Gaussian distri- butions (orange) of POD modes 1, 3, 5, 61 and 252.	78
4.6	Plots of a_1 against a_2 ; a_3 against a_4 ; and a_1 against a_3	78

4.6a	a_1 against a_2	78
4.6b	a_3 against a_4	78
4.6c	a_1 against a_3	78
4.7	Velocity fluctuations at the point of maximum turbulent kinetic energy.	79
4.8	Power spectrum of the x velocity component at probe point 1 for Re=600.	79
4.9	System identification using velocity data from probe point 1 (6th order ROM): DNS (blue) and N4SID (orange).	80
4.9a	Training	80
4.9b	Validation	80
4.10	Effect of model order on the accuracy of reconstruction for the train- ing and validation datasets using 1 (a) and 6 (b) probe points.	82
4.10a	1 probe point	82
4.10b	6 probe points	82
4.11	Contour plots of $\sqrt{\mathbf{u}'_{1&2}{}^2}$ computed using (a) the true and (b) the es- timated temporal coefficients $a_1(t)$ and $a_2(t)$. The black dot indicates the single sensor used.	82
4.11a	True coefficients (DNS)	82
4.11b	Estimated coefficients (N4SID)	82
4.12	Comparison of temporal coefficients for the 3rd and 4th POD modes using an 8th order model with 6 point input: DNS (blue) and N4SID (orange).	85
4.12a	Training	85

4.12b	Validation	85
4.13	Contour plots of $\sqrt{\mathbf{u}'_{3\&4}{}^2}$ 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.	86
4.13a	True coefficients (DNS)	86
4.13b	Estimated coefficients (N4SID)	86
4.14	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&2 the input is only from sensor point 1, for modes 3&4 from all 6 points).	86
4.14a	DNS	86
4.14b	ROM Modes 1 to 4	86
4.15	Turbulent kinetic energy obtained from the DNS simulations and the ROM along a line bisecting two blades at a distance of $0.353T$ from the tank bottom.	87
4.16	Isosurface of $\lambda_2/\omega^2 < -1.5$ in the vicinity of a blade at 20 rotations into validation dataset.	87
4.16a	Estimated (modes 1 and 2)	87
4.16b	DNS	87
4.17	True (DNS) and estimated velocity magnitudes at 20 rotations into validation dataset.	88
4.17a	DNS	88
4.17b	Estimated (Modes 1&2)	88

4.18	Temporal coefficients for modes 1 and 2 at two off-design conditions using a single probe point and a 6th order model; DNS (blue) and N4SID (orange). The temporal coefficients are normalised by the respective V_{tip} for each case.	90
4.18a	Re=500	90
4.18b	Re=700	90
4.19	Contour plots of $\sqrt{\mathbf{u}'_{1\&2}{}^2}$ 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.	90
4.19a	True coefficients (DNS)	90
4.19b	Estimated coefficients (N4SID)	90
5.1	Contours of instantaneous pressure obtained directly from DNS and from solving the Poisson equation on a vertical plane midway between two blades.	97
5.1a	Poisson Equation	97
5.1b	DNS	97
5.2	Power Number for Re=600	103
5.2a	Time Signal	103
5.2b	Spectrum of Power Number	103
5.3	Power number signal for individual blades and blade combinations. .	103
5.4	Spectrum of power number for individual blades and blade combinations.	105

5.5	Power number due to diagonal partial pressures.	106
5.5a	Power number.	106
5.5b	Cumulative power number.	106
5.6	Contour plots of mean and diagonal partial pressure terms on a ver- tical section bisecting the blades.	107
5.6a	Mean P	107
5.6b	P_{00}	107
5.6c	$\lambda_1 P_{11}$	107
5.6d	$\lambda_2 P_{22}$	107
5.6e	$\lambda_3 P_{33}$	107
5.6f	$\lambda_4 P_{44}$	107
5.7	Contour plots of the mean and diagonal partial pressure terms on horizontal section 0.353T from the tank bottom.	108
5.7a	Mean P	108
5.7b	P_{00}	108
5.7c	$\lambda_1 P_{11}$	108
5.7d	$\lambda_2 P_{22}$	108
5.7e	$\lambda_3 P_{33}$	108
5.7f	$\lambda_4 P_{44}$	108
5.8	Contour plots of the linear partial pressure terms on vertical section bisecting the blades.	109
5.8a	$\sqrt{\lambda_1} P_{01}$	109

5.8b	$\sqrt{\lambda_1}P_{02}$.	109
5.8c	$\sqrt{\lambda_1}P_{03}$.	109
5.8d	$\sqrt{\lambda_1}P_{04}$.	109
5.9	Contour plots of the linear partial pressure terms on horizontal section 0.353T from the tank bottom.	110
5.9a	$\sqrt{\lambda_1}P_{01}$.	110
5.9b	$\sqrt{\lambda_2}P_{02}$.	110
5.9c	$\sqrt{\lambda_3}P_{03}$.	110
5.9d	$\sqrt{\lambda_4}P_{04}$.	110
5.10	$a_1(P_{01} + P_{10}) + a_2(P_{02} + P_{20})$	110
5.11	Partial Power Number	111
5.11a	$a_1(P_{01} + P_{10}) + a_2(P_{02} + P_{20})$	111
5.12	Surface Plots of Partial Pressures on Blades.	111
5.12a	$\sqrt{\lambda_1}P_{01}$	111
5.12b	$\sqrt{\lambda_2}P_{02}$	111
5.12c	$\sqrt{\lambda_3}P_{03}$	111
5.12d	$\sqrt{\lambda_4}P_{04}$	111
5.13	Contour plots of velocity magnitude and partial pressure correspond- ing to Mode 9.	112
5.13a	ϕ_9	112
5.13b	ϕ_9	112
5.13c	$\sqrt{\lambda_9}P_{09}$	112

5.13d	$\sqrt{\lambda_9}P_{09}$	112
5.13e	$\sqrt{\lambda_9}P_{09}$	112
5.14	Power spectra of contributions to the power number by linear pressure modes.	113
5.14a	$a_9(P_{09} + P_{90})$	113
5.14b	$\sum_{i=1}^{i=300} a_i(P_{i0} + P_{0i})$	113
5.14c	$\sum_i a_i(P_{i0} + P_{0i})$ where i corresponds to the 30 modes with the largest contributions from the first 300 modes.	113
5.15	Contours of $\sqrt{a_1^2 a_2^2} P_{12}$	116
5.15a	Vertical section bisecting two blades.	116
5.15b	Horizontal section 0.353T from the tank bottom.	116
5.15c	Impeller surface.	116
5.16	Mean and RMS Pressure on impeller surface. Location of max RMS Pressure and max RMS Pressure on the outer tank wall shown in green and magenta, at coordinates (r, θ, z) of $(0.125T, 57^\circ, 0.304T)$ and $(0.500T, 74^\circ, 0.372T)$ respectively.	116
5.16a	Mean Pressure	116
5.16b	RMS Pressure	116
5.17	Pressure at max PRMS on the impeller.	117
5.17a	Time Series	117
5.17b	Power Spectrum averaged over 19 windows of approximately 20 rotations	117

5.18	Estimated (orange) and true (blue) coefficients for modes 1 and 2 using 1 probe point based on max PRMS (5th order model).	118
5.18a	$Re = 600$ Training	118
5.18b	$Re = 600$ Validation	118
5.19	Estimated (orange) and true (blue) coefficients for modes 1 and 2 using 6 probe points based on max PRMS (6th order model).	119
5.19a	$Re = 600$ Training	119
5.19b	$Re = 600$ Validation	119
5.19c	$Re = 500$	119
5.19d	$Re = 700$	119
5.20	Cross correlation coefficients as a function of lag.	123
5.20a	Impeller probe point pressure to a_1	123
5.20b	Wall probe point pressure to a_1	123
5.20c	Impeller probe point pressure to wall probe point pressure. . .	123
5.21	Pressure after 40 rotations in validation dataset using 6 points at Pressure RMS maximum reconstruction.	124
5.21a	DNS	124
5.21b	ROM	124
6.1	Contours of mean velocity magnitude and projection of the mean velocity vectors onto a vertical plane bisecting two blades.	129
6.2	POD mode eigenvalues.	130
6.2a	Eigenvalues 1 to 9999.	130

6.2b	Cumulative eigenvalue energy.	130
6.3	Power spectra of u velocity at the locations of the maxima of POD modes 1 and 3.	131
6.3a	Mode 1.	131
6.3b	Mode 3.	131
6.4	Compensated power spectra of u velocity at the locations of the max- ima of POD modes 1 and 3.	132
6.4a	Mode 1.	132
6.4b	Mode 3.	132
6.5	Temporal coefficients of POD modes.	132
6.6	Spectra of temporal coefficients of POD modes.	133
6.7	Contours of POD velocity magnitudes at a plane 0.357T above tank bottom.	134
6.7a	Mode 1	134
6.7b	Mode 2	134
6.7c	Mode 3	134
6.7d	Mode 4	134
6.8	Contours of POD velocity magnitudes at a vertical plane bisecting blades.	135
6.8a	Mode 1	135
6.8b	Mode 2	135
6.8c	Mode 3	135

6.8d	Mode 4	135
6.9	Contours of axial vorticity superimposed on velocity vectors at a horizontal plane 0.11T above the bottom of the tank.	136
6.9a	Mode 1	136
6.9b	Mode 2	136
6.9c	Mode 3	136
6.9d	Mode 4	136
6.10	Contours of mode 1 velocity magnitudes on horizontal planes at various heights above the tank bottom.	137
6.10a	0.21T from tank bottom	137
6.10b	0.26T from tank bottom	137
6.10c	0.32T from tank bottom	137
6.10d	0.37T from tank bottom	137
6.10e	0.42T from tank bottom	137
6.10f	0.47T from tank bottom	137
6.11	Contours of mode 3 velocity magnitudes on horizontal planes at various heights above the tank bottom.	137
6.11a	0.21T from tank bottom	137
6.11b	0.26T from tank bottom	137
6.11c	0.32T from tank bottom	137
6.11d	0.37T from tank bottom	137
6.11e	0.42T from tank bottom	137

6.11f	0.47T from tank bottom	137
6.12	Power spectra of the velocity components corresponding to Modes 1+2 in an inertial frame at the height and radius of the max of POD mode 1.	138
6.12a	Tangential Velocity	138
6.12b	Radial Velocity	138
6.12c	Axial Velocity	138
6.13	Power spectra of the velocity components corresponding to Modes 3+4 in an inertial frame at the height and radius of the max of POD mode 3.	138
6.13a	Tangential Velocity	138
6.13b	Radial Velocity	138
6.13c	Axial Velocity	138
6.14	Locations of the maxima of the first 4 POD modes. Mode 1 (green point), Mode 2 (magenta point), Mode 3 (black point) and Mode 4 (blue point).	141
6.15	Average fit value dependence on the number of timesteps for a 2 hidden layer LSTM neural network.	142
6.16	N4SID with 1 point at max of first POD mode.	144
6.16a	$Re = 30000$ Training Dataset	144
6.16b	$Re = 30000$ Validation Dataset	144
6.16c	$Re = 20000$	144
6.16d	$Re = 40000$	144

6.17 LSTM 2 with 6 point input.	145
6.17a $Re = 30000$ Training Dataset	145
6.17b $Re = 30000$ Validation Dataset	145
6.17c $Re = 20000$	145
6.17d $Re = 40000$	145
6.18 Velocity magnitudes of LES and ROM with mean plus first 2 modes using 1 point input N4SID at 20 rotations into validation dataset on vertical plane between two blades.	147
6.18a LES	147
6.18b Mean + Modes 1+2	147
6.19 Velocity magnitudes of LES and ROM with mean plus first 2 modes using 1 point input N4SID at 20 rotations into validation dataset on a horizontal plane 0.357T from the tank bottom.	148
6.19a LES	148
6.19b Mean + Modes 1+2	148
6.20 N4SID+POD model with 6 inputs for modes 3 and 4.	149
6.20a $Re = 30000$ Training Dataset	149
6.20b $Re = 30000$ Validation Dataset	149
6.20c $Re = 20000$	149
6.20d $Re = 40000$	149
6.21 Sum of modes 1 to 4 magnitudes of LES and POD+N4SID ROM using 12 point input points at 20 rotations into validation dataset for modes 1+2 and 45 rotations into the validation dataset for modes 3+4.	150

6.21a	LES	150
6.21b	ROM	150

Nomenclature

Roman Symbols

$\mathbf{A}', \mathbf{B}' \mathbf{H}'$ System matrices for linear model

a_i Temporal coefficient of the i th POD mode

$a_{i,e}$ Estimated temporal coefficient of the i th POD mode from ROM

C Impeller bottom clearance

$\mathbf{C}_s$ Smagorinsky constant

D Impeller diameter

d Impeller disk width

F_N Impeller frequency

H Tank height

L Impeller blade length

$\bar{N}_p$ Time averaged power number

N Impeller speed

N_p Instantaneous power number

$\bar{p}$ Time averaged pressure

p Pressure

P_{ij}	Partial pressure for the i th and j th POD mode
Re	Reynolds number
S	Shaft diameter
$\mathbf{S}_{ij}$	Strain rate tensor
$\mathbf{s}(t_j)$	Output data at time t_j for linear model
T	Tank diameter
t	Impeller blade thickness
$\bar{\mathbf{u}}$	Time averaged Velocity vector
$\tilde{\mathbf{u}}$	Filtered velocity vector
$\mathbf{u}$	Velocity vector
$\mathbf{v}(t_j), \mathbf{w}(t_j)$	White noise terms at time t_j for linear model
v_{tip}	Blade tip velocity
W	Impeller blade width
$\mathbf{X}(t_j)$	Estimator state at time t_j for linear model

Greek symbols

Δ_m	Cell size in the m th direction
δ_{ij}	Kronecker delta
λ_i	Eigenvalue of the i th POD mode
μ	Dynamic viscosity
ν	Kinematic viscosity
ν_e	Eddy viscosity

ρ	Fluid density
τ	Sub-grid stress tensor
Φ_i	Spatial component of the i th POD mode
ω	Angular velocity vector

Acronyms

CFD	Computational Fluid Dynamics
DMD	Dynamic Mode Decomposition
DNS	Direct Numerical Simulation
EPOD	Extended Proper Orthogonal Decomposition
LES	Large Eddy Simulation
LSTM	Long-Short Term Memory
MI	Macro Instability
N4SID	Numerical algorithm for subspace state space system identification
NN	Neural Networks
PIV	Particle Image Velocimetry
POD	Proper Orthogonal Decomposition
PRMS	Pressure Root Mean Square
PSD	Power Spectral Density
RMS	Root Mean Square
ROM	Reduced Order Model
SSG	Speziale, Sarkar and Gatski

SST Shear Stress Transport

TKE Turbulent Kinetic Energy

Chapter 1

Introduction

1.1 Motivation

There are two key elements that motivate the present work. The first is the in depth investigation of the large scale periodic structures in the industrially relevant geometry of a stirred tank. Large scale flow structures can have a significant impact on the mixing performance and the quality of products and hence understanding their behaviour is important for better mixing control [5, 6]. While these have been analysed experimentally and computationally in 2D slices [4, 7–13], there have only been a few studies that have considered their shape and periodic behaviour in 3D [14–17] and none computationally in unbaffled stirred tanks agitated by a Rushton turbine.

While CFD approaches can provide 3D flow fields throughout the system considered, they come with a considerable computational cost. This limits their application in both industry and academia to short temporal timescales and/or low spatial resolutions that are only satisfactory to compute the time averaged flow field. On the other hand, experiments can be used to obtain data over a long time, but are limited in their spatial extent. Typically, data is collected at a few locations or on small 2D planes.

A method to combine the key strength of experiments - the quick data acquisition times, and the key strength of CFD - 3D flow fields throughout the tank domain would provide a valuable tool for both industrial and academic applications, in stirred tanks, and other fluid mechanical systems. Therefore, the second element of this work is to use data-driven and machine learning methods that have recently started to be used in fluid mechanics applications to identify and to reconstruct the behaviour of large scale structures from limited measurements. This would enable previously unfeasible approaches to flow and mixing control, such as the synchronisation of timing of reactant injection to the local behaviour of periodic structures in a flow, to be used.

The data-driven and machine learning methods used in the reconstruction of these large scale structures have been largely limited to simple 2D flow fields such as the flow behind a cylinder, or when a 3D problem is considered, only 2D slices of the velocity are reconstructed. Hence, this study also seeks to extend and demonstrate the capability of these methods in fully 3D complex flows at both transitional and fully turbulent Reynolds numbers.

1.2 Literature Review

1.2.1 Flows in Stirred Tanks

Stirred tanks are a common mixing device, used at multiple scales and applications in industry, from mixing together liquid-liquid mixtures [18, 19], to precipitation reactions [20, 21] and biochemical applications such as cell culture growth [22]. The most common configuration examined in academic studies is a cylindrical tank agitated by the Rushton Turbine, typically consisting of 6 blades equidistantly spaced around a central disk [19–21, 23], with a bottom clearance of $1/3$ to $1/2$ of the tank height. Also, 4 equispaced baffles on the outer wall are typically present. Rushton

turbines are characterised by a strong radial flow and the baffles break up the circumferential flow resulting in better mixing [23]. Other commonly examined impellers are pitched blade turbines with 3 to 8 blades [14, 23] and flat blade impellers that are similar to the Rushton Turbine but without a central disk. In addition, other impeller designs, such as fractal [24], porous [25], inclined [26] and sawtooth impellers [27] have also been examined. Multiple other impeller shapes and variations also exist [28, 29].

The impeller types can be grouped into two broad categories. Axial impellers such as pitched blade turbines are characterised by having a strong axial flow component. These are typically used for applications such as solids suspension, blending, gas inducement and heat transfer [19].

Radial impellers such as Rushton and flat blade turbines have a primarily radial flow profile. These are used for both single and multiple-phase mixing and are most effective for gas–liquid and liquid–liquid dispersion [19]. They have the advantage of providing higher shear and turbulence levels with lower pumping compared to axial flow impellers. The addition of the disk allows the Rushton impeller to have a more uniform radial flow profile and the ability to increase the number of blades more easily than a flat blade impeller [19].

While baffled tanks have been more commonly investigated in academic studies, unbaffled tanks have also been considered [15, 30–37]. Due to the ease of cleaning, unbaffled tanks are preferred in many industrial applications [38], at laminar and transitional flow regimes due to the baffles creating dead zones [28] and in bioreactors due to the baffles causing cell death [39, 40]. A common modification when using an unbaffled stirred tank is to place the impeller off-centre to break up the circumferential flow and to improve mixing [34–37].

The power number in a stirred tank is defined as $N_p = \frac{\overline{P_o}}{\rho N^3 D^5}$, where $\overline{P_o}$ is the time-averaged power input to the impeller or equivalently the total energy dissipation integrated over the tank volume. This quantifies the average energy dissipation rate

in the tank. Measurements of the power number for an unbaffled stirred tank [1] and a baffled stirred tank [2] are shown in Figure 1.1. For unbaffled tanks, N_p has been experimentally found to have values of ~ 70 at very low Reynolds numbers of ~ 1 and then decreases with increasing Reynolds numbers to ~ 2 for $Re \sim 1000$ and to 0.6 for $Re \sim 100000$ [1]. For baffled tanks, the power number behaves in the same way as for unbaffled tanks up to $Re \sim 100$, but then starts to increase with increasing Reynolds number, reaching an asymptote of $\sim 5 - 6$ for $Re > 10000$ [1]. The difference is due to the baffles that convert the tangential flow generated by the impeller into axial motion, increasing the power draw [28]. The power number can also be influenced by the impeller dimensions such as the blade thickness, impeller diameter and number of blades [1, 29, 41]. In particular, an increase in impeller diameter [1] or increase in impeller blade thickness [29, 41] has been observed to decrease the measured power number.

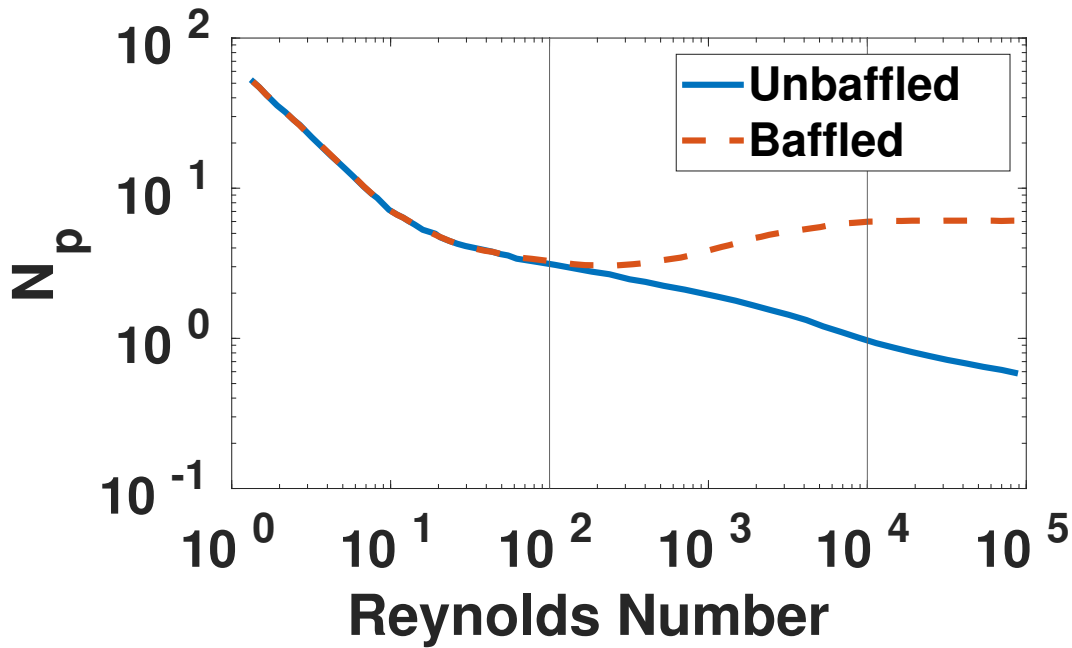


Figure 1.1: Measurements of the power number for an unbaffled stirred tank [1] and a baffled stirred tank [2]. The approximate transition points from laminar to transitional flow at $Re = 100$ and from transitional to turbulent flow at $Re = 10000$ are denoted by the vertical lines.

Trailing Vortices

Trailing vortices have long been observed in the wake of impeller blades in Rushton and other flat blade turbines. They consist of two vortices on the upper and lower parts of the suction side of the blade and contribute to a strong periodic component to the flow in the vicinity of the impeller [23]. These were first observed experimentally in the 1970s [42–45] and they have been further studied experimentally and computationally since [3, 31, 46, 47]. A schematic based upon [3] is shown in Figure 1.2a. These studies have mainly focused on identifying the mean path of the trailing vortices and the time averaged velocity and pressure profiles in the vortex cores [48, 49]. On the other hand, the periodic nature of the trailing vortices has received much less attention. In the inertial frame, the blade passage frequency has been found to dominate the velocity power spectra, but in a recent study [31] the trailing vortices behind a flat blade impeller with 1, 2 and 4 blades were investigated in the rotating reference frame. For the 4 blade impeller, frequencies of $1.58F_N$ and $3.16F_N$ were observed in the power spectra of the tangential and axial velocities at probe points in the impeller discharge region. The power number was also found to have fluctuations with a dominant frequency of $3.16F_N$ when all blades were considered and an additional frequency of $1.58F_N$ was observed when considering the power number for an individual blade.

Macro-Instabilities

Macro-instabilities are large scale periodic structures that have been observed in stirred tanks and can be classified into two categories [50]. The first is associated with the fluctuation of the jet generated by the motion of the impeller due to the pressure reflection off of the wall. The second is due to precessing vortices about the impeller shaft [4, 11]. A schematic based upon [4] is shown in Figure 1.2b. In baffled stirred tanks, agitated by a Rushton turbine, the first type has been found to have characteristic frequencies of the order of $\sim 0.1F_N$ [50], while the second type

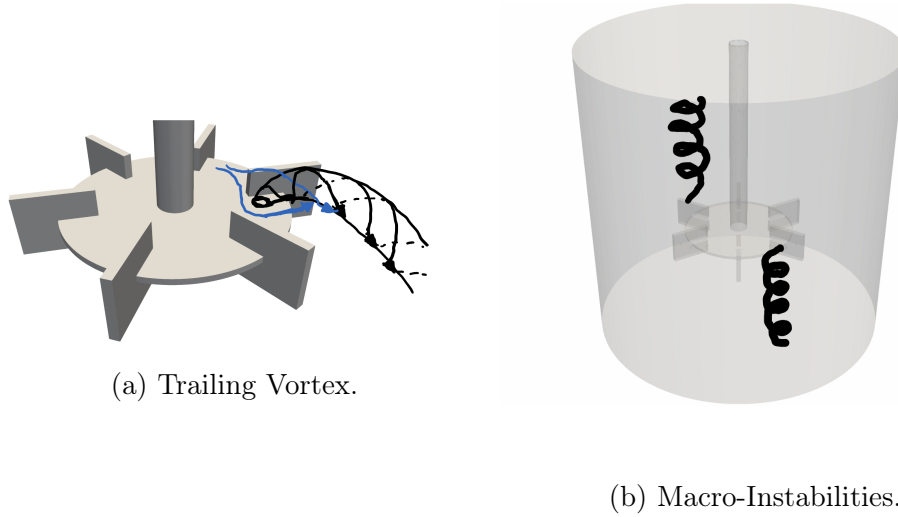


Figure 1.2: Schematics of a Trailing Vortex based upon [3] and precessing Macro Instabilities based upon [4]

has frequencies of an order of a magnitude smaller i.e. $\sim 0.01F_N$ [4, 11, 50].

While many studies have examined macro-instabilities for a Rushton turbine in a baffled tank, for unbaffled tanks there have been fewer studies and these have been primarily focused on tanks with an off-centre impeller [34–37]. Interestingly, in unbaffled tanks stirred by eccentrically located impellers, MIs of the second type have been observed, but, with a frequency closer to that of the first, of the order of $\sim 0.1F_N$ [36, 51]. This frequency has also been found to increase from $0.116F_N$ to $0.27F_N$ as the eccentricity of the impeller is reduced from a distance of $0.32T$ to $0.1T$ [36], where T is the tank diameter. However, no studies with measurements for the MI frequencies in an unbaffled stirred tank with a centred Rushton turbine have been found in the literature. A recent study [52] of the frequency of the oscillation of the fluid surface in an uncovered, unbaffled stirred tank agitated by a centred, 4 blade flat turbine found oscillations of $0.22F_N$ which they attributed to a precessing MI. Another study [51] investigating the oscillations of the fluid surface for an unbaffled, uncovered tank agitated by a Rushton turbine found oscillations frequencies at $0.74F_N, 1.03F_N, 2.07F_N, 3.10F_N$ and $4.14F_N$ at $Re \approx 15000$ and at $0.50F_N, 0.73F_N, 1.45F_N, 2.18F_N$ and $2.90F_N$ at $Re \approx 95000$. The length of the

dataset captured meant that the minimum resolved frequency was $\sim 0.3F_N$ which is greater than those typically associated with MIs.

Large Scale Structures and Mixing

Knowledge and prediction of the behaviour of macro instabilities and trailing vortices has a large potential for mixing enhancement. In [53] it was observed experimentally that the insertion of a passive scalar in the vortex core of the MI leads to a reduction in mixing time of 20-30% compared to injection outside the vortex core.

The authors in [38,54,55] investigated the effect on mixing of placing feed tubes in an iodine-iodate reaction in the path of the trailing vortices compared to the standard location at the tank surface and other locations in the flow. It was observed that the amount of waste product was reduced to 5% when placing static feed tubes in the path of the trailing vortices compared to 20% when placing the feed tubes at the tank surface. In [38,55] the case of feed tubes placed in the path of the trailing vortices that rotate with the impeller was investigated. A 7-fold decrease in the amount of waste product was observed compared to a feed tube at the tank surface case and a reduction of waste product to 80% of that of the static feed tube in the impeller wake.

These works suggest that a targeted feeding location and rate based on the behaviour of the large scale flow structures could lead to a reduction of mixing time and waste product. However, for these approaches to be possible a method for reconstructing the 3D behaviour of these is first required. Hence, this work aims to lay the groundwork for this to be explored in future studies.

CFD Modelling

Traditionally CFD studies of flows in stirred tanks were conducted using Unsteady RANS models, typically the well-known $k - \epsilon$ model [33,56] and these remain the

typical models in industrial applications. While these models are successful at predicting the mean flows in baffled stirred tanks satisfactorily, they are unable to reproduce the turbulence levels of the flow. In addition, a recent study examining the performance of $k - \omega$ SST, $k - \epsilon$ and SSG turbulence models in baffled and unbaffled tanks [33] found that while the results for the mean flow were consistent between the turbulence models, and at laminar and transitional Reynolds numbers also with DNS for baffled tanks, this was not the case for unbaffled tanks.

An improvement to the RANS-based turbulence models is LES. While the grid resolution for LES is still insufficient to resolve small scale turbulence, it can nevertheless resolve the large scale periodic structures in the flow. The effect of the small scales on the resolved large scales is typically modelled by adding a spatially dependent turbulent viscosity (obtained from modelling the subgrid scales) to the physical viscosity. The first LES simulations in stirred tanks appeared in the middle 1990s [57–59] and are now becoming progressively more prominent as the available computational power increases. More recent works have used LES to investigate a range of topics such as the mixing time [60, 61], solid suspension [62], free surface behaviour in an unbaffled uncovered tank [63] and macro-instabilities [64].

DNS resolves all the turbulent scales, with grid resolutions of the order of the Kolmogorov length scale. While still limited to transitional Reynolds numbers, these have recently started to be used in studies of stirred tanks [30, 31, 65]. DNS has the significant advantage of not requiring any sub-grid turbulence models and can provide accurate results for the mean as well as Reynolds stresses. In particular DNS has been used to investigate the mixing time of the stirred tank in [66] and the periodic behaviour of the trailing vortices and power number in [15].

1.2.2 Data-Driven and Machine Learning Methods

Data-driven and machine learning approaches employ algorithms to fit the model parameters purely based upon data. Recently, there has also been a renewed interest

in the application of these methods to fluid flows, including to the resolution enhancement of under-resolved flow fields [67], coherent structure tracking [68] and the reconstruction of periodic flow features from limited sensor measurements [69, 70]. A recent review of the application of machine learning to fluid dynamics problems is provided in [71], but a brief overview will also be provided below for completeness.

Proper Orthogonal Decomposition and Model Order Reduction Algorithms

Data-driven methods have been applied to fluids for a considerable period of time. The most prominent of these is the Proper Orthogonal Decomposition (POD). POD is an example of unsupervised learning, where a set of velocity field snapshots is taken as an input and an optimal orthogonal basis to represent the data is identified [71]. The POD modes consist of a temporal coefficient and spatial component and provide a linear decomposition of the velocity field. As the identified POD modes are ordered with respect to their energy content, this method can be used for model order reduction by keeping only the leading modes and truncating the energetically smaller modes. This method was first applied to fluid flows by Lumely [72] and a more computationally and memory-efficient adaptation, the method of snapshots, was introduced by Sirovic [73]. In addition, it has close similarities to the Principle Component Analysis (PCA) algorithm employed for model order reduction in computer science [74, 75] and the resulting modes are identical to those produced from the Singular Value Decomposition (SVD) [76].

The POD has been used in many fluids applications, including stirred tanks. The aforementioned trailing vortices have been extensively characterised by applying (POD) mainly to two-dimensional PIV data captured on horizontal or vertical planes [4, 7–13]. 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 [44, 48, 49]. A second pair of POD modes with half the length scale and twice the frequency has also been found [7–10]; these are formed by the nonlinear interaction between the first pair of modes. The energy exchange between modes has also been analysed in [10] and was used to derive the observed frequency relationship between the first and second pair of modes.

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 [8, 9]. In this case, the setup comprises a shaft with two Rushton turbines 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 [11, 77] that have been observed in stirred tank flows [78] as mentioned earlier.

Prior to the present work, only one study has applied POD to the full three-dimensional (3D) flow field in an unbaffled stirred tank [14] over the whole tank volume. The author used LES to simulate the flow agitated by a 3-blade pitched turbine using a combined rotating/static mesh approach. In total, 1500 snapshots of the 3D velocity field were stored over 4 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.

A more recent study applied POD to the 3D flow field in a baffled tank and used 386 snapshots over 28 rotations [79]. The authors demonstrated a method to compute the POD over the entire domain (for the static and rotating mesh together). Expressing all of the velocities in the inertial frame, the POD was carried out with the samples stored at a particular cell at each timestep. As the position of the cells in the rotating zone changes over time, this makes the physical interpretation

of the POD modes more challenging. However, from the perspective of flow field reconstruction, it provides a tool to reconstruct the entire flow field with one set of POD modes in the case where the flow domain is required to be separated into a static and rotating part, such as for a baffled stirred tank.

While the POD identifies the most optimal orthogonal modes by energy, these modes do not necessarily represent the most dynamically important features. Hence, several modifications and extensions to the POD have been proposed. An example is the balanced proper orthogonal decomposition [80] where two sets of modes are identified for specified inputs and outputs. Typically, the former are external disturbances or actuations used for flow control and the latter are sensor measurements or quantities that must be captured by the model. Balanced POD aims to balance the properties of controllability (the modes most excited by the inputs) and observability (those easiest to observe at the selected sensor points).

Another variation is the Dynamic Mode Decomposition (DMD) [81–83], where time-resolved data is decomposed into modes with a single characteristic frequency and a growth or decay rate. This takes as an input, pairs of velocity field snapshots separated by a constant time interval. While this leads to modes that can be more representative of the dynamics than POD, the computation of DMD modes requires a greater computational cost. However, a POD pre-processing step prior to applying the DMD can reduce this and provide better results for noisy data [83, 84]. In the case of a stirred tank, the largest structures are periodic, and in such cases, the POD and DMD have also been found to identify near identical modes [8]

Spectral POD is another alternative [83, 85] that is effective at extracting coherent flow structures from statistically stationary flows. Here the POD is applied in the frequency domain. A set of time series snapshots are segmented into a number of blocks. A temporal Fourier transform is then performed on each block and all realisations of the Fourier transform at a specific frequency are then used to form a new data matrix. The POD is then performed on this new data matrix. This leads

to spectral POD modes that each oscillate at a single frequency.

Non-linear model order reduction methods based upon neural network auto-encoders have also recently appeared [71]. They consist of two parts, the encoder and the decoder. The encoder takes the full velocity field as an input and using non-linear activation functions encodes the data into a low dimensional space. The decoder then takes the data from the low dimensional space and reconstructs the fully dimensional velocity field. The basis of this method is the universal approximation theorem [86]. This theorem states that provided a Neural Network is large enough, with enough learnable parameters, it can represent an arbitrarily complex input-output function. These methods can represent the velocity data better than the POD, however, as the resulting modes are purely data-driven, they can be more difficult to interpret. Auto-encoders have been used for the reduced order modelling of the flow behind a cylinder [87] and the flows around bluff bodies [88].

In this thesis, the POD is used as the model order reduction method due to its relative simplicity and smaller computational requirements than the other methods. The derivation of the POD modes from the SVD of the snapshot matrix allows the well-established, optimised, and parallelised algorithms in the SLEPC library to be used [89].

In addition to the common application of POD to identify large scale periodic structures in the velocity fields, the obtained velocity decomposition can be used to decompose the pressure field. This allows the pressure field to be expressed as a sum of partial pressures, with terms constant, linear and quadratic with respect to the velocity POD mode temporal coefficients [90], [91]. This was applied in [90] to 2D and 3D shear layers and in [91] to the flow behind a cylinder. In this study, for the first time, the shape of these pressure modes for the large scale structures in the stirred tank is presented.

N4SID, LSTM NN and Supervised Learning Algorithms

Another application of data-driven methods is in the reconstruction of the full 2D or 3D flow field, or of large scale flow features using sparse sensor measurements in the flow. They fall into the category of supervised learning, with an input of sensor data such as velocities or pressure and an output of the velocity field [71], or when model order reduction methods are used the temporal coefficient from which the full flow field can be reconstructed [69–71, 92]. In this study, we use POD as the order reduction method and as this decomposes the flow field into temporal coefficients and spatial components, we require a model that produces as an output the temporal coefficients of the POD modes.

One method to obtain such a model is to use system identification to construct a linear estimator in the form of a multiple-input multiple-output linear, time-invariant system. The form of the model is derived by projecting the Navier-Stokes equations onto POD modes, as will be explained in Chapter 3. In the present thesis, the system matrices and the order of the system are determined optimally using the N4SID algorithm [93]. The latter belongs to the family of subspace identification methods; the reader may refer to [94] for a detailed description and to [95] for an overview.

While system identification has been applied to many different areas (see examples in [94]), its application to fluid flows is relatively recent. It was first applied in [92] 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, linearised around the Blasius velocity profile. The assumption of small perturbations was relaxed in [69], where the method was applied to reconstruct non-linear 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 above, the flow was 2D and laminar. In the present thesis the method is applied to complex transitional and fully turbulent 3D flows for the first time. In addition, the use of pressure sensor data as the input is also considered for the first time in this study.

A second approach based upon Long-Short Term Memory Neural Networks (LSTM) [96] has been proposed in [70]. The authors reconstructed the temporal coefficients of the first 50 2D POD modes of a flow around an oscillating flag that had been computed from 2D PIV data. LSTM neural networks are particularly suited to reconstructing sequential and temporal data as they have a memory capability that allows them to store information from past data in a sequence [68, 70, 97, 98].

In contrast to the linear system identification method described, the structure of the LSTM has no direct link to the Navier-Stokes equations. Instead, it relies on a considerably larger number of parameters and the universal approximation theorem mentioned above; it is thus much more difficult to interpret physically. LSTM neural networks have also been used to predict the evolution of the flow directly with no additional input measurements. In [97], 2 datasets consisting of 2D slices of the x component of the velocity fields of forced isotropic turbulence and of magnetohydrodynamic turbulence in a cube are considered. POD is applied to the dataset and the temporal evolution of the x component of the velocity is reconstructed by predicting the evolution of the temporal POD coefficients using an LSTM neural network taking the past values of the temporal POD coefficients as an input. In [88] the authors applied a similar idea to the x and y components of velocity and the pressure field in 2D flows behind bluff bodies, but a Convolutional Neural Network Auto-Encoder is used as the model order reduction method instead of POD. The LSTM is then used to predict the evolution of the flow in the reduced basis from past data. Finally, the full flow field prediction can be recovered from the reduced basis data by using the decoder.

Most papers on LSTM applications to fluids have been published very recently

and the large computational requirements mean that the flows studied have been primarily confined to simple, and typically 2D, cases. This thesis presents one of the first attempts to apply these methods to a fully turbulent 3D complex turbulent flow field in a more industrially relevant geometry.

1.3 Objectives

The central aim of this thesis is to investigate the use of data-driven and machine learning methods to analyse and reconstruct the dynamic behaviour of large scale structures in an unbaffled stirred tank at both design and off-design conditions. This presents an initial step in the direction of quick 3D flow field prediction through the combination of CFD and experimental methods and in the future can lead to more precise control of mixing behaviour.

To summarise, the key objectives are as follows:

- To investigate the large scale periodic flow structures in an unbaffled stirred tank for transitional and turbulent flows.
- To explore the link between the structures identified to the experimentally and computationally observed macro-instabilities and their impact on the instantaneous power number.
- To reconstruct the evolution of the large scale structures in the flow from limited point measurements of velocity in the flow or pressure on blade and tank surfaces for both design and off-design conditions.
- To extend the application of data-driven and machine learning based methods to a complex 3D flow, beyond the simple 2D flow cases typically investigated in the literature.

1.4 Thesis Outline

This thesis is divided into chapters as follows. The present chapter is dedicated to the introduction and motivation of the thesis, as well as a relatively brief literature review. Chapter 2 describes the numerical methodology used for the flow simulation, the validation of the employed code and the methodology against results from the literature. The machine learning algorithms used in this work are presented in Chapter 3. Chapter 4 describes the results for the decomposition and reconstruction of large scale flow structures using velocity field data from sparse measurements for the transitional Reynolds number of 600. Chapter 5 builds on the results presented in Chapter 4 and decomposes the pressure field and power number, and shows the reconstruction of the leading order POD modes using pressure data from blade and wall surfaces. Chapter 6 considers the fully turbulent case ($Re=30000$), further explores the nature of the leading order modes and discusses them in the context of precessing MIs. The performance of the N4SID algorithm used in the previous two chapters is compared to a modified N4SID algorithm with an additional pre-processing step and to LSTM NN based methods. Finally, the main findings of this work are summarised and some future research directions are proposed in Chapter 7.

Chapter 2

CFD Methodology and Validation

2.1 Introduction

In this chapter, the methodology for the computational fluid dynamics simulations is presented. The flow configuration considered is that of an unbaffled top-covered stirred tank, agitated by a 6-blade Rushton impeller. A schematic is shown in Figure 2.1 and the physical dimensions are shown in Table 2.1. The flows studied are of two types, transitional flows are examined in Chapters 4 and 5 with a base case of $Re=600$ and off-design condition cases of $Re=500$ and 700 , and fully turbulent flows are examined in Chapter 6 with a base case of $Re=30000$ and off-design conditions of $Re=20000$ and 40000 . The Reynolds number is defined in the usual way as $Re = \rho ND^2/\mu$, where ρ is the density, N is the impeller speed in rotations per second and D is the impeller diameter. Note, the impeller diameter in this study ($T/2$) is larger than the more typical impeller diameter ($T/3$) investigated in baffled tanks. In the transitional flow cases, an approximation to a stirrer shaft guide and sealing mechanism is made by making the top $0.02T$ of the shaft irrotational. The fluid is assumed to be incompressible throughout.

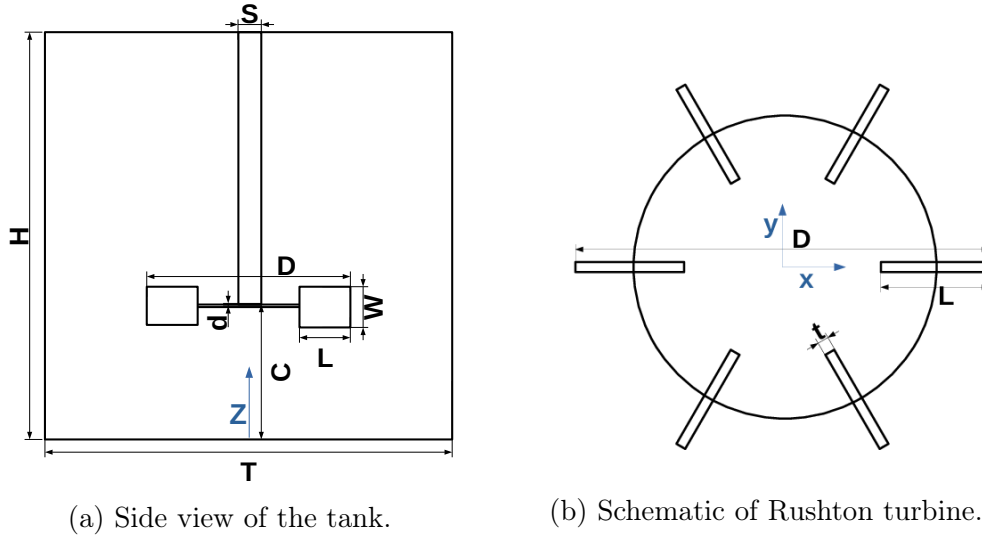


Figure 2.1: Sketch of the stirred tank and the impeller geometry. The axes of the Cartesian coordinate system are shown in blue.

D	S	H	C	L	W	d	t
0.500	0.060	1.000	0.333	0.125	0.100	0.010	0.015

Table 2.1: Normalised Dimensions (T) of the stirred tank considered.

2.2 DNS Simulations

The results for the transitional flows presented in Chapters 4 and 5 have been obtained with DNS. The same geometry and Reynolds number as in the present study has been simulated with DNS in [32] and investigated experimentally in [1]. Simulations for two additional cases, with $Re = 500$ and 700 , were conducted in order to assess the ability of the multi input-multi output model that was identified for $Re = 600$ to reconstruct the flow structures at off-design conditions, thus providing truly predictive capabilities.

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 2.1, the impeller is held at

a fixed position and the governing equations take the form

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} - [\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{x}) + 2\boldsymbol{\omega} \times \mathbf{u}], \quad (2.1a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.1b)$$

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. This vector has a non-zero component only in the axial z direction, i.e. $\boldsymbol{\omega} = (0, 0, \omega)$, 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 (2.1a) 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, peripheral walls and the stirrer shaft guide. This is known as the ‘‘Frozen Rotor’’ approach [99] 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 [30, 31, 60, 100–102]. The convection terms are discretised in space using a second order central scheme. The orthogonal viscous terms are treated implicitly, while the convection, non-orthogonal viscous, Coriolis and centrifugal terms are all treated explicitly in time using quadratic extrapolation. A third-order accurate scheme is used to discretise the transient term. The pressure-velocity coupling is treated using the fractional step method. The code is parallelised with the aid of PETSc [103] and HYPRE libraries [104].

The computational domain was discretised using $\sim 7 \times 10^6$ internal cells. This is between the 3rd and 4th finest grids considered in [32], with the authors choosing the 3rd 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

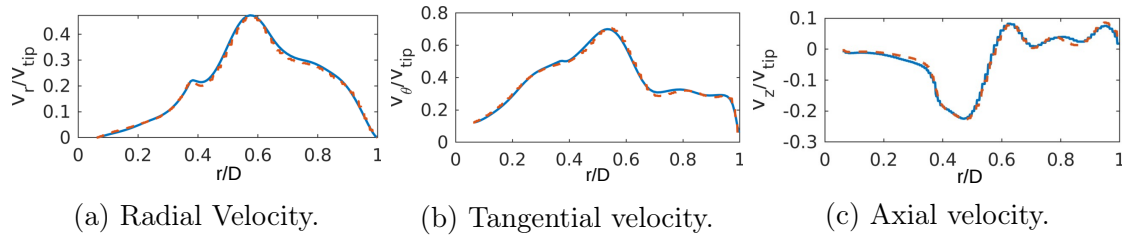


Figure 2.2: Time-averaged velocities along a line bisecting two blades at a distance of $0.353T$ from the tank bottom. Dashed line (---) high resolution grid. Solid line (—) low resolution grid.

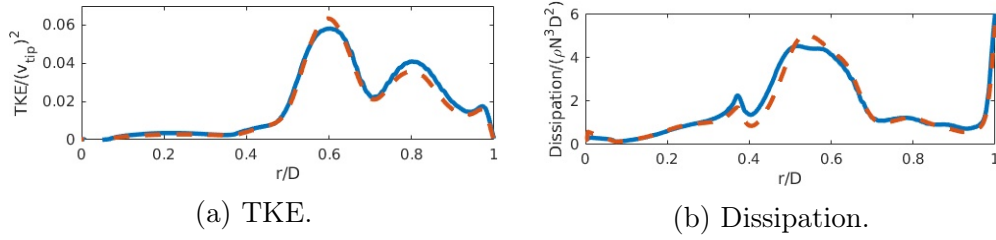


Figure 2.3: Turbulent kinetic energy and dissipation rate of kinetic energy along a line bisecting two blades at a distance of $0.353T$ from the tank bottom. Dashed line (---) high resolution grid. Solid line (—) low resolution grid.

resolution of the flow. The corresponding CFL number was ~ 0.2 , significantly lower than 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 of $\sim 19 \times 10^6$ internal cells. In Figure 2.2 we compare the profiles of time-averaged radial, tangential and axial velocities along a line bisecting two blades at a height of $0.353T$, 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. In Figure 2.3 we compare the profiles of the TKE and dissipation rate of kinetic energy along the aforementioned line. The results from the two grids match well.

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 [56]. The local ratio of the grid spacing to η was estimated from $V_{cell}^{1/3}/\eta_{cell}$, where V_{cell} is the cell volume. For the lower grid resolution, this ratio was found to be less than 1.82 everywhere, while

the average value in the domain was 0.69. These metrics satisfy fully the resolution requirements for DNS [56] and are comparable to those of other DNS studies in stirred tanks [32, 65].

As a final check, we computed the power number, $N_p = \frac{\overline{Po}}{\rho N^3 D^5}$, where $\overline{Po}$ 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 matches well with the experimentally measured value of 2.2 in [1] and obtained using DNS in [32]. While this study considered finite-width impeller blades with a thickness of 0.015T, [32] considered the case of infinitely thin blades. A larger blade thickness has been observed to reduce the power number [29, 41] and is the likely explanation for the small discrepancy between the result from our study and [32]. All the above tests confirm that the lower resolution grid is sufficient, and this grid was used to obtain all the results presented for the DNS simulations in Chapters 4 and 5.

2.3 LES Simulations

A DNS simulation of the fully turbulent flow at $Re=30000$ considered in Chapter 6 is not feasible, thus an LES simulation is performed.

As for the transitional Reynolds number case, the equations are solved in the rotating frame of reference with the so-called "Frozen rotor" approach. A Dirichlet boundary condition of $\mathbf{u}_w = -\boldsymbol{\omega} \times \mathbf{x}$ is imposed on the top, bottom and outer moving walls, where $\boldsymbol{\omega}$ is the angular velocity vector and $\mathbf{x} = (x, y, z)$ is the position vector. In a LES only the large turbulent scales are resolved, due to the insufficient resolution of the grid. The filtered governing equations in the rotating frame for these scales

are given by

$$\frac{\partial \tilde{\mathbf{u}}}{\partial t} + (\tilde{\mathbf{u}} \cdot \nabla) \tilde{\mathbf{u}} = -\frac{1}{\rho} \nabla(\tilde{p} + tr(\boldsymbol{\tau})) + \nu \nabla^2 \tilde{\mathbf{u}} + 2 \nabla \cdot (\nu_e S) - [\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{x}) + 2 \boldsymbol{\omega} \times \tilde{\mathbf{u}}], \quad (2.2a)$$

$$\nabla \cdot \tilde{\mathbf{u}} = 0, \quad (2.2b)$$

where $\tilde{u}$ denotes the filtered variable and the unknown turbulent stress tensor $\boldsymbol{\tau}$ has been replaced by the model

$$\tau_{ij} = \widetilde{u_i u_j} - \tilde{u}_i \tilde{u}_j = -2\nu_e S_{ij} + tr(\boldsymbol{\tau}) \quad (2.3)$$

and

$$S_{ij} = \frac{1}{2} \partial_i \tilde{u}_j + \frac{1}{2} \partial_j \tilde{u}_i \quad (2.4)$$

is the strain rate tensor. The $tr(\boldsymbol{\tau})$ is absorbed into the pressure term. As the mesh used in the simulation is unstructured, the eddy viscosity is estimated using the model proposed in [105] that takes into account the cell non-uniformity. The eddy viscosity is computed from

$$\nu_e = c \sqrt{\frac{B_\beta}{\alpha_{ij} \alpha_{ij}}}, \quad (2.5)$$

with

$$\alpha_{ij} = \partial_i \tilde{u}_j, \quad (2.6)$$

$$\beta_{ij} = \Delta_m^2 \alpha_{mi} \alpha_{mj}, \quad (2.7)$$

$$B_\beta = \beta_{11} \beta_{22} - \beta_{12}^2 + \beta_{11} \beta_{33} - \beta_{13}^2 + \beta_{22} \beta_{33} - \beta_{23}^2. \quad (2.8)$$

The model constant c is related to the Smagorinsky constant C_s through $c \approx 2.5 C_s^2$ and is set to 0.025 corresponding to $C_s = 0.1$ which while smaller than the theoretical value of 0.17 has been shown to have a good performance in stirred tanks [106], [100], [65]. Δ_m is the cell size in the m -th direction. As the grid is unstructured, this is computed by finding the distance between the maximum and minimum cell face

centre locations in the m -th direction.

As for the DNS, the equations are solved using our in-house code "Pantarhei". A grid convergence study was performed using three mesh sizes, a coarse mesh of $\sim 7 \times 10^6$ internal cells, a finer mesh of $\sim 13 \times 10^6$ internal cells and a finest mesh of approximately 19×10^6 internal cells. The time and azimuthally averaged tangential, axial and radial velocities at a height of $z/T = 0.211$ from the tank bottom were compared for the three mesh sizes. The tangential velocity results were further compared to LES data from [106] ($\sim 7.6 \times 10^5$ cells) and [107] ($\sim 2.7 \times 10^6$ cells), and PIV data from [108] and to SSG data at a slightly higher Reynolds number of 33000 from [33]. As the meshes are unstructured and non-uniform, the results in the plane at the considered height were interpolated onto a uniform mesh with 200 and 100 equispaced points in the azimuthal and radial directions respectively and then the time and azimuthal averages were performed.

The results for the tangential velocity are displayed in Figure 2.4a. It can be seen that the velocity profiles for the two finer grids are nearly identical, while the low resolution grid shows good agreement for $r/D > 0.35$. For $r/D > 0.35$ the results for all three meshes considered agree well with the data from the literature. For $r/D < 0.35$ the low resolution grid appears to overpredict the velocity profile compared to other studies, however, both the medium and fine resolution grids show a slight underprediction.

The results for the axial velocity are displayed in Figure 2.4b. All grids show very similar results and the medium and high resolution grids are nearly identical. The results also match well with those obtained at the slightly lower height of $z/T = 0.2$ in [106]. The results for the radial velocity are displayed in Figure 2.4c. The values are quite small, but all three mesh sizes show very similar results.

As the focus of this study is on large periodic structures, the prediction of the correct path of the trailing vortices is important. Following the approach of [3], the path of the trailing vortex predicted was identified by the points where the mean axial

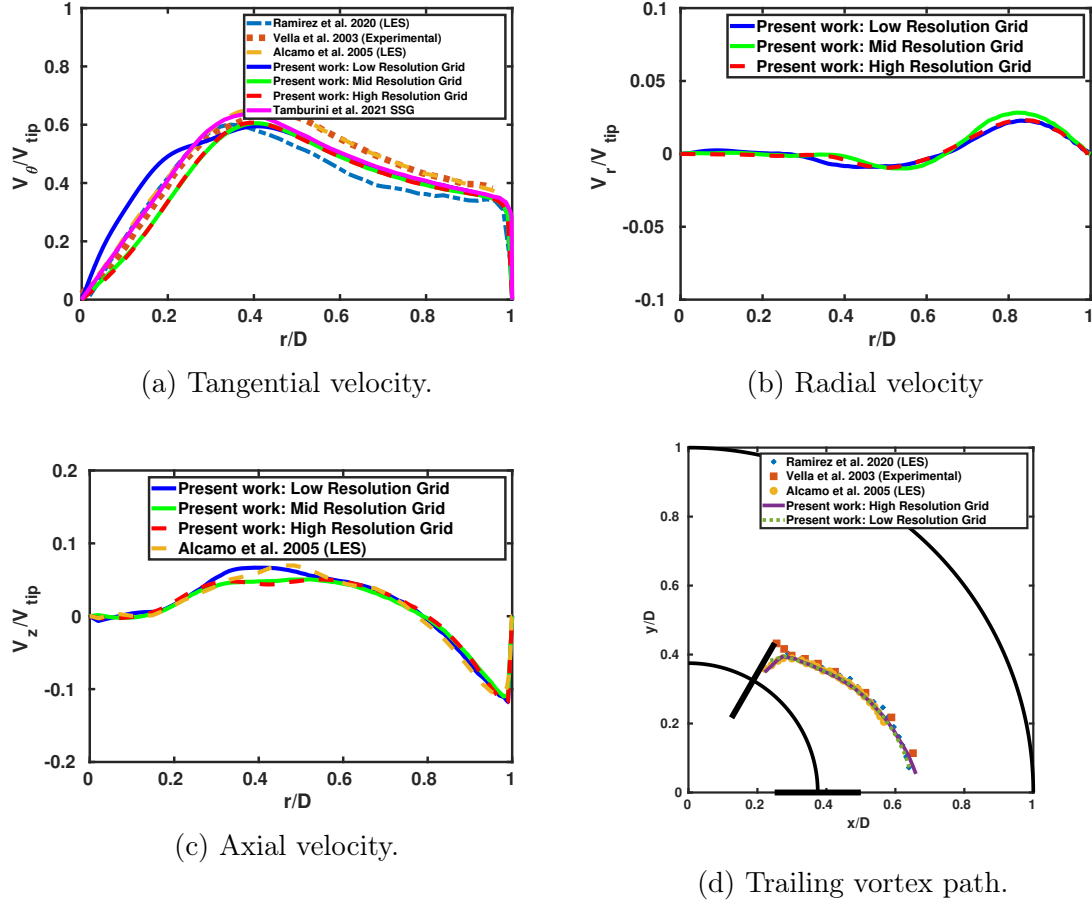


Figure 2.4: Time and tangentially averaged velocity at $z/T = 0.211$ and trailing vortex path comparison for different meshes against data from other studies.

velocity is equal to zero in the horizontal plane near the lower part of the blade. At the centre of a vortex the velocity is close to zero in all directions, however, the path of the centre of the vortex is subject to fluctuations. Hence, the axial velocity is considered as the displacement of the vortex axis in the axial direction is small [43]. The comparison of the high and low resolution meshes to the results obtained from [106], [108] and [107] is shown in Figure 2.4d. There is an excellent agreement between the reference data and the meshes.

Finally, the power number is examined. Calculating the integral of the torque over the blades gives a power number of 1.0 for the high resolution mesh. This is an overestimate compared to the experimental measurement of ~ 0.7 in [1], however, is very similar to the values of 1.01 and 1.05 obtained in [106] and [109] respectively and to the power numbers of ~ 0.85 , ~ 1.3 and ~ 1.5 obtained using SSG, $k-\omega$ SST and $k-\epsilon$ turbulence models for a slightly higher Reynolds number of 33000 in [33]. It should be noted that all the computational studies ([33, 106, 109]) considered the case of infinitely thin impeller blades. On the other hand, our work considers impeller blades of thickness $0.015T$.

On the whole, the results match well with the literature. The results for the medium and high resolution grids matched very well.

We also checked the resolution of the finest grid with respect to the Taylor length scale $\lambda_T \approx \sqrt{(10\nu k/\epsilon)}$ where k is the TKE and ϵ is the dissipation rate of TKE. The local ratio of the grid spacing to λ_T was estimated from $V_{cell}^{1/3}/\lambda_{T,cell}$, where V_{cell} is the cell volume. The volume average of this value was found to be 0.32 and the maximum was 2.5. However, the maximum values are found in the vicinity of the wall and blade surfaces where the resolution of the grid perpendicular to the surfaces is finer compared to the other dimensions. Therefore, the high resolution grid is deemed to have a sufficient resolution and is the one used in this study.

Chapter 3

Data-Driven and Machine learning Methods

3.1 Introduction

In this chapter the data-driven and machine learning methods used in this work are presented. For all three results chapters (4, 5, 6) the POD is used as a model order reduction method to extract the large and most energetic flow structures that are then reconstructed. A brief description of this method will be presented first. Next, the N4SID algorithm that is used for the reconstruction of the large scale flow structures from limited flow measurements will be presented. Finally, the modification to the N4SID with an additional pre-processing step and the LSTM NN method that are considered in Chapter 6 are described.

3.2 Proper Orthogonal Decomposition

We use Proper Orthogonal Decomposition (POD) to construct a reduced order model of the velocity fluctuations around the mean flow. POD is a modal decompo-

sition 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, the reader may refer to the book [110] and the review [83].

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, for the DNS case, the governing equations of the fluctuations, obtained by subtracting the time-averaged RANS equations from set (2.1), are provided below,

$$\frac{\partial \mathbf{u}'}{\partial t} + (\bar{\mathbf{u}} \cdot \nabla) \mathbf{u}' + (\mathbf{u}' \cdot \nabla) \bar{\mathbf{u}} + \frac{1}{\rho} \nabla p' - \nu \nabla^2 \mathbf{u}' + 2\boldsymbol{\omega} \times \mathbf{u}' = \overline{(\mathbf{u}' \cdot \nabla) \mathbf{u}'} - (\mathbf{u}' \cdot \nabla) \mathbf{u}', \quad (3.1a)$$

$$\nabla \cdot \mathbf{u}' = 0. \quad (3.1b)$$

We have arranged the equations so that all terms linear in the fluctuating variables are on the left hand side. The centrifugal term cancels out, while the Coriolis term is linear and thus appears on the left hand side. The corresponding expression for the LES case will also contain terms associated with the turbulent viscosity. The velocity fluctuations can be expanded as

$$\mathbf{u}(\mathbf{x}, t) = \bar{\mathbf{u}} + \mathbf{u}'(\mathbf{x}, t) = \bar{\mathbf{u}} + \sum_{i=1}^{i=\infty} a_i(t) \phi_i(\mathbf{x}) = \sum_{i=0}^{i=\infty} a_i(t) \phi_i(\mathbf{x}) \approx \sum_{i=0}^{i=n} a_i(t) \phi_i(\mathbf{x}) \quad (3.2)$$

where $a_i(t)$ is the temporal coefficient, $\phi_i(\mathbf{x})$ is the spatial distribution of mode i , and n is the order of the truncated expansion. The second equality follows from writing the mean as the 0th mode of the decomposition with $a_0 = 1$ and $\phi_0 = \bar{\mathbf{u}}$. The final term is an approximation to the full flow field obtained by truncating the series. POD orders the modes by decreasing energy content, hence the majority of the energy is captured in the truncated series.

We define the inner product between two fluctuating vectors $\mathbf{u}'_i$ and $\mathbf{u}'_j$ as

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \int_v \mathbf{u}'_i{}^\top(x) \mathbf{u}'_j(x) dV \quad (3.3)$$

where $^\top$ 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 $\phi_i(\mathbf{x})$ are obtained by maximising the average projection of $\mathbf{u}'(\mathbf{x}, t)$ to $\phi_i(\mathbf{x})$; the projection is defined using the inner product (3.3). It can be shown [110] that this optimisation problem can be reduced to an eigenvalue problem, which we solve using the method of snapshots. More specifically, we store three velocity fluctuations around the time-average $u' = u - \bar{u}$, $v' = v - \bar{v}$ and $w' = w - \bar{w}$ at each cell centroid $\mathbf{x}_k$ ($k = 1 \cdots m$) at time t_j to form the vector

$$\mathbf{y}(t_j) = (u'(\mathbf{x}_1, t_j), \cdots, u'(\mathbf{x}_m, t_j), v'(\mathbf{x}_1, t_j), \cdots, v'(\mathbf{x}_m, t_j), \\ w'(\mathbf{x}_1, t_j), \cdots, w'(\mathbf{x}_m, t_j))^\top, \quad (3.4)$$

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

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}(t_1), \mathbf{y}(t_2), \dots, \mathbf{y}(t_K) \end{pmatrix} \quad (3.5)$$

It is shown in [83, 110] that the POD modes can be obtained from the singular value decomposition (SVD) of the matrix $\mathbf{V}^{1/2}\mathbf{Y}$, i.e.

$$\mathbf{V}^{1/2}\mathbf{Y} = \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Psi}^\top. \quad (3.6)$$

where $\mathbf{V}$ is a $3m \times 3m$ weighting matrix that accounts for the grid non-uniformity. For a second order finite volume discretisation, $\mathbf{V}$ is a diagonal matrix with the cell volume in the main diagonal. In (3.6), $\mathbf{\Sigma}$ 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 $a_i(t_j)$ are obtained from the projection $\mathbf{Y}^\top \mathbf{V}^{1/2} \Phi$.

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 in Chapters 4 and 5, the flow was simulated for another 120 rotations and 210 rotations respectively and for the main $Re = 30000$ case for Chapter 6 for another 100 rotations. During the simulations, we saved approximately 100 snapshots of the velocity field in the whole domain per rotation. This provided a dataset of approximately 12000 and 21000 snapshots for the $Re = 600$ case for Chapters 4 and 5 respectively, and 10000 snapshots for the $Re = 30000$ case from which the spatial distribution of the POD modes $\phi_i(\mathbf{x})$ ($i = 1 \dots n$) were computed using SLEPC [89, 111], a library that provides parallel subroutines for solving SVD problems efficiently.

From the above it is clear that the same grid used for the flow simulation was also used for the POD mode extraction, i.e. 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 an additional 20-40 revolutions for model validation. In the present study, we used conservatively long datasets (120, 210 and 100 rotations) to ensure that we also capture slowly varying features (this is important for Chapter 5 where we find low frequencies in the spectrum of the power number).

In this study and also in other works that have performed a 3D POD in a stirred tank [14, 79] the POD is performed on the full tank volume. However, the computational cost of the decomposition could be reduced by performing the POD only on the sub-regions of interest, such as in the vicinity of the impeller. Nevertheless, as the

extent of the leading modes was not known a priori, the choice to perform the POD over the entire tank volume was made. In particular, while the leading order modes obtained for the $Re=600$ case investigated in Chapters 4 and 5 were concentrated in the region around the impeller, the second pair of leading order modes in the $Re=30000$ examined in Chapter 6 were found to have the largest energies near the top of the tank. Therefore, one should be careful not to miss important dynamical features by limiting the decomposition to a specific subvolume.

3.3 Flow Field Reconstruction

After performing the POD and obtaining the modes corresponding to the large scale flow structures we now consider their reconstruction from velocity or pressure data from sparse locations in the flow or on the impeller blade surfaces.

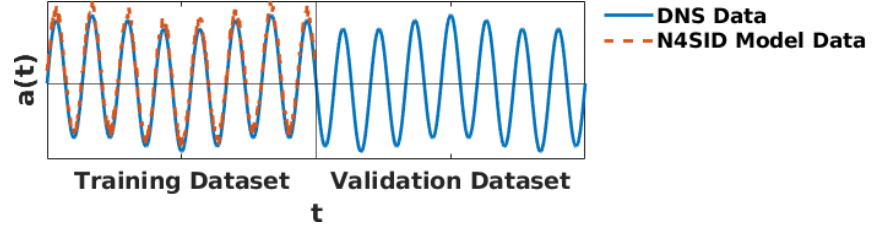
The identification process consists of three procedural steps, which are shown schematically in Figure 3.1. In the first step, the DNS simulations are performed to generate the input and output data, $\mathbf{s}(t_j)$ and $\mathbf{a}(t_j)$, respectively. In the second step, the optimal model parameters are computed by maximising 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. This follows the train-validation split convention from [69].

3.3.1 Linear Estimator Using System Identification

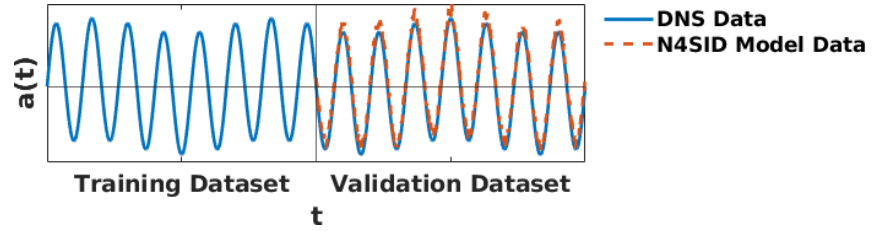
We first discuss the method that is applied in Chapters 4, 5 and 6, which is based on the construction of a linear multiple-input, multiple-output system. We define the state vector of the temporal POD coefficients $\mathbf{a}(t) = (a_1(t), \dots, a_n(t))^T$ and obtain its evolution equation by performing Galerkin projection, i.e. by substituting 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 3.1: Procedural steps for system identification.

expansion (3.2) into the fluctuating equation set (3.1) and taking into account that the basis is divergence-free and orthonormal. The resulting system takes the form

$$\frac{d\mathbf{a}}{dt} = \mathcal{A}\mathbf{a}(t) + \mathbf{F}(t) \quad (3.7)$$

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 non-linear terms that appear in the right hand side of (3.1a) as well as the fact that we retain only a finite number of POD modes (for more details on the derivation of (3.7) see [69]). 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.

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 (3.2) 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 [112], it is known that an optimal estimation $\mathbf{a}_e(t)$ can be obtained from

$$\frac{d\mathbf{a}_e}{dt} = \mathcal{A}\mathbf{a}_e(t) + \mathcal{L}(\mathbf{s}(t) - \mathbf{S}\mathbf{a}_e(t)) \quad (3.8)$$

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 (3.8) to get

$$\frac{d\mathbf{a}_e}{dt} = (\mathcal{A} - \mathcal{L}\mathbf{S})\mathbf{a}_e(t) + \mathcal{L}\mathbf{s}(t) \quad (3.9)$$

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 (3.9) motivates the more general form of the estimator

$$\frac{d\mathbf{X}_e}{dt} = \mathbf{A}\mathbf{X}_e(t) + \mathbf{B}\mathbf{s}(t), \quad (3.10a)$$

$$\mathbf{a}_e(t) = \mathbf{H}\mathbf{X}_e(t), \quad (3.10b)$$

where we have introduced the internal 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 (3.9) 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 (3.10), i.e. 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 below. A similar approach for the pressure measurements results in the same equations. This shall be discussed further in Chapter 5.

System identification using the N4SID algorithm from DNS or LES data

Since the available data $\mathbf{a}(t_j)$ and $\mathbf{s}(t_j)$ from the DNS or LES simulations are known at the discrete time instants t_j , we write system (3.10) in discrete form as

$$\mathbf{X}_e(t_{j+1}) = \mathbf{A}'\mathbf{X}_e(t_j) + \mathbf{B}'\mathbf{s}(t_j), \quad (3.11a)$$

$$\mathbf{a}_e(t_j) = \mathbf{H}'\mathbf{X}_e(t_j) \quad (3.11b)$$

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 $\mathbf{A}'$, $\mathbf{B}'$, $\mathbf{H}'$ 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 (3.11), 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, i.e.

$$\mathbf{X}_e(t_{j+1}) = \mathbf{A}'\mathbf{X}_e(t_j) + \mathbf{B}'\mathbf{s}(t_j) + \mathbf{w}(t_j), \quad (3.12a)$$

$$\mathbf{a}_e(t_j) = \mathbf{H}'\mathbf{X}_e(t_j) + \mathbf{v}(t_j) \quad (3.12b)$$

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

$$\mathcal{E} \left[\begin{pmatrix} \mathbf{w}(t_p) \\ \mathbf{v}(t_p) \end{pmatrix} \begin{pmatrix} \mathbf{w}(t_q)^\top & \mathbf{v}(t_q)^\top \end{pmatrix} \right] = \begin{pmatrix} \mathbf{Q}' & \mathbf{S}' \\ \mathbf{S}'^\top & \mathbf{R}' \end{pmatrix} \delta_{pq}, \quad (3.13)$$

where $\mathcal{E}$ 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.

At the heart of the N4SID algorithm for the computation of the linear model matrices 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 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 $\mathbf{A}', \mathbf{B}', \mathbf{H}', \mathbf{Q}', \mathbf{S}', \mathbf{R}'$ are computed using least squares. For more details refer to [94, 113].

3.3.2 Point-POD+N4SID

The Point-POD+N4SID approach is a modification of the previously described method with an additional pre-processing step and is shown schematically in Figure 3.2. A second unweighed POD is performed on the training dataset of the input data. This identifies temporal and spatial coefficients in the same manner as for the full tank POD. The first few temporal coefficients of the Point-POD are then used as the inputs to the N4SID algorithm. This method leverages the idea that, provided the input data points are chosen in places with strong signals from the periodic structures corresponding to the POD modes identified from the full tank POD, the leading Point-POD modes have similar temporal coefficients to the leading order modes of the full tank POD. Selecting just the first few Point-POD temporal coefficients then allows the periodic signals associated with these structures to be identified and the noise to be removed. For the validation dataset and off-design conditions, the Point-POD temporal coefficients can then be computed by projecting the point velocity component data onto the Point-POD spatial coefficients. These are then used as the input data to the identified model as before.

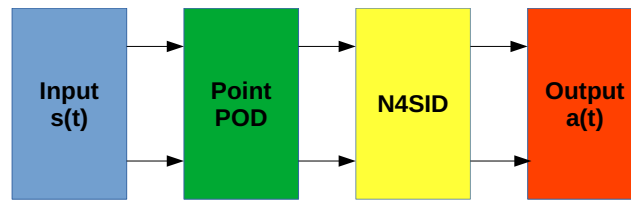


Figure 3.2: Schematic of the Point-POD+N4SID.

This approach is very similar to that applied in extended POD (EPOD) where two PODs are performed, one on the full flow field and one on a set of measurement quantities such as point measurements [114, 115] or wall quantity measurements [116]. The full flow field POD temporal coefficients are then computed by projecting the measurement quantity POD modes onto the global POD modes. Further, the retention of only some measurement quantity POD modes to reduce the noise in the input measurement was proposed in [114]. Here, this approach is modified by using N4SID as the algorithm to connect the Point-POD mode temporal coefficients to the global POD temporal coefficients, rather than using the static estimator proposed in EPOD.

3.3.3 LSTM Neural Network

Long-Short Term Memory (LSTM) neural networks are a class of neural networks that were first developed in [96] for the problem of speech recognition. They have subsequently been successfully applied to problems involving sequential data, including time series data [70, 71]. As mentioned in Chapter 2, the LSTM neural network approach has no direct link to the Navier-Stokes equations and instead relies on a large number of fittable parameters (the weights and biases) and the universal approximation theorem.

A neural network is typically composed of an input layer, one or more hidden layers and an output layer. Each layer consists of a set of neurons or cells. The input data is fed into the input layer and then layer by layer, the information is propagated through the network until it reaches the output layer. The input to a hidden, or

output, layer is determined by the application of weights, biases and an activation function on the output of cells in the preceding layer. In the case of a LSTM neural network, a further input to the cell comes from its previous cell state, allowing the cells to have an intrinsic memory functionality. The mathematical formulation of this will be described below.

The LSTM neural network approach and architecture used in this work is based upon the structure used in [70]. The input data is again taken to be the velocity components at limited data points. The LSTM approach is found to perform better when the input data is taken at not just the current timestep, but also for past timesteps [70], thus explicitly including a memory effect. In the present study, we use the data from the current and preceding 74 timesteps (1% of the training dataset and approximately 75% of an impeller rotation) for the input data at each timestep. We shall discuss this choice further in Chapter 6. The output data are the temporal coefficients of the POD modes as before. For neural networks to perform well the data must be pre-processed [117]. First, the time average is subtracted from the input signal to centre the data around 0. As we are considering velocity fluctuations, this step can be skipped in our case. Next, both the input and output data are normalised to be between 0 and 1 using min-max normalisation as

$$k' = (k - k_{min}) / (k_{max} - k_{min}) \quad (3.14)$$

where k and k' represent the raw and normalised input or output respectively, and k_{max} and k_{min} represent the maximum and minimum values of k in the dataset.

The LSTM neural networks are implemented using the Keras and Tensorflow libraries in Python [118]. We consider two neural networks in this study, with one or two hidden layers, which we denote as LSTM 1 and LSTM 2 respectively. Both neural networks consist of an input layer with the number of neurons equal to the number of inputs, one or two hidden LSTM layers each followed by a dropout layer and separated by a batch normalisation layer, and finally a dense layer with the

number of neurons equal to the number of outputs. A schematic of the LSTM 2 network with two hidden layers is shown in Figure 3.3.

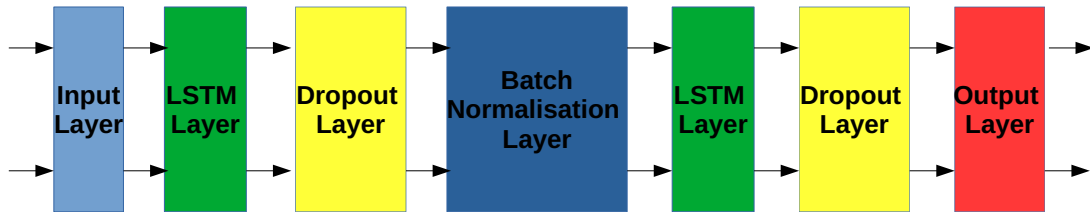


Figure 3.3: Schematic of a LSTM with two hidden layers.

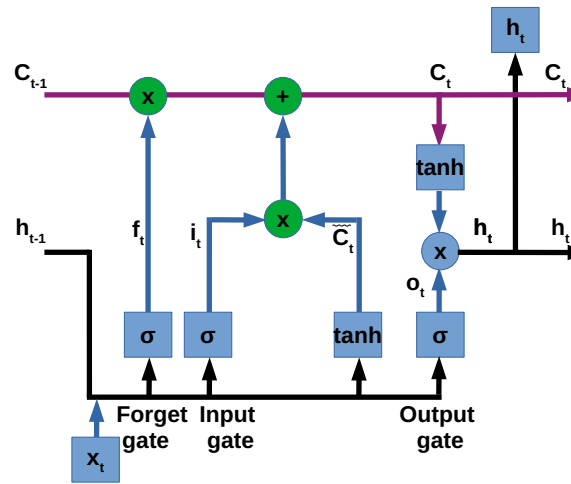


Figure 3.4: Schematic of a LSTM Cell.

The LSTM layers are composed of LSTM cells that consist of three gates: input, output and forget. The gates allow the cell to obtain and retain information for long periods of time, with the forget gate determining whether the information is retained or discarded [70]. A schematic of an LSTM cell is shown in Figure 3.4. The

equations governing a LSTM cell are given by [97]

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3.15a)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3.15b)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (3.15c)$$

$$C_t = i_t \otimes \tilde{C}_t + f_t \otimes C_{t-1} \quad (3.15d)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3.15e)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (3.15f)$$

where x_t and h_t are the cell input and output, C_t and $\tilde{C}_t$ the current and updated cell state, i_t , o_t and f_t take values between 0 and 1 and correspond to whether the input, output and forget gates respectively are fully closed (0), fully open (1), or in between. The W and U are weights and b are biases that are learned during the training process, σ denotes the logistic sigmoid function and $\otimes$ represents element wise multiplication.

The drop-out layers act as a way to reduce potential overfitting of the model by switching off a random 20% of connections from the preceding layer during every training epoch (defined in the following paragraph). The batch normalisation layer accelerates the training of the neural network and reduces overfitting by normalising its per batch output to have a mean of 0 and a standard deviation of 1. Finally, the dense output layer maps the hidden layer outputs to the model outputs.

The weights and biases of the network are computed using back-propagation [117]. This involves first feeding the input data through the network to provide an estimation of the true outputs. The loss function, taken to be the root mean squared error (RMSE) between the true and estimated outputs, is then computed. The gradients of the loss function with respect to the connection weights and biases are then back-propagated and the weights and biases are adjusted to minimise the error. This process of feeding data forwards and backwards (one epoch) is then repeated until

the loss function is no longer reducing. The Adam optimiser [119] is used to perform the back-propagation in an optimised way.

3.3.4 Fit Quality

The performance of the models in reconstructing the temporal POD coefficients was evaluated using the percentage fit between the modelled and the true output for the i -th mode defined as

$$\text{FIT}_i[\%] = 100 \times \left(1 - \frac{\|a_i(t_k) - a_{i,e}(t_k)\|}{\|a_i(t_k) - \overline{a_i(t_k)}\|} \right), \quad (3.16)$$

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 $a_{i,e}(t)$.

3.3.5 Velocity Field Reconstruction

After the temporal coefficients of the leading POD modes have been computed using one of the above methods, the corresponding 3D velocity field can be reconstructed using (6.1). This allows the time varying 3D field to be visualised and opens the possibility of using the reconstructed velocity field in chemical reaction and mixing process computations.

Chapter 4

Reconstruction of large scale flow structures in a stirred tank from sparse sensor data *

4.1 Introduction

In the present chapter, reduced order modelling and system identification are combined to reconstruct the temporal evolution of large scale vortical structures behind the blades of a Rushton impeller. Direct Numerical Simulations are performed at Reynolds number 600 as described in Chapter 2 and the POD described in Chapter 3 was employed to extract the dominant modes and their temporal coefficients. The identification algorithm, N4SID, described in Chapter 3, is applied to construct an estimator that captures the relation between the velocity signals at sensor points (input) and the POD coefficients (output). The effect of the number of input locations used on the fit quality is investigated. Finally, the estimator derived at $Re=600$ is applied to the flows at the off-design Reynolds numbers $Re=500$ and 700 .

*Results presented in this chapter have been published as K. Mikhaylov, S. Rigopoulos, and G. Papadakis, "Reconstruction of large-scale flow structures in a stirred tank from limited sensor data," *AIChE J.*, vol. 67, no. 10, p. e17348, 2021 [120].

The aim of this chapter is to answer the following open questions. Can periodic structures in a stirred tank agitated by a Rushton turbine be identified and extracted using 3D POD? What do they look like? Can the leading order POD modes be reconstructed using data from a single or more limited number of sensor locations? What effect does the number of sensor locations have on the model performance? Can the identified model be successfully applied to off-design conditions?

4.2 Characterisation of POD Modes

Figure 4.1a shows the energy content of each POD mode as a fraction (percentage) of the turbulent kinetic energy integrated over the vessel. Modes 1 and 2 have identical contributions, equal to 14% each. Modes 3 and 4 also have nearly identical energy, but much smaller contributions, equal to 2.6% 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 4.1b 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 86%, of the total kinetic energy of the flow, and together with the first 100 modes captures more than 95% of the energy.

These results are very similar to those obtained for both a single Rushton turbine [7,12] and a pair of Rushton turbines [8,9]. 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 [7–10].

Contour plots of the velocity magnitudes of modes 1-4 in a horizontal slice just above the impeller disk (0.353T from the bottom of the tank) are shown in Figure 4.2. Both pairs modes have rotational symmetry of 180° and show vortex structures that have also been observed in other studies [7, 8, 10] and have what is considered a mesoscale length scale [5]. An animation of the velocity field $\mathbf{u}'_{1\&2}(\mathbf{x}, t) =$

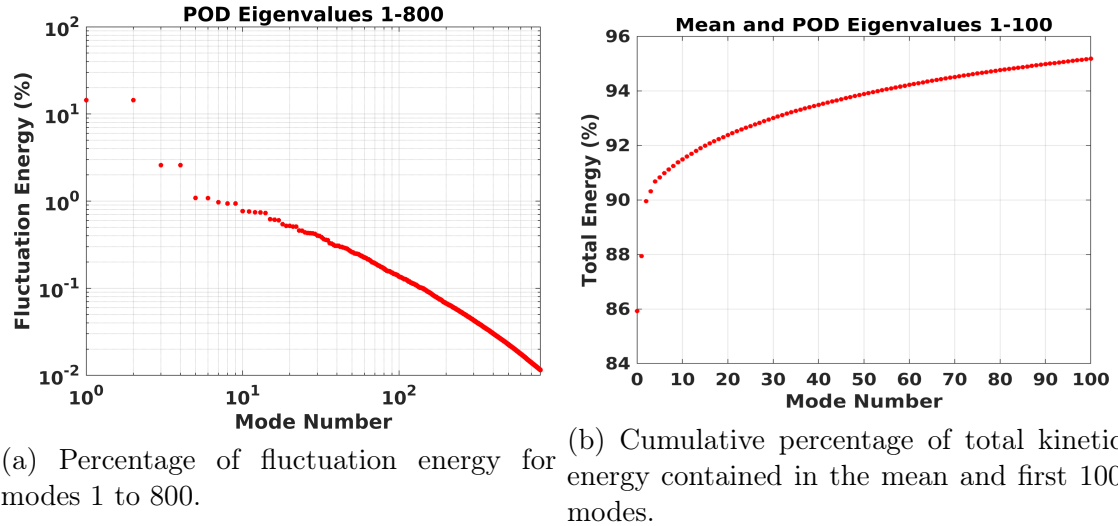


Figure 4.1: Energy content of POD modes.

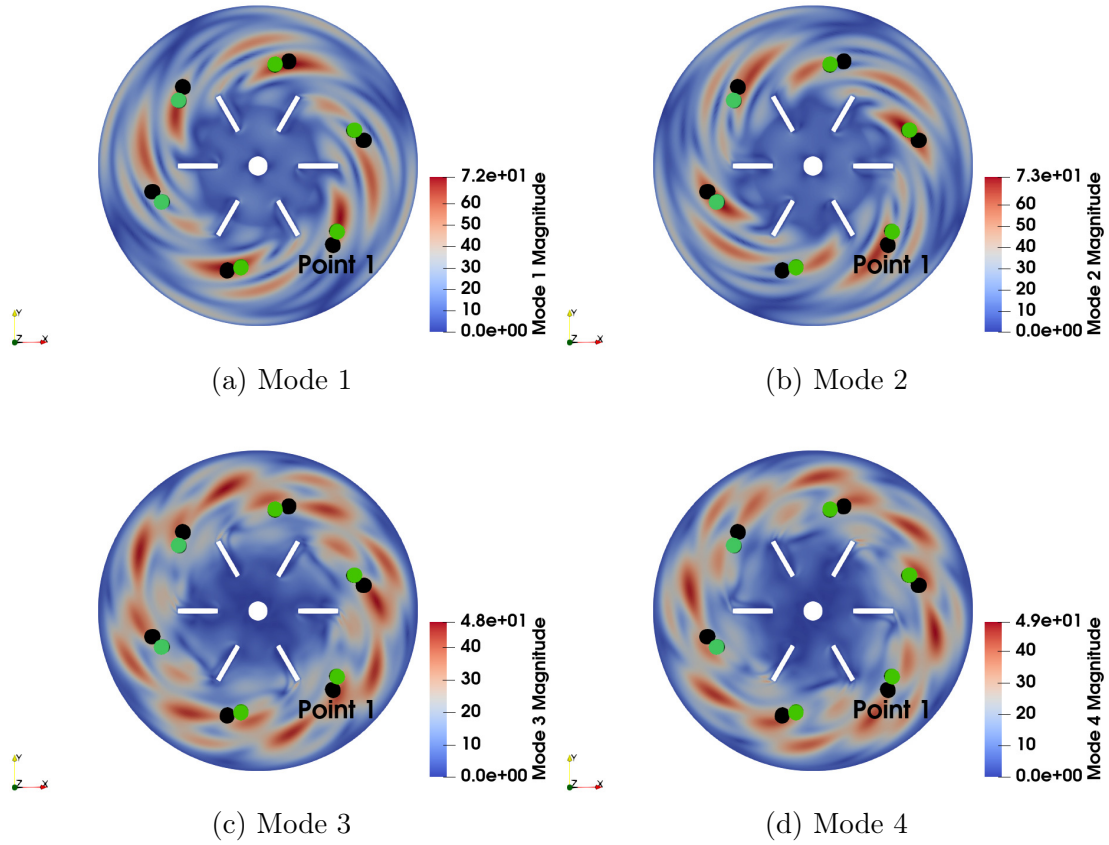


Figure 4.2: Contours of the velocity magnitudes of the first 4 POD modes on a z plane just above the impeller disk at the height of the maximum RMS of velocity fluctuations ($0.353T$ from the tank bottom). Black and green points denote the locations of maxima of the RMS of the velocity fluctuations and the RMS of the sum of Modes 1 and 2 respectively.

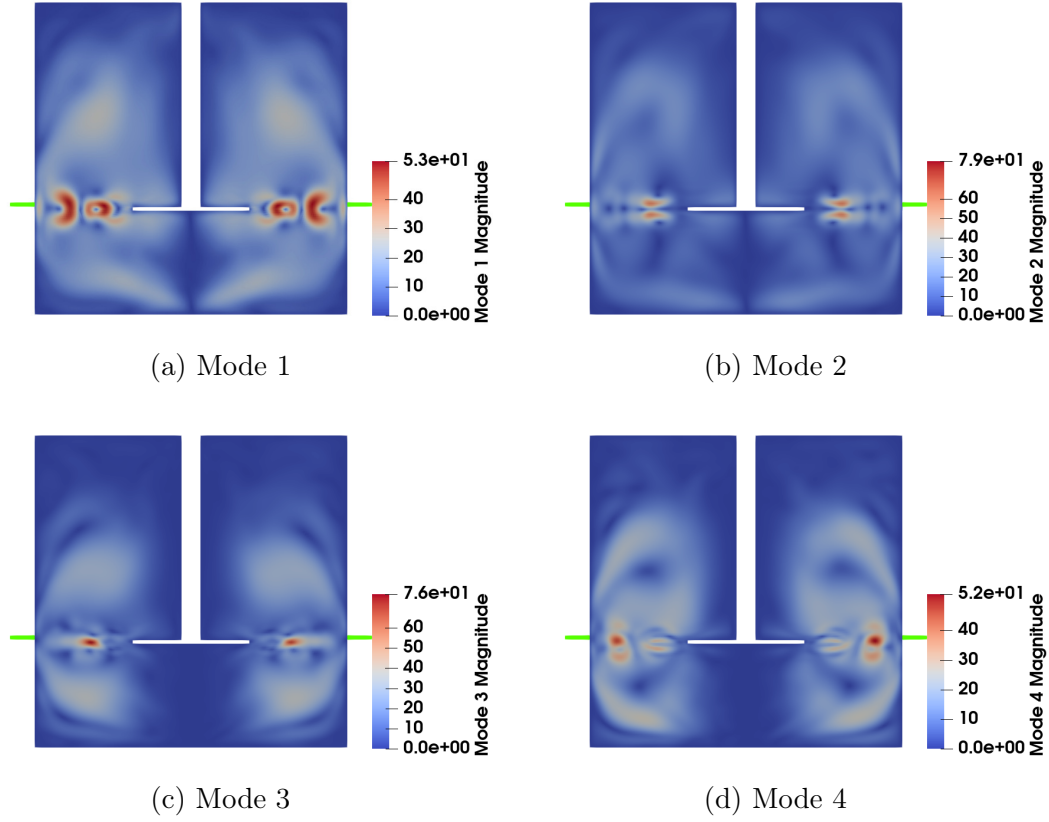


Figure 4.3: Contours of the velocity magnitudes of the first 4 POD modes on a vertical plane bisecting the blades. The horizontal green lines indicate the height of maximum $\sqrt{\mathbf{u}'^2}$ and $\sqrt{\mathbf{u}'_{1\&2}^2}$ values.

$a_1(t)\phi_1(\mathbf{x}) + a_2(t)\phi_2(\mathbf{x})$ shows the periodic formation and shedding of vortices in the wake of each blade; once released, these vortices rotate around the vessel. The length scale of modes 3 and 4 is approximately half that of the first two modes, a feature also noted in [7]. The locations of the maxima of the 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 4.3. The modes are symmetric with respect to the z -axis, as expected, and this indicates that a large enough number of samples was used. As can be seen, the modes impinge on the wall and are deflected in the

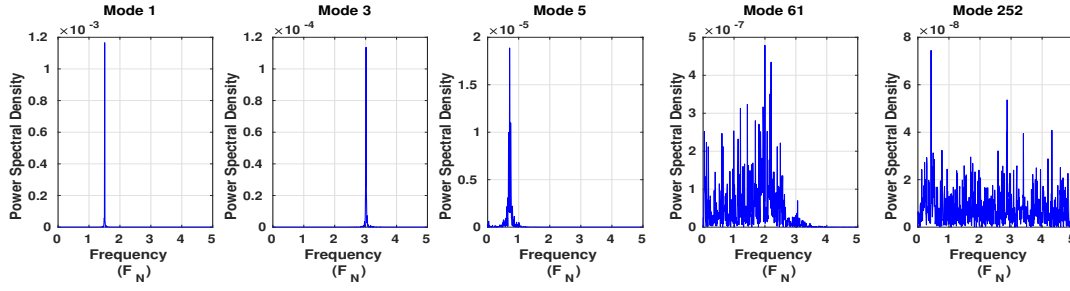


Figure 4.4: Power spectral density of POD modes 1, 3, 5, 61 and 252.

vertical direction both upwards and downwards, thereby occupying a relatively large fraction of the vessel, but their presence is strongest 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 $z = 0.353T$, i.e. the plane shown in the Figure 4.2. The presented mode plots are very similar to the turbulent kinetic energy contour plots of [121], 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 and have identical spectra. In Figure 4.4 we plot the power spectral density of $a_i(t)$ for modes $i = 1, 3, 61$ and 252. Modes 1 and 2 both have a very clear dominant frequency of $1.5F_N$, where F_N is the impeller rotation frequency, while the frequency for Modes 3 and 4 is the first harmonic, i.e. $3F_N$.

The spectra of $a_5(t)$ and $a_6(t)$ peak at $0.7F_N$ (close to the sub-harmonic of the dominant $1.5F_N$) but spectral content starts to appear in other frequencies as well. Modes 1 to 6 all show clear evidence of structure. As the mode number increases, the power spectrum becomes increasingly 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 a subset of homogeneous isotropic turbulence. We examined also the probability density function plots of $a_i(t)$ and found that these morph smoothly to Gaussian distribution with increasing mode number (Figure 4.5).

Plots of a_1 against a_2 , a_3 against a_4 , and a_1 against a_3 are shown in Figure 4.6. The

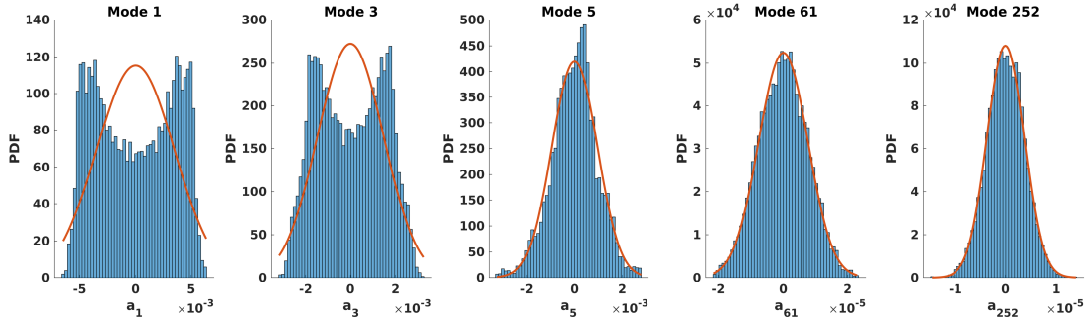


Figure 4.5: Probability density function plots (blue) and fitted Gaussian distributions (orange) of POD modes 1, 3, 5, 61 and 252.

plots of a_1 against a_2 , and of a_3 against a_4 both have a circular shape, confirming that the modes are indeed paired [7, 12, 14]. The plot of a_1 against a_3 has a "butterfly" shape that has been observed for cases when one mode is a harmonic of the other [7].

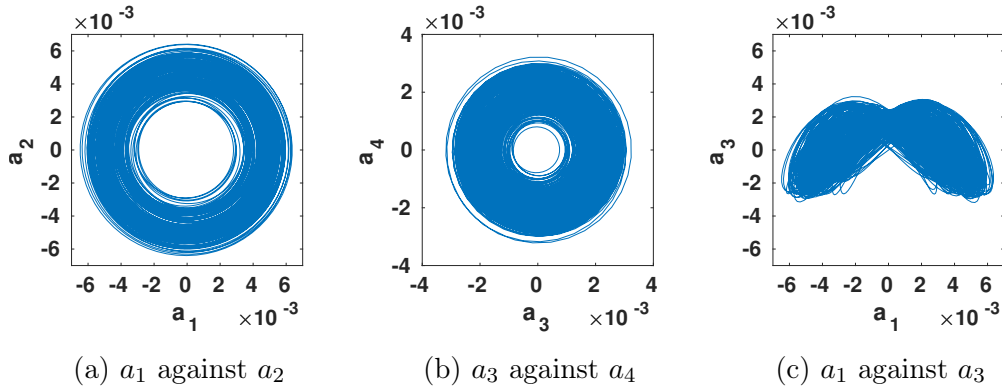


Figure 4.6: Plots of a_1 against a_2 ; a_3 against a_4 ; and a_1 against a_3 .

4.3 Reconstruction of the 1st and 2nd 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 1 and 2, i.e. the output vector in equation (3.11) which consists of the temporal coefficients $\mathbf{a}_e(t) = [a_{1,e}(t), a_{2,e}(t)]^\top$. We start by employing as input the velocity signals from a sin-

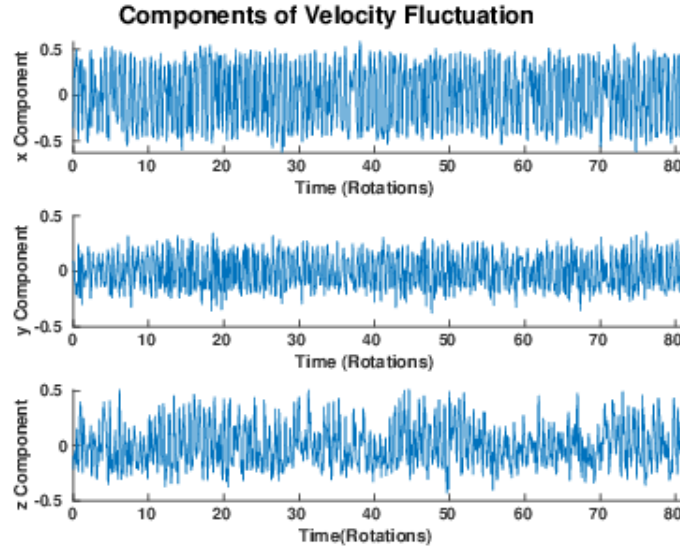


Figure 4.7: Velocity fluctuations at the point of maximum turbulent kinetic energy.

gle probe point, so $\mathbf{s}(t) = [u_1(t), v_1(t), w_1(t)]^\top$, 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, i.e. $\mathbf{s}(t) = [u_1(t), v_1(t), w_1(t), \dots, u_6(t), v_6(t), w_6(t)]^\top$.

4.3.1 Reconstruction Using the Velocity Signals from a Single Sensor Point

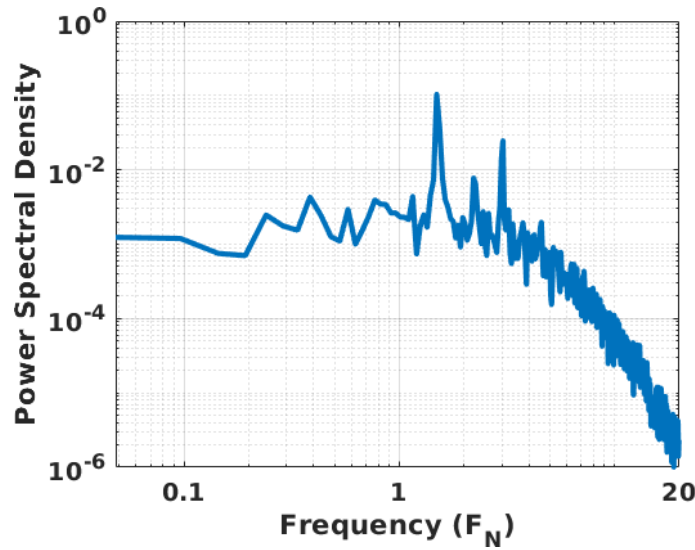


Figure 4.8: Power spectrum of the x velocity component at probe point 1 for $Re=600$.

The location of the probe point is important for the quality of reconstruction [69]. In

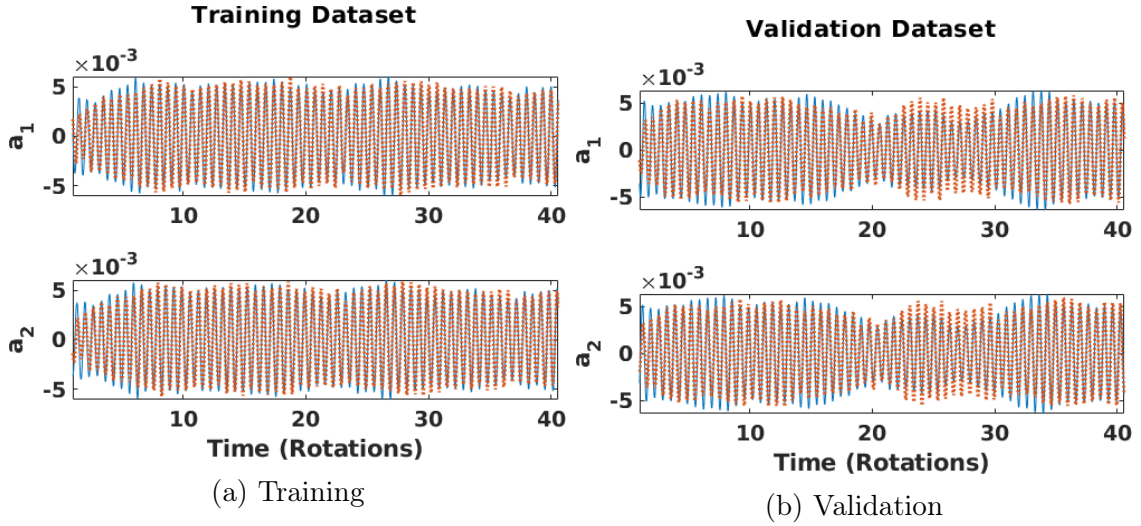


Figure 4.9: System identification using velocity data from probe point 1 (6th order ROM): DNS (blue) and N4SID (orange).

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 4.2. The time signals of the velocity fluctuations for 80 revolutions at this point are shown in Figure 4.7 and the corresponding power spectrum for the x component in Figure 4.8 (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 of the flow at the examined Re number. The spectral peaks at frequencies $1.5F_N$ and $3F_N$ are clearly visible in Figure 4.8.

Velocity data from 40 revolutions was used for training and extraction of the matrices $\mathbf{A}'$, $\mathbf{B}'$, $\mathbf{H}'$, $\mathbf{Q}'$, $\mathbf{S}'$, $\mathbf{R}'$. In Figure 4.9, we plot the variation of the temporal coefficients $a_1(t)$ and $a_2(t)$ as obtained from DNS (grey line) and as predicted by the estimator (using a 6th order model). It can be seen that for the training dataset, Figure 4.9a, 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 4.8, 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 4.9b. For the initialisation of system (3.11) 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 localised 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 (3.11), i.e. the number of states of the vector $\mathbf{X}_e$, on the accuracy of the reconstruction as determined by the FIT[%] value, shown in (3.16). The results are shown in Figure 4.10a for both the training and the validation datasets (in computing the fit for the former we have assumed correct initial model conditions, the MATLAB default for training data, while for the latter, we have excluded the first 500 data points to allow the model to "spin-up" from an initial zero state). 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 8th order model as optimal. This is determined by examining the variation of the singular values, which are drawn with an orange line in Figure 4.10a. It can be seen that the singular values drop by two orders of magnitude between the 8th and 9th order, indicating that orders higher than 8 do not offer a significant advantage. In Figure 4.11 we plot the spatial distribution of the velocity RMS due to the first two modes, $\sqrt{\mathbf{u}'_{1\&2}{}^2}$, 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 4.9, the time-average distributions are almost indistinguishable.

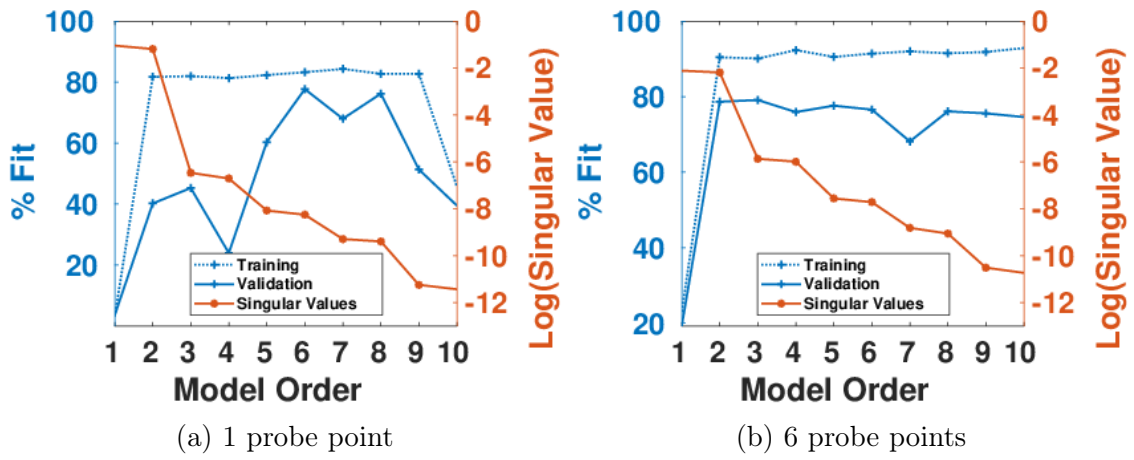


Figure 4.10: Effect of model order on the accuracy of reconstruction for the training and validation datasets using 1 (a) and 6 (b) probe points.

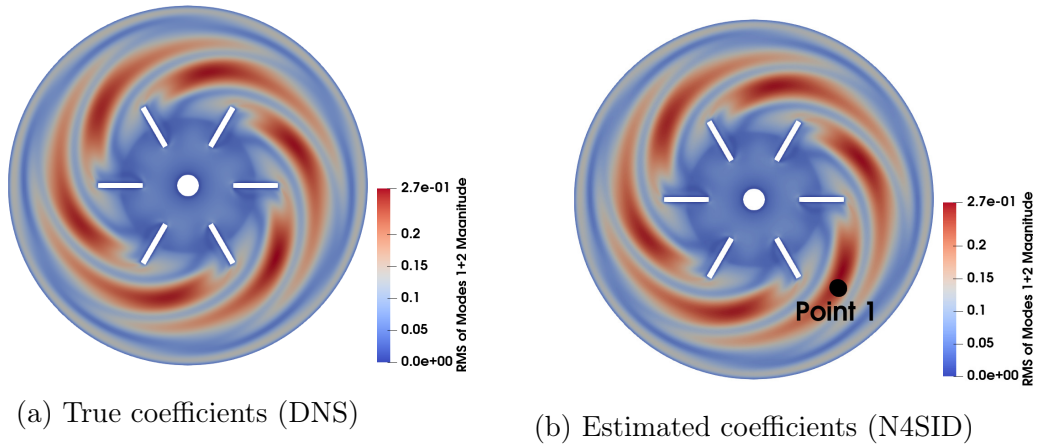


Figure 4.11: Contour plots of $\sqrt{\mathbf{u}'_{1\&2}{}^2}$ computed using (a) the true and (b) the estimated temporal coefficients $a_1(t)$ and $a_2(t)$. The black dot indicates the single sensor used.

4.3.2 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 4.2.

It should be noted that there are multiple ways to approach the placement of the sensor points. An alternative would be a data driven method such as that proposed in the recent paper [122] using the QR decomposition. However, this has not been tested in 3D and would require a large amount of data to be processed to have a representative number of point locations considered. Hence, this was not deemed to be feasible for this study.

The remaining approaches are more heuristic, such as the choice of a symmetrical configuration of sensors chosen here. An alternative could be to choose 6 points in a line, in the path of one vortex, or some combination of the two. It is important to consider the nature of the POD temporal coefficients that we are attempting to reconstruct. Together, these two coefficients represent the 6 vortices that are shed from each of the respective blades at different times. In essence, this implies that they represent an average of the behaviour over the 6 vortices.

For example, if at a particular time, the vortex shed off 1 blade was slightly weaker than usual (hence contributing less to the POD coefficient), this could be compensated by a vortex shedding of another blade being stronger than usual (hence contributing more to the POD coefficient). If the choice was made to place all 6 points in the path of the weaker vortex, then the POD coefficient would be under-

predicted, or overpredicted if they are placed in the path of the stronger vortex. In contrast, by placing a sensor in the path of each vortex, the cancellation of the effects of the variations in the vortices can be replicated in the input data to the model.

As this work is a first attempt at the flow reconstruction in a stirred tank, we place the 6 sensors in a simple symmetrical configuration that can capture the average vortex behaviour. However, more sophisticated arrangements such as placing a point in the path of each vortex, but at different heights and/or radii for each vortex, could also capture this and potentially lead to better reconstruction performance. The investigation of optimal sensor placement is an objective for future work.

The fit for the training dataset is very high for all cases considered. For example, using 6 probes, the fit is almost 90% as shown in Figure 4.10b. 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 4.10a. 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.

4.4 Reconstruction of the 3rd and 4th Modes

The ability of the identification algorithm to estimate the 3rd and 4th order modes was also examined. For the training step, the same velocity signals from the same 6 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 4.4), was predicted very

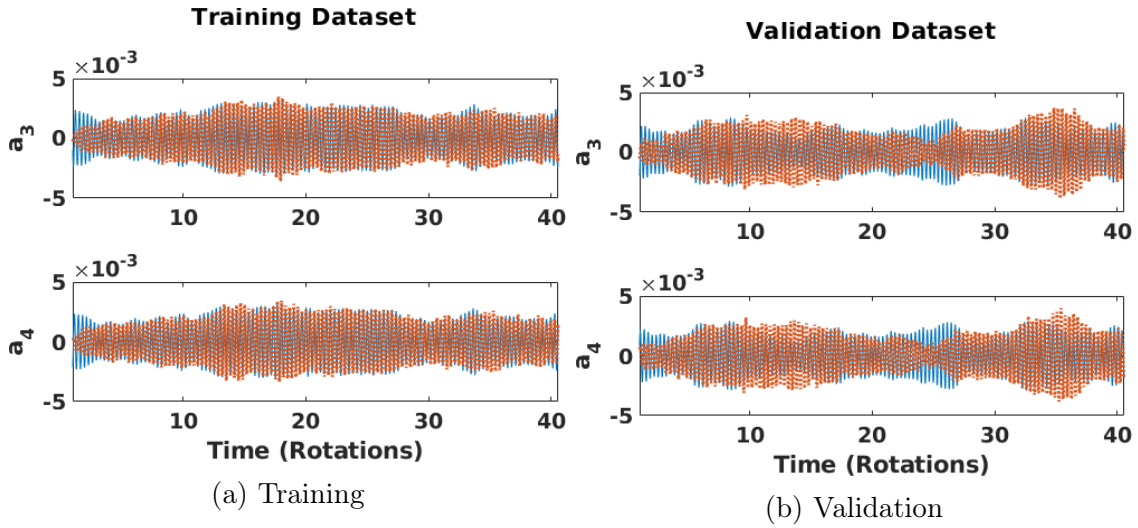


Figure 4.12: Comparison of temporal coefficients for the 3rd and 4th POD modes using an 8th order model with 6 point input: DNS (blue) and N4SID (orange).

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 8th order model are shown in Figure 4.12.

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 Figures 4.2c and 4.2d. Additionally, the energy content of these modes is substantially smaller (about 5 times) than that of the first pair, refer to Figure 4.1a (this can be also noticed by comparing the peaks at frequencies $1.5F_N$ and $3F_N$ in the velocity spectrum plot shown in Figure 4.8).

The velocity RMS due to modes 3 and 4, $\sqrt{\mathbf{u}'_{3\&4}{}^2}$, 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 4.13. 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 4.14 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).

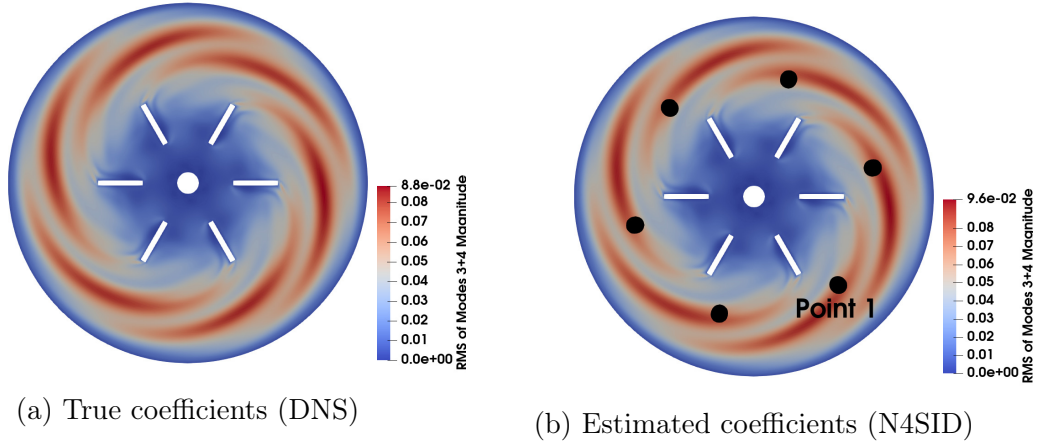


Figure 4.13: Contour plots of $\sqrt{\mathbf{u}'_{3\&4}{}^2}$ computed using (a) the true and (b) the estimated temporal coefficients $a_3(t)$ and $a_4(t)$. The black dots indicate the sensor points used.

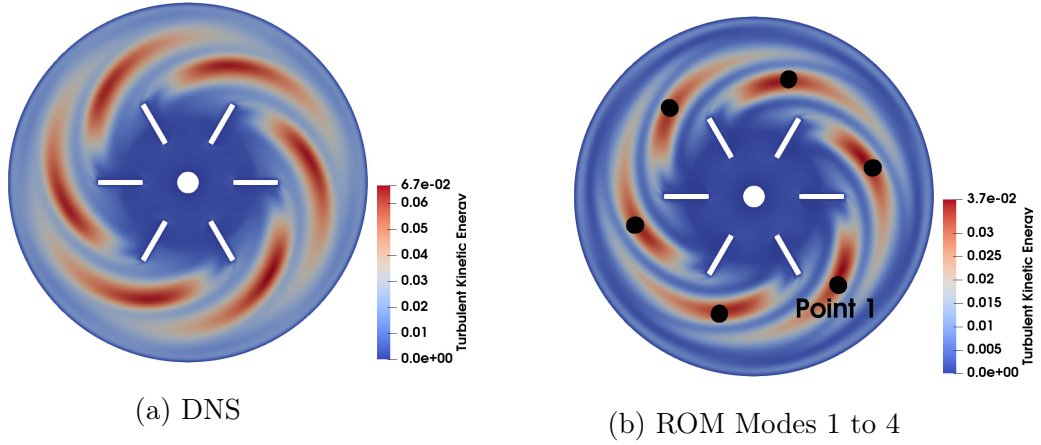


Figure 4.14: Contour plots of turbulent kinetic energy as (a) computed from the DNS simulation and (b) estimated by superposition of modes 1-4 (for modes 1&2 the input is only from sensor point 1, for modes 3&4 from all 6 points).

The plots are qualitatively identical, but quantitatively the estimated peak value is about 55% of the DNS value. This is expected since only 4 modes were used for reconstruction. It is interesting to notice, that although in the full tank these 4 modes account for $\approx 34\%$ of the total kinetic energy (Figure 4.1a), the peak value at the plane considered is captured with significantly better accuracy.

To further examine the performance of the ROM we compare the turbulent kinetic energy obtained from the DNS simulations and the ROM along a line bisecting two blades at a distance of $0.353T$ from the tank bottom (Figure 4.15). As for the

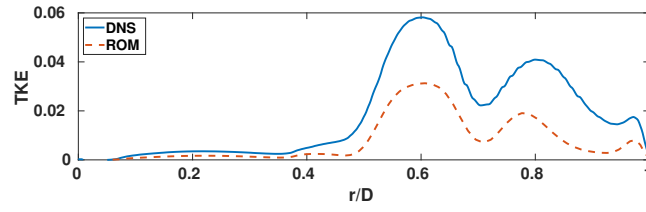
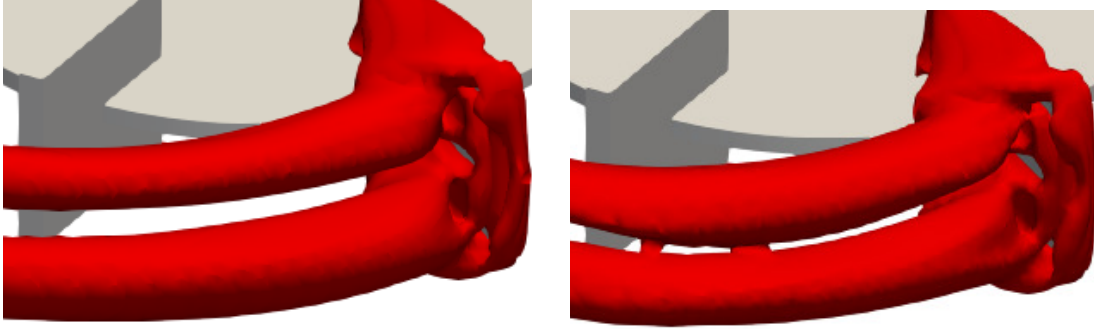


Figure 4.15: Turbulent kinetic energy obtained from the DNS simulations and the ROM along a line bisecting two blades at a distance of $0.353T$ from the tank bottom.



(a) Estimated (modes 1 and 2)

(b) DNS

Figure 4.16: Isosurface of $\lambda_2/\omega^2 < -1.5$ in the vicinity of a blade at 20 rotations into validation dataset.

contour plots, the ROM reconstructs the trends in the TKE well, but underpredicts compared to the DNS.

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 4.4), thereby representing random, rather than organised, motions. The contribution of the white noise terms $\mathbf{w}(t_j)$ and $\mathbf{v}(t_j)$ in model (3.12) becomes dominant, preventing meaningful estimation.

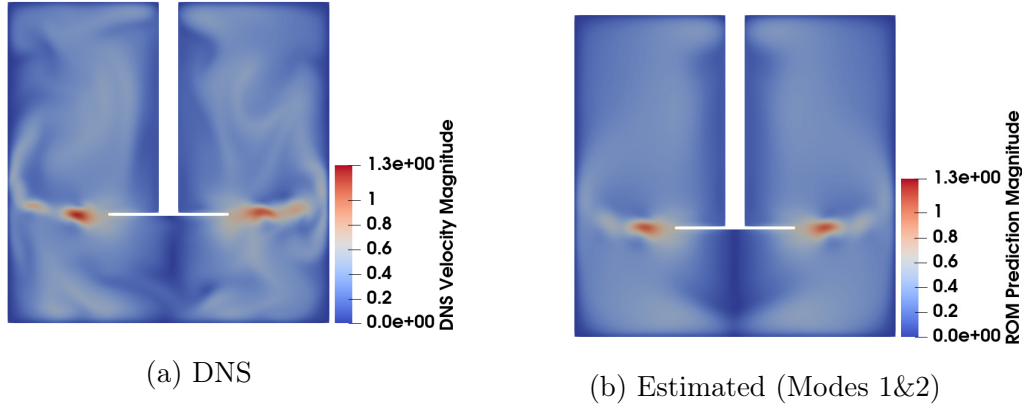


Figure 4.17: True (DNS) and estimated velocity magnitudes at 20 rotations into validation dataset.

4.5 Towards 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 $\mathbf{u}_e(\mathbf{x}, t) \approx \bar{u}(\mathbf{x}) + a_{1,e}(t)\phi_1(\mathbf{x}) + a_{2,e}(t)\phi_2(\mathbf{x})$. 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 4.1b.

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 data set. 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 [123]. Isosurface plots of negative λ_2 (that visualise vortex structures) are shown in Figure 4.16. A cutoff of -1.5 is used to more easily identify the vortex structures. 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 4.10a), 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 4.17. 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 modes 1 and 2 are predominantly present around the impeller and have also a smooth but weak footprint in the rest of the vessel (Figures 4.3a and 4.3b). For this reason, the estimated velocity field in Figure 4.17b 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).

4.6 Performance at Off-Design Conditions

The robustness and accuracy of reconstruction were 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 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, i.e. $Re=600$. Finally, we employed the model described by (3.11), 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.

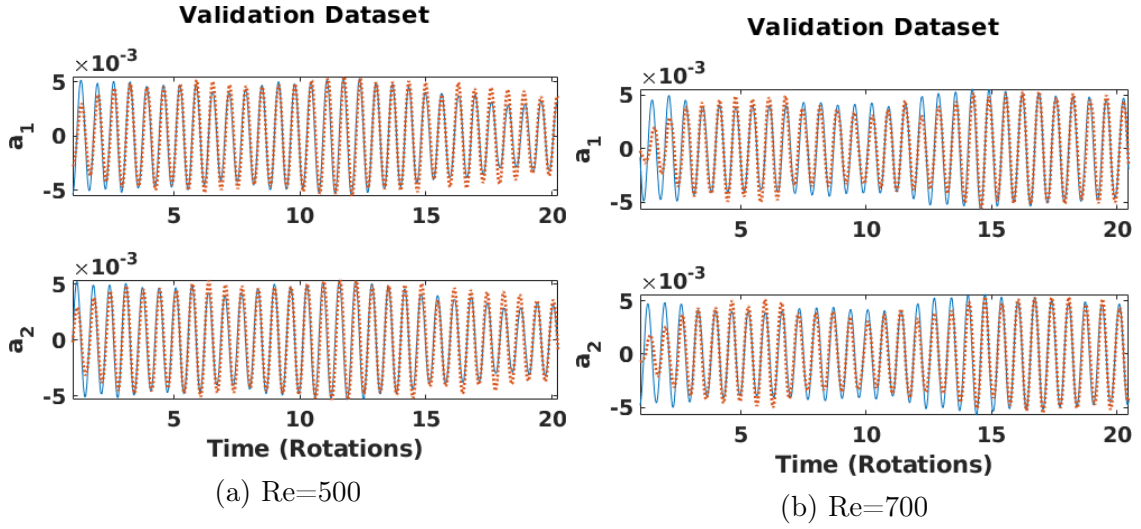


Figure 4.18: Temporal coefficients for modes 1 and 2 at two off-design conditions using a single probe point and a 6th order model; DNS (blue) and N4SID (orange). The temporal coefficients are normalised by the respective V_{tip} for each case.

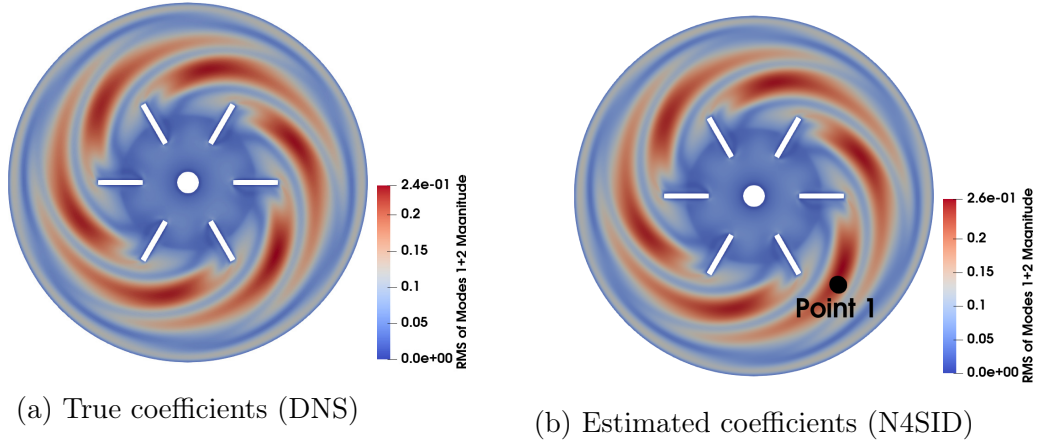


Figure 4.19: Contour plots of $\sqrt{\mathbf{u}'_{1+2}{}^2}$ computed using (a) the true and (b) the estimated temporal coefficients $a_1(t)$ and $a_2(t)$ for the off-design condition of Re=500. The black dot indicates the single sensor point used.

The results using a single probe point and a 6th order model are shown in Figure 4.18. The fits for modes (1,2) are (75%, 74%) for Re=500 and (77%, 77%) 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 3rd order model with improved fits, equal to (83%, 83%) and (84, 83%) for modes 1&2 for Re=500 and 700 respectively.

Contour plots of the velocity RMS due to modes 1 and 2, $\sqrt{\mathbf{u}'_{1\&2}{}^2}$, 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 4.19. 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 4.11a and 4.19a. 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).

4.7 Conclusions

In the present chapter, 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 N4SID identification algorithm was then used to construct a data-driven estimator, i.e. 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 6-point input predicted well the frequency of the mode, and yielded a satisfactory reproduction of the time signal 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 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? The first two questions will be addressed in the following chapters and the rest are left as topics for future research.

Chapter 5

Decomposition of Power Number and Reconstruction of Flow Structures from Surface Pressure Measurements

5.1 Introduction

This chapter builds on the work of the preceding chapter and considers the same geometry and Reynolds numbers. The POD of the velocity field is now used to obtain pressure modes. The latter are then employed to decompose the power number (total and per blade) to identify periodic features. The identification algorithm N4SID is again used to construct an estimator that captures the relation between pressure signals at sensor points on the blade surface or the outer wall (input) and the velocity POD coefficients (output). Finally, the estimator derived at $Re=600$ is applied to the flows at off-design $Re=500$ and 700 .

The aim of this chapter is to answer the following open questions. How can the modes identified in the preceding chapter be used to decompose the pressure field and power number? Using this decomposition can the periodic fluctuations in the

power number (total and per blade) be identified and their origin explained? Can the N4SID reconstruction approach be adapted to use non-intrusive pressure measurements rather than intrusive velocity measurements?

5.2 Methodology

5.2.1 Governing Equations for the Pressure Field

As mentioned in Chapter 2, the flow velocity field in a rotating frame is described by

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} - [\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{x}) + 2\boldsymbol{\omega} \times \mathbf{u}], \quad (5.1a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (5.1b)$$

where the last two terms in the momentum equation are the Centrifugal and Coriolis forces respectively. The centrifugal and pressure terms can be combined to form the modified pressure $\hat{P} = p - \frac{1}{2}\rho((\boldsymbol{\omega} \times \mathbf{x})^2)$ [124]

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla \hat{P} + \nu \nabla^2 \mathbf{u} - 2\boldsymbol{\omega} \times \mathbf{u}. \quad (5.2)$$

Taking the divergence of Equation 5.2 yields

$$\frac{\partial(\nabla \cdot \mathbf{u})}{\partial t} + \nabla \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) = -\frac{\nabla^2 \hat{P}}{\rho} + \nu \nabla \cdot (\nabla^2 \mathbf{u}) - \nabla \cdot (2\boldsymbol{\omega} \times \mathbf{u}). \quad (5.3)$$

Setting the first and fourth terms to 0 (due to the continuity equation) results in the so-called simplified pressure Poisson equation (SPPE) [125]. However, it has been observed that, for numerical reasons, it is advantageous to retain the viscous term [125], [126]. This simplifies the equation to the so-called consistent pressure

Poisson equation (CPPE)

$$\nabla \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) = -\frac{\nabla^2 \hat{P}}{\rho} - \nabla \cdot (2\boldsymbol{\omega} \times \mathbf{u}) + \nu \nabla \cdot \nabla^2 \mathbf{u} \quad (5.4)$$

and rearranging gives

$$\nabla^2 \hat{P} = -\rho \nabla \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) - \rho \nabla \cdot (2\boldsymbol{\omega} \times \mathbf{u}) + \mu \nabla \cdot \nabla^2 \mathbf{u}. \quad (5.5)$$

Boundary Conditions

Projecting Equation 5.2 onto a direction of the unit vector normal to the wall, $\mathbf{n}$, gives

$$\frac{\partial u_n}{\partial t} + \mathbf{n} \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) = -\frac{1}{\rho} \frac{\partial \hat{P}}{\partial n} + \nu \frac{\partial^2 u_n}{\partial n^2} - 2\mathbf{n} \cdot (\boldsymbol{\omega} \times \mathbf{u}). \quad (5.6)$$

where $u_n = \mathbf{u} \cdot \mathbf{n}$. As the wall normal velocity is zero, the first term vanishes, giving the boundary condition

$$\frac{\partial \hat{P}}{\partial n} = -\rho \mathbf{n} \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) + \mu \frac{\partial^2 u_n}{\partial n^2} - 2\rho \mathbf{n} \cdot (\boldsymbol{\omega} \times \mathbf{u}). \quad (5.7)$$

In the case of stationary walls, this simplifies to

$$\frac{\partial \hat{P}}{\partial n} = \mu \frac{\partial^2 u_n}{\partial n^2}. \quad (5.8)$$

In the case of the outer walls, $\mathbf{n}$ is in the radial direction, $\boldsymbol{\omega}$ is the axial direction and $\mathbf{u}$ is in the tangential direction, hence the Coriolis term simplifies to $-2R\omega^2$, where R is the radius of the wall. In this case, the boundary condition becomes

$$\frac{\partial \hat{P}}{\partial n} = -\rho \mathbf{n} \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) + \mu \frac{\partial^2 u_n}{\partial n^2} - 2\rho R\omega^2. \quad (5.9)$$

The component of the advection term normal to the wall in this case is in the radial direction. In cylindrical coordinates, this is given by

$$-\rho \mathbf{n} \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) = -\rho(u_r \frac{\partial u_r}{\partial r} + \frac{u_\theta}{r} \frac{\partial u_r}{\partial \theta} + u_z \frac{\partial u_r}{\partial z} - \frac{u_\theta^2}{r}). \quad (5.10)$$

The first three terms on the right hand side are zero, as the radial and axial velocities at the wall are zero. Finally, $u_\theta = -R\omega$. This means that Equation 5.10 reduces to

$$-\rho \mathbf{n} \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) = \rho R\omega^2. \quad (5.11)$$

Hence the boundary condition at the outer walls is given by

$$\frac{\partial \hat{P}}{\partial n} = \mu \frac{\partial^2 u_n}{\partial n^2} - \rho R\omega^2. \quad (5.12)$$

For the top and bottom walls $\mathbf{n}$ is in the same direction as $\boldsymbol{\omega}$, hence the Coriolis term is zero. The boundary condition then becomes

$$\frac{\partial \hat{P}}{\partial n} = -\rho \mathbf{n} \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) + \mu \frac{\partial^2 u_n}{\partial n^2}. \quad (5.13)$$

In this case $\mathbf{n}$ is along the positive axial direction for the top wall and along the negative axial direction for the bottom wall. The component of the advection term in the axial direction in cylindrical coordinates is given by

$$\rho \mathbf{n} \cdot (\mathbf{u} \cdot \nabla \mathbf{u}) = \rho(u_r \frac{\partial u_z}{\partial r} + \frac{u_\theta}{r} \frac{\partial u_z}{\partial \theta} + u_z \frac{\partial u_z}{\partial z}) = 0 \quad (5.14)$$

as all three components of velocity are equal to zero at the wall. Hence, the boundary condition reduces to that of stationary walls

$$\frac{\partial \hat{P}}{\partial n} = \mu \frac{\partial^2 u_n}{\partial n^2}. \quad (5.15)$$

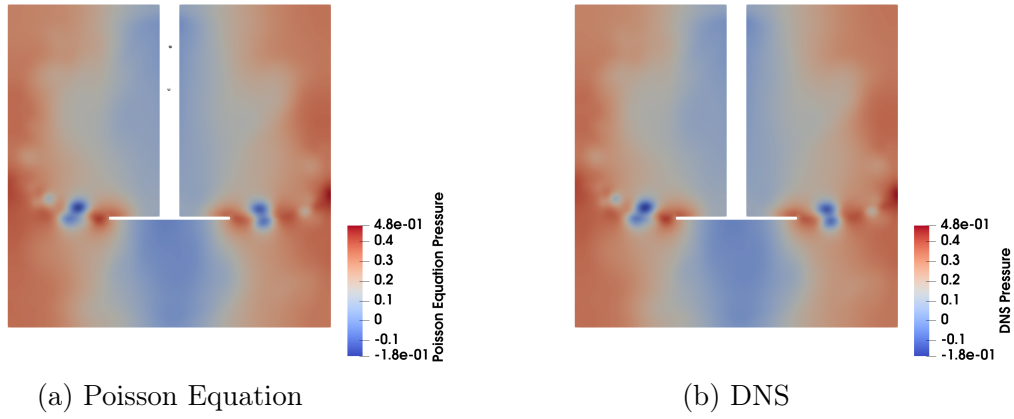


Figure 5.1: Contours of instantaneous pressure obtained directly from DNS and from solving the Poisson equation on a vertical plane midway between two blades.

Equation 5.5 and the above boundary conditions can then be used to obtain $\hat{P}$ relative to a reference pressure. In this study, we set the zero pressure to be at a point located at the centre at a height $0.005T$ from the tank bottom. The Poisson equation is solved numerically using an adapted version of the Poisson equation solver of the "Pantarhei" code used for the main simulations. The pressure is normalised by $0.5\rho v_{tip}^2$ throughout, following [127]. The instantaneous pressure field obtained with this method is identical to the one obtained by DNS as shown in Figure 5.1.

5.3 Proper Orthogonal Decomposition

5.3.1 Velocity POD Modes

In the previous chapter, using data from 120 rotations, it was found that the first two modes were paired with a contribution of 14% each to the turbulent kinetic energy, with a frequency of $1.5F_N$, where F_N is the impeller rotation frequency. The third and fourth modes were also found to be paired, with smaller contributions of 2.6% each to the turbulent kinetic energy and a frequency of $3.0F_N$. Both pairs of modes were found to correspond to the trailing vortices in the wake of the blades and were primarily concentrated in the areas close to the impeller. In this chapter

data from 210 rotations comprising of the original 120 rotations and a further 90 rotations is used. Near identical results were found, with the first two pairs of modes having frequencies of $1.5F_N$ and $3.0F_N$ respectively and 14% and 2.6% per mode of TKE respectively. It should be noted that due to the 60° symmetry, the paired POD modes obtained can be rotated by 60° relative to those obtained with different length datasets, despite otherwise being identical. As such, the results for contours of the partial pressure on the vertical slices between blades presented below may not directly correspond to the velocity fields presented for these modes in the preceding chapter.

5.3.2 Governing Equations for the Partial Pressures

Substituting the POD expansion into Equation 5.5 and summing over i and j , where both indices start from zero and hence include the mean flow, gives

$$\nabla^2 \hat{P} = - \sum_{i=0}^{i=\infty} \sum_{j=0}^{j=\infty} \rho a_i a_j \nabla \cdot (\phi_i \cdot \nabla (\phi_j)) - \sum_{i=0}^{i=\infty} \rho a_i \nabla \cdot (2\boldsymbol{\omega} \times \phi_i) + \sum_{i=0}^{i=\infty} \mu a_i \nabla \cdot \nabla^2 \phi_i, \quad (5.16)$$

or writing the mean, linear and quadratic terms explicitly

$$\begin{aligned} \nabla^2 \hat{P} = & -\rho \nabla \cdot (\bar{\mathbf{u}} \cdot \nabla \bar{\mathbf{u}}) + \mu \nabla^2 \bar{\mathbf{u}} \\ & - \sum_{i=1}^{i=\infty} a_i [\rho \nabla \cdot (\bar{\mathbf{u}} \cdot \nabla (\phi_i)) - \rho \nabla \cdot (\phi_i \cdot \nabla (\bar{\mathbf{u}})) - \rho \nabla \cdot (2\boldsymbol{\omega} \times \bar{\mathbf{u}}) - \rho \nabla \cdot (2\boldsymbol{\omega} \times \phi_i) + \mu \nabla \cdot \nabla^2 \phi_i] \\ & - \sum_{i=1}^{i=\infty} \sum_{j=1}^{j=\infty} \rho a_i a_j \nabla \cdot (\phi_i \cdot \nabla (\phi_j)). \end{aligned} \quad (5.17)$$

where the first line on the right hand side represents the time averaged terms, the second line the terms linear in a_i and the final line the terms quadratic in a_i . The

partial pressures p_{ij} are defined as follows

$$\begin{aligned} P(\mathbf{x}, t) &= \sum_{i=0}^{i=\infty} \sum_{j=0}^{j=\infty} P_{ij}(\mathbf{x}) a_i(t) a_j(t) \\ &= P_{00} + 2.0 \sum_{i=1}^{i=\infty} P_{i0}(\mathbf{x}) a_i(t) + \sum_{i=1}^{i=\infty} \sum_{j=1}^{j=\infty} P_{ij}(\mathbf{x}) a_i(t) a_j(t) \end{aligned} \quad (5.18)$$

and related to the modified pressure as

$$\hat{P}(\mathbf{x}, t) = \sum_{i=0}^{i=\infty} \sum_{j=0}^{j=\infty} \hat{P}_{ij}(\mathbf{x}) a_i(t) a_j(t), \quad (5.19)$$

where $P_{ij} = \hat{P}_{ij}$ for all terms apart from the mean term ($i = j = 0$) which includes the centrifugal term

$$\hat{P}_{00} = P_{00} - \frac{1}{2} \rho (\boldsymbol{\omega} \times \mathbf{x}^2). \quad (5.20)$$

The time averaged pressure component satisfies

$$\nabla^2 \hat{P}_{00} = \rho \nabla \cdot (\boldsymbol{\phi}_0 \cdot \nabla(\boldsymbol{\phi}_0)) - \rho \nabla \cdot (\boldsymbol{\omega} \times \boldsymbol{\phi}_0) + \mu \nabla \cdot \nabla^2 \boldsymbol{\phi}_0. \quad (5.21)$$

The linear pressure terms satisfy

$$\begin{aligned} \nabla^2 P_{i0} = \nabla^2 P_{0i} &= \rho \nabla \cdot (\boldsymbol{\phi}_0 \cdot \nabla(\boldsymbol{\phi}_i)) - \rho \nabla \cdot (\boldsymbol{\omega} \times \boldsymbol{\phi}_i) \\ &= \rho \nabla \cdot (\boldsymbol{\phi}_i \cdot \nabla(\boldsymbol{\phi}_0)) - \rho \nabla \cdot (\boldsymbol{\omega} \times \boldsymbol{\phi}_i) + \frac{1}{2} \mu \nabla \cdot \nabla^2 \boldsymbol{\phi}_i. \end{aligned} \quad (5.22)$$

Finally, the quadratic pressure terms satisfy

$$\nabla^2 P_{ij} = \nabla^2 P_{ji} = \rho \nabla \cdot (\boldsymbol{\phi}_i \cdot \nabla(\boldsymbol{\phi}_j)) \quad \text{where } i, j \neq 0. \quad (5.23)$$

5.3.3 Boundary Conditions

For stationary walls and the top and bottom moving walls the boundary conditions for the total pressure are

$$\frac{\partial \hat{P}}{\partial n} = \mu \frac{\partial^2 u_n}{\partial n^2}. \quad (5.24)$$

Substituting in the pressure and velocity expansions we obtain

$$\sum_{i=0}^{i=\infty} \sum_{j=0}^{j=\infty} a_i a_j \frac{\partial \hat{P}_{ij}}{\partial n} = \sum_{i=0}^{i=\infty} a_i \mu \frac{\partial^2 \phi_{i,n}}{\partial n^2}, \quad (5.25)$$

where $\phi_{i,n}$ denotes the projection of ϕ_i on the wall normal.

Taking the time average, the boundary condition for the mean is given by

$$\frac{\partial \hat{P}_{00}}{\partial n} = \mu \frac{\partial^2 \phi_{0,n}}{\partial n^2}. \quad (5.26)$$

As the right hand side is linearly dependent on a_i for the relation to hold at all times, the boundary condition for the quadratic terms must be

$$\frac{\partial P_{ij}}{\partial n} = 0, \quad \text{where } i, j > 0. \quad (5.27)$$

Finally, as $p_{0i} = p_{i0}$, the boundary conditions for the linear terms are given by

$$\frac{\partial P_{0i}}{\partial n} = \frac{\partial P_{i0}}{\partial n} = \frac{1}{2} \mu \frac{\partial^2 \phi_{i,n}}{\partial n^2}, \quad \text{where } i > 0. \quad (5.28)$$

For the outer moving walls, the boundary conditions are

$$\frac{\partial P}{\partial n} = \mu \frac{\partial^2 u_n}{\partial n^2} - \rho R \omega^2. \quad (5.29)$$

The boundary conditions for the linear and quadratic terms are as before, but the

mean term is modified to incorporate the effects of the moving wall

$$\frac{\partial \hat{P}_{00}}{\partial n} = \mu \frac{\partial^2 \phi_{0,n}}{\partial n^2} - \rho R \omega^2. \quad (5.30)$$

These boundary conditions match those in the literature [90] for $\omega = 0$.

5.4 Construction of a Data-Driven Linear Flow Estimator from Pressure Measurements

We now seek to construct a system that estimates the temporal coefficients, $a_i(t)$, and thus the whole velocity field from sparse pressure measurements. While the pressure has terms that are both linear and quadratic in $a(t)$, the linear terms have been found to be dominant throughout the domain in previous studies [90]. The investigation of the partial pressures for our case presented below reveals a similar behaviour, in particular in the vicinity of the blades. In addition, as only pressure readings from particular locations are used, these can be chosen such that the linear terms are dominant in these locations.

Thus, we seek an estimate, $a_e(t)$, of the state vector, $a(t)$, from a set of pressure measurements $\mathbf{s}(t) = (p_1(t), \dots, p_p(t))^T$ at p sensor points located at $\mathbf{x}_1, \dots, \mathbf{x}_p$ within the stirred tank. From 5.18 we can write $\mathbf{s}(t) = \mathbf{S}_p \mathbf{a}(t) + \mathbf{g}(t)$, where the matrix $\mathbf{S}_p$ depends on the sensor locations and $\mathbf{g}(t)$ represents the noise due to considering a finite number of POD modes and all of the pressure terms quadratic in $a(t)$. This is the same form as for the case of velocity input data and the derivation proceeds in an identical fashion to that described in Chapter 2. The resulting system has the form

$$\mathbf{X}(t_{j+1}) = \mathbf{A}' \mathbf{X}(t_j) + \mathbf{B}' \mathbf{s}(t_j) + \mathbf{w}(t_j), \quad (5.31a)$$

$$\mathbf{a}(t_{j+1}) = \mathbf{H}' \mathbf{X}(t_j) + \mathbf{v}(t_j), \quad (5.31b)$$

where $\mathbf{X}'(t_j)$ is the estimator state at a time t_j . The input $\mathbf{s}(t_j)$ is taken to be pressure readings at selected sensor points on the blades and the output $\mathbf{a}(t_j)$ are the temporal coefficients of the first few POD modes corresponding to the large-scale structures. The matrices $\mathbf{A}'$, $\mathbf{B}'$ and $\mathbf{H}'$ are unknown and are computed using the N4SID algorithm [93]. $\mathbf{v}(t_j)$ and $\mathbf{w}(t_j)$ represent white noise terms.

The data collection, training and testing of the model is performed using a similar procedure to that for the velocity probe case. The training dataset is taken to be approximately 100 rotations long and the validation dataset is approximately 110 rotations long for the base case. The validation datasets for the off-design conditions are taken to be approximately 20 rotations long.

5.5 Power Number Decomposition

The power number in a stirred tank is defined as $\bar{N}_p = \frac{\overline{Po}}{\rho N^3 D^5}$, where Po is the power provided to the flow field by the rotating impeller and the overbar denotes the time average. Equivalently, $\bar{N}_P$ is the total energy dissipation rate integrated over the tank volume. The instantaneous power number is defined as $N_p(t) = \frac{Po(t)}{\rho N^3 D^5}$ and can be computed by integrating the pressure on the blade surface multiplied by the rotational velocity of the impeller.

The time signal of the instantaneous power number computed from this method, as well as by the volume integral of dissipation rate, is shown in Figure 5.2a. The agreement between the two methods is very good with $\bar{N}_p = 2.1$ for both and the same general trends in the time series. The power spectrum of the $N_p(t)$ is shown in Figure 5.2b. We split the signal into several windows with a 50% overlap, compute the power spectra and average. The results for three window lengths give similar results. Hence, to obtain the highest possible resolution, we use a window size of approximately 82 impeller rotations for all subsequent power spectra in this chapter.

Interestingly, the frequencies of $1.5F_N$ and $3.0F_N$ corresponding to the two leading pairs of modes are not present. In addition, the spectrum contains peaks at many low frequencies and in particular at $\sim 0.04F_N$ and $\sim 0.43F_N$. Similar low frequency oscillations in the power number were observed in [128] for the case of a Rushton turbine of diameter $T/3$ and clearance $0.15T$ in a baffled tank. In the dataset of 210 rotations, the period corresponding to the slowest frequency of $\sim 0.04F_N$ is repeated approximately 8 times. We will examine the absence of the $1.5F_N$ and $3.0F_N$ frequencies and the origin of the small frequencies below. The power number

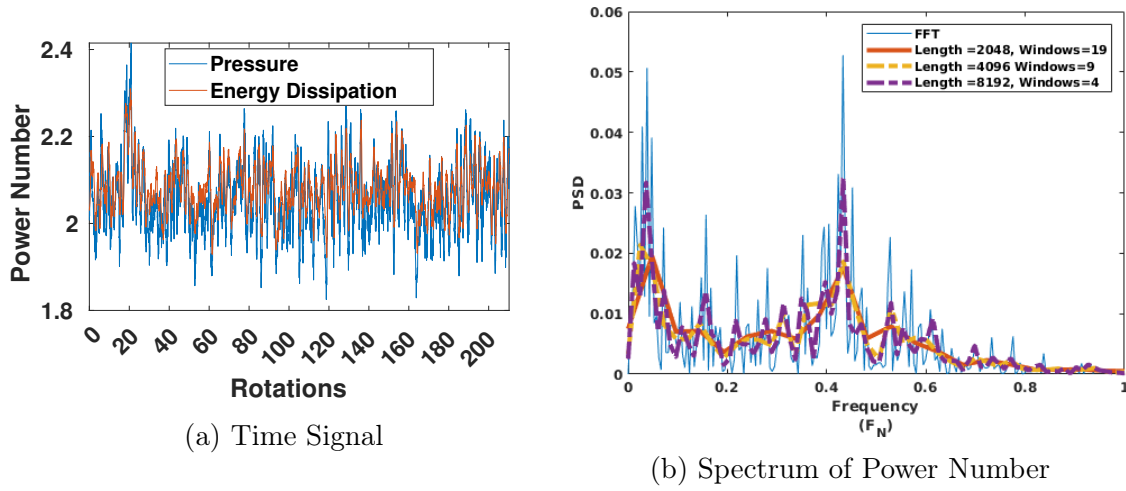


Figure 5.2: Power Number for $Re=600$

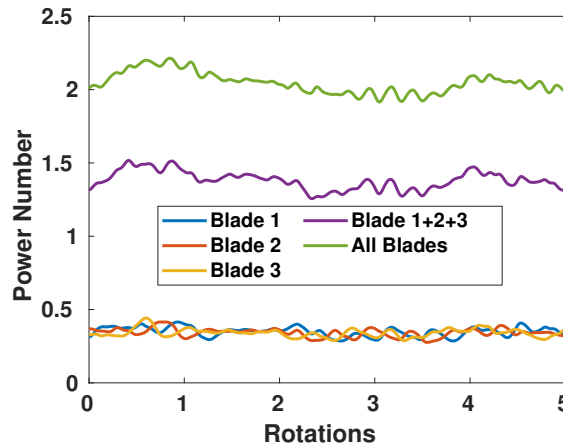


Figure 5.3: Power number signal for individual blades and blade combinations.

signal for individual blades and blade contributions for 5 rotations is shown in Figure 5.3. It can be seen that, as suggested by the power spectra, the time signal for the combined power number from the three blades has fewer periodic features compared

to the time signals of the individual blades.

The power number spectra computed from individual blades and from combinations of blades are shown in Figure 5.4. The results for the 6 individual blades (top row) are very similar as expected but show small differences in the peak magnitudes.

In the bottom row, we average the results for individual blades and blade combinations. For a single blade, there is strong spectral content at frequencies $1.5F_N$, $3.0F_N$, $2.2F_N$, as well as at the earlier mentioned low frequencies. The $1.5F_N$, $3.0F_N$ and $2.2F_N$, match the frequencies of the first, second and fourth POD mode pairs respectively. It can be seen that the frequencies of $1.5F_N$ and $3.0F_N$ are present in all spectra from pairs of blades. On the other hand, the $2.2F_N$ frequency vanishes when considering adjacent or blades separated by 180° , but increases to become stronger than the $1.5F_N$ and $3.0F_N$ signals when the separation angle is 120° . When considering three adjacent blades, both the $3.0F_N$ and $1.5F_N$ signal vanish, while the $2.2F_N$ remains but is relatively weak. Finally, when considering all blades, all frequencies greater than $1.0F_N$ vanish.

A similar cancellation of frequencies corresponding to trailing vortices was observed in [24], for a 4 bladed impeller. More specifically, the $1.70F_N$ frequency observed for individual blades cancelled out in the power number due to the pressure distribution in adjacent blades being exactly 180° out of phase. Notably, when breaking the symmetry of the blades by considering fractal blades, the authors found that this frequency was restored.

The cancellation of the dominant frequencies of $1.5F_N$ and $3.0F_N$ in the power number fluctuations indicates that it is not possible to recover the evolution of the large scales by measuring just the power number fluctuation in this study. However, in the case of a different geometry and in particular, one in which the symmetry is (even slightly) broken, the approach that we demonstrate below using point pressure data could potentially also be applied using power number fluctuation data.

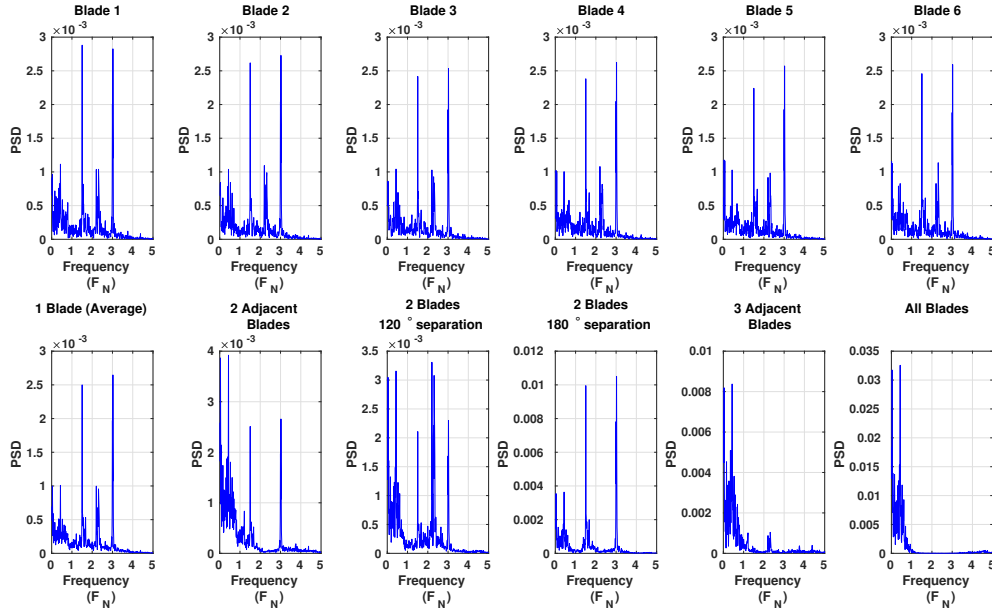


Figure 5.4: Spectrum of power number for individual blades and blade combinations.

5.6 Characterisation of Partial Pressures and Power Number Decomposition

Due to the presence of quadratic terms, the mean pressure has contributions from not just the mean flow mode P_{00} , but also from all of the diagonal partial pressure modes P_{ii} . Hence, the mean power

$$\begin{aligned} \bar{N}_{pi} &= \sum_{i=0}^{i=\infty} \frac{\overline{P_{o_i}}}{\rho N^3 D^3} = - \sum_{i=0}^{i=\infty} \frac{\overline{a_i a_i}}{\rho N^3 D^3} \iint_{\text{impeller}} P_{ii} \vec{u} \cdot \vec{ds} \\ &= - \sum_{i=0}^{i=\infty} \frac{\lambda_i}{\rho N^3 D^3} \iint_{\text{impeller}} P_{ii} r \omega \hat{n} \cdot \vec{ds} \quad (5.32) \end{aligned}$$

where λ_i is the i -th eigenvalue of the POD mode, corresponding to twice the time averaged kinetic energy per unit mass and $\hat{n}$ is the unit vector perpendicular to the blade surface. The partial power numbers and their cumulative percentage of the total power number are shown in Figure 5.5. The mean partial pressure is found to account for 97% of the total power number, the first pair of POD modes 0.6%, the second pair of POD modes 0.06% and other modes $< 0.03\%$. The first

300 modes contain 99.7% of the total power number. The contribution of the 0-th

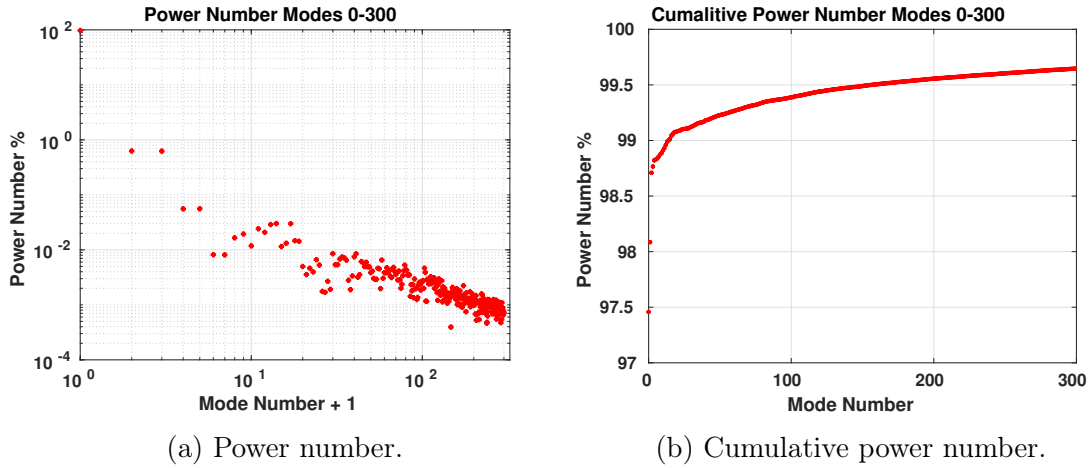


Figure 5.5: Power number due to diagonal partial pressures.

mode (the mean velocity) to the power number is larger compared to that for the kinetic energy (86% for the kinetic energy). This is expected as the power number considers the pressure differences between the two sides of the blades. Considering the higher order modes that represent homogeneous turbulence, we would expect the contributions to the average power number to be negligible, as we would expect the time averaged pressure distribution on both sides of the blades to be approximately equal. Similarly, the large scale structures represented by the leading modes could both increase the pressure difference across one blade, while decreasing it on another. This would result in a small contribution to the time averaged power number for the whole impeller.

Contour plots of the mean pressure and the partial pressures on a vertical plane bisecting the blades are shown in Figure 5.6. The mean pressure is primarily contained in the P_{00} . In particular, in regions away from the impeller the P_{00} term almost exactly matches the mean pressure. However, in areas close to the impeller height there are noticeable differences in the regions where the other partial pressure modes are large in magnitude. This agrees with the results of the previous chapter where the 4 leading order modes have a substantial presence in the vicinity of the impeller height and a smaller footprint elsewhere. The contours of $\lambda_1 P_{11}$ and $\lambda_1 P_{22}$

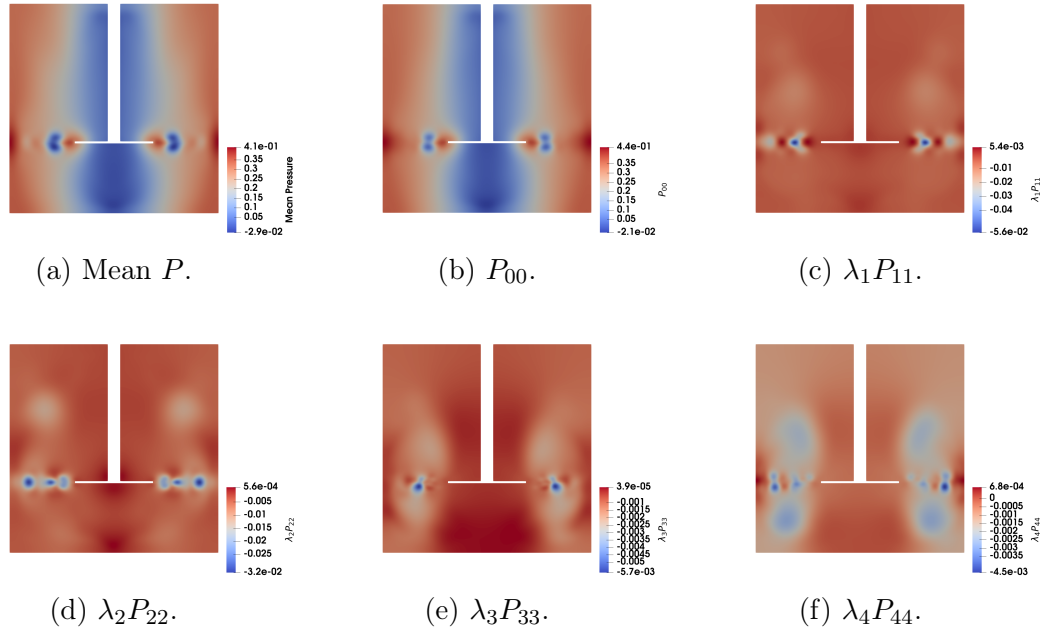


Figure 5.6: Contour plots of mean and diagonal partial pressure terms on a vertical section bisecting the blades.

are symmetric about the blade disk height as was the case for the magnitude of the first two velocity POD modes, but this symmetry is broken for the third and fourth modes in agreement with the observation for the velocity magnitudes of these modes in the previous chapter.

Contour plots of the mean pressure and the partial pressures on a horizontal plane $0.353T$ above the tank bottom (plane of maximum RMS velocity) are shown in Figure 5.7. As for the vertical plane, the largest contribution to the mean pressure comes from the P_{00} term, but with noticeable contributions from other modes as the trailing vortices get further from the blades. The $\lambda_1 P_{11}$, $\lambda_1 P_{22}$, $\lambda_1 P_{33}$ and $\lambda_1 P_{44}$ have a small difference in the pressure values between the pressure and suction sides of the blades, compared to the P_{00} term, which is in agreement with the earlier finding that the largest contribution of the power number comes from the P_{00} term. The 180° symmetry in the terms corresponding to modes 1 and 2 observed for the velocity magnitudes in the previous chapter is again seen. The length scale of the structures corresponding to the first pair of modes is also again observed to be twice that for those corresponding to the second pair.

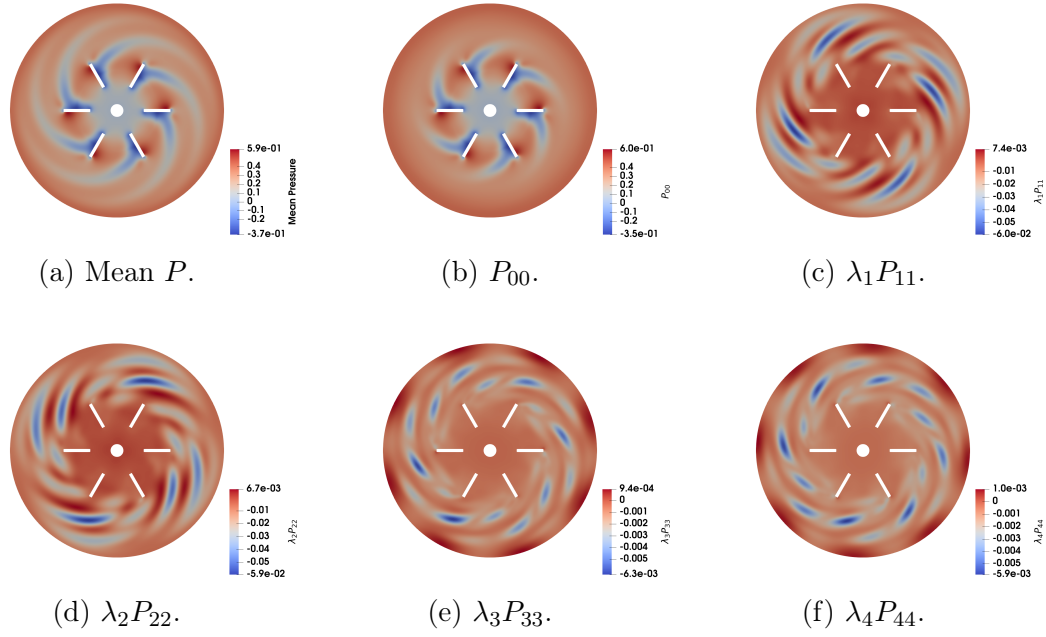


Figure 5.7: Contour plots of the mean and diagonal partial pressure terms on horizontal section 0.353T from the tank bottom.

Contours of the linear (with respect to a_i) pressure terms contributions to the pressure for a vertical plane bisecting two blades and a horizontal plane 0.353T from the tank bottom are shown in Figures 5.8 and 5.9 respectively. It should be noted that the full contribution to the pressure from the linear term for each POD mode is in fact twice this as $P_{i0} = P_{0i}$. In the vertical plane, the $\sqrt{\lambda_1}P_{01}$ and $\sqrt{\lambda_2}P_{02}$ are antisymmetric about the impeller disk (or symmetric if considering the RMS). Similarly to the diagonal terms, the second pair of modes do not have this symmetry. In the horizontal plane, the structure sizes for the terms for the first two modes again appear to be twice the length of those corresponding to the second pair of modes.

5.7 Analysis of Power Number Spectrum Using Partial Pressure Modes

In Figure 5.4, it was observed that although peaks in the spectra of the power number at 1.5 and 3.0 F_N were present when considering individual blades, they were absent

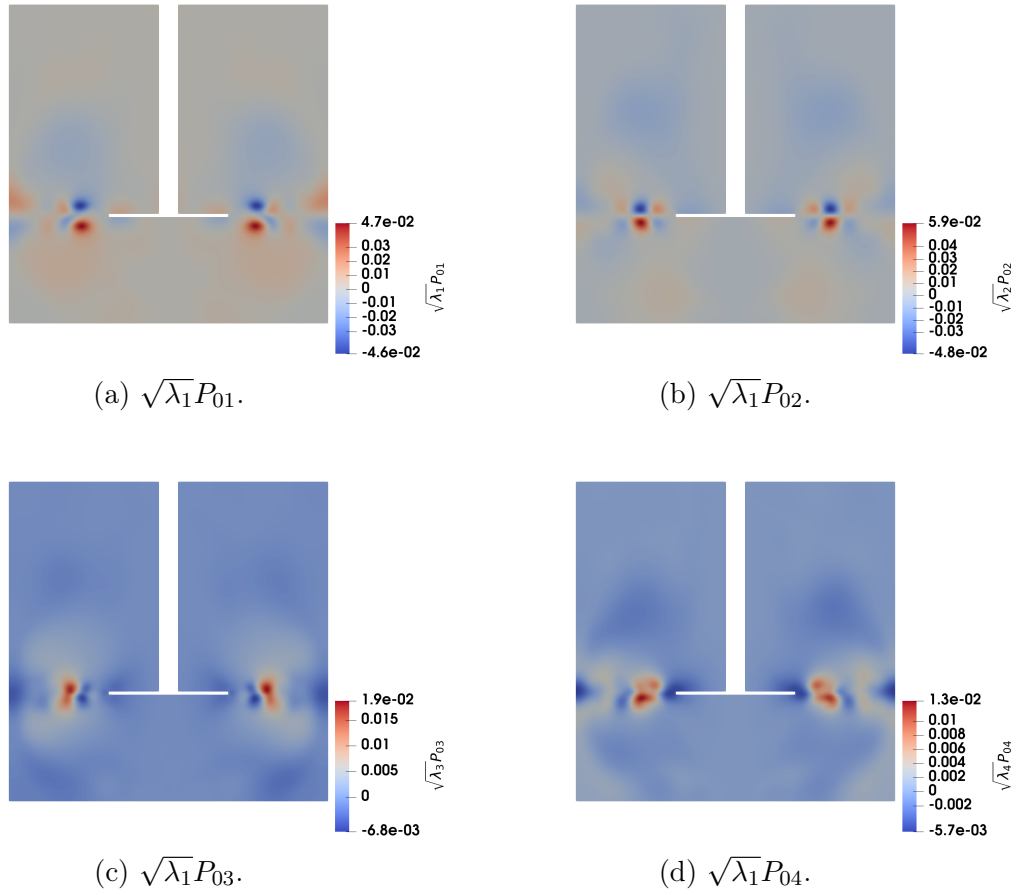


Figure 5.8: Contour plots of the linear partial pressure terms on vertical section bisecting the blades.

for the total power number. The spectra of the power number computed from the linear partial pressures corresponding to the first pair of modes are shown in Figure 5.10. The peak at $1.5F_N$ for individual blades is again present, but, when considering the total corresponding power number, its magnitude is significantly reduced (by a factor of over 150 when comparing to an individual blade, compare the leftmost and rightmost plots). This suggests that there is a cancellation between the blades. The time series of the partial power number in Figure 5.11a for 2 rotations, further demonstrates this with the amplitude of the oscillation significantly reduced when the contribution from three adjacent blades is considered. Examining the partial P_{01} and P_{02} distribution on the blades (Figure 5.12a and 5.12b) it can be seen that for both the pressure and suction sides, the pressure distribution for the blade has the opposite sign compared to the adjacent blade, leading to cancellation when

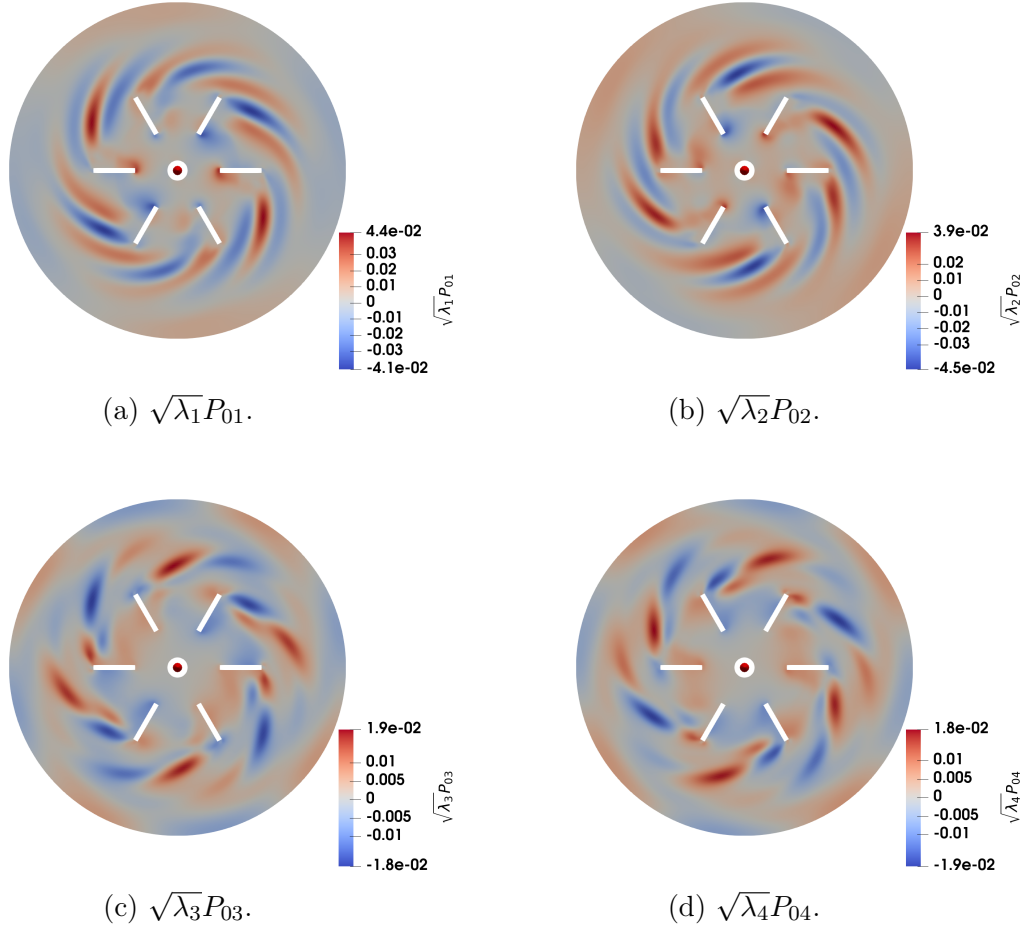


Figure 5.9: Contour plots of the linear partial pressure terms on horizontal section 0.353T from the tank bottom.

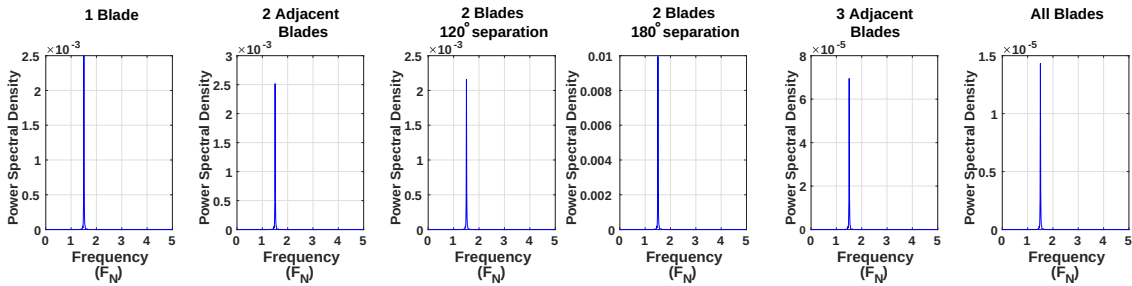
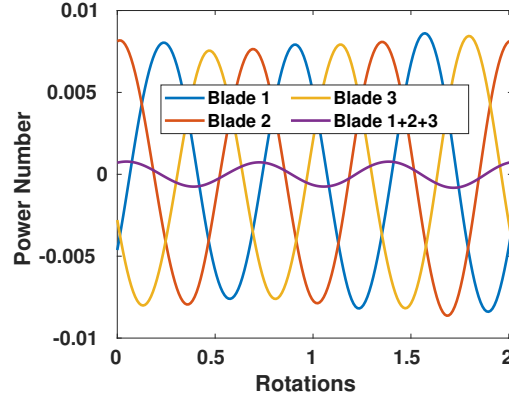


Figure 5.10: $a_1(P_{01} + P_{10}) + a_2(P_{02} + P_{20})$

considering the total power number.

For the spectra of the partial power number computed from P_{03} and P_{04} a similar effect to the first pair of modes is observed, with the large magnitude peaks in the power number power spectrum at $3.0F_N$ for individual significantly compressed when the total power number is considered (reduction by a factor of over 400). The

corresponding time series also shows the same cancellation behaviour as observed for the first pair of modes. The correspondence between adjacent blades observed for the first pair of modes can also be seen for the second pair of modes in Figures 5.12c and 5.12d.



$$(a) a_1(P_{01} + P_{10}) + a_2(P_{02} + P_{20})$$

Figure 5.11: Partial Power Number

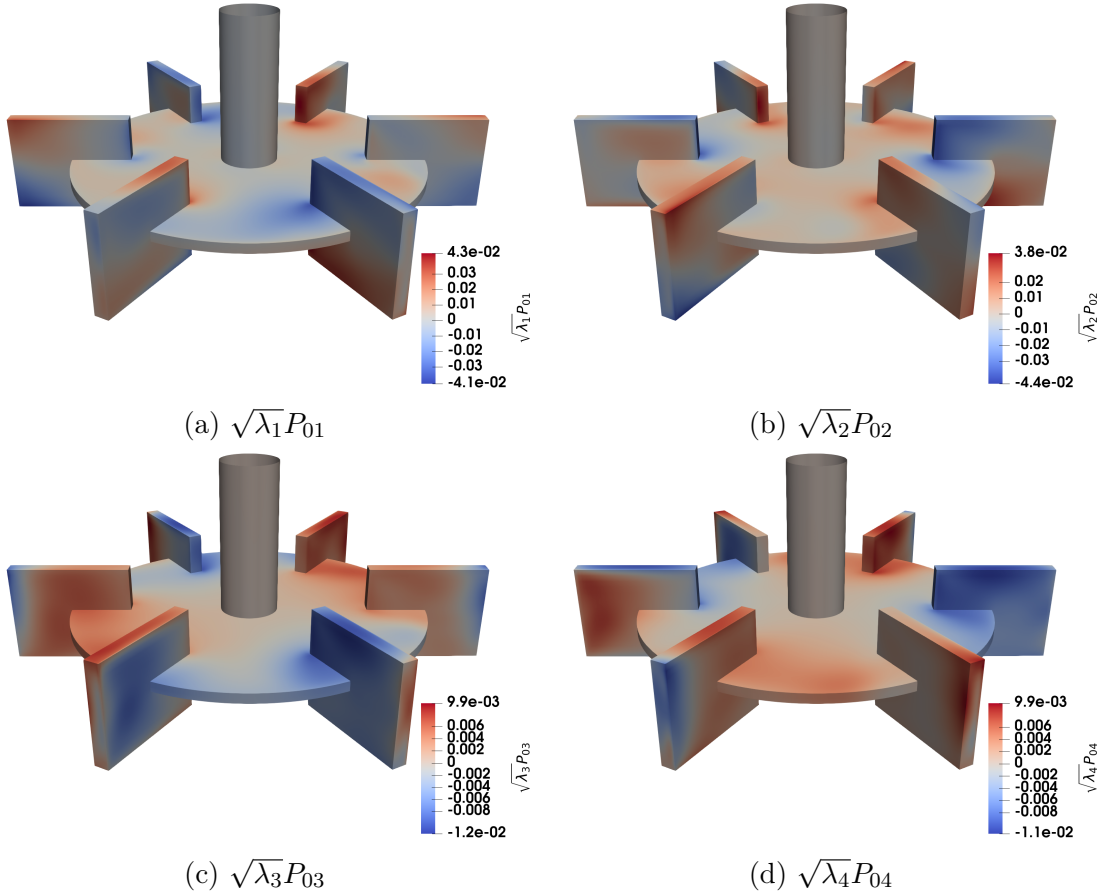


Figure 5.12: Surface Plots of Partial Pressures on Blades.

Due to the negligible contribution of the first two pairs of modes to N_P , higher

modes that contribute smaller percentages to the turbulent kinetic energy have to be considered. The 5th and 6th, and 7th and 8th modes are also paired and the power number computed from the corresponding linear partial pressure terms is again found to be negligible in size.

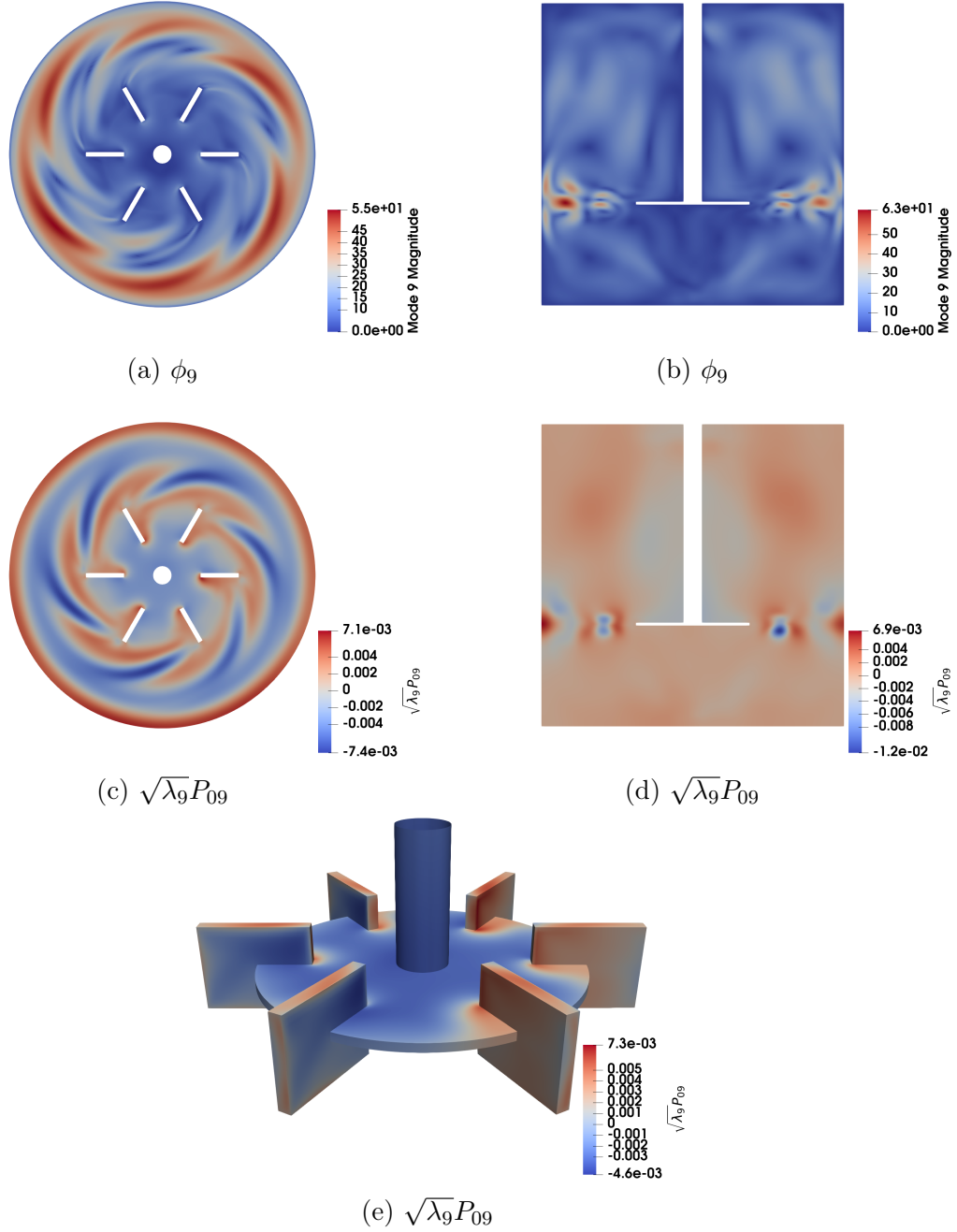


Figure 5.13: Contour plots of velocity magnitude and partial pressure corresponding to Mode 9.

The next mode, 9, is found to be unpaired, with a low frequency peak at $\sim 0.04F_N$. Contours of the spatial component of the POD mode on a horizontal plane 0.353T

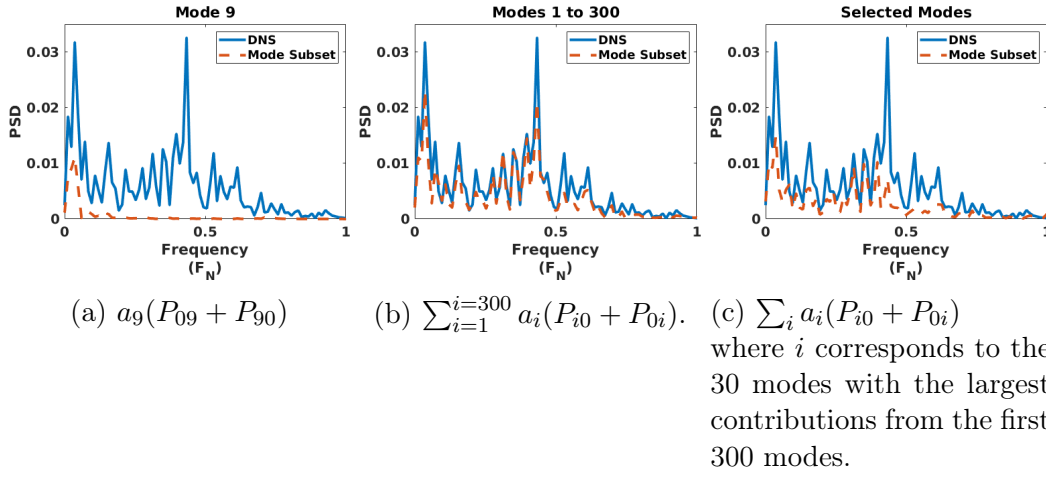


Figure 5.14: Power spectra of contributions to the power number by linear pressure modes.

above the bottom of the tank and on a vertical plane bisecting the blades are shown in Figures 5.13a and 5.13b respectively. The expected 60° symmetry of the mode can be observed, but due to the low frequency of this mode, some discrepancies can be seen. Modes with low frequencies have been observed in previous studies [11, 77, 78, 129–131] and have been associated with the macro-instabilities in the stirred tank.

The power spectrum of the contribution of the partial pressure terms linear in a_9 to the power number is shown in Figure 5.14a. There is a peak at the $0.04F_N$ that was observed in the spectra of the temporal POD coefficient and in the overall total power number (Figure 5.2b). The contribution of this mode to the power number can be explained by examining the spatial distribution of P_{09} mode. Contours of P_{09} at a horizontal plane $0.353T$ above the bottom of the tank, on a vertical plane bisecting the blades, and on the surface of the impeller are shown in Figures 5.13c, 5.13d and 5.13e respectively. The symmetry between the blades means that the power number contribution from each blade will be synchronised and so will sum up to contribute to the total power number with no cancellation. Further, considering a single blade, the pressure on the two sides of the blade has different signs, giving the significant contribution to the power number.

The spectrum of power number computed from the linear partial pressure terms corresponding to the first 300 modes is shown in Figure 5.14b. It matches well with the spectrum of the true power number, suggesting that the majority of the contribution is captured in these first 300 linear terms. The discrepancies suggest that there are contributions from higher linear partial pressure terms, and from the quadratic partial pressure terms. Examination of the first 300 linear modes, revealed that most have no significant peaks in the spectrum of the power number.

The spectrum of N_P computed from the 30 modes with the greatest contribution to RMS of N_P is shown in Figure 5.14c. While some peaks in the full spectrum are reproduced, a significantly simpler spectrum is obtained. In particular, the peak at $\sim 0.43F_N$ is not as distinct. This suggests that while individual modes may have small contributions to the kinetic energy, their contribution should not be neglected as their combination can give noticeable peaks in the spectrum of the power number. It should be noted that a single POD mode can have more than one frequency and that a physical phenomenon with a particular frequency can be spread across multiple modes. In addition, the ordering of the POD modes is based on their contribution to the kinetic energy in the whole tank, and this order may not be the same in the vicinity of the impeller. An approach to extract the modes corresponding to individual frequencies in the power number could be to apply Dynamic Mode Decomposition (DMD) [81] or Spectral POD [83,85] to obtain DMD or Spectral POD pressure modes from the application to velocity modes.

5.8 Reconstruction of Large Scale Vortical Structures from Surface Pressure Measurements

In this section, pressure data taken on the blade surfaces is used to reconstruct the evolution of the temporal coefficients for the first two POD modes for the base $Re = 600$ and off-design conditions $Re = 500$ and $Re = 700$.

As we are constructing a linear estimator for the evolution of the temporal POD coefficients from pressure readings, it is important to consider the relative sizes of the linear and quadratic cross terms in the partial pressures. We find that the magnitudes of the linear partial pressure terms corresponding to the first two modes P_{01} and P_{02} (shown in Figures 5.8, 5.9 and 5.12) have a maximum magnitude of approximately double that of the largest cross term P_{12} (shown in Figure 5.15) in the fluid domain. Importantly, these two modes have a substantially larger footprint in the vicinity of the blades than P_{12} with a maximum magnitude over 10 times greater than that of P_{12} in this subregion. This suggests that, for points chosen on the surfaces of the blades, the assumption of a linear relationship between the pressure evolution and the evolution of the temporal coefficient of the first two POD modes is valid. Considering the linear terms in the second pair of modes (P_{03} and P_{04}), their magnitude is of comparable size to the P_{12} cross term and substantially smaller than that of the first pair of modes. This suggests that a reconstruction of these terms in most locations is unlikely to be successful due to the difficulty in extracting their signal from a pressure time series because that contains not only linear signals corresponding to the first two POD modes but also their interaction.

First, a single sensor point located at the maximum of the pressure RMS on a surface is considered. This was found to be on the suction edge of the blade (indicated with a green dot in Figure 5.16). Next, making use of the rotational symmetry of the geometry, 5 additional points are placed at 60° intervals from the first point. Finally, we add a further 6 points based upon the location of the maximum of the pressure RMS on the outer wall of the tank (indicated with a magenta dot in Figure 5.16) and again use the rotational symmetry of the problem to get 5 more points. Combining these 6 points with the 6 points on the blade surfaces, we consider the final case of input data from pressure sensors at 12 input locations. The pressure time series are pre-processed by subtracting the mean and dividing by the standard deviation of the pressure time series at each sensor location. The quality of the reconstruction

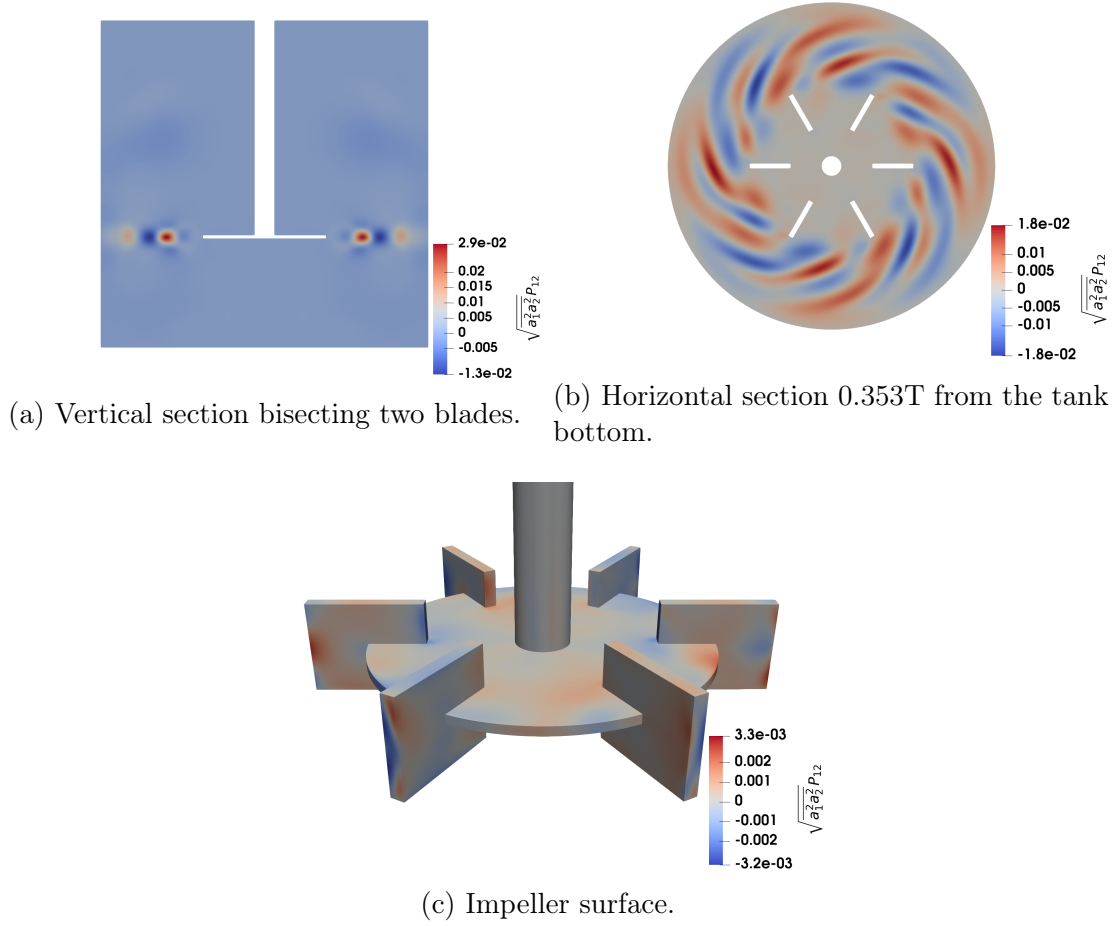


Figure 5.15: Contours of $\sqrt{a_1^2 a_2^2} P_{12}$.

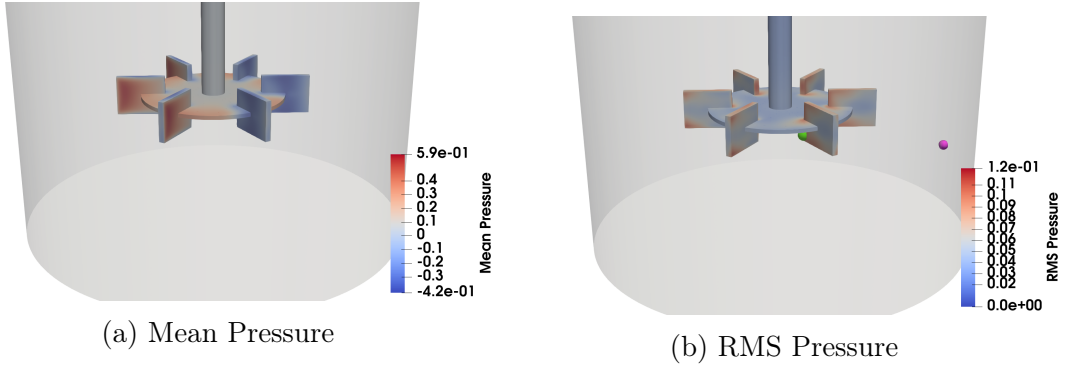


Figure 5.16: Mean and RMS Pressure on impeller surface. Location of max RMS Pressure and max RMS Pressure on the outer tank wall shown in green and magenta, at coordinates (r, θ, z) of $(0.125T, 57^\circ, 0.304T)$ and $(0.500T, 74^\circ, 0.372T)$ respectively.

is quantified by considering the percentage fit defined as

$$\text{FIT}_i[\%] = 100 \times \left(1 - \frac{\|a_i(t_k) - a_{i,e}(t_k)\|}{\|a_i(t_k) - \overline{a_i(t_k)}\|} \right), \quad (5.33)$$

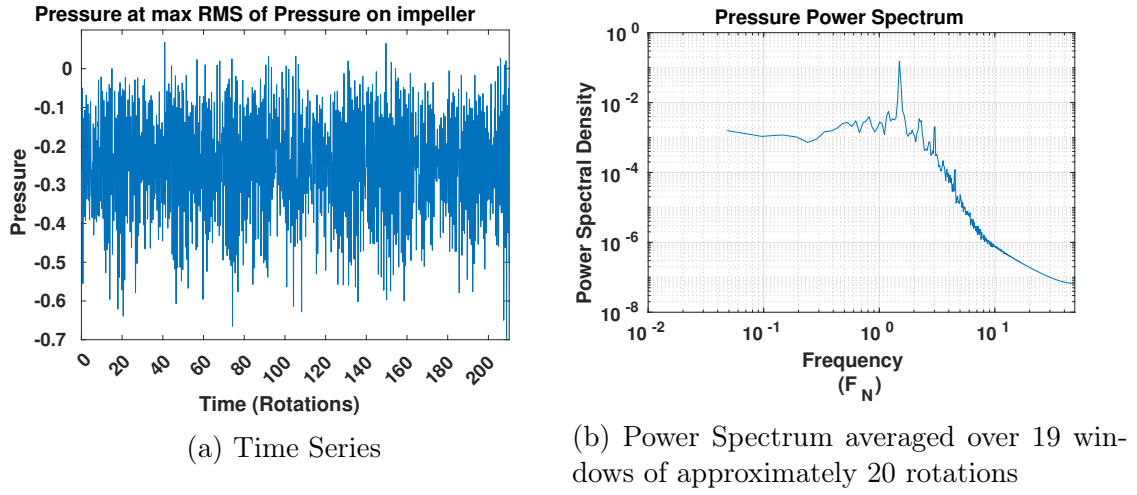


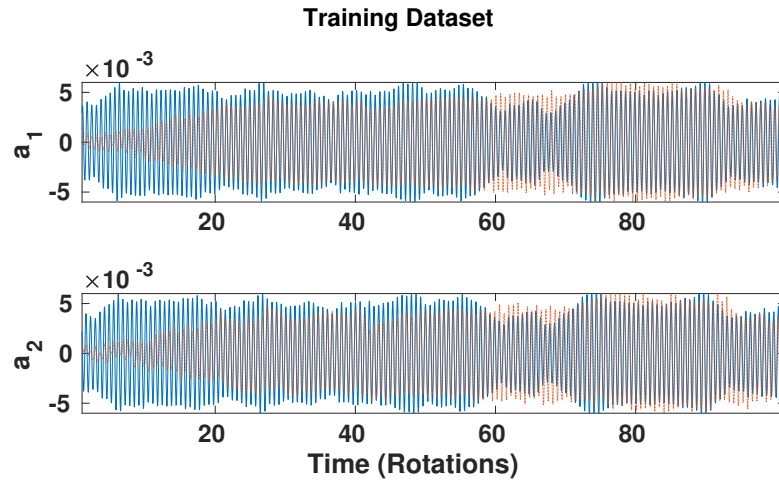
Figure 5.17: Pressure at max PRMS on the impeller.

where $\|\cdot\|$ is the 2-norm of a vector, the subscript e denotes the estimated coefficient, 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 $a_{i,e}(t)$. For both the training and validation dataset, the model is initialised with initial conditions of zero. The fit percentage is evaluated after discarding the first 5 rotations. It should be noted that in some of the below results, the model spin-up takes longer than 5 rotations and as such the fit percentage can be considered as a measure of both model accuracy and speed of convergence of the model to the correct state.

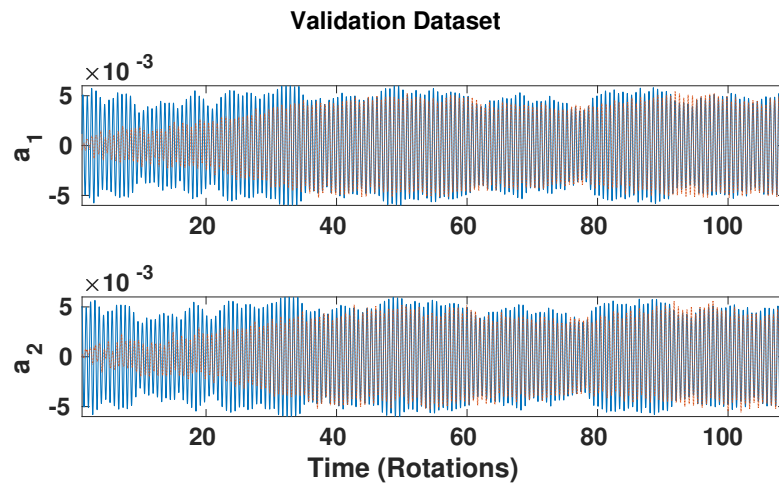
5.8.1 1 Point on the Blade Surface at Maximum Pressure RMS

The time series and spectrum of the pressure at the point with the maximum RMS on the surfaces are shown in Figure 5.17. The spectrum shows a strong peak at $1.5F_N$ corresponding to the frequency of the first two pairs of modes. The time series for the pressure shows that while having a strong periodic nature, the signal is chaotic. The results for $Re = 600$ for this case are shown in Figure 5.18. For the base case, the training and validation fit values for modes (1,2) were found to be (49%,48%) and (53%,52%) respectively. These values reflect that while the period of both

modes is predicted correctly, the amplitude is underestimated locally. Increasing the model spin-up time by discarding the first 20 rotations, rather than the first 5 rotations, increased both the training and validation fit values by approximately 7-9%. This suggests that with substantial spin-up time, a reasonable fit quality of the reconstructed signal to the DNS signal can be achieved.



(a) $Re = 600$ Training



(b) $Re = 600$ Validation

Figure 5.18: Estimated (orange) and true (blue) coefficients for modes 1 and 2 using 1 probe point based on max PRMS (5th order model).

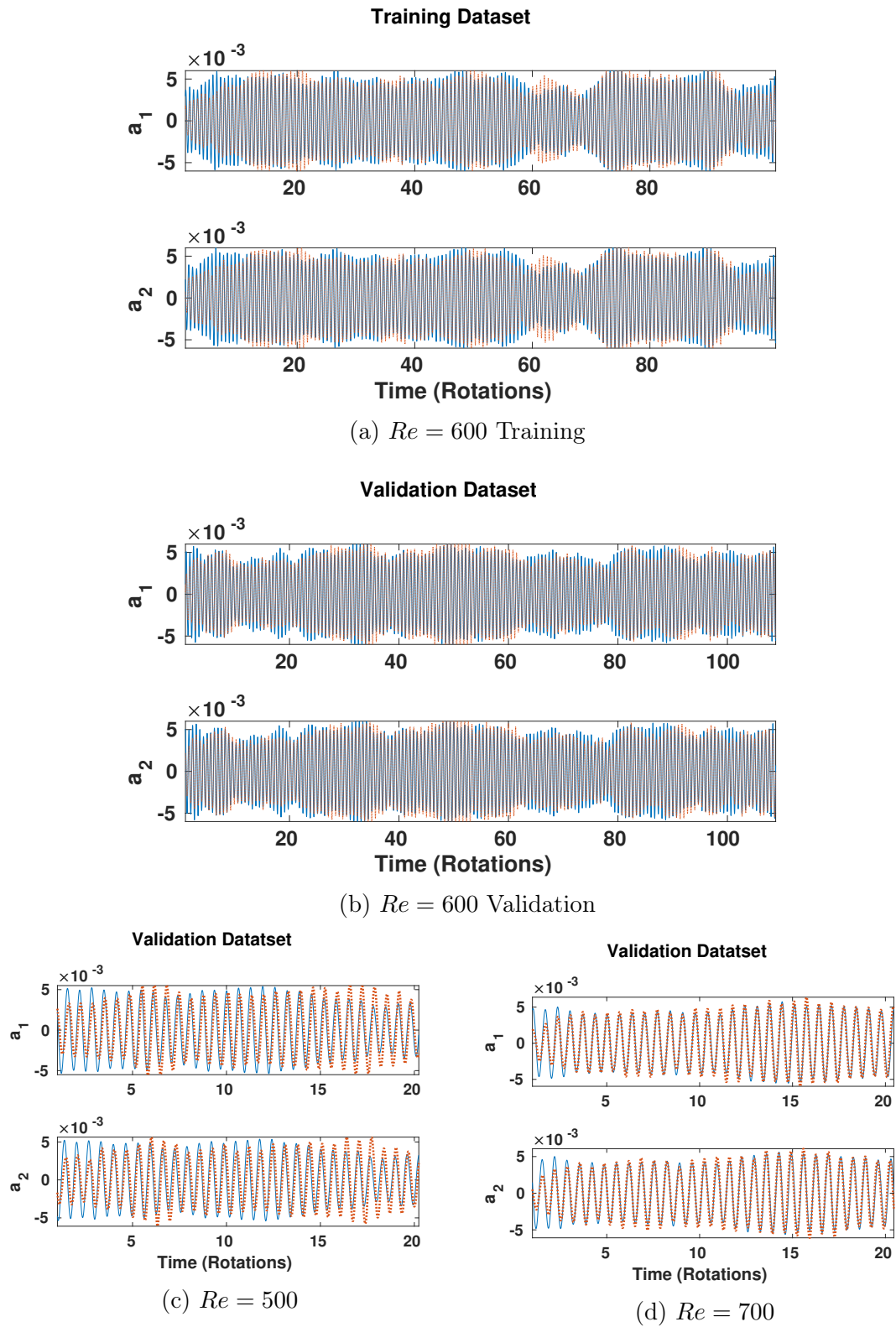


Figure 5.19: Estimated (orange) and true (blue) coefficients for modes 1 and 2 using 6 probe points based on max PRMS (6th order model).

5.8.2 6 Points on Blade Surface Based on the Maximum of the Pressure RMS

The results for the second case of 6 sensors with locations based on the max pressure RMS on the blades and the rotational symmetry of the geometry are shown in Figure 5.19. In this case, a very good fit is obtained for both the training and validation dataset for the base $Re = 600$ case of 76% for both modes for the training dataset and (76%,75%) for the validation dataset. Both the period and the magnitude of the signals are predicted relatively well and this is reflected in the fit values. The predicted signal convergences onto the DNS signal considerably more quickly than the 1 point case. For the $Re = 700$ case, the model output matches the DNS values very well, with fits of (80%,79%). In the $Re = 500$ case, the fit is relatively low, with fit values of (32%,31%). This appears to be due to the DNS signal lagging the reconstructed signal. Introducing a delay of 5 snapshots to the reconstructed signal increases the fit values to (70%,69%) suggesting that both the period and the amplitude of the signal are in fact reconstructed well. The better performance of the reconstruction with an increased Reynolds number suggests that the coupling of the pressure signal at the blade to the evolution of the structures represented by the modes increases as the impeller speed increases.

5.8.3 12 Points based on the Maximum of the Pressure RMS on the Blades and on the Outer Wall

We now consider the third case of 12 points with locations based upon the max pressure RMS on the blades and the max pressure RMS on the outer tank walls, and the rotational symmetry of the geometry. In this case a good fit is obtained for both the training and validation dataset for the base $Re = 600$ case with (84%, 84%) fits for the training dataset and (78%,77%) fits for the validation dataset. Both the period and the magnitude of the signals are predicted well and this is reflected in

the fit values. For the $Re = 700$ case the model output matches the DNS values very well, with fits of (79%,77%). In the $Re = 500$ case the reconstruction is also good with fits of (66%,68%). Notably, the phase shift between the DNS and reconstructed signals is no longer present. This suggests that the addition of the probe points on the outer walls corrects for this. A potential explanation for this behaviour is that the vortex in the $Re = 500$ case is shed slightly earlier than for the $Re = 600$ and $Re = 700$ cases, but then reaches the outer wall slightly later. Hence, this suggests that the vortex propagation speed for the $Re = 500$ case is slightly slower than the other two cases considered.

We investigate this further by computing the cross-correlation coefficient between the pressure time series at the first blade point considered on the blade surface and a_1 . This is defined by

$$\hat{R}_{xy}(m) = \begin{cases} \sum_{n=0}^{S_L-m-1} G_{n+m}H_n, & m \geq 0, \\ \hat{R}_{yx}(-m), & m < 0. \end{cases} \quad (5.34)$$

where G and H represent the two time series, in this case, the pressure time series at the first blade point and a_1 respectively. S_L is the signal length and m is the lag.

The cross-correlation coefficient is then normalised such that autocorrelations at zero lag are equal to 1 by

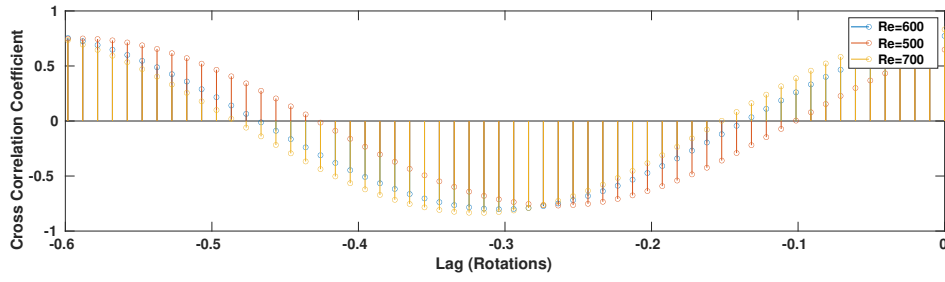
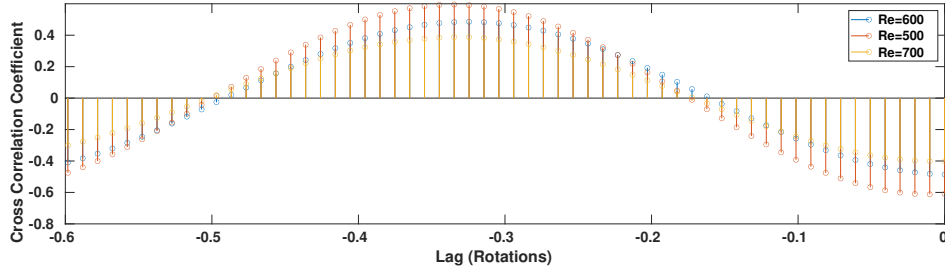
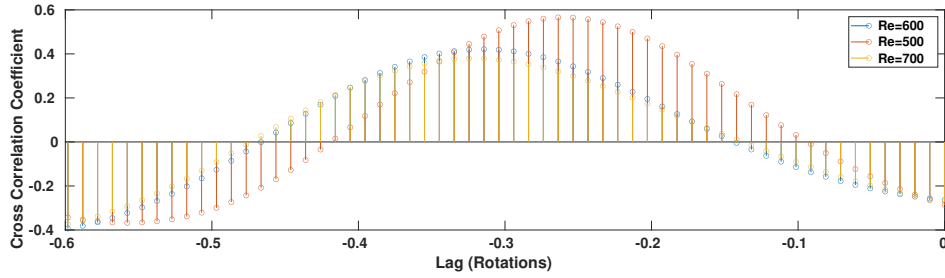
$$\hat{R}_{xy, \text{norm}}(m) = \frac{1}{\sqrt{\hat{R}_{xx}(0)\hat{R}_{yy}(0)}}\hat{R}_{xy}(m). \quad (5.35)$$

The normalised cross-correlation coefficient as a function of signal lag is shown in Figure 5.20a. As N4SID only receives information from the past readings of the pressure we consider the first negative value of lag that leads to a maximum or minimum cross-correlation (as the algorithm is agnostic to multiplication by -1). For the $Re=600$ and $Re=700$ cases, the lag for the first minimum is very similar at -0.30 and -0.31 impeller rotations respectively. On the other hand, for the $Re=500$

case, the lag is only -0.26 impeller rotations. As we train the algorithm on the $Re=600$ case, it is expecting a lag of -0.30 impeller rotations between the input signal and the output signal. However, when this is applied to the $Re=500$ case it leads to an early prediction of a_1 by approximately 0.04 impeller rotations due to the shorter lag between input and output signal. This matches the observed increase in the fit quality for the $Re = 500$ case when a delay of 0.05 rotations is introduced to the input signal.

Next, we consider the normalised cross-correlation coefficient between the pressure time series from a point on the wall from the second set of 6 points. We choose the point that measures the vortex shedding from the same blade as the point considered in the preceding paragraph. The cross correlation as a function of lag between this pressure reading and the a_1 coefficient. We again have to consider the negative lags only as the algorithm only receives information from past timesteps. However, it should be noted that when we use the set of 6 points on the tank wall we are using information from the previous vortex to reconstruct the behaviour of the current vortex. In this case, the lag to the first maximum in the cross-correlation occurs at very similar values for all three cases with values of -0.33 rotations for both $Re=500$ and $Re=700$ and -0.32 rotations for $Re=600$. Hence, this matching explains the improvement in fit value observed in $Re=500$ when the input signals from the 6 points at the tank wall are also considered.

Finally, the normalised cross-correlation coefficient as a function of lag from the point on the blade to the point on the tank wall is shown in Figure 5.20c. For $Re=600$ the first maximum in cross-correlation occurs at -0.31 rotations, for $Re=700$ at -0.32 rotations and for $Re=500$ at -0.26 rotations. This suggests that as the Reynolds number is increased, the normalised propagation time for the vortices decreases. Hence, if the pressure is to be used as an input signal as a reconstruction at off design conditions, then probe points at both the impeller and wall would lead to improved results. Alternatively, the difference in vortex propagation time at off design conditions to the base case could be computed from readings at the

(a) Impeller probe point pressure to a_1 .(b) Wall probe point pressure to a_1 .

(c) Impeller probe point pressure to wall probe point pressure.

Figure 5.20: Cross correlation coefficients as a function of lag.

impeller and wall. Then, only the probe points at the impeller could be used for reconstruction, with a lag introduced based on the difference in vortex propagation time.

5.9 Reconstruction of Instantaneous Pressure

The reconstructed temporal coefficients $a_1(t)$ and $a_2(t)$, as well as the mean velocity, can be used to reconstruct the instantaneous velocity field from

$$\mathbf{u}(\mathbf{x}, t) = \bar{\mathbf{u}} + \mathbf{u}'(\mathbf{x}, t) = \bar{\mathbf{u}} + \sum_{i=1}^{i=\infty} a_i(t) \phi_i(\mathbf{x}) = \sum_{i=0}^{i=\infty} a_i(t) \phi_i(\mathbf{x}) \approx \sum_{i=0}^{i=n} a_i \phi_i(\mathbf{x}). \quad (5.36)$$

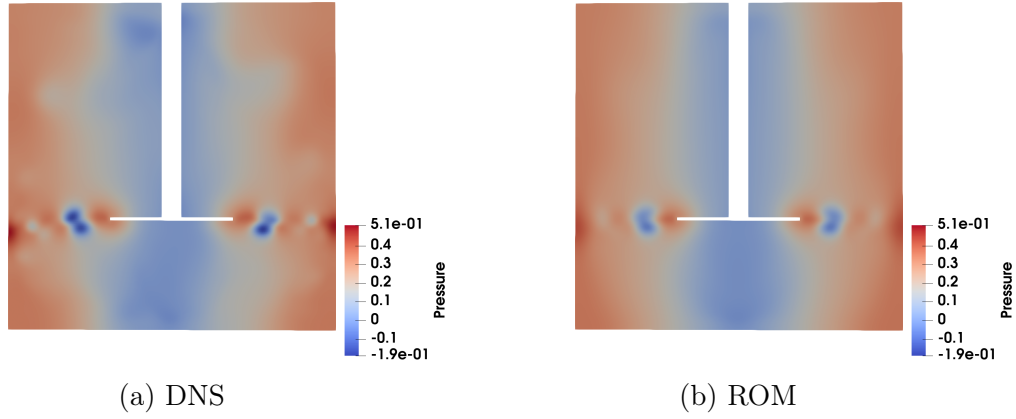


Figure 5.21: Pressure after 40 rotations in validation dataset using 6 points at Pressure RMS maximum reconstruction.

As in our earlier study [120], the reconstruction works well in reproducing the large flow structures in the vicinity of the impeller where the footprint of the first two modes is large. In addition, using the mean pressure and the partial pressure terms that only depend on $a_1(t)$ and $a_2(t)$, the instantaneous pressure can be reconstructed by applying the pressure decomposition (Equation 5.18). Contours of the DNS and estimated pressure on a vertical slice midway between the impeller blades at the same time instant are shown in Figure 5.21. As for the velocity field, the pressure field is reproduced well in the vicinity of the impeller, with the large structures appearing in the reconstructed pressure field. However, the reconstructed pressure further from the impeller looks smooth compared to the DNS pressure field which has random, small scale, features. This is expected as these features are represented by higher order modes that are truncated here.

5.10 Conclusion

It is customary in the literature to decompose the velocity field into a set of POD modes. In this chapter, we extended the decomposition to the pressure field as a sum of mean pressure and terms linear and quadratic in the temporal coefficients of the velocity POD. The partial pressure modes corresponding to the combinations

of the first four velocity POD modes in a stirred tank were visualised for the first time. The contributions of the diagonal partial pressures to the Power Number were examined and 97% was found to come from the P_{00} mode.

The spectrum of the total and per blade Power Number was examined. It was found that although the frequencies $1.5F_N$ and $3.0F_N$ corresponding to the first four modes were strongly present in the per blade Power Number, they were absent in the total. Further examination of the power number computed from the partial pressures associated with these modes found that the contributions from the first two pairs of modes cancelled out when considering all blades together. The first significant contribution to the power number fluctuation came from the unpaired mode 9 incorporating low frequency peaks in the spectrum of the power number. The latter was reproduced well by considering the first 300 linear partial pressure modes, suggesting that the dominant frequencies are distributed over many POD modes.

The temporal coefficients of the leading two modes were then reconstructed using the N4SID algorithm by using pressure measurements at 1 and 6 points on the impeller blade surfaces and 12 points (with 6 points on the impeller blade surfaces and an additional 6 located on the outer tank walls). The first point on the impeller surface was placed at the maximum fluctuating pressure RMS and found to be on the suction side of the blade. An additional 5 points were placed based upon the location of the initial point and the rotational symmetry of the system. Similarly, the first point on the tank wall was placed at the maximum fluctuating pressure RMS on the outer walls, with 5 further points placed by using the rotational symmetry of the system. In the base case, the temporal coefficients were reconstructed using a single data point, with the period correctly identified. However, the validation fit values were limited to only (53%, 52%) for modes (1,2) due to the slow convergence of the model to the DNS results. It should be noted that unlike the velocity based reconstruction method proposed in Chapter 4, the input data at each probe point consists of just one input signal. Increasing the number of points to 6 or 12 points

led to an increase in the fit values to (76%, 75%) or (78%, 77%) respectively.

For the $Re = 700$ case, similar results were obtained to the base case for both 6 and 12 point locations, with the fit values being (80%,79%) and (79%,77%) respectively. For the $Re = 500$ case, for 6 point measurement locations, the fit was (32%,31%), with the magnitude and period reconstructed correctly, but with a phase shift. However, for the 12 point case the reconstruction was good, with fit values of (66%, 68%), with no phase difference between the DNS and reconstructed results.

The extension of the method to using pressure measurements increases its capabilities due to the non-intrusive nature of the measurements. However, this does come with some performance degradation.

The examination of the Power Number revealed some low frequency dependence. These features require some further exploration. For example, when using a longer dataset, would it be possible to use just the instantaneous power number measurement to reconstruct the evolution of these low frequency modes? Further, while the frequencies for the first two pairs of modes were not present in the power number in this case, it would be interesting to examine whether this would also hold for other geometries. For example, a differently sized or shaped impeller, a different impeller clearance from the tank bottom and the addition of baffles may break the cancellation of the power number fluctuations from different blades. In addition, while we have focused on the reconstruction of the two leading order modes in this study, further work needs to be done to identify a method to incorporate the lower energy effects to improve the instantaneous velocity and pressure field reconstruction.

Chapter 6

Identification and Reconstruction of Macro-Instabilities in a Turbulent Stirred Tank Using Data-Driven and Machine Learning Methods

6.1 Introduction

This chapter extends the work of the previous two chapters to consider the fully turbulent case and other machine learning approaches for the reconstruction of the temporal POD coefficients. The POD is applied to the velocity field for the same stirred tank geometry as the previous two chapters, but a base case of $Re = 30000$ and off-design conditions of $Re = 20000$ and $Re = 40000$ are considered and modelled using Large Eddy Simulations (LES). In this chapter, the leading POD modes are interpreted in the context of macro-instabilities (MI) in a stirred tank.

In total, 4 algorithms (N4SID, N4SID with a preprocessing POD step on the probe data, and a 1 and 2 hidden layer LSTM Neural Network) are used to reconstruct

the first pair of POD modes using velocity data from 1, 2 and 6 input points.

The aim of this chapter is to answer the following open questions. Can the methods used to identify and extract the large scale structures for the transitional Reynolds number case be applied to the fully turbulent case? How do the identified modes change with the increased Reynolds number, and how do they link to the previously observed MIs in stirred tanks? How does the performance of the linear, physics-inspired method of N4SID compare to a purely data-driven non-linear method such as a LSTM Neural Network and can an extra pre-processing step improve the N4SID performance?

6.2 Results and Discussion

6.2.1 Mean Flow

We first examine the mean flow field. The results from the earlier grid validation section 2.3 suggest that the mean flow field is dominated by a strong tangential flow, with the form of a solid body rotation close to the shaft, i.e. for $r/D < 0.35$. Contours of the magnitude of the mean velocity and the projection of the mean velocity vectors onto a 2D plane bisecting two blades are shown in Figure 6.1. The mean flow magnitude is small near to the tank centre and increases up to a peak at $\sim 0.35D$ and then decreases closer to the wall. The vertical plane vectors clearly show the strong radial discharge flow at the impeller height, with two counter-rotating vortices above and below the impeller. Strong radial velocities appear at the top and bottom walls as well.

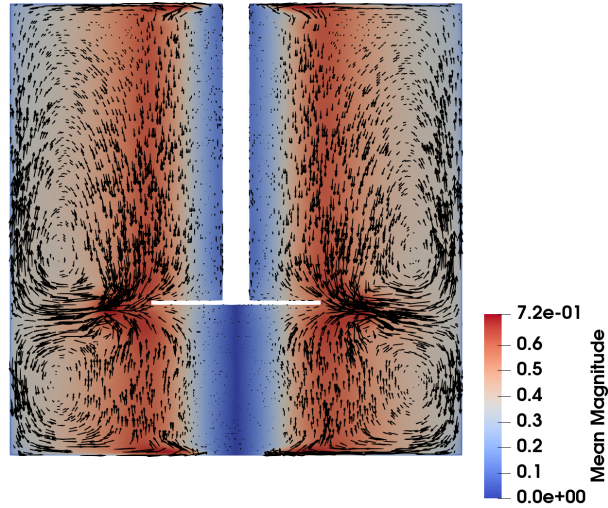


Figure 6.1: Contours of mean velocity magnitude and projection of the mean velocity vectors onto a vertical plane bisecting two blades.

6.2.2 POD Mode Characterisation

Eigenvalues

The eigenvalues from 1 to 9999 as a function of mode number are shown in Figure 6.2a. Mode 10000 has an eigenvalue of 0 to the order of machine precision as we have performed the POD on the velocity fluctuations in this study. The first two modes have the same energy, equal to about 5.5% of the turbulent kinetic energy (TKE) of the flow. They are considerably larger than the other modes, the next two of which each have an energy of approximately 0.9%. Modes higher than 4 have a small contribution to the TKE of under 1%.

The cumulative percentage of the total energy for modes 0 to 100 is shown in Figure 6.2b. The majority of the energy ($\approx 93\%$) is contained in mode 0, corresponding to the mean flow. This is larger than the $\approx 86\%$ observed in the $Re=600$. However, the total energy contained in the first 100 modes is similar to that found for the $Re=600$ case, corresponding to over 95% with a long tail again being observed.

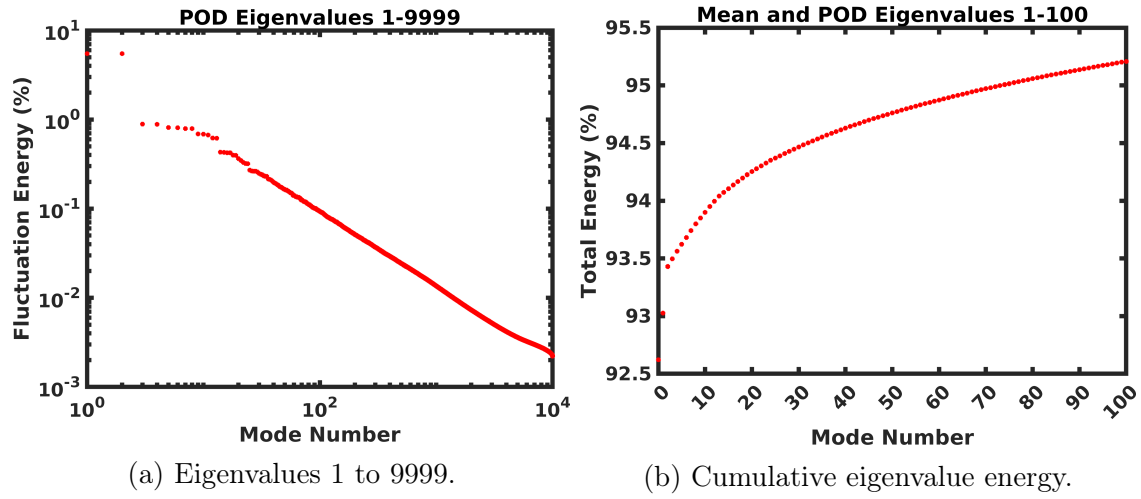


Figure 6.2: POD mode eigenvalues.

Temporal Coefficients

The signals of the temporal coefficients of the first 6 POD modes are shown in Figure 6.5 and spectra of the first 12 modes are shown in Figure 6.6. The first two modes have an identical peak frequency of $1.4F_N$. This frequency also matches the frequency of $1.4F_N$ observed experimentally by [132] in the pressure fluctuation on the pressure and suction sides of the blades in an octagonal unbaffled stirred tank agitated by an impeller that has four blades with regular and fractal shapes at $Re = 150000$ and 200000 . Frequencies of $1.57F_N$ for $Re = 320$ and $1.5F_N$ for $Re = 1600$ in the velocity spectra in the vicinity of the blades in a rotating frame for the same impeller types, but with a cylindrical unbaffled tank were observed in [15]. However, the corresponding signals reveal a more complicated behaviour with a clear beating behaviour visible. The 2 modes have a phase difference of a quarter of a period, but are otherwise identical, suggesting that they are again paired.

Modes 3 and 4 also have an identical peak frequency of $0.8F_N$ and their time signals have a phase shift of a quarter of a period. This suggests that these modes are also paired and like the first two modes, the time series reveals a complicated beating behaviour. Most importantly, however, unlike for the $Re = 600$ case these modes are not the harmonics of the first pair of modes. An identical frequency of $0.8F_N$

in the pressure fluctuation on the pressure and suction sides of the blades was also observed experimentally by [132].

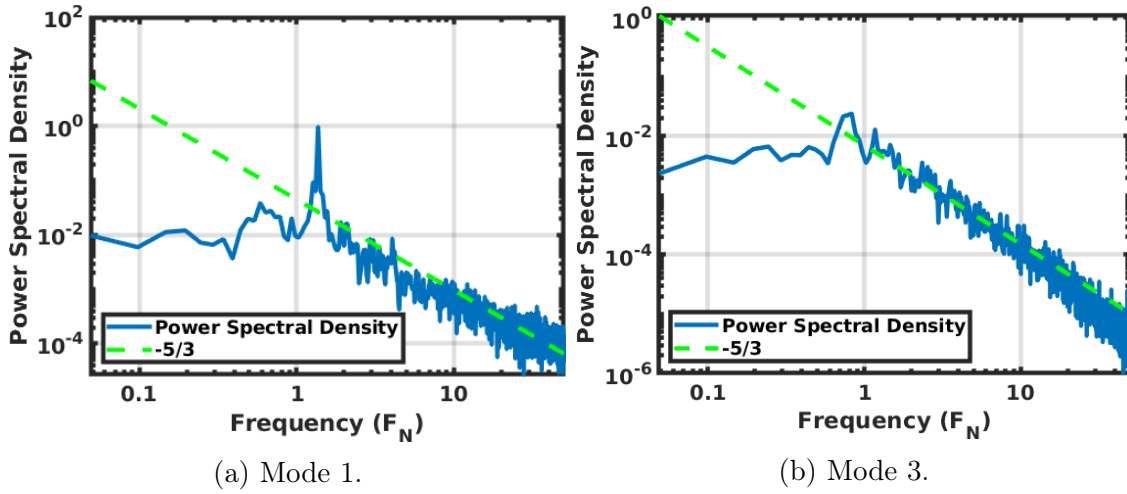


Figure 6.3: Power spectra of u velocity at the locations of the maxima of POD modes 1 and 3.

The power spectra of the velocity at the maximum of POD mode 1 and 3 are shown in Figure 6.3a. A peak at $1.4F_N$, corresponding to the first and second POD mode frequency can be clearly seen in the power spectrum at the maximum of POD mode 1. While less distinct, a peak at $0.8F_N$, corresponding to the third and fourth POD mode frequency, is present in the power spectrum at the maximum of POD point 3. The $-5/3$ power law can also be observed for both power spectra, suggesting that despite not having a DNS resolution, the LES nevertheless captures a significant proportion of the inertial range. We further verify this by considering the compensated power spectra [56] at the same two locations. These are shown in Figure 6.4. The compensated power spectra are computed by multiplying the power spectrum by $\epsilon^{-2/3} f^{5/3}$, where ϵ is the dissipation of TKE and f is the frequency. Both compensated power spectra approximately display the horizontal line corresponding to the $-5/3$ power law.

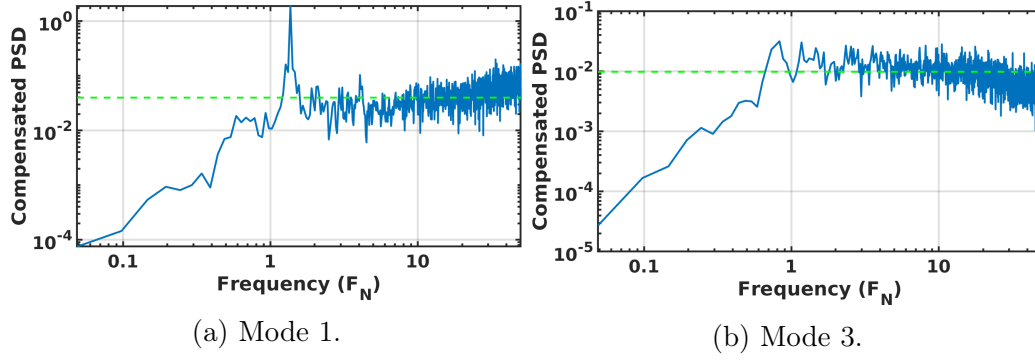


Figure 6.4: Compensated power spectra of u velocity at the locations of the maxima of POD modes 1 and 3.

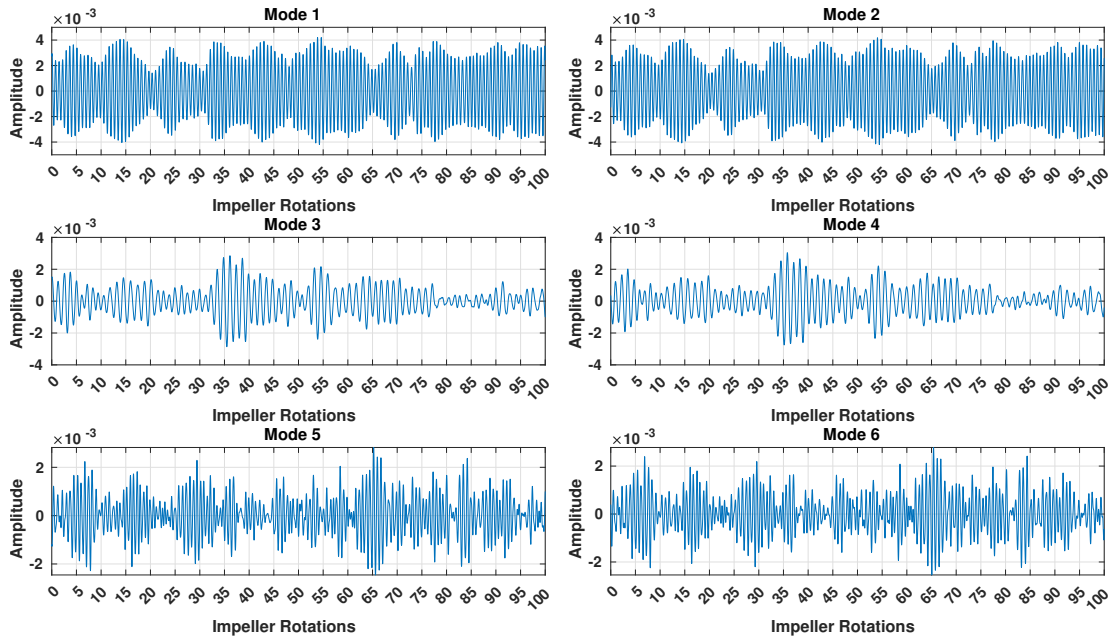


Figure 6.5: Temporal coefficients of POD modes.

Spatial Distribution of Modes.

Contours of the velocity magnitude on a horizontal plane for the first 4 modes at the height of the maximum TKE are shown in Figure 6.7. Modes 1 and 2 both have a symmetry of 180° and show structures behind the blades corresponding to rotating trailing vortices. Modes 3 and 4 appear to be concentrated around the impeller shaft and also have a 180° phase difference in the spatial distribution. Contours of velocity magnitude of the spatial mode components at vertical planes bisecting two blades are shown in Figure 6.8. Modes 1 and 2 have the largest magnitude at the

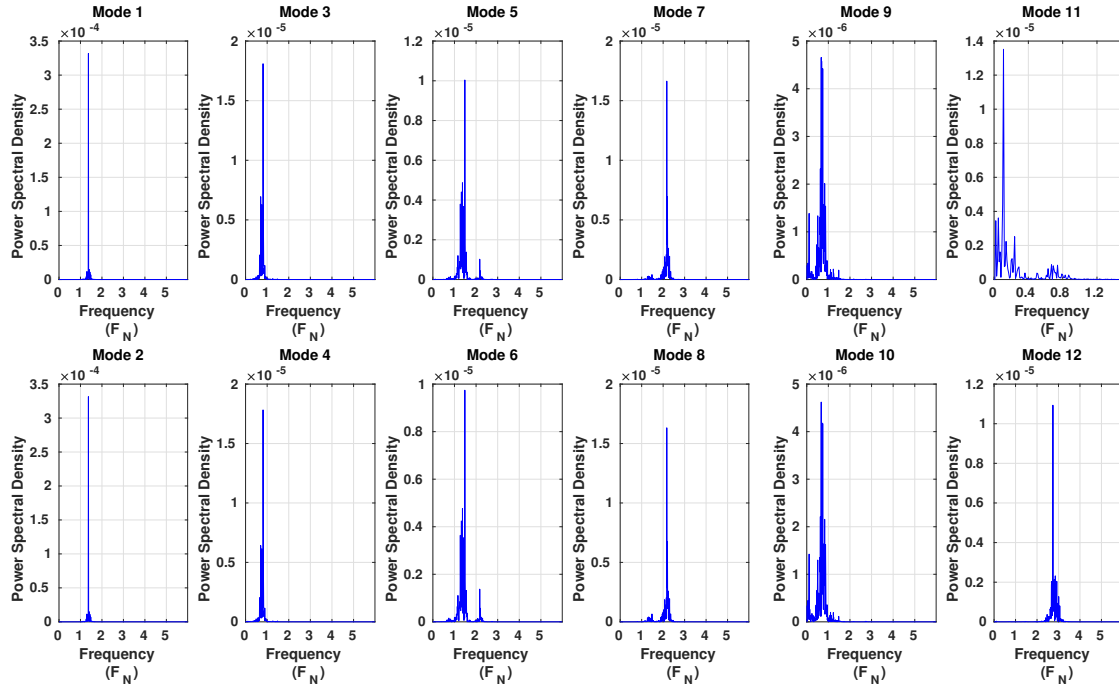


Figure 6.6: Spectra of temporal coefficients of POD modes.

impeller blade height. On the other hand modes 3 and 4 show large structures that extend throughout the height of the vessel, suggesting that they could correspond to a precessing MI. In particular, they appear to have a large footprint at the top and bottom of the vessel. Similar regions of high TKE were observed in [106] for the same Reynolds number and geometry. These regions were also found at the bottom of the tank in [131], where an unbaffled tank agitated by a flat blade impeller with 8 blades was investigated. They were attributed to the "Ekman pumping effect", arising from the interaction between the swirling flow generated by the impeller and the stationary bottom surface. In [131] only the bottom of the tank was assumed to be stationary, while the lid was modelled as a free surface to approximate an uncovered tank. In contrast, in the present study both the top and bottom surfaces are stationary (in the inertial frame) and hence, the feature is present at both walls. The centres of the two vortices forming the macro instability can also be seen in the vertical slice for mode 4, corresponding well to that observed experimentally in [4]. We shall investigate these modes further below.

To provide more insight into the behaviour of the first 4 modes below the impeller,

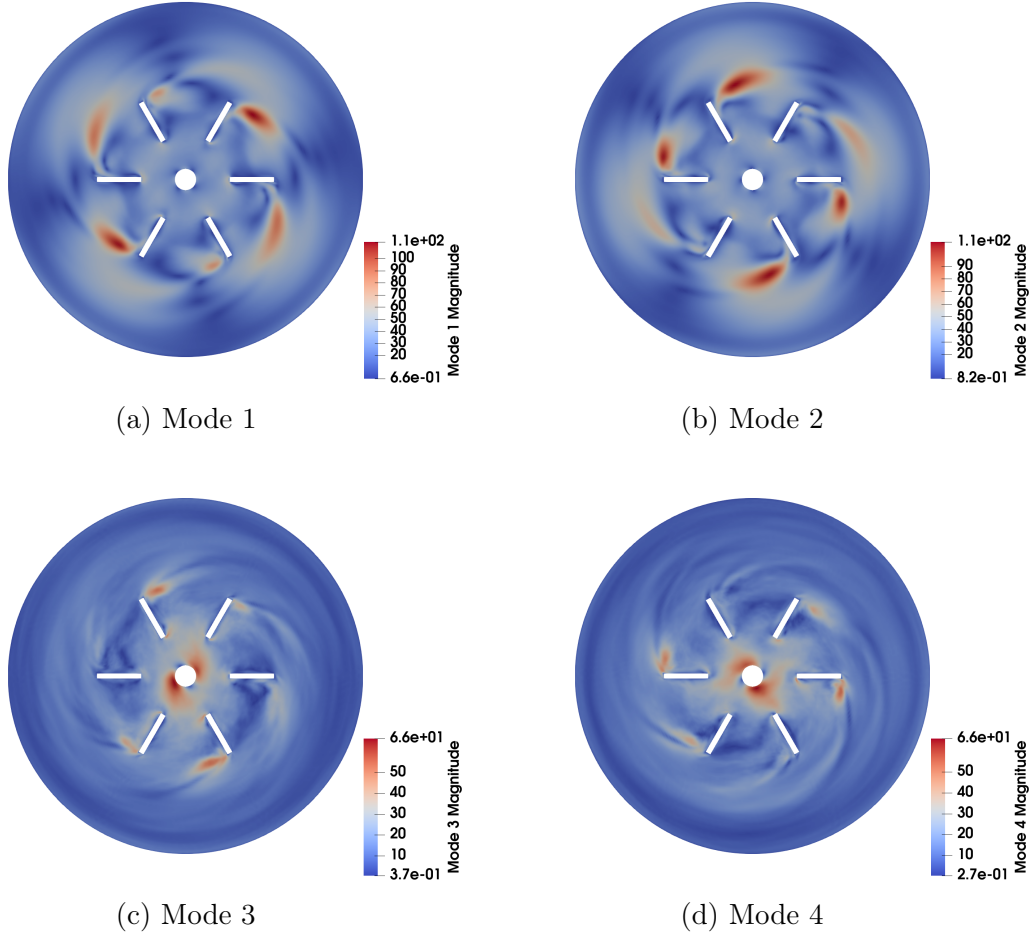


Figure 6.7: Contours of POD velocity magnitudes at a plane $0.357T$ above tank bottom.

we consider the axial component of the vorticity and the velocity vectors at a 2D horizontal slice $0.11T$ above the tank bottom in Figure 6.9. This is similar to the slices considered in [4, 11]. Modes 1 and 2 each show 2 regions of positive and 2 of negative vorticity and 180° symmetry. These bear a close resemblance to modes 3 and 4 obtained by [11] from POD on PIV data on a 2D slice below the impeller in a baffled stirred tank agitated by a $T/3$ Rushton impeller with a clearance of $T/2$ at a $Re = 8000$ and 27000 . The authors attributed these modes to a MI. In the reference [4] the same geometry as [11] was studied, but with a $Re = 33600$ and a horizontal plane closer to the bottom of the impeller was considered. There the first pair of modes obtained were attributed to the trailing vortices and show a similarity to the structures of modes 1 and 2 in our study in the vicinity of the impeller, as

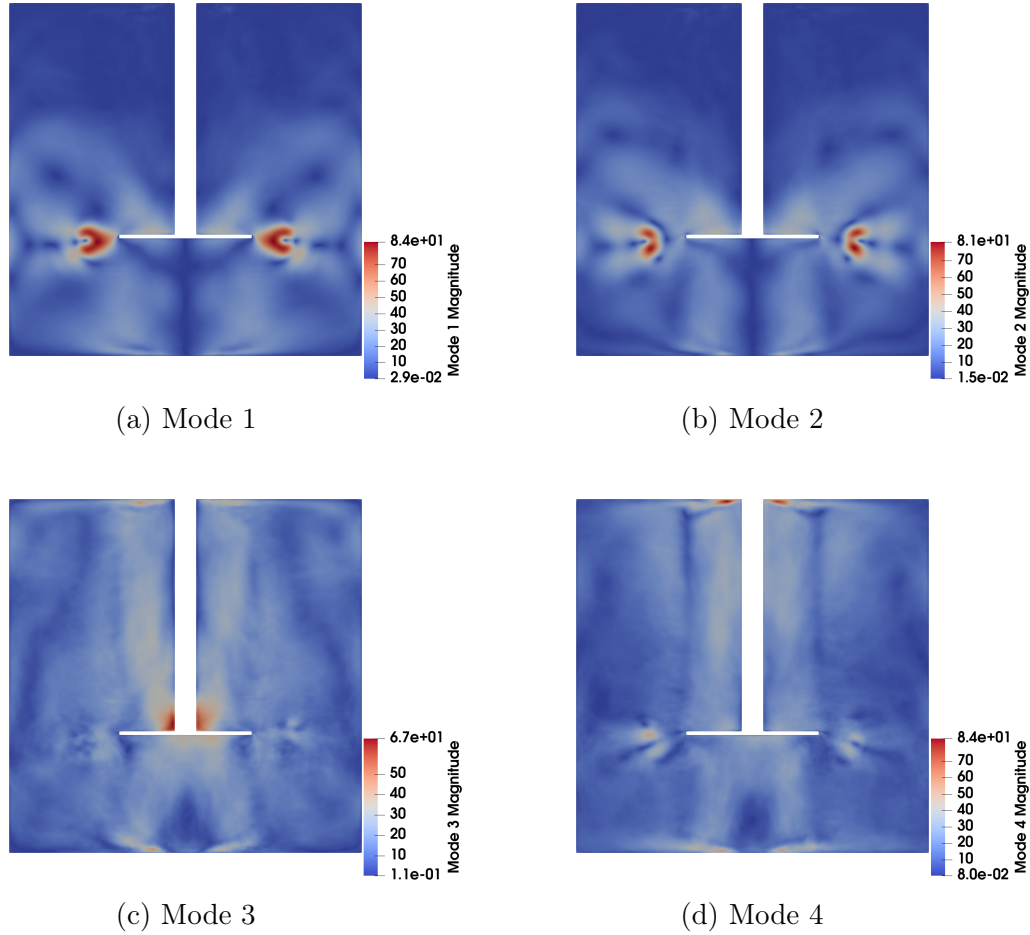


Figure 6.8: Contours of POD velocity magnitudes at a vertical plane bisecting blades.

shown for the POD spatial mode magnitude in Figure 6.7. This suggests that in our study, the first two modes correspond to rotating trailing vortices that appear to form a MI as they rotate around the vessel. As a further check of this hypothesis, we plot the contours of the magnitude of the spatial component of mode 1 at different heights from the tank bottom in Figure 6.10. In the region around the impeller, the trailing vortex structures behind the blades can be seen. However, as the radius increases these start to correspond to the 4 vortex structure that is also observed at heights above and below the impeller. This 4 vortex structure has also been observed in the lower Reynolds number case of $Re = 150$ with the same geometry by [133] with a slightly higher frequency of $1.66F_N$.

Modes 3 and 4 each have a large region of positive vorticity and a large region

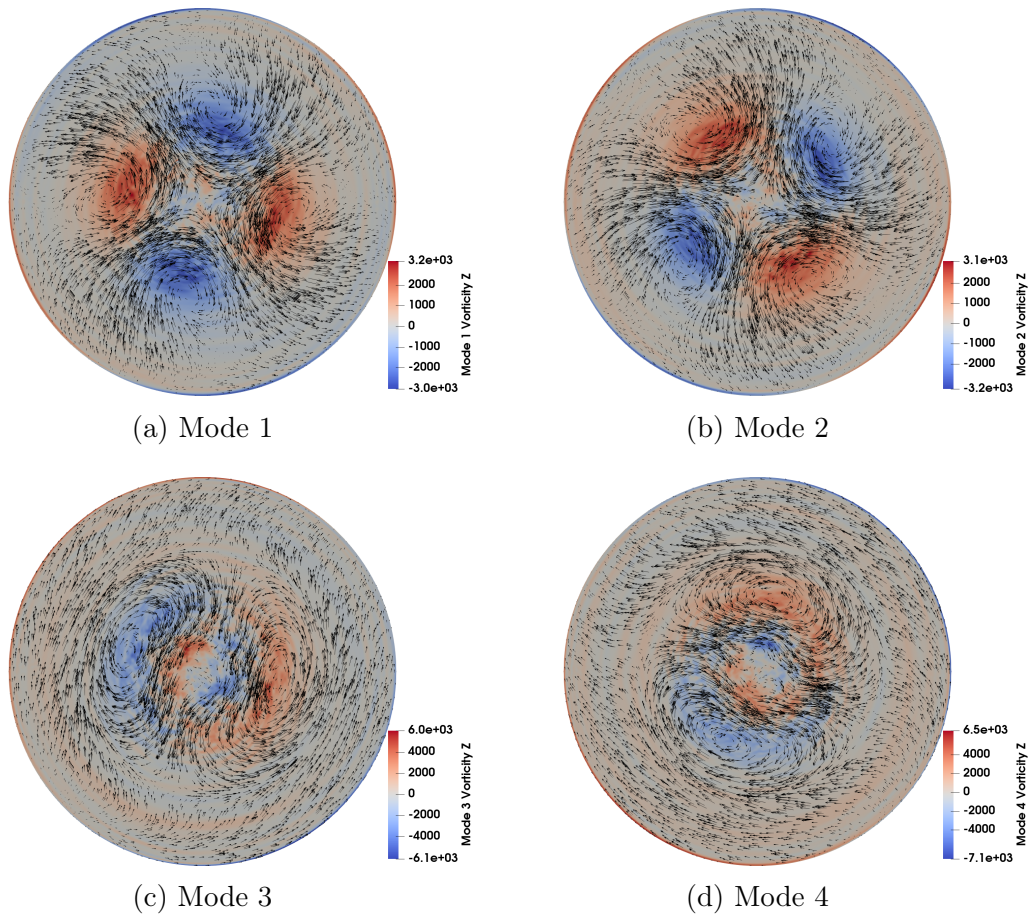


Figure 6.9: Contours of axial vorticity superimposed on velocity vectors at a horizontal plane 0.11T above the bottom of the tank.

of negative vorticity and show that while the magnitude of the POD modes has a 180° symmetry, the vortices are in fact locally rotating in opposite directions about their respective centres (while also both rotating in the same direction around the vessel centre). These show close similarity to modes 1 and 2 in [11] and modes 3 and 4 in [4], where they were attributed to a precessing MI. Indeed, considering the combination of modes 3 and 4, it can be seen that they correspond to precession of large structures about the impeller shaft. Contours of the magnitude of the spatial component of mode 3 at different heights from the tank bottom are shown in Figure 6.11. Here the 2 vortex structure is observed at all impeller heights.

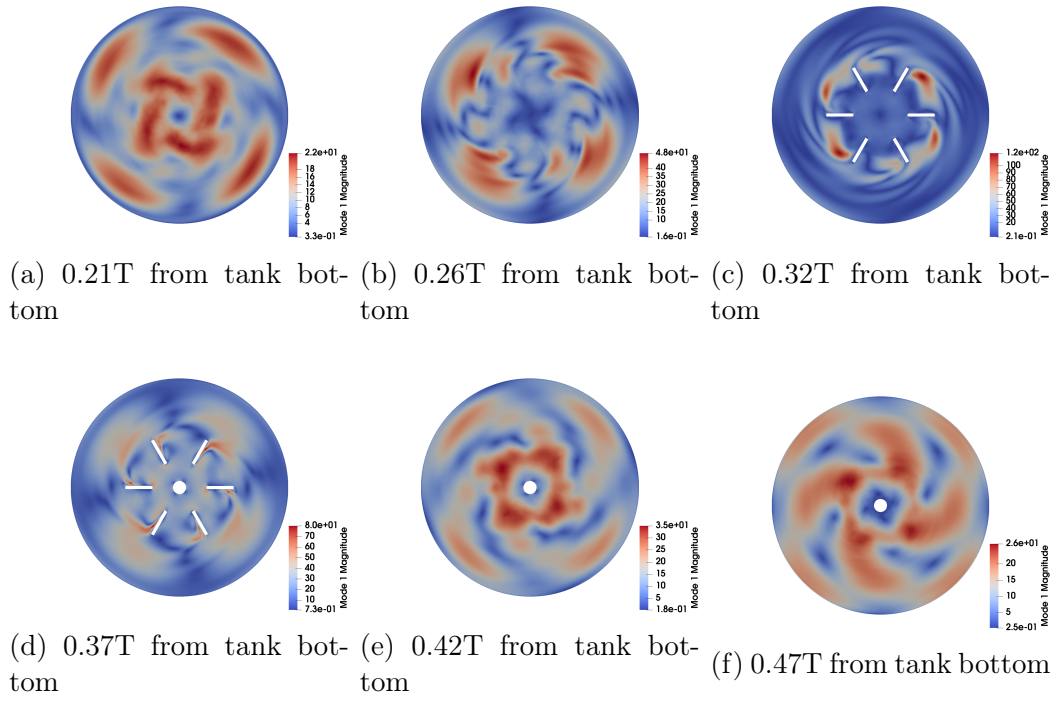


Figure 6.10: Contours of mode 1 velocity magnitudes on horizontal planes at various heights above the tank bottom.

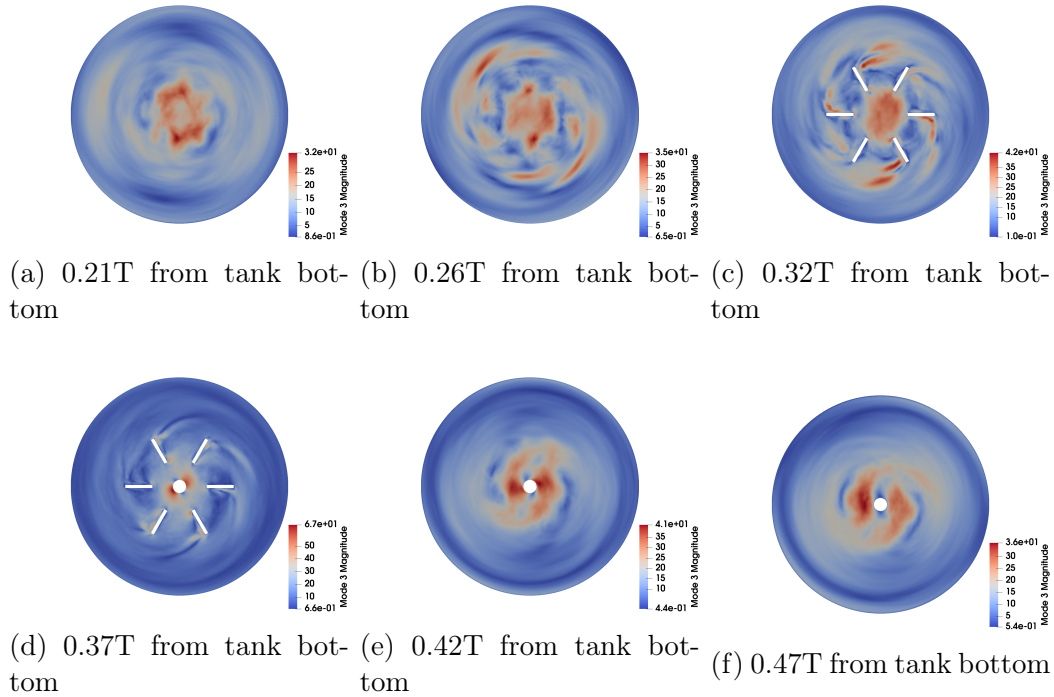


Figure 6.11: Contours of mode 3 velocity magnitudes on horizontal planes at various heights above the tank bottom.

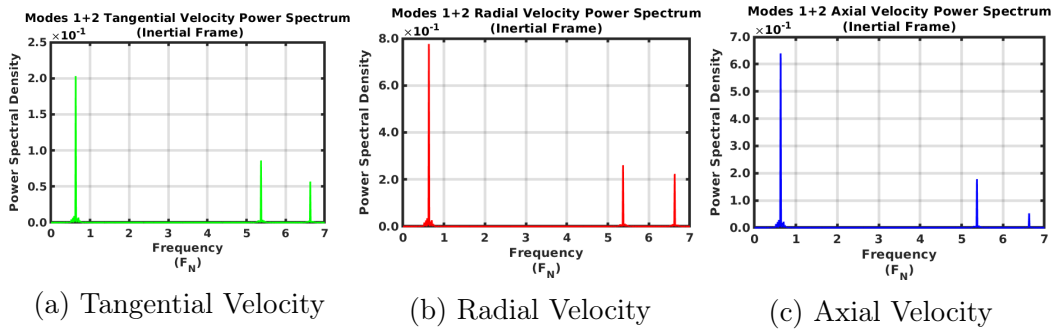


Figure 6.12: Power spectra of the velocity components corresponding to Modes 1+2 in an inertial frame at the height and radius of the max of POD mode 1.

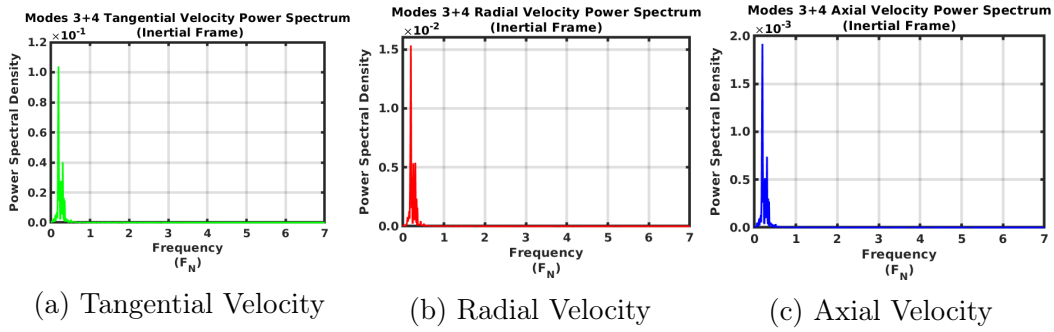


Figure 6.13: Power spectra of the velocity components corresponding to Modes 3+4 in an inertial frame at the height and radius of the max of POD mode 3.

Frequency of POD Modes in the Inertial Frame

The majority of studies on the frequencies present in stirred tanks have been conducted in an inertial i.e. laboratory reference frame. As such, to compare our results with these studies, we compute the frequencies of the first 4 modes in the inertial reference frame. We start with the first pair of POD modes. As these modes are paired, they correspond to travelling structures and following [32] we consider the flow feature represented by the first two modes as a travelling wave of wavelength $2\pi/k$, where k is an integer that represents the azimuthal wave number, and an angular celerity relative to the impeller of $\omega = \pm c \times (2\pi N)$, where $c = F_N/k$. In an inertial frame, this corresponds to an angular celerity of $\omega + 2\pi N = 2\pi N(1 \pm c)$. As mode 1 is a quarter of a period ahead of mode 2, we can see that the travelling wave is in the clockwise direction in the rotating frame (opposite direction to the impeller rotation) and hence has an angular celerity of $-c(2\pi N)$ and $(1 - c)(2\pi N)$.

in the rotating and inertial frames respectively. As both modes 1 and 2 have a 180° rotational symmetry, as in [32], there are two wavelengths per 2π . Hence $k = 2$ and the angular celerity in the rotating and inertial frames is $-0.7(2\pi N)$ and $0.3(2\pi N)$ respectively. Finally, as $k = 2$, the frequency of the modes in the inertial frame is $\approx 0.6F_N$.

To verify our calculation we compute the power spectrum of the 3 components of velocity from the first two modes at a fixed probe point located at the radius and height of the maximum of the first POD mode in a rotating frame. This was done by selecting the initial point to be located in the rotating frame at the maximum of the first POD mode. At the next timestep the new location of the point in the rotating frame, corresponding to the fixed point in the inertial frame, is computed. The cell in which the point lies is then identified and the radial and tangential velocities are computed and all velocity components are recorded. The process is then repeated for the next timestep. The results are shown in Figure 6.12. The largest peaks in all 3 spectra are located at the frequency of $\approx 0.6F_N$ matching our theoretical calculation and peaks located at the blade passage frequency of $6F_N \pm 0.6F_N$ are also present.

These frequencies are considerably higher than the values typically associated with processing MIs in baffled stirred tanks of $\sim 0.01F_N$ and in unbaffled stirred tanks of $\sim 0.1F_N$. Nevertheless, they are very similar to the frequencies observed in the oscillation of the surface of the liquid in an unbaffled stirred tank agitated by a T/3 sized Rushton turbine observed in [51]. Peaks at $0.74F_N$ and $1.03F_N$ were observed for $Re \approx 15000$ and at $0.5F_N$ and $0.73F_N$ for $Re = 95000$, as well as harmonics of the latter frequency for both Reynolds numbers.

Next, we consider modes 3 and 4 in an inertial frame. As mode 4 is a quarter of a period ahead of mode 3, we can see that the travelling wave is also in the clockwise direction (opposite to the impeller rotation). The argument applied for modes 1 and 2 can be applied here also, but due to the 180° antisymmetry of the modes

$k = 1$. Hence, the angular celerity in the inertial frame of the structures represented by these modes is equal to $0.2(2\pi N)$ and the frequency is $0.2F_N$. To verify this we again consider the power spectra of the three velocity components, at a fixed probe point in the inertial frame located at the height and radius of the max of the third POD mode; the results are shown in Figure 6.13. The frequency of $0.2F_N$ is present in all three power spectra.

This agrees with frequencies observed for precessing MIs in eccentric impeller unbaffled tanks [34–36, 52]. Although not attributed to MIs by the authors, frequencies of $\sim 0.2F_N$ can also be observed in the presented radial velocity spectra for the unbaffled centred Rushton impeller case investigated in [129, 130]. In addition, the similarity of the vorticity contours on a horizontal plane underneath the impeller to those observed and attributed to a MI in [4, 11], suggests that these two modes correspond to a second precessing MI. Finally, this value agrees with the frequency of $0.162F_N$ observed for the Ekman pumping feature in [131]. Hence, we conclude that modes 3 and 4 do indeed represent a precessing MI which is active around the impeller shaft.

6.2.3 Reconstruction of POD Modes 1 and 2

Validation of past number of time steps for LSTM

Prior to examining the performance of the algorithms for the reconstruction of the temporal coefficients we first validate the choice of using data from 75 timesteps (1% of the training dataset length) for the input into the LSTM neural network. This is approximately the number of timesteps corresponding to a full oscillation of the first 2 POD modes. In [70] it was suggested that using more than 1 past timestep as an input leads to better performance in the reconstruction of the temporal coefficients as the neural network can more precisely identify the trends in the data. We consider LSTM 2 for the case of 6 input points with 1 point located at the maximum of the

first POD mode as shown in Figure 6.14 and a further 5 points at 60° intervals from the first. The outputs are the temporal POD coefficients of the first two modes.

At the start of the training process of the neural network, the weights are initialised with random values, hence the final fit values can show small variations from one training run to another. Hence, all of the results presented for the LSTM are the average fit values over 10 sets of training and validation runs. The average of the fit values for the training and validation datasets for the first two modes over a range of number of past input timesteps are shown in Figure 6.15. As observed in [70], the performance is poorest when using data from just the current timestep with an average training data fit of 81% and validation data fit of 68%. Increasing the number of timesteps used to 10 increases the training data fit to 95% with no further significant increase for a larger number of timesteps. On the other hand, the fit for the validation data shows a small increase in fit up to 30 timesteps where the fit reaches approximately 79%. Hence, the 75 timesteps used in this study are considered to be sufficient to obtain the best possible performance of the LSTM.

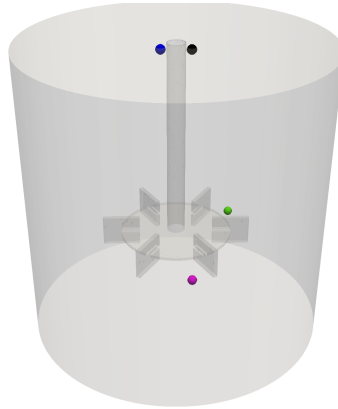


Figure 6.14: Locations of the maxima of the first 4 POD modes. Mode 1 (green point), Mode 2 (magenta point), Mode 3 (black point) and Mode 4 (blue point).

Temporal Coefficients

Having identified the large scale periodic structures in the flow, we now reconstruct their behaviour by using data from sparse locations in the flow. We initially focus

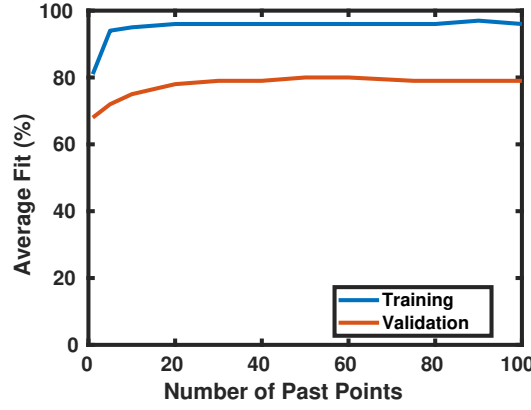


Figure 6.15: Average fit value dependence on the number of timesteps for a 2 hidden layer LSTM neural network.

on the reconstruction of the first pair of POD modes. Input data from 1, 2, and 6 data points are considered with 7500 timesteps used as a training dataset and 2500 timestep validation datasets for $Re = 20000$, 30000 and 400000. For the single point case, the point is placed at the location of the maximum of the first POD mode. For the 2-point case, a second point is added at the location of the maximum of the second POD mode. Finally, for the 6-point case, the first point is located at the maximum of the spatial component of the first POD mode and a further 5 points are placed at 60° intervals from the first. Figure 6.14 shows the location of the maximums of the first 4 POD modes, with modes 1, 2, 3 and 4 shown in green, magenta, black and blue respectively. For the Point POD+N4SID algorithm only the first 2 Point POD temporal coefficients are used as inputs to N4SID.

The off design conditions of $Re = 20000$ and 40000 represent a larger change in Reynolds number from the base $Re = 30000$ case compared to the previous two chapters. However, as we are in the turbulent regime here (compared to the transitional regime in the previous two chapters), the flows for all three Reynolds numbers are more self-similar. On the other hand, the much smaller percentage of TKE represented by the leading order modes for the $Re = 30000$ case (5.5% compared to 14% for each of modes 1 and 2), presents an additional challenge to the algorithm.

The results for the design condition of $Re = 30000$ are shown in Table 6.1. The fit

Algorithm	Number of Sensors	Mode 1 Fit Training	Mode 2 Fit Training	Mode 1 Fit Validation	Mode 2 Fit Validation
N4SID	1	62	63	67	66
N4SID	2	73	73	78	78
N4SID	6	85	85	81	80
Point POD+N4SID	1	60	61	62	64
Point POD+N4SID	2	61	59	66	63
Point POD+N4SID	6	82	82	82	82
LSTM 1	1	93	94	62	60
LSTM 1	2	94	94	72	71
LSTM 1	6	95	93	77	75
LSTM 2	1	95	95	64	61
LSTM 2	2	96	96	74	74
LSTM 2	6	96	96	79	79

Table 6.1: Training and Validation Fits for the Temporal Coefficients of the First 2 POD Modes.

Algorithm	Number of Sensors	Re=20000 Mode 1 Fit	Re=20000 Mode 2 Fit	Re=40000 Mode 1 Fit	Re=40000 Mode 2 Fit
N4SID	1	56	58	46	47
N4SID	2	64	64	60	59
N4SID	6	74	74	64	63
Point POD+N4SID	1	52	53	51	52
Point POD+N4SID	2	57	56	57	54
Point POD+N4SID	6	77	77	75	74
LSTM 1	1	49	47	51	49
LSTM 1	2	60	58	63	63
LSTM 1	6	66	66	68	67
LSTM 2	1	52	50	48	46
LSTM 2	2	61	60	63	62
LSTM 2	6	70	71	69	70

Table 6.2: Fits for Off-design Conditions for the Temporal Coefficients of the First 2 POD Modes.

quality for the training and validation datasets is reasonably good for all numbers of input points and algorithms, with the validation fit quality increasing with the number of points for all algorithms. The training fit quality also increases with the number of sensors for N4SID and Point POD+N4SID, while being over 93% for all of the LSTM cases. The results for 1 input point using the N4SID algorithm for

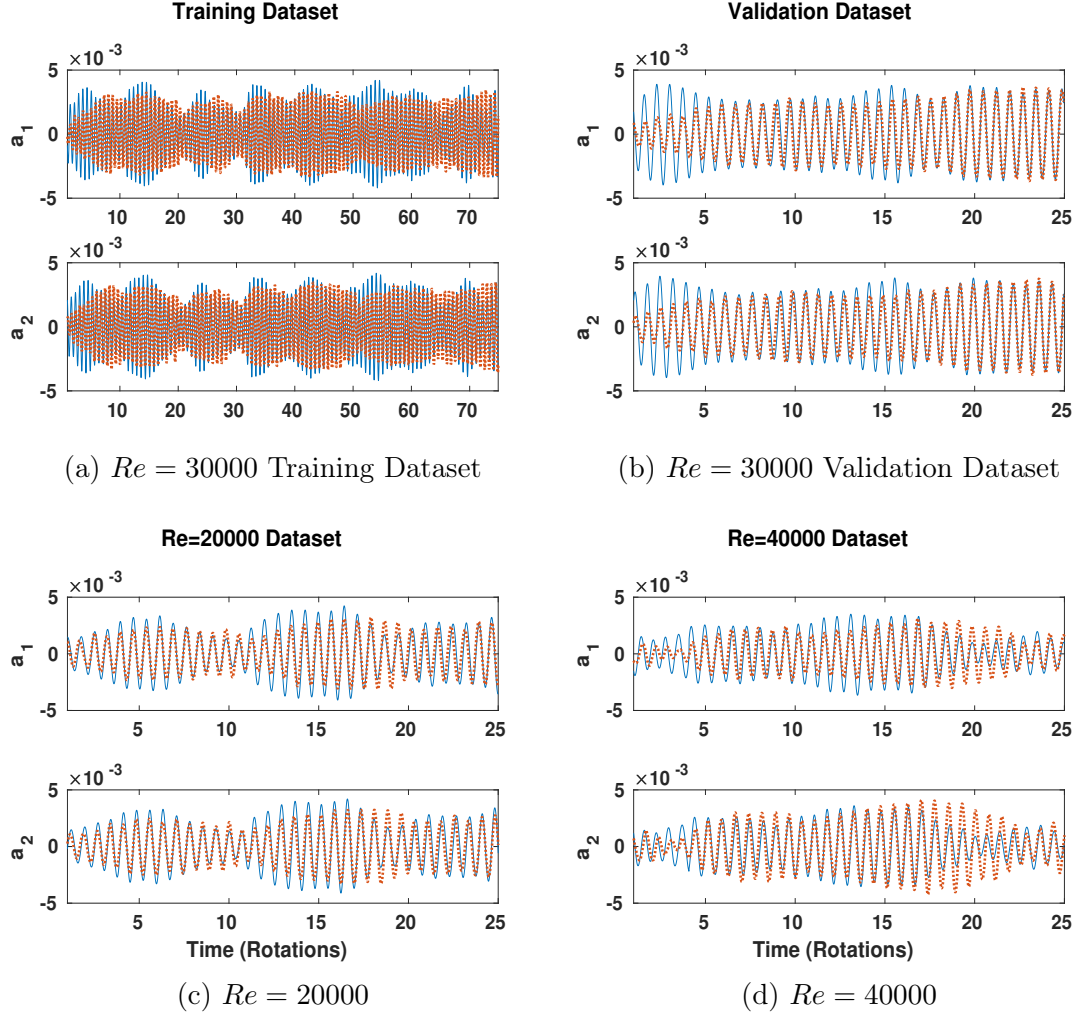


Figure 6.16: N4SID with 1 point at max of first POD mode.

the $Re = 30000$ case are shown in Figure 6.16a and Figure 6.16b. The training fit quality is 62% and 63%, the validation fit quality 67% and 66% for the first and second modes respectively and both the period and amplitude of the temporal POD coefficients are identified well. Similar results are obtained for the other algorithms, demonstrating the ability of machine learning to reconstruct the evolution of the most energetic periodic flow structures with data from a single point even at turbu-

lent flow conditions. These fit values represent a decrease from the $\sim 77\%$ obtained using N4SID with one input point in [120] for $Re = 600$, however, the performance is nevertheless still very good.

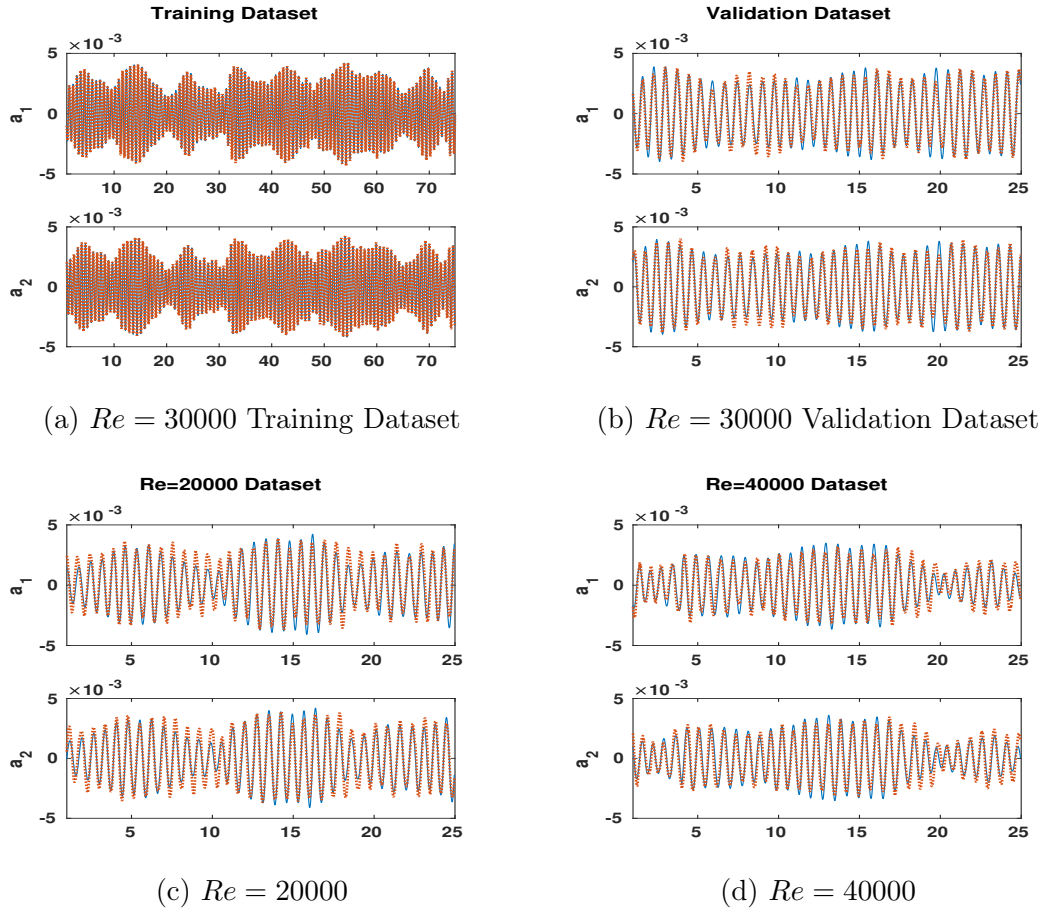


Figure 6.17: LSTM 2 with 6 point input.

The Point POD+N4SID method has lower training and validation fits than the N4SID methods for 1 and 2 input points and broadly similar results for 6 input points. This suggests that for the first two point POD modes to accurately extract the flow features captured by the first two full volume POD modes, input data from 6 points is required.

The results for LSTM 1 and LSTM 2 algorithms are broadly similar, with LSTM 2 having slightly higher fit values than LSTM 1. This agrees with the assumptions made in the derivation of the N4SID model that the relationship between the inputs and outputs can be captured by a linear model, with non-linear terms making minor

contributions. The results for the $Re = 30000$ case for the LSTM 2 algorithm with input data from 6 points are shown in Figure 6.17a and Figure 6.17b. The training dataset is reconstructed near perfectly and while the validation dataset fits are lower, they are nevertheless 79% and reconstruct the temporal POD coefficients very well. The LSTM methods obtain near-perfect training fits for all cases, which can be explained by the large number of tunable parameters in the model. However, the lower validation fits suggest that the model is probably over-fitting. Hence, the validation fit could be improved further by a longer training dataset.

The results for the off-design conditions are shown in Table 6.2. For the majority of the cases considered, the fits for the off-design condition are $\sim 10 - 15\%$ lower than those for the base validation case, with data from 6 input points required to obtain fits of over 70%. The results for the off-design conditions for the N4SID with 1 input point and LSTM 2 with 6 input points are shown in Figures 6.16c and 6.16d and Figures 6.17c and 6.17d respectively. The results for both algorithms, for both Reynolds numbers show a good qualitative fit, even in the case for the N4SID with 1 input point for $Re = 40000$ where the fit values are (46%, 47%) for modes (1,2). For the LSTM 6 input case, the reconstructed time signal almost perfectly matches the true time signal. The best results obtained are with the Point POD+N4SID algorithm with 6 input points. This suggests that there is a good correspondence between the first two modes of the point POD and the first two modes of the full volume POD when 6 input points are used.

The above analysis demonstrates the applicability of the model to reconstruct the flow in a wide range of flow conditions beyond those initially used for training and opening up possibilities of substantial computational savings in the computation of flows dominated by large scale structures. This would allow the modelling of mixing and chemical reactions while accounting for both the mean flow and large scale periodic structures over considerably longer timescales than is currently possible.

Reconstruction of the Full Flow Field

Once the temporal coefficients of the POD modes have been identified, the velocity field can then be computed using

$$\mathbf{u}(\mathbf{x}, t) = \bar{\mathbf{u}} + \mathbf{u}'(\mathbf{x}, t) \approx \sum_{i=0}^{i=n} a_i \phi_i(\mathbf{x}). \quad (6.1)$$

Figures 6.18 and 6.19 show the magnitude of the instantaneous velocity field 20 rotations into the validation dataset for the base $Re = 30000$ case. Figures 6.18a and 6.19a show the flow field computed using LES, and 6.18b and 6.19b show the reconstructed flow field composed of the mean flow and the first two modes, with the temporal coefficients reconstructed from the 2 input point case using the N4SID algorithm. As would be expected, the large scale flow features are reproduced well by the ROM. The downward direction of the jet in the vertical plane and the jet extension in the horizontal plane are predicted well. However, the ROM velocity field is considerably smoother than the LES velocity field, due to the truncation of higher modes that represent small scale turbulence. The maximum in the ROM velocity field is also underpredicted compared to the LES velocity field.

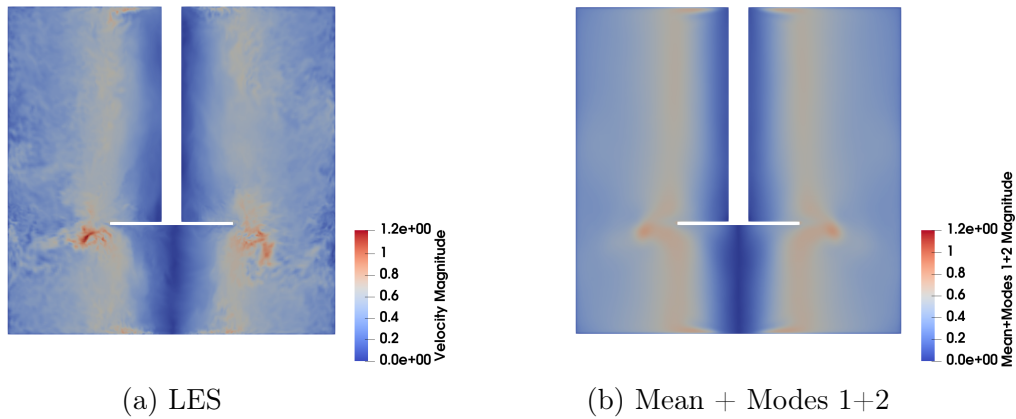


Figure 6.18: Velocity magnitudes of LES and ROM with mean plus first 2 modes using 1 point input N4SID at 20 rotations into validation dataset on vertical plane between two blades.

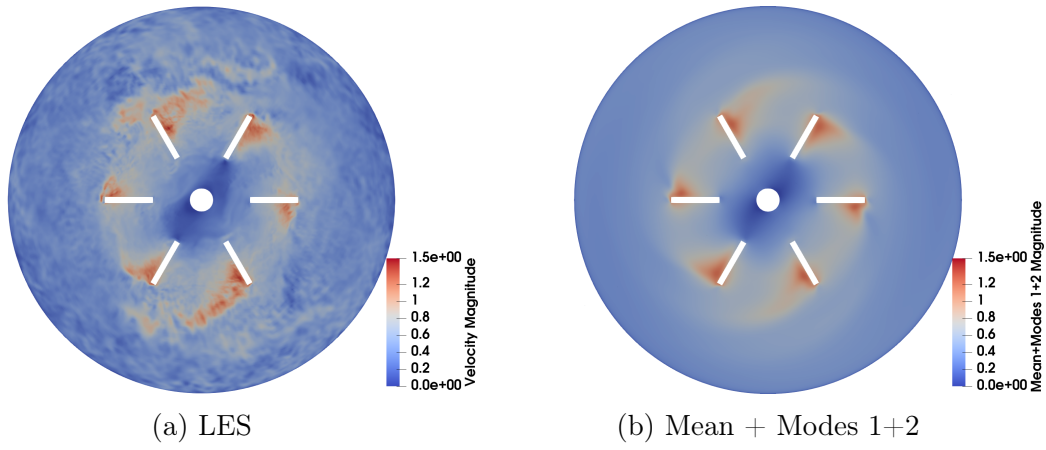


Figure 6.19: Velocity magnitudes of LES and ROM with mean plus first 2 modes using 1 point input N4SID at 20 rotations into validation dataset on a horizontal plane 0.357T from the tank bottom.

6.2.4 Reconstruction of Modes 3 and 4

Temporal Coefficients

We now consider the reconstruction of the 3rd and 4th mode, which although they only carry smaller percentages of the flow energy, nevertheless represent significant periodic components in the flow. Here we take 5000 timesteps as the training dataset and 5000 timesteps as the validation dataset for the base case. While this comes with a reduction in training data, considering the last 50 rotations as a validation dataset gives a more representative test sample due to the high variability in the time series amplitude. We consider input data from 6 points, but this time based upon the location of the maximum of the third POD mode. The N4SID and LSTM models all obtained high fit values for the training dataset, however, they failed to accurately reproduce the evolution of the modes for the validation dataset and for the off-design conditions. The fit for the Point POD + N4SID method has the lowest fit values for the training dataset. The results for this case for the design and off-design conditions are shown in Figure 6.20. Interestingly, despite the low fit values, the general trends in the data are predicted well and there is a reasonable qualitative agreement between the true and reconstructed signal. The reconstructed mode 3

also appears to show more noise than mode 4, suggesting that it corresponds better to the modes of the point POD. Nevertheless, the preprocessing step of performing a POD on the point data and selecting the two leading order modes prior to using the N4SID algorithm appears to allow a reasonable qualitative reconstruction of the signal, despite the failure of the LSTM and N4SID only models. While these results show substantial error, they nevertheless show the potential of this approach. Future improvements to the model, such as a better choice of point locations, an increased number of points and longer training datasets may lead to substantially better results.

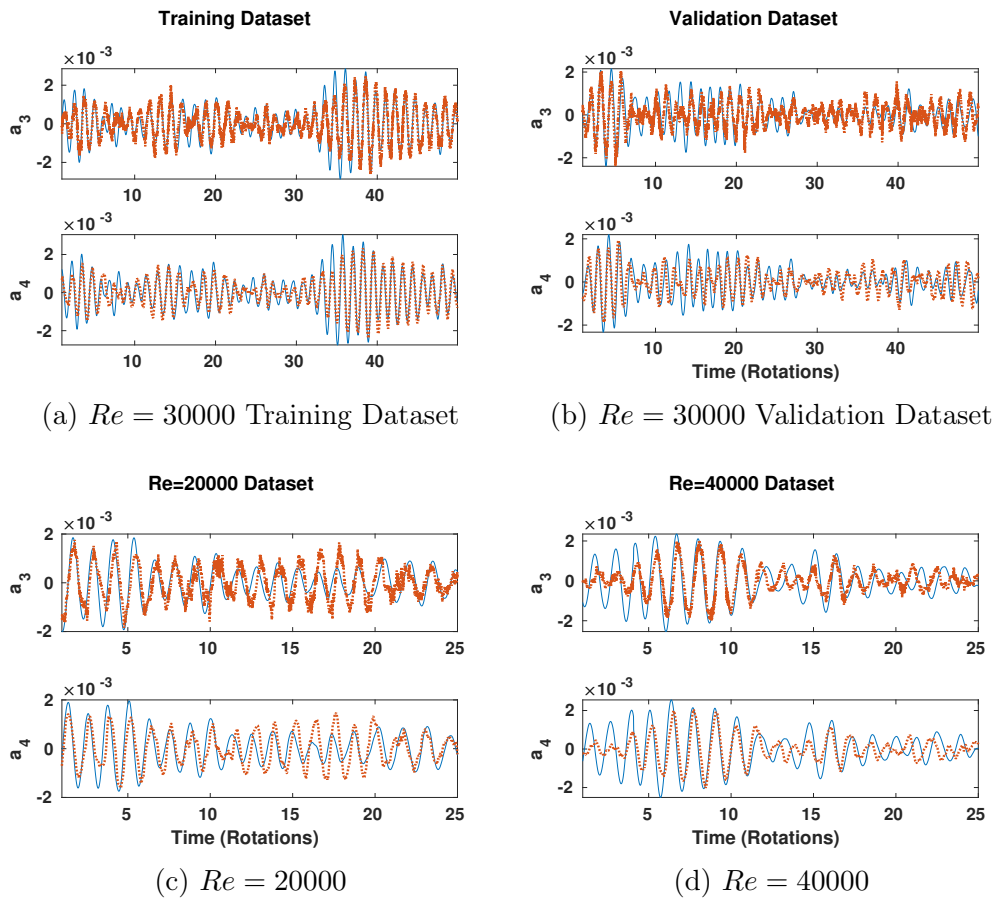


Figure 6.20: N4SID+POD model with 6 inputs for modes 3 and 4.

Flow field

We now attempt to reconstruct the velocity field corresponding to the first 4 POD modes from the reconstructed temporal coefficients of the POD modes using Equa-

tion 6.1. Figure 6.21 shows the magnitude of the instantaneous velocity field at 95 rotations (20 rotations into the validation dataset for modes 1+2 and 45 rotations into the validation dataset for modes 3+4) for the base $Re = 30000$ case computed from the first 4 POD modes. Figure 6.21a shows the LES results and 6.21b shows the reconstruction using a ROM. The first pair of temporal coefficients are reconstructed from the N4SID+POD model with 6 inputs based upon the max of the first POD mode. The second pair of temporal coefficients are reconstructed from the N4SID+POD model with 6 inputs based upon the max of the third POD mode. While the reconstructed flow field underpredicts the magnitude of the fluctuations the results are otherwise near identical.

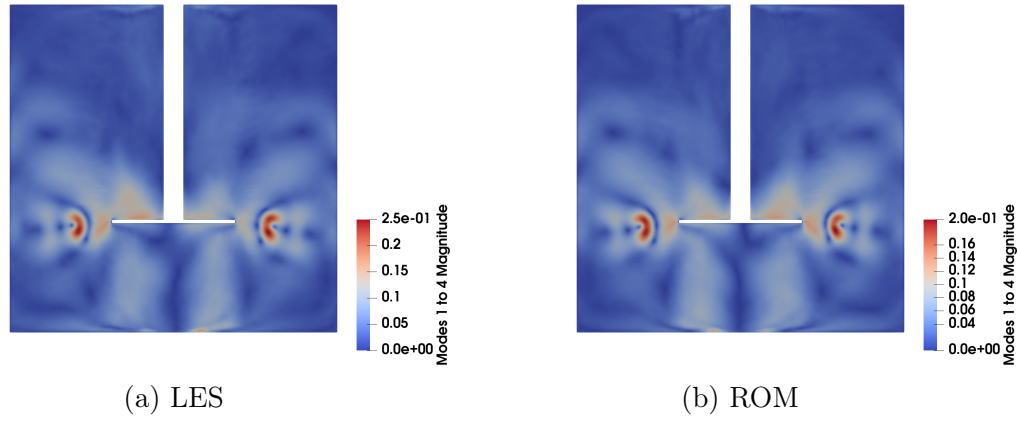


Figure 6.21: Sum of modes 1 to 4 magnitudes of LES and POD+N4SID ROM using 12 point input points at 20 rotations into validation dataset for modes 1+2 and 45 rotations into the validation dataset for modes 3+4.

6.3 Conclusions

In this study, data-driven machine learning methods were applied to the turbulent flow in a stirred tank to identify and reconstruct large scale coherent structures. This work makes two key contributions. First, the characteristics of the large scale structures in an unbaffled stirred tank at a turbulent Reynolds number was analysed. Second, the performance of the reconstruction approach investigated in Chapters 4 and 5 was examined at more industrially relevant Reynolds numbers and compared

to the performance of two other methods Point POD+N4SID and an LSTM neural network approach.

POD was used to extract the coherent structures in the flow. The 4 leading modes have been found to come in two pairs. The first pair accounted for $\sim 5.5\%$ of the TKE, with the next 2 accounting for $\sim 0.9\%$ each. Despite the small energies, these flow structures can account for velocity fluctuations of $\sim \pm 0.4V_{tip}$ in the case of the first pair of modes and $\sim \pm 0.1V_{tip}$ for the second pair of modes.

The first pair of modes were found to have a frequency of $1.4F_N$ and to correspond to a clockwise travelling wave in the rotating frame. In the inertial frame these corresponded to a frequency of $0.6F_N$. Examining the shape of the modes near to the blades suggests that these modes correspond to rotating trailing vortices. On the other hand, the shape of the modes underneath the impeller reveals a similarity to the shape of a precessing MI observed in other studies. In addition, the observed frequency in the inertial frame, while greater than observed in baffled tanks, corresponds well to frequencies of oscillation of the free surface for an unbaffled tank observed in the literature. This indicates that rather than being separate large scale effects, the rotating trailing vortices and precessing MIs are connected. Trailing vortices from different blades appear to interact to form a travelling wave structure further away from the blades.

The second pair of modes were found to have frequencies of $0.8F_N$ and $0.2F_N$ in the rotating and inertial frames respectively. Similarly to the first pair of modes, these were found to correspond to a clockwise travelling wave in the rotating frame. These modes also have a close resemblance to processing MI structure, as observed in [4, 11] and are stronger in the vicinity of the shaft. While again corresponding to a higher frequency than observed for baffled tanks, the values fit well with the frequencies observed for precessing MIs observed in unbaffled tanks.

The first 2 modes were reconstructed using 4 different algorithms: N4SID, Point POD+N4SID and LSTM Neural Networks with one or two hidden layers. The

modes were reconstructed well by all algorithms using velocity data from 1, 2, and 6 sensors. The fit improved with an increase in the number of sensors. The Point POD+N4SID with 6 sensors showed the best performance with fits of $\sim 82\%$ for the base $Re = 30000$ case validation dataset, $\sim 77\%$ for the off-design $Re = 20000$ case, and $\sim 75\%$ for the $Re = 40000$ case. A ROM constructed from the mean and the first 2 modes showed a good reconstruction of the largest features of the flow, but as expected did not reproduced the finer flow features.

The reconstruction of the second pair modes was attempted with all 4 algorithms, but only the Point POD+N4SID approach generated a qualitatively satisfactory model. The reconstruction of the fluctuating velocity field corresponding to the first 4 modes showed a very good agreement to the fluctuating velocity field corresponding to the first 4 modes extracted from the LES.

This study has successfully demonstrated the ability of data-driven machine learning methods to both identify and reconstruct the large scale flow structures in a stirred tank. It has also demonstrated the feasibility of these methods beyond the 2D simple flow cases typically considered in the literature to a 3D fully turbulent flow. Nevertheless, many questions and extensions to this work remain. For example, what is the origin of the beating behaviour observed in the leading order modes? What is the performance of the ROMs in predicting particle distributions and chemical reaction products compared to using just the mean flow? How accurately can the reconstructed flow using a few modes predict the mixing compared to full LES? Can the performance of the reconstruction of the higher order modes be improved by using a longer training dataset and better sensor placement? Can the small scale turbulence be represented in the ROM, for example using a statistical method? How can the optimal location of the sensor be identified?

Chapter 7

Conclusions and Future Work

7.1 Summary of Main Achievements

7.1.1 Reconstruction of Large Scale Flow Structures in a Stirred Tank from Sparse Sensor Data

An unbaffled stirred tank agitated by a six-blade Rushton impeller at a transitional Reynolds number of 600 was simulated using DNS. Approximately 12000 3D velocity field snapshots were collected over approximately 120 impeller rotations. POD was used to identify the large scale structures of the flow. The first 4 modes were found to come in two pairs with frequencies of $1.5F_N$ and $3.0F_N$, respectively, with 14% and 2.6% each of the TKE, respectively. The modes were found to correspond to large vortical structures rotating around the vessel.

The identification algorithm N4SID was used to construct an estimator that captures the relation between the velocity signals at sensor points (input) and the temporal POD coefficients (output) of the first four modes. The first pair of modes were reconstructed very well using data from just a single point at both the design conditions of $Re = 600$ and the off-design conditions of $Re = 500$ and $Re = 700$. This

demonstrated the robustness of the estimator to changes in operating conditions. Increasing the number of points to 6 led to an increase in the accuracy and in the robustness of the reconstruction and also led to a better reconstruction of the second pair of POD modes. To the best of our knowledge, this is the first time that the 3D instantaneous large scale structure flow field inside a stirred tank has been reconstructed using data from a single point.

7.1.2 Decomposition of the Pressure Field and Power Number, and Reconstruction of Large Scale Flow Structures in a Stirred Tank from Blade Surface Pressure Measurements

In this chapter, the work of the previous chapter was extended to consider the pressure decomposition, power number and reconstruction of the velocity and pressure fields from pressure point measurements on the impeller blade and outer wall surfaces.

The same transitional flow cases were considered as in the previous chapter. A further ~ 90 rotations of simulations were performed to give a dataset of ~ 21000 3D velocity field snapshots corresponding to ~ 210 rotations. Applying the POD decomposition on the source term of the pressure Poisson equation, the pressure field was decomposed into a time averaged component, and components that were linear and quadratic with respect to the temporal coefficients of the velocity POD modes. The mean pressure and mean power number were dominated by the term corresponding to the mean velocity.

The power number of each blade was found to contain peaks at the $1.5F_N$ and $3F_N$ frequencies of the leading POD modes. However, when considering the spectrum of the fluctuations in the total power number, these frequencies were very suppressed. Examination of the spectra of the power number of blade combinations found that

the cancellation occurred when 3 adjacent blades were considered. The largest peaks in the spectra of the total power number were found to be at $0.04F_N$ and $0.43F_N$. The former peak matches that of the unpaired 9th velocity POD mode, corresponding to oscillations in the impeller jet. The origin of the latter peak was not clear and not found to be related in a particular POD mode, suggesting that a technique such as DMD or Spectral POD may be required to shed light on its origins. However, the spectrum from the superposition of the first 300 linear POD pressure modes approximates the power number spectrum well. Reducing this to a subset of 30 modes with the largest contributions to the power number among the first 300 linear POD pressure modes also gave a reasonable approximation.

The identification algorithm N4SID was used to construct an estimator that captures the relation between the pressure signals at sensor points (input) and the temporal POD coefficients (output) of the first two modes. This allowed the extension of the methodology of the previous chapter to input data from non-intrusive surface pressure measurements. Using 1 input point located on the impeller blade at the maximum PRMS provided a reasonable reconstruction of the evolution of the first 2 POD modes for the base $Re = 600$ case, albeit with a long spin-up period for the model. Increasing the number of pressure sensors to 6 led to a good fit quality for the base and $Re = 700$ off-design condition case, with fast model spin-up times. While the $Re = 500$ case was qualitatively well reconstructed, the reconstructed signal was found to have a delay relative to the DNS, leading to a reduced fit value. Adding 6 more points located on the outer walls led to a good reconstruction of the first pair of modes at all three transitional Reynolds numbers.

7.1.3 Identification and Reconstruction of MIs in a Turbulent Stirred Tank Using Data-Driven and Machine Learning Methods

In this chapter, the Reynolds numbers considered were increased to $Re = 30000$ for the base case and $Re = 20000$ and 40000 for the off-design conditions. These correspond to more industrially relevant fully turbulent scenarios. A LES simulation was performed, with 10000 snapshots saved for the base $Re = 30000$ case, corresponding to 100 impeller revolutions. The first two POD modes were paired and very similar to those observed for the transitional flow cases, but with a decreased energy, each corresponding to $\sim 5.5\%$ of the TKE and a lower frequency of $1.4F_N$. The third and fourth modes were also paired, with a frequency of $0.8F_N$.

The corresponding frequencies to the first two pairs of modes in the laboratory frame were found to be $0.6F_N$ and $0.2F_N$. Investigation of their spatial components in several locations in the flow suggests that they correspond to precessing MIs observed in other works. However, in contrast to the $\sim 0.01F_N$ values observed for MIs in baffled tanks, these have frequencies an order of magnitude greater. This matches the results from other studies on unbaffled stirred tanks.

The reconstruction of the first two pairs of POD modes from velocity measurements was considered. The performance of the N4SID-based method from the previous chapters was compared against three other architectures. One was a N4SID-based method with an additional preprocessing step of performing a POD on the point data and retaining only the leading 2 modes as an input. The second and third methods were both based on a LSTM neural network, with one and two hidden layers respectively.

For the first pair of modes, velocity input data from 1, 2 and 6 points located at the position of the maximum of the first POD mode was used. The fits for both the design and off-design conditions were generally good for all numbers of sensors

considered for all 4 algorithms. Increasing the number of sensor points led to an increase in the fit values for all architectures. The performance of the examined algorithms was broadly similar, with some slightly out-performing the others for the base case, while under-performing for the off-design conditions. The lack of a substantial difference between the two LSTM NN architectures supported the assumption that a linear model was sufficient for good reconstructions.

For the second pair of modes, velocity input data from 1, 2 and 6 points located at the position of the maximum of the third POD mode was used. The fits were poor for all 4 algorithms. However, the N4SID+POD approach had a reasonably good qualitative agreement with the LES data, with a good reconstruction of the general behaviour of the signals. This suggests that the point POD has the capability to identify and extract periodic signals from noisy data collected at a limited number of data points. On the other hand, the small data size and noisy data lead to both the LSTM NN based approaches and the direct N4SID approach overfitting the training data and being unable to perform well on the validation datasets.

7.2 Future Work

This work has used the POD and several types of supervised learning algorithms to successfully decompose and reconstruct the velocity and pressure fields in an unbaffled tank at both transitional and turbulent Reynolds numbers. Nevertheless many open questions and avenues for future work remain both in respect to the study of flows in a stirred tank and to the application of data-driven and machine learning methods to fluid flows.

The stirred tank configuration explored in this study was an unbaffled, centred configuration with a relatively large impeller size of $T/2$. In particular, it would be interesting to explore the baffled tank case and explore whether the frequencies of the POD modes match the frequencies of the MIs in baffled tanks of $\sim 0.01F_N$.

In addition, the POD method could be used to explore the changes in energy and frequency of the large scale structures as a function of impeller clearance and impeller diameter. Further, the case of the eccentrically placed impeller in an unbaffled tank and whether the energy contained in the large scale periodic structures increases or decreases could be examined.

While POD was chosen as the model order reduction method in this study due to its simplicity and computational efficiency, other model reduction methods could be explored to both improve the reconstruction of the flow and to better understand the dynamical flow structures. In particular DMD or Spectral POD could be used to better identify and extract the periodic structures. Unlike POD which can have multiple frequencies in one mode, the DMD and Spectral POD modes have a single frequency allowing a dynamical feature to be more accurately attributed to an extracted structure. This, for example, would allow the beating behaviour observed in the first pair of modes in the fully turbulent cases to be further investigated and its origin understood. Non-linear autoencoder methods could also be explored as these would be able to more optimally represent the flow features, with more energy captured in a smaller number of modes compared to the POD which required ~ 100 modes to capture 95% of the turbulent kinetic energy.

The effect of the number and location of the probes on the quality of the reconstruction, as well as their optimal location, could be further examined. For example, by adapting the data driven procedure developed in [122], or through further heuristic point placement choices.

The reconstruction of the flow field could also be improved by representing the effect of the small scales. For example, the unresolved modes could be represented by a random flow field as suggested in [134, 135] or by using a machine learning based methods as has been suggested for adding smaller scale turbulent flow features to a coarse grained flow field [136, 137].

The performance of the reconstructed flow field in the simulation of scalars dispersed

in the flow and of chemical reactions such as precipitation and aggregation. The mixing time for flow fields corresponding to the mean flow, mean flow and particular periodic structures and the full flow field could be examined by allowing the impact of the previously observed macro, meso and micro mixing in a more quantitative fashion.

Finally, the sensor data used in this study was all generated from CFD simulations. An important extension of this work would be to use experimentally collected data using pressure or temperature probes, or PIV. This line of research would bring the techniques explored in this study closer to use in an industrial setting. The additional challenge of more noisy experimental data could lead to modifications of the methodology discussed here to account for this. For example, through additional data cleaning steps, the use of more (and perhaps optimally placed) input locations or the concurrent use of pressure and velocity input data.

Bibliography

- [1] F. Scargiali, A. Tamburini, G. Caputo, and G. Micale, “On the assessment of power consumption and critical impeller speed in vortexing unbaffled stirred tanks,” *Chem. Eng. Res. Des.*, vol. 123, pp. 99–110, 2017.
- [2] J. Rushton, “Power characteristics of mixing impellers part 1,” *Chem. Eng. Prog.*, vol. 46, pp. 395–404, 1950.
- [3] M. Yianneskis, Z. Popielek, and J. H. Whitelaw, “An experimental study of the steady and unsteady flow characteristics of stirred reactors,” *J. Fluid Mech.*, vol. 175, pp. 537–555, feb 1987.
- [4] Z. Doulgerakis, M. Yianneskis, and A. Ducci, “On the interaction of trailing and macro-instability vortices in a stirred vessel-enhanced energy levels and improved mixing potential,” *Chem. Eng. Res. Des.*, vol. 87, no. 4, pp. 412–420, 2009.
- [5] J. Baldyga and J. Bourne, “Interactions between mixing on various scales in stirred tank reactors,” *Chem. Eng. Sci.*, vol. 47, no. 8, pp. 1839–1848, 1992.
- [6] J. Baldyga, J. Bourne, and S. Hearn, “Interaction between chemical reactions and mixing on various scales,” *Chem. Eng. Sci.*, vol. 52, no. 4, pp. 457–466, 1997.
- [7] A. Liné, J.-C. Gabelle, J. Morchain, D. Anne-Archard, and F. Augier, “On POD analysis of PIV measurements applied to mixing in a stirred vessel with

- a shear thinning fluid,” *Chem. Eng. Res. Des.*, vol. 91, no. 11, pp. 2073–2083, 2013.
- [8] A. de Lamotte, A. Delafosse, S. Calvo, and D. Toye, “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.*, vol. 178, pp. 348–366, 2018.
- [9] A. de Lamotte, A. Delafosse, S. Calvo, and D. Toye, “Identifying dominant spatial and time characteristics of flow dynamics within free-surface baffled stirred-tanks from CFD simulations,” *Chem. Eng. Sci.*, vol. 192, pp. 128–142, 2018.
- [10] J.-C. Gabelle, J. Morchain, and A. Liné, “Kinetic energy transfer between first proper orthogonal decomposition modes in a mixing tank,” *Chem. Eng. Technol.*, vol. 40, no. 5, pp. 927–937, 2017.
- [11] A. Ducci, Z. Doulgerakis, and M. Yianneskis, “Decomposition of flow structures in stirred reactors and implications for mixing enhancement,” *Ind. Eng. Chem. Res.*, vol. 47, no. 10, pp. 3664–3676, 2008.
- [12] J.-C. Gabelle, J. Morchain, D. Anne-Archard, F. Augier, and A. Liné, “Experimental determination of the shear rate in a stirred tank with a non-Newtonian fluid: Carbopol,” *AIChE J.*, vol. 59, no. 6, pp. 2251–2266, 2013.
- [13] J. Moreau and A. Liné, “Proper orthogonal decomposition for the study of hydrodynamics in a mixing tank,” *AIChE J.*, vol. 52, no. 7, pp. 2651–2655, 2006.
- [14] G. Janiga, “Large-eddy simulation and 3D proper orthogonal decomposition of the hydrodynamics in a stirred tank,” *Chem. Eng. Sci.*, vol. 201, pp. 132 – 144, 2019.

- [15] S. Başbuğ, G. Papadakis, and J. C. Vassilicos, “DNS investigation of the dynamical behaviour of trailing vortices in unbaffled stirred vessels at transitional reynolds numbers,” *Phys. Fluids*, vol. 29, no. 6, p. 064101, 2017.
- [16] K. V. Sharp, D. Hill, D. Troolin, G. Walters, and W. Lai, “Volumetric three-component velocimetry measurements of the turbulent flow around a rushton turbine,” *Exp. Fluids*, vol. 48, no. 1, pp. 167–183, 2010.
- [17] Z. Chara, B. Kysela, J. Konfrst, and I. Fort, “Study of fluid flow in baffled vessels stirred by a rushton standard impeller,” *Appl. Math. Comput.*, vol. 272, pp. 614–628, 2016.
- [18] C. A. Coulaloglou and L. L. Tavlarides, “Drop size distributions and coalescence frequencies of liquid-liquid dispersions in flow vessels,” *AIChE J.*, vol. 22, no. 2, pp. 289–297, 1976.
- [19] E. L. Paul, V. A. Atiemo-Obeng, and S. M. Kresta, *Handbook of industrial mixing: science and practice*, vol. 1. John Wiley & Sons, 2003.
- [20] J. Baldyga, W. Podgórska, and R. Pohorecki, “Mixing-precipitation model with application to double feed semibatch precipitation,” *Chem. Eng. Sci.*, vol. 50, no. 8, pp. 1281–1300, 1995.
- [21] Z. Jaworski and A. W. Nienow, “CFD modelling of continuous precipitation of barium sulphate in a stirred tank,” *Chem. Eng. J.*, vol. 91, no. 2, pp. 167–174, 2003. Chemreactor - 15 S.I.
- [22] Y. Chisti, “Animal cell culture in stirred bioreactors: observations on scale-up,” *Bioprocess Eng.*, vol. 9, no. 5, pp. 191–196, 1993.
- [23] H. A. Jakobsen, *Chemical reactor modeling - Multiphase Reactive Flows 2nd ed.* Cham. Springer, 2012.
- [24] S. Basbug, *Flow field and mixing in stirred vessels with regular and fractal impellers*. PhD thesis, Imperial College London, 2017.

- [25] K. Steiros, P. J. K. Bruce, O. R. H. Buxton, and J. C. Vassilicos, “Effect of blade modifications on the torque and flow field of radial impellers in stirred tanks,” *Phys. Rev. Fluids*, vol. 2, p. 094802, Sep 2017.
- [26] J. M. Vasconcelos, S. C. Orvalho, A. M. Rodrigues, and S. S. Alves, “Effect of blade shape on the performance of six-bladed disk turbine impellers,” *Ind. Eng. Chem. Res.*, vol. 39, no. 1, pp. 203–213, 2000.
- [27] H. Unadkat, C. D. Rielly, and Z. K. Nagy, “Piv study of the flow field generated by a sawtooth impeller,” *Chem. Eng. Sci.*, vol. 66, no. 21, pp. 5374–5387, 2011.
- [28] S. Nagata, *Mixing: principles and applications*. Halsted press, 1975.
- [29] R. L. Bates, P. L. Fondy, and R. R. Corpstein, “Examination of some geometric parameters of impeller power,” *Ind. Eng. Chem. Process Des. Dev.*, vol. 2, no. 4, pp. 310–314, 1963.
- [30] S. Başbuğ, G. Papadakis, and J. C. Vassilicos, “Reduced mixing time in stirred vessels by means of irregular impellers,” *Phys. Rev. Fluids*, vol. 3, p. 084502, 2018.
- [31] S. Başbuğ, G. Papadakis, and J. C. Vassilicos, “Reduced power consumption in stirred vessels by means of fractal impellers,” *AIChE J.*, vol. 64, no. 4, pp. 1485–1499, 2018.
- [32] A. Tamburini, G. Gagliano, G. Micale, A. Brucato, F. Scargiali, and M. Ciofalo, “Direct numerical simulations of creeping to early turbulent flow in unbaffled and baffled stirred tanks,” *Chem. Eng. Sci.*, vol. 192, pp. 161–175, 2018.
- [33] A. Tamburini, A. Brucato, M. Ciofalo, G. Gagliano, G. Micale, and F. Scargiali, “CFD simulations of early- to fully-turbulent conditions in unbaffled and baffled vessels stirred by a Rushton turbine,” *Chem. Eng. Res. Des.*, vol. 171, pp. 36–47, 2021.

- [34] C. Galletti and E. Brunazzi, "On the main flow features and instabilities in an unbaffled vessel agitated with an eccentrically located impeller," *Chem. Eng. Sci.*, vol. 63, no. 18, pp. 4494–4505, 2008.
- [35] C. Galletti, S. Pintus, and E. Brunazzi, "Macro-instabilities in Eccentrically Agitated Vessels," *Chem. Eng. Trans.*, vol. 12, pp. 483–488, 2009.
- [36] C. Galletti, S. Pintus, and E. Brunazzi, "Effect of shaft eccentricity and impeller blade thickness on the vortices features in an unbaffled vessel," *Chem. Eng. Res. Des.*, vol. 87, no. 4, pp. 391 – 400, 2009.
- [37] F. L. Yang, S. J. Zhou, and G. C. Wang, "Experimental study and detached eddy simulation of the macro-instability in an eccentric stirred tank," *Appl. Mech. Mater.*, vol. 66-68, pp. 20–26, 2011.
- [38] M. Assirelli, W. Bujalski, A. Eaglesham, and A. W. Nienow, "Macro- and micromixing studies in an unbaffled vessel agitated by a rushton turbine," *Chem. Eng. Sci.*, vol. 63, no. 1, pp. 35–46, 2008.
- [39] F. Scargiali, A. Busciglio, F. Grisafi, and A. Brucato, "Mass transfer and hydrodynamic characteristics of unbaffled stirred bio-reactors: Influence of impeller design," *Biochem. Eng. J.*, vol. 82, pp. 41–47, 2014.
- [40] A. Busciglio, F. Grisafi, F. Scargiali, and A. Brucato, "Mixing dynamics in uncovered unbaffled stirred tanks," *Chem. Eng. J.*, vol. 254, pp. 210–219, 2014.
- [41] K. Rutherford, S. Mahmoudi, K. Lee, and M. Yianneskis, "The influence of rushton impeller blade and disk thickness on the mixing characteristics of stirred vessels," *Chem. Eng. Res. Des.*, vol. 74, no. 3, pp. 369–378, 1996.
- [42] I. Takashima and M. Mochizuki, "Tomographic observations of the flow around agitator impeller," *J. Chem. Eng. Jpn.*, vol. 4, no. 1, pp. 66–72, 1971.

- [43] K. Van't Riet and J. M. Smith, "The behaviour of gas—liquid mixtures near rushton turbine blades," *Chem. Eng. Sci.*, vol. 28, no. 4, pp. 1031–1037, 1973.
- [44] K. V. Riet and J. M. Smith, "The trailing vortex system produced by Rushton turbine agitators," *Chem. Eng. Sci.*, vol. 30, no. 9, pp. 1093 – 1105, 1975.
- [45] K. Van't Riet, W. Bruijn, and J. M. Smith, "Real and pseudo-turbulence in the discharge stream from a rushton turbine," *Chem. Eng. Sci.*, vol. 31, no. 6, pp. 407–412, 1976.
- [46] C. M. Stoots and R. V. Calabrese, "Mean velocity field relative to a rushton turbine blade," *AIChE J.*, vol. 41, no. 1, pp. 1–11, 1995.
- [47] K. Lee and M. Yianneskis, "Turbulence properties of the impeller stream of a rushton turbine," *AIChE J.*, vol. 44, no. 1, pp. 13–24, 1998.
- [48] R. Escudié, D. Bouyer, and A. Liné, "Characterization of trailing vortices generated by a Rushton turbine," *AIChE J.*, vol. 50, no. 1, pp. 75–86, 2004.
- [49] R. Escudié and A. Liné, "A simplified procedure to identify trailing vortices generated by a Rushton turbine," *AIChE J.*, vol. 53, no. 2, pp. 523–526, 2007.
- [50] A. Paglianti, G. Montante, and F. Magelli, "Novel experiments and a mechanistic model for macroinstabilities in stirred tanks," *AIChE J.*, vol. 52, no. 2, pp. 426–437, 2006.
- [51] A. Busciglio, F. Scargiali, F. Grisafi, and A. Brucato, "Oscillation dynamics of free vortex surface in uncovered unbaffled stirred vessels," *Chem. Eng. J.*, vol. 285, pp. 477–486, 2016.
- [52] T. Yamamoto, Y. Fang, and S. V. Komarov, "Surface vortex formation and free surface deformation in an unbaffled vessel stirred by on-axis and eccentric impellers," *Chem. Eng. J.*, vol. 367, pp. 25–36, 2019.
- [53] A. Ducci and M. Yianneskis, "Vortex tracking and mixing enhancement in stirred processes," *AIChE J.*, vol. 53, no. 2, pp. 305–315, 2007.

- [54] M. Assirelli, W. Bujalski, A. Eaglesham, and A. Nienow, "Study of micromixing in a stirred tank using a rushton turbine: comparison of feed positions and other mixing devices," *Chem. Eng. Res. Des.*, vol. 80, no. 8, pp. 855–863, 2002.
- [55] M. Assirelli, W. Bujalski, A. Eaglesham, and A. Nienow, "Intensifying micromixing in a semi-batch reactor using a rushton turbine," *Chem. Eng. Sci.*, vol. 60, no. 8-9, pp. 2333–2339, 2005.
- [56] S. B. Pope, *Turbulent Flows*. Cambridge University Press, 2000.
- [57] J. G. Eggels, "Direct and large-eddy simulation of turbulent fluid flow using the lattice-boltzmann scheme," *Int. J. Heat Fluid Flow.*, vol. 17, no. 3, pp. 307–323, 1996.
- [58] J. Derksen and H. Van den Akker, "Parallel simulation of turbulent fluid flow in a mixing tank," in *High-Performance Computing and Networking* (P. Sloot, M. Bubak, and B. Hertzberger, eds.), pp. 96–104, Springer Berlin Heidelberg, 1998.
- [59] J. Derksen and H. E. A. Van den Akker, "Large eddy simulations on the flow driven by a rushton turbine," *AIChE J.*, vol. 45, no. 2, pp. 209–221, 1999.
- [60] S. Yeoh, G. Papadakis, and M. Yianneskis, "Determination of mixing time and degree of homogeneity in stirred vessels with Large Eddy Simulation," *Chem. Eng. Sci.*, vol. 60, no. 8-9, pp. 2293–2302, 2005.
- [61] H. Hartmann, J. Derksen, and H. Van den Akker, "Mixing times in a turbulent stirred tank by means of les," *AIChE J.*, vol. 52, no. 11, pp. 3696–3706, 2006.
- [62] J. Derksen, "Numerical simulation of solids suspension in a stirred tank," *AIChE J.*, vol. 49, no. 11, pp. 2700–2714, 2003.

- [63] N. Lamarque, B. Zoppé, O. Lebaigue, Y. Dolias, M. Bertrand, and F. Ducros, “Large-eddy simulation of the turbulent free-surface flow in an unbaffled stirred tank reactor,” *Chem. Eng. Sci.*, vol. 65, no. 15, pp. 4307–4322, 2010.
- [64] H. Hartmann, J. Derksen, and H. Van den Akker, “Macroinstability uncovered in a rushton turbine stirred tank by means of les,” *AIChE J.*, vol. 50, no. 10, pp. 2383–2393, 2004.
- [65] J. J. Gillissen and H. E. A. Van den Akker, “Direct numerical simulation of the turbulent flow in a baffled tank driven by a Rushton turbine,” *AIChE J.*, vol. 58, no. 12, pp. 3878–3890, 2012.
- [66] J. J. Derksen, “Direct simulations of mixing of liquids with density and viscosity differences,” *Ind. Eng. Chem. Res.*, vol. 51, no. 19, pp. 6948–6957, 2012.
- [67] K. Fukami, K. Fukagata, and K. Taira, “Super-resolution reconstruction of turbulent flows with machine learning,” *J. Fluid Mech*, vol. 870, pp. 106–120, 2019.
- [68] J. J. Aguilar-Fuertes, F. Noguero-Rodríguez, J. C. Jaen Ruiz, L. M. García-RAffi, and S. Hoyas, “Tracking Turbulent Coherent Structures by Means of Neural Networks,” *Energies*, vol. 14, no. 4, 2021.
- [69] J. Guzmán-Inigo, M. A. Sodar, and G. Papadakis, “Data-based, reduced-order, dynamic estimator for reconstruction of nonlinear flows exhibiting limit-cycle oscillations,” *Phys. Rev. Fluids*, vol. 4, p. 114703, Nov 2019.
- [70] 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*, vol. 31, no. 7, 2019.
- [71] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, “Machine Learning for Fluid Mechanics,” *Annu. Rev. Fluid Mech.*, vol. 52, pp. 477–508, 2020.

- [72] J. Lumley, “The structure of inhomogeneous turbulent flows,” in *Atmospheric Turbulence and Radio Wave Propagation* (M. Yaglom and V. Tatarski, eds.), pp. 166–178, Nauka Moscow, 1967.
- [73] L. Sirovich, “Turbulence and the dynamics of coherent structures. II. Symmetries and transformations,” *Q. Appl. Math.*, vol. 45, no. 3, pp. 573–582, 1987.
- [74] P. Baldi and K. Hornik, “Neural networks and principal component analysis: Learning from examples without local minima,” *Neural Netw.*, vol. 2, no. 1, pp. 53–58, 1989.
- [75] Y. Liang, H. Lee, S. Lim, W. Lin, K. Lee, and C. Wu, “Proper orthogonal decomposition and its applications—part I: Theory,” *J. Sound Vib.*, vol. 252, no. 3, pp. 527–544, 2002.
- [76] K. Taira, S. L. Brunton, S. T. M. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, “Modal analysis of fluid flows: An overview,” *AIAA Journal*, vol. 55, no. 12, pp. 4013–4041, 2017.
- [77] P. Hasal, I. Fort, and J. Kratena, “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. Des.*, vol. 82, no. 9, pp. 1268–1281, 2004.
- [78] L. Nikiforaki, G. Montante, K. Lee, and M. Yianneskis, “On the origin, frequency and magnitude of macro-instabilities of the flows in stirred vessels,” *Chem. Eng. Sci.*, vol. 58, no. 13, pp. 2937–2949, 2003.
- [79] C. Mayorga, J. Morchain, and A. Liné, “Reconstruction of the 3d hydrodynamics in a baffled stirred tank using proper orthogonal decomposition,” *Chem. Eng. Sci.*, vol. 248, p. 117220, 2022.

- [80] C. W. Rowley, “Model reduction for fluids, using balanced proper orthogonal decomposition,” *Int. J. Bifurc. Chaos Appl. Sci. Eng.*, vol. 15, no. 03, pp. 997–1013, 2005.
- [81] P. J. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *J. Fluid Mech.*, vol. 656, pp. 5–28, 2010.
- [82] C. W. Rowley, I. Mezic, S. Bagheri, P. Schlatter, and D. S. Henningson, “Spectral analysis of nonlinear flows,” *J. Fluid Mech.*, vol. 641, p. 115–127, 2009.
- [83] K. Taira, S. L. Brunton, S. T. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, “Modal analysis of fluid flows: An overview,” *AIAA Journal*, pp. 4013–4041, 2017.
- [84] M. S. Hemati, M. O. Williams, and C. W. Rowley, “Dynamic mode decomposition for large and streaming datasets,” *Phys. Fluids*, vol. 26, no. 11, p. 111701, 2014.
- [85] A. Towne, O. T. Schmidt, and T. Colonius, “Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis,” *J. Fluid Mech.*, vol. 847, pp. 821–867, 2018.
- [86] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
- [87] L. Agostini, “Exploration and prediction of fluid dynamical systems using auto-encoder technology,” *Phys. Fluids*, vol. 32, no. 6, p. 067103, 2020.
- [88] K. Hasegawa, K. Fukami, T. Murata, and K. Fukagata, “Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes,” *Theor. Comput. Fluid Dyn.*, vol. 34, no. 4, pp. 367–383, 2020.
- [89] V. Hernandez, J. E. Roman, and V. Vidal, “SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems,” *ACM Trans. Math. Software*, vol. 31, no. 3, pp. 351–362, 2005.

- [90] B. R. Noack, P. Papas, and P. A. Monkewitz, “The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows,” *J. Fluid Mech.*, vol. 523, pp. 339–365, 2005.
- [91] B. R. Noack, K. Afanasiev, M. Morzyński, G. Tadmor, and F. Thiele, “A hierarchy of low-dimensional models for the transient and post-transient cylinder wake,” *J. Fluid Mech.*, vol. 497, pp. 335–363, 2003.
- [92] J. Guzmán-Iñigo, D. Sipp, and P. J. Schmid, “A dynamic observer to capture and control perturbation energy in noise amplifiers,” *J. Fluid Mech.*, vol. 758, pp. 728–753, 2014.
- [93] P. Van Overschee and B. De Moor, “N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems,” *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.
- [94] P. Van Overschee and B. L. De Moor, *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science & Business Media, 2012.
- [95] S. Qin, “An overview of subspace identification,” *Comp. & Chem. Eng.*, vol. 30, no. 10, pp. 1502–1513, 2006.
- [96] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [97] A. T. Mohan and D. V. Gaitonde, “A Deep Learning based Approach to Reduced Order Modeling for Turbulent Flow Control using LSTM Neural Networks,” *arXiv*, 2018.
- [98] K. H. Manohar, C. Morton, and P. Ziadé, “Sparse sensor-based cylinder flow estimation using artificial neural networks,” *Phys. Rev. Fluids*, vol. 7, p. 024707, Feb 2022.

- [99] J. B. Joshi, N. K. Nere, C. V. Rane, B. Murthy, C. S. Mathpati, A. W. Patwardhan, and V. V. Ranade, “CFD simulation of stirred tanks: Comparison of turbulence models. Part I: Radial flow impellers,” *The Can. J. Chem. Eng.*, vol. 89, no. 1, pp. 23–82, 2011.
- [100] S. Yeoh, G. Papadakis, and M. Yianneskis, “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. Des.*, vol. 82, no. 7, pp. 834–848, 2004.
- [101] H. Yao, F. Alves-Portela, and G. Papadakis, “Evolution of conditionally averaged second-order structure functions in a transitional boundary layer,” *Phys. Rev. Fluids*, vol. 5, p. 093902, 2020.
- [102] N. Thomareis and G. Papadakis, “Resolvent analysis of separated and attached flows around an airfoil at transitional Reynolds number,” *Phys. Rev. Fluids*, vol. 3, p. 073901, 2018.
- [103] 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, 2018.
- [104] R. D. Falgout and U. M. Yang, “Hypre: A library of high performance preconditioners,” in *Computational Science — ICCS 2002* (P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, and J. J. Dongarra, eds.), (Berlin, Heidelberg), pp. 632–641, Springer Berlin Heidelberg, 2002.
- [105] A. W. Vreman, “An eddy-viscosity subgrid-scale model for turbulent shear flow: Algebraic theory and applications,” *Phys. Fluids*, vol. 16, no. 10, pp. 3670–3681, 2004.

- [106] R. Alcamo, G. Micale, F. Grisafi, A. Brucato, and M. Ciofalo, “Large-eddy simulation of turbulent flow in an unbaffled stirred tank driven by a rushton turbine,” *Chem. Eng. Sci.*, vol. 60, no. 8-9, pp. 2303–2316, 2005.
- [107] J. Ramírez-Cruz, M. Salinas-Vázquez, G. Ascanio, W. Vicente-Rodríguez, and C. Lagarza-Córtés, “Mixing dynamics in an uncovered unbaffled stirred tank using Large-Eddy Simulations and a passive scalar transport equation,” *Chem. Eng. Sci.*, vol. 222, 2020.
- [108] A. Vella, C., Grisafi, F., Micale, G., Rizzuti, L., Brucato, “Near impeller flow field in an unbaffled stirred tank,” *Proceedings of the 11th European Conference on Mixing*, pp. 629–636, 2003.
- [109] A. Tamburini, G. Gagliano, F. Scargiali, G. Micale, A. Brucato, and M. Ciofalo, “CFD simulation of radially stirred baffled and unbaffled tanks,” *Chem. Eng. Trans.*, vol. 74, no. November 2018, pp. 1033–1038, 2019.
- [110] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics, Cambridge University Press, 2 ed., 2012.
- [111] V. Hernández, J. E. Román, and A. Tomás, “A robust and efficient parallel SVD solver based on restarted Lanczos bidiagonalization,” *Electron. Trans. Numer. Anal.*, vol. 31, pp. 68–85, 2008.
- [112] T. Kailath, B. Hassibi, and A. H. Sayed, *Linear estimation*. Prentice-Hall International, 2000.
- [113] P. Van Overschee and B. De Moor, “A unifying theorem for three subspace system identification algorithms,” *Automatica*, vol. 31, no. 12, pp. 1853–1864, 1995.
- [114] S. Discetti, M. Raiola, and A. Ianiro, “Estimation of time-resolved turbulent fields through correlation of non-time-resolved field measurements and time-

- resolved point measurements,” *Exp. Therm. Fluid Sci.*, vol. 93, pp. 119–130, 2018.
- [115] S. Discetti, G. Bellani, R. Örlü, J. Serpieri, C. Sanmiguel Vila, M. Raiola, X. Zheng, L. Mascotelli, A. Talamelli, and A. Ianiro, “Characterization of very-large-scale motions in high-re pipe flows,” *Exp. Therm. Fluid Sci.*, vol. 104, pp. 1–8, 2019.
- [116] L. Guastoni, A. Güemes, A. Ianiro, S. Discetti, P. Schlatter, H. Azizpour, and R. Vinuesa, “Convolutional-network models to predict wall-bounded turbulence from wall quantities,” *J. Fluid Mech*, vol. 928, p. A27, 2021.
- [117] A. Graves, *Long Short-Term Memory*. Berlin, Heidelberg: Springer, 2012.
- [118] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” 2016.
- [119] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
- [120] K. Mikhaylov, S. Rigopoulos, and G. Papadakis, “Reconstruction of large-scale flow structures in a stirred tank from limited sensor data,” *AIChE J.*, vol. 67, no. 10, p. e17348, 2021.
- [121] H. Hartmann, J. Derksen, C. Montavon, J. Pearson, I. Hamill, and H. van den Akker, “Assessment of large eddy and rans stirred tank simulations by means of lda,” *Chem. Eng. Sci.*, vol. 59, pp. 2419–2432, June 2004.

- [122] F. Savarino and G. Papadakis, “Reconstruction of irregular flow dynamics around two square cylinders from sparse measurements using a data-driven algorithm,” *arXiv preprint arXiv:2209.03290*, 2022.
- [123] J. Jeong and F. Hussain, “On the identification of a vortex,” *J. Fluid Mech.*, vol. 285, pp. 69–94, 1995.
- [124] P. A. Davidson, *Turbulence: an introduction for scientists and engineers*. Oxford university press, 2015.
- [125] R. L. Sani, J. Shen, O. Pironneau, and P. M. Gresho, “Pressure boundary condition for the time-dependent incompressible Navier–Stokes equations,” *Int. J. Numer. Methods Fluids*, vol. 50, no. 6, pp. 673–682, 2006.
- [126] J. L. Guermond, P. Mineev, and J. Shen, “An overview of projection methods for incompressible flows,” *Comput. Methods Appl. Mech. Eng.*, vol. 195, no. 44, pp. 6011–6045, 2006.
- [127] G. L. Lane, G. D. Rigby, and G. M. Evans, “Pressure distribution on the surface of Rushton turbine blades-experimental measurement and prediction by CFD,” *J. Chem. Eng. Jpn.*, vol. 34, no. 5, pp. 613–620, 2001.
- [128] K. Rave, M. Lehmenkühler, D. Wirz, H.-J. Bart, and R. Skoda, “3d flow simulation of a baffled stirred tank for an assessment of geometry simplifications and a scale-adaptive turbulence model,” *Chem. Eng. Sci.*, vol. 231, p. 116262, 2021.
- [129] S. Roy and S. Acharya, “Perturbed turbulent stirred tank flows with amplitude and mode-shape variations,” *Chem. Eng. Sci.*, vol. 66, no. 22, pp. 5703–5722, 2011.
- [130] S. Roy and S. Acharya, “Effect of Impeller Speed Perturbation in a Rushton Impeller Stirred Tank,” *J. Fluids Eng.*, vol. 134, 06 2012. 061104.

- [131] V. Lavezzo, R. Verzicco, and A. Soldati, “Ekman pumping and intermittent particle resuspension in a stirred tank reactor,” *Chem. Eng. Res. Des.*, vol. 87, no. 4, pp. 557–564, 2009.
- [132] K. Steiros, P. J. K. Bruce, O. R. H. Buxton, and J. C. Vassilicos, “Power consumption and form drag of regular and fractal-shaped turbines in a stirred tank,” *AIChE J.*, vol. 63, no. 2, pp. 843–854, 2017.
- [133] A. Tamburini, A. Brucato, A. Busciglio, A. Cipollina, F. Grisafi, G. Micale, F. Scargiali, and G. Vella, “Solid-liquid suspensions in top-covered unbaffled vessels: Influence of particle size, liquid viscosity, impeller size, and clearance,” *Ind. Eng. Chem.*, vol. 53, no. 23, pp. 9587–9599, 2014.
- [134] E. Mémin, “Fluid flow dynamics under location uncertainty,” *Geophys. Astrophys. Fluid Dyn.*, vol. 108, no. 2, pp. 119–146, 2014.
- [135] V. Resseguier, E. Mémin, and B. Chapron, “Geophysical flows under location uncertainty, part I random transport and general models,” *Geophys. Astrophys. Fluid Dyn.*, vol. 111, no. 3, pp. 149–176, 2017.
- [136] B. Liu, J. Tang, H. Huang, and X.-Y. Lu, “Deep learning methods for super-resolution reconstruction of turbulent flows,” *Phys. Fluids*, vol. 32, no. 2, p. 025105, 2020.
- [137] K. Fukami, K. Fukagata, and K. Taira, “Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows,” *J. Fluid Mech*, vol. 909, p. A9, 2021.