University of Naples Federico II
Naples, Italy

On the data-driven reduced order modelling in fluid dynamics

Antonio Colanera

Thesis for the Degree of
Doctor of Philosophy
in Industrial Engineering
Cycle XXXVI

Supervisors:
prof. Luigi de Luca
dr. Matteo Chiatto

Coordinator:
prof. Michele Grassi

October 2023

To my family

Il Punto Interrogativo è il simbolo del Bene,
così come quello Esclamativo è il simbolo del Male.

Luciano De Crescenzo

On the data-driven reduced order modelling in fluid dynamics

Ph.D. Thesis presented
for the purpose of obtaining the Degree of Doctor of Philosophy
in Industrial Engineering
by

Antonio Colanera

October 2023

Supervisors:
prof. Luigi de Luca
dr. Matteo Chiatto

University of Naples Federico II
Ph.D. Program in Industrial Engineering
XXXVI cycle - Coordinator: prof. Michele Grassi

Abstract

Fluid dynamics is characterized by several physical phenomena, leading to complex spatial and temporal structures characterized by different spatial and temporal scales. This thesis is a comprehensive exploration of data-driven approaches for modal analysis, stability analysis and reduced-order modeling in fluid dynamics, aiming to enhance the understanding of complex flow phenomena.

As regarding the modal analysis, this work delves into the extraction of coherent flow structures using conventional techniques like Spectral Proper Orthogonal Decomposition (SPOD) and introduces innovative approaches like Robust SPOD to deal with noisy or corrupt data and Gappy POD for handling two-phase PIV measurements. These methods are applied to various flow configurations, including the vertical liquid jet, the turbulent jet, the open cavity flow and the two-phase mixing layer.

A data-driven approach to estimate the global spectrum of gravitational liquid jet is presented. The underlying linear operator has been extracted with the Dynamic Mode Decomposition (DMD), considering random perturbations of the base flow. This analysis has shed light on sinuous and varicose modes, their interaction, and the influence of the main governing parameters.

Conventional and novel Reduced-order models (ROM) are presented, including Extended Cluster-based Network Modeling (eCNM) and functional based CNM. These methods offer efficient ways to capture flow dynamics, forecast fluid behaviours and handle undersampled data.

This thesis advances data-driven approaches in fluid dynamics, providing valuable tools for the comprehension of complex flow phenomena.

Keywords: Data-Driven analysis, Reduced order modelling, Stability, Modal decomposition.

Contents

Abstract ii
Acknowledgements vii
List of Publications vii
List of Figures xiv
List of Tables xxvii

1 Introduction **1**
1.1 Modal analysis 2
1.2 Stability analysis 4
1.3 Reduced order modelling 6
1.4 Outline 8

2 Modal analysis **9**
2.1 Spectral Proper Orthogonal Decomposition 10
2.1.1 Application to vertical liquid jet 12
2.1.2 Application to 3D vertical liquid jet 37
2.2 Robust SPOD 43
2.2.1 Differences between the proposed algorithms . . 49
2.2.2 Application of RSPOD to the turbulent jet 50

2.2.3 Application of RSPOD to the open cavity flow . 63
2.3 Gappy POD of two-phase PIV data 71

3 Data driven Stability analysis **77**
3.1 Data-driven global stability analysis 78
3.1.1 1D global analysis 81
3.1.2 2D analysis 87

4 Reduced Order Modelling **107**
4.1 Cluster-based approaches 108
4.1.1 Standard Clustering 108
4.1.2 Extended Clustering 110
4.1.3 Cluster-based Network Model 114
4.1.4 Clustering distance metric and parameters choice 117
4.1.5 Toy model: SEIRD Model for Disease Spread . . 121
4.2 Application to swirl flame 124
4.2.1 Model flow: Swirl Flame 125
4.2.2 CNM analysis 129
4.2.3 Extended CNM of the heat release 134
4.2.4 Extended CNM analysis on spatial subdomains . 137
4.2.5 Extended CNM analysis for temporally undersampled data reconstruction 140
4.3 Comparison between CNM and POD-ROM 144
4.3.1 Full order model of an airfoil configuration . . . 144
4.3.2 POD-ROM 146
4.3.3 Results 147
4.4 Trajectory based CNM 151
4.4.1 Functions clustering 155

4.4.2 Filtering approaches 156
4.4.3 Trajectories network model 160
4.4.4 Validation 162
4.5 Functional CNM results 164
4.5.1 Landau's system 164
4.5.2 Fluidic Pinball 176

5 Conclusions **185**
5.1 Modal analysis 186
5.2 Stability analysis 188
5.3 Reduced order modelling 189
5.4 Recommendations for future work 191

A Insights into liquid curtains **193**
A.1 Numerical simulation layout 195
A.2 Linear stability analysis 200

B Functional data analysis **203**
B.1 Functional principal component analysis 203
B.2 Basis expansions 205
B.2.1 Short time Fourier transform 206
B.2.2 B-splines decomposition 207
B.2.3 Wavelet Decomposition 208

References **211**

Acknowledgements

The study presented in this dissertation has been financially supported by Università degli Studi di Napoli "Federico II", and it has been carried out under the supervision of prof. Luigi de Luca and dr. Matteo Chiatto. Part of the work has been done at Harbin Institute of Technology (Shenzhen, China), in collaboration with prof. Bernd R. Noack. Moreover a period has been spent in Technische Universität Berlin (Berlin, Germany) under the supervision of prof. Kilian Oberleithner. The candidate is deeply grateful to his Supervisors for the human and scientific guidance they have provided during the whole time of the Ph.D. activity.

List of Publications

Published journal articles:

- A. Colanera, A. Della Pia and M. Chiatto, *Data-driven global stability of vertical planar liquid jets by dynamic mode decomposition on random perturbations*, Physics of Fluids, 34, 012355 (2022);
- A. Della Pia, A. Colanera and M. Chiatto, *Surface tension-induced instability in spatially developing subcritical liquid curtains*, Physics of Fluids, 34, 042122 (2022);
- A. Della Pia, A. Colanera, M. Chiatto and L. de Luca, *Energy insights into the unsteady dynamics of a viscous gravitational liquid sheet*, Physics of Fluids, 33, 092118 (2021);
- A. Colanera, A. Della Pia, M. Chiatto, L. de Luca and F. Grasso, *Modal decomposition analysis of unsteady viscous liquid sheet flows*, Physics of Fluids, 33, 092117 (2021);

Journal articles under review/in preparation:

- A. Colanera, O.T. Schmidt and M. Chiatto, *Robust spectral proper orthogonal decomposition*, in preparation for Journal of Computational Physics

- A. Colanera, M. Reumschuessel, J. Beuth, M. Chiatto, L. de Luca and K. Oberleithner, *Extended cluster-based network modelling for coherent structures in turbulent flows*, in preparation for Theoretical and Computational Fluid Dynamics

- A. Colanera, N. Deng, M. Chiatto, L. de Luca and B. R. Noack, *Functional cluster-based network modelling*, in preparation for Journal of Fluid Mechanics

- A. Colanera, E. Di Costanzo, M. Chiatto and L. de Luca, *Reduced order model approaches for airfoil performances prediction*, in preparation for Fluids

Conference proceedings:

- A. Colanera, A. Della Pia, M. Chiatto and L. de Luca, *Computation of liquid jets BiGlobal spectrum via randomly perturbed numerical data*, 2023 AIAA Aviation and Aeronautics Forum and Exposition (AIAA AVIATION Forum), Manchester Grand Hyatt, San Diego, California & Online (USA), 12-16 June 2023;

- M. Acquaviva, A. Della Pia, A. Colanera, M. Chiatto and L. de Luca, *Numerical investigation of hole propagation in a three-dimensional*

liquid curtain, 2023 AIAA Aviation and Aeronautics Forum and Exposition (AIAA AVIATION Forum), Manchester Grand Hyatt, San Diego, California & Online (USA), 12-16 June 2023;

- E. Di Costanzo, A. Colanera, M. Chiatto and L. de Luca, *Cluster-based network reduced order modeling for flow fields around airfoil profiles*, 2023 AIAA Aviation and Aeronautics Forum and Exposition (AIAA AVIATION Forum), Manchester Grand Hyatt, San Diego, California & Online (USA), 12-16 June 2023;

- A. Colanera, M. Chiatto and L. de Luca, *Reduced Order Modelling of Multiphase Fluid Flows*, 14^{th} European Fluid Mechanics Conference, EFMC14 Conference (2022), Athens (Greece), 13-16 September 2022;

- A. Della Pia, L. Grande, A. Colanera, M. Chiatto and L. de Luca, *Volume-of-fluid simulation of three-dimensional gravitational liquid sheet flows*, European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2022), Oslo (Norway), 5-9 June 2022;

- A. Colanera, A. Della Pia, M. R. Acquaviva, M. Chiatto and L. de Luca, *Modal analysis of a 3D gravitational liquid sheet*, European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2022), Oslo (Norway), 5-9 June 2022;

- A. Colanera, M. Chiatto and L. de Luca, *Data-driven reduced order modelling for parametrized simulations of vertical liquid sheet flow*, 9^{th} International Symposium on Scale Modelling, ISSM9 Conference (2021), Naples (Italy), 2-4 March 2022;

- A. Colanera, M. Chiatto and L. de Luca, *POD and SPOD analysis of vertical liquid sheet.*, Associaione Italiana Di Aeronautica ed Astronautica AIDAA Conference (2021), Pisa (Italy) 31 August - 3 September 2021.

List of Figures

2.1 Sketch of physical and computational domains of gravitational liquid jet 13

2.2 SPOD spectrum in supercritical conditions 15

2.3 Leading SPOD modes in supercritical conditions 17

2.4 Sketch of sinuous disturbance 18

2.5 Supercritical conditions (case (a) of Table A.1). (a) u' mode, (b) v' mode, (c) C' mode at $St \approx 0.01$. 19

2.6 POD analysis of forced vertical liquid jet in supercritical conditions 20

2.7 SPOD spectrum in subcritical conditions 21

2.8 Leading SPOD modes in subcritical conditions 22

2.9 SPOD spectrum in subcritical conditions at several Reynolds numbers 23

2.10 Cumulative energy distribution of SPOD modes at various Reynolds numbers. 24

2.11 Sketch of varicose disturbance 25

2.12 Decomposition in sinuous and varicose components. . . 26

2.13 Normalized SPOD spectra at various forcing Strouhal numbers St_f 27
2.14 Spatial distribution of the leading SPOD modes at $St = 0.137$ and $St = 0.182$ 28
2.15 Cumulative energy distribution of SPOD modes at various forcing Strouhal numbers. 28
2.16 Normalized SPOD spectra at various Weber numbers We 29
2.17 Spatial distribution of the 1^{st} SPOD mode at $St = 0.076$ and $St = 0.152$ 30
2.18 Energy content $(\%)$ of the sinuous and varicose components 32
2.19 Distributions of the first SPOD mode at the leading frequency ($St = 0.076$) and its first harmonic ($St = 0.152$) 33
2.20 Comparison of the original interface shape and the reconstructed one with SPOD 35
2.21 Subcritical regime (case (j)) of Table A.1). SPOD reconstruction error. 36
2.22 Subcritical regime (case (e)) of Table A.1). SPOD reconstruction error. 37
2.23 Subcritical regime (case (j)) of Table A.1). Comparison of POD- and SPOD-based reconstruction of the liquid sheet shape. Dashed red line, POD; solid blue line, SPOD. 38
2.24 3D vertical liquid jet: snapshot of interface shape for $We = 2.5$ and 0.75. 39
2.25 SPOD analysis of 3D vertical liquid jet. 41
2.26 Effect of the Weber number on the 3D vertical liquid jet. 42

2.27 Comparison between SPOD and robust SPOD reconstructions of a toy signal. In panel (a) is reported the analyzed signal. Panel (b) contains the perturbed signal with salt and pepper noise. Panel (c) reports the SPOD reconstruction by the whole spectrum of 1^{st} SPOD mode. Panel (d) and (e) contain the robust SPOD reconstruction with de-noising stage before and past FFT, respectively. 50
2.28 Mean pressure distribution in the turbulent jet 51
2.29 SPOD spectrum for the turbulent jet case. Red shaded area highlights the separation between the first and second mode. 52
2.30 Real part of 1^{st} and 2^{nd} mode at different St. Modes have been normalized with respect their maximum. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j and radial coordinate r/D_j, respectively, omitted for clarity. (blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$) 53
2.31 Comparison between SPOD and robust SPOD leading modes at different St. Salt and pepper noise with the 10% of corrupted data points. Panels in first line refer to SPOD modes of corrupted data. Other panels show the effect of the de-noising parameter α on the robust SPOD modes. Data corruption modelled as salt and pepper noise $\pm 4p_{\max}$. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j and radial coordinate r/D_j, respectively, omitted for clarity. Field variables have been normalized with respect to their maximum. (blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$) 55

2.32 Comparison between SPOD and robust SPOD leading modes at different St. Gaussian noise with the 10% of corrupted data points. Panels in first line refer to SPOD modes of corrupted data. Other panels show the effect of the de-noising parameter α on the robust SPOD modes. Data corruption modelled as Gaussian noise with $\pm 4 p_{\max}$ variance. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j and radial coordinate r/D_j, respectively, omitted for clarity. Field variables have been normalized with respect to their maximum. (blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$) 56

2.33 Relative error ε_{f_k} computed with SPOD and robust SPOD techniques at $St = 0.4$ for different levels and types of noise, considering several values of α. Solid lines represent the ε_{f_k} at several α, color-coded according to the corrupted data points percentage. Dashed lines represent the SPOD error, namely α approaching infinity. 57

2.34 Effect of de-noising parameter α on the error ε. Panels (a) to (d) goes from the 1^{st} to the 4^{th} SPOD mode. Salt and pepper noise with 10% of corrupted data points. 58

2.35 Effect of de-noising parameter α on the robust SPOD spectrum. Panels (a) refers to 1^{st} SPOD mode, panel (b) to the 2^{nd} one. Black lines represent the robust SPOD spectra and red lines the original SPOD spectrum. Dashed magenta lines are the SPOD spectra of corrupted data. Salt and pepper noise with 10% of corrupted data points. . . . 59

2.36 De-noising with robust SPOD. First panel contains a sample snapshot of $\hat{p}_0$. Second panel shows the 10% of corrupted data points with salt and pepper noise. Third panel reports an attempt of de-noising with SPOD by employing the whole spectrum of the first 3 modes. Lower panels show the robust SPOD reconstruction by employing the whole spectrum of all the robust SPOD modes, highlighting the de-noising effect of the parameter α. 60
2.37 Comparison between robust SPOD algorithms using the turbulent jet data with a syntetic cosine signal at $x/D_j = 0.3$. Salt and pepper noise with the 0.2% of corrupted data points has been considered. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j ($0 < x/D_j < 1$) and radial coordinate r/D_j ($0 < r/D_j < 0.7$), respectively, omitted for clarity. The first panel shows the leading SPOD mode at $St = 1.2$ near the inlet section. The second panel displays the same SPOD mode obtained from the corrupt dataset. The lower panels illustrate the application of the two robust SPOD algorithms with the same denoising parameter $\alpha = 2$. Field variables have been normalized with respect to their maximum. (■ ■ blue, white, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$) 62
2.38 Open cavity flows. Geometrical sketch; panel and mean distributions. 64
2.39 SPOD spectrum of open cavity flows 65
2.40 Leading SPOD modes of open cavity flows 66
2.41 Comparison between the SPOD spectrum and the robust SPOD one for the cavity flow. 67

2.42 Open cavity: comparison between SPOD and robust SPOD leading u modes at different St. 68
2.43 Open cavity: comparison between SPOD and robust SPOD leading v modes at different St. 69
2.44 Spatial standard deviation of the velocity mode gradient magnitude. Comparison between SPOD and robust SPOD. 70
2.45 Visualization of the splitter plate experimental setup . . . 72
2.46 Gappy POD analysis for the splitter plate PIV data. . . . 74
2.47 PSD of the leading modes temporal coefficients for the splitter plate. 75

3.1 Comparison between the supercritical spectrum evaluated with the linear stability analysis and DMD approach. . . 84
3.2 Comparison between leading LST eigenfunctions and DMD modes. 85
3.3 Convergence of DMD algorithm. 86
3.4 Comparison between sheet centerline snapshots of the forced configuration and the same results obtained by means of the DMD approach. 88
3.5 Evolution of a random initial disturbance. 89
3.6 Data driven BiGlobal spectrum. 90
3.7 Spatial distributions of the DMD modes. 92
3.8 Varicose component of DMD modes 94
3.9 Energy distribution for the modes. 95
3.10 Convergence of DMD algorithm. 96
3.11 Effect of the We number on the global spectrum in supercritical regime. 97

3.12 DMD spectra comparison between supercritical and subcritical regimes. 99
3.13 Varicose components of DMD modes. 100
3.14 Leading lateral displacement modes. 101
3.15 Density ratio r_ρ effect on DMD spectrum. 103
3.16 Varicose components of the DMD leading mode 104
3.17 Spatial distribution of the local norm E_x. 106
3.18 DMD reconstruction for the C field, considering the leading unstable mode only. 106

4.1 Comparison of Standard and Extended Clustering Techniques. 112
4.2 Schematic representation of a directed network of clusters. 115
4.3 Time evolution of the SEIRD model. The simulation showcases the dynamics of Susceptible ($\mathscr{S}$), Exposed ($\mathscr{E}$), Infected ($\mathscr{I}$), Recovered ($\mathscr{R}$), and Deceased ($\mathscr{D}$) populations over time. 123
4.4 Comparison of Infectious population dynamics using original SEIRD model, standard CNM, Extended eCNM, eCNM with undersampled $\mathscr{I}$. 124
4.5 Mean (Time-averaged) flow. Panel (a)-(b): mean axial ($\bar{u}$) and transversal ($\bar{v}$) velocity components. Panel (c): mean turbulent kinetic energy (TKE) distribution. Panel (d): Abel deconvoluted temporal mean of heat release distribution $\bar{q}$. Blue lines are the $\bar{q}/\bar{q}_{\max} = 0.5$ isolines. All variables are normalized with respect to their respective maxima. 127
4.6 Pressure signals of a swirl flame with PVC. 128

4.7 Swirl flame with PVC: Phase portraits of the three leading temporal modes for various N_f values. 130
4.8 Bayesian information criterion (BIC) for parameter selection. 131
4.9 Clustered phase potrait. 132
4.10 CNM's direct transition probabilities (Q_{ij}) and times (T_{ij}) matrices. 133
4.11 Spatial distribution of the backward finite time Lyapunov exponent (FTLE) for selected velocity centroids. The figure also includes an overlay of heat release cluster centroids. Each distribution has been normalized with respect to its maximum. 134
4.12 Three-dimensional reconstruction of heat release rate centroids, representing the dynamic behavior of the flame. This visualization combines the Abel-deconvoluted symmetric mean flame with symmetric (Abel-deconvoluted) and antisymmetric heat release rate fluctuations, representing the azimuthal wavenumber $\alpha = 1$ component. In all panels, the isosurfaces $q/q_{\max} = 0.5$ are depicted in semi-transparent red, while $q/q_{\max} = 0.8$ in yellow. . . . 135
4.13 Reconstruction error $\varepsilon(t)$. 136
4.14 Autocorrelation function R for velocity components and I. 137
4.15 Spatial subdomain for extended CNM. 138
4.16 Comparison between ε for the full domain and sub-domain. 139
4.17 Leading local frequency. 139
4.18 CNM of pressure azimuthal modes. 141
4.19 Extended cluster centroids for v' velocity component derived using temporally undersampled data. 142

4.20 Analysis of the reconstruction error for temporally undersampled velocity data. 143
4.21 Comparison between mean reconstruction error of uniform and random sampling. 144
4.22 Mean flow and leading POD modes for a NACA 0012 at $\alpha = 8°$ and $Re_\infty = 7000$. 147
4.23 Comparison between FOM, POD-ROM, and CNM-ROM snapshots. 149
4.24 Comparison between FOM, POD-ROM, and CNM-ROM force coefficients. 150
4.25 Comparison of the autocorrelation function R between the FOM, POD-ROM, and CNM-ROM. 151
4.26 Prediction of the effect of α using CNM-ROM. 152
4.27 Sketch of functional clustering on an ensemble of n two-dimensional trajectories presenting noise. 153
4.28 Schematic representation of a directed network of clusters in the trajectory-based CNM. 161
4.29 Phase space of the Landau equations 165
4.30 Phase space plots with the true dynamics of the Landau system. 166
4.31 Demonstration of the Trajectory-based CNM with $K = 10$ trajectory clusters. 167
4.32 Comparison of different Cluster Network Model (CNM) variants in the three-dimensional space of a_1, a_2 and σ for the Landau's system. 169
4.33 Comparison of representation error E_r for different Cluster Network Model (CNM) variants. 170
4.34 Analysis of a time-modulated signal. 171

4.35 Comparison between original b_1 data, standard CNM, and STFT-CNM. 171
4.36 RMSE values vs K. 172
4.37 Comparison of spectrograms for the original b_1 signal, the standard CNM, and the STFT CNM. 173
4.38 Normalized PSD of trajectory cluster centroids. 174
4.39 Temporal dynamics and the corresponding phase space for even-numbered centroids $\tilde{\mathbf{c}}_i$. 175
4.40 Direct transition probability matrix Q for standard CNM and STFT-CNM. 176
4.41 Comparison of the normalized unbiased autocorrelation function for several models. 177
4.42 Numerical grid and a sample snapshot for fluidic pinball at $Re = 160$. 178
4.43 Phase portrait of force coefficients at different values of Re. 179
4.44 Comparison between standard CNM and functional based CNM with Fourier basis. 181
4.45 Leading POD modes in terms of vorticity for the post-transient solution of the fluidic pinball at $Re = 160$. All the field variables have been normalized with respect to its maximum. 182
4.46 Comparison between the reconstructions of leading POD temporal coefficients using conventional CNM and functional-based CNM with Fourier basis. 183

A.1 Typical numerical arrangement and sample output of a supercritical case. 197
A.2 Supercritical conditions: Mean flow and generic snapshot. 199

A.3 Typical eigenvalue spectrum in supercritical conditions. . 201

B.1 The 'db10' wavelet and its decomposition of a time series signal. 209

List of Tables

2.1 Sinuonus and varicose energy content, for subcritical vertical liquid jet. 32
2.2 Cumulative energy ratio ER and mean reconstruction error $\bar{\varepsilon}$. 35

3.1 Selected eigenvalues of the vertical liquid jet for a supercritical case. 90
3.2 Sinuous and varicose energy contributions of selected DMD modes. 93
3.3 Comparison of the leading dimensionless frequency of the DMD and 1D model at various supercritical Weber numbers. 98
3.4 Sinuous and varicose energy contents of the first two leading eigenvalues. 98
3.5 Effect of the r_ρ on the leading frequency. 105

A.1 Analyzed cases of the vertical liquid jet. 198

Chapter 1

Introduction

Fluid dynamics involves quantities that vary in space and time that are driven by several physical phenomena, leading to complex spatial and temporal structures characterized by different spatial and temporal scales. Recent fluid dynamics research is therefore the identification of ordered motion within field data and the analysis of coherent structures.

These structures, often seen as organized patterns, can impact the transport of momentum, heat, and mass in a fluid flow (Hussain [1]). In engineering systems, they can significantly affect the operation by causing noise and vibrations (Assi et al. [2]), affecting fluid mixing (Bobusch et al. [3]) and perturbing chemical reactions within the flow (Poinsot et al. [4]). As an example, the shedding of vortices from a cylinder, the von Kármán vortex street, can induce vibrations in structures (Ausoni et al. [5]). In aerodynamics, coherent structures can affect drag and noise (Pujals et al. [6], Bastin et al. [7]). It is therefore crucial to understand and control coherent structures in order to increase efficiency of technical systems (Nair et al. [8]).

The emergence of data-driven methods in fluid dynamics can be attributed to the remarkable advancements in computational power. As computational techniques have evolved, they have paved the way for novel approaches in data analysis (Taira et al. [9]), simulations (Hesthaven et al. [10]), and flow control (Brunton and Noack [11]). In particular, the synergy between computational simulations and experimental measurements

has enabled the acquisition of large-scale, high-fidelity data in the field of modal analysis. The compression of a vast amount of flow field data to a low-dimensional form is ever more important in studying complex fluid flows and in developing models for understanding and modeling their dynamical behavior. Furthermore, the application of global stability analysis, which investigates the linear stability of both two- and three-dimensional base flows, has recently undergone extensive review and is in a continuous phase of expansion (Theofilis [12]). Additionally, reduced order modelling, where the primary objective is to reduce the computational complexity of fluid dynamic models while maintaining a high level of accuracy, can facilitate the creation of efficient surrogates for complex systems (Noack et al. [13]).

This thesis deals with the employment and development of data-driven approaches for modal analysis, stability analysis and reduced order modelling.

1.1 Modal analysis

Modal decomposition techniques represent a powerful tool to gain physical insights on fluid dynamics problems through the identification and the analysis (in terms of spatio-temporal evolution) of the main coherent structures of the flow (Taira et al. [14]).

Among the modal analysis techniques, spectral proper orthogonal decomposition (SPOD) has gained much interest in the last years due to its ability to extract the main spectral features of a flow field, optimally capturing the two-point space–time correlations, thus providing modes that evolve coherently in space and time (Schmidt et al. [15], Towne et al. [16]). SPOD algorithm has been employed in a wide range of fluid dynamics applications, mainly for the post-processing of numerical and experimental data. This technique is a valuable tool to investigate the flow topology itself (Chiatto et al. [17], Nidhan et al. [18], He et al. [19], Chu et al. [20], Abreu et al. [21]) and even to evaluate the flow receptivity and the effectiveness of an applied control strategy (Chiatto et al. [22]).

In this framework, the topology and the dynamics of the vertical liquid jet, harmonically forced at inlet section, has been analyzed by means of the SPOD analysis. Also the three dimensional (3D) configuration has been studied. This part of the work has been carried out in collaboration with prof. Francesco Grasso of the Princeton University (Princeton, USA).

Nowadays, one of the most used experimental measurement techniques is the particle image velocimetry (PIV), which can suffer from erroneous measurements due to inadequate illumination, optical issues, reflections and sharp gradients in field properties (Huang et al. [23], Mendez et al. [24]). Experimental measurements have to deal with the trade-off between the quantity and quality of PIV data; thus, acquired flow fields often have corrupt and even missing measurements. Standard filtering and reconstruction techniques involve interpolation methods, that rely only on local flow information (Westerweel and Scarano [25]), least square methods and Kriging (Oliver and Webster [26]).

The problem of gappy measurements has traditionally been addressed with approaches based on proper orthogonal decomposition, POD (Everson and Sirovich [27], Venturi and Karniadakis [28], Gunes et al. [29]). Recently, Nekkanti and Schmidt [30] have demonstrated the gappy SPOD capability in reconstructing flow fields. Even machine learning techniques such as Physics-informed neural networks, PINNs (von Saldern et al. [31]), deep generative adversarial model, deep-GAN (Buzzicotti et al. [32]) and autoencoders (Discetti and Liu [33]) can be efficiently employed to generate missing data in turbulence and processing PIV data. On the other hand, the problem of the presence of outliers in data, whose location is not known a priori, was efficiently faced using matrix completion, which separates the sparse noise from meaningful data (Candes et al. [34]) through different algorithmic approaches such as augmented Lagrange multiplier (ALM) and alternating directions method (ADM) (Yuan and Yang [35] and Lin et al. [36]). These techniques have been proven helpful for experimental data post-processing and modal decompositions (Scherl et al. [37]).

Such limitations are overcome by a novel approach here proposed, the robust spectral proper orthogonal decomposition (robust SPOD), which

implements the robust principal component analysis within the SPOD technique. The new technique is firstly presented with details on its algorithm and its effectiveness is tested on two different fluid dynamics problems: the subsonic jet flow field numerically simulated and the flow within an open cavity analyzed experimentally. This novel technique has been developed in collaboration with prof. Oliver T. Schmidt of the University of San Diego (San Diego, USA).

Furthermore, two-phase gappy PIV measurements of the two-phase mixing layer past a splitter plate have been analyzed by means of a tailored alghoritm inspired by the Gappy POD (Gunes et al. [29]).

1.2 Stability analysis

The global stability analysis focuses on the temporal and spatial evolution of small-amplitude perturbations superimposed upon a base flow (Huerre and Monkewitz [38]). This method offers insight into the eigenvalues spectrum of the system, which helps us identify the onset of flow fluctuations and instabilities. Consequently, it enables us to predict the dominant frequencies within the flow field.

The recent increase in computing power encouraged the development of many data-driven analysis tools (Hou et al. [39], Ma et al. [40]); in particular, the rediscovery of Koopman spectral analysis by Mezic [41] and ROWLEY et al. [42] traced the road-map for the formulation of the Dynamic Modes Decomposition (DMD) by Schmid [43] and Tu et al. [44]. DMD generates the best linear fit data-driven model and provides structures ranked according to the dynamics of the system. Its application field is quite wide including the analysis of fluid flows configurations (Schmid [45]), face recognition (Tirunagari et al. [46]), neuroscience (Brunton et al. [47]) and infectious diseases (Proctor and Welkhoff [48]).

Following the approach outlined by Ranjan et al. [49] for the data-driven stability analysis of one-phase flows, in the present work this method is extended to two-phase flows. The global spectra of viscous gravitational jets is estimated by applying the DMD technique on randomly perturbed

base flow.

Vertical planar gravitational liquid jets (sheet or curtain flows) are an interesting industrial class of flows, whose modeling has been investigated since the middle of the last century. Typical applications are coating layer deposition (Weinstein and Ruschak [50]) and paper production (Soderberg [51]); the ability to control and eventually suppress undesired oscillations (or non-uniformities) in the steady configurations is fundamental for the final quality of an industrial process. Vertical curtain flows are also related to the nappe configuration, i.e. the flow of waterfalls generated by the overflow of water over long crests of dams or weirs (Binnie [52], Girfoglio et al. [53], Lodomez et al. [54]); one of the major drawback of these flows is related to the onset of self-sustained oscillations which can lead to significant noise levels and structural damages.

In recent years there has been a renewed interest towards the unsteady dynamics of the gravitational liquid sheets also promoted by the possibility of performing direct numerical simulations of this relevant two-phase flow field (Schmidt and Oberleithner [55], Arote et al. [56, 57], Agbaglah [58], Liu et al. [59],Torsey et al. [60]).

Due to gravitational effects, the flow field is not parallel and the determination of the oscillation frequencies of the liquid jet involves a global problem essentially governed by the Weber number. The global stability analysis has been until now theoretically carried out by using a linear one-dimensional (1D) model (Della Pia et al. [61], Della Pia et al. [62]) and taking advantage from an energy budget approach (Della Pia et al. [63]). A recent application of data-driven techniques to liquid jets has been presented by Arote et al. [57], who focused on the effect of the Reynolds number on the topology of the flow; moreover, Colanera et al. [64] and Della Pia et al. [65] explored the effect of a continuous harmonic perturbation enforced at the inlet section on the main coherent structures of the unsteady field and the related temporal frequencies. Recently, Schmidt and Oberleithner [55] studied the stability of planar liquid jets, harmonically forcing the transversal velocity at the inlet section without gravitational effects.

1.3 Reduced order modelling

Over the last decade, reduced order modeling (ROM) has gained significant attention in the field of computational science and engineering. It has become a crucial tool for achieving high-performance computing in various industries, including mechanical and electronic engineering, as well as in fundamental and applied sciences such as neuroscience, medicine, biology, and chemistry (Quarteroni et al. [66]). These methods are also gaining importance in emerging areas that deal with complex problems (Papaioannou et al. [67]) spanning multiple physical phenomena and scales (Mendez et al. [68]), as well as in addressing uncertainties (Cinnella et al. [69], Edeling et al. [70], Xiao and Cinnella [71]).

A typical way to obtain a ROM is to extract physically important features or modes which characterize the flow topology and project the Navier-Stokes equations (Galerkin projection) onto a subset of these modes, resulting in a system of ordinary differential equations (Noack et al. [13], Rowley and Dawson [72]). Choosing a restricted set of modes among those available, it is possible to built a reduced model able to predict the flow field behavior, with a lower computational cost.

Recently, the integration of Machine Learning has opened new perspectives for understanding and modeling complex behaviors (Lee et al. [73], Fabiani et al. [74], Alvarez et al. [75]). For instance, Physics-Informed Neural Networks (PINNs) can help to estimate mean fields in linearized approaches when the governing equations are known but data is sparse (Raissi et al. [76], von Saldern et al. [31]).

Cluster-based Network Modeling (CNM) emerges as a powerful approach for data-driven modeling of complex nonlinear dynamics. This technique employs unsupervised machine learning to reduce a collection of snapshots into selected few flow states that best represent the system (Fernex et al. [77]). By means of statistical approaches a probabilistic reduced order model is then obtained. CNM's strength lies in its automation, offering a comprehensive framework for understanding system dynamics (Vogel et al. [78], Hou et al. [39]), estimating parameters, and enabling model-based control (Wang et al. [79]). Furthermore, CNM proves to be

highly effective in analyzing phenomena like bifurcations and the emergence of secondary flow patterns (Deng et al. [80]).

Generally, the methodology is applicable to any time series. In complex configurations, however, challenges may arise. Inherent variable heterogeneity and measurement noise, both in time and space, can complicate interpretations and model training. Additionally, researchers might be interested in specific subdomains of the field, such as shear layers or inflow/outflow regions and how they affect the overall dynamics. Within the framework of Proper Orthogonal Decomposition, extended approaches have been proposed to address these challenges, like Extended POD (EPOD) (Borée [81]) and Extended SPOD (ESPOD) (Sieber et al. [82]).

The present work introduces an analogous approach in the clustering procedure to focus on specific, potentially more informative, subspaces or subsets of variables. Furthermore, an extension is proposed to leverage a high-temporal-resolution quantity to identify the coherent motions in a temporally undersampled dataset. This work has been carried out in collaboration with prof. Kilian Oberleithner of the Technische Universität Berlin (Berlin, Germany), where the candidate spent spent a Visiting Scholar period of three months during his Ph.D. Program.

Standard CNM can affect the estimation evolution by experiencing diffusion of the state vector to post transient attractors (Fernex et al. [77], Kaiser et al. [83], Ishar et al. [84]). To overcome this limitation a method presented in this dissertation will focus on the analysis of piecewise trajectories instead of temporal local states. The standard Galerkin-ROM techniques and previous CNM approaches consider a state vector composed of the realizations of flow field variables. The new method consists in considering a state vector composed of trajectories of the dynamical system. This methodology has been developed during a Visiting Scholar period of two months spent in Harbin Institute of Technology (Shenzhen, China), under the supervision prof. Bernd R. Noack.

1.4 Outline

The rest of the thesis is structured as follows. Chapter 2 discusses in detail the modal analysis techniques that have been employed and developed during the Ph.D. activities. In particular SPOD analysis of the vertical liquid jet, the novel Robust SPOD algorithm and the Gappy POD of two-phase PIV data. Chapter 3 contains a detailed description of the data-driven stability analysis of the 2D vertical liquid jet (whose insights are given in appendix A.1). Chapter 4 is dedicated to the reduced order models developed. Specifically, the Extended CNM approach is described, a comparison between POD-Galerkin ROM and CNM is reported for airfoil flow fields and the functional CNM is introduced. The thesis ends with Chapter 5, containing a summary of the main findings and suggestions for possible future research activities.

Chapter 2

Modal analysis

It is known that the comprehension of complex flows takes advantage from the recognition of physically important features, associated with the modes, that characterize the spatial topology. Modes can also be used to construct reduced-order modeling and to control the flow itself as shown in Schmidt and Colonius [85]. Several techniques have been developed to extract the main features of the flow, as reported in the reviews of Rowley and Dawson [86], Taira et al. [9], Towne et al. [16], and others. POD (Berkooz et al. [87]), dynamic mode decomposition (DMD; Schmid [43], [44]), and SPOD (Towne et al. [16]) are the most used techniques for this purpose. In particular, the POD provides the minimal number of basis functions (the modes) to capture as much energy as possible, whereas the SPOD allows one to characterize the spatial and temporal evolution of coherent structures through the decomposition in the frequency domain of the flow in various modes, ordered by their energy content. Recent works on this topic have been carried out to extract the features of swirling coaxial jets (Kadu et al. [88]) and of the flow around a square-section building (Zhang et al. [89]), to characterize the dynamics of flows past finite curved cylinders by Chiatto et al. [17] and the dynamics of actively controlled flow past a backward facing ramp by the same authors (Chiatto et al. [22]). Recently, in the framework of two-phase flows, Barreiro-Villaverde et al. [90] analyzed numerical simulations of the jet wiping process by means of multiscale proper orthogonal decomposition (mPOD), to characterize

the interaction between the gas flow and the liquid film, by detecting characteristic timescales of the phenomenon and related dominant spatial structures.

The algorithm of the SPOD technique represents the frequency domain version of the standard space-only POD (Lawrence [91]), which is not robust to outliers and corrupted data. the present work aims to overcome this limitation by introducing the robust spectral proper orthogonal decomposition (robust SPOD), and describing its algorithm in detail. The technique is applied to classical fluid dynamics problems to analyze different aspects of the algorithm: the subsonic jet flow field numerically computed by large-eddy simulations (LES) (Schmidt et al. [15]) and the flow within an open cavity, obtained by means of PIV measurements.

The chapter is organized as follows: Section 2.1 provides an introduction to the Spectral Proper Orthogonal Decomposition (SPOD) technique, highlighting its key features. This is followed by its application in the analysis of vertical liquid jets. In Section 2.2, the robust SPOD procedure is introduced and subsequently applied to two distinct test cases: a subsonic turbulent jet and an open cavity flow. Finally, Section 2.3 introduces the Gappy Proper Orthogonal Decomposition (Gappy POD) method for two-phase flows, with an illustration of its application in the context of the two-phase mixing layer past a splitter plate.

2.1 Spectral Proper Orthogonal Decomposition

In this section the main elements of the spectral proper orthogonal decomposition are recalled.

Defining the fluctuation of a stochastic field $\boldsymbol{q}(\boldsymbol{x}, t)$, with $\boldsymbol{x}$ the position and t the discrete time, as:

$$\boldsymbol{q'}(\boldsymbol{x}, t) = \boldsymbol{q}(\boldsymbol{x}, t) - \bar{\boldsymbol{q}}(\boldsymbol{x}) \tag{2.1}$$

where $\bar{\boldsymbol{q}} = E\left\{\boldsymbol{q}(\boldsymbol{x}, t)\right\}$ is the ensemble average (and $E\left\{\cdot\right\}$ the expectation operator), in accordance with the works by [16], the SPOD decomposition provides the best representation of $\boldsymbol{q'}(\boldsymbol{x}, t)$, with the least number of

modes, in the space-time domain(Schmidt and Colonius [85]).

The SPOD modes satisfy the eigenvalue problem of the cross-spectral density tensor $\boldsymbol{S}$, and are computed with the Welch periodogram method (Schmidt and Colonius [85]). Accordingly, the flow snapshots (each snapshot representing a flow field realization at a given time instance) are grouped in N_b blocks of equal length N_f (N_f being the number of snapshots per block): $\boldsymbol{Q}^{(j)} = \left[\boldsymbol{q}_1^{(j)}, \boldsymbol{q}_2^{(j)}, ..., \boldsymbol{q}_{N_f}^{(j)}\right]$, with $j = 1, ..., N_b$. For clarity, the superscript $\{\cdot\}'$ (which represents the fluctuation) has been removed. Each block overlaps with adjacent ones.

For each block j, the weighted discrete Fourier transform (Welch [92]), $\hat{\boldsymbol{Q}}^{(j)} = \mathscr{F}\left\{\boldsymbol{Q}^{(j)}\boldsymbol{W}_T\right\}$ (where $\boldsymbol{W_T}$ is the matrix of the window weights) is first computed. Then all Fourier transform realizations at the same kth frequency are collected in the matrices as $\hat{\boldsymbol{Q}}_k = \left[\hat{\boldsymbol{q}}_k^{(1)}, \hat{\boldsymbol{q}}_k^{(2)}, ..., \hat{\boldsymbol{q}}_k^{(N_b)}\right]$, and the cross-spectral density tensor $\boldsymbol{S}_k = \hat{\boldsymbol{Q}}_k\hat{\boldsymbol{Q}}_k^*$ is computed one frequency at a time, thus obtaining the SPOD modes ($\boldsymbol{\Phi}_k$) and eigenvalues ($\boldsymbol{\Lambda}_k$).

Once the SPOD modes have been identified, a low rank reconstruction of the fluctuation field can be obtained through the inverse weighted (discrete) Fourier transform $\boldsymbol{Q^{(j)}} = \mathscr{F}^{-1}\left\{\hat{\boldsymbol{Q}}^{(j)}\boldsymbol{W}_T^{-1}\right\}$ to come back in the time domain. This approach was recently presented by Nekkanti and Schmidt [93] and applied by Chiatto et al. [22] to reconstruct the flow past a backward facing ramp (with and without active flow control) considering a limited number of SPOD modes.

More in detail, the Fourier realizations matrix is expressed as $\hat{\boldsymbol{Q}}_k = \boldsymbol{\Phi}_k\boldsymbol{\Lambda}_k\boldsymbol{\Psi}_k^*$, where $\boldsymbol{\Psi}_k$ is the matrix containing the eigenvectors of $\hat{\boldsymbol{Q}}_k^*\hat{\boldsymbol{Q}}_k$. For low rank reconstruction only r eigenvalues are retained and $\hat{\boldsymbol{Q}}_k$ is approximated as

$$\hat{\boldsymbol{Q}}_k \approx \tilde{\boldsymbol{\Phi}}_{\boldsymbol{k}}\tilde{\boldsymbol{\Lambda}}_k\tilde{\boldsymbol{\Psi}}_k^*, \tag{2.2}$$

where $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{k}}$ and $\tilde{\boldsymbol{\Psi}}_k$ are, respectively, the first r columns of $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Psi}_k$.

As previously stated, once the reconstruction procedure of the $\hat{\boldsymbol{Q}}_k$ has been completed, the Fourier realizations are re-arranged into the blocks form $\hat{\boldsymbol{Q}}^{(j)} = \left[\hat{\boldsymbol{q}}_1^{(j)}, \hat{\boldsymbol{q}}_2^{(j)}, ..., \hat{\boldsymbol{q}}_{N_f}^{(j)}\right]$, and, the snapshots in the time domain are obtained through a inverse weighted Fourier transform. Particular at-

tention needs to be paid to avoid the ambiguity for the snapshots belonging to two different block. To solve the ambiguity, as proposed by Nekkanti and Schmidt [93], the snapshot with the higher windowing weight $w(j)$ is chosen.

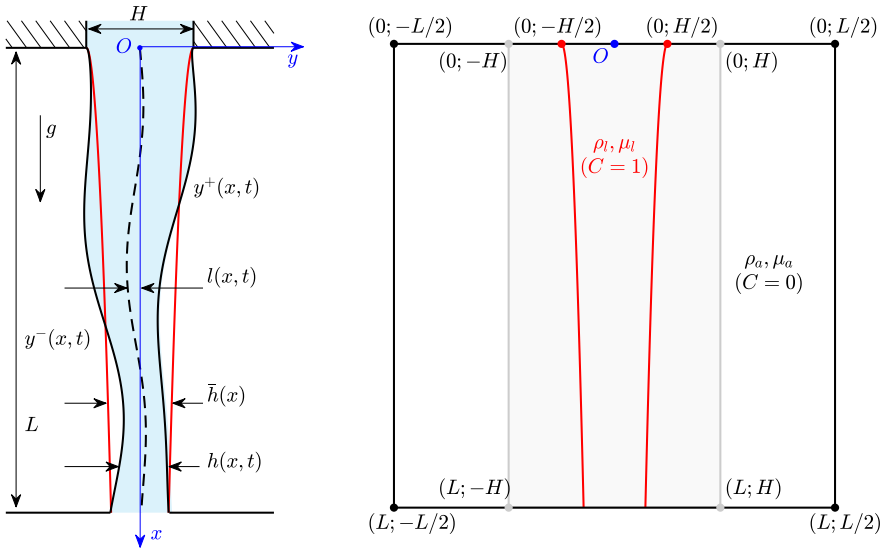
2.1.1 Application to vertical liquid jet

In this section a modal decomposition analysis based on the spectral proper orthogonal decomposition (SPOD) is applied to viscous thin liquid sheets to extract the spatially and dynamically most relevant flow structures. For all cases a POD (Proper Orthogonal Decomposition) analysis has also been carried out. Numerical data of two-dimensional liquid sheets flow fields computed by means of the two-phase code BASILISK (Popinet [94, 95]).

The flow here considered is a gravitational viscous liquid jet interacting on both sides with an unconfined gaseous ambient. Panel (a) of Fig. 2.1 contains a sketch of the system, together with the reference frame: its origin is coincident with the center of the inlet flow section (whose width is denoted by H), the x axis is directed along the streamwise direction, while the y axis is across the width of the sheet. Red lines represent the symmetric unperturbed flow, corresponding to the Torricellian solution whose thickness ($\bar{h}(x)$) is described by the law

$$\frac{\bar{h}(x)}{H} = \left(1 + \frac{2g}{U^2}x\right)^{-1/2}, \tag{2.3}$$

where U is the mean axial velocity at the inlet section and g is the gravity acceleration. Black lines, instead, correspond to the instantaneous left and right interface locations, and are denoted with $y^-(x,t)$ and $y^+(x,t)$, respectively. Further details on the liquid sheet configuration and numerical simulations are reported in appendix A.1. For the subsequent analysis dimensionless parameters have been conveniently introduced, the Weber number (($We = \rho_l U^2 H/(2\sigma)$, σ being the surface tension), the Reynolds number ($Re = \rho_l U L/2\mu_l$), the Froude number ($Fr = U^2/(gL)$), the



(a) Physical domain. **(b)** Computational domain.

Figure 2.1. Sketch of physical (a) and computational (b) domains. The red lines represent the steady (or mean) interface locations corresponding to the Torricellian shape.

Strouhal number ($St = fH/U$) and the density ratio ($r_\rho = \rho_a/\rho_l$), where subscripts a and l stand, respectively, for ambient and liquid.

The configuration consists in a forced liquid sheet flow, with a fully developed parabolic axial velocity profile at the inlet, which is perturbed by a sinusoidal forcing of variable amplitude in the lateral velocity component. Accordingly, Dirichlet boundary conditions are imposed at the inlet, for $y \in [-H/2; H/2]$:

$$u = \frac{3}{2}U\left[1 - \left(\frac{2y}{H}\right)^2\right], \tag{2.4a}$$

$$v = \hat{A}U \sin\left(2\pi f_f(t - t_s)\right) \Theta\left(t - t_s\right), \tag{2.4b}$$

$$C = 1, \tag{2.4c}$$

where U and H are, respectively, the mean axial velocity and the thickness of the liquid sheet at the inlet, $\hat{A}$ is the dimensionless amplitude of the forcing lateral velocity perturbation, f_f is the forcing frequency and $\Theta\left(t - t_s\right)$ the Heaviside function acting at the forcing starting time t_s. On the remaining part of the inlet side ($|y| > H/2$) no-slip conditions are imposed.

Note that here and in the following any variable ϕ is decomposed as:

$$\phi(x, y, t) = \bar{\phi}(x, y) + \phi'(x, y, t) \tag{2.5}$$

where $\bar{\phi}$ denotes the temporal mean and ϕ' the fluctuation.

A parametric study is carried out to further elucidate the influence of the Weber, Strouhal and Reynolds numbers on the dynamics of planar liquid sheets subjected to lateral perturbations at the inlet section in both supercritical and subcritical regimes.

For the SPOD analysis the state vector $\boldsymbol{q(x, t)}$ is obtained by stacking the fluctuations of the velocity components (u' and v') and of the volume fraction (C') in a column vector for a given time instance. $N = 2500$ snapshots, corresponding at least to 150 travel times, have been considered, subdividing the data in $N_b = 12$ blocks of equal length ($N_f = 374$). The Strouhal number ranges from 0 to 0.30, with a resolution of $\Delta St = 0.0017$.

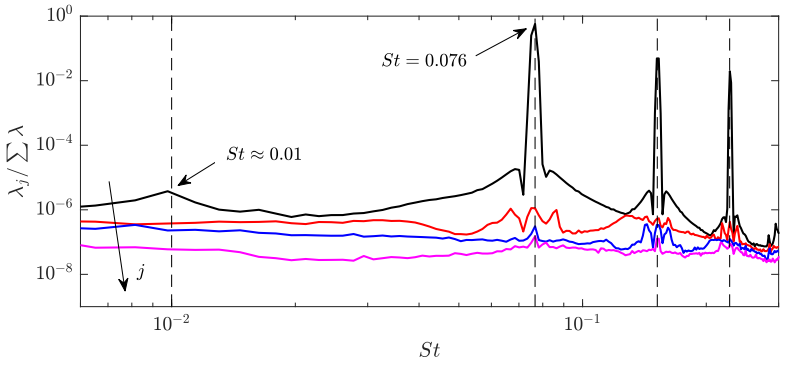


Figure 2.2. Supercritical conditions (case (a) of Table A.1). Normalized SPOD spectrum. Only the first 4 modes are reported. Arrow, labeled j, indicates increasing order of modes, with j the running index associated with the modes.

For all cases we have also carried out a POD analysis. However, the results are not discussed systematically unless relevant for the physical understanding of the flow structure and dynamics.

Supercritical regime

The SPOD spectrum for the supercritical condition ($We > 1$, case (a) of Table A.1) corresponding to $We = 2.5$, $Re = 400$, $St_f = 0.076$ (dimensionless forcing frequency corresponding to $f_f = 25$ Hz), and $\hat{A} = 0.02$ is shown in Fig. 2.2.

The flow presents the largest modal separations at the forcing frequency and its harmonics, meaning that the greatest amount of energy content is concentrated at these frequencies. Another separation region occurs at a lower frequency, $St \approx 0.01$, which is associated with the natural frequency predicted by the linearized 1D inviscid analytical model of Della Pia et al. [61]. As shown by Della Pia et al. [61, 65], supercritical regimes are characterized by natural oscillation frequencies which are one order of magnitude lower than the corresponding frequencies in subcritical

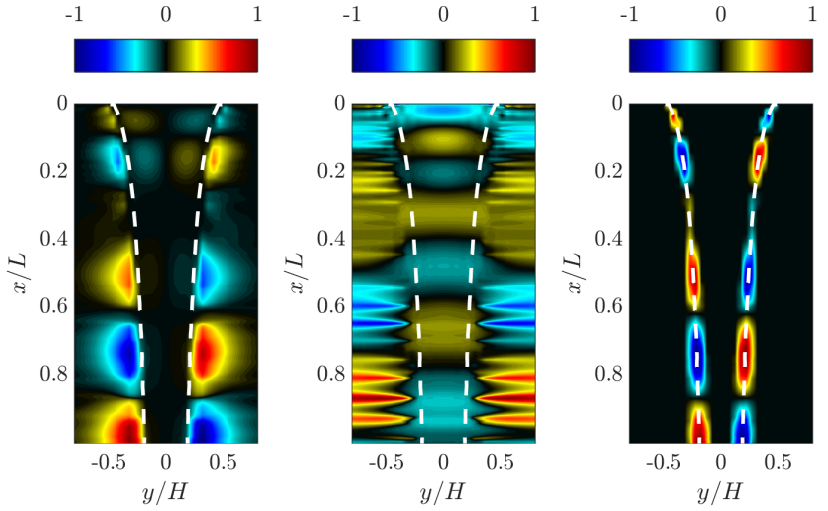
conditions, as will be recovered by the modal analysis hereafter reported.

The spatial structures at the leading frequency $f_f = 25$ Hz ($St_f = 0.076$) of the 1^{st} SPOD mode are reported in Fig. 2.3. Panels (a)-(c) show the real part of the mode, whereas panels (d)-(e) refer to the corresponding imaginary parts. Both real and imaginary structures show a sinuous behavior of the interface, exhibiting an antisymmetric distribution (with respect to the axis $y = 0$) for u' and C' and a symmetric one for v'. By comparing the real and imaginary parts of the spatial structures, one can recover the traveling nature of the disturbances, as the imaginary structures are shifted upstream by a quarter of wavelength along the advection direction with respect the corresponding real structures. A sketch of sinuous disturbance, obtained from the previous distributions, is reported in Fig. 2.4, panel (a). Fig. 2.4, panel (b) depicts the streamlines pattern as viewed by an observer moving with the mean flow; the color map refers to the spatial distribution of C' mode.

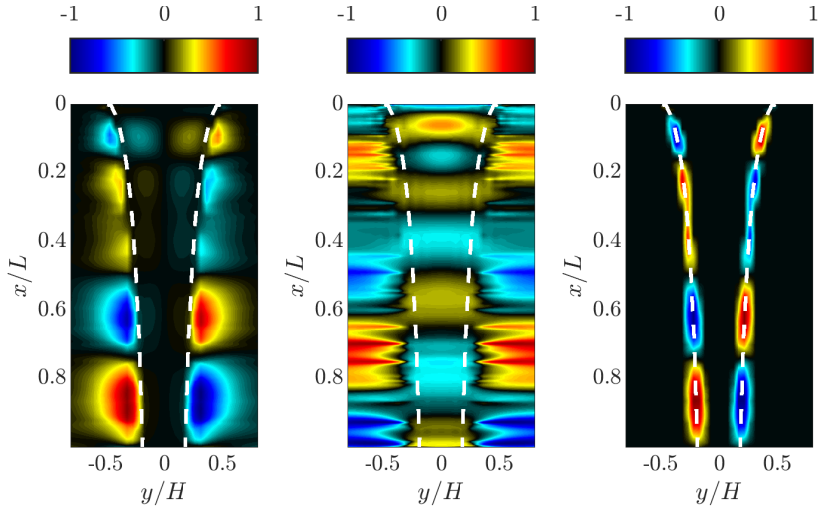
The structures at $St \approx 0.01$, reported in Fig. 2.5, confirm a dominant sinuous deformation of the interface, in agreement with the previous study (Della Pia et al. [61]) carried out with the simplified 1D model.

The topology of the POD modes (not reported) is practically similar to that of the SPOD ones. In Fig. 2.6 we report the energy content (panel a) and the cumulative energy (panel b) of the POD modes. The figure shows that the most energetically relevant modes are the first two that contain approximately the 85% of the total energy, to be compared with the 99% cumulative energy of the first 10 modes.

The power spectral density (PSD) of the POD modes is reported in panel (c) of Fig. 2.6. The figure shows that at the natural frequency ($St \approx 0.01$) the dominant mode is the 9^{th} one, even though its total energy content is far less than the first mode. Therefore, one could consider the first 2 modes for which the cumulative energy is about 85% of the total energy, thus producing the same loss of space-wise information of the SPOD decomposition. However, the low frequency features would be lost, indeed to recover low frequency effects one should include the 9^{th} POD mode.

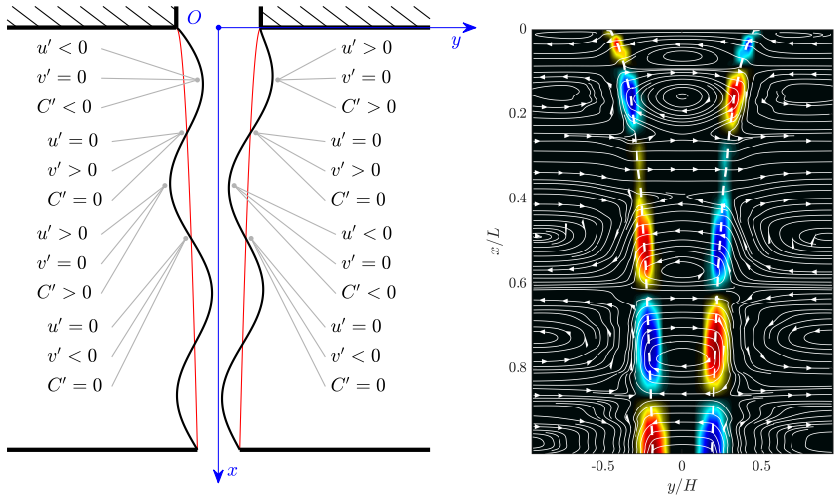


(a) Real part u' mode **(b)** Real part v' mode **(c)** Real part C' mode



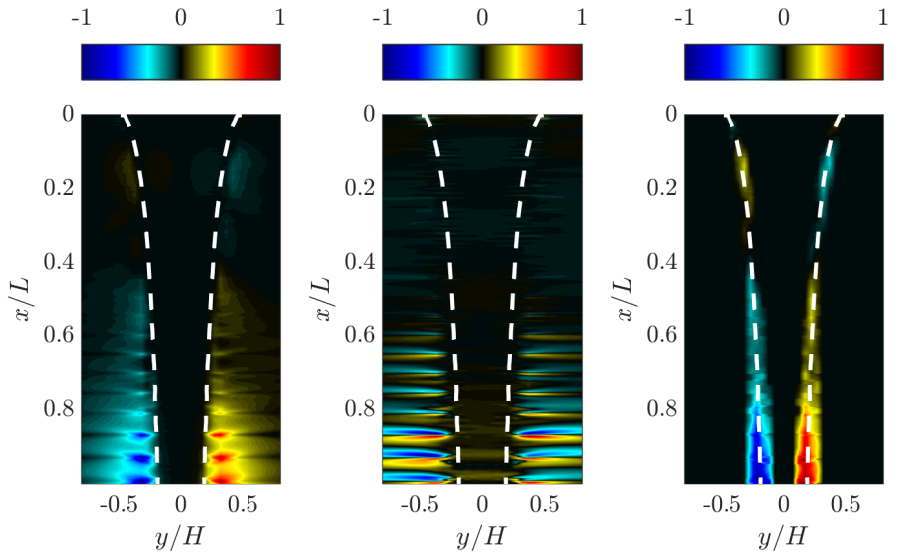
(d) Imaginary part u' mode **(e)** Imaginary part v' mode **(f)** Imaginary part C' mode

Figure 2.3. Supercritical conditions (case (a) of Table A.1). 1^{st} SPOD mode at $St = 0.076$, real parts in panels (a)-(c), and imaginary parts in panels (d)-(f).



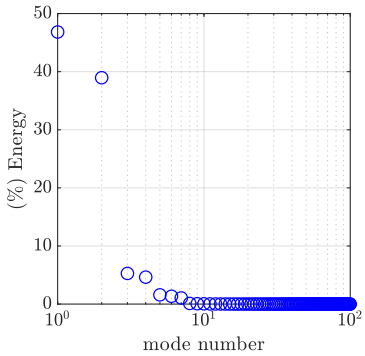
(a) **(b)**

Figure 2.4. Panel (a): Sketch of sinuous disturbance. The red lines denote the interface of the mean field. Panel (b): Streamlines pattern as viewed by an observer moving with the mean flow. The color map refers to C' mode.

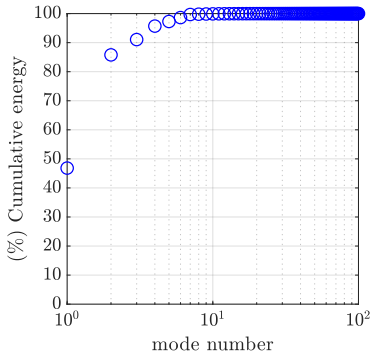


(a) u' mode, at $St = 0.01$ **(b)** v' mode, at $St = 0.01$ **(c)** C' mode, at $St = 0.01$

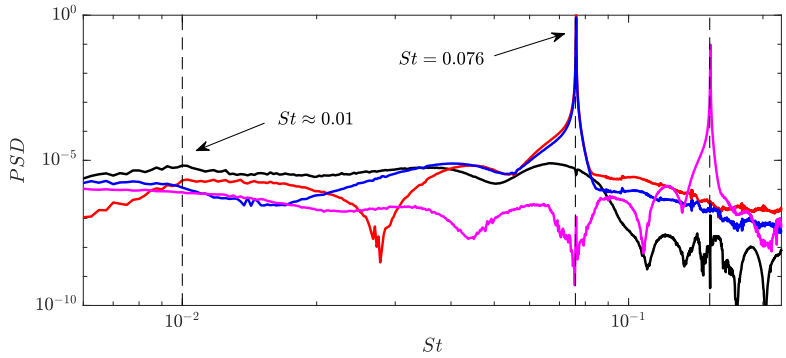
Figure 2.5. Supercritical conditions (case (a) of Table A.1). (a) u' mode, (b) v' mode, (c) C' mode at $St \approx 0.01$.



(a) Energy content of POD modes.



(b) Cumulative energy.



(c) normalized PSD of POD modes.

Figure 2.6. Supercritical regime (case (a) of Table A.1). Energy content (panel (a)), cumulative energy (panel b) and power spectral density (panel (c)) of POD modes. Red line, 1^{st} mode; blue line, 2^{nd} mode; magenta line, 3^{rd} mode; black line, 9^{th} mode.

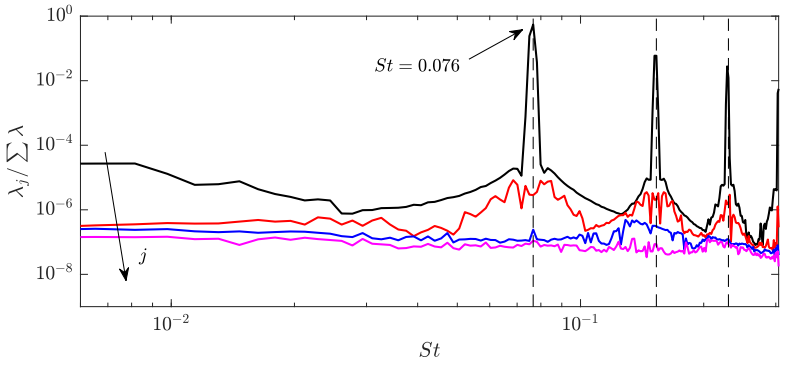


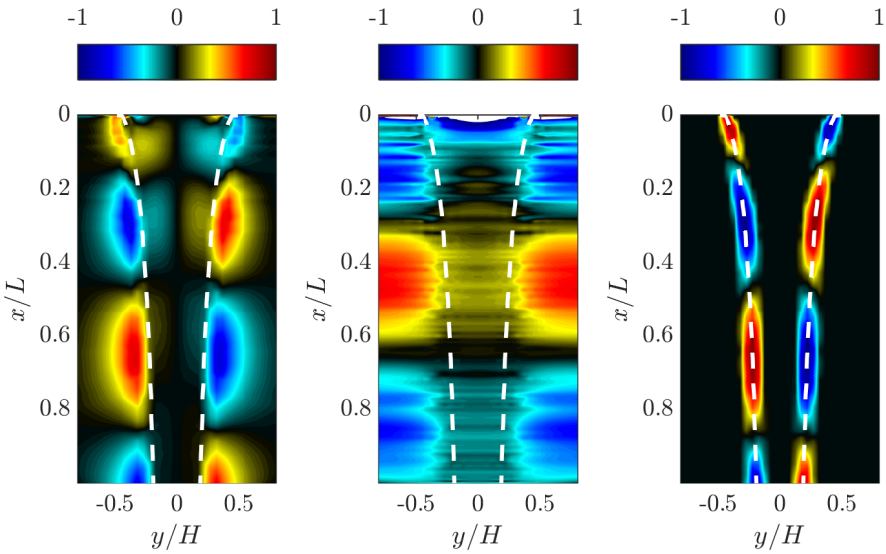
Figure 2.7. Subcritical conditions (case (c) of Table A.1). Normalized SPOD spectrum. Only the first 4 modes are reported. Arrows, labeled j, indicate increasing order of modes, with j the running index associated with the modes.

Subcritical regime

In order to further elucidate the prominent role played by the surface tension in the subcritical regime ($We < 1$), we have carried out a parametric study at various We, Re and forcing frequency (St_f) (cases (b)-(k) of Table A.1). Because of the stabilizing effect of the surface tension and in order to detect the liquid sheets oscillations on the adopted computational grid, we have set the initial disturbance amplitude to $\hat{A} = 0.1$.

Four different values of the Reynolds number have been considered at a given Weber number ($We = 0.75$) and a forcing frequency $f_f = 25$ Hz ($St = 0.76$) (cases (b)-(e) of Table A.1). Note that for the subcritical regime 25 Hz coincides with the natural frequency (Della Pia et al. [65]) (i.e. the planar sheet is forced in resonance conditions). The SPOD spectrum of case (c) is reported in Fig. 2.7.

At low frequencies ($St < 0.076$), unlike the supercritical regime, the spectrum does not exhibit any local maximum (corresponding to leading eigenvalues), confirming the occurrence of a discontinuous frequency behavior between the subcritical and supercritical regimes(Girfoglio et al. [53]).



(a) u' mode at $St = 0.076$ **(b)** v' mode at $St = 0.076$ **(c)** C' mode at $St = 0.076$

Figure 2.8. Subcritical conditions (case (c) of Table A.1). 1^{st} SPOD mode at $St = 0.076$. (a), u'; (b), v'; (c), C'.

The leading SPOD mode ($St = 0.076$ of the 1^{st} SPOD mode) for case (c) of Table A.1 is reported in Fig. 2.8 (panel (a), u' mode; panel (b), v'; panel (c), C'). The sheet exhibits only a sinuous motion; u' and C' modes show an antisymmetric behavior whereas v' is symmetric.

The spatial topology is similar to the one of the corresponding supercritical regime (case (a) shown in Fig. 2.3, panels (a)-(c)). However, the dimensionless wavelength of the subcritical sinuous deformation is greater.

Fig. 2.9 reports the normalized spectra at various Re. The flow presents a clear modal separation at the forcing frequency and its harmonics at all Reynolds numbers. At the highest Reynolds number ($Re = 1600$) a peak occurring at the sub-harmonic frequency $St = 0.038$ is also detectable, this being a typical behavior of nonlinear phenomena (Hayashi [96]),

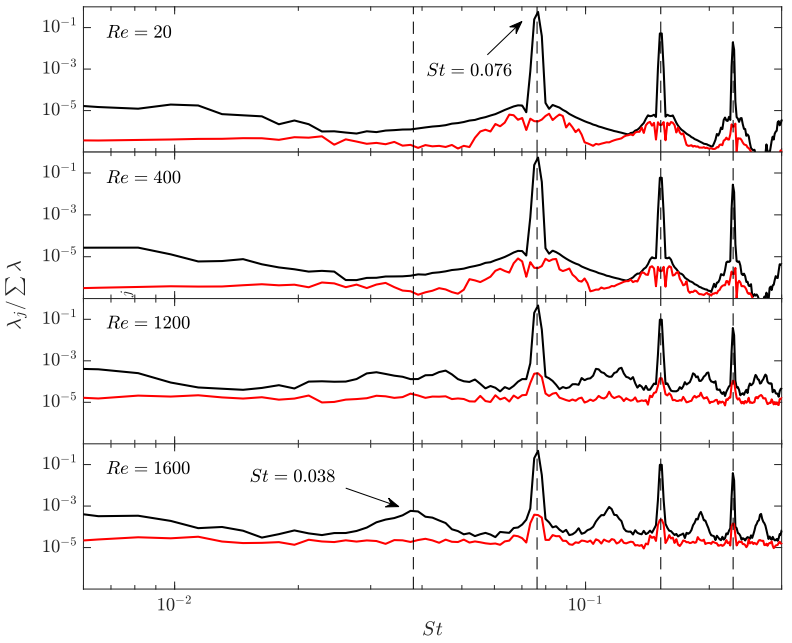


Figure 2.9. Subcritical regime (cases (b)-(e) of Table A.1). Normalized SPOD spectra at various Reynolds numbers Re. The spectra report only the frequency content of the first 2 modes.

associated with the quadratic phase coupling(Schmidt [97]).

Fig. 2.10 shows the normalized cumulative energy distribution of SPOD C' modes at various Reynolds numbers for a driving frequency close to the natural one. At low Re, the first 10 modes contribute for approximately 99% to the total energy of the flow. As the Reynolds number increases, due to nonlinear effects, higher modes are excited and about 100 modes are needed to capture the same amount of energy at $Re > 1000$. The same results are retrieved when applying the POD technique (not reported).

In Della Pia et al. [65] it has been shown that forcing the liquid sheet with a (forcing) frequency close to the natural one produces a maximum amplitude of the centerline deflection. They also observed that at high

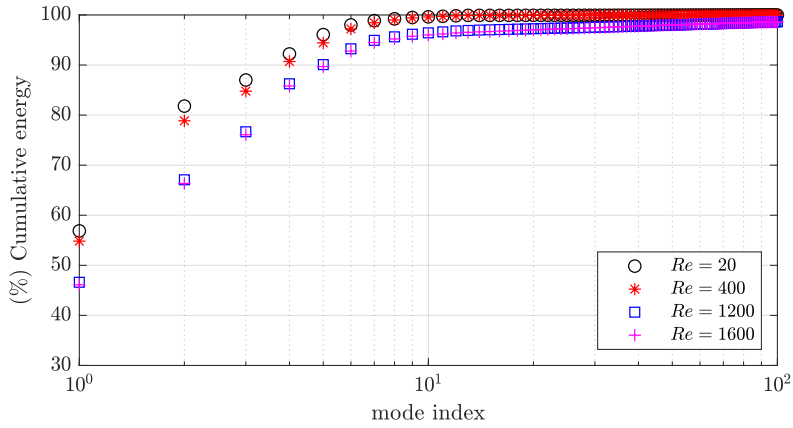


Figure 2.10. Subcritical regime (cases (b)-(e) of Table A.1). Cumulative energy distribution of SPOD modes at various Reynolds numbers.

Reynolds numbers, in subcritical regime the flow experiences resonance due to nonlinear coupling, exhibiting a varicose deformation. As previously discussed, sinuous contributions show an antisymmetric behavior for u' and C' and a symmetric behavior for v' (Fig. 2.4), whereas the opposite behavior holds true for the varicose ones as sketched in Fig. 2.11, panel (a). The panel (b) depicts the streamlines pattern as viewed by an observer moving with the mean flow. The color map refers the spatial distribution of varicose part of C' mode.

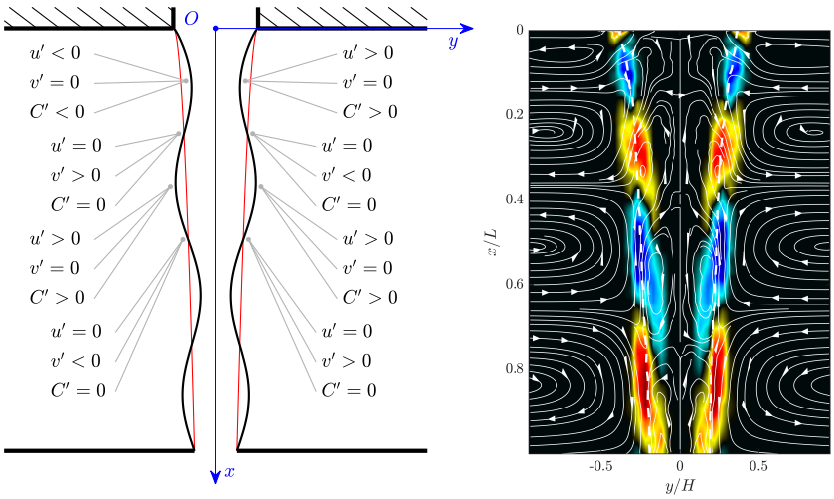
To recover sinuous and varicose structures, any mode $\boldsymbol{\varphi}_j$ is decomposed as:

$$\boldsymbol{\varphi}_j(\boldsymbol{x}) = \boldsymbol{\varphi}_j^s(\boldsymbol{x}) + \boldsymbol{\varphi}_j^v(\boldsymbol{x}) \tag{2.6}$$

where superscripts s and v stand for sinuous and varicose, respectively.

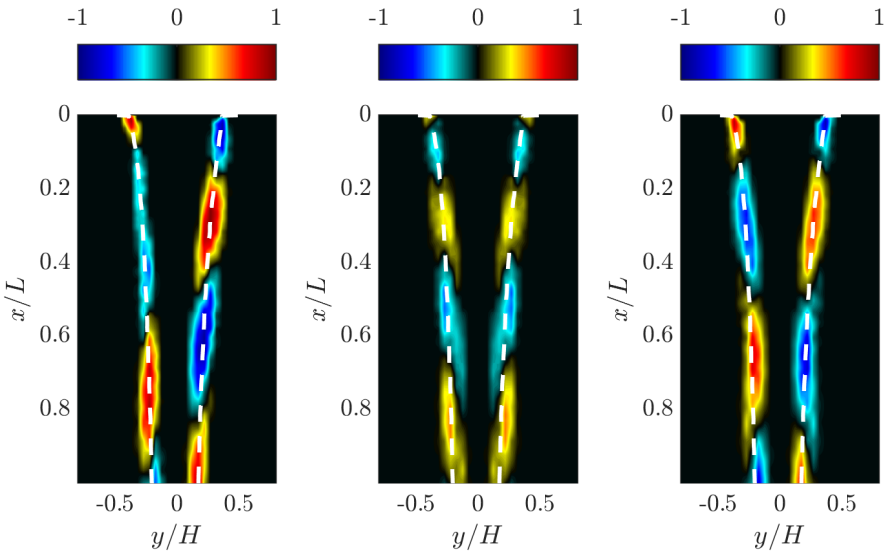
Fig. 2.12 reports the varicose and sinuous components extracted for the C' mode (case (e)), where $\boldsymbol{\varphi}_j^s$ and $\boldsymbol{\varphi}_j^v$ are defined as:

$$\boldsymbol{\varphi}_{C'}^s(x,y) = \big[\boldsymbol{\varphi}_{C'}(x,y) - \boldsymbol{\varphi}_{C'}(x,-y)\big]/2 \tag{2.7a}$$
$$\boldsymbol{\varphi}_{C'}^v(x,y) = \big[\boldsymbol{\varphi}_{C'}(x,y) + \boldsymbol{\varphi}_{C'}(x,-y)\big]/2 \tag{2.7b}$$



(a) **(b)**

Figure 2.11. Panel (a): Sketch of varicose disturbance. The red lines denote the interface of the mean field. Panel (b): Streamlines pattern as viewed by an observer moving with the mean flow; the color map refers to the varicose part of C' mode.



(a) C' mode at $St = 0.076$

(b) varicose component (symmetric).

(c) sinuous component (antisymmetric).

Figure 2.12. Subcritical regime (case (e) of Table A.1). (a) C' mode, (b) varicose component and (c) sinuous component at $St = 0.076$ for $Re = 1600$. Decomposition in sinuous and varicose components.

.

The eigenvalues (normalized by the total energy) of the first and second SPOD modes at three forcing frequencies and at a given Re and We (cases (e)-(g) of Table A.1) are reported in Fig. 2.13 as a function of the Strouhal number. A clear mode separation is observed at the forcing frequency and its superharmonics for all forcing frequencies.

The spatial distributions of the leading u' and C' modes are reported in Fig. 2.14. Panels (a)-(b) and (c)-(d) refer to cases (f) and (g), respectively. The figure shows that forcing at frequencies greater than the natural one (25 Hz) induces a sinuous deformation of the liquid sheet.

Looking at the cumulative energy distribution reported in Fig. 2.15, one observes that forcing at resonance conditions the first leading SPOD

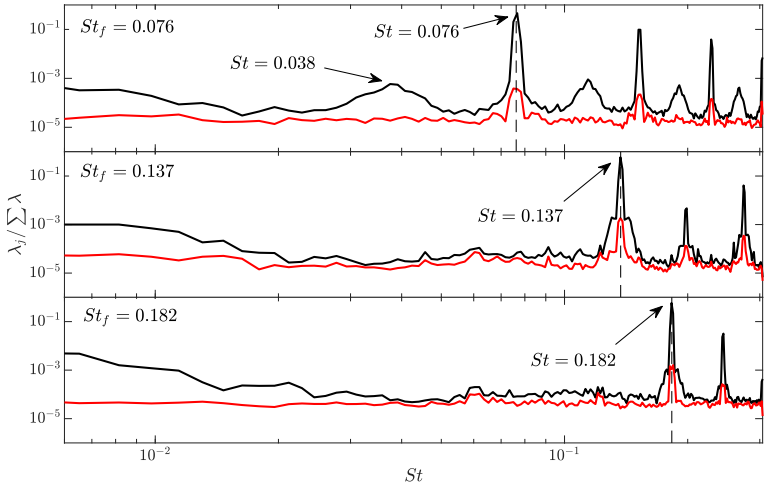


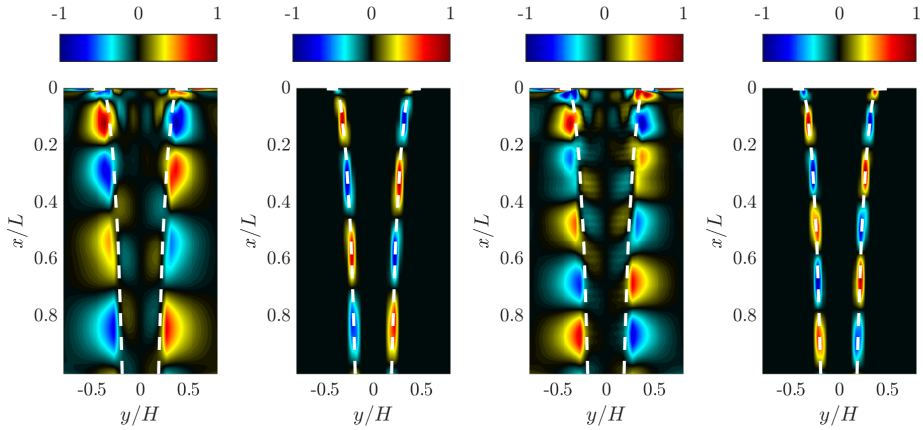
Figure 2.13. Subcritical regime (cases (e)-(g) of Table A.1). Normalized SPOD spectra at various forcing Strouhal numbers St_f. The spectra report only the frequency content of the first 2 modes.

modes (modes 1-3) are energetically less significant than the corresponding ones at higher forcing Strouhal numbers. This is due to the nonlinear interactions between sinuous and varicose components, exhibiting different frequencies and producing a large number of energetically relevant modes.

Four different values of the Weber number have been considered for a given Re at resonant frequency $f_f = 25$ Hz ($St_f = 0.076$) (cases (e) and (h)-(k)). The spectra of the first and second SPOD modes are reported in Fig. 2.16.

For all Weber numbers, the spectra peak at the forcing frequency (the leading one) and its harmonics. For $We = 0.75$ and 0.8 the spectra also peak at the subharmonic ($St = 0.038$). To understand the response of the sheet subjected to a sinuous forcing as a function of the Weber number, in Fig. 2.17 we report the u' and C' SPOD modes corresponding to $We = 0.85$ and $Re = 1600$ (case (j)).

The figure shows that the SPOD modes have a dominant varicose



(a) u' mode at $St = 0.137$. **(b)** C' mode at $St = 0.137$. **(c)** u' mode at $St = 0.182$. **(d)** C' mode at $St = 0.182$.

Figure 2.14. Subcritical regime (cases (f) and (g) of Table A.1). Spatial distribution of the leading SPOD modes at $St = 0.137$ (panels (a)-(b)) and $St = 0.182$ (panels (c)-(d)), respectively, in terms of u' and C'.

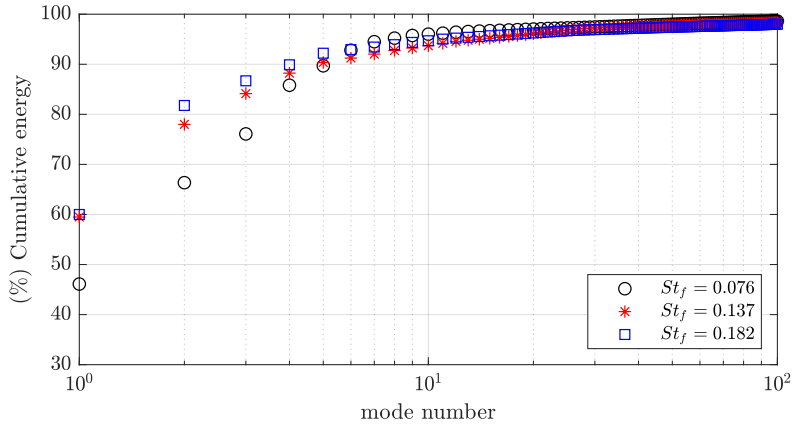


Figure 2.15. Subcritical regime (cases (e)-(g) of Table A.1). Cumulative energy distribution of SPOD modes at various forcing Strouhal numbers.

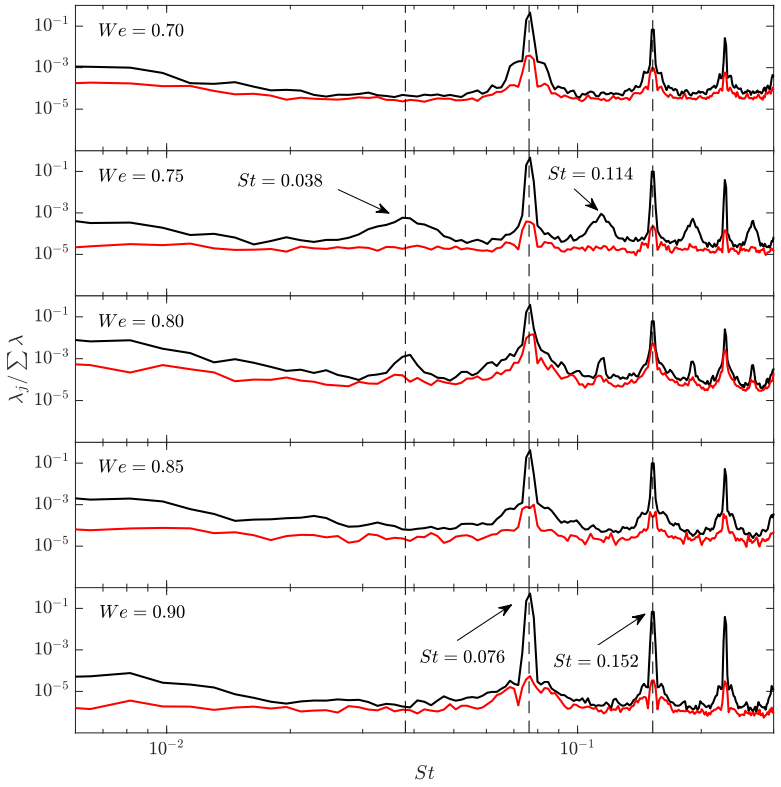
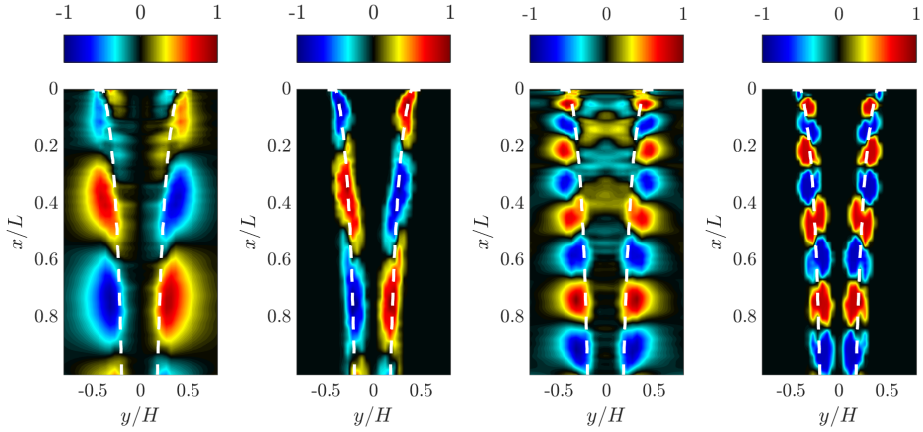


Figure 2.16. Subcritical regime (cases (e) and (h)-(k) of Table A.1). Normalized SPOD spectra at various Weber numbers We. The spectra report only the frequency content of the first 2 modes.



(a) u' mode at $St = 0.076$. **(b)** C' mode at $St = 0.076$. **(c)** u' mode at $St = 0.152$. **(d)** C' mode at $St = 0.152$.

Figure 2.17. Subcritical regime (case (j) of Table A.1). Spatial distribution of the 1^{st} SPOD mode at $St = 0.076$ (panel (a)-(b)) and $St = 0.152$ (panel (c)-(d)), in terms of u' and C'.

character at the harmonic $St = 0.152$ (panels (c) and (d)), whereas at the forcing frequency they exhibit a dominant sinuous character. Decreasing the Weber number, at $We = 0.75$ and 0.80, the varicose component of the modes becomes significant at the leading frequency (see panel (b) of Fig. 2.12) and at its subharmonic ($St = 0.038$) (not shown herein).

The role of the sinuous and varicose components in the dynamics and topology of the flow can be further highlighted in terms of their energy content. For that purpose we introduce the energy ratio of the first SPOD mode at a frequency k defined as $E_k = \lambda_{1,k} / \sum_{i,j} \lambda_{i,j}$. Then, the sinuous and varicose energy contents (respectively, E^s and E^v) can be obtained from the antisymmetric and symmetric parts of the energy ratio.

The energy contents ($\%$) of the sinuous and varicose components of the first SPOD mode are given in Fig. 2.18 and Table 2.1 as a function of the Weber number at the leading frequency ($St = 0.076$) and its first harmonic ($St = 0.152$). The figure shows that E^s and E^v have opposite trends. At $St = 0.076$ E^v becomes progressively more important (energy-wise) as We decreases, attaining a maximum at $We = 0.75$, where E^s is minimum and then increases with We. At $St = 0.152$ the sinuous contribution is always smaller than the varicose one. This is further confirmed by the streamwise distributions of the sinuous and varicose components along the liquid sheet axis reported in Fig. 2.19. From the figure we also observe that disturbances associated with the leading frequency and its first harmonic propagate with different velocities (in one case due to disturbances having both sinuous and varicose character, in the other predominantly varicose).

The excitation of the varicose mode when the Weber number is decreased, with a progressive shift from higher harmonics towards the leading frequency, can be interpreted as a prelude to the rupture experienced by the liquid sheet and experimentally observed when We is reduced by progressively reducing the inlet flow rate (e.g., de Luca and Meola [98]). [99] have also concluded that the progressive thinning of the sheet observed when the inlet Weber number is reduced down to the subcritical regime, can indeed be one of the rupture mechanisms of a three-dimensional liquid curtain configuration.

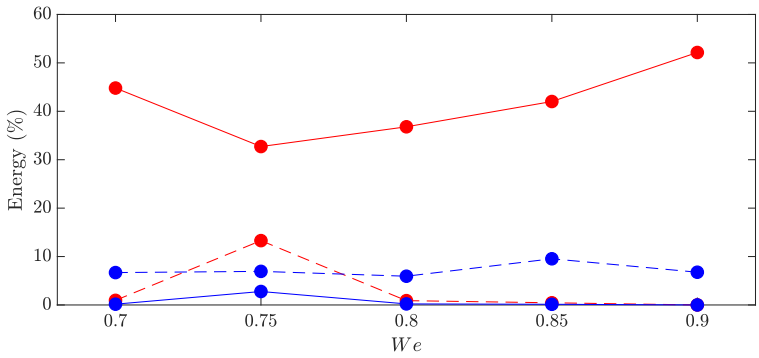
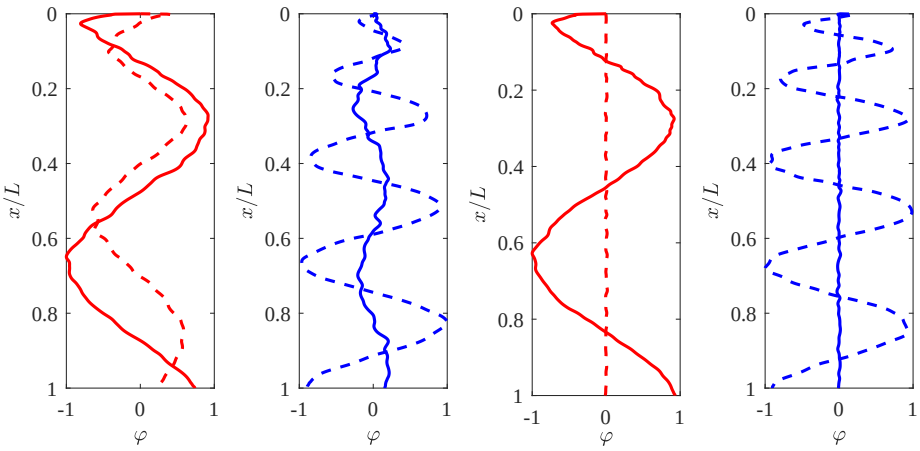


Figure 2.18. Subcritical regime (cases (e) and (h)-(k) of Table A.1). Energy content $(\%)$ of the sinuous (continuous lines) and varicose (dashed lines) components of the first SPOD mode as a function of the Weber number at $St = 0.076$ (red curves) and $St = 0.152$ (blue curves).

		We				
		0.7	0.75	0.8	0.85	0.9
$St = 0.076$	E^s $(\%)$	44.80	32.72	36.80	42.02	52.14
	E^v $(\%)$	0.94	13.28	0.88	0.44	0.00
$St = 0.152$	E^s $(\%)$	0.16	2.78	0.20	0.14	0.00
	E^v $(\%)$	6.68	6.94	5.94	9.54	6.76

Table 2.1. Subcritical regime (cases (e) and (h)-(k) of Table A.1). Sinuonus and varicose energy content of 1^{st} SPOD mode at leading frequency and its first harmonic.



(a) $St = 0.076$ **(b)** $St = 0.152$ **(c)** $St = 0.076$ **(d)** $St = 0.152$

Figure 2.19. Subcritical regime (cases (e) and (j) of Table A.1). Distributions of the first SPOD mode at the leading frequency ($St = 0.076$) and its first harmonic ($St = 0.152$). Solid lines represent the sinuous components, whereas dashed lines the varicose ones. Panels (a) and (b) refer to $We = 0.75$; panels (c) and (d) to $We = 0.85$.

Field reconstruction

From the spectral decomposition, any variable can be reconstructed by applying the weighted inverse Fourier transform to the blocks as described in section 2.1. In order to make some considerations on low-order reconstruction, we have applied both SPOD and POD to reconstruct the volume fraction C for the subcritical flow conditions corresponding to cases (e) and (j), which exhibit a more relevant dynamics (with respect the supercritical conditions), presenting a varicose deformation of the liquid sheet. For the analysis, the field is reconstructed considering either a limited number of frequencies of the first SPOD-mode or its full spectrum.

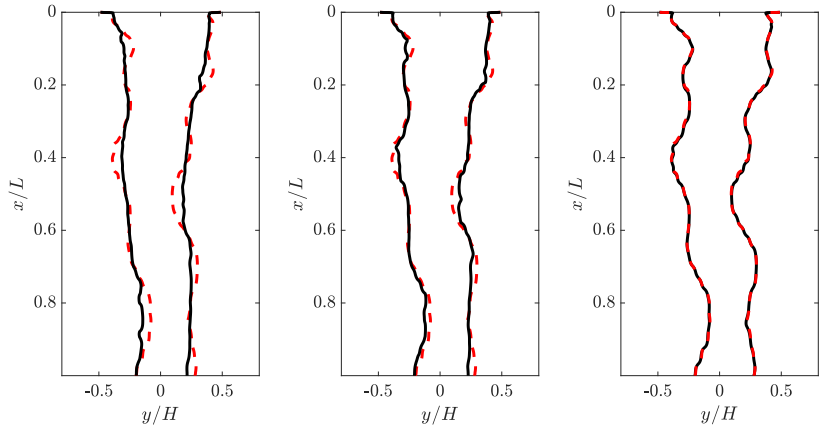
The reconstructed interface location at a given time instant considering only the leading frequency of the 1^{st} SPOD mode is reported in panel (a) of Fig. 2.20 in black solid line and compared with the original one (red dashed line); panels (b) and (c) show, respectively, the reconstructed field using two frequencies (the leading and its first harmonic) and the full spectrum (i.e. 188 frequencies) of the 1^{st} SPOD C' mode.

Let ER be the cumulative energy ratio defined as $\sum_{\tilde{m}} \lambda_{1,\tilde{m}} / \sum_{i,j} \lambda_{i,j}$, where $\tilde{m}$ is the running index of the selected frequencies. Table 2.2 shows that ER varies between 42.6% and 98.1% when considering, respectively, only the leading frequency and the whole spectrum of the 1^{st} mode and a (qualitatively) better interface reconstruction is obtained (Fig. 2.20). Let ε be the reconstruction error defined as

$$\varepsilon(t) = \frac{||\boldsymbol{q}(\boldsymbol{x},t) - \tilde{\boldsymbol{q}}(\boldsymbol{x},t)||}{||\boldsymbol{q}(\boldsymbol{x},t)||} \tag{2.8}$$

where $||\cdot||$ is the L_2 norm, $\boldsymbol{q}$ is the snapshot of the disturbance field at a given time and $\tilde{\boldsymbol{q}}$ its reconstruction; $\bar{\varepsilon}$ represents its temporal mean value.

The evolution of the reconstruction error for the case (j), considering one (leading frequency), two (leading frequency and its first harmonic) and all frequencies of the first SPOD mode, the first two and five SPOD modes, is reported in Fig. 2.21. As expected, as the number of selected frequencies and modes increases, ε progressively decreases. Due to windowing effects in the SPOD computation (Nekkanti and Schmidt [93]),



(a) 1 frequency of the 1^{st} SPOD mode.

(b) 2 frequencies of the 1^{st} SPOD mode.

(c) full spectrum of the 1^{st} SPOD mode.

Figure 2.20. Subcritical regime (case (j)) of Table A.1). Comparison of the original interface shape (dashed red line) and the reconstructed one with 1, 2 and all (panels (a) to (c), respectively) frequencies of the first SPOD mode.

	1 frequency of the 1^{st} SPOD mode	2 frequencies of the 1^{st} SPOD mode	all frequencies 1^{st} SPOD mode
ER(%)	42.6	52.4	98.1
$\bar{\varepsilon}$(%)	57.6	49.3	12.9

Table 2.2. Subcritical regime (case (j)) of Table A.1). Cumulative energy ratio ER and mean reconstruction error $\bar{\varepsilon}$.

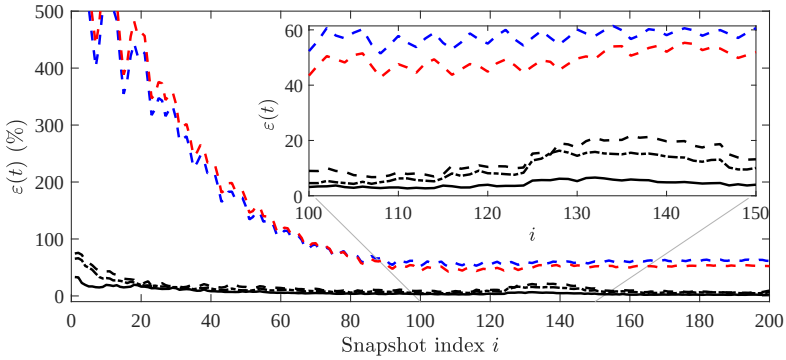


Figure 2.21. Subcritical regime (case (j)) of Table A.1). SPOD reconstruction error ε considering 1 (blue dashed line), 2(red dashed line) and all (black dashed line) frequencies of the first SPOD mode, the first two (black dash-dot line) and five modes (black solid line).

the error at the initial snapshots is largest. Case (j) exhibits a strong modal separation (Fig. 2.16) and using the full spectrum of the first SPOD mode suffices to reconstruct the original snapshot with an error of about 13%. Considering more than one SPOD mode for the reconstruction does not add any additional relevant information regarding the topology of the flow field.

At a lower Weber number ($We = 0.75$, case (e)), for which the liquid sheet has a dominant varicose character, the reconstruction error (see Fig. 2.22) varies between $O(10\%)$ and $O(5\%)$ when using, respectively, only the first mode (full spectrum) and the first five modes (full spectrum).

Fig. 2.23 compares the POD- and SPOD-based low order reconstruction of the liquid sheet shape of case (j). For the comparison, the number of modes and frequencies (for the SPOD) have been selected imposing a cumulative energy of about 52% of the total energy and considering only two frequencies (the leading and its first harmonic) of the first mode. For the POD we have selected the first two energetically leading modes for which $ER \approx 63\%$. The figure shows that at low order POD does not capture the varicose component, even though its energy content is greater.

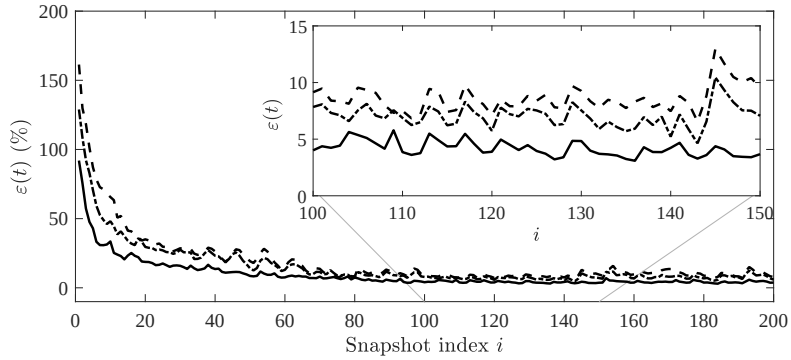


Figure 2.22. Subcritical regime (case (e)) of Table A.1). SPOD reconstruction error ε considering the full spectrum of the first SPOD mode (black dashed line), the first two (black dash-dot line) and five (black line) SPOD modes.

2.1.2 Application to 3D vertical liquid jet

In this section, the analysis of a three-dimensional (3D) configuration of the vertical liquid jet is presented. The SPOD technique introduced in section 2.1 has been employed for the investigation. Specifically, considering non-uniformity of the numerical grid and the heterogeneity of the variables included into the state vector $\boldsymbol{q}$, the eigenvalue problem of the cross-spectral density matrix $\boldsymbol{S}_k$ is modified including a weight matrix: $\boldsymbol{S}_k \boldsymbol{W} \boldsymbol{\Phi}_k = \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_k$. The weight matrix $\boldsymbol{W}$ is computed according to the energy norm:

$$E = \int_{\Omega} \left[\bar{\rho} \left(u'^2 + v'^2 + w'^2 \right) + \left(\bar{u}^2 + \bar{v}^2 + \bar{w}^2 \right) \frac{\Delta\rho^2}{\bar{\rho}} C'^2 \right] d\Omega, \qquad (2.9)$$

where $\Delta\rho = \rho_l - \rho_a$.

At moderately high We, the flow is stationary. In panel (a) of Figure 2.24 is reported a snapshot of the interface location for $We = 2.5$. In subcritical regime ($We < 1$) the flow in unsteady. Panel (b) of Figure 2.24 contains a snapshot of the interface location for $We = 0.75$ in which it is

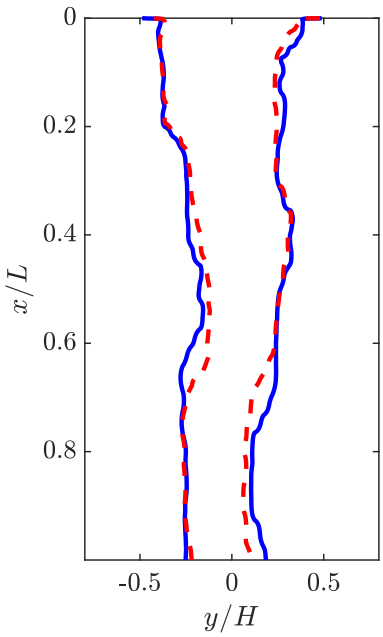
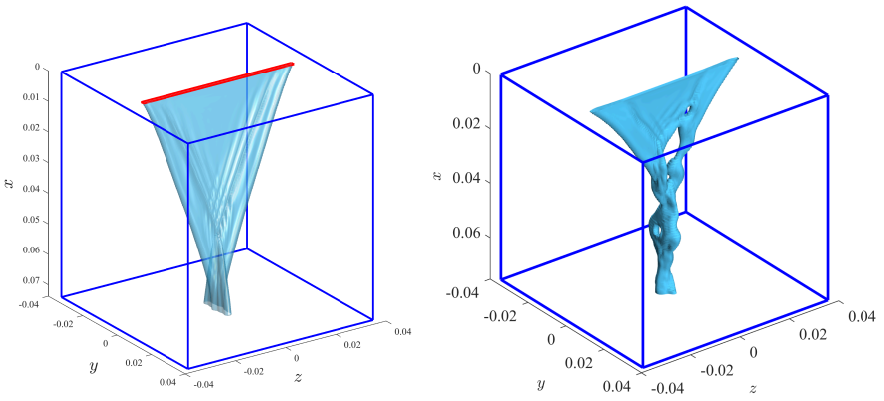


Figure 2.23. Subcritical regime (case (j)) of Table A.1). Comparison of POD- and SPOD-based reconstruction of the liquid sheet shape. Dashed red line, POD; solid blue line, SPOD.



(a) $We = 2.5$ **(b)** $We = 0.75$

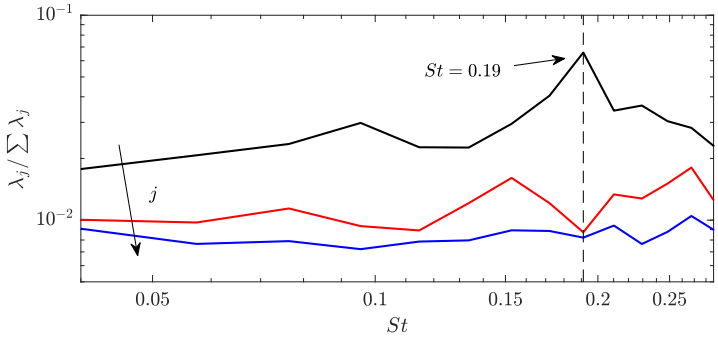
Figure 2.24. 3D vertical liquid jet: snapshot of interface shape for $We = 2.5$ and 0.75.

possible to notice that the flow regime characterized by three-dimensional asymmetric holes enucleation and advection and by irregular dynamics of the columnar curtain tail.

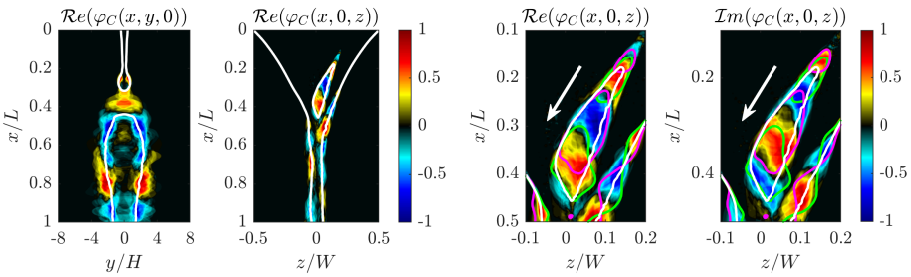
The application of SPOD highlights some interesting features of the flow field. In panel (a) of Figure 2.25 is reported the SPOD spectrum for the base case analysis ($We = 0.75$). The largest modal separation occurs at $St = 0.19$. In this section the Strouhal number is defined as $St = fH/U$, where H is the thickness of the curtain and U the mean inlet axial velocity. The inspection of the leading mode of C', reported in panels (b) and (c) for the principal planes xy and xz, reveals that the columnar curtain tail exhibits an irregular dynamics and, according to the definitions given in Colanera et al. [64], it reports different symmetries in the two main planes. Panels (d) and (e) contain a zoom of C' leading mode (real and imaginary parts) in the holes region. Green and magenta lines are two snapshots of the interface location. The spatial shift of real and imaginary parts highlights the advection direction.

The effect of the variation of the We is explored. We influences

both the leading frequency and the spatial topology of the leading modes. By looking at SPOD spectra at several We, reported in panel (a) of Figure 2.26, it is possible to notice that the flow presents leading frequency at $St = 0.06$ for $We = 0.9$. Decreasing the flow rate, that correspond to a decrease of the We, the leading frequency firstly increases and then decreases again. This occurrence can be further investigated by looking at the leading modes reported in panels (b)-(g). At $We = 0.9$ the flow if slightly unsteady and the modes exhibit a shape very close to that of the supercritical configurations (panels (b) and (e)). The region with $We \approx 0.75$ is characterized by rapid holes formation and advection. With a further decrease in We, holes dimension increase and the characteristic frequency decreases. Further details on the analysis can be found in Colanera et al. [100].



(a) SPOD spectrum



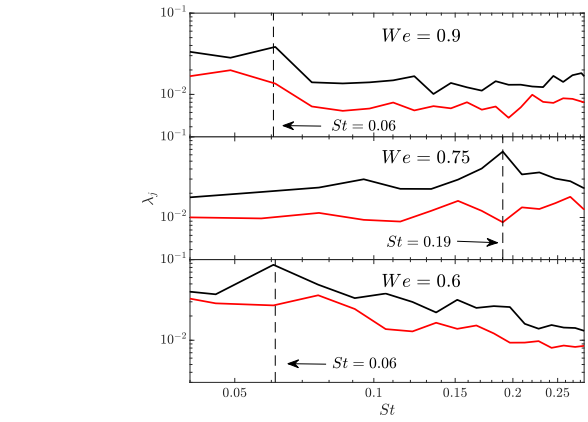
(b) C' mode (xy plane)

(c) C' mode (xz plane)

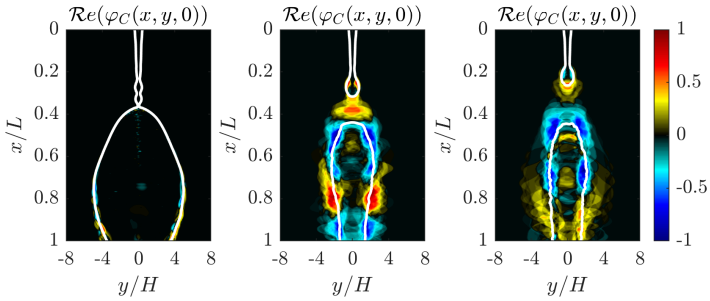
(d) Zoom of C' mode real part

(e) Zoom of C' mode imag. part

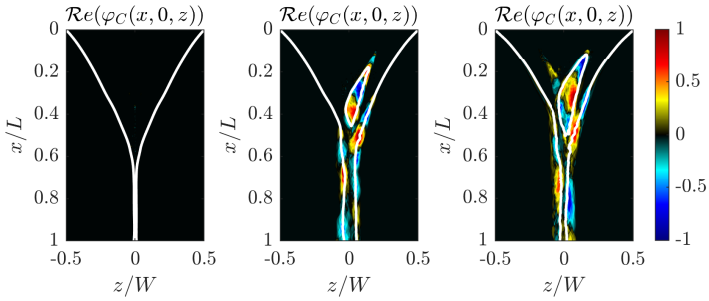
Figure 2.25. 3D vertical liquid jet: (a) SPOD spectrum for $We = 0.75$. SPOD parameters: $N_t = 230$, $N_b = 6$, $N_f = 64$, Hamming window for DFT, j is the running index associated with the modes. (b)-(c) Leading mode of C' in the planes xy and xz. (d)-(e) Zoom of C' leading mode (real and imaginary parts) in the holes region. The white arrows identify the advection direction.



(a)



(b) $We = 0.9$ **(c)** $We = 0.75$ **(d)** $We = 0.6$



(e) $We = 0.9$ **(f)** $We = 0.75$ **(g)** $We = 0.6$

Figure 2.26. Effect of the Weber number We. (a) SPOD spectra at several We. SPOD parameters: $N_t = 230$, $N_b = 6$, $N_f = 64$, Hamming window for DFT, only first 2 modes are reported. (b)-(g) Leading C' mode at several We.

2.2 Robust SPOD

SPOD algorithm, as seen in the section 2.1, is based on the eigenvalue decomposition od the cross spectral density (CSD) matrix. In this way, if the original data include outliers, the latter will be retrieved (and even amplified) in the Fourier realizations leading to corrupted and noisy modes. To avoid this occurrence, we propose the use of the robust PCA, introduced by Candes et al. [34] and reviewed by Scherl et al. [37], within the SPOD technique.

Specifically, we present two distinct procedures: one applies robust PCA to the Fourier realization $\hat{\mathbf{Q}}_{f_k}$, while the other directly applies it to the snapshots blocks $\mathbf{Q}^{(j)}$. The features and algorithms of these procedures will be discussed in detail hereafter.

The robust PCA can be used to decompose the matrix of the Fourier realizations $\hat{\mathbf{Q}}_{f_k}$ (or equivalently the snapshots blocks $\mathbf{Q}^{(j)}$), into

$$\hat{\mathbf{Q}}_{f_k} = \hat{\mathbf{L}}_{f_k} + \hat{\mathbf{H}}_{f_k}, \tag{2.10}$$

with $\hat{\mathbf{L}}_{f_k}$ being a low rank structure and $\hat{\mathbf{H}}_{f_k}$ a sparse matrix containing outliers and corrupted data, Candes et al. [34]. In this way the principal components of $\hat{\mathbf{L}}_{f_k}$ would not be affected by the presence of the incorrect data. To highlight how the robust PCA isolates outliers into the matrix $\hat{\mathbf{H}}_{f_k}$, it is insightful to consider that, for the standard POD, the mathematical problem for the extraction of the desired low rank (r) structures $\hat{\mathbf{L}}_{f_k}$ consists in the minimization of the Frobenius norm ($\|\cdot\|_F$) of the term $\hat{\mathbf{H}}_{f_k} = \hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{L}}_{f_k}$:

$$\min_{\hat{\mathbf{L}}_{f_k}} \|\hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{L}}_{f_k}\|_F \quad \text{subject to} \quad \text{rank}(\hat{\mathbf{L}}_{f_k}) \leq r. \tag{2.11}$$

The choice of the Frobenius norm makes the results of the computation highly sensitive to outliers.

For the robust PCA, instead, the optimization problem consists in:

$$\min_{\hat{\mathbf{L}}_{f_k},\hat{\mathbf{H}}_{f_k}} \operatorname{rank}(\hat{\mathbf{L}}_{f_k}) + \|\hat{\mathbf{H}}_{f_k}\|_0 \quad \text{subject to} \quad \hat{\mathbf{L}}_{f_k} + \hat{\mathbf{H}}_{f_k} = \hat{\mathbf{Q}}_{f_k}, \tag{2.12}$$

in which $\|\hat{\mathbf{H}}_{f_k}\|_0$ is the cardinality of nonzero elements in $\hat{\mathbf{H}}_{f_k}$ (featuring the sparsity of $\hat{\mathbf{H}}_{f_k}$). Problem (2.12) is non convex and typically is solved considering its *convex relaxation* (Candes et al. [34]):

$$\min_{\hat{\mathbf{L}}_{f_k},\hat{\mathbf{H}}_{f_k}} \|\hat{\mathbf{L}}_{f_k}\|_* + \alpha_0\|\hat{\mathbf{H}}_{f_k}\|_1 \quad \text{subject to} \quad \hat{\mathbf{L}}_{f_k} + \hat{\mathbf{H}}_{f_k} = \hat{\mathbf{Q}}_{f_k}, \tag{2.13}$$

in which $\|\cdot\|_*$ is the nuclear norm (i.e. the sum of the singular values) and $\|\cdot\|_1$ is the L_1 norm (i.e. the sum of the magnitudes of each entry in the matrix). The coefficient α_0 is defined as,

$$\alpha_0 = \alpha/\sqrt{\max(N, N_b)}, \tag{2.14}$$

where α has to be tuned and represents the filter intensity (Scherl et al. [37]).

The problem (2.13) is known as *principal component pursuit*, and may be solved using the augmented Lagrange multiplier (ALM) algorithm, see Lin et al. [36] and Yuan and Yang [35]. Following Candes et al. [34], in this work the augmented Lagrangian has been defined as:

$$\begin{aligned}\ell(\hat{\mathbf{L}}_{f_k}, \hat{\mathbf{H}}_{f_k}, \mathbf{Y}) = \|\hat{\mathbf{L}}_{f_k}\|_* + \alpha_0\|\hat{\mathbf{H}}_{f_k}\|_1 + \langle \mathbf{Y}, \hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{L}}_{f_k} - \hat{\mathbf{H}}_{f_k}\rangle + \\ + \frac{\mu}{2}\|\hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{L}}_{f_k} - \hat{\mathbf{H}}_{f_k}\|_F,\end{aligned} \tag{2.15}$$

in which $\mathbf{Y}$ is the matrix of Lagrange multipliers, $\langle\cdot,\cdot\rangle$ is the standard trace inner product and μ a parameter that counts the error in (2.10). In this work, the relaxation parameter μ has been chosen according to $\mu = 0.25NN_b/\|\hat{\mathbf{Q}}_{f_k}\|_1$. It is worth to notice that μ does not affect the solution of the problem (2.12) but only the convergence speed. The set of $(\hat{\mathbf{L}}_{f_k}, \hat{\mathbf{H}}_{f_k}, \mathbf{Y})$ that minimizes the (2.15) can be found in different ways,

as reported in Candes et al. [34], Lin et al. [36] and Yuan and Yang [35]. In the present work the alternating directions method (ADM) has been employed. Once the robust PCA algorithm converges, the so-obtained de-noised Fourier realizations matrix $\hat{\mathbf{L}}_{f_k}$ is employed instead of $\hat{\mathbf{Q}}_{f_k}$ in the standard SPOD procedure.

In case of the application of robust PCA directly on the snapshots blocks $\mathbf{Q}^{(j)}$, it extracts a the low rank blocks matrices $\mathbf{L}^{(j)}$ and the sparse matrices $\mathbf{H}^{(j)}$ containing the noise. In this case the denoised matrices $\mathbf{L}^{(j)}$ are employed for the computation of the CSD matrix following the standard SPOD algorithm.

It is worth to notice that standard ADM method deals with real matrices, whereas in case of the procedure considering the decomposition of the matrices $\hat{\mathbf{Q}}_{f_k}$ it has to deal with complex matrices. For this reason shrinkage operator $\mathscr{S}$ is not defined as usual with $\mathscr{S}(\tau, \mathbf{B}) = \text{sign}(\mathbf{B}) \max(|\mathbf{B}| - \tau, 0)$ but as shown in the following Equation (2.16).

The robust SPOD algorithms proposed here are reported in the following section. Algorithm 1, shows the robust SPOD algorithm with denoising of the $\hat{\mathbf{Q}}_{f_k}$, the robust PCA procedure is reported in Algorithm 2, while Algorithm 3 illustrates the robust SPOD with Principal Component Pursuit on each block.

Robust SPOD algorithms

In this section the mayor features of the proposed methodology will be hereafter outlined.

Algorithm 1 Robust spectral proper orthogonal decomposition algorithm.

1. For each data block $j = 1, ..., N_b$:

   (a) assemble data matrices

$$\mathbf{Q}^{(j)} = \left[\boldsymbol{q}^{(j)}_{1+(j-1)(N_f-N_o)}, \boldsymbol{q}^{(j)}_{2+(j-1)(N_f-N_o)}, ..., \boldsymbol{q}^{(j)}_{N_f+(j-1)(N_f-N_o)} \right],$$

with N_o being the number of overlapping snapshots.

(b) Within each block perform weighted FFT

$$\hat{\mathbf{Q}}^{(j)} = \mathscr{F}\left\{\mathbf{Q}^{(j)}\boldsymbol{W}_T\right\},$$

where $\boldsymbol{W_T}$ is the matrix of the window weights. Columns of $\hat{\mathbf{Q}}^{(j)}$, $\hat{\boldsymbol{q}}_k^{(j)}$ are the j_{th} realizations of the Fourier mode at the k^{th} discrete frequency f_k.

2. For each frequency f_k, with $k = 1, ..., N_f$:

(a) Collect Fourier transform realizations in the matrices $\hat{\mathbf{Q}}_{f_k} = \sqrt{\kappa}\left[\hat{\boldsymbol{q}}_k^{(1)}, \hat{\boldsymbol{q}}_k^{(2)}, ..., \hat{\boldsymbol{q}}_k^{(N_b)}\right]$, where $\kappa = \Delta t/(N_b\|\boldsymbol{W}_T\|_F^2)$.

(b) Split $\hat{\mathbf{Q}}_{f_k}$ in $\hat{\mathbf{L}}_{f_k}$ and $\hat{\mathbf{H}}_{f_k}$ with robust PCA algorithm (see Algorithm 2). Store $\hat{\mathbf{H}}_{f_k}$ if needed.

(c) With the de-noised matrix $\hat{\mathbf{L}}_{f_k}$ calculate

$$\mathbf{M}_{f_k} = \hat{\mathbf{L}}_{f_k}^* \mathbf{W} \hat{\mathbf{L}}_{f_k}.$$

(d) Perform the eigendecomposition of $\mathbf{M}_{f_k}$

$$\mathbf{M}_{f_k} = \boldsymbol{\Theta}_{f_k}\boldsymbol{\Lambda}_{f_k}\boldsymbol{\Theta}_{f_k}^*.$$

For the k^{th} frequency, store de-noised SPOD eigenvalues $\boldsymbol{\Lambda}_{f_k}$.

(e) Compute de-noised SPOD modes for the k^{th} frequency with

$$\boldsymbol{\Phi}_{f_k} = \hat{\mathbf{L}}_{f_k}\boldsymbol{\Theta}_{f_k}\boldsymbol{\Lambda}_{f_k}^{-1/2}$$

3. Return $\boldsymbol{\Phi}_{f_k}$ and $\boldsymbol{\Lambda}_{f_k}$.

Algorithm 2 Robust PCA algorithm with Principal Component Pursuit by ADM (from Lin et al. [36] and Yuan and Yang [35]).

Given a matrix $\hat{\mathbf{Q}}_{f_k}$ and the parameters α_0 and μ:

1. Initialize matrices $\hat{\mathbf{H}}_{f_k}$, $\hat{\mathbf{L}}_{f_k}$ and Lagrange multipliers matrix $\mathbf{Y}$:

$$\hat{\mathbf{H}}^i_{f_k} = \mathbf{0},$$

$$\hat{\mathbf{L}}^i_{f_k} = \mathbf{0},$$

$$\mathbf{Y}^i = \mathbf{0}.$$

2. Update matrices

$$\hat{\mathbf{L}}^{i+1}_{f_k} = \mathscr{T}(1/\mu\,,\, \hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{H}}^i_{f_k} - \mathbf{Y}^i/\mu),$$

$$\hat{\mathbf{H}}^{i+1}_{f_k} = \mathscr{S}(\alpha_0/\mu\,,\, \hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{L}}^{i+1}_{f_k} + \mathbf{Y}^i/\mu),$$

$$\mathbf{Y}^{i+1} = \mathbf{Y}^i + \mu(\hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{L}}^{i+1}_{f_k} - \hat{\mathbf{H}}^{i+1}_{f_k}),$$

in which operators $\mathscr{T}$ is the singular value thresholding operator:

$$\mathbf{A} = \mathscr{T}(\tau, \mathbf{B}) = \mathbf{U}\mathscr{S}(\tau, \mathbf{S})\mathbf{U}^H \quad \text{with} \quad [\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{B}),$$

and $\mathscr{S}$ is the shrinkage operator defined as:

$$\mathbf{A} = \mathscr{S}(\tau, \mathbf{B}) = e^{\mathrm{i}\,arg(\mathbf{B})} \max(|\mathbf{B}| - \tau\,,\, 0) \tag{2.16}$$

3. Compute the convergence parameter ϱ:

$$\varrho = \frac{\|\hat{\mathbf{Q}}_{f_k} - \hat{\mathbf{L}}^{i+1}_{f_k} - \hat{\mathbf{H}}^{i+1}_{f_k}\|_F}{\|\hat{\mathbf{Q}}_{f_k}\|_F}. \tag{2.17}$$

4. If ϱ is greater than an assigned tolerance, go to point (ii), else return $\hat{\mathbf{L}}^{i+1}_{f_k}$ and $\hat{\mathbf{H}}^{i+1}_{f_k}$.

The algorithm corresponding to the application of the robust PCA algorithm before those of the FFT, directly on matrices $\mathbf{Q}^{(j)}$ is reported in algorithm 3. Indeed, if the FFT algorithm is carried out before the PCA, then the technique completely preserves the spectral information related to the dynamics occurring in small regions of the domain. Contrary, the application of the robust PCA technique before the FFT results in a complete filtering of the local dynamics.

Algorithm 3 robust SPOD with Principal Component Pursuit on each block.

1. For each data block $j = 1, ..., N_b$:

   (a) assemble data matrices

   $$\mathbf{Q}^{(j)} = \left[\boldsymbol{q}^{(j)}_{1+(j-1)(N_f-N_o)}, \boldsymbol{q}^{(j)}_{2+(j-1)(N_f-N_o)}, ..., \boldsymbol{q}^{(j)}_{N_f+(j-1)(N_f-N_o)} \right].$$

   (b) Split $\mathbf{Q}^{(j)}$ in $\mathbf{L}^{(j)}$ and $\mathbf{H}^{(j)}$ with RPCA algorithm (see Algorithm 2). Store $\mathbf{H}^{(j)}$ if needed.

   (c) Within each block perform weighted FFT

   $$\hat{\mathbf{Q}}^{(j)} = \mathscr{F}\left\{ \mathbf{L}^{(j)} \boldsymbol{W}_T \right\},$$

   where $\boldsymbol{W_T}$ is the matrix of the window weights. Columns of $\hat{\mathbf{Q}}^{(j)}$, $\hat{\boldsymbol{q}}^{(j)}_k$ are the j_{th} realizations of the Fourier mode at the k^{th} discrete frequency f_k.

2. For each frequency f_k, with $k = 1, ..., N_f$:

   (a) Collect Fourier transform realizations in the matrices $\hat{\mathbf{Q}}_{f_k} = \sqrt{\kappa} \left[\hat{\boldsymbol{q}}^{(1)}_k, \hat{\boldsymbol{q}}^{(2)}_k, ..., \hat{\boldsymbol{q}}^{(N_b)}_k \right]$, where $\kappa = \Delta t/(N_b \|\boldsymbol{W}_T\|^2_F)$.

   (b) Calculate

   $$\mathbf{M}_{f_k} = \hat{\mathbf{Q}}^*_{f_k} \mathbf{W} \hat{\mathbf{Q}}_{f_k}.$$

(c) Perform the eigendecomposition of $\mathbf{M}_{f_k}$

$$\mathbf{M}_{f_k} = \mathbf{\Theta}_{f_k} \mathbf{\Lambda}_{f_k} \mathbf{\Theta}^*_{f_k}.$$

For the k^{th} frequency, store de-noised SPOD eigenvalues $\mathbf{\Lambda}_{f_k}$.

(d) Compute de-noised SPOD modes for the k^{th} frequency with

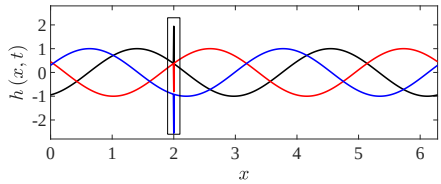
$$\mathbf{\Phi}_{f_k} = \hat{\mathbf{L}}_{f_k} \mathbf{\Theta}_{f_k} \mathbf{\Lambda}_{f_k}^{-1/2}$$

3. Return $\mathbf{\Phi}_{f_k}$ and $\mathbf{\Lambda}_{f_k}$.

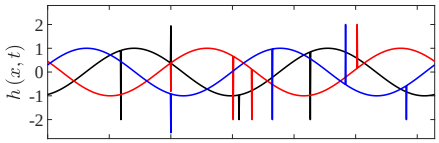
2.2.1 Differences between the proposed algorithms

To elucidate the differences between the latter two approaches, a toy signal denoted as $h(x,t) = \cos(x - \omega t)$, with the angular frequency $\omega = 2\pi f$ and the frequency $f = 1$, has been hereafter considered. The original signal has been reported in panel (a) of Figure 2.27 for different time instances with a phase delay of $45°$ (the temporal order of the reported curves is black, red and blue); in $x = 2$ the signal has been set to the value $\cos(2\omega t)$ and its location is highlighted by a black rectangle. Perturbing the signal with a salt and pepper noise (panel (b)), obtained summing to the 0.2% of corrupted data points a value equal to ± 2 at each time step, one can appreciate the differences in the local dynamics. Despite this low level of perturbation, the SPOD reconstruction (panel (c)) based on the entire first mode already appears particularly noisy. Panels (d) and (e) reports the robust SPOD application considering the de-noising phase before and after the FFT application, respectively. As expected, the local dynamics introduced in $x = 2$ has been completely removed in the panel (d).

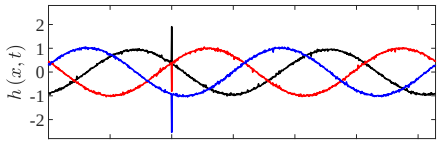
If one is not interested in preserving local sparse dynamics Algorithm 3 is preferred. In the following sections the latter is employed to real world data, the turbulent jet and the open cavity flow.



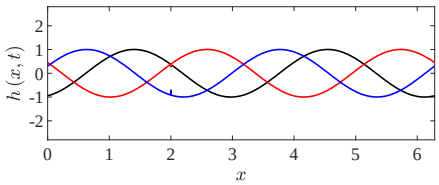
(a) Signal



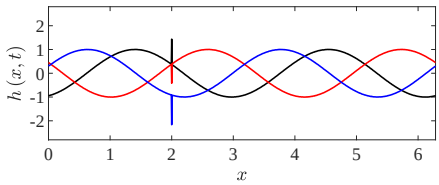
(b) Perturbed signal



(c) SPOD reconstruction



(d) Robust SPOD with RPCA before FFT

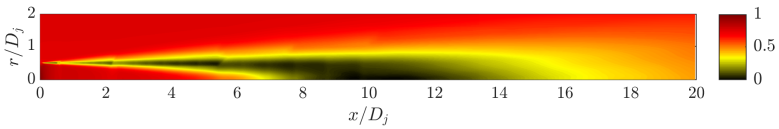


(e) Robust SPOD with RPCA past FFT

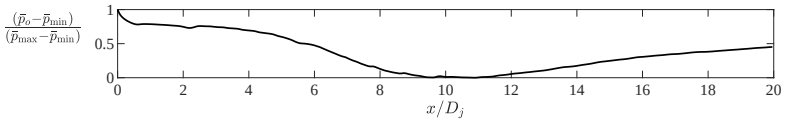
Figure 2.27. Comparison between SPOD and robust SPOD reconstructions of a toy signal. In panel (a) is reported the analyzed signal. Panel (b) contains the perturbed signal with salt and pepper noise. Panel (c) reports the SPOD reconstruction by the whole spectrum of 1^{st} SPOD mode. Panel (d) and (e) contain the robust SPOD reconstruction with de-noising stage before and past FFT, respectively.

2.2.2 Application of RSPOD to the turbulent jet

The first application considered in this work regards a turbulent jet presented in the work by Schmidt et al. [15]. The jet was numerically simulated by means of the unstructured flow solver "Charles" (Brès et al. [101]) at a subsonic Mach number (defined as the ratio between the mean inlet jet velocity U_j and the speed of sound a_j) equal to $M = 0.4$ and Reynolds number ($Re = U_j D_j/\nu$, with D_j the nozzle diameter and ν the kinematic viscosity) of $Re = 450000$; the large-eddy computations



(a)



(b)

Figure 2.28. Mean pressure distribution in the turbulent jet. Panel (a) depicts the mean pressure ($\bar{p_0}$) for azimuthal wavenumber $m = 0$, normalized as $(\bar{p}_0 - \bar{p}_{\min})/(\bar{p}_{\max} - \bar{p}_{\min})$. Panel (b) shows the normalized mean pressure $\bar{p}_0$ variation along the axis of the jet. (■■■ black, yellow, red, $0 < (\bar{p}_0 - \bar{p}_{\min})/(\bar{p}_{\max} - \bar{p}_{\min}) < 1$)

include also the nozzle geometry where synthetic turbulence combined with a wall model is applied to obtain a fully turbulent boundary layer. The data were interpolated onto a structured cylindrical grid of $[0, 30] \times [0, 6] \times [0, 2\pi]$, respectively, in the axial (x), the radial (r) and the azimuthal (θ) directions.

The analysis considers the pressure field p, which is decomposed in

$$p(x, r, \theta, t) = \bar{p}(x, r, \theta) + p'(x, r, \theta, t), \tag{2.18}$$

where $\bar{(\cdot)}$ is the long time mean, $(\cdot)'$ represents the fluctuating part. Moreover, p' is further decomposed in azimuthal Fourier modes

$$p'(x, r, \theta, t) = \sum_m \hat{p}_m(x, r, t) \mathrm{e}^{\mathrm{i}m\theta}, \tag{2.19}$$

being m the azimuthal wavenumber.

Panel (a) of figure 2.28 presents the spatial distribution of the mean pressure $\bar{p}_0$ ($m = 0$ Fourier component) within the turbulent jet; whereas

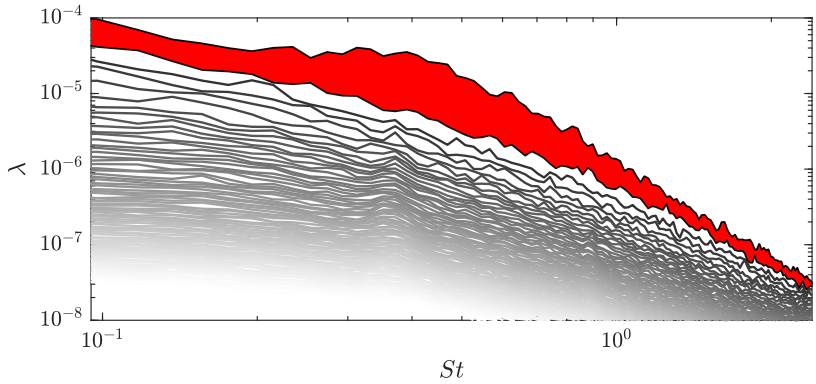


Figure 2.29. SPOD spectrum for the turbulent jet case. Red shaded area highlights the separation between the first and second mode.

panel (b) reports the normalized pressure profile along the jet axis. The terms $\bar{p}_{\max}$ and $\bar{p}_{\min}$ represent the maximum and the minimum of $\bar{p}_0$, respectively. The pressure is almost constant in the potential core region ($0 < x/D_j < 5$); then it decreases in the developing jet region and finally starts to increase near the beginning of the self-similar region ($5 < D_j < 25$).

The present analysis is carried out considering only the $m = 0$ Fourier component; at each time instance t_i the state vector $\boldsymbol{q}$ contains the evaluation of $\hat{p}_0$ in each point of the domain (x, r).

The standard SPOD algorithm is initially used to determine the principal frequencies of the turbulent jet, reporting also the main coherent structures of the 1^{st} and 2^{nd} mode; later, the data is artificially corrupted comparing the performance of the SPOD and robust SPOD procedures separately.

The snapshots have been sampled with a dimensionless time equal to 0.2, and a total of $N_t = 10000$ snapshots were considered. For the SPOD analysis, $N_f = 128$ snapshots has been used, considering a Hamming temporal window and a 50% overlap among blocks. This resulted in $N_b = 155$ blocks. It's worth noting that in this particular test, where only the $\hat{p}_0$ component was considered, the spatial weights were determined by

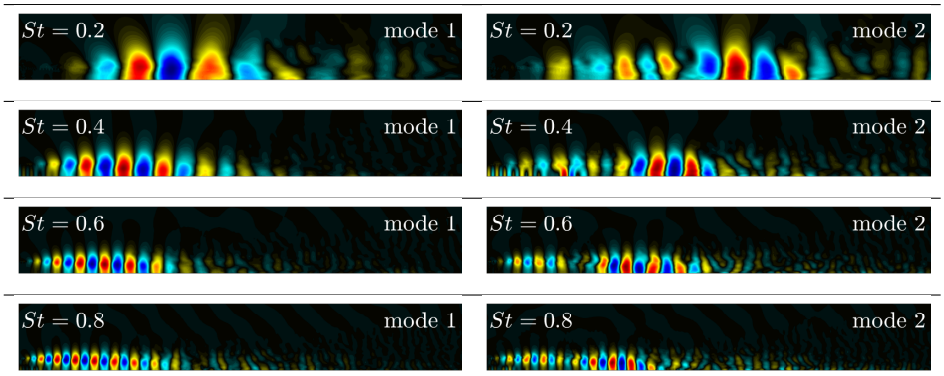


Figure 2.30. Real part of 1^{st} and 2^{nd} mode at different St. Modes have been normalized with respect their maximum. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j and radial coordinate r/D_j, respectively, omitted for clarity. (■■■■■ blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$)

taking into account only the integration quadrature, following the approach outlined by Schmidt and Colonius [85]. Figure 2.29 displays the SPOD spectrum as a function of the Strouhal number $St = fD_j/U_j$. The most prominent modal separation is observed at $St \approx 0.4$. Figure 2.30 exhibits the real part of the first two SPOD modes at different values of the Strouhal number. The pressure field takes the form of a compact wavepacket in the initial shear-layer region of the jet similar to a the Kelvin–Helmholtz shear-layer instability as shown by Schmidt et al. [15].

To assess the robustness of the SPOD method, artificial data corruption was introduced by randomly selecting a certain percentage of corrupted data points across the spatial domain at each time step. The pressure field hereafter represented is limited to the shear layer region only, $x/D_j < 4$, with a grid of $n_x \times n_y = 137 \times 45 = 6115$ points. Note that the robust SPOD algorithm was unaware of the specific locations of these corrupted data points. Two kinds of data corruption have been considered:

- Gaussian noise with zero mean and $4p_{\max}$ variance, $p_{\max}$ being the maximum of the $\hat{p}_0$ component;

- Salt and pepper noise with values of $\pm 4p_{\max}$.

This kind of noises introduce a high perturbation in the L_2 norm of the flow even with a low percentage of corrupted data points.

The differences between the two techniques are highlighted in Figure 2.31 which reports a comparison between the coherent structures of the 1^{st} mode at different St values, computed with the SPOD and robust SPOD technique; the results refer to the salt and pepper corruption of the 10% of the data and consider different values of α, defined in Eq.(2.14). Panels in first line contain the real part of the leading SPOD mode at $St = 0.4$, 0.8, and 1.2, considering the corrupted data; they show how a quite small percentage of corrupted data results in noisy modes, particularly at higher frequencies. The use of robust SPOD relevantly improves the modes quality, as confirmed by panels from (g) to (r). Here, the influence of the α parameter, as discussed in Scherl et al. [37], can be observed in terms of noise filtering. The higher the α the lower the filtering effects. The value $\alpha = 2$ gives the modes corresponding to the uncorrupted configuration (not shown herein), see Schmidt et al. [15].

The effectiveness of the robust SPOD technique is confirmed also with a Gaussian noise. Indeed, Figure 2.32 clearly show how the robust SPOD algorithm enhances the quality and the accuracy of the modes.

To quantify the validity of the robust SPOD, the relative error, denoted as ε_{fk}, is introduced as:

$$\varepsilon_{fk} = \frac{\|\boldsymbol{\phi}_{f_k}^{corr} - \boldsymbol{\phi}_{f_k}\|_2}{\|\boldsymbol{\phi}_{f_k}\|_2}, \tag{2.20}$$

in which $\boldsymbol{\phi}_{f_k}$ denotes the original SPOD modes and $\boldsymbol{\phi}_{f_k}^{corr}$ refers to the modes obtained from corrupted data using either SPOD or robust SPOD. This relative error provides a measure of the accuracy and reliability of the robust SPOD technique. The dependency of ε_{fk} with respect α and St has been reported in Figures 2.33 and 2.34, respectively. In particular Figures 2.33 refers to two different types of noise, various percentages of corrupted data points, and different values of α, specifically at the leading frequency $St = 0.4$. In both panels solid lines represent the ε_{f_k} at several

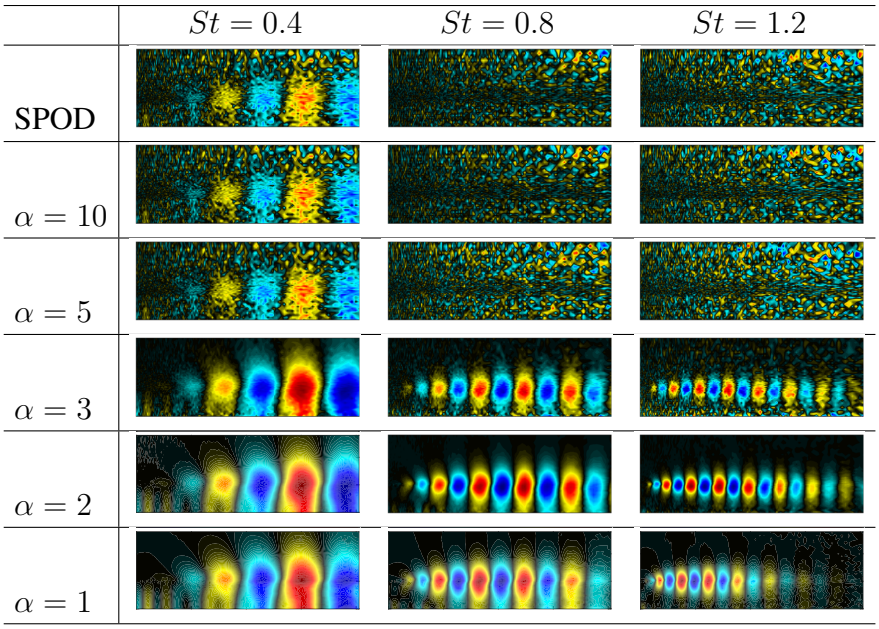


Figure 2.31. Comparison between SPOD and robust SPOD leading modes at different St. Salt and pepper noise with the 10% of corrupted data points. Panels in first line refer to SPOD modes of corrupted data. Other panels show the effect of the de-noising parameter α on the robust SPOD modes. Data corruption modelled as salt and pepper noise $\pm 4p_{\max}$. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j and radial coordinate r/D_j, respectively, omitted for clarity. Field variables have been normalized with respect to their maximum.
(blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$)

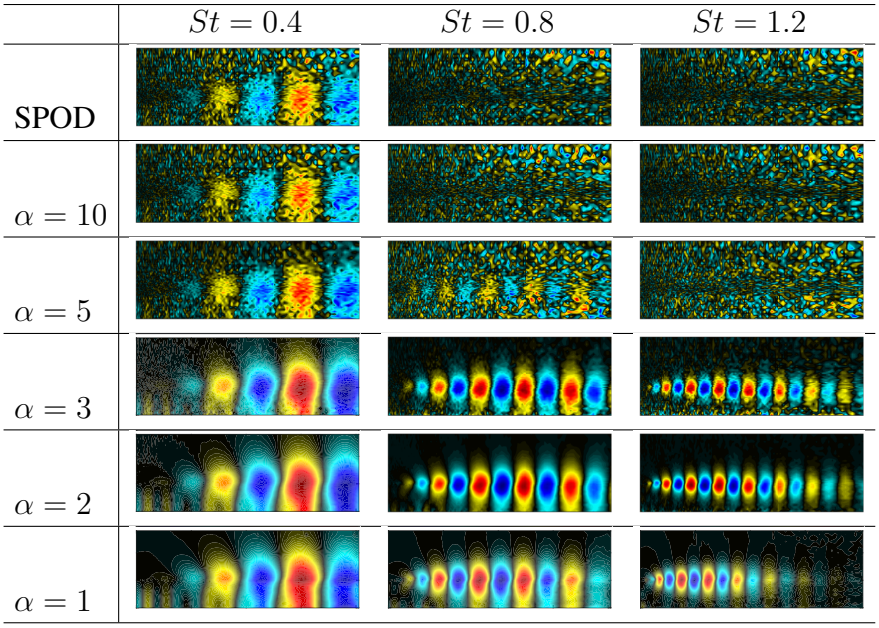
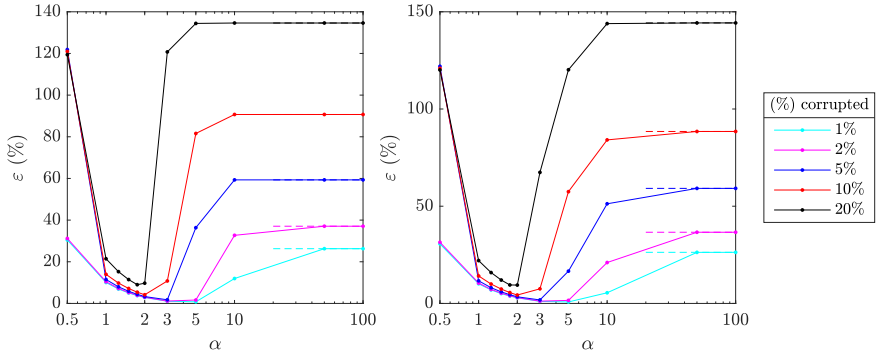


Figure 2.32. Comparison between SPOD and robust SPOD leading modes at different St. Gaussian noise with the 10% of corrupted data points. Panels in first line refer to SPOD modes of corrupted data. Other panels show the effect of the de-noising parameter α on the robust SPOD modes. Data corruption modelled as Gaussian noise with $\pm 4p_{\max}$ variance. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j and radial coordinate r/D_j, respectively, omitted for clarity. Field variables have been normalized with respect to their maximum.
(blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$)



(a) Salt and pepper noise **(b)** Gaussian noise

Figure 2.33. Relative error ε_{f_k} computed with SPOD and robust SPOD techniques at $St = 0.4$ for different levels and types of noise, considering several values of α. Solid lines represent the ε_{f_k} at several α, color-coded according to the corrupted data points percentage. Dashed lines represent the SPOD error, namely α approaching infinity.

α, color-coded according to the corrupted data points percentage. The dashed lines represent the SPOD error, specifically when α approaches infinity. Each curve exhibits its unique minimum for α, but regardless the error type or the percentage of corrupted data points, as in Scherl et al. [37], the lowest values of ε_{fk} are found when α is $\mathrm{o}(1)$, approximately in the range of $1 \sim 3$. Figures 2.34, instead, reports the results for the salt and pepper noise only, highlighting the influence of the de-noising parameter α on the mode error ε for the first four modes as a function of St. The effect of the de-noising parameter α on the mode error varies with the frequency of interest; the relative performance of robust SPOD tends to be more pronounced and advantageous at higher St values. For this case the optimal choice of α is 2.

The SPOD spectrum is strongly affected by α. Figure 2.35 reports both the SPOD spectra of the clean (red lines) and corrupted (dashed magenta curves) configurations and those obtained by the robust SPOD at various filter levels (black curves). The introduction of noise, even in

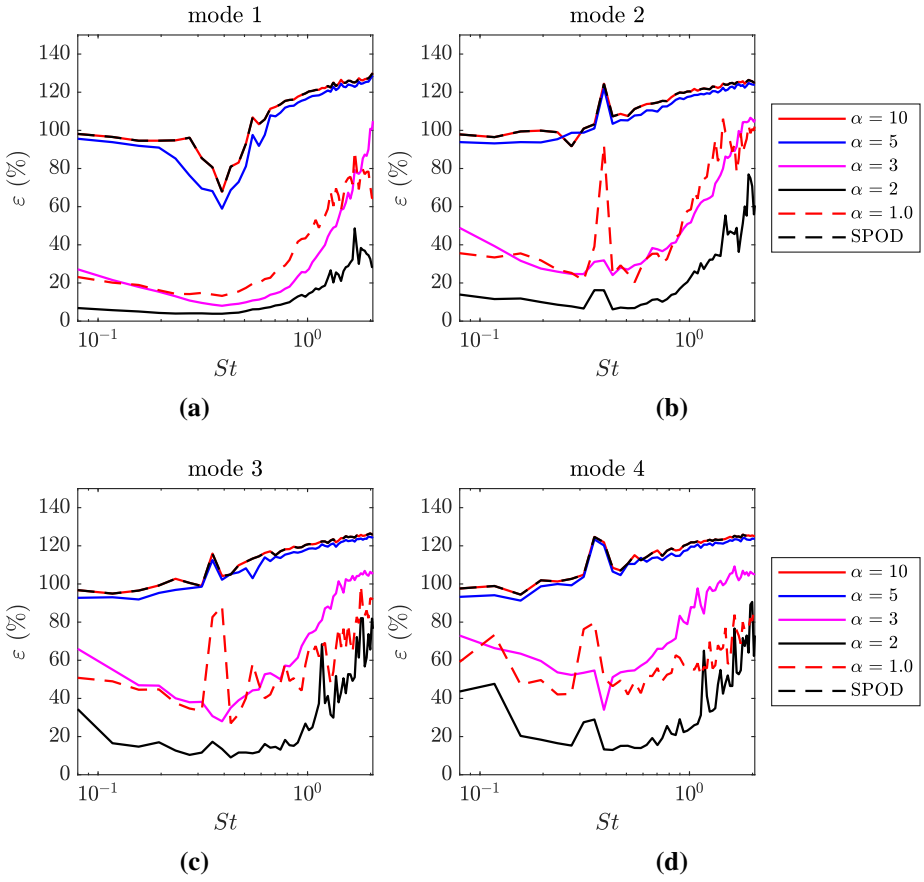
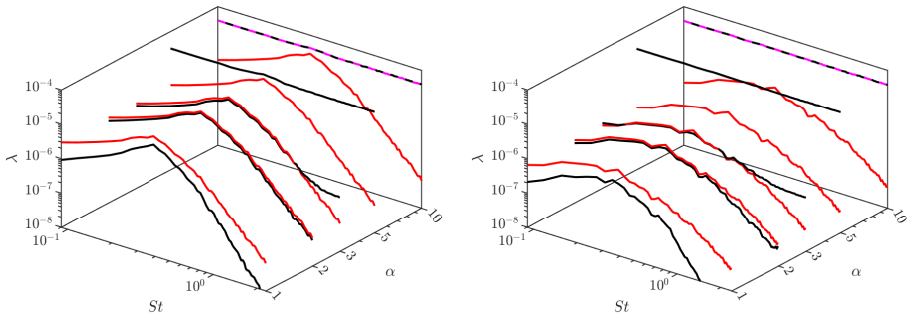


Figure 2.34. Effect of de-noising parameter α on the error ε. Panels (a) to (d) goes from the 1^{st} to the 4^{th} SPOD mode. Salt and pepper noise with 10% of corrupted data points.



(a) mode 1 **(b)** mode 2

Figure 2.35. Effect of de-noising parameter α on the robust SPOD spectrum. Panels (a) refers to 1^{st} SPOD mode, panel (b) to the 2^{nd} one. Black lines represent the robust SPOD spectra and red lines the original SPOD spectrum. Dashed magenta lines are the SPOD spectra of corrupted data. Salt and pepper noise with 10% of corrupted data points.

a limited number of data points, significantly perturbs the spectrum. At high α values the robust SPOD produces exactly the same spectrum of the SPOD technique; a reduction of α is accompanied by a decrease in the noise levels and, thus, the robust spectrum tends towards the clean one (the best result is obtained for $\alpha = 2$). This optimal value strikes a balance between noise reduction and preservation of relevant flow features.

Robust SPOD can be also a valuable tool for de-noising data and reconstructing the signal from de-noised modes. This ability is reported in Figure 2.36, which contains a snapshot of the original pressure fluctuation (first panel), the corresponding perturbed field (second panel) considering the 10% of corrupted data points, the SPOD reconstruction (third panel) and various robust SPOD reconstructions (lower panels). The SPOD reconstruction, based on the entire spectrum of the first three modes, shows noticeable noise levels. Panels (d) to (i) demonstrate the de-noising effect achieved by adjusting the parameter α, which increases as α reduces. For values lower than 5 the reconstructed snapshots exhibit a remarkable

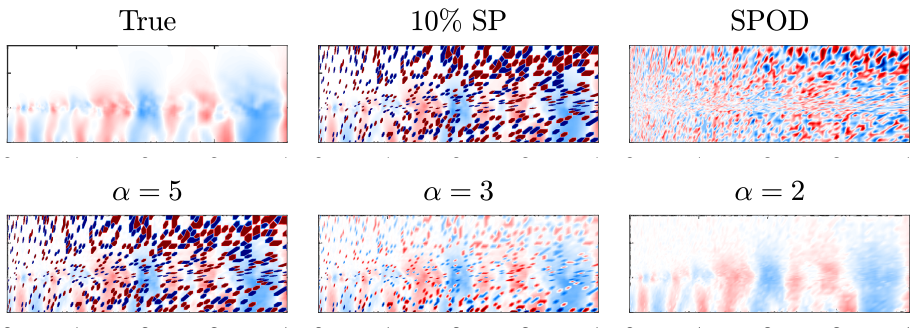


Figure 2.36. De-noising with robust SPOD. First panel contains a sample snapshot of $\hat{p}_0$. Second panel shows the 10% of corrupted data points with salt and pepper noise. Third panel reports an attempt of de-noising with SPOD by employing the whole spectrum of the first 3 modes. Lower panels show the robust SPOD reconstruction by employing the whole spectrum of all the robust SPOD modes, highlighting the de-noising effect of the parameter α.

reduction in noise, leading to a clearer representation of the underlying flow structures.

Finally, the robust spod algortithms using the turbulent jet data are compared. At the location $x/D_j = 0.3$ along the axis $r = 0$ a synthetic cosine signal with a frequency corresponding to $St_f = 1.2$ has been implemented. The first panel of Figure 2.37 presents a zoomed-in view near the inlet section ($0 < x/D_j < 1$) of the leading pressure SPOD mode at $St = 1.2$. This mode exhibits the oscillating pressure pattern previously observed in the paper and captures the synthetic signal. Subsequently, salt and pepper noise is introduced to corrupt only 0.2% of the dataset. The second panel of the figure displays the same SPOD mode obtained from the corrupt dataset. While the synthetic signal remains detectable, the noise introduced by even such a low number of corrupted data points results in significantly noisier modes. The lower panels of the figure demonstrate the application of both robust SPOD algorithms, both with the same denoising parameter $\alpha = 2$. It is evident that Algorithm 1, while providing less smooth modes, preserves the local sparse dynamics. On

the other hand, Algorithm 3 should be preferred if local behaviors are not of a primary interest.

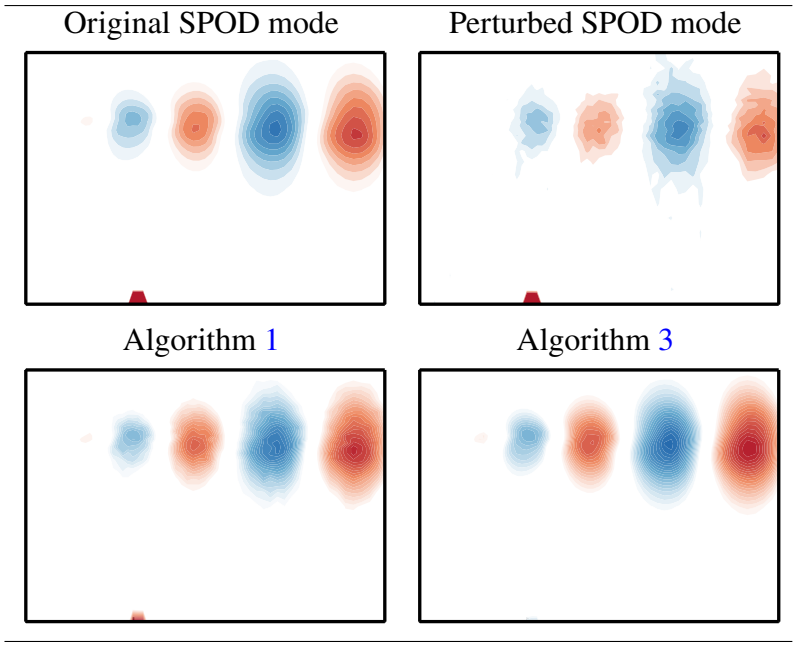


Figure 2.37. Comparison between robust SPOD algorithms using the turbulent jet data with a syntetic cosine signal at $x/D_j = 0.3$. Salt and pepper noise with the 0.2% of corrupted data points has been considered. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_j ($0 < x/D_j < 1$) and radial coordinate r/D_j ($0 < r/D_j < 0.7$), respectively, omitted for clarity. The first panel shows the leading SPOD mode at $St = 1.2$ near the inlet section. The second panel displays the same SPOD mode obtained from the corrupt dataset. The lower panels illustrate the application of the two robust SPOD algorithms with the same denoising parameter $\alpha = 2$. Field variables have been normalized with respect to their maximum.
(■ ■ blue, white, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$)

2.2.3 Application of RSPOD to the open cavity flow

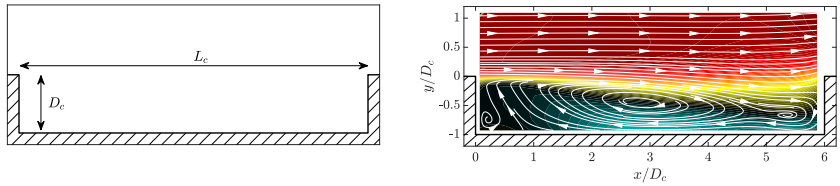
The second investigation concerns the flow field within an open cavity, whose measurements have been acquired by means of PIV technique. Such a configuration is of significant industrial relevance, given its complex spatio-temporal dynamics, and raises several unresolved questions. The standard SPOD tends to produce noisy outcomes, especially for non-leading modes. As demonstrated later in this dissertation, the application of robust SPOD enhances the smoothness and physical interpretability of these modes. The dataset reported in Zhang et al. [102] has been considered here limited to a free stream Mach number (defined as the ration between the incoming flow velocity U_c and the speed of sound a_c) equal to $M = 0.6$.

Panel (a) of Figure 2.38 reports a sketch of the rectangular cavity. It has a length of $L_c = 158.8$ mm, a depth of $D_c = 26.4$ mm, and a width of $W_c = 101.6$ mm; the resulting nondimensional ratios are $L_c/D_c = 6$ and $W_c/D_c = 3.85$.

In accordance with the panel (b) of Figure 2.38, reporting the temporal mean distribution of the streamwise velocity component, the incoming boundary layer separates at the cavity leading edge forming a shear layer, which can induce acoustic resonance phenomena. Indeed, the shear layer formed at the cavity leading edge can generate small disturbances, that grow moving downstream and then interact with the trailing edge; the resulting acoustic field radiates waves and those traveling upstream can feed and generate further disturbances in the shear layer (Zhang et al. [102], Sun et al. [103, 104]). This feedback process results in an aeroacoustic resonance whose main tones are known as Rossiter modes (Rossiter [105]). The characteristic dimensionless frequencies (in terms of the Strouhal number, St) associated with these modes can be predicted using the empirical relation from Heller and Bliss [106]:

$$St_n = \frac{f_n L_c}{U_\infty} = \frac{n - a}{1/\kappa_r + M/\sqrt{1 + (\gamma - 1)M^2/2}}. \tag{2.21}$$

Here, U_∞ represents the freestream velocity, $\kappa_r = 0.65$ is an empirical



(a) **(b)**

Figure 2.38. Open cavity flows. Panel (a): geometrical sketch; panel (b): mean distribution of streamwise velocity component and streamlines. (■■■■■ blue, cyan, black, yellow, red, $-1 < \bar{u}/\|\bar{u}\|_\infty < 1$)

coefficient, $a = 0.38$ is the phase lag, and $n = 1, 2, \ldots$ corresponds to the Rossiter mode index. Equation (2.21) provides a prediction for the dimensionless frequencies St_n associated with the resonance modes in open cavities.

The open flow cavity data-set consists of velocity component measurements on a uniform grid with dimensions $n_x \times n_y = 156 \times 55$, with a sampling frequency of $f_s = 16000$ Hz. The analysis is based on $N_t = 16000$ snapshots, with $N_b = 30$ blocks and $N_f = 1024$ frequencies. The corresponding spectrum has been reported in Figure 2.39; a good agreement between the leading frequencies and the Rossiter frequencies, predicted by Equation (2.21) and indicated by vertical dashed lines, can be appreciated.

Figure 2.40 presents the real part of the first SPOD mode at the two leading frequencies, $St = 0.75$ and $St = 1.22$. Panels (a) and (b) show the streamwise component of the mode, while panels (c) and (d) report the transversal component. The spatial structures at higher frequencies exhibit finer scales compared to those at lower frequencies. Panels (d) and (e), instead, contains the magnitude of velocity mode gradients, defined as the square root of the sum of squares of the individual velocity mode

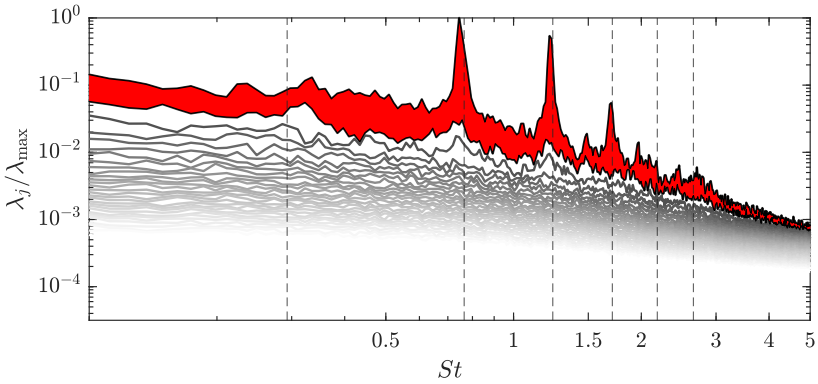


Figure 2.39. SPOD spectrum of open cavity flows from Zhang et al. [102]. Case with $M = 0.6$. Vertical dashed lines represents Rossiter frequencies from equation (2.21). Red shaded area highlights the separation between the first and second mode.

gradient components:

$$\xi_{f_k}(x,y) = \sqrt{\left(\frac{\partial \varphi_{u_{f_k}}}{\partial x}\right)^2 + \left(\frac{\partial \varphi_{u_{f_k}}}{\partial y}\right)^2 + \left(\frac{\partial \varphi_{v_{f_k}}}{\partial x}\right)^2 + \left(\frac{\partial \varphi_{v_{f_k}}}{\partial y}\right)^2}. \tag{2.22}$$

It provides important information about the spatial distribution and intensity of flow structures within the open cavity. Indeed, these distributions have the largest magnitude in the shear layer, which indicates that Rossiter modes are mainly driven by the shear layer of the flow, as highlighted in Sun et al. [103].

For this case, the robust SPOD analysis is based on the algorithm 3, with a de-noising parameter $\alpha = 1$. Figure 2.41 presents the spectra obtained from both SPOD (red lines) and robust SPOD (black lines) analysis. The robust SPOD analysis primarily focuses on reducing noise at high frequencies, meaning that the obtained reconstructed signals tend to exhibit less noise contamination compared to those obtained from SPOD, particularly in the higher frequency range. To provide a reference for the spectral behavior, the blue dashed line represents the Kolmogorov

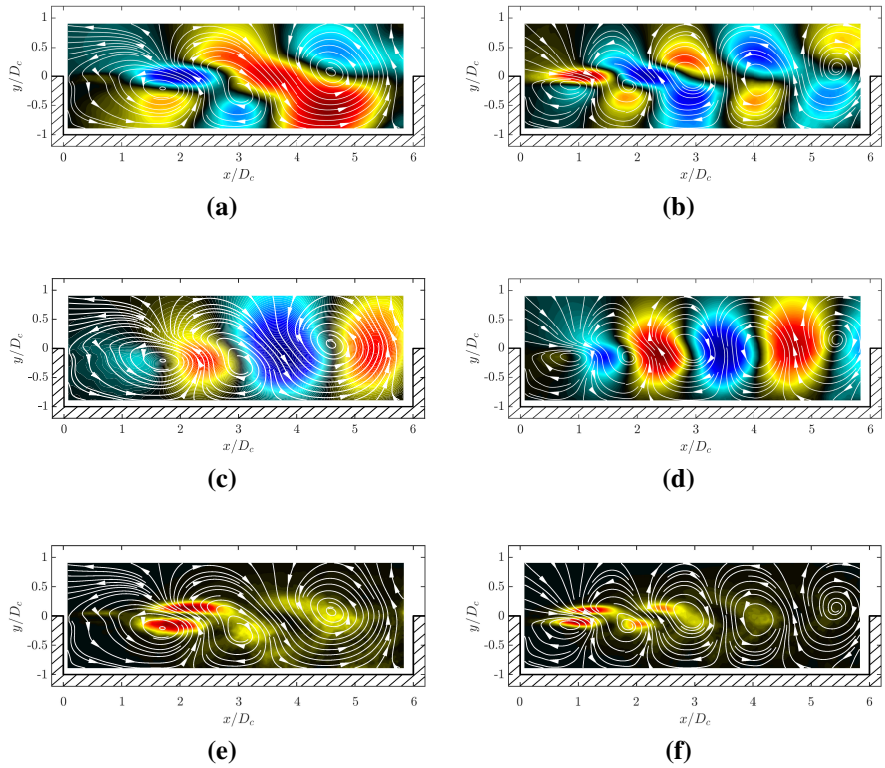


Figure 2.40. Leading SPOD modes of open cavity flows data from Zhang et al. [102]. Panel (a): real part of leading SPOD mode of u at $St = 0.75$. Panel (b): real part of leading SPOD mode of u at $St = 1.22$. Panel (c): real part of leading SPOD mode of v at $St = 0.75$. Panel (d): real part of leading SPOD mode of v at $St = 1.22$. Panel (e): magnitude of velocity mode gradient m_{f_k} at $St = 0.75$. Panel (f): magnitude of velocity mode gradient ξ_{f_k} at $St = 0.75$. White lines are the sreamlines pattern as viewed by an observer moving with the mean flow. Field variables have been normalized with respect to their maximum.

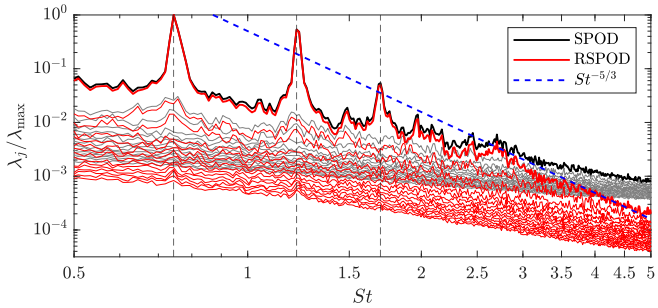


Figure 2.41. Comparison between the SPOD spectrum (black) and the robust SPOD one (red) for the cavity flow. The reference blue dashed curve represents $St^{-5/3}$. Vertical dashed lines represents Rossiter frequencies. Note that for robust SPOD the Algorithm 3 has been employed.

power law ($St^{-5/3}$), which is a well-established representation of the energy spectrum in turbulent flows and serves as a benchmark for assessing the spectral characteristics of the analyzed data (Wilczek et al. [107]), showing how the curve tends to align with this well-known benchmark. As expected, the power law scaling is not valid at low frequencies, where the flow is characterized by relatively larger coherent structures.

Figures 2.42 and 2.43 provide a comparison between the leading u and v modes, respectively, obtained from both SPOD and robust SPOD analysis. The comparison is done for the first four Rossiter frequencies as well as a generic high Strouhal number ($St = 4$).

It is worth noting that the application of robust SPOD leads to considerable improvements in the smoothness of the extracted modes. Specifically, for the first mode at the leading frequencies ($St = 0.75$ and $St = 1.70$), SPOD already performs well, producing relatively smooth modes. However, for the remaining modes, the employment of robust SPOD significantly enhances the smoothness of the modes. Moreover, at high frequencies, such as $St = 4$, the coherent structures, extracted with SPOD technique, are not easily recognizable; whereas the robust SPOD modes exhibit clearer and more distinguishable coherent structures.

To quantify the roughness of the modes, following Grohmann et al.

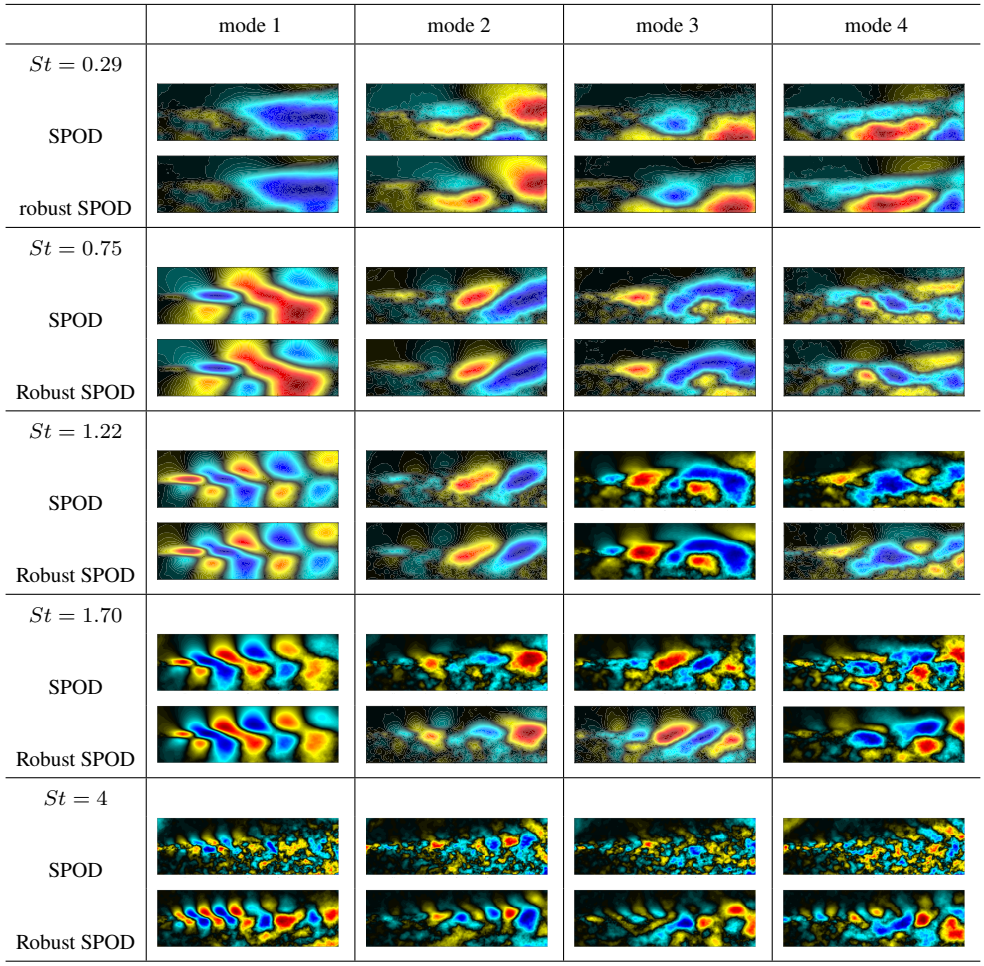


Figure 2.42. Comparison between SPOD and robust SPOD leading u modes at different St. Field variables have been normalized with respect to their maximum. Note that for robust SPOD the Algorithm 3 has been employed. De-noising parameter $\alpha = 1$. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_c and radial coordinate r/D_c, respectively, omitted for clarity. (■■■■ blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$)

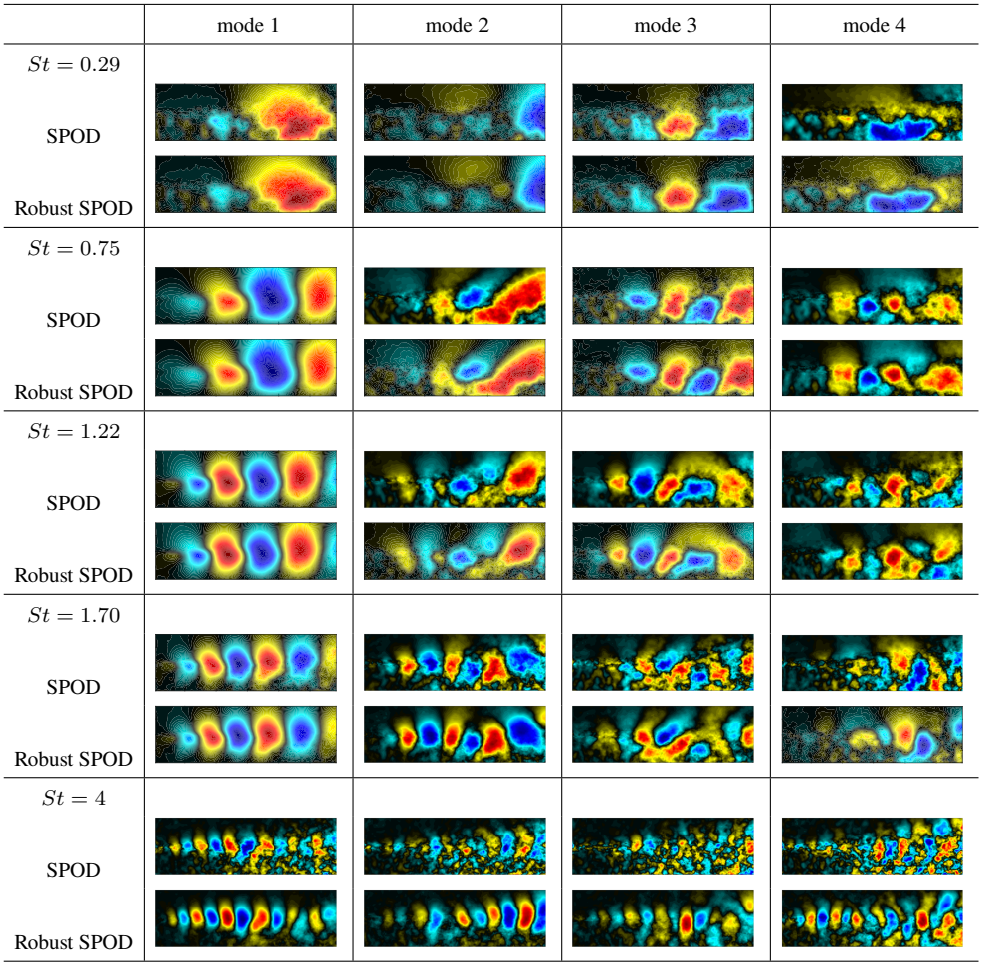
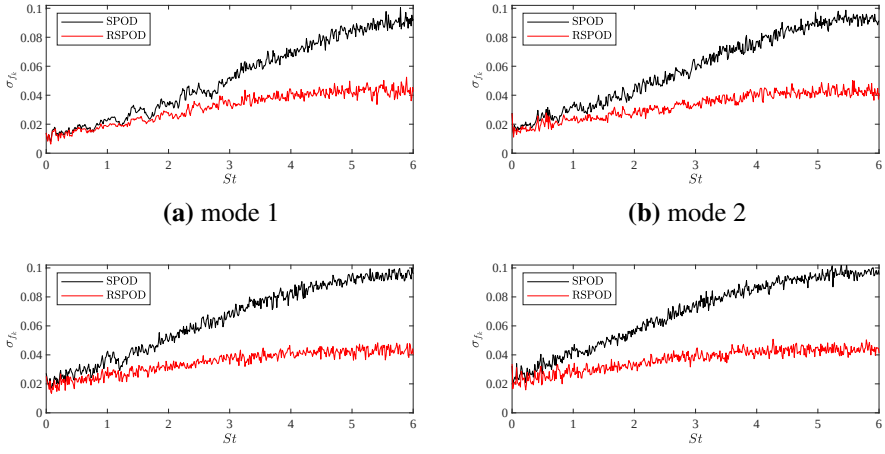


Figure 2.43. Comparison between SPOD and robust SPOD leading v modes at different St. Field variables have been normalized with respect to their maximum. Note that for robust SPOD the Algorithm 3 has been employed. De-noising parameter $\alpha = 1$. In all the panels the abscissa and ordinate are the dimensionless axial coordinate x/D_c and radial coordinate r/D_c, respectively, omitted for clarity. (blue, cyan, black, yellow, red, $-1 < \phi_{f_k}/\|\phi_{f_k}\|_\infty < 1$)



(a) mode 1 **(b)** mode 2

(c) mode 3 **(d)** mode 4

Figure 2.44. Spatial standard deviation of the velocity mode gradient magnitude. Comparison between SPOD and robust SPOD.

[108], it is convenient to compute the spatial standard deviation of the velocity mode gradient magnitude defined in Eq.(2.22):

$$\sigma_{f_k} = \mathrm{std}(\xi_{f_k}) = \sqrt{\frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\left(\xi_{f_k}(x_i, y_j) - \bar{\xi}_{f_k}\right)^2} \tag{2.23}$$

where $\bar{\xi}_{f_k} = \frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\xi_{f_k}(x_i, y_j)$. Figure 2.44 displays the values of σ_{f_k} for the leading four modes obtained from both SPOD and robust SPOD analysis. Comparing the two sets of modes, it is evident that the robust SPOD modes exhibit greater smoothness, particularly at high Strouhal numbers (St). The lower values of σ_{f_k} for robust SPOD modes indicate reduced roughness and enhanced coherence in flow field structures.

2.3 Gappy POD of two-phase PIV data

Experimental measurements often contain corrupted data. This section presents a modal analysis that explores the spatially dominant coherent structures and their corresponding temporal dynamics in presence of missing data. This analysis is based on time-resolved particle image velocimetry (TR-PIV) applied to a wake-mixing layer flow, where the velocity field is simultaneously measured in both gas and liquid phases. To identify the leading modes, we have utilized the Gappy POD (Proper Orthogonal Decomposition) iterative algorithm, initially introduced in Everson and Sirovich [27] and subsequently reviewed in Venturi and Karniadakis [28] and Gunes et al. [29].

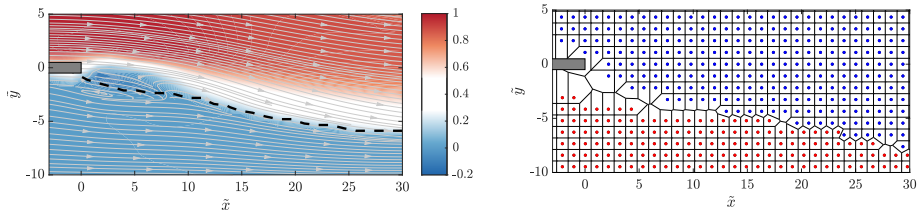
Specifically, for this analysis, the state vector $\mathbf{q}$ comprises fluctuations in both velocity components at each point $\mathbf{x}$ within the flow field. It's important to note that measurement locations in $\mathbf{x}$ originate from two separate cameras: one for the liquid phase and the other for the gaseous phase. To address this heterogeneity within the modal analysis, a spatial weight matrix, which defines the norm of the state vector, has been designed to account for the varying densities of the two phases, ρ_a for the gaseous phase and ρ_l for the liquid phase, as well as the Voronoi area A_v associated with each point.

In Figure 2.45 are presented the mean axial velocity $\bar{u}$ in Panel (a) and the Voronoi cells corresponding to the measurements points in Panel (b). Blue dots represent measurements from the gaseous phase camera (GC), while red dots represent measurements from the liquid phase camera (LC). In the Figure, $\tilde{x} = x/e$ and $\tilde{y} = y/e$, where e represents the thickness of the splitter plate.

The spatial weight function $w(\mathbf{x})$ is defined as follows:

$$w(\mathbf{x}) = \begin{cases} \rho_a A_v(\mathbf{x}) & \text{if the field variable is recorded by GC,} \\ \rho_l A_v(\mathbf{x}) & \text{if the field variable is recorded by LC.} \end{cases} \tag{2.24}$$

To address the presence of corrupt measurements, we introduce the



(a) $\bar{u}$ **(b)** Voronoi cells

Figure 2.45. Visualization of the experimental setup: Panel (a) shows the mean axial velocity $\bar{u}$, while Panel (b) displays Voronoi cells corresponding to the measurements. Black ashed lines represent the mean interface location. Blue dots represent measurements from the gaseous phase camera, and red dots represent measurements from the liquid phase camera.

variable $\upsilon(\mathbf{x}, t)$, which tracks spatio-temporal missing data:

$$\upsilon(\mathbf{x}, t) = \begin{cases} 1 & \text{if the field variable is known at } (\mathbf{x}, t), \\ 0 & \text{if the field variable is missing at } (\mathbf{x}, t). \end{cases} \tag{2.25}$$

Algorithm 4 outlines the Gappy POD procedure employed for our analysis.

Algorithm 4 Gappy POD algorithm for two-phase flows.

Given a set of M corrupt snapshots $\mathbf{q}_i$

1. Use time-average values as initial guesses at the locations where $\upsilon(\mathbf{x}, t) = 0$ to obtain M snapshots of an initial complete field $\widetilde{\mathbf{q}}(\mathbf{x}, t)$.

2. Perform weighted POD of $\widetilde{\mathbf{q}}(\mathbf{x}, t)$ to obtain M guessed spatial modes $\widetilde{\boldsymbol{\varphi}}_i(\mathbf{x})$ and the corresponding temporal modes $\widetilde{a}_k(t)$.

3. Select the number of modes N_m ($N_m < M$) to be employed in the reconstruction.

4. Construct a new vector field as follows:

$$\widetilde{\mathbf{w}}(\mathbf{x},t) = \sum_{k=1}^{N_m} \widetilde{\boldsymbol{\varphi}}_k(\mathbf{x},t)\widetilde{a}_k(t)$$

and overwrite the previous guess, i.e., set

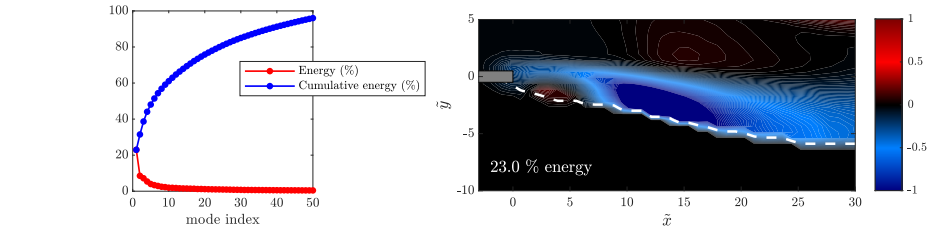
$$\widetilde{\mathbf{q}}(\mathbf{x},t) = \widetilde{\mathbf{w}}(\mathbf{x},t) \quad \text{only where} \quad \upsilon(\mathbf{x},t) = 0.$$

5. Upon convergence of the POD eigenvalues return $\widetilde{\mathbf{q}}$, otherwise go to 2.

The key dimensionless quantities for the analyzed configuration are as follows: $Re_a = \rho_a U_a \delta_a / \mu_a = 768$, where U_a represents the reference velocity of the gaseous phase, δ_a denotes the vorticity thickness (Fuster et al. [109]), and μ_a is the viscosity of the gaseous phase. Additionally, $We_l = \rho_l U_l^2 \delta_l / \sigma = 0.19$, where σ stands for surface tension, and the subscript l pertains to liquid phase properties. Other important quantities include $M_u = \rho_g U_g^2 / \rho_l U_l^2 = 25.56$ and $e/\delta_a = 2.6$.

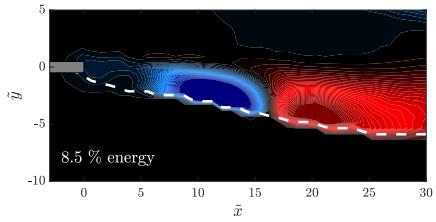
In Fig. 2.46 is reported the Gappy POD analysis for the analyzed case. Panel (a) contains the energy distribution per mode (red) and cumulative (blue). It is possible to notice that with only about 50 modes more that 95% of the fluctuations energy is retrieved. In panels from (b) to (f) are presented the leading POD modes for this configuration; white dashed lines represent the mean interface location. The inspection of this modes reveals that there is a superposition of a flapping motion of the separation region and an interfacial advection behavior. In particular, the first mode (panel (a)), with about 23% of the total energy, is related to the flapping of the wake separation region. By looking at the 2^{nd} and 3^{rd} modes (panels (c) and (d)) it is possible to notice a spatial shift between them that highlights the advective nature of this modes couple. The same behavior is retrieved for 4^{th} and 5^{th} modes, shown in panels (e) and (f).

PSDs of the temporal coefficients a_i of the leading modes are reported in Fig. 2.47. All the modes present a dimensionless frequency peak at 2.8×10^{-4}.

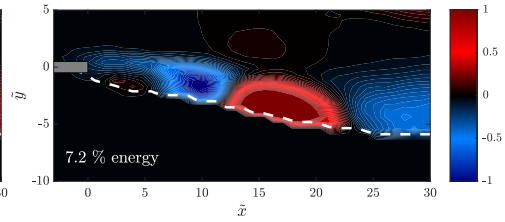


(a) POD energy

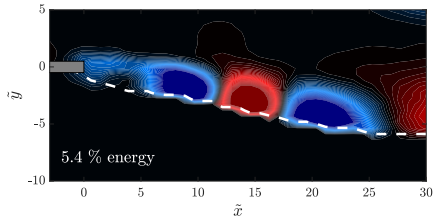
(b) 1^{st} u' POD mode



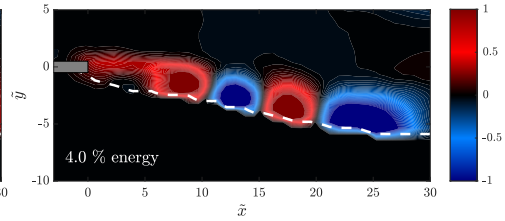
(c) 2^{nd} u' POD mode



(d) 3^{rd} u' POD mode



(e) 4^{th} u' POD mode



(f) 5^{th} u' POD mode

Figure 2.46. Gappy POD analysis. Panel (a), POD energy distribution. Panels (b) to (f), leading u' modes. Field variables have been normalized with respect to their maximum. White dashed lines represent the mean interface location.

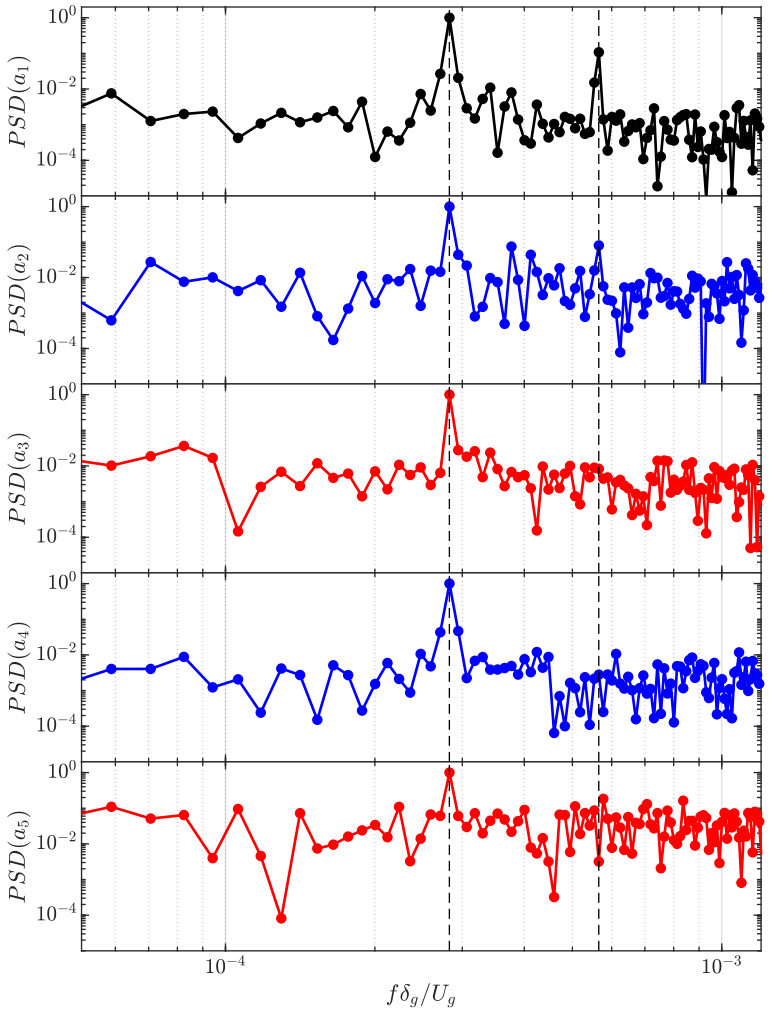


Figure 2.47. Power spectral densities (PSD) of the leading modes temporal coefficients a_i.

Chapter 3

Data driven Stability analysis

One classical way of examining fluid flows is by linearizing the governing equations, and studying the mathematical evolution of disturbances (of small-amplitude) superposed on a base flow. This approach provides the eigenvalues spectrum of the system, revealing the emergence of the flow unsteadiness and thus allowing the prediction of flow field dominant frequencies (Drazin and Reid [110], Schmid and Henningson [111], Criminale et al. [112] and Chen et al. [113]). Many effort, both in modal (Theofilis [12]) and non-modal (Schmid [114]) analysis, have been made in the framework of Linear Stability Analysis (LSA) on the identification of how the small-amplitude perturbations develop in space and time. However, spectral information of the system can be also obtained by means of the DMD technique, that can be conveniently applied instead of the linearization procedure, as shown in Ferrer et al. [115] and Ranjan et al. [49].

In this case, the DMD provides the linear relationship between a generic snapshot and the subsequent one, considering random perturbations of the base flow. Other very recent applications involve data-driven resolvent analysis (Herrmann et al. [116]) and optimally time-dependent modes (Kern et al. [117]). This approach avoids the linearization process of the governing equations, together with the related numerical and theoretical complexity.

In this work, the global spectra of viscous gravitational two-dimensional

(2D) liquid sheets has been obtained by applying the DMD technique on randomly perturbed base flows. The DMD is used to estimate the linearized evolution operator and to extract the main spatial structures, together with information on the system dynamics. The flow fields have been obtained by means of numerical simulations performed with the two-phase code BASILISK (Popinet [95]).

The chapter is structured as follows: Section 3.1 covers the methodology, with a focus on base flow perturbation and the DMD (Dynamic Mode Decomposition) algorithm for estimating the underlying linear operator, along with the associated eigenspectra and eigenfunctions. In Section 3.1.1, the results for the 1D model of the vertical liquid jet are presented. Subsequently, Section 3.1.1 examines the 2D jet configurations.

3.1 Data-driven global stability analysis

The global stability analysis focuses on the temporal and spatial evolution of small-amplitude perturbations superimposed upon a base flow. With this assumption, the governing equations can be recast in the form of a classic first order dynamical system,

$$\dot{\mathbf{q}} = f(\mathbf{q}), \tag{3.1}$$

with f being a non-linear operator depending on $\mathbf{q}$.

From a practical point of view, the analysis is carried out through the decomposition of $\mathbf{q}$ in the sum of a base flow $\bar{\mathbf{q}}$ and the fluctuations $\mathbf{q}'$, such that $\mathbf{q} = \bar{\mathbf{q}} + \mathbf{q}'$; in this way it is possible to obtain the linearized form of the equation (3.1):

$$\dot{\mathbf{q}}' = \mathbf{L}_{\bar{\mathbf{q}}}\mathbf{q}', \tag{3.2}$$

with $\mathbf{L}_{\bar{\mathbf{q}}}$ being the Jacobian operator evaluated at the base flow.

The computation of $\mathbf{L}_{\bar{\mathbf{q}}}$ is numerically prohibitive. Theofilis [118] and Bagheri et al. [119], in the hypothesis of small perturbations, showed that its evaluation can be simplified when based on instantaneous fields

of $\mathbf{q}'$ (snapshots), obtained by a high-order solver (in present case the BASILISK). Moreover, following Gomez et al. [120] a proper set of snapshots is constructed through the Fréchet derivative:

$$\dot{\mathbf{q}}' = \mathbf{L}_{\bar{\mathbf{q}}}\mathbf{q}' = \left.\frac{\partial f(\mathbf{q})}{\partial \mathbf{q}}\right|_{\bar{\mathbf{q}}} \mathbf{q}' \approx f(\bar{\mathbf{q}} + \mathbf{q}') - f(\bar{\mathbf{q}}), \tag{3.3}$$

that when integrated from a generic time instance $t = t_n$ to the subsequent one $t = t_n + \Delta t$, leads to

$$\mathbf{q}'_{n+1} = e^{\mathbf{L}_{\bar{\mathbf{q}}}\Delta t}\mathbf{q}'_n = \mathbf{A}\mathbf{q}'_n \approx \int_{t_n}^{t_n+\Delta t} f(\bar{\mathbf{q}}+\mathbf{q}')dt - \int_{t_n}^{t_n+\Delta t} f(\bar{\mathbf{q}})dt. \tag{3.4}$$

The inspection of the previous equation highlights the linear relationship between two subsequent snapshots $\mathbf{q}'_n$ and $\mathbf{q}'_{n+1}$ through the matrix $\mathbf{A}$; note that $\mathbf{A}\mathbf{q}'_n$ approximates the difference between the variation of the perturbed field and that of the base flow.

In the framework of two-phase flows, it is convenient to define a state vector $\mathbf{q}$ by stacking the velocity components and the volume fraction at every time instance. For the vertical liquid jet, introduced in section 2.1.1 and appendix A.1, the base flow $\bar{\mathbf{q}}$ has been assumed coincident with the flow field after an evolving time $t_0 = 6t_r$, where t_r is the reference time $t_r = L/U$. The perturbed field has been obtained by adding a purely random perturbation to the whole flow field at $t = t_0$. In particular, this perturbation is modeled as white noise, with an amplitude of 1% of U for the velocity components (u and v) and 0.01 for the volume of fraction in the interface region.

The matrix $\mathbf{A}$ has been obtained by means of the DMD technique (Ranjan et al. [49]), which decomposes time-resolved data into modes, each of one with its own frequency and growth rate. Considering a standard DMD algorithm (Tu et al. [44]), it is convenient to define a physically meaningful norm for $\mathbf{q}'$ related to the fluctuations energy of the velocity components and of the volume fraction. In particular, the norm is $\|\mathbf{q}'\|^2_{\mathbf{P}} = \mathbf{q}'^*\mathbf{P}\mathbf{q}'$, where $(\cdot)^*$ denotes the conjugate transpose operator and $\mathbf{P}$ represents a weight matrix whose element have been computed

according to work of Chu [121] providing the following energy norm:

$$E = \int_{\Omega} \left[\bar{\rho} \left(u'^2 + v'^2 \right) + \left(\bar{u}^2 + \bar{v}^2 \right) \frac{\Delta \rho^2}{\bar{\rho}} C'^2 \right] d\Omega, \tag{3.5}$$

in which Ω is the computational domain. In this way, $\mathbf{P}$ takes into account the integration quadratures and appropriate scaling of heterogeneous variables of the problem. Finally, a Cholesky decomposition $\mathbf{P} = \mathbf{F}^*\mathbf{F}$, has been computed to obtain $E = \|\mathbf{q}'\|_{\mathbf{P}}^2 = \mathbf{q}'^*\mathbf{F}^*\mathbf{F}\mathbf{q}' = \|\mathbf{F}\mathbf{q}'\|_2^2$.

The DMD algorithm considers the snapshots ($\mathbf{F}\mathbf{q}'_k$) collected as columns in the matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$

$$\begin{aligned} \mathbf{Q}_1 &= \mathbf{F} \begin{bmatrix} | & & | & & | \\ \mathbf{q}'_1 & \cdots & \mathbf{q}'_k & \cdots & \mathbf{q}'_M \\ | & & | & & | \end{bmatrix}, \\ \mathbf{Q}_2 &= \mathbf{F} \begin{bmatrix} | & & | & & | \\ \mathbf{q}'_2 & \cdots & \mathbf{q}'_{k+1} & \cdots & \mathbf{q}'_{M+1} \\ | & & | & & | \end{bmatrix}, \quad \mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{R}^{N \times M}, \end{aligned} \tag{3.6}$$

where N is the cardinality of the state vector $\mathbf{q}'$, and M the number of snapshots. In this view, the linear operator approximates the temporal dynamics of the data such that

$$\mathbf{Q}_2 \approx \mathbf{\Theta}\mathbf{Q}_1. \tag{3.7}$$

It is worth noting that the matrix $\mathbf{\Theta}$ is related to the matrix $\mathbf{A}$ by $\mathbf{\Theta} = \mathbf{F}\mathbf{A}\mathbf{F}^{-1}$; it shares the same eigenvalues of $\mathbf{A}$ and has eigenvectors scaled by $\mathbf{F}$. Moreover, snapshots collected within the matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ can also refer to several simulations with different random initializations (Herrmann et al. [116]). The following analysis considers 200 simulations each one composed of 200 snapshots, with $\Delta t = 10\,\Delta t_s \approx 0.003\,t_r$.

The leading coherent structures (namely the DMD modes) are evaluated through the SVD decomposition of $\mathbf{Q}_1$, according to which $\mathbf{Q}_1 = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$. In this way, Eq. (3.7) becomes $\mathbf{\Theta} = \mathbf{Q}_2\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^*$, which generally requires a large computational cost. To mitigate this effort, it is

convenient to introduce the matrix $\tilde{\mathbf{\Theta}} = \mathbf{U}^* \mathbf{Q}_2 \mathbf{V} \mathbf{\Sigma}^{-1}$ that shares the same non-zero eigenvalues of $\mathbf{\Theta}$ (and of $\mathbf{A}$), but has a lower dimension and thus requires a reduced numerical cost. Finally, the DMD modes $\mathbf{\Phi}$ are retrieved as $\mathbf{\Phi} = \mathbf{Q}_2 \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{W}$, with $\mathbf{W}$ being the matrix of eigenvectors of $\tilde{\mathbf{\Theta}}$.

The dimensionless spectrum of the underlying linear operator $\mathbf{L}_{\bar{\mathbf{q}}}$ is computed by

$$\lambda_k = \lambda_r + \lambda_i \, i = \frac{\log(\mu_k)}{\Delta t} 2\pi t_r, \tag{3.8}$$

where μ_k is the k^{th} eigenvalue of $\tilde{\mathbf{\Theta}}$, λ_r is the dimensionless growth rate and λ_i the dimensionless frequency. Note that throughout this work relevant eigenvalues are represented with two decimal digits only.

The convergence and the saturation of the DMD algorithm are evaluated by means of the L_2 norm of the residual $\mathbf{r}$. Indeed, considering the qr factorization of the snapshots matrix ($\mathbf{Q}_1 = \mathbf{S}\mathbf{R}$) the residual $\mathbf{r}$ is defined as:

$$\mathbf{r} = \mathbf{q}'_{M+1} - \mathbf{Q}_1 \mathbf{R}^\dagger \mathbf{S}^* \mathbf{q}'_{M+1}, \tag{3.9}$$

in which $\mathbf{q}'_{M+1}$ is the last snapshot and $(\cdot)^\dagger$ is the pseudoinverse operator. Note that $\mathbf{r}$ is the reconstruction error field of the last snapshot $\mathbf{q}'_{M+1}$ and simply represents how well the last snapshot can be represented as a linear combination of the first M ones.

3.1.1 1D global analysis

The methodology illustrated in section 3.1 is first tested on the simplified 1D model of the curtain flow described by Della Pia et al. [61], which provides the streamwise distribution of both the sheet centerline lateral displacement (ℓ_s) and the lateral velocity on the centerline (v_s) for sinuous perturbations of the jet. Results of the linear stability analysis carried out by means of the 1D model will be hereafter compared with those obtained by the application of the DMD technique to the same simplified flow data.

The dimensionless 1D model is based on the following governing

equations:

$$\frac{\partial v_s}{\partial \hat{t}} - RU_s \int_0^1 \frac{\partial v_s}{\partial \hat{t}} \ln|\hat{x}-\xi|\, d\xi + RU_s \int_0^1 U_s \frac{\partial^2 \ell_s}{\partial \hat{t}\partial \hat{x}} \ln|\hat{x}-\xi|\, d\xi = -U_s \frac{\partial v_s}{\partial \hat{x}} + \frac{U_s}{We}\frac{\partial^2 \ell_s}{\partial \hat{x}^2} + \frac{\varepsilon}{2Re}\frac{\partial^2 v_s}{\partial \hat{x}^2}, \tag{3.10}$$

$$\frac{\partial \ell_s}{\partial \hat{t}} = v_s - U_s \frac{\partial \ell_s}{\partial \hat{x}}, \tag{3.11}$$

where $\hat{t} = t/t_r$ and $\hat{x} = x/L$, coupled with only homogeneous boundary conditions at the inlet section, $\ell_s(0) = 0$, $v_s(0) = 0$, being the free-outflow condition at the bottom edge self-guaranteed (Girfoglio et al. [53]). Eq. (3.10) represents the momentum equation averaged along the lateral coordinate y and contains the unsteady Bernoulli-based pressure model by Kornecki et al. [122]; Eq. (3.11), instead, is the linearized kinematic boundary condition imposed at the free interfaces (Girfoglio et al. [53]). The parameter R takes in account the effects of the density ratio, r_ρ, and the sheet slenderness, ε, while U_s is the dimensionless axial velocity (corresponding to the free-fall Torricellian base flow):

$$R = \frac{2}{\pi}\frac{1}{\varepsilon} r_\rho, \quad U_s(\hat{x}) = \sqrt{1 + 2\frac{\hat{x}}{Fr}}. \tag{3.12}$$

The analysis starts by considering supercritical flow conditions ($We > 1$). Note that in this one-dimensional analysis, unlike the subsequent two-dimensional one, the state vector $\mathbf{q}'$ consists of a collection of the values of ℓ_s ans v_s in each streamwise location, and snapshots of the simulation are taken by integrating the system (3.10)-(3.11). Accordingly, the weight matrix is computed by considering the perturbations energy per unit length given by Olsson and Henningson [123] and Della Pia et al. [63]:

$$E_{1D} = \frac{1}{2}\int_0^1 \left[\frac{v_s^2}{U_s} + \frac{1}{We}\left(\frac{\partial \ell_s}{\partial \hat{x}}\right)^2\right] \mathrm{d}\hat{x} \tag{3.13}$$

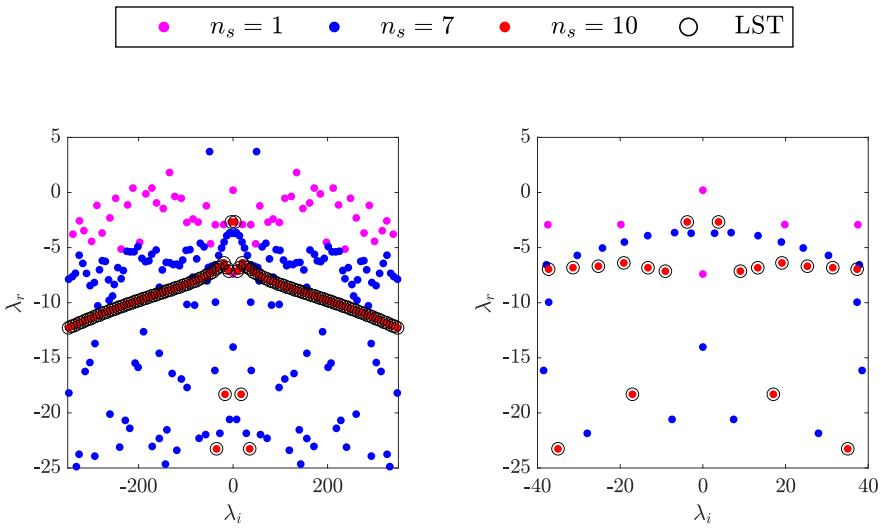
in which the first term, proportional to v_s^2, is related to the kinetic energy; whereas, the second one, proportional to $\left(\frac{\partial \ell_s}{\partial \hat{x}}\right)^2$, involves the surface tension.

Panel (a) of Fig. 3.1 reports the global spectrum computed with the Linear Stability Theory (LST), together with the ones estimated through the DMD technique for several numbers of random initializations (n_s). The physical conditions are represented by the following set of parameters: $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. The LST spectrum, as reported by Della Pia et al. [61], exhibits two branches (see the zoom reported in panel (b)), with an almost constant spacing between the imaginary part of the eigenvalues, which is directly associated with the crossing time of slow (upper branch) and fast (lower branch) traveling waves.

The cases represented consider $n_s = 1$, 7 and 10. It is interesting to observe that with a few random snapshots ($n_s = 7$) the spectrum exhibits spurious unstable eigenvalues, but with a slight increase of n_s it quickly tends to converge to the standard one (open circles) for both branches. For $n_s = 10$ a good match can be appreciated, the leading eigenvalues being properly captured both in real part (growth/decay rate) and imaginary one (frequency). Note that the convergence of the residuals comes with the convergence of the eigenvalues, and the disappearance of the spurious ones. For the analyzed case the leading dimensionless frequency (imaginary part of the leading eigenvalue) is $\lambda_i = 3.64$.

The leading eigenfunctions of displacement (φ_l, panel (a)) and lateral velocity (φ_v, panel (b)) for the cases $n_s = 7$ and 10 together with those based on the LST method are reported in Fig. 3.2. The DMD modes computed with $n_s = 10$ exhibit a very good match with the theoretical ones; the case with $n_s = 7$ is interesting because even if the leading modes exhibit spurious oscillations due to the randomness of the perturbations, the streamwise trend of these modes is correctly predicted.

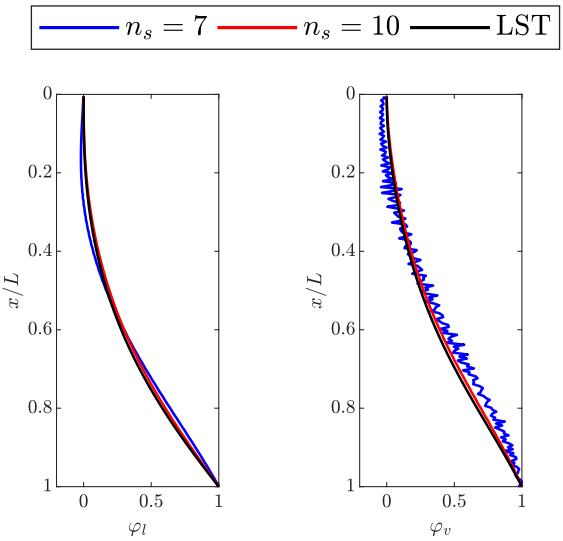
The convergence and the saturation of the algorithm are evaluated by means of the L_2 norm of the residual, $\|\mathbf{r}\|$, which measures how much the last snapshot $\mathbf{q}'_{M+1}$ can be expressed as a linear combination of first M



(a) Global spectrum.

(b) Zoom of the inner spectrum.

Figure 3.1. Comparison between the supercritical spectrum evaluated with the linear stability analysis and various spectra computed with the DMD approach for different values of n_s. Panel (a) contains the global spectrum and panel (b) a zoom of its inner region. $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$.



(a) Lateral displacement ℓ_s.

(b) Transverse velocity v_s.

Figure 3.2. Comparison between leading LST eigenfunctions and DMD modes of lateral displacement (a) and transverse velocity (b).

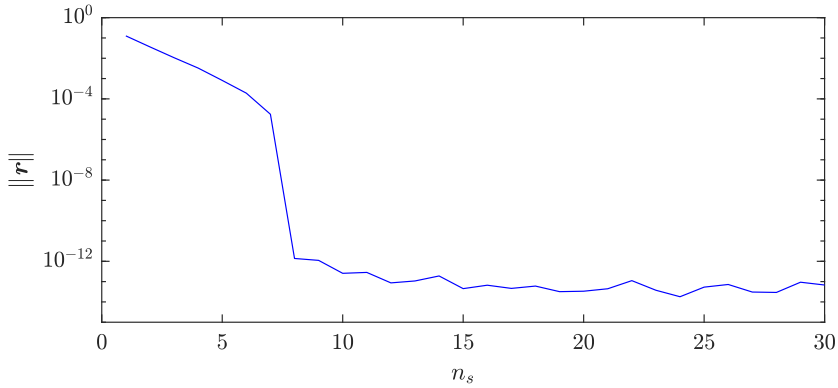


Figure 3.3. Convergence of DMD algorithm. Residual value history versus number of random simulations n_s

ones (Schmid [43]). The trend of the residue is reported in Fig. 3.3 as a function of the number of simulations n_s. In accordance with the spectrum considered above, a rapid convergence can be observed for $n_s > 8$, for which $\|\mathbf{r}\| \approx 10^{-12}$, thus providing a sufficient accuracy to determine the spectrum and the corresponding eigenmodes.

Once the capability of the DMD algorithm in determining the eigenfunctions/eigenvalues of the underlying linear system has been proven, it is interesting to test its ability to properly represent the evolution of the system by comparing results of its integration with the theoretical ones (Eq.s (3.10)-(3.11)).

Indeed, the linear operator $\mathbf{A}$ can be used to predict the behavior of the system both in free and forced configurations; to do that, the system (3.10)-(3.11) is recast in the form

$$\mathbf{M}\dot{\mathbf{q}}' = \mathbf{L}\mathbf{q}' + \mathbf{b}_1\mathbf{f} + \mathbf{b}_2\dot{\mathbf{f}}, \tag{3.14}$$

where $\mathbf{f}$ models the effect of an external forcing at the inlet ($\hat{x} = 0$), through non-homogeneous boundary conditions: $\mathbf{f} = [v_s(0,\hat{t}); \ell_s(0,\hat{t})]$.

For discrete times the solution reads

$$\mathbf{q}'_{n+1} = \mathbf{A}_{LST}\mathbf{q}'_n + \mathbf{B}_1\mathbf{f}_n + \mathbf{B}_2\dot{\mathbf{f}}_n, \tag{3.15}$$

in which $\mathbf{A}_{LST} = e^{(\mathbf{M}^{-1}\mathbf{L})\Delta t}$, $\mathbf{f}_n$ and $\dot{\mathbf{f}}_n$ are the forcing and its derivative at the n^{th} time step, respectively; the matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ are equal to

$$\mathbf{B}_1 = \int_0^{\Delta t} e^{(\mathbf{M}^{-1}\mathbf{L})(\Delta t-\tau)}\mathbf{b}_1 \mathrm{d}\tau, \quad \mathbf{B}_2 = \int_0^{\Delta t} e^{(\mathbf{M}^{-1}\mathbf{L})(\Delta t-\tau)}\mathbf{b}_2 \mathrm{d}\tau. \tag{3.16}$$

Fig. 3.4 shows the solution obtained enforcing homogeneous conditions for ℓ_s and a harmonic forcing ($f_f = 25$ Hz) for the transverse velocity $v_s(0,\hat{t}) = \sin(2\pi t_r f_f \hat{t})$. A comparison between the integration of (3.15) by using $\mathbf{A}_{LST}$ and the DMD estimations of $\mathbf{A}$ is reported. Again, as n_s increases the DMD modes tend to behave strictly as the eigenfunctions of the LST model. Note that for $n_s = 7$, despite the quite noisy modes of the DMD model, an acceptable averaged representation of the centerline deflection (blue lines) is achieved.

3.1.2 2D analysis

The methodology described earlier is now applied to 2D numerical simulations, providing an estimation of the BiGlobal spectrum in a non-intrusive way, namely without the linearization of the governing equations.

As stated before, the perturbation is imposed at the initial time in a random way (white noise) in both the velocity components and volume fraction fields. Its evolution, for the axial velocity (u') field only, can be appreciated considering the snapshots reported in Fig. 3.5. Panel (a) contains the white noise added to the velocity component at the initial time ($t = 0$); panels (b) and (c) show the evolution of these disturbances that, while reducing in amplitude, excite dynamics with different scales.

The study exploits the DMD technique to obtain the global spectrum; in this case 200 simulations, corresponding to 40000 snapshots, have been used for the analysis. The DMD spectrum of the base case has been reported in Fig. 3.6, together with the corresponding one of the 1D model

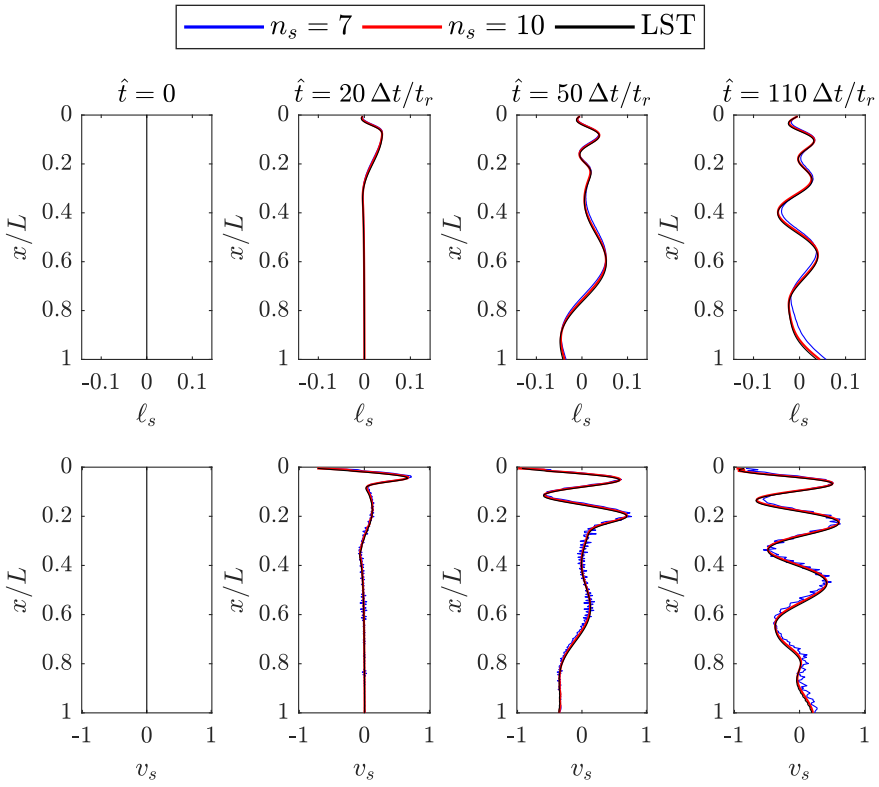
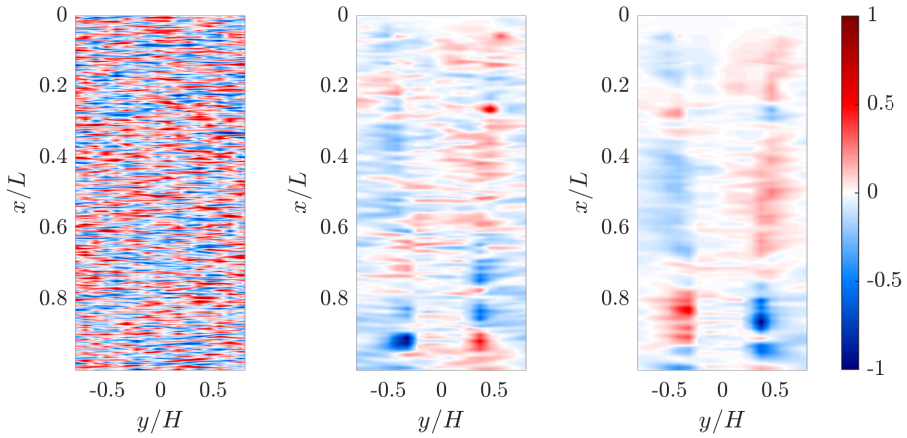


Figure 3.4. Comparison between sheet centerline snapshots of the forced configuration and the same results obtained by means of the DMD approach. $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. Sinusoidal forcing of v_s at inlet, $v_s(0, \hat{t}) = sin(\hat{f}_f t)$ with $\hat{f}_f = 25$.



(a) $t = 0$ **(b)** $t = 5\,\Delta$t **(c)** $t = 10\,\Delta$t

Figure 3.5. u' random initial disturbance (a) and its temporal evolution (b) and (c). $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. Field variables have been normalized with respect to the maximum.

(red circles), for a direct comparison. Here, only 400 DMD modes are reported, which retain more than the 99.9% of the POD (proper orthogonal decomposition) total energy, avoiding the introduction of spurious modes (Schmid [43]).

Generally speaking, the BiGlobal spectrum highlights a wider frequency content than the LST results, with a further branch in between the upper and lower branches previously found for the 1D model. The physical relevance of the middle branch will be discussed hereafter. All the eigenvalues have a negative real part due to the stable nature of the flow; the less damped eigenvalues are located at low frequencies. The leading mode, labeled A, is equal to $\lambda = -0.54 + \mathrm{i}\,3.64$, which perfectly matches the value predicted by the low order analysis; other interesting eigenvalues are highlighted with blue circles in the spectrum and are located on different branches: modes A and B belong to the upper branch, modes C, D and E are on the middle one, and mode F lies on the lower branch. The corresponding values are reported in Table 3.1.

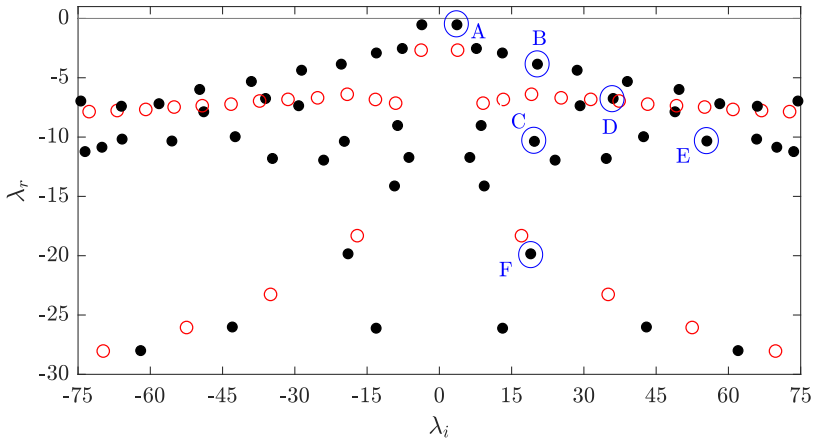


Figure 3.6. Data driven BiGlobal spectrum. $We = 2.5, Fr = 0.33, \varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. Black dots refer to the 2D simulations, red circles represent 1D model results. Letters denote selected modes.

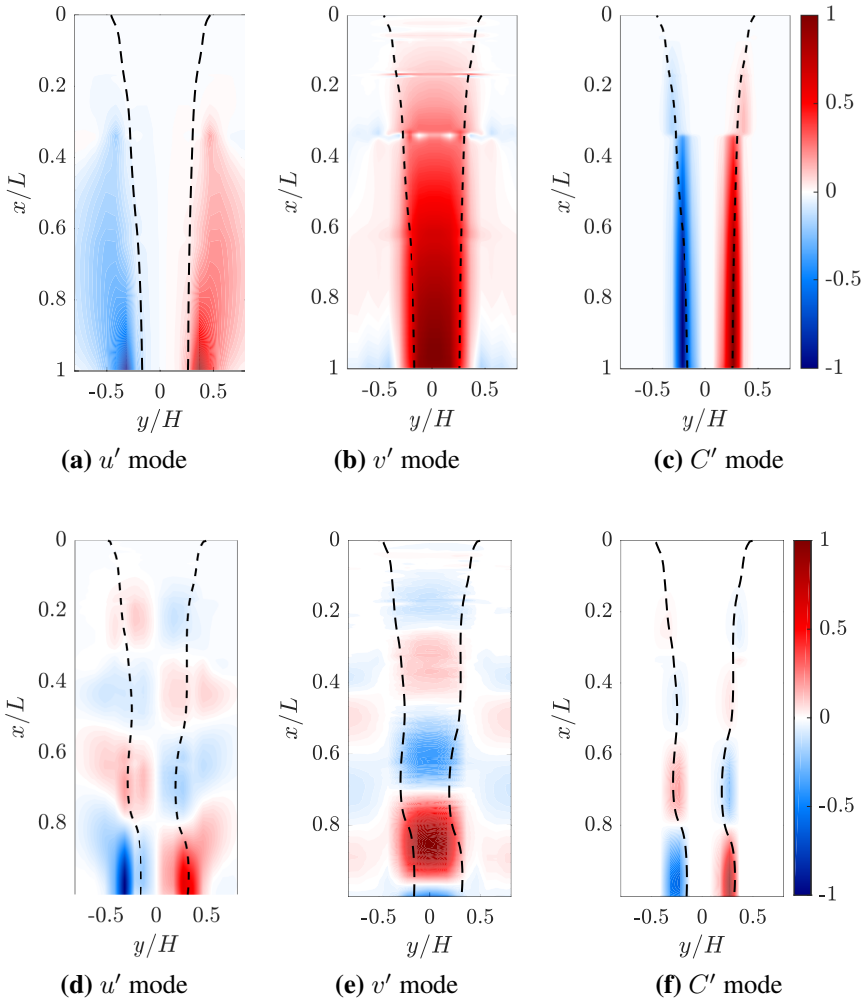
	λ_k	
A	-0.54	$+ 3.64\,i$
B	-3.86	$+ 20.36\,i$
C	-10.36	$+ 19.73\,i$
D	-6.75	$+ 36.08\,i$
E	-10.33	$+ 55.51\,i$
F	-18.96	$+ 19.84\,i$

Table 3.1. Selected eigenvalues of Fig. 3.6 for the case of $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$.

The DMD technique also provides insights on the topology of the flow through the analysis of the coherent structures associated with the main modes. In particular, for the liquid sheet flow under study, the DMD provides the spatial distributions of both velocity components (u', v') and volume fraction (C'). Fig. 3.7 reports the real part of the structures corresponding to modes A and B, scaled with respect their maximum. Note that the black dashed line represents the interface location due the mode's action, that is computed considering only the correspondent mode in the C reconstruction according to Tu et al. [44] and Schmid [43]. Panels (a) and (c) show antisymmetric distributions of u' and C', while panel (b) contains a symmetric distribution of v'. In accordance with Colanera et al. [64], this is typical of a sinuous motion. The v' distribution (panel (b)) along the axis $y/H = 0$ features the same behavior as the 1D leading eigenfunction reported in Fig. 3.2, namely an almost monotonic increase of the transverse velocity component along the downstream direction. A similar behavior can be also observed for the mode B, which is a generic mode representative of the flow topology in this branch. Indeed, panels from (d) to (f) report structures with smaller spatial scales with respect the previous ones, but still characterized by a sinuous motion. The topology described above is retrieved, with different scales, in all modes of this branch, which is therefore characterized by a purely sinuous behavior. The same characteristics are obtained for the mode F (not reported herein) belonging to the lower branch, whose eigenvalues physically are associated with fast traveling waves with a sinuous behavior. Thus, bearing in mind that the 1D model is purely sinuous, it is not surprising that for these two sinuous branches there is a good agreement between both the spectra.

As will be detailed in the following, the additional 2D middle branch is associated with a varicose flow behavior, which could not be predicted by the 1D model, accounting for the sinuous dynamics only.

To proceed further with the discussion of the spectrum, it is convenient to recall that each mode $\boldsymbol{\varphi}_j$ can be decomposed in sinuous and varicose contributions (Colanera et al. [64]): $\boldsymbol{\varphi}_j(x, y) = \boldsymbol{\varphi}_j^s(x, y) + \boldsymbol{\varphi}_j^v(x, y)$, where $\boldsymbol{\varphi}_j^s(x, y)$ and $\boldsymbol{\varphi}_j^v(x, y)$ are the sinuous and varicose components of the mode, respectively. $\boldsymbol{\varphi}_j^s(x, y)$ contains the antisymmetric part (with



(a) u' mode **(b)** v' mode **(c)** C' mode

(d) u' mode **(e)** v' mode **(f)** C' mode

Figure 3.7. Spatial distributions of the DMD modes A (panels (a)-(c)) and B (panels (d)-(f)). $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. The black dashed line represents the interface location. Field variables are normalized with respect to the maximum.

respect to the axis $y/H = 0$) of u' and C' modes and the symmetric one of v', while $\boldsymbol{\varphi}_j^v(x, y)$ the complementary distributions. Note that sinuous and varicose components are orthogonal to each other by definition. This decomposition can be directly extended to the perturbations energy, E, of Eq. (3.5). In this way, the energy reads $E = E^s + E^v$, where E^s and E^v represent the sinuous and varicose contributions, respectively. Table 3.2 reports the energy contents for the selected DMD modes.

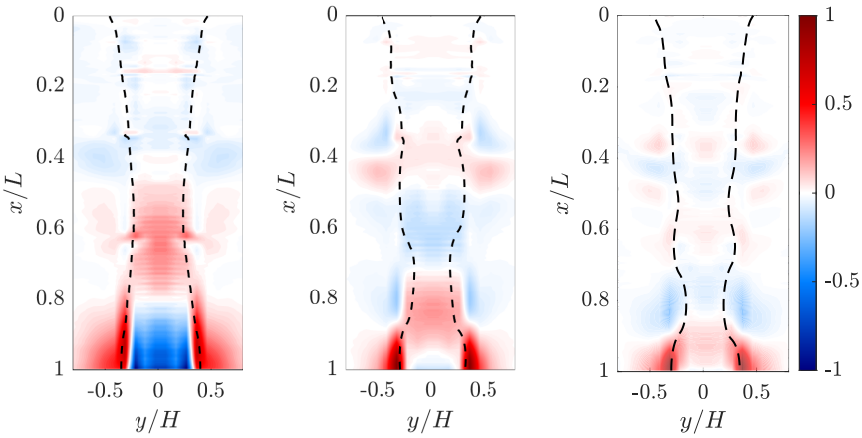
λ_k	$E^s(\%)$	$E^v(\%)$
A	93.5%	6.5%
B	93.5%	6.5%
C	78.0%	22.0%
D	47.9%	52.1%
E	22.9%	77.1%
F	89.0%	11.0%

Table 3.2. Sinuous and varicose energy contributions of the selected DMD modes. $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$.

Looking at the table, one can note that, as expected, the first two modes (A and B) contain a sinuous energy contribution around the 93%; a similar energy distribution can be observed also for the mode F, for which $E^s = 89\%$. Different energy contents are observed in the central branch for increasing frequencies; indeed, modes C, D and E exhibit a growing varicose energy content, respectively equal to $E^v = 22.0\%$, 52.1% and 77.1%.

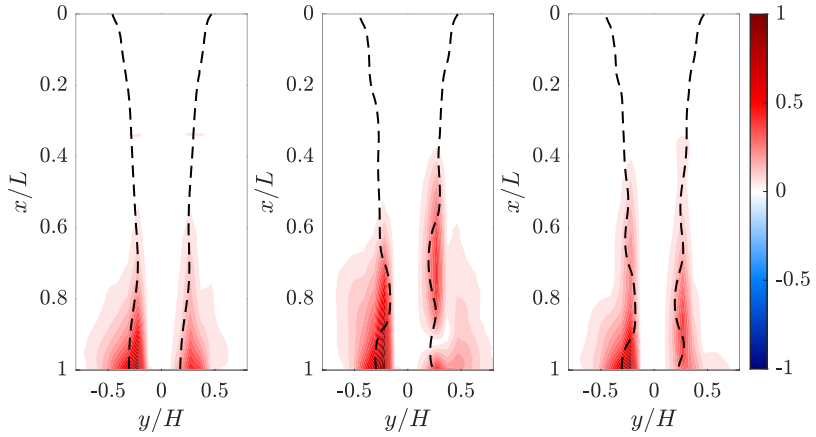
Bearing in mind the previous energy distributions, Fig. 3.8 reports the varicose component of the u' modes for eigenvalues C, D and E. The black dashed line represents the interface location, highlighting the cross-section area variation in varicose regime.

Moving from C to E, a frequency increase is accompanied by a global decrease of the spatial scales. The resulting wavelength reduction determines a more uniform energy distribution along the sheet length. This can



(a) mode C **(b)** mode D **(c)** mode E

Figure 3.8. Varicose component of DMD modes of u' corresponding to eigenvalues C (panel (a)), D (panel (b)) and E (panel (c)). $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. The black dashed line represents the interface location. Field variables are normalized with respect to the maximum.



(a) mode C **(b)** mode D **(c)** mode E

Figure 3.9. Energy distribution for the modes C (panel (a)), D (panel (b)) and E (panel (c)). $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. The black dashed line represents the interface location. Field variables are normalized with respect to the maximum.

be inferred considering the spatial distributions of the modes' amplitude, which represents the local contribution of the mode to the energy defined in (3.5). The corresponding energy distributions are reported in Fig. 3.9; the energy of the mode C is mainly concentrated downstream of the station $x/L = 0.6$, whereas the modes D and E show a more homogeneous distribution starting from $x/L = 0.5$ and 0.4, respectively.

As discussed also for the 1D case, the convergence and the saturation of the algorithm have been evaluated by means of the analysis of the residual $\mathbf{r}$. Its trend is reported in Fig. 3.10 which shows that the residual rapidly tends towards 10^{-13} using about 95 random initializations.

Weber number effect

This section investigates how the Weber number affects the eigenspectra and the leading frequencies both in supercritical and subcritical

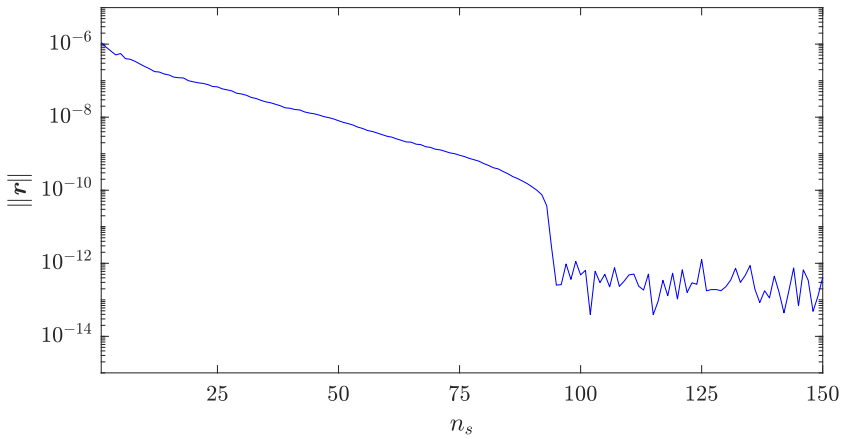


Figure 3.10. Convergence of DMD algorithm. Residual history of the DMD versus number of random simulations n_s for the base case ($We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$).

regimes. Changes in the We number have been obtained by varying the surface tension (σ), keeping constant the inlet axial velocity U in order to not include effects related to Reynolds and Froude numbers variations. The analysis has been carried out considering the following reference values: $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. Note that here only the effects of the Weber number and the density ratio (next sub-section) have been considered, as previous works (Della Pia et al. [61, 65]) have shown that these parameters are the ones with the greatest influence on the behavior of the flow.

The analysis starts considering the flow in supercritical conditions for three different We numbers, equal to $We = 1.5$, 2.5 and 3.5. The corresponding dimensionless spectra are reported in Fig. 3.11, which reveals that, for all these We values, the flow is stable. When We increases, at high frequencies the middle branch moves towards the upper one, and the corresponding modes tend to lose their varicose characteristic assuming a pure sinuous behavior. The leading frequencies are in good agreement with the corresponding ones of the 1D model; the frequency values (λ_i)

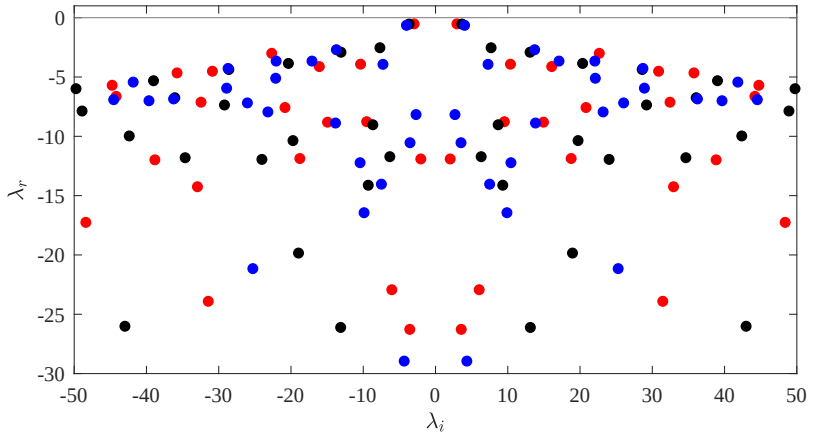


Figure 3.11. Effect of the We number on the global spectrum in supercritical regime. Red filled markers represent the DMD spectrum at $We = 1.5$; black ones at $We = 2.5$ and blue ones at $We = 3.5$. $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$.

and the relative spreads (normalized with respect to the 1D model values) are summarized in Table 3.3.

When the Weber number decreases under the unit threshold, the surface tension has a great influence on the flow field, and the behavior of the curtain radically changes. Indeed, passing from supercritical to subcritical conditions, the damping rate increases together with the leading frequency which presents a jump of about an order of magnitude. The experimental evidence of this phenomenon has been recently presented in the work by Chiatto and Della Pia [124].

Fig. 3.12 contains the dimensionless spectra for $We = 0.75$ (red dots) and $We = 1.5$ (black dots). The leading modes are highlighted with vertical solid lines and the corresponding frequencies are equal to $\lambda_i = 2.95$ and 16.69 for the supercritical and subcritical case, respectively. For the subcritical spectrum the second leading eigenvalue is also highlighted (vertical dashed line at $\lambda_i = 31.64$); the first two leading eigenvalues have a similar growth rate, and the spectrum does not reveal the slow branch.

Table 3.3. Comparison of the leading dimensionless frequency (λ_i) of the DMD and 1D model at various supercritical Weber numbers. The spread refers to the percentage difference between the values normalized with respect to the 1D model one.

	We		
	1.5	2.5	3.5
DMD	2.95	3.64	4.02
1D model	2.99	3.78	4.07
% spread	1.4%	3.7%	1.2%

Table 3.4 reports the first two leading eigenvalues together with their sinuous and varicose energy contents. Unlike the previous cases, here the second leading eigenvalue has already a remarkable varicose energy content, about the 53%.

Table 3.4. Sinuous and varicose energy contents of the first two leading eigenvalues for the case at $We = 0.75$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$.

λ_r	λ_i	$E^s(\%)$	$E^v(\%)$
-5.51	16.69	93.5%	6.5%
-5.56	31.64	46.3%	53.7%

The spatial distributions of the leading DMD mode (with $\lambda_i = 16.69$) highlight a sinuous behavior, similar to those of the supercritical conditions (see panels (a)-(c) of Fig. 3.7) but with a lower spatial wavelength, therefore they have not been reported here. The second leading eigenvalue, as already said, presents a relevant amount of energy related to the varicose motion; the corresponding spatial distributions of u' and C' are reported in panels (a) and (b) of Fig. 3.13. Panel (c) contains both real (black line) and imaginary (red line) parts of the thickness mode φ_h that has been computed as: $\varphi_h = \int_{-1}^{1} \varphi_C \, \mathrm{d}\hat{y}$, where $\hat{y} = y/H$.

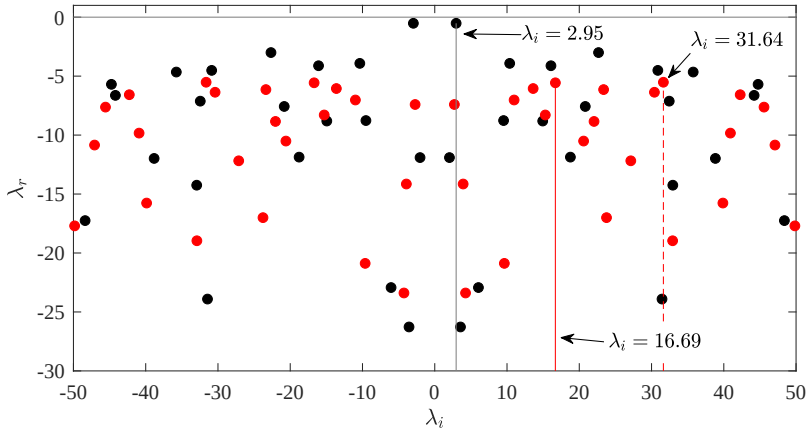
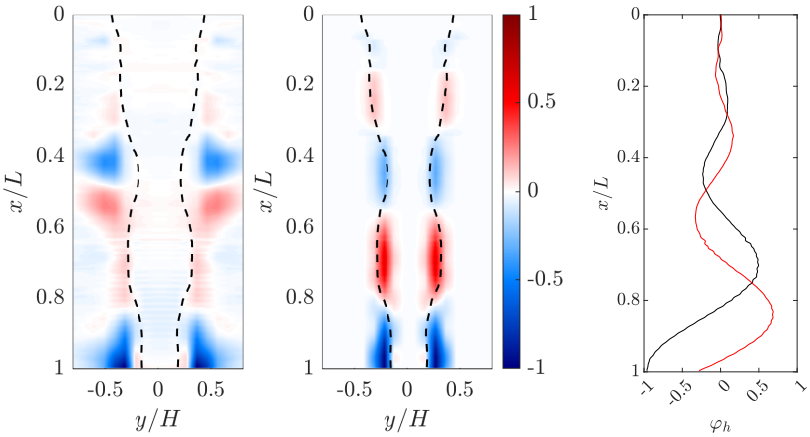
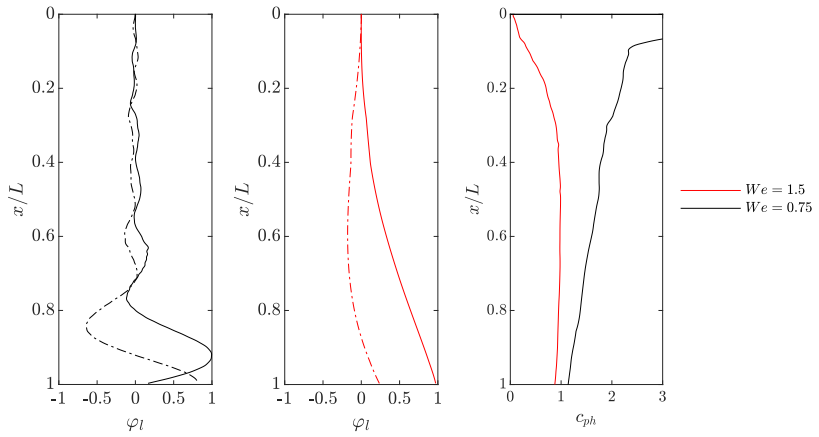


Figure 3.12. DMD spectra comparison between supercritical and subcritical regimes. Black filled markers represent the DMD spectrum at $We = 1.5$ and red ones at $We = 0.75$. Vertical solid lines highlight the leading eigenvalues. The red dashed line refers to the second most unstable eigenvalue of the $We = 0.75$ case. $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$.



(a) u' mode **(b)** C' mode **(c)** thickness mode

Figure 3.13. Varicose components of DMD modes with $\lambda_i = 31.63$ (panels (a) and (b)). $We = 0.75$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.01$, $Re = 413$. The black dashed line represents the interface location. Field variables have been normalized with respect the maximum. Panel (c) reports the real (black line) and imaginary (red line) parts of the thickness mode.



(a) $We = 0.75$ **(b)** $We = 1.5$ **(c)** phase velocity

Figure 3.14. Leading modes lateral displacement (φ_ℓ) for $We = 0.75$, panel (a), and $We = 1.5$, panel (b). The solid lines are the real parts, the dash-dotted lines are the imaginary ones. Modes have been normalized with respect the maximum amplitude. Panel (c) reports the phase velocity c_{ph} of the modes.

To give further insights on the behavior of the sheet in the subcritical regime, it is convenient to recover the centerline displacement mode from the spatial distribution of the volume of fluid C' one. In particular, given a C' mode distribution $\varphi_C(x, y)$, the lateral displacement mode $\varphi_\ell(x)$ is calculated as

$$\varphi_\ell(\hat{x}) = \frac{\int_{-1}^{1} \varphi_C \hat{y} \, \mathrm{d}\hat{y}}{\int_{-1}^{1} \bar{C} \, \mathrm{d}\hat{y}}. \tag{3.17}$$

Panels (a) and (b) of Fig. 3.14 report real (solid line) and imaginary parts (dash-dotted line) of the leading lateral displacement mode (φ_ℓ) for $We = 0.75$ and $We = 1.5$, respectively.

Apart from the different spatial wavelength, the main difference between these modes is the disturbances traveling time. To clarify this, it is useful to compute the dimensionless phase velocity distributions $c_{ph}(\hat{x}) = \lambda_i/\alpha$ for both modes, where $\alpha(\hat{x})$ is the leading spatial wave-

length, depending on the position $\hat{x}$, and it is equal to $\alpha(\hat{x}) = \mathrm{d}\theta/\mathrm{d}\hat{x}$, where the phase distribution is $\theta(\hat{x}) = \arg(\varphi_\ell(\hat{x}))$. The corresponding phase velocities for $We = 0.75$ and 1.5 are reported in the panel (c) in black and red lines, respectively. Considering that the phase velocity of the leading mode at the lower We is always higher than that at the higher We, the crossing time in the subcritical configuration will be smaller. This finding agrees with the results of the 1D model discussed in the previous works of Della Pia et al. [61] and Girfoglio et al. [53].

Density ratio effect

The density ratio r_ρ has a great impact on the stability of the flow. Here, three different r_ρ values, equal to $r_\rho = 0.005$, 0.01 and 0.05 have been considered, whose spectra are reported in Fig. 3.15. The analysis has been carried out considering $We = 2.5$. Increasing the density ratio up to $r_\rho = 0.05$, the flow shows an unstable behavior. The most unstable eigenvalue is equal to $\lambda = 2.69 + 20.85\,i$, with a characteristic frequency in good agreement with that reported in the work by Della Pia et al. [61] (Table VI), that regards the frequency response to an impulse disturbance of the transversal velocity component.

A comparison between frequencies of the leading DMD modes and the most unstable eigenvalues of the 1D model is presented in Table 3.5. A good agreement is retrieved for $r_\rho \leq 0.01$, with a relative spread of about the 5%. For the case $r_\rho = 0.05$ the frequencies of the first two most unstable eigenvalues are reported, because, according to the work by Della Pia et al. [61], a better agreement between VOF simulations and 1D model is found for the imaginary part of the second most unstable eigenvalue ($\lambda_i = 22.60$) rather than the leading one ($\lambda_i = 34.47$). In fact the ripples of shorter wavelength are likely damped in the more realistic 2D VOF simulations.

The DMD structures for the stable cases at $r_\rho = 0.005$ and 0.01 feature a purely sinuous shape and, being quite similar to those shown in Fig. 3.7, they are not reported herein. On the contrary, the unstable condition found for $r_\rho = 0.05$ presents leading modes with a different and very interesting

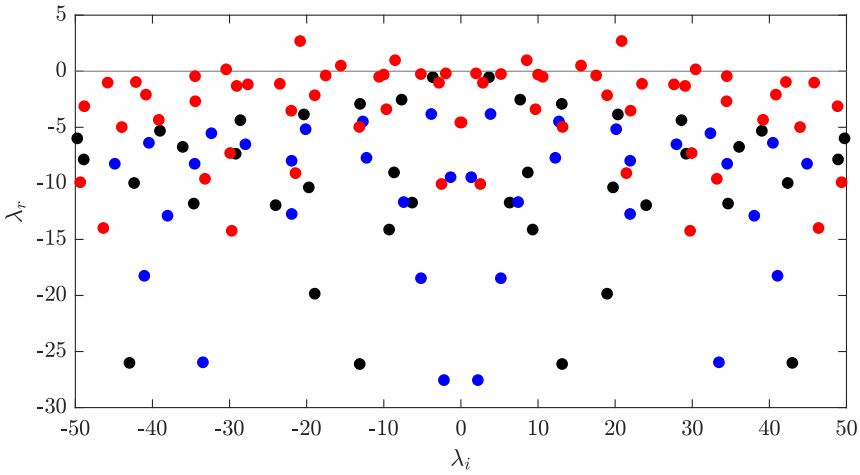
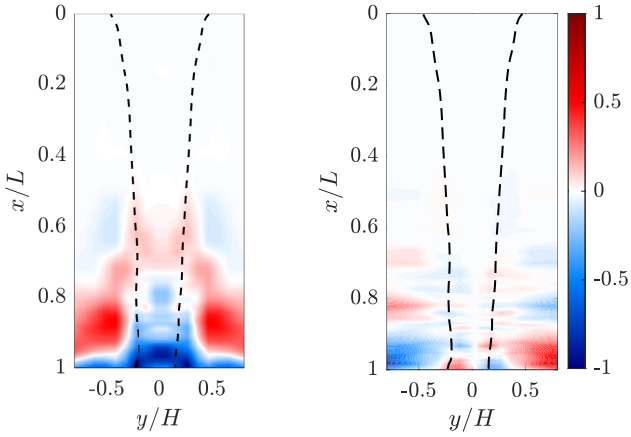


Figure 3.15. Density ratio r_ρ effect on DMD spectrum. Red filled markers represent the DMD spectrum at $r_\rho = 0.05$; black ones at $r_\rho = 0.01$ and blue ones at $r_\rho = 0.005$. $We = 2.5$, $Fr = 0.33$, $\varepsilon = 0.02$, $Re = 413$.

spatial distribution. In this this case last case the investigation is carried out decomposing the flow field into sinuous and varicose contributions as shown in Eq. (2.6).

The energy content of the sinuous and varicose components of the leading mode are equal to $E^s = 57.0\%$ and $E^v = 43.0\%$, respectively. This unstable case is therefore characterized by a relevant varicose content, unlike what has been observed in the previous stable cases, exhibiting a predominant sinuous behavior. Moreover, the amplitude of the varicose components increases downstream, as can be observed by the corresponding spatial distributions reported in Fig. 3.16.

Further insights on the topology of the leading mode can be obtained analyzing the local contribution E_x of the sinuous and varicose components to the total energy E of the mode. This contribution is obtained for each mode by integrating the Eq. (3.5) along the spanwise direction y only, and therefore E_x depends on the x/L station. Furthermore, as the orthogonality between the sinuous and varicose components is also found



(a) u' varicose component

(b) v' varicose component

Figure 3.16. Varicose components of the DMD leading mode of u' (panel (a)) and v' (panel (b)) for $\lambda_i = 20.85$. $We = 2.5$, $Re = 413$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.05$. Field variables are normalized with respect to the maximum.

Table 3.5. Effect of the r_ρ on the leading frequency. For the case $r_\rho = 0.05$ the frequencies of the first two most unstable eigenvalues are reported.

	r_ρ		
	0.005	0.01	0.05
DMD	3.85	3.64	20.85
1D model	4.07	3.78	34.47 - 22.60
% spread	5.4%	3.7%	39.5% - 7.7%

locally, one can write $E_x = E_x^s + E_x^v$, where E_x^s and E_x^v are the sinuous and varicose contributions to the total energy, respectively. Fig. 3.17 contains the spatial distribution along the longitudinal direction x of the local norm E_x together with E_x^s and E_x^v. Curves have been normalized with respect to the E_x maximum. The inspection of this figure reveals that the rupture mechanism of an unstable liquid jet is deeply related to the amplification of the varicose mode. Indeed, the leading mode is mainly sinuous until $x/L \approx 0.9$, while the varicose behavior gains prominence moving downstream. The shift from sinuous to varicose shape is strictly related to the curtain instability, leading to the rupture of the sheet.

Fig. 3.18 reports the DMD reconstructions of the volume fraction field C at different time instances for this unstable case. By looking at the panel (b), which refers to a time instance before the curtain rupture, it is possible to note that the liquid phase (where $C = 1$, dark red) reports a greater thickness reduction around $x/L = 0.9$, that corresponds to the aforementioned varicose behavior. The sheet rupture can be appreciated in the panel (c).

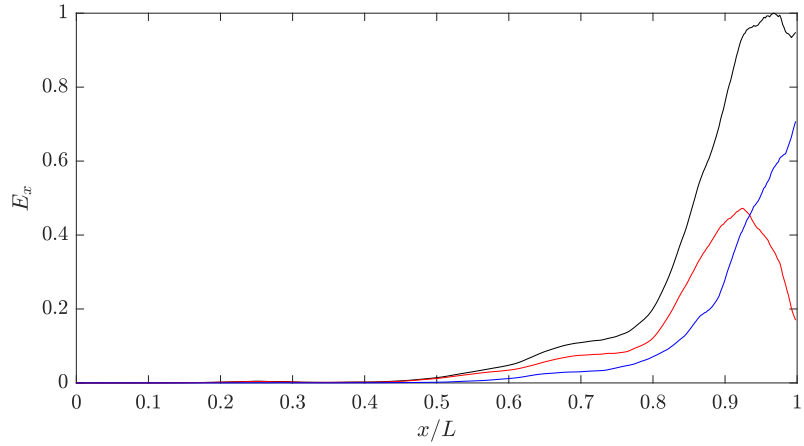
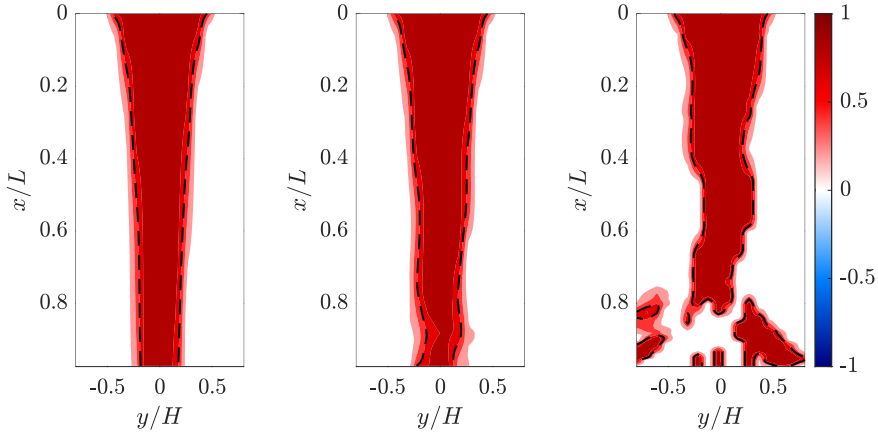


Figure 3.17. Spatial distribution of the local norm E_x (black line) together with the corresponding sinuous (red) and varicose (blue) contributions normalized with respect the maximum of E_x. $We = 2.5$, $Re = 413$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.05$.



(a) $t = 0$ **(b)** $t = 1.5\,t_r$ **(c)** $t = 2.5\,t_r$

Figure 3.18. DMD reconstruction for the C field, considering the leading unstable mode only. Panels from (a) to (c) report its temporal evolution. $We = 2.5$, $Re = 413$, $Fr = 0.33$, $\varepsilon = 0.02$, $r_\rho = 0.05$.

Chapter 4

Reduced Order Modelling

As highlighted in recent works (Rowley and Dawson [72]), the formulation of Reduced Order Models (ROM), inheriting the main features of the Full Order Model (FOM) but with a far lesser computational effort, is a crucial step to reduce the complexity of the governing equations, to gain further insights on the features of the flow field and even to design control strategies. Modern data-driven modal techniques as POD (Proper Orthogonal Decomposition) and SPOD (Spectral Proper Orthogonal Decomposition) can be effectively used in the field of multiphase flows (Colanera et al. [64]). A way to obtain a ROM is to extract physically important features, or modes, characterizing the flow topology, and to project (Galerkin projection) the governing equations onto a subset of these modes, obtaining a system of ordinary differential equations. POD, that gives a set of orthonormal modes, has been widely employed for this purpose in the fluid dynamic community (Stabile et al. [125], Carlberg et al. [126], Stabile and Rozza [127], Sanderse [128]), while SPOD, that gives modes that are spatio-temporal functions, has been recently showed to be effective for the formulation of Galerkin reduced models (Towne [129]).

A fundamentally different data-centric avenue, cluster-based modeling (Fernex et al. [77]), rapidly has gained traction. While the Galerkin method is based on a linear superposition of modes, the cluster-based approach conceptualizes the dynamics as ‘flights’ in a state-space which

is discretized by selected centroids.

In the present work new approaches have been proposed to further improve the capability of cluster-based methods to capture the dynamics of fluid systems dealing with both numerical and empirical data.

This chapter is organized as follows. In section 4.1, standard clustering procedures are recalled and an innovative approach to deal with heterogeneous variables, spatial subdomains and temporal undersamplig is presented. Section 4.2 showcases the application of the latter approach to a swirl flame configuration. In section 4.3 POD-Galerkin ROM and cluster based models are compared for an airfoil configuration. In section 4.4 trajectory based CNM is described.

4.1 Cluster-based approaches

4.1.1 Standard Clustering

Let us consider a field $\mathbf{u}(\mathbf{x}, t)$ within a steady domain Ω. This field might arise from experimental results or computational simulations and could include variables such as velocity, pressure, and so forth.

An ensemble of M time-resolved snapshots, which are statistically representative of the system, serve as the foundation for the analysis. The field is sampled at equidistant time steps, Δt, such that $t^m = m\Delta t$ represents the time instance for the mth snapshot. The corresponding snapshot field at this time is denoted as $\mathbf{u}^m(\mathbf{x}) := \mathbf{u}(\mathbf{x}, t^m)$, where $m = 1, \ldots, M$.

Cluster analysis, an unsupervised method of data organization, aggregates similar entities into groups known as clusters, all without the need for prior data labeling or classification. In the context of our set of M snapshots $\mathbf{u}^m(\mathbf{x})$, the method coarsens the data into K clusters, each represented by a centroid $\mathbf{c}_k(\mathbf{x})$, where $k = 1, \ldots, K$. These centroids are determined via the unsupervised k-means++ algorithm ([130]; MacQueen [131]; Lloyd [132]), representing the characteristic flow patterns or modes within each cluster.

The cluster-affiliation function plays a crucial role here, associating a specific velocity field $\mathbf{u}$ with the index of its nearest centroid:

$$k(\mathbf{u}) = \arg\min_i \|\mathbf{u} - \mathbf{c}_i\|_\Omega, \tag{4.1}$$

where $\|\cdot\|_\Omega$ denotes an Hilbert space norm within the domain Ω. Details on the spatial norm choice will be given in section 4.1.4. The function $k(\mathbf{u})$ enables the definition of cluster regions as Voronoi cells around the centroids:

$$\mathscr{C}_i = \{\mathbf{u} \in \mathscr{L}^2(\Omega) : k(\mathbf{u}) = i\}. \tag{4.2}$$

It is useful to establish a characteristic function χ_i^m to indicate if the mth snapshot belongs to the ith centroid:

$$\chi_i^m = \begin{cases} 1, & \text{if } i = k(\mathbf{u}^m). \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

The effectiveness of a given set of centroids $\mathbf{c}_k$, relative to a given set of snapshots $\mathbf{u}^m$, can be evaluated based on the mean variance of the snapshots in relation to their nearest centroid. This inner-cluster variance gives the cost function:

$$J(\mathbf{c}_1, \ldots, \mathbf{c}_K) = \frac{1}{M} \sum_{m=1}^{M} \|\mathbf{u}^m - \mathbf{c}_{k(m)}\|_\Omega^2, \tag{4.4}$$

where $k(m) := k(\mathbf{u}^m)$. The optimal centroids $\mathbf{c}_k^\star$ are those that minimize this inner-cluster variance:

$$(\mathbf{c}_1^\star, \ldots, \mathbf{c}_K^\star) = \arg\min J(\mathbf{c}_1, \ldots, \mathbf{c}_K). \tag{4.5}$$

It is worth to notice that this algorithm is indeterminate with respect to both snapshots and clusters reordering. In the following, the first cluster is determined as the one with the highest population, that is, it has the highest number of associated snapshots. Subsequent clusters, starting from the second one $(k+1)$, are chosen based on the maximum transition

probability from the preceding cluster (k).

To solve the optimization problem indicated by equation (4.5), the k-means++ algorithm is employed (Lloyd [132]). This algorithm begins by randomly initializing the K centroids. It then iteratively adjusts these centroids until J falls below a predefined threshold, indicating that the clustering is sufficiently accurate. The k-means++ algorithm repeats the clustering process several times and selects the best set of centroids.

The cluster population n_k, that is the total number of snapshots within the kth cluster, is calculated as follows:

$$n_k = \sum_{m=1}^{M} \chi_k^m. \tag{4.6}$$

Moreover, the centroids $\mathbf{c}_k$ that are representative of each cluster, are computed as the mean of all snapshots within the respective cluster:

$$\mathbf{c}_k = \frac{1}{n_k} \sum_{\mathbf{u}^m \in \mathscr{C}_k} \mathbf{u}^m = \frac{1}{n_k} \sum_{m=1}^{M} \chi_k^m \mathbf{u}^m. \tag{4.7}$$

In this equation, $\mathscr{C}_k$ represents the set of snapshots that belong to the kth cluster.

4.1.2 Extended Clustering

In the Extended Clustering approach, similarly to Extended Spectral Proper Orthogonal Decomposition (Extended SPOD), see Sieber et al. [133], only a subset of variables from the vector $\mathbf{u}$ or a spatial subspace $S \subseteq \Omega$ is considered. In the Extended SPOD methodology, this subset forms the basis for calculating spatial/temporal correlations, and the Extended SPOD modes are subsequently derived from projecting the original snapshots onto these temporal modes.

The present Extended Clustering approach leverages the distance function $\|\cdot\|_\Omega^2$ choice in the evaluation of the inner-cluster variance in (4.4). In this approach the snapshots $\mathbf{u}^m$ now consist only in a subset of the

variables (or the spatial subspace), and is denoted as $\mathbf{u}_s^m(\mathbf{x})$, where $\mathbf{x}$ belongs to the subspace S. This necessitates modifying the definition of the cluster-affiliation function, the characteristic function and the inner-cluster variance as follows.

The cluster-affiliation function becomes:

$$k_s(\mathbf{u}) = \arg\min_i \|\mathbf{u} - \mathbf{c}_{si}\|_S, \tag{4.8}$$

where $\|\cdot\|_S$ denotes the norm in the spatial subspace S. Note that $k_s(\mathbf{u}) \equiv k_s(\mathbf{u}_s)$. The characteristic function can be rewritten as:

$$\xi_i^m = \begin{cases} 1, & \text{if } i = k_s(\mathbf{u}^m), \\ 0, & \text{otherwise} \end{cases} \tag{4.9}$$

The modified inner-cluster variance in the subspace S is:

$$J(\mathbf{c}_{s1}, \ldots, \mathbf{c}_{sK}) = \frac{1}{M} \sum_{m=1}^{M} \|\mathbf{u}_s^m - \mathbf{c}_{sk_s(m)}\|_S^2. \tag{4.10}$$

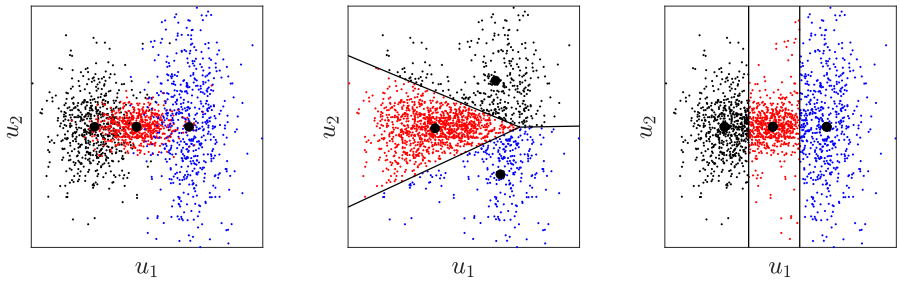
The optimization problem then becomes to find the optimal centroids set $\{\mathbf{c}_{sk}^\star\}_{k=1}^K$ that minimize this inner-cluster variance (4.10), and again is solved using the k-means++ algorithm. Thus, the centroid computation in the subspace is:

$$\mathbf{c}_{sk} = \frac{1}{n_{sk}} \sum_{\mathbf{u}_s^m \in \mathscr{C}_{sk}} \mathbf{u}_s^m = \frac{1}{n_{sk}} \sum_{m=1}^{M} \xi_k^m \mathbf{u}_s^m, \tag{4.11}$$

where, similarly to standard clustering, $n_{sk} = \sum_{m=1}^M \xi_k^m$ and $\mathscr{C}_{sk}$ are the clusters. The extended centroids can be obtained with:

$$\mathbf{c}_k^e = \frac{1}{n_{sk}} \sum_{m=1}^{M} \xi_k^m \mathbf{u}^m. \tag{4.12}$$

Figure 4.1 provides a visual comparison between Standard and Ex-



(a) Sample Data **(b)** Standard Clustering **(c)** Extended Clustering

Figure 4.1. Comparison of Standard and Extended Clustering Techniques. Panel (a) displays a 2D sample dataset generated from three distinct Gaussian distributions. In Panel (b), Standard Clustering is applied with $K = 3$, resulting in Voronoi clusters $\mathscr{C}_i$, and the corresponding centroids $\mathbf{c}_k$. Panel (c) illustrates the results of Extended Clustering, wherein only the only dimension u_1 from the data is considered. This results in spatial-subspace Voronoi clusters $\mathscr{C}_{si}$, and the corresponding extended centroids $\mathbf{c}_k^e$.

tended Clustering techniques, applied to a simple 2D dataset generated from three Gaussian distributions, shown in Panel (a). Panel (b) presents the results of the Standard Clustering technique with a chosen number of clusters, $K = 3$. The resulting Voronoi clusters, $\mathscr{C}_i$, and their corresponding centroids, $\mathbf{c}_k$, illustrate the conventional approach of identifying clusters based on minimizing inner-cluster variance over the entire dataset. In contrast, Panel (c) showcases the outcomes of the Extended Clustering method, which focuses on a specific subspace, in this case, the single dimension u_1 from the data. The derived spatial-subspace Voronoi clusters, $\mathscr{C}_{si}$, and the corresponding subspace centroids, $\mathbf{c}_k^e$, underline the flexibility of Extended Clustering to focus on specific, potentially more informative, subspace or subsets of variables.

This Extended Clustering approach allows for more refined control over the clustering process by focusing on a subset of variables, providing a flexible and robust tool for handling complex, noisy and high-dimensional data.

Temporal undersampling

The extended clustering approach allows the processing of quantities acquired at different sampling rates. Namely, it allows the aggregation of undersampled data based on another variable that was sampled at a higher rate. Therefore each snapshot $\mathbf{u}^m$ (the same holds for $\mathbf{u}_s^m$) can be regarded as a combination of its associated cluster centroid and a residual value $\boldsymbol{\gamma}^m$, resulting in the relationship $\mathbf{u}^m = \mathbf{c}^e_{k(m)} + \boldsymbol{\gamma}^m$. This can be expressed component-wise for every state vector component, with the spatial index i which ranges from 1 to N and temporal index from 1 to M:

$$u_i^m = c^e_{i,k(m)} + \gamma_i^m = \xi_k^m c^e_{i,k} + \gamma_i^m. \tag{4.13}$$

Similarly to Manohar et al. [134], it is possible to define $\mathbf{U}$ as a matrix whose lines are the $\mathbf{u}^m$, $\mathbf{C}$ as a matrix containing the K values of $\mathbf{c}_k^e$ along the lines, $\boldsymbol{\Gamma}$ holding the $\boldsymbol{\gamma}^m$, and Ξ with elements of ξ_k^m. The optimization problem:

$$\min_{\mathbf{C}} \|\boldsymbol{\Gamma}\|_F \quad \text{subject to} \quad \mathbf{U} - \Xi\mathbf{C} = \boldsymbol{\Gamma}, \tag{4.14}$$

in which $\|\cdot\|_F$ is the matrix Frobenius norm, yields the simple relationships described in (4.12) for each centroid.

In situations in which there are temporally undersampled field variables, denoted as $\tilde{\mathbf{u}}^p$ (where p spans from 1 to $M_u < M$), it is possible to use the characteristic function ξ_k^m (and thus Ξ) derived from clustering of the available high-resolution measurements. The M_u dimensional ensemble of the $\tilde{\mathbf{u}}^p$, stored in $\tilde{\mathbf{U}}$, can be linked to the supposed high-resolution snapshot ensemble $\mathbf{u}^m$ using a rectangular projection matrix $\mathbf{P}$. This matrix could either be a basic sampling matrix or could hold a windowed mean of the $\mathbf{u}^m$:

$$\tilde{\mathbf{U}} = \mathbf{P}\mathbf{U}. \tag{4.15}$$

It's important to note that the matrix $\mathbf{P}$ has dimensions $M_u \times M$. The optimization problems can then be written as:

$$\min_{\mathbf{C}} \left\|\tilde{\boldsymbol{\Gamma}}\right\|_F \quad \text{subject to} \quad \tilde{\mathbf{U}} - \mathbf{P}\Xi\mathbf{C} = \tilde{\boldsymbol{\Gamma}}, \tag{4.16}$$

where $\tilde{\mathbf{\Gamma}} = \mathbf{P}\mathbf{\Gamma}$. The solution of the problem (4.16) from a practical point of view is achieved by applying the Moon-Penrose pseudoinverse of the matrix $\mathbf{P}\Xi$. This provides an estimation of the cluster centroids for temporally undersampled data, using the available high-resolution data as reference.

4.1.3 Cluster-based Network Model

After the original snapshots are coarse-grained into K clusters, either via the Standard or Extended method, each cluster can be interpreted as a distinct state of the system dynamics. This representation not only reduces the data dimensionality but also provides an abstracted view of the complex data structure, facilitating the understanding and analysis of the underlying dynamics.

Following the works by Fernex et al. [77] and Li et al. [135], these clusters are then used to conceptualize the system dynamics as a directed network. In this network, the centroids represent the nodes, each signifying a unique state of the system. The directed edges, on the other hand, denote potential finite-time transitions between these states. This network-based representation aids in uncovering the temporal flow or transition tendencies among the clusters, revealing the structure and progression of the system's states.

Figure 4.2 shows and example a directed network of clusters. Each node in the network represents a cluster, more specifically the centroid of that cluster, which is characterized by typical flow patterns. The directed arrows between nodes indicate transitions between different states of the system, with the direction of the arrow indicating the shift from one state to another. The transitions in the cluster-based network model are characterized by two parameters: the probability Q_{ij} and the transition time T_{ij} from cluster j to cluster i. The residence time in a cluster is defined as:

$$\tau_n = t^{n+1} - t^n, \tag{4.17}$$

where t^n is the time associated with the first snapshot of a consecutive snapshots sequence assigned to cluster n.

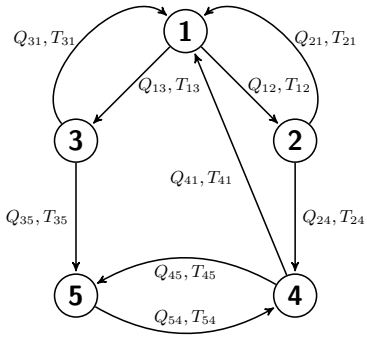


Figure 4.2. Schematic representation of a directed network of clusters. The nodes (1 to 5) represent the centroids of the identified clusters, and the arrows indicate the transitions between them. Each transition is associated with a transition probability Q_{ij} and mean transition time T_{ij}.

Consider j and i as the indices of the clusters at times t^n and t^{n+1}, respectively. The definition of the transition time from cluster j to cluster i is given as half the total residence time in both clusters:

$$\tau_{ij} = \frac{\tau_n + \tau_{n+1}}{2} = \frac{t^{n+2} - t^n}{2}. \tag{4.18}$$

From the dataset, the direct transition probability, denoted as Q_{ij}, and transition time, denoted as T_{ij}, can be deduced. The equations for these quantities are as follows:

$$Q_{ij} = \frac{n_{ij}}{n_j}, \quad i, j = 1, \ldots, K, \tag{4.19}$$

where n_{ij} represents the count of transitions from centroid $\mathbf{c}_j$ to centroid $\mathbf{c}_i$. Moreover, n_j refers to the total count of transitions originating from centroid $\mathbf{c}_j$, regardless of the destination:

$$n_j = \sum_{i=1}^{K} n_{ij}, \quad i, j = 1, \ldots, K. \tag{4.20}$$

The probabilities Q_{ij} are stored into the direct transition matrix (DTM) $Q = \left[Q_{ij}\right] \in \mathscr{R}^{K\times K}$. It is worth noticing that, in contrast to Cluster Markov Models (CMM, Kaiser [83]), in the CNM framework, it is not possible to stay in the same cluster, read $n_{ii} = Q_{ii} = 0$ for $i = 1, \ldots, K$, as only non-trivial transitions from cluster j to cluster i are considered.

The direct transition time T_{ij} from cluster j to cluster i is defined as the mean of all observed transition times, which is represented as:

$$T_{ij} = \langle \tau_{ij} \rangle. \tag{4.21}$$

These averaged transition times are stored into the matrix $T = \left[T_{ij}\right] \in \mathscr{R}^{K\times K}$.

The asymptotic probability p_i^∞ to be in a cluster i can be estimated with

$$p_i^\infty = \frac{\sum \tau_i}{T_f}, \tag{4.22}$$

where T_f being a sufficiently long time horizon simulated by the model, and $\sum \tau_i$ being the cumulative residence time from (4.17).

The Cluster-based Network Model (CNM) is based on centroid visits at discrete times. The clusters, denoted as $k_0, k_1, k_2, \ldots$, are visited in sequence at the following times:

$$t_0 = 0, \quad t_1 = T_{k_1 k_0}, \quad t_2 = t_1 + T_{k_2 k_1} \ldots \tag{4.23}$$

This visitation sequence is in accordance with the direct transition matrix (Q_{ij}) and the transition times T_{ij}. In the following, a continuous motion is assumed between these visits by means of a linear interpolation. However, it should be noted that the application of splines could potentially result in a more smooth motion. Hence, the reconstructed field $\mathbf{u}_r(\mathbf{x}, t)$ for $t \in [t_n, t_{n+1}]$ is formulated as:

$$\mathbf{u}_r(\mathbf{x}, t) = w_n(t)\mathbf{c}_{k_n}(\mathbf{x}) + \left[1 - w_n(t)\right]\mathbf{c}_{k_{n+1}}(\mathbf{x}), \quad w_n = \frac{t_{n+1} - t}{t_{n+1} - t_n}. \tag{4.24}$$

An important observation to make here is that standard clusters were

utilized in (4.24); however, when extended clusters $\mathbf{c}_k^e$ are used to estimate the extended reconstruction $\mathbf{u}_r^e(\mathbf{x}, t)$, the equation remains identical in form:

$$\mathbf{u}_r^e(\mathbf{x}, t) = w_n(t)\mathbf{c}_{k_n}^e(\mathbf{x}) + [1 - w_n(t)]\,\mathbf{c}_{k_{n+1}}^e(\mathbf{x}). \tag{4.25}$$

In Section 4.2, the effectiveness of the Extended Cluster-based Network Model (eCNM) will be examined through its application to several datasets.

4.1.4 Clustering distance metric and parameters choice

Choice of the distance metric

The choice of an appropriate distance metric influences the clustering algorithm. In sections 4.1.1 and 4.1.2, the choice of the distance metric $\|\cdot\|_\Omega$ (and its subset-limited counterpart $\|\cdot\|_S$) significantly impacts various aspects of the analysis, including the cluster affiliation function $k(\mathbf{u})$ (4.1), the cost function J (4.4), and their subset-limited equivalents (4.8)-(4.10). These choices lead to distinct computations for cluster centroids and ultimately influence the structure of the network models that are constructed. The most commonly employed approach involves using the standard Euclidean distance metric. This choice not only finds utility in application but also helps in terms of computational efficiency, as demonstrated in Li et al. [135]. To enhance the separation in the phase space of spatially and temporally coherent structures, in this work a filtered version of the Euclidean distance has been employed. This filtered distance metric is derived as follows.

The squared standard Euclidean distance between two snapshots $\mathbf{u}^i$ and $\mathbf{u}^j$ is fundamentally linked to the elements of the correlation matrix R_{ij} (Dattorro [136]). Mathematically, this relationship is given by

$$\|\mathbf{u}^i - \mathbf{u}^j\|_2^2 = R_{ii} + R_{jj} - 2R_{ij}, \tag{4.26}$$

where the elements R_{ij} are defined as

$$R_{ij} = \frac{\langle \mathbf{u}^i, \mathbf{u}^j \rangle}{M}, \tag{4.27}$$

and $\langle \cdot, \cdot \rangle$ represents the L_2 scalar inner product, defined as $\langle \mathbf{u}^i, \mathbf{u}^j \rangle = \mathbf{u}^{iT}\mathbf{W}\mathbf{u}^j$, $\mathbf{W}$ being a weight matrix taking into account of the grid non-uniformity and possibles variables inhomogeneity, and $(\cdot)^T$ the transpose operator.

In analogy to Spectral Proper Orthogonal decomposition (SPOD), see Sieber et al. [137], it is possible to introduce a filtered version of the correlation matrix. The elements of this filtered matrix, denoted S_{ij}, are formulated as:

$$S_{ij} = \sum_{k=-N_f}^{N_f} g_k R_{i+k,j+k}, \tag{4.28}$$

where g_k is a filter function that can be of various forms, such as a box filter or a Gaussian filter and N_f is the filter span. Given the filtered correlation matrix, it is possible to define a filtered distance metric, $\|\mathbf{u}^i - \mathbf{u}^j\|^2_{\text{filt}}$, based on the filtering parameters N_f and g_k:

$$\|\mathbf{u}^i - \mathbf{u}^j\|^2_{\text{filt}} = S_{ii} + S_{jj} - 2S_{ij} = \sum_{k=-N_f}^{N_f} g_k \|\mathbf{u}^{i+k} - \mathbf{u}^{j+k}\|_2^2. \tag{4.29}$$

It is worth noting that $N_f = 0$, $g_k = 1$ leads to standard euclidean clustering.

From the practical point of view, similarly to the standard euclidean distance in which POD temporal coefficients are employed instead of snapshots to accelerate the distances computation (Li et al. [135]), here the same is made for the filtered distance. In particular, considering the SPOD temporal modes $a_m(t)$ that lead to the decomposition of the snapshots:

$$\mathbf{u}^i = \sum_{m=1}^{d} \boldsymbol{\varphi}_m a_m(t^i), \tag{4.30}$$

they are the eigenvectors (mutually orthogonal) of the filtered correlation matrix whose elements are S_{ij} (Sieber et al. [137]), it is possible to show that:

$$\|\mathbf{u}^i - \mathbf{u}^j\|^2_{\text{filt}} = \|\mathbf{a}^i - \mathbf{a}^j\|^2_2, \tag{4.31}$$

where $\mathbf{a}^i$ are the vectors containing all the SPOD temporal modes $a_m(t^i)$, $m = 1, \ldots, d$, with d being the number of non-zero eigenvalues of the filtered correlation matrix (lossless SPOD). Typically $d < N$ and this lead to relevant computational savings during the distances computation. For further details on SPOD the reader is referred to [137].

Parameters choice

Choosing the right number of clusters, K, and the optimal SPOD filter size, N_f, is pivotal to the integrity and accuracy of clustering analysis. One possible approach is to use the Bayesian Information Criterion (BIC), see Pelleg [138], Schwarz [139] and Priestley [140].

The BIC score is quantified as:

$$\text{BIC} = M\log(J) + K\log(M) - \frac{2}{d}\sum_{1}^{K} n_k \log(n_k/M), \tag{4.32}$$

where d (with $d \leq N$) is the dimension of the state vector. The BIC is a measure wherein a lower score is preferred, indicating a more favorable model (Konishi [141], Wit et al. [142]). This score, when considering the number of snapshots M, penalizes a high value of the inner-cluster variance J and a high number of clusters K. A high J implies that the points in the cluster are spread out, which might indicate the need for further clusters to describe the data or a different value of N_f to better describe the coherent spatio/temporal structures. On the other hand, an high number of clusters may result in overfitting and in a not converged probability distributions (4.22) of the dataset. BIC score offers a trade off between these purposes. It is worth to notice that the last term of (4.32) becomes negligible for high dimensional state vectors, such as the lossless SPOD reduction in which d is equal to the rank of the filtered correlation

matrix.

Finite Time Lyapunov Exponent for flow visualization

The Finite Time Lyapunov Exponent (FTLE) is a Lagrangian metric used to identify coherent structures within complex fluid flows, as discussed by [143]. It quantifies the divergence of pathlines within the flow, and these pathlines can be computed in either forward or backward time, each emphasizing attracting or repelling structures within the flow, respectively. In this study, we focus on the backward time FTLE, which highlights regions where fluid from different flow regions converges, revealing vortex structures within shear layers.

To compute the FTLE, we consider a set of initial conditions, denoted as $\mathbf{x}_0$, which are uniformly distributed. We then calculate the pathlines of these initial conditions using a fourth-order Runge-Kutta solver (RK4). The backward integration time, denoted as τ, is chosen to be 1.5 periods of the PVC oscillation. The resulting distribution of final points represents a flow map of the initial state, denoted as $\mathbf{\Phi}_\tau(\mathbf{x}_0)$. The spatial distribution of the FTLE, which depends on the time instance t^m and τ, can be calculated as [144]:

$$\lambda(\mathbf{x}_0, t^m, \tau) = \frac{1}{\tau} \ln \sqrt{\sigma_{\max}}, \tag{4.33}$$

where $\sigma_{\max}$ represents the maximum eigenvalue of the right Cauchy-Green deformation tensor $\mathbf{G}$, defined as:

$$\mathbf{G} = (\nabla \mathbf{\Phi}_\tau)^T \nabla \mathbf{\Phi}_\tau. \tag{4.34}$$

Here, ∇ denotes the gradient operator.

Within the framework of the CNM, the flow map $\mathbf{\Phi}_\tau$ is constructed by considering a realization of the clusters centroids sequence, and equation (4.24) provides a smooth temporal function suitable for the backward-in-time RK4 solver.

4.1.5 Toy model: SEIRD Model for Disease Spread

In the present investigation the SEIRD model has been employed as a representative toy model to test the efficacy of our Extended Clustering approach before applying it to high-dimensional flow data. The acronym SEIRD stands for Susceptible, Exposed, Infected, Recovered, and Deceased, which represent the five categories, or compartments, into which the population is divided. SEIRD is a commonly used compartmental model in epidemiology, designed to provide insights into the way infectious diseases can spread through a population (Hethcote [145], Piccolomini and Zama [146]).

- Susceptible ($\mathscr{S}$): This category includes individuals who are not infected with the disease but could potentially become infected if exposed.
- Exposed ($\mathscr{E}$): These are individuals who have been exposed to the disease and are carrying it, but are not yet infectious.
- Infected ($\mathscr{I}$): This category includes individuals who have been exposed to the disease and are now capable of transmitting it to others.
- Recovered ($\mathscr{R}$): These individuals have been infected and have fully recovered, and are now immune to the disease.
- Deceased ($\mathscr{D}$): This category includes individuals who have died from the disease.

The transitions between these categories are described by a system of differential equations. Let $\mathscr{H} = \mathscr{S} + \mathscr{E} + \mathscr{I} + \mathscr{R} + \mathscr{D}$ be the total population. The model assumes that the population size is fixed (i.e., birth and natural death are not considered), and is given by the following equations:

$$
\begin{aligned}
\frac{d\mathscr{S}}{dt} &= -\beta\frac{\mathscr{S}\,\mathscr{I}}{\mathscr{H}},\\
\frac{d\mathscr{E}}{dt} &= \beta\frac{\mathscr{S}\,\mathscr{I}}{\mathscr{H}} - \delta\mathscr{E},\\
\frac{d\mathscr{I}}{dt} &= \delta\mathscr{E} - (\gamma+\mu)\mathscr{I},\\
\frac{d\mathscr{R}}{dt} &= \gamma\mathscr{I},\\
\frac{d\mathscr{D}}{dt} &= \mu\mathscr{I},
\end{aligned}
\tag{4.35}
$$

where β is the transmission rate, δ is the rate at which exposed individuals become infectious (the inverse of the incubation period), γ is the recovery rate (the inverse of the infectious period), μ is the disease-induced death rate, t is the time expressed in days. The SEIRD model provides a useful tool for studying the dynamics of infectious diseases and can be used to predict the course of an outbreak under different conditions. The transmission rate (β) is set to 0.8, suggesting a relatively high likelihood of the disease spreading from infected to susceptible individuals. The recovery rate (γ) is fixed at 0.1, implying a mean infectious period of 10 days. The incubation rate (δ), which governs the speed at which exposed individuals become infectious, is set at 0.2. Lastly, a disease-induced death rate (μ) of 0.05 is chosen, indicating a 5% mortality rate among infected individuals. The ratio between transmission rate and the recovery rate is the well known R_0 index, $R_0 = \beta/\gamma$, that in this example is equal to $R_0 = 8$.

The simulation results presented in Figure 4.3 provide a visual representation of the dynamics of the SEIRD model with the specific parameters detailed previously. Over time, the Susceptible population decreases due to exposure to the disease, while the Exposed, Infected, and Deceased populations show respective increases based on the disease's transmission rate, incubation period, and death rate. Concurrently, the Recovered population rises as infected individuals overcome the disease.

Both the standard CNM and the eCNM have been applied to the

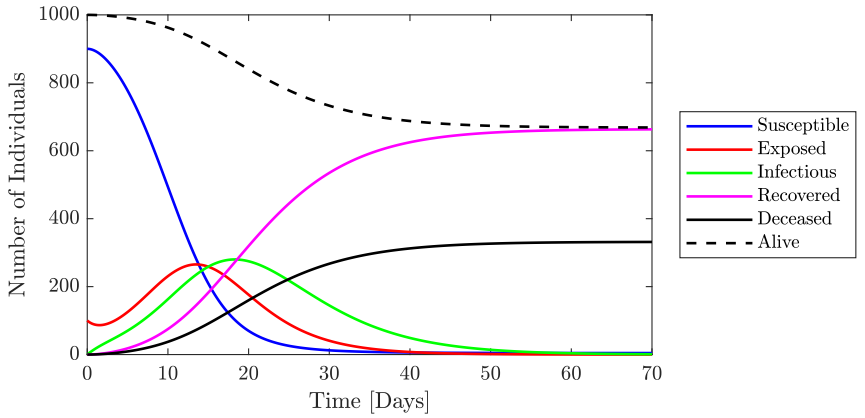


Figure 4.3. Time evolution of the SEIRD model. The simulation showcases the dynamics of Susceptible ($\mathscr{S}$), Exposed ($\mathscr{E}$), Infected ($\mathscr{I}$), Recovered ($\mathscr{R}$), and Deceased ($\mathscr{D}$) populations over time.

SEIRD dataset with a selected number of clusters K=10. For the eCNM, the Infectious population has been excluded from the clustering procedure. The simulation results are presented in Figure 4.4. This figure compares the original Infectious population curve, as predicted by the SEIRD model, drawn in black, with the discrete dynamics predicted with CNM, in red, and with the eCNM in blue. A continuous motion between the discrete jumps can be obtained by means of (4.25) or (4.24).

The standard CNM effectively replicates the time evolution of the Infectious population as it takes into account all five compartments of the SEIRD model during the clustering process. However, it's worth noting that the eCNM also successfully captures the dynamics of the Infectious population despite excluding this compartment from the clustering process. This is indicative of the inherent interconnections within the system variables, suggesting that the dynamics of the Infectious population can be inferred from the dynamics of the other compartments.

The eCNM remains effective even when analyzing the infectious population sampled at a reduced frequency. In Figure 4.4, the undersampled $\mathscr{I}$, with a sampling period of 5 days, depicted as a piecewise constant

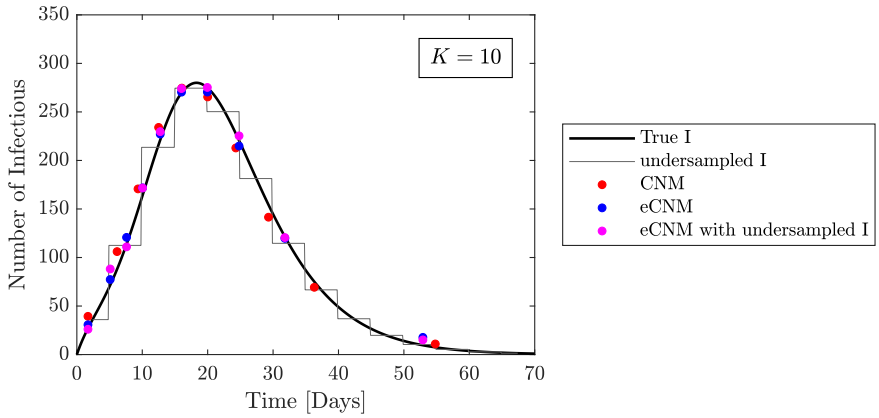


Figure 4.4. Comparison of Infectious population dynamics using original SEIRD model, standard CNM, Extended eCNM, eCNM with undersampled $\mathscr{I}$.

function, is shown in gray. By employing the method outlined in Section 4.1.2 by using a basic sampling matrix $\mathbf{P}$, the discrete dynamics of $\mathscr{I}$ with the undersampled data can be estimated, indicated by the magenta dots in the figure. The results showcase satisfactory performance.

The effectiveness of eCNM in capturing the dynamics of the Infectious population, as showcased by the SEIRD model simulation, is encouraging for its applicability in more intricate scenarios. This potential will be further investigated in subsequent sections, where eCNM will be applied to a real-world fluid dynamics system, demonstrating its robustness and adaptability in handling complex dynamical systems.

4.2 Application to swirl flame

In this section, we put to test the methodologies introduced earlier by applying them to real-world data. Our primary test case involves a specific fluid dynamics scenario: a turbulent swirl flame under self-excited oscillations.

In Subsection 4.2.1, we introduce the swirl flame and provide context on its operating conditions in the presence of a PVC. In Subsection 4.2.2, clustering is applied to the velocity components obtained using Particle Image Velocimetry (PIV). In Subsection 4.2.3, we use extended clustering to extend the PIV-based clusters to simultaneously recorded flame images, that capture the distribution of heat release in terms of OH* chemiluminescence. Subsection 4.2.4 takes a more specialized approach, focusing on the application of extended clustering on a subdomain of the flow field. Concluding the exploration, Subsection 4.2.5 delves into extended clustering on temporal undersampled data, where clustering is applied to pressure measurements and subsequently extended to the undersampled velocity components.

4.2.1 Model flow: Swirl Flame

The dataset analyzed in this section comes from a combustor with a swirl-stabilized natural gas flame. The flow in the combustion chamber exhibits a precessing vortex core, a helically shaped coherent flow structure, commonly found in both reacting and nonreacting swirling flows when a vortex breakdown happens [147, 148].

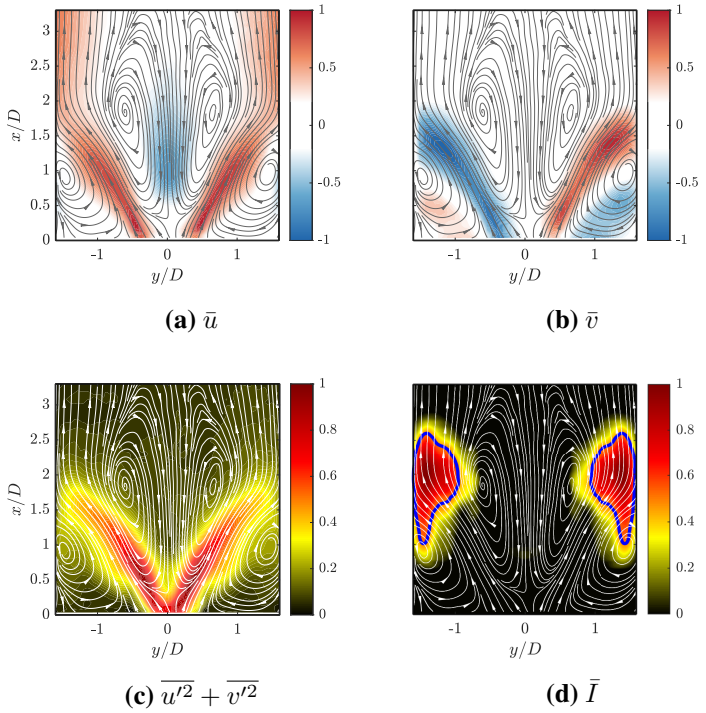
In the context of combustion systems, the interplay between oscillations of the flow field and flame behavior is crucial. In many combustion systems, such as those used in aero-engines or gas turbines, swirling flow field is generated to stabilize the flame (Candel et al. [149]). These configurations often feature a Precessing Vortex Core (PVC), a large-scale coherent flow structure frequently observed in swirling flows (Syred [150]). The presence of a PVC can influence the stability of the combustion process (Oberleithner et al. [151]). Its effect on flame anchoring can also impact combustion efficiency (Stöhr et al. [152], Lückoff et al. [153], An et al. [154]). Depending on its characteristics, the PVC can either stabilize or destabilize the flame, which can be used for flow control (Lückoff and Oberleithner [155], Müller et al. [156]).

The first part of the dataset consists in Particle Image Velocimetry (PIV) measurements, which provide time data for both velocity compo-

nents and synchronous flame images [157]. The latter are given in the form of UV-images filtered around the characteristic wavelength of OH* chemiluminescence. The chemiluminescence emission indicates the presence of excited OH molecules, a species which is known to be produced in heat-releasing reactions within the flame. In particular, perfectly premixed flows exhibit a linear correlation between the spatial distributions of OH* chemiluminescence and heat release rate [158]. The sampling rate of the UV images and the PIV recordings is $f_s = 2\,\mathrm{kHz}$. The PIV grid consist in 73×77 measurements while the UV images are 251×251. By studying the velocity components together with the heat release, denoted by q, it is possible to gain insights into the combustion dynamics when a PVC is present.

Figure 4.5 depicts the time-averaged flow field within the combustor in its natural state. Panels (a) and (b) show the temporal mean of the velocity components u and v, respectively. These panels reveal a radial expansion of the swriling jet flow with a conical vortex breakdown shape [159]. The measurement domain spans up to the combustor side walls, making evident the attachment and deflection of the jet upon the wall. Within the conical jet, the streamlines reveal the presence of a large inner recirculation zone (IRZ), induced by vortex breakdown. This distinct zone is delimited at its upstream extremity by a stagnation point located along the axis of the jet. Adjacent to the jet and the outer confinement, an outer recirculation zone (ORZ) is visible. The formation of this ORZ can be attributed to the sudden expansion in cross-sectional area occurring between the mixing tube and the combustion chamber. Moving forward to Panel (c), it illustrates the turbulent kinetic energy (TKE), defined as $u'^2 + v'^2$. Panel (d) then showcases the radial distribution of time-averaged heat release, $\bar{q}$, obtained from an Abel deconvolution of the UV images [160] . It displays a detached, M-shaped flame configuration starting from the combustor inlet. In all panels, D denotes the diameter of the inlet section, x and y are the axial and transversal coordinate respectively; every field variable has been normalized with respect to its maximum value.

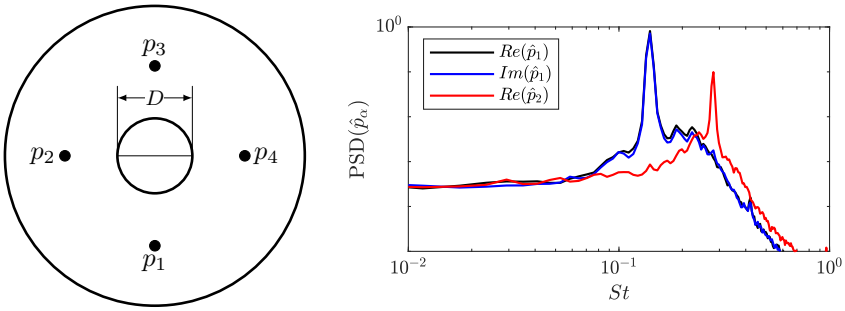
In addition to the PIV dataset, pressure signals have been collected. Four pressure probes, located on the burner base plate are circumferentially



(a) $\bar{u}$ **(b)** $\bar{v}$

(c) $\overline{u'^2} + \overline{v'^2}$ **(d)** $\bar{I}$

Figure 4.5. Mean (Time-averaged) flow. Panel (a)-(b): mean axial ($\bar{u}$) and transversal ($\bar{v}$) velocity components. Panel (c): mean turbulent kinetic energy (TKE) distribution. Panel (d): Abel deconvoluted temporal mean of heat release distribution $\bar{q}$. Blue lines are the $\bar{q}/\bar{q}_{\max} = 0.5$ isolines. All variables are normalized with respect to their respective maxima.



(a) Pressure probes locations

(b) PSD of azimuthal pressure component

Figure 4.6. Pressure signals analysis.

arranged around the inlet at $x = 0$. Their location is shown in panel (a) of figure 4.6. The data from these probes have been sampled at a rate of $f_{s_p} = 16384\,\mathrm{Hz}$.

To analyze these signals, they have been decomposed into Fourier modes characterized by azimuthal wavenumbers α:

$$\hat{p}_\alpha(t) = \sum_{k=1}^{4} p_k(t) \exp\left(i\frac{2\pi\alpha}{4}(k-1)\right). \tag{4.36}$$

The resultant Fourier modes are complex and it is worth noticicing that

$$\begin{aligned} Re(\hat{p}_1) &= p_1 - p_3, \\ Im(\hat{p}_1) &= p_2 - p_4, \\ Re(\hat{p}_2) &= (p_1 + p_3) - (p_2 + p_4), \\ Im(\hat{p}_2) &= 0. \end{aligned} \tag{4.37}$$

Further insights can be drawn from panel (b) of figure 4.6, where the Power Spectral Densities (PSD), computed with Welch's method [161], are plotted. The real part of $\hat{p}_1$ is shown in black, its imaginary part in blue, and the real part of $\hat{p}_2$ in red. The first Fourier mode $\hat{p}_1$ is related to

the PVC with normalized frequency of $St = 0.14$, while $\hat{p}_2$ is associated to its first harmonic. Here, the Strouhal number, St, is expressed as $St = fD/U$, where U denotes the reference velocity.

For the subsequent analyses each field variable ϕ has been decomposed in its temporal mean $\bar{\phi}$ plus the fluctuation ϕ':

$$\phi(x, y, t) = \bar{\phi}(x, y) + \phi'(x, y, t). \tag{4.38}$$

All the methodologies have been applied onto the fluctuating parts.

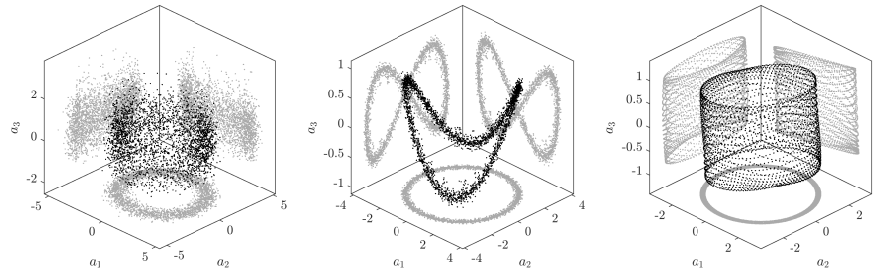
This comprehensive dataset, characterized by both its detailed flow patterns and spectral features, will serve as the foundational basis for the analyses presented in the subsequent sections.

4.2.2 CNM analysis

The analysis starts with the application standard clustering to velocity components PIV measurements. The computational load of clustering is significantly reduced by an effectively lossless SPOD compression detailed in section 4.1.4. In fact, all operations are performed on the SPOD amplitude vector $\mathbf{a}^i = [a_1(t^i), a_2(t^i), \ldots, a_d(t^i)]^T$ instead of the snapshots $\mathbf{u}_s^i$.

Preliminary remarks

The preliminary step in this process is to examine the phase portrait of the temporal modes $a_m(t)$ across different values of the SPOD filtering parameter N_f. As depicted in Figure 4.7, the phase portrait of the three leading modes for various N_f values is presented. For the case of $N_f = 0$, as shown in panel (a) which corresponds to the application of standard POD, the phase portrait appears to be somewhat chaotic, especially in the third dimension. On the contrary, when $N_f = n_t$ that correspond to a DFT, as seen in panel (c), the first two dimensions of the phase portrait become distinctly clear, although the third dimension remains ambiguous. A good comprimise is realized at $N_f = 75$, demonstrated in panel (b). Here, the motions in all the three dimensions are both distinct



(a) $N_f = 0$ **(b)** $N_f = 75$ **(c)** $N_f = M$

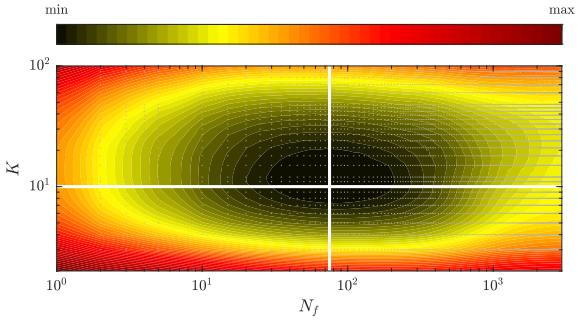
Figure 4.7. Phase portraits of the three leading temporal modes for various N_f values.

and interpretable. Notably, from an inspection of this figure, it's evident that the third dimension is representative of a harmonic corresponding to the primary frequency.

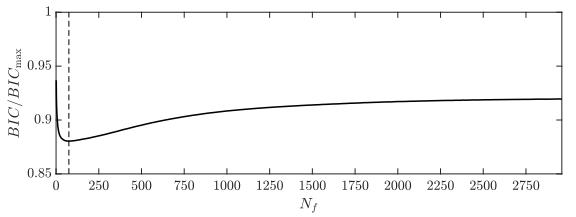
The first step for the clustering approach is to determine is the sizing of the cluster-based model, namely the choice the proper number of clusters K (Nair et al. [8]) and the optimal value of the filter dimension for the best distance metric. The optimal values for both N_f and K are determined according to the Bayesian information criterion (BIC score), delineated in equation (4.32). Figure 4.8 show the evaluation of the BIC score. Specifically, panel (a) showcases the contour plot of the normalized BIC score plotted against N_f and K. Subsequent panels offer sectional views: one delineating BIC against N_f for a fixed $K = 10$ and another versus K at a specified $N_f = 75$. A distinct minimum in the BIC score is found at $K = 10$ and $N_f = 75$. The latter correspond to about 7.5 period times of PVC oscillations. Based on this analysis, these values are adopted for the subsequent analyses.

CNM of velocity components

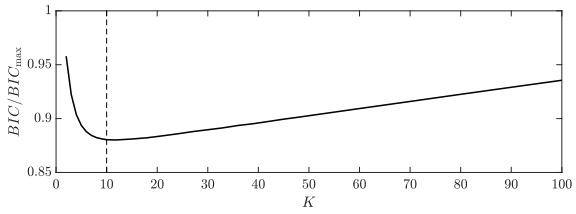
Once the ideal parameters are identified, the subsequent step involves partitioning the phase space. For this purpose, the Kmeans++ algorithm is



(a) BIC score



(b) $K = 10$



(c) $N_f = 75$

Figure 4.8. Bayesian information criterion (BIC) for parameter selection. The minimum is observed at $K = 10$ and $N_f = 75$.

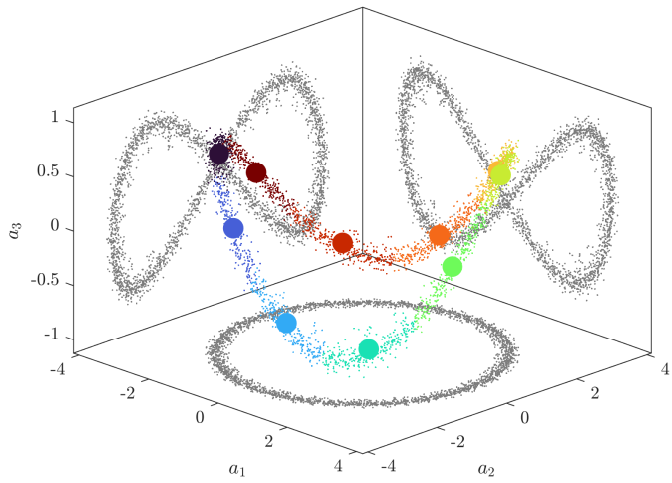
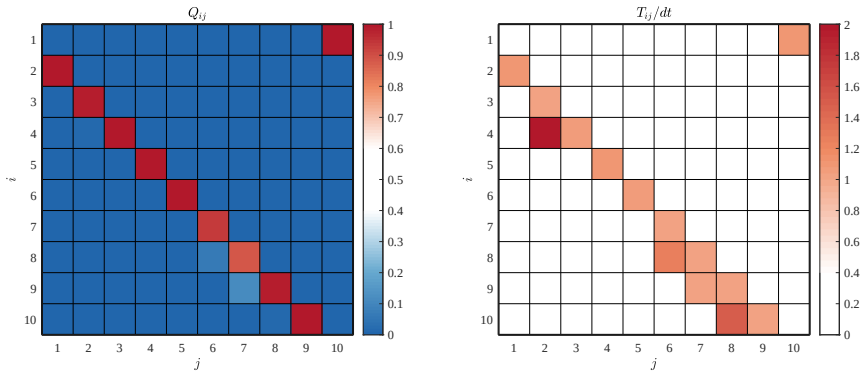


Figure 4.9. Clusters of the a_i phase potrait. K-means++ algorithm hase been employed for the clustering. $K = 10$ and $N_f = 75$ are used.

employed. A visual representation of this clustering is illustrated in Figure 4.9. The depicted phase portrait differentiates the centroids of each cluster, represented by larger dots, from the individual snapshots, represented by smaller dots. Each snapshot is color-coded according to its pertaining centroid, highlighting the cluster boundaries and affiliations in the phase space. The phase portrait vividly displays the evolution corresponding to a limit cycle.

The limit cycle behavior is highlighted with the construction of the Cluster Network Model described in section 4.1.3. In particular direct transition probabilities matrix Q_{ij} and direct transition times matrix T_{ij} can describe the system dynamics by telling how often and how quickly our system moves from one state (or cluster) to another. In figure 4.10 are reported Q_{ij}, panel (a), and T_{ij}, panel (b). The most probable transitions are consistently to the next cluster in the sequence, corroborating the limit cycle nature of the dynamics.

Moreover, from the transition times matrix the leading frequency can



(a) Q_{ij} **(b)** T_{ij}

Figure 4.10. CNM's direct transition probabilities (Q_{ij}) and times (T_{ij}) matrices.

be estimated by means of:

$$St_{\mathrm{PVC}} \approx \frac{1}{T_{2,1} + T_{3,2} + \ldots + T_{1,10}} \frac{D}{U} \approx 0.14. \tag{4.39}$$

The cluster centroids for the velocity components u and v can be examined to discern patterns and underlying dynamics. Using the sequence of cluster centroids smoothed according to (4.24), the backward finite time Lyapunov exponent (FTLE), detailed in appendix 4.1.4, serves to elucidate the structures influenced by the velocity centroids. The backward integration time for the FTLE calculation was set to 1.5 times the period of the PVC oscillation. In Figure 4.11, the spatial distribution of FTLE for selected centroids is presented. These distributions distinctly highlight the vortex structures formed by the PVC. The FTLE in the xy-plane reveals vortex roll-up, characteristic of Kelvin-Helmholtz instability. However, the PVC's helical nature results in alternating vortices forming a helix. These vortices, moving downstream along the inner shear layer, eventually collide with the combustor wall. Additionally, an overlay of the heat release cluster centroids is included for further analysis in the subsequent

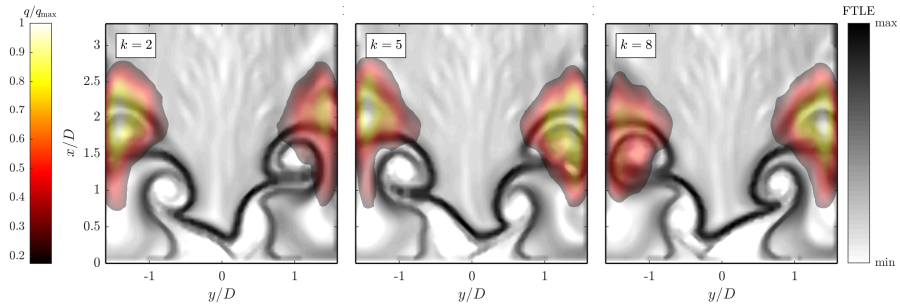


Figure 4.11. Spatial distribution of the backward finite time Lyapunov exponent (FTLE) for selected velocity centroids. The figure also includes an overlay of heat release cluster centroids. Each distribution has been normalized with respect to its maximum.

section.

4.2.3 Extended CNM of the heat release

Having successfully clustered the velocity components and constructed the associated CNM, it is possible to proceed to the construction of the extended clusters. Leveraging the methodology introduced in section 4.1.2, the extended clusters $\mathbf{c}_k^e$ are derived using the function ξ_k^m, as detailed in Eq. (4.9). This allows for the incorporation of the heat release data into our clustering model, providing a comprehensive representation that combines both velocity and flame patterns. For the analysis the heat release has been decomposed in temporal mean $\bar{q}$ plus fluctuation q':

$$q = \bar{q} + q', \tag{4.40}$$

where the fluctuation term q' has been subjected to the clustering procedure.

The analysis of the heat release centroids can provide insights into the flame dynamics and the structures that arise within the combustor. Figure 4.11 displays an overlay of the backward FTLE analysis of velocity centroids and the heat release reconstruction. The heat release reconstruction

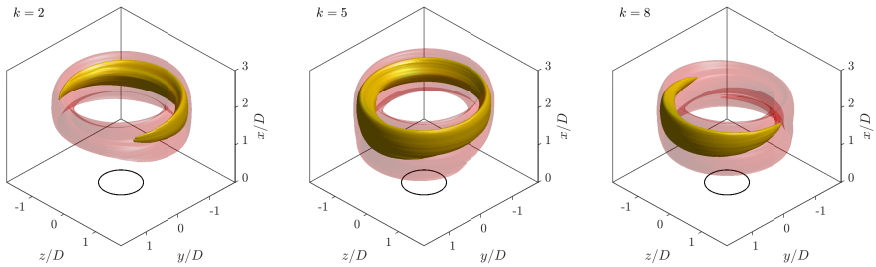


Figure 4.12. Three-dimensional reconstruction of heat release rate centroids, representing the dynamic behavior of the flame. This visualization combines the Abel-deconvoluted symmetric mean flame with symmetric (Abel-deconvoluted) and antisymmetric heat release rate fluctuations, representing the azimuthal wavenumber $\alpha = 1$ component. In all panels, the isosurfaces $q/q_{\max} = 0.5$ are depicted in semi-transparent red, while $q/q_{\max} = 0.8$ in yellow.

combines the Abel-deconvoluted symmetric mean flame with the fluctuating heat release rate centroids. This overlay reveals the predominant localization of heat release (q) within areas influenced by the PVC vortices roll-up. The progression of these vortices extends from the center to the sides of the combustor chamber, reflecting the PVC's impact on flame dynamics [**?**].

Further emphasizing the flame's rotational motion, Figure 4.12 includes a 3D plot showcasing the dynamic behavior of the flame. This 3D reconstruction integrates the symmetric mean flame, obtained through Abel-deconvolution, with symmetric (Abel-deconvoluted) and antisymmetric fluctuations in the heat release rate, the latter representing the azimuthal wavenumber $\alpha = 1$ component. Specifically, the isosurface representing $q/q_{\max} = 0.5$ is drawn in semi-transparent red, while the isosurface for $q/q_{\max} = 0.8$ is rendered in yellow. The sequence of clusters centroids vividly illustrates the dynamic of the combustion, showing the rotating pattern of the flame.

The efficacy of the extended CNM in predicting q dynamics is evaluated using an error measure, denoted by $\varepsilon(t)$. This metric is formulated

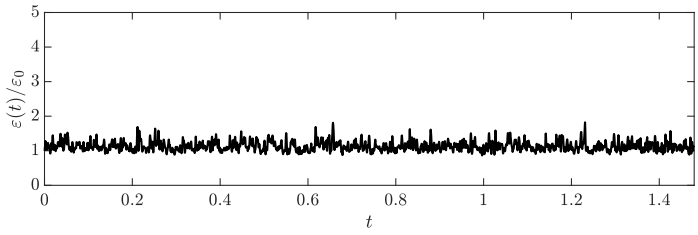


Figure 4.13. Reconstruction error $\varepsilon(t)$.

as:

$$\varepsilon(t) = \sqrt{\frac{1}{A}\int_{\Omega}(q(x,y,t) - q_r(x,y,t))^2 dA}, \tag{4.41}$$

where A represents the measurement area and q_r the reconstructed distribution of heat release rate. Given the inherent error introduced by the clustering procedure, it is insightful to consider the error in relation to its initial value, $\varepsilon_0 = \varepsilon(0)$. Figure 4.13 displays the evolution of the error $\varepsilon(t)$. Notably, the error remains close to the initial value ε_0 ($\varepsilon/\varepsilon_0 \approx 1$), indicating the model's reliable predictive capability for the q temporal evolution. It should be noted that the initial error ε_0 is inherent to the clustering process and is unavoidable unless the number of clusters equals the number of snapshots. An effective model is characterized by its ability to not accumulate further errors arising from incorrect modeling of the dynamics [10].

The model's performance can be also assessed using the unbiased autocorrelation function R, as suggested by Deng et al. (Deng et al. [80]). This function, defined by Protas et al. [162] as:

$$R(\tau) = \frac{1}{T_f - \tau}\int_{\tau}^{T_f} \mathbf{b}(t-\tau)^T \mathbf{b}(t) dt, \qquad \tau \in [0, T_f), \tag{4.42}$$

provides insights into the temporal relationships within the dataset. Here, T_f is the final time, τ stands for time delay, and $\mathbf{b}$ is a generic vector that contains the values of the field variable under examination at all spatial locations. In Figure 4.14, the autocorrelation function for the velocity

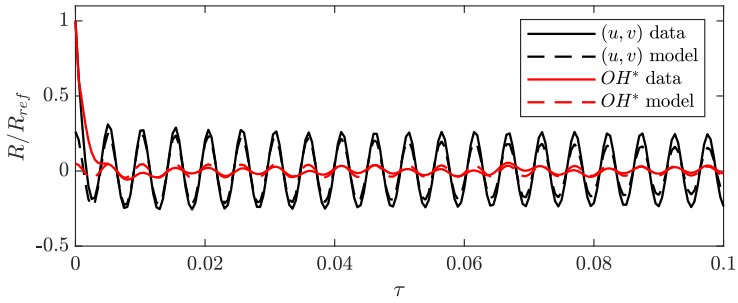


Figure 4.14. Autocorrelation function R for velocity components and I. Comparison between the real data and the model outputs. R_{ref} corresponds to the $R(0)$ of the actual dataset.

components and I is shown. Both real and model-generated data are included. The reference autocorrelation value, R_{ref}, used in these plots, corresponds to the $R(0)$ of the actual dataset. By comparing the actual and model-derived R, it's evident that the model proficiently reproduces the temporal statistics of the flow.

The presented results underscore that the extended clustering method performs effectively when there's a pronounced correlation between the data incorporated into the clustering process and the data left unclustered. This correlation ensures that the method can accurately capture and represent underlying dynamics.

4.2.4 Extended CNM analysis on spatial subdomains

In many situations, rather than considering the entire domain, it may be of interest to focus on a specific subdomain which exhibits unique or dominant features. By focusing on such a subdomain, one can often uncover specific behaviors or patterns that might be obscured in a full domain analysis. In this section the extended CNM is tested considering only a spatial subdomain of the velocity components to describe the dynamic of I in the entire domain.

Figure 4.15 offers a visualization of the time-averaged streamwise

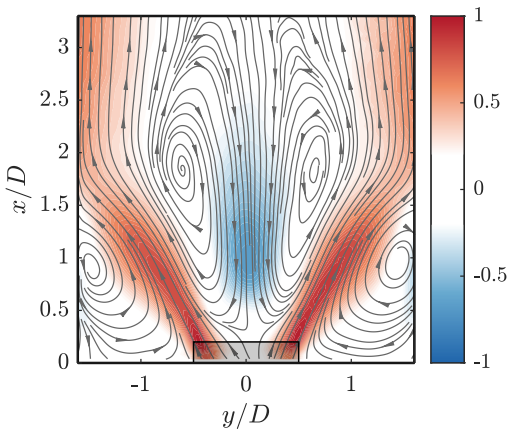


Figure 4.15. Spatial subdomain for extended CNM. Normalized $\bar{u}$ distribution with respect to its maximum.

velocity component, $\bar{u}$, together with the mean flow streamlines. The region of interest, adjacent to the inlet section, is highlighted in gray.

With the same procedure as previous section, the function ξ_k^m is constructed considering only data from the subdomain. This leads to the determination of the cluster sequence and the extended cluster centroids. The latter due to their resemblance to the ones presented in figures 4.11 and 4.12, are not reported herein.

A crucial step in this process is the quantitative evaluation of the model's ability to capture the dynamics of I. To this end, figure 4.16 displays a comparison of the error ε, described in Eq.(4.41). Specifically, the black curve represents the error for the full domain model, while the red curve corresponds to the subdomain model. Remarkably, the deviation between these two curves is negligible, indicating that both models perform equivalently in capturing the dynamics of the system.

The observed congruence in performance between the full domain and subdomain models can be attributed to the strong correlation between the

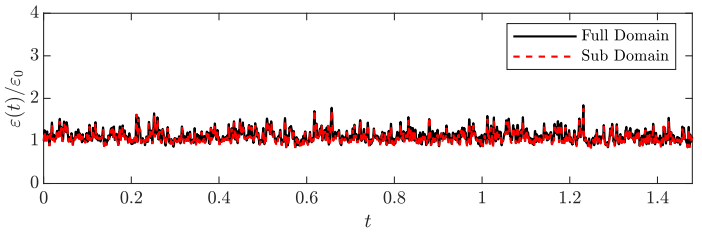
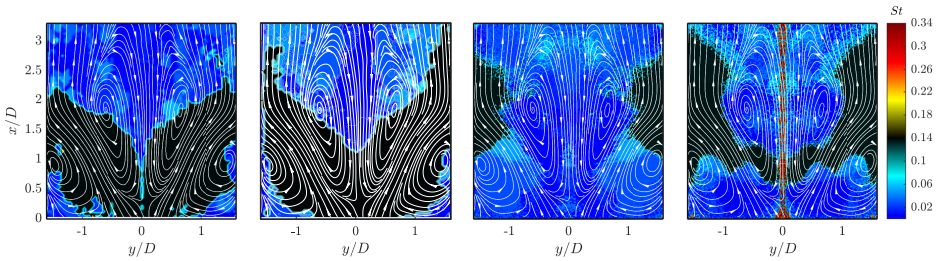


Figure 4.16. Comparison between ε for the full domain, black line, and sub-domain, red dashed line.



(a) u' **(b)** v' **(c)** I'_s **(d)** I'_a

Figure 4.17. Leading local frequency. The white lines are the streamline pattern of the mean flow.

flow at the inlet section and the regions where most dynamics occur. This relationship is clarified in figure 4.17, which showcases the spatial distribution of the predominant local frequency across different field variables. Specifically, panel (a) of the figure details the primary frequency for u', panel (b) for v', and panels (c) and (d) for I'_s and I'_a respectively. A crucial observation is that the inlet section exhibits the same leading frequency as the PVC region in both u and v distribution. Similarly, there's a match in the leading frequencies associated with the helical structures of I'_a and I'_s. This synchronization in frequencies underlines the strong correlation between the regions, and it provides an explanation for the comparable results presented in figure 4.16.

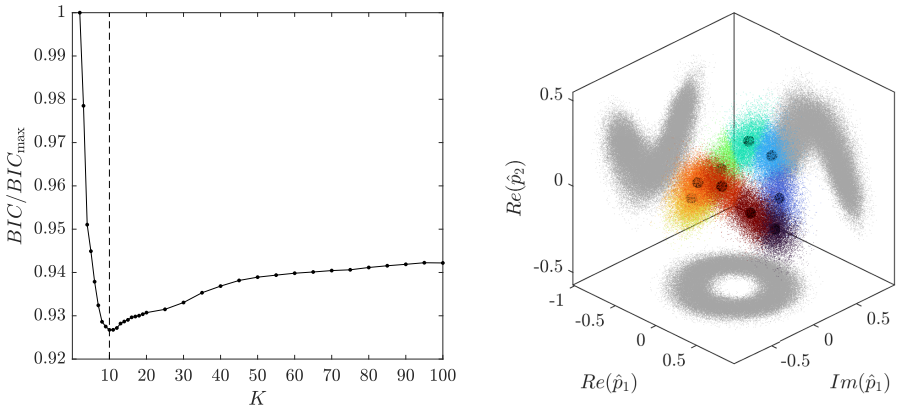
4.2.5 Extended CNM analysis for temporally undersampled data reconstruction

In this section the effectiveness of the approach described in Section 4.1.2 to deal with temporal undersampled data is evaluated. The Fourier azimuthal modes of pressure measurements detailed in Section 4.2.1 served as available high-resolution data, while the velocity components acted as the temporally undersampled dataset.

The first step involved applying the standard CNM to the $\alpha = 1$ and 2 pressure azimuthal modes. The BIC score, defined in (4.32), for this application is displayed in panel (a) of Figure 4.18. The optimal number of clusters are identified where the BIC is at its minimum. A notable dip in the BIC score is evident at $K = 10$. In panel (b) of the same figure, a phase portrait is presented, with the cluster centroids distinctly marked. The various colors in the figure represent the cluster affiliations. The transition probability Q_{ij} and the time T_{ij} matrices (not reported herein) identify a behavior consistent with a limit cycle, showing patterns analogous to that seen in Figure 4.10.

In this study, after determining the cluster sequence from the temporally high-resolution pressure data, the matrix Ξ, containing each ξ_k^m was extracted. Here, m ranges from 1 to the total number of pressure measurements, M_p, as detailed in section 4.1.2. By addressing the optimization problem given by Equation (4.16), where undersampled data $\tilde{\mathbf{u}}_i$ coincide with the available undersampled velocity components, the extended clusters' centroids for the velocity were estimated. The difference in sampling frequencies between pressure and velocity measurements has been accounted with the matrix $\mathbf{P}$ that in this case has dimension $M_u \times M_p$.

In panel (a) of figure 4.19, the first v' cluster centroid derived using this approach is shown. Interestingly, this centroid aligns closely with the one in panel (a) of figure 4.11, which was obtained through standard clustering directly on velocity components. Panels (b) through (f) in figure 4.19 display the estimates of the first centroid at different sampling frequencies f_s^e, by artificially further undersampling the velocity field, achieved by using distinct configurations of the sampling matrix $\mathbf{P}$. As expected,



(a) BIC **(b)** Phase portrait

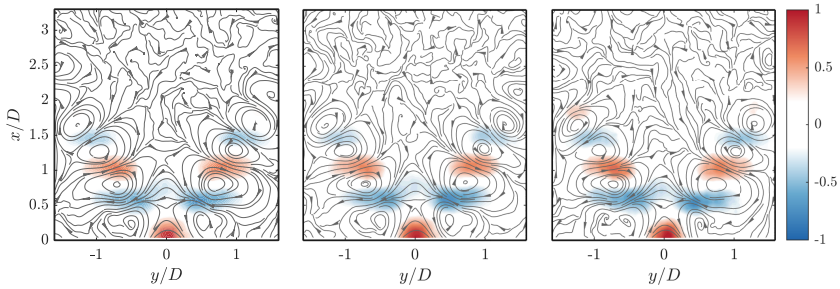
Figure 4.18. CNM of pressure azimuthal modes. Panel (a): BIC score as a function of the number of clusters. Panel (b): Phase portrait of the $m = 1$ and 2 pressure azimuthal modes.

a decrease in the sampling frequency is associated with a diminished quality of the clusters centroids. Considering that the PVC dynamics are associated with a frequency of approximately $200\,\mathrm{Hz}\,(St = 0.14)$, even when panels (d) to (f) represent frequencies below the Nyquist limit, the derived clusters remain physically interpretable. It is worth to notice that a sampling frequency below the PVC Nyquist limit would not allow any standard analysis of such dynamics.

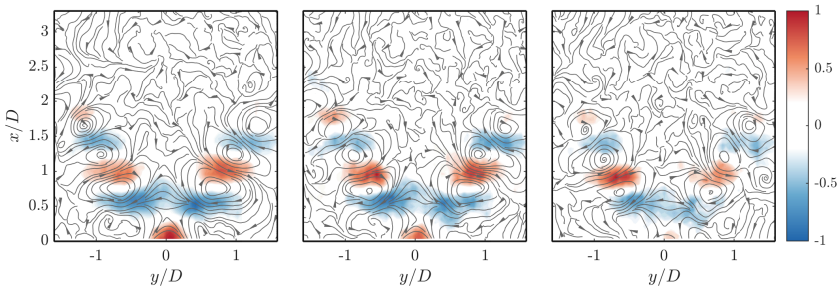
The effectivenes of the proposed method was assessed using the mean reconstruction error for the velocity field, denoted by $\bar{\varepsilon}_{\tilde{\mathbf{u}}}$. This mean error is calculated as:

$$\bar{\varepsilon}_{\tilde{\mathbf{u}}} = \sqrt{\frac{1}{AT_f}\int_0^{T_f}\int_\Omega (\tilde{\mathbf{u}}(t) - \tilde{\mathbf{u}}_r^e(t))^T(\tilde{\mathbf{u}}(t) - \tilde{\mathbf{u}}_r^e(t))\,dA\,dt}, \qquad (4.43)$$

where $\tilde{\mathbf{u}}_r^e$ is the velocity field constructed from the discrete sequence of extended centroids, subsequently smoothed according to (4.25). Panel (a)



(a) $f_s^e = 2\,\text{kHz}$ **(b)** $f_s^e = 1\,\text{kHz}$ **(c)** $f_s^e = 400\,\text{Hz}$



(d) $f_s^e = 200\,\text{Hz}$ **(e)** $f_s^e = 100\,\text{Hz}$ **(f)** $f_s^e = 50\,\text{Hz}$

Figure 4.19. Extended cluster centroids for v' velocity component derived using temporally undersampled data.

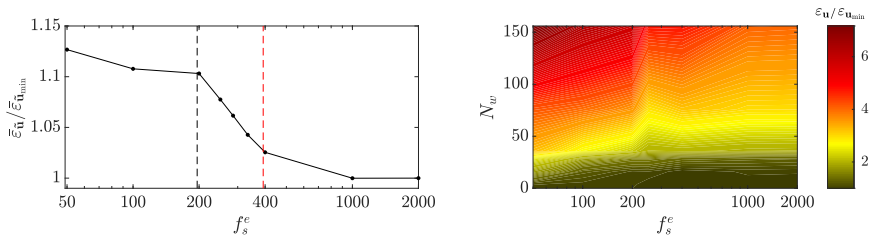


Figure 4.20. Analysis of the reconstruction error for temporally undersampled velocity data.

of Figure 4.20 plots $\varepsilon_{\tilde{\mathbf{u}}}$ against f_s^e. The error $\varepsilon_{\tilde{\mathbf{u}}}$ is normalized with respect to the error value at $f_s^e = 2\,\mathrm{kHz}$, namely when using all the available velocity snaphots. As anticipated, when the sampling frequency decreases, the error rises. However, it stays below 10% even for $f_s^e < 2f_{\mathrm{PVC}}$, f_{PVC} being the PVC frequency. In the plot, vertical black dashed line represents the PVC frequency, and the red line indicates the Nyquist limit at $2f_{\mathrm{PVC}}$. In panel (b) of Figure 4.20, the influence of a box windowing parameter N_w during the computation of $\mathbf{P}$ is examined. The results indicate that windowing adversely impacts $\varepsilon_{\tilde{\mathbf{u}}}$. Such an outcome is attributed to the fact that, in this test, $\tilde{\mathbf{u}}$ was procured through straightforward subsampling, making simple sampling the most effective window choice.

When exploring subsampling techniques, it's interesting to compare uniform subsampling, which was previously utilized, with random subsampling. Random subsampling can be beneficial when the sampling frequency approaches, or is less than, the Nyquist limit (see Manohar et al. [134]). In Figure 4.21, a comparison of the error $\varepsilon_{\tilde{\mathbf{u}}}$ is presented, as a function of the percentage of sampled velocity data relative to the available data (set at $f_s^e = 2\,\mathrm{kHz}$). The black line represents the results from uniform undersampling, while the red line showcases the effects of random subsampling. Black and red vertical dashed line represent the percentage of sampled snapshots corresponding the the PVC frequency and the Nyquist limit respectively. Notably, for a sampling frequency near or

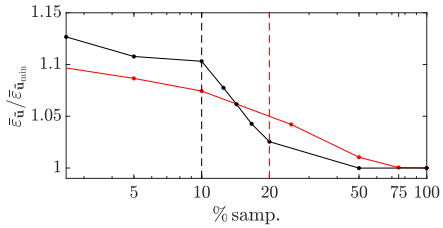


Figure 4.21. Comparison between mean reconstruction error of uniform (black) and random (red) sampling. Black and red vertical dashed line are % sampled corresponding the the PVC frequency and the Nyquist limit respectively.

lower than the Nyquist limit (around $f_s^e \approx 400\,\mathrm{Hz}$), random subsampling offers superior results compared to its uniform counterpart.

4.3 Comparison between CNM and POD-ROM

The aim of this section is to develop ROMs for the flow around an airfoil configuration. In particular two kinds of low order models are built: proper orthogonal decomposition (POD) Galerkin ROMs (Stabile and Rozza [163]), based on the projection of the governing equations onto selected POD modes, and the cluster-based network models (CNM) (Kaiser et al. [83] and Li et al. [135]).

This section is organized as follows: section 4.3.1 contains a brief introduction the analyzed configuration and the FOM. In section 4.3.2 POD-ROM is recalled. Section 4.3.3 reports the obtained results and the comparison between the different low order models.

4.3.1 Full order model of an airfoil configuration

Drag reduction and improvement of the lift-drag ratio of an aircraft has always been an important goal for aircraft design (Couto and Bergada [164] and Chiatto et al. [165, 166]). The implementation of flow control devices such as synthetic jets (Ceglia et al. [167], Girfoglio et al. [168],

de Luca et al. [169]) can be helpful for both delay separation and turbulence transition (Tousi et al. [170] and Tousi et al. [171]). By optimizing the shape of the airfoil and the control parameters, the available lift and resulting drag were tailored for the desired flight envelope. Aerodynamics optimization usually involves a lot of numerical simulations that can be time-consuming and computationally expensive.

The physical problem analyzed in the present work is the two dimensional, unsteady, incompressible flow past a NACA0012 airfoil. Numerical solutions of the Unsteady Raynolds-Averaged Navier-Stokes (URANS) equations have been obtained by using pimpleFoam, a one-phase solver from the open-source software OpenFOAM, which uses the PIMPLE algorithm for pressure–velocity coupling and allows the modeling of the 3D transient flow on a discretized domain by solving the Reynolds-averaged Navier–Stokes equations. Turbulence has been modeled using the widely validated $\kappa - \omega$ turbulence model. In particular, the equations system related to as FOM are reported in Eq.(4.44).

$$\nabla \cdot \mathbf{u} = 0, \tag{4.44a}$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} + \nabla p - \nabla \cdot \left(\nu_{\text{eff}}(\nabla \mathbf{u} + [\nabla \mathbf{u}]^T)\right) = 0, \tag{4.44b}$$

where t is the time, p is the kinematic pressure, $\mathbf{u}$ is the velocity vector $\mathbf{u} = [u, v, w]$, $\nu_{\text{eff}} = \nu + \nu_t$ with ν and ν_t being the kinematic and turbulent eddy viscosity, respectively. The free-stream condition of Reynolds number is $Re_\infty = U_\infty c/\nu = 7000$, where U_∞ is the free-stream velocity and c is the airfoil chord. The angle of attack, α, has been set to $\alpha = 8°$ for the present investigation. The FOM unsteady CFD simulations are run until the time-averaged flow variable values (e.g., velocity, drag, and lift) become constant. Then snapshots of flow field variables are saved for the construction of the low order models.

The FOM, incorporating the URANS approach and the $\kappa - \omega$ turbulence model, provides a detailed representation of the unsteady turbulent flow around the NACA0012 airfoil. It serves as the foundation for further analysis and the development of reduced-order models.

4.3.2 POD-ROM

Following the approach described by Hijazi et al. [172], a systematic procedure combining the POD technique and Galerkin projections can be utilized to obtain ROMs from simulation or experimental data.

Specifically, the truncation of the POD expansions for the discretized fields of velocity $\mathbf{u}$, pressure p, and eddy viscosity ν_t can be expressed as follows:

$$\begin{bmatrix} u(x_i, y_i, t) \\ v(x_i, y_i, t) \end{bmatrix} = \tilde{\mathbf{\Phi}}_{\mathbf{u}} \mathbf{a}_r, \quad \left[p(x_i, y_i, t)\right] = \tilde{\mathbf{\Phi}}_p \mathbf{b}_r, \quad \left[\nu_t(x_i, y_i, t)\right] = \tilde{\mathbf{\Phi}}_{\nu_t} \mathbf{c}_r, \tag{4.45}$$

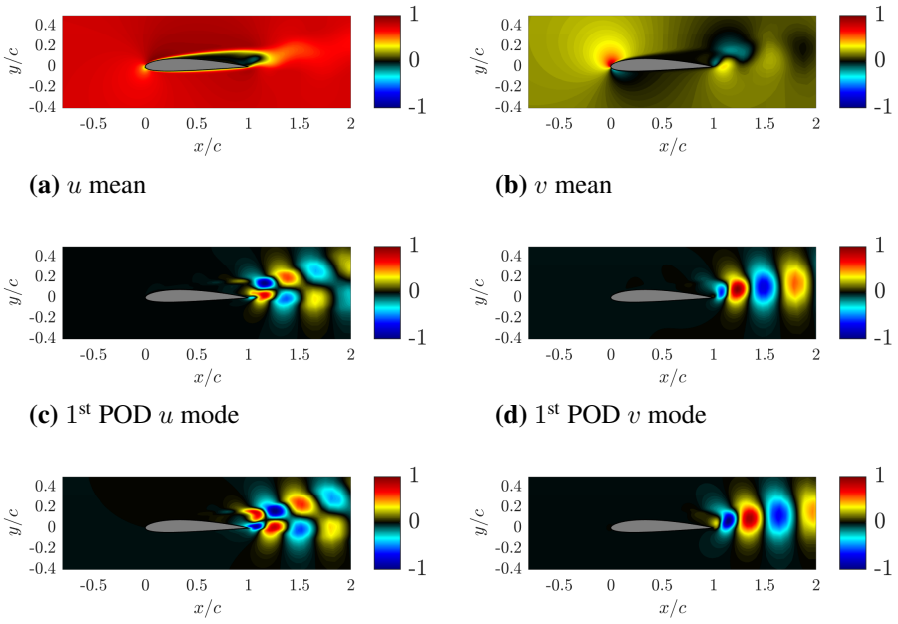
where $\tilde{\mathbf{\Phi}}_{\mathbf{u}}$, $\tilde{\mathbf{\Phi}}_p$ and $\tilde{\mathbf{\Phi}}_{\nu_t}$ represent the POD mode matrices of velocity, pressure, and eddy viscosity, respectively. These mode matrices are computed using the procedure outlined by Stabile and Rozza [163] based on the numerical data obtained from the FOM.

The momentum equation is then projected onto the $\tilde{\mathbf{\Phi}}_{\mathbf{u}}$ modes, resulting in the following low order system:

$$\frac{d\mathbf{a}_r}{dt} = f_{\mathbf{u}}(\mathbf{a}_r, \mathbf{c}_r, \mathbf{b}_r), \tag{4.46}$$

where $\mathbf{a}_r$, $\mathbf{c}_r$, and $\mathbf{b}_r$ represent the coefficients associated with the velocity, eddy viscosity, and pressure fields, respectively. The projections of the pressure Poisson equation and the eddy viscosity model provide additional equations for the determination of the $\mathbf{b}_r$ and $\mathbf{c}_r$ coefficients. Initial conditions for $\mathbf{a}_r$ and $\mathbf{c}_r$ are obtained by projecting $\mathbf{u}(0)$ and $\nu_t(0)$ onto the POD modes. To stabilize the pressure, a supremizer enrichment strategy of the velocity space is employed, as described in detail by Ballarin et al. [173]. The entire procedure is implemented using ITHACA-FV, a C++ library developed in the OpenFOAM framework (Stabile and Rozza [163]).

The combination of POD and Galerkin projections offers an effective approach for constructing ROMs, enabling the efficient analysis and prediction of flow behavior while significantly reducing computational costs.



(a) u mean

(b) v mean

(c) 1st POD u mode

(d) 1st POD v mode

(e) 2nd POD u mode

(f) 2nd POD v mode

Figure 4.22. Mean flow and leading POD modes for a NACA 0012 at $\alpha = 8°$ and $Re_\infty = 7000$. All field variables have been normalized with respect to their maximum.

4.3.3 Results

The methodologies described above have been applied to 2D numerical data of an unsteady, incompressible flow past a NACA0012 airfoil. These data have provided valuable insights into the flow dynamics and the performance of the ROMs. For both the POD-based and CNM-ROM approaches, a total of 400 snapshots have been considered for the training stage, with a sampling time of $\Delta t = 0.025$ s.

Panels (a) and (b) of Figure 4.22 present contours of the time-averaged streamwise velocity component u and transversal velocity component v, respectively. These contours provide a visual representation of the mean flow behavior. Panels (c) to (f) display the leading velocity Proper Or-

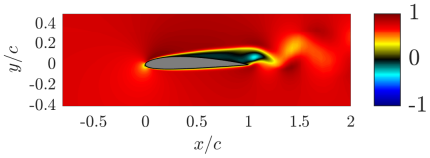
thogonal Decomposition (POD) modes, with the first two modes capturing the most dominant fluctuations in the flow field. It is noteworthy that the spatial shift along the advection direction depicted in the POD modes corresponds to the traveling pattern of the emitted eddies.

Figure 4.23 compares the velocity contours obtained from the FOM, the POD-ROM (considering 10 velocity modes and 12 pressure and viscosity modes), and the CNM-ROM (with 10 clusters and a 4th order ROM) at $t = 1$ s. This comparison allows us to assess the accuracy and fidelity of both the reduced-order models in capturing the flow features. The reconstructed flow fields obtained from both the POD-ROM and CNM-ROM are in good agreement with the original FOM results, indicating the effectiveness of the reduced-order models in capturing the essential flow dynamics.

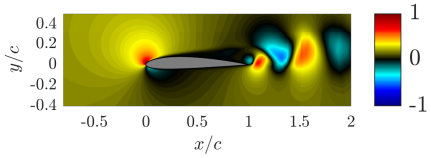
Figure 4.24 provides a comprehensive comparison of the lift and drag coefficients (C_l and C_d) obtained from the FOM, the POD-ROM, and the CNM-ROM at various ROM parameters. This comparison allows us to evaluate the accuracy and predictive capability of the reduced-order models in capturing the aerodynamic forces acting on the airfoil. The black lines represent the lift and drag coefficients obtained from the FOM simulations. The red lines correspond to the lift and drag coefficients obtained from the POD-ROM, which utilizes 10 velocity modes and 12 pressure and viscosity modes. The blue lines correspond to the lift and drag coefficients obtained from a 2nd order CNM-ROM, while the red lines correspond to a 4th order CNM-ROM.

This comparison enables us to assess the performance of the reduced-order models in replicating the aerodynamic forces evaluated in the full-scale simulations. The figure clearly demonstrates the capabilities of the CNM-ROM technique in accurately predicting the lift and drag coefficients, surpassing the performance of the POD-ROM. Therefore the CNM-ROM approach (with a higher order) captures the complex flow phenomena more effectively, resulting in improved predictive accuracy of the aerodynamic forces.

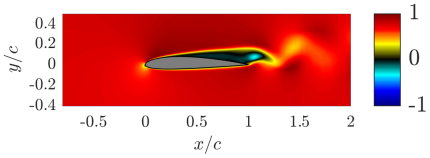
To further analyze the statistical properties of the flow, the unbiased autocorrelation function $R(\tau)$ (4.42) between the FOM, POD-ROM, and



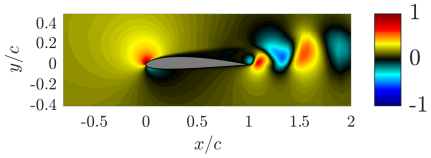
(a) u snapshot from FOM



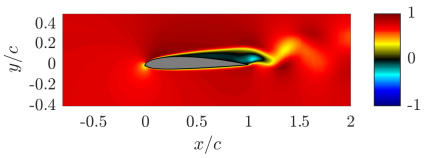
(b) v snapshot from FOM



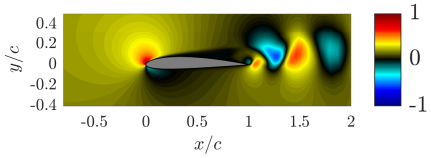
(c) POD-ROM u snapshot



(d) POD-ROM v snapshot

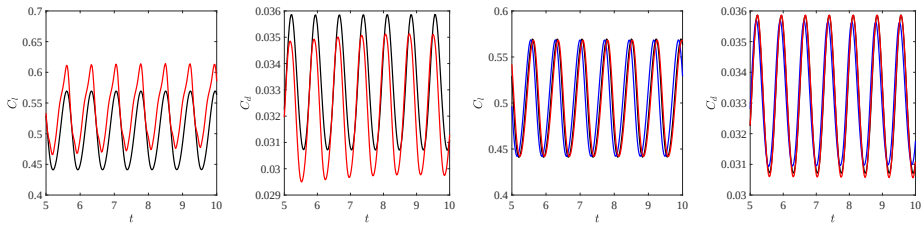


(e) CNM-ROM u snapshot



(f) CNM-ROM v snapshot

Figure 4.23. Comparison between FOM, POD-ROM, and CNM-ROM snapshots. All field variables have been normalized with respect to their maximum.



(a) C_l POD-ROM **(b)** C_d POD-ROM **(c)** C_l CNM-ROM **(d)** C_d CNM-ROM

Figure 4.24. Comparison between FOM, POD-ROM, and CNM-ROM force coefficients. The black lines represent the FOM force coefficients, the red lines refer to the POD-ROM with 10 velocity modes and 12 pressure and viscosity modes; the blue lines refer to a 2nd order CNM-ROM, while the red ones correspond to a 4th order CNM-ROM.

CNM-ROM approaches, is shown in Figure 4.25.

Figure 4.25 clearly demonstrates that the ROM approaches perform quite well in capturing the statistical properties and temporal correlations of the flow. However, the CNM-ROM exhibits a closer match to the auto-correlation function obtained from the FOM data, indicating its superior capability in reproducing the flow statistics.

The CNM-ROM approach proves its versatility by allowing predictions of the system behavior under different parameter and control settings. Figure 4.26 presents the C_l (panel (a)) and C_d (panel (b)) predictions for $\alpha = 6°$ by training the CNM-ROM with snapshots from different α values, excluding $\alpha = 6°$. This demonstrates the CNM-ROM's ability to extrapolate and generalize the learned dynamics to unseen parameter settings.

In Figure 4.26, the red and blue lines correspond to $\alpha = 5°$ and $\alpha = 7°$, respectively. The black dashed lines represent the lift and drag coefficients obtained from the FOM simulations, while the black solid lines represent the CNM-ROM predictions. Remarkably, the CNM-ROM accurately captures both the amplitude and frequency of the force coefficients for varying α values, including the unseen $\alpha = 6°$. This showcases the robustness

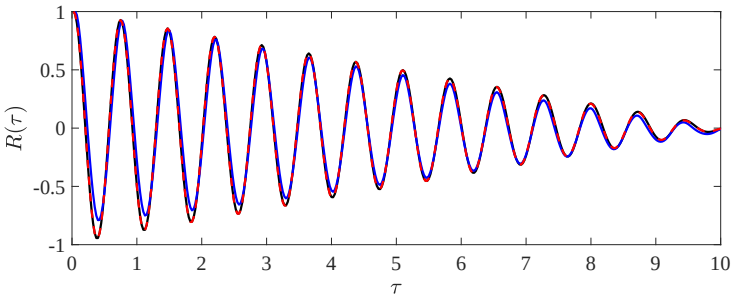


Figure 4.25. Comparison of the autocorrelation function R between the FOM, POD-ROM, and CNM-ROM. The black line represents the FOM, the blue line represents the POD-ROM, and the red dashed line corresponds to the CNM-ROM.

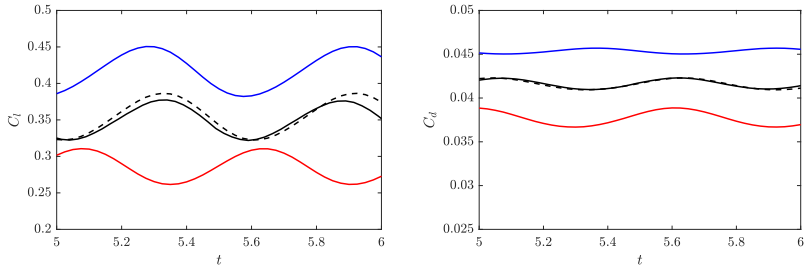
and generalization capability of the CNM-ROM, as it effectively predicts the aerodynamic forces even for parameter values that were not included in its training data.

Overall, the results presented in this section validate the effectiveness of the POD-ROM and CNM-ROM approaches in accurately capturing the flow dynamics, reproducing the flow fields, predicting force coefficients, and capturing the statistical properties of the flow. These findings highlight the potential of reduced-order modeling techniques in enhancing computational efficiency and enabling rapid design iterations for airfoil configurations.

4.4 Trajectory based CNM

The standard Galerkin-ROM techniques and previous CNM approaches consider a state vector composed of the realizations of flow field variables. Here we propose a new approach considering a state vector composed of trajectories of the dynamical system.

The basic idea behind trajectory-based CNMs is to cluster piecewise trajectories from a given dataset, dealing with data in terms of curves or



(a) C_l prediction **(b)** C_d prediction

Figure 4.26. Prediction of the effect of α using CNM-ROM. The red and blue lines correspond to $\alpha = 5°$ and $\alpha = 7°$, respectively. The black dashed lines represent the FOM simulations, while the black solid lines represent the CNM-ROM predictions.

functions. Using functional approaches, data clustering can be performed more efficiently and effectively, as the shape and features of trajectories can be more accurately represented by curves. Compared to standard CNM, in which snapshots order does not influences the clustering, this allows a more meaningful and informative clustering of trajectories, leading to improved insights and decision-making.

In order to understand the aim of trajectories clustering, in figure 4.27 is reported a functional clustering sketch. An ensemble of n two-dimensional trajectories presenting noise are analyzed, in figure are reported in grey. In standard clustering approach, a centroid would be a two-dimensional vector, and in figure are reported 3 centroids in black. In the functional framework the clusters are no more state vector realizations but function-s/trajectories.

While working with functional data one could assume that observations exist in an infinite-dimensional space, but, actually, we only have samples that are observed in a finite number of points. It is common to work with discrete observations $\mathbf{u}_{i,j}$ of each sample path $\mathbf{u}_i(t)$ in a finite set of knots, j. Therefore, the first step in functional data analysis frequently involves reconstructing the functional shape of the data from discrete ob-

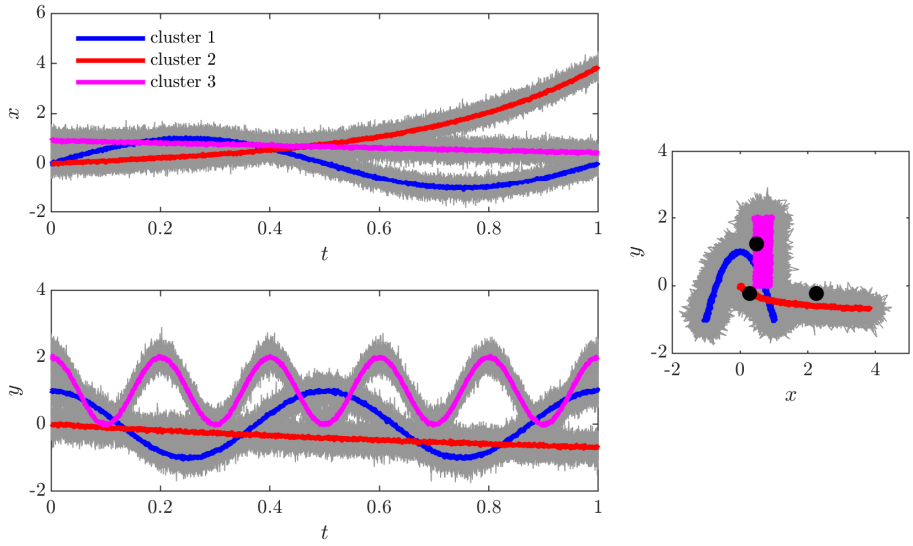


Figure 4.27. Sketch of functional clustering on an ensemble of n two-dimensional trajectories presenting noise. Black marker denotes standard centroids. Coloured lines trajectories clusters.

servations. Depending on the shape reconstruction method, functional data analysis can be broadly classified into three categories: raw data approaches, filtering approaches and adaptive methods. Raw data analysis involve analysing the functional data without any preprocessing or transformation; this approach is suitable for data that are relatively smooth and continuous over the entire domain. In the filtering approach, the curve shape reconstruction is addressed by assuming that the sample paths belong to a finite-dimensional space that is generated by a basis of functions. Finally, adaptive methods combine the previous ones depending on the data local variability.

The first step of the analysis involve the segmentation of the original d dimensional time series denoted with $\mathbf{u}(t)$, where t represents the time defined in the interval $t \in [0, T]$. The dimensions d corresponds to the number of variables times the number of points of the spatial discretization. Let $\mathbf{u}_i(t)$ be the $i-th$ segment of the time series, where $i = 1, 2, ..., N$. The segments can be overlapping between each other. It is possible to consider $\mathbf{u}_1(t'), \ldots, \mathbf{u}_N(t')$ to be a sample of N trajectories, with $t' = t - t_{0i}$ and defined in $t' \in [0, T_t]$. In this domain $\mathbf{u}_i(t')$ is further discretized into n_i points giving the discrete observations $\mathbf{u}_{i,j}$. T_t is chosen to be proportional to τ^*, with τ^* being the first time instance in which $R(\tau) = 0$ (4.42). The foundation of the analysis is that the data $\mathbf{u}_i(t')$, in the framework of Functional principal component analysis (FPCA), can be approximated as a linear combination of N basis functions, $\phi_1(t'), \ldots, \phi_N(t')$, such that

$$\mathbf{u}_i(t') \approx \boldsymbol{\mu}(t') + \sum_{k=1}^{N} c_{ik} \boldsymbol{\phi}_k(t'), \qquad i = 1, \ldots, N, \tag{4.47}$$

where $\boldsymbol{\mu}(t') = E(\mathbf{u}_i(t'))$ is estimated by $\boldsymbol{\mu}(t') = \frac{1}{n} \sum_{i=1}^{N} \mathbf{u}_i(t')$ and represents the mean trajectory observed. Coefficients c_{ik} are the principal component scores. It is worth to notice that at this stage $\boldsymbol{\mu}(t')$ and $\boldsymbol{\phi}_k(t')$ are d-dimensional. The assumption (4.47) implies that the functional data lies in a finite dimensional subspace spanned by the $\boldsymbol{\phi}_k$ basis functions.

Functions $\boldsymbol{\phi}_k$ can be chosen *a priori* (filtering approaches) and based on raw data by means of Functional principal component analysis. Appendix

B.1 reports the description of FPCA analysis.

The truncation of the series (4.47) to the order $P < N$

$$\tilde{\mathbf{u}}_i(t') = \boldsymbol{\mu}(t') + \sum_{k=1}^{P} c_{ik}\boldsymbol{\phi}_k(t'), \tag{4.48}$$

leads to a representation of the trajectories ensemble containing the essential features of the data. Once piecewise trajectories have been reconstructed it is possible to merge the segments to reconstruct the original data. Outside their domain $t' = t - t_{0i} \in [0, T_t]$, the $\tilde{\mathbf{u}}_i(t')$ can be either zero-padded or periodically extended. The original data can be reconstructed with the introduction of the weights $w_i(t)$:

$$\tilde{\mathbf{u}}(t) = \sum_{i=1}^{N} w_i(t)\tilde{\mathbf{u}}_i(t). \tag{4.49}$$

The inspection of the Eq. (4.49) reveals that $w_i(t)$ must satisfy the condition

$$\sum_{i=1}^{N} w_i(t) = 1, \qquad \forall t. \tag{4.50}$$

In particular the Eq. (4.50) simply comes from the fact that Eq. (4.49) must be valid also for a perfect reconstruction involving not only the first P eigenfunctions but the whole basis set $\boldsymbol{\phi}_k$, namely when $\tilde{\mathbf{u}}_i \equiv \mathbf{u}_i$.

4.4.1 Functions clustering

The goal of functions clustering is to identify groups of functions that share similar characteristics, such as shape, amplitude, or frequency (Jacques and Preda [174]). The first step to cluster functions consists in the definition of a metric to quantify the similarity or dissimilarity between pairs of functions. The choice of the distance metric can have a significant impact on the results of the clustering analysis; different data types or specific research goals can require different metrics.

A very common choice considers the Euclidean distance between two d-dimensional functions $f(t)$ and $g(t)$. It is defined as:

$$d_E^2(f,g) = \int_a^b (f(t) - g(t))^T (f(t) - g(t)) dt. \tag{4.51}$$

This metric calculates the root-mean-square difference between the two functions over the specified domain [a,b]. However, the Euclidean distance is generally used for functions with continuous domains, and can be sensitive to outliers (Scherl et al. [37]). To overcome this issue and deal with discontinuous domains or sparse data the Manhattan distance can be employed

$$d_M(f,g) = \int_a^b \|f(t) - g(t)\|_1 dt. \tag{4.52}$$

Once the distance matrix between all pairs of functions has been computed, it is possible to use standard clustering algorithms to group the functions together. Most popular choices can be the hierarchical clustering (Nielsen [175]), the spectral clustering (von Luxburg [176]) or the K-means clustering (MacQueen [131]). Here, the *K-means++* algorithm (Arthur and Vassilvitskii [177]) has been employed.

4.4.2 Filtering approaches

Filtering methods are a popular class of techniques used in functional data clustering to pre-process the raw data and reduce its complexity. In particular they leverage on the fact that the functional data $\mathbf{u}(t)$, and the trajectories segments $\mathbf{u}_i(t')$, lie themselves on a finite dimensional space spanned by some temporal basis of functions,

$$\mathbf{u}_i(t') = \sum_{j=1}^{P} \boldsymbol{\alpha}_{ij} f_j(t'), \tag{4.53}$$

with $\boldsymbol{\alpha}_{ik}$ being d-dimensional.

Depending on the choice of the f_k, there are several types of filtering

methods used in functional data clustering, including smoothing (Ramsay and Silverman [178]), B-splines (Abraham et al. [179]), wavelet-based methods (Giacofci et al. [180]), Fourier-based methods, and others. Each of these methods has its own strengths and weaknesses and is suited for specific types of functional data.

Short time Fourier transform

The short time Fourier transform (STFT) based method is particularly suited for periodic and quasi-periodic datasets. In appendix B.2.1 STFT is introduced and described. In the STFT framework, referring to the Eq. (4.53), each i^{th} trajectory is modelled as a superposition of complex exponentials $e^{\mathrm{i}2\pi ft}$.

$$\mathbf{u}_i(t) = \sum_{k=1}^{n_f} \hat{\boldsymbol{\alpha}}_{ik} e^{\mathrm{i}2\pi f_k t} \,, \tag{4.54}$$

where n_f is the number of frequencies.

Complex valued coefficients $\hat{\boldsymbol{\alpha}}_{ik}$ (d-dimensional) are the discrete STFT component and are computed according to the Eq. (B.9). It is worth to notice that for the STFT case, the weights w for the signal reconstruction are already contained in the definition of $\hat{\boldsymbol{\alpha}}_{ik}$, see Eq. (B.9), and the reconstruction of the original data simply consists in the sum of the local trajectories as in Eq. (4.54):

$$\mathbf{u}(t) = \sum_{i=1}^{N} \mathbf{u}_i(t). \tag{4.55}$$

For the clustering procedures the Euclidean distance between two trajectories i, j can be defined as:

$$d_{ij}^2 = \sum_{k=1}^{n_f} (\boldsymbol{\xi}_{ik} - \boldsymbol{\xi}_{jk})^T (\boldsymbol{\xi}_{ik} - \boldsymbol{\xi}_{jk}), \tag{4.56}$$

where $\boldsymbol{\xi}_{ik}$ is

$$\boldsymbol{\xi}_{ik} = [\mathrm{Re}(\hat{\alpha}_{ik}); \mathrm{Im}(\hat{\alpha}_{ik})]. \tag{4.57}$$

As the standard clustering algorithms deal with real-valued data, the introduction of $\boldsymbol{\xi}$ gives the possibility to employ such techniques for clustering complex-valued data.

B-splines

Functional data clustering using B-splines is a common approach for analyzing complex data sets where the observations are functions defined over a continuous domain. B-splines are a type of basis function that can be used to represent the functions in the data set, allowing for clustering based on the coefficients of the B-spline expansion. Each trajectory $\mathbf{u}_i(t')$ is represented as

$$\mathbf{u}_i(t') = \sum_{j=0}^{p} \boldsymbol{s}_{ij} B_j(t'), \tag{4.58}$$

where $B_j(t')$ is the B-spline orthonormal basis and p the order. Further details on the derivation of the $\boldsymbol{s}_{ij}$ and $B_j(t')$ can be found in appendix B.2.2 in which B-splines fitting is introduced and described.

The most commonly used distance metric for this approach is the squared L_2 norm (Euclidean), which measures the difference between two functions as the sum of the squared differences between their corresponding B-spline coefficients c_{ij}:

$$d_{ij}^2 = \sum_{k=1}^{p} (\boldsymbol{s}_{ik} - \boldsymbol{s}_{jk})^T (\boldsymbol{s}_{ik} - \boldsymbol{s}_{jk}) \tag{4.59}$$

This distance metric captures the difference between the shape of the two functions, regardless of their magnitude or location.

Wavelets

In functional data clustering, wavelets can be used as a powerful tool for data analysis, particularly for non-stationary signals and data with sharp changes or discontinuities. The wavelet transform is a mathematical technique used to decompose a signal into its frequency components while maintaining the time location of these components. In this framework, each function $\mathbf{u}_i(t')$ can be represented as

$$\mathbf{u}_i(t') = \sum_{j=-N_{\text{lev}}}^{N_{\text{lev}}} \sum_{k=-N_{\text{tra}}}^{N_{\text{tra}}} \boldsymbol{\beta}_{ijk} \psi \left(\frac{t' - 2^j k}{2^j} \right), \tag{4.60}$$

where $\boldsymbol{\beta}_{ijk}$ are the coefficients of the wavelet expansion, N_{lev} is the number of scales levels, N_{tra} is the number of the translations, and $\psi(t')$ is the mother wavelet function. For a detailed derivation of the wavelet transform, please refer to appendix B.2.3.

As with Fourier and B-spline methods, the squared L_2 norm (Euclidean distance) is often used as the distance metric, which measures the difference between two functions as the sum of the squared differences between their corresponding wavelet coefficients $\boldsymbol{\beta}_{ijk}$:

$$d_{ij}^2 = \sum_{j=-N_{\text{lev}}}^{N_{\text{lev}}} \sum_{k=-N_{\text{tra}}}^{N_{\text{tra}}} (\boldsymbol{\beta}_{ijk} - \boldsymbol{\beta}_{jik})^T (\boldsymbol{\beta}_{ijk} - \boldsymbol{\beta}_{jik}). \tag{4.61}$$

This distance metric quantifies the difference in the frequency content and temporal location of the features of the two functions, making it particularly suitable for non-stationary and discontinuous signals. The clustering algorithm is then applied to the wavelet coefficients, which capture the most important features of the data. This approach is advantageous in many applications where the raw functional data are too complex to be directly analyzed, and the key features of the data can be effectively captured by the wavelet transform.

However, the choice of the mother wavelet function can have a significant effect on the clustering results, making it crucial to select a wavelet function that is appropriate for the analyzed data. In the present study,

the Daubechies wavelets of order 10 were used, as discussed in appendix B.2.3.

4.4.3 Trajectories network model

As already said, in this framework the clusters are no more state vector realizations but functions/trajectories. The result of the clustering procedure is a set of K functional clusters, denoted as $\tilde{\mathbf{c}}_k(t')$ with $k = 1, \ldots, K$. This approach offers a richer and more dynamic perspective on the system's evolution, capturing not only the instantaneous state but also the temporal trends and patterns over a specific interval.

Similarly to the standard CNM, in the trajectory-based Cluster Network Model the transitions between these clusters or trajectories are modelled as a directed network, as sketched in Figure 4.28. The calculations of transition probabilities Q_{ij} and mean transition times T_{ij} between clusters follow a similar concept as in standard CNM. Q_{ij} is calulated as the number of transitions from centroid $\tilde{\mathbf{c}}_j$ to $\tilde{\mathbf{c}}_i$ over the total number transitions departing from $\tilde{\mathbf{c}}_j$. The mean transition times T_{ij} are the average time that the system spends in one trajectory-cluster before transitioning to another.

According to Q_{ij} and T_{ij}, a visitation sequence can be created, denoted as $\tilde{\mathbf{c}}_{k_i}$. The signal segments can be reconstructed by means of:

$$\tilde{\mathbf{u}}_i(t) = \sum_{i=1}^{N} \alpha_j(t) \tilde{\mathbf{c}}_{k_j}(t), \tag{4.62}$$

where $\alpha_j(t)$ can be linear interpolation coefficients or spline ones for a higher smoothness. The total time series can be reconstructed by means of (4.49). It must be noticed that reconstruction (4.62) may create discrete jumps at the trajectories overlap times. In analogy to the standard CNM, this procedure can also be conveniently extended to an order L. This involves considering transition probabilities that depend on the preceding L trajectory clusters, expressed as $Pr(\tilde{\mathbf{c}}_{k_{n+1}} = i | \tilde{\mathbf{c}}_{k_n}, ..., \tilde{\mathbf{c}}_{k_{n-L}})$. The actual flow computations are based on a lossless FPCA, as elaborated in appendix B.1.

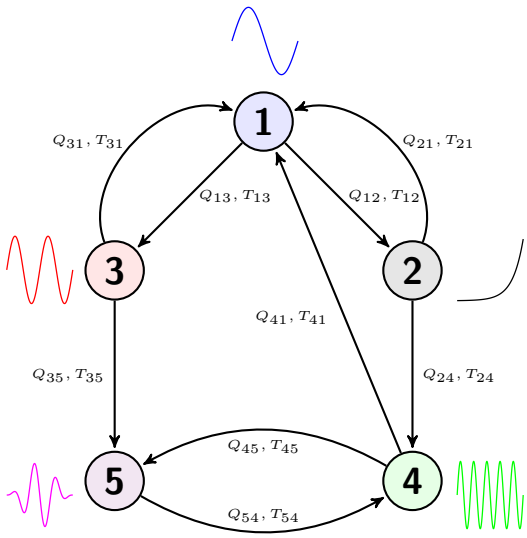


Figure 4.28. Schematic representation of a directed network of clusters in the trajectory-based CNM. The directed network represents the potential transitions between different trajectory-clusters, each symbolizing a distinct temporal behavior of the system. The small curves drawn next to each node illustrate the general temporal behaviour of each cluster. The arrows between the clusters are labeled with the corresponding transition probabilities Q_{ij} and mean transition times T_{ij}.

4.4.4 Validation

The performance of the model is evaluated using several metrics: the autocorrelation function $R(\tau)$, its root mean square error (RMSE), the asymptotic cluster probability, and the representation error.

The autocorrelation function, defined in Equation (4.42), is employed to avoid issues of comparing two trajectories directly due to phase mismatch (Hou et al. [39], Fernex et al. [77]). The RMSE of the autocorrelation function is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_R}\sum_{n=1}^{N_R}\Big(R(\tau_n) - \hat{R}(\tau_n)\Big)^2}, \tag{4.63}$$

where $\hat{R}$ is the autocorrelation of the modesl and N_R is the number of time lags τ_n (fernex2020).

The asymptotic probability p_i^∞ to be in a cluster i can be estimated with

$$p_i^\infty = \frac{\sum \tau_i}{T_0}, \tag{4.64}$$

where T_0 being a sufficiently long time horizon simulated by the model, and $\sum \tau_i$ being the cumulative residence time from (4.17). The vector containing al the p_i^∞ indicates whether the predicted trajectories could populate the phase space similarly to the original data, thus providing insight into the transition error of the modeling method (Hou et al. [39]).

The representation error of the reconstructed dynamics, denoted by E_r, is defined as:

$$E_r = \frac{1}{M}\sum_{m=1}^{M} D_{\mathscr{T}}^m, \tag{4.65}$$

where $D_{\mathscr{T}}^m$ is defined as the minimal distance from the snapshot $\mathbf{u}^m$ to all the states $\mathbf{u}^n$ of the reconstructed trajectory $\mathscr{T}$:

$$D_{\mathscr{T}}^m = \min_{\mathbf{u}^n \in \mathscr{T}} \|\mathbf{u}^m - \mathbf{u}^n\|_\Omega . \tag{4.66}$$

This parameter gives an indication of how well the model reconstructs the dynamics.

A key parameter for the model is the number of clusters K. A model with a small number of clusters can effectively capture the dominant transition dynamics behaviour. However, larger representation errors are induced as fewer clusters represent the dynamics and the snapshot details are eliminated. Conversely, a large K allows the CNM to model the transition dynamics with more detail, but the transition relationships become more complex and often meaningless (Li et al. [135]).

Nair et al. [8] proposed an optimal choice for the cluster number determined by the F-test (Hand [181]) which considers the ratio of inter-cluster variance to the total variance. In the framework of functional clusters $\tilde{\mathbf{c}}_i(t)$, the inner-cluster variance is similar to that of standard CNM, read

$$J = \frac{1}{N} \sum_{k=1}^{K} \sum_{\mathbf{u}_i \in \mathscr{C}_k} \int_0^{T_t} \|\mathbf{u}_i - \tilde{\mathbf{c}}_k\|^2 \, dt, \tag{4.67}$$

and the inter-cluster variance $\bar{J}$ can be defined as

$$\bar{J} = \frac{1}{N} \sum_{k=1}^{K} N_k \int_0^{T_t} \|\tilde{\mathbf{c}}_k - \bar{\mathbf{c}}\|^2 \, dt, \tag{4.68}$$

where N_k is the number of segments pertaining to the cluster k and $\bar{\mathbf{c}} \equiv \boldsymbol{\mu}(t)$. It is worth to notice that in the definitions (4.67) and (4.68), when using filtering methods the expansion coefficients $\boldsymbol{\alpha}_{ik}$ are employed. Except when otherwise specified, in this work the minimum number of clusters has been chosen such that $\bar{J}/(\bar{J} + J) > 0.9$, that corresponds to resolve at least 90% of the flow fluctuations after the cluster-coarse graining.

4.5 Functional CNM results

In this section the practical implementation and effectiveness of our method is reported, focusing on the comparison between standard CNM and trajectory based CNM, and presenting the outcomes from both synthetic and real-world scenarios.

In the first part, our method is applied to Landau's equation (Noack et al. [13], Luchtenburg et al. [182]) and a synthetic multi-scale signal. These carefully controlled synthetic environments provide a clear understanding of the method's purpose and behavior under known conditions.

The second part transitions to real-world fluid dynamic systems. Our methodologies are put to the test with force coefficients and flow fields past the fluidic pinball (Deng et al. [183, 184]), underlining the applicability and usefulness of the method in dealing with complex, practical situations.

4.5.1 Landau's system

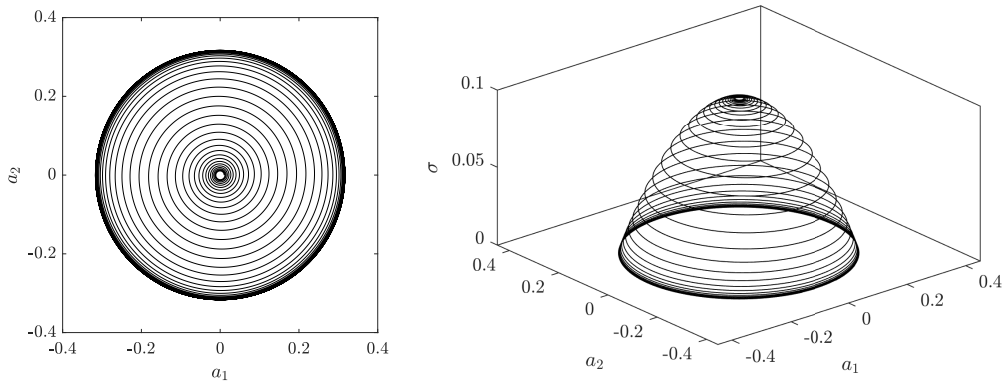
The Landau equations, serving as a model system, exhibit the rich dynamics characteristic of many real-world phenomena. They are utilized in this work as a "toy problem" to demonstrate the purpose and the capabilities of our trajectory-based Cluster Network Model. The equations are defined as:

$$\begin{aligned} \dot{a}_1 &= \sigma a_1 - \omega a_2, \\ \dot{a}_2 &= \sigma a_2 + \omega a_1, \end{aligned} \tag{4.69}$$

where a_1 and a_2 are the state variables, σ is a damping term computed as

$$\sigma = l - a_1^2 - a_2^2, \tag{4.70}$$

ω is a constant angular frequency and l parameter that influences the system behaviour. This system exhibits a range of rich dynamics, including sustained oscillations for certain parameter values, providing a comprehensive yet tractable testcase for validating our trajectory-based cluster network model. In this section ω has been set equal to π and $l = 0.1$. Considering the initial conditions $a_1(0) = 0.01$ and $a_2(0) = 0$ it is pos-

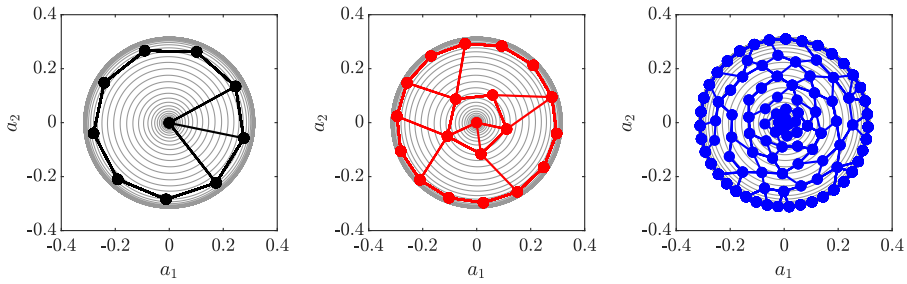


(a) **(b)**

Figure 4.29. (a) Phase space of the Landau equations with initial conditions $a_1 = 0.01$ and $a_2 = 0$, showing the trajectory spiraling out to a limit cycle. (b) Three-dimensional plot of the trajectory in the σ-a_1-a_2 space. It illustrates that as the system settles into the limit cycle, forming a paraboloid manifold, the variable σ approaches zero.

sible to highlight a distinctive feature of these equations: the trajectory converges to a limit cycle with a radius equal to $\sqrt{l}$, indicating the system's tendency to oscillate. Figure 4.29 provides a visual representation of this behavior. In panel (a), the phase space depicts the path of the trajectory toward the limit cycle. It is interesting to note how the system starts from nearly the origin and gradually spirals out to the limit cycle. Panel (b) showcases a three-dimensional plot of the variables σ, a_1, and a_2, forming a paraboloid manifold. This clearly indicates that the trajectory settles into the limit cycle as σ approaches zero, reinforcing the notion of the system entering a state of sustained oscillation.

After obtaining the data from the Landau system, the standard CNM has been applied on it. For this analysis, we select a model order of $L = 1$. 8000 snapshots have been collected with a $\Delta t = 0.01$. We explore the performance of the model with different number of clusters K, specifically, $K = 10$, $K = 20$, and $K = 100$. With $K = 10$, the intercluster variance



(a) $K = 10$ **(b)** $K = 20$ **(c)** $K = 100$

Figure 4.30. Phase space plots (a_1 versus a_2) with the true dynamics of the Landau system represented in grey. The model dynamics from the standard CNM with $K = 10$, $K = 20$, and $K = 100$ are overlaid in black, red, and blue respectively. Panel (a): $K = 10$, Panel (b): $K = 20$, Panel (c): $K = 100$. It can be observed that the characteristic behavior of amplitude selection is not well captured, even with the highest number of clusters ($K = 100$).

$\bar{J}/(\bar{J} + J)$ is approximately 90%.

In Figure 4.30, three panels display the phase space with the true dynamics represented in grey. The model dynamics with $K = 10$, $K = 20$, and $K = 100$ are overlaid in black, red, and blue respectively. The motion between cluster sequences (represented by continuous lines) is smoothed using Eq. (4.24). However, it is noteworthy that the standard CNM often leads to non-physical transitions, making it less reliable for accurately modelling the dynamics of the Landau's system. It is noticeable that even with the highest number of clusters ($K = 100$), the standard CNM struggles to adequately capture the characteristic behavior of amplitude selection in the Landau's equations.

Transitioning from standard CNM to trajectory-based CNM affords an improvement in capturing the Landau system's intrinsic dynamics, including transitions and amplitude selection mechanisms. Each trajectory segment is composed of $n_i = 200$ snapshots. For the spline basis, we choose a third-order spline, with comprehensive details available in Ap-

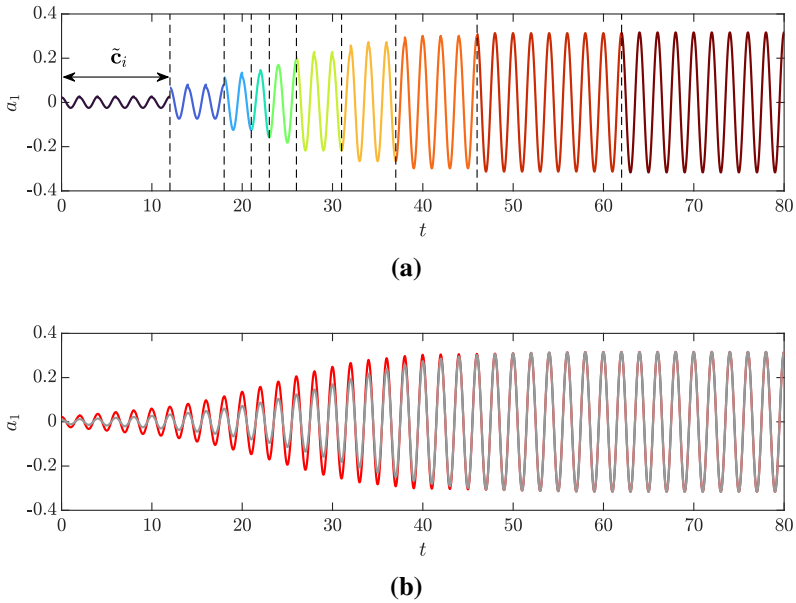


Figure 4.31. Demonstration of the Trajectory-based CNM with $K = 10$ trajectory clusters. Panel (a): Sequence of jumps among the trajectory cluster centroids (different colors) with the accompanying residence times denoted by vertical dashed lines. Panel (b): Comparison of the original dynamics (in gray) with the smoothed representation (in red) by means of Eq. (4.62).

pendix B.2.2. For the STFT basis, a rectangular window is selected, refer to Appendix B.2.1. The wavelet CNM employs a Daubechies wavelet, specifically 'db10', configured with 5 levels; for an in-depth discussion on this, refer to Appendix B.2.3.

Figure 4.31 demonstrates the proficiency of the trajectory CNM using spline basis functions and adopting 10 trajectory clusters. In Panel (a), the trajectory CNM coarse-graining is illustrated through a sequence of jumps between the trajectory cluster centroids, represented by distinct colors. These jumps are characterized by the residence times in each cluster, which are visualized through vertical dashed lines. To transition smoothly between the discrete jumps of trajectory segments, Eq. (4.62)

is utilized. As a result of this smoothing technique, Panel (b) shown the refined representation (in red) with the original data (in gray). Even with only 10 clusters, the trajectory CNM captures the underlying dynamics of the system.

In Figure 4.32, a three-dimensional plot provides a comparative perspective on the performance of different CNM variants in capturing the temporal dynamics of the Landau system. Panel (a) represents the standard CNM, while panels (b), (c), and (d) depict the spline CNM, STFT CNM, and Wavelets CNM respectively. For all these representations, a clusters number of $K = 10$ has been used. The true dynamics of the system are displayed in gray across all panels, providing a baseline for comparison. It becomes evident that the trajectory CNM, across its various adaptations, demonstrates superior capability in accurately capturing the temporal behavior and the amplitude selection mechanism of this system, as compared to the standard CNM.

The distinctions between the standard CNM and the trajectory-based CNMs are accentuated when examining Figure 4.33. In this figure, the representation error E_r , as defined in Eq. (4.65), is plotted. It becomes evident that the trajectory-based CNMs consistently yield a representation error that is at least an order of magnitude smaller compared to the standard CNM. Such a dramatic reduction in error underscores the efficacy and superiority of the trajectory-based approaches in capturing the inherent dynamics of the system.

To further underscore the strengths of the trajectory CNM, it is instructive to analyze the following time-modulated signal:

$$b_1 = A(t)\cos(2\pi f_1 t + \theta(t)) \tag{4.71}$$

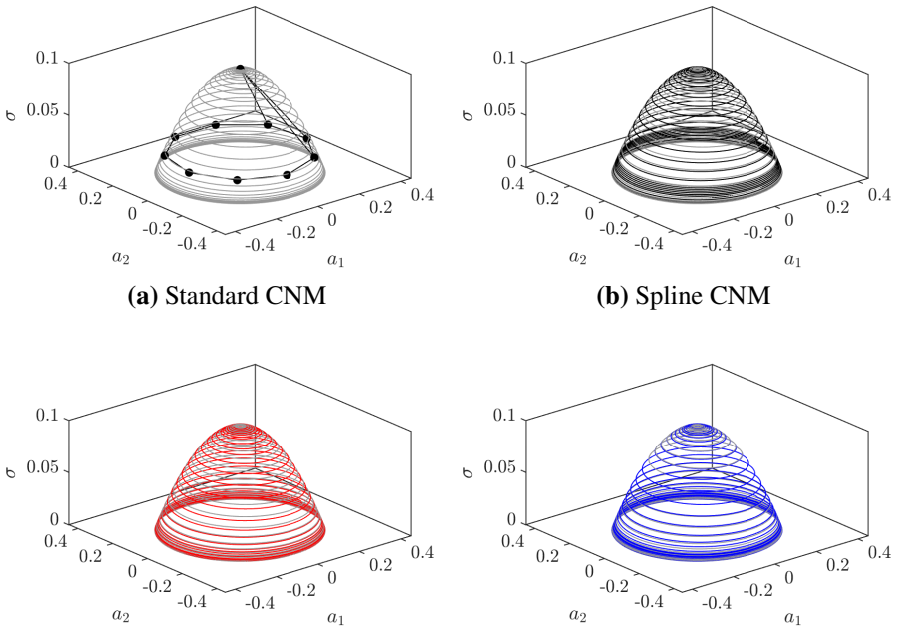
$$b_2 = A(t)\sin(2\pi f_1 t + \theta(t)) \tag{4.72}$$

where

$$A(t) = \cos(2\pi f_2 t), \tag{4.73}$$

$$\theta(t) = c_0 2\pi \cos(2\pi f_3 t) \tag{4.74}$$

and c_0 is a constant. The chosen parameters are $f_1 = 100$, $f_2 = 2$,



(a) Standard CNM

(b) Spline CNM

(c) STFT CNM

(d) Wavelets CNM

Figure 4.32. Comparison of different Cluster Network Model (CNM) variants in the three-dimensional space of a_1, a_2 and σ for the Landau's system. The true data is represented in gray across all panels. Panel (a) showcases the standard CNM (solid-point black), Panel (b) illustrates the spline CNM (black), Panel (c) demonstrates the STFT CNM (red), and Panel (d) displays the Wavelets CNM (blue). This visual representation underscores the enhanced capability of the Trajectory CNM in its various forms over the standard CNM.

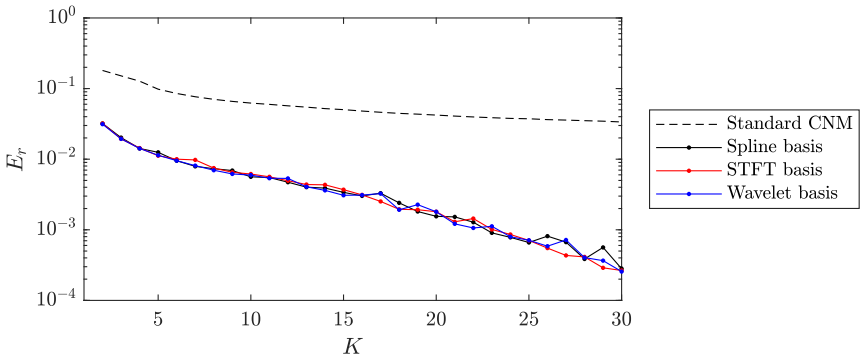


Figure 4.33. Comparison of representation error E_r for different Cluster Network Model (CNM) variants. The standard CNM error is represented by the dashed curve. The spline CNM is shown in black, the STFT CNM in red, and the Wavelets CNM in blue.

$f_3 = 0.25$. A number of samples $n_t = 10^6$ has been saved within a simulation time $t_{\text{fin}} = 100$, and with a $\Delta t = 10^{-4}$.

This particular signal is characterized by both frequency and amplitude modulation. It can be interpreted as a post-transient solution to the Landau's equations (4.69) if one assumes that the parameters ω, and l vary over time. Figure 4.34 showcases the phase space representation of this signal and the temporal evolution of $b_1(t)$.

Following the representation in Figure 4.34, we investigate the application of both the standard CNM and the trajectory-based CNM. For this analysis, the STFT-CNM is chosen. Both methods employ $K = 16$ clusters. For the STFT-CNM, segment length is set at $n_i = 5000$, with segment overlaps accounting for 75% of the segment length. Additionally, a standard Hamming window is employed, ensuring that the COLA (Constant Overlap-Add) constraint is satisfied. Figure 4.35 showcases the original b_1 data in black, the output from the standard CNM in blue, and the STFT-CNM results in red. Notably, while the standard CNM struggles to accurately capture the dynamics along with the amplitude and frequency modulation, the STFT-CNM, even with only $K = 16$, is indistinguishable from the original data.

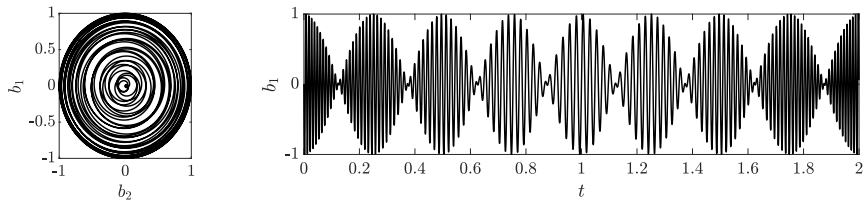


Figure 4.34. Analysis of the time-modulated signal. The phase space representation (left) illustrates the evolution of the system in the b_1 vs. b_2 plane. The right panel depicts the variation of the signal b_1 over time, revealing its temporal complexities. This signal offers insights into the behavior of the Landau's equations (4.69) with time-dependent parameters.

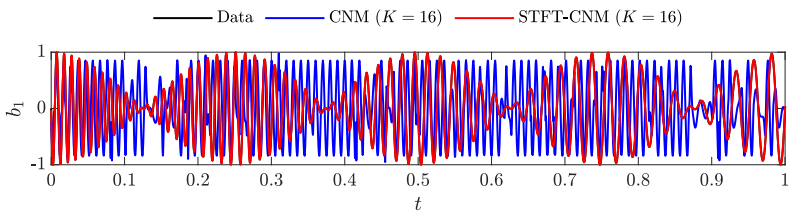


Figure 4.35. Comparison between original b_1 data (in black), standard CNM (in blue), and STFT-CNM (in red). Both methods employ $K = 16$ clusters.

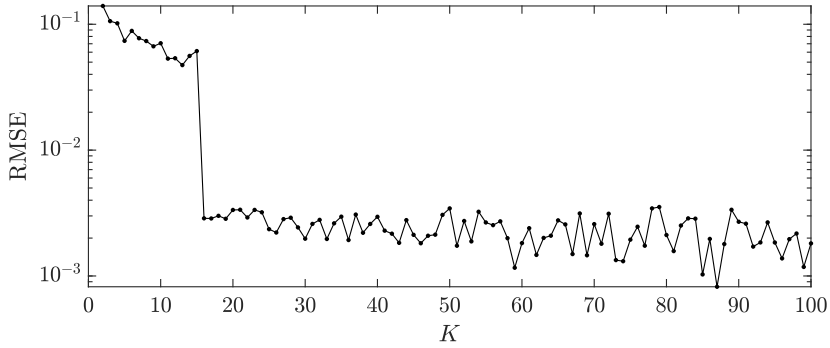
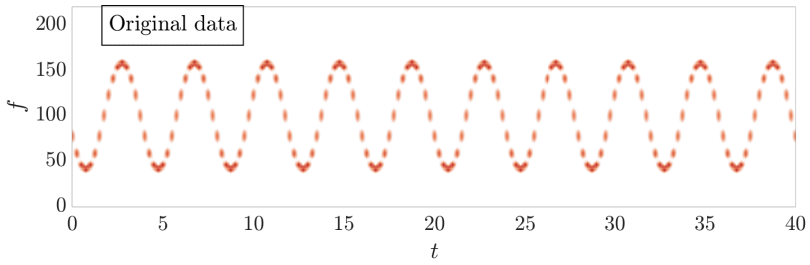


Figure 4.36. RMSE values vs K. The sharp decrease at $K = 16$ underscores its optimal choice for this analysis.

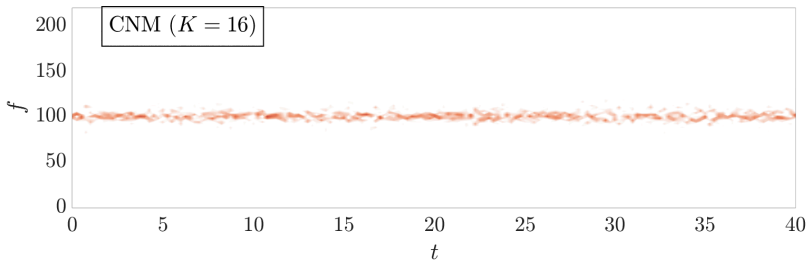
The selection of $K = 16$ is due to the Root Mean Square Error (RMSE) as defined in Eq. (4.63). Figure 4.36 presents the RMSE values across a range of K. It is evident that at $K = 16$, there's a significant drop in RMSE, which is then followed by a gradual decrease as K is further incremented.

As previously mentioned, the signal defined by Eqs. (4.71) and (4.72) exhibits frequency modulation over time. Insight into the temporal behavior of the signal can be obtained from its spectrogram, that is the amplitude of ξ_{ik}, where the index i denotes the frequencies and k spans time. Figure 4.37 depicts the spectrogram of the original b_1, compared to those from the standard CNM and the STFT CNM. Through this representation, it is evident that the STFT CNM perfectly captures the frequency modulation of the original signal, while the standard CNM is not able to achieve this purpose.

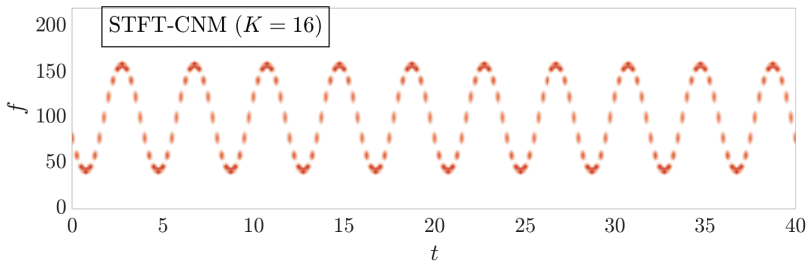
It is instructive to examinate the trajectory cluster centroids, $\tilde{\mathbf{c}}_i$. The spectral content intrinsic to each $\tilde{\mathbf{c}}_i$ can be discerned from its power spectral density (PSD). Figure 4.38 presents the PSD of each centroid, normalized with respect to its respective peak value. This representation clearly illustrates that each centroid is associated with a distinctive frequency. In this context, the centroids are presented in the sequence corresponding to



(a) Original data



(b) Standard CNM



(c) STFT CNM

Figure 4.37. Comparison of spectrograms for the original b_1 signal, panel (a), the standard CNM, panel (b), and the STFT CNM, panel (c).

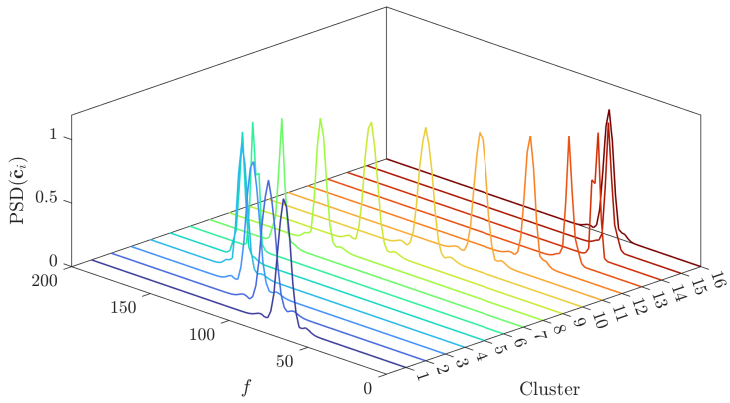


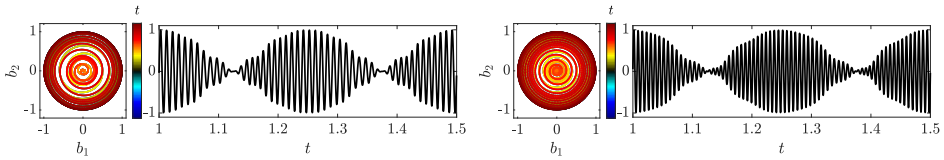
Figure 4.38. Normalized PSD of each trajectory cluster centroid $\mathrm{PSD}(\tilde{\mathbf{c}}_i)$.

their manifestation in the data.

The temporal dynamics within a given segmentation interval, as well as the phase space corresponding to the centroids $\tilde{\mathbf{c}}_i$, are depicted in Figure 4.39. Specifically, even-numbered centroids are showcased in panels spanning from (a) to (h). Each characteristic trajectory intrinsically encapsulates the amplitude modulation. This is due to the fact that within each segment, a full cycle of the amplitude oscillation is inherently contained. Moreover, the distinct frequency characteristics of these centroids can be readily observed. Transitioning between these clusters offers a coherent description that accurately reflects the frequency modulation in the original dataset.

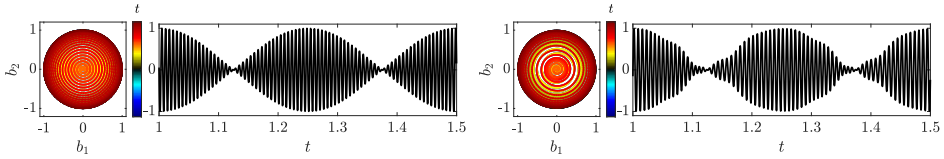
The frequency modulation's limit cycle-like behaviour is evident in the structure of the direct transition probability matrix Q. Figure 4.40 illustrates Q for the standard CNM in panel (a) and for the STFT-CNM in panel (b). Notably, the Q matrix in the STFT-CNM scenario showcases a shape that is typical of a limit cycle, as discussed in Hou et al. [39].

To provide deeper insights into the statistical properties of the original and modeled data, it is interesting to compare the unbiased autocorrelation function R for the two CNM methodologies. The function R is formally defined in (4.42). Figure 4.41 presents the normalized R on the left, with



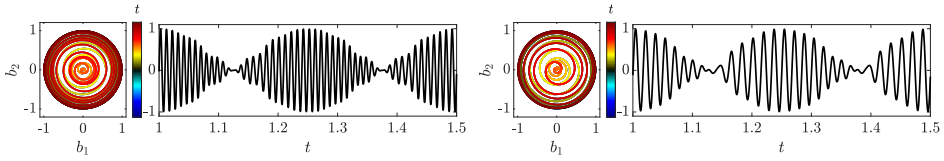
(a) $\tilde{\mathbf{c}}_i$: $i = 2$

(b) $\tilde{\mathbf{c}}_i$: $i = 4$



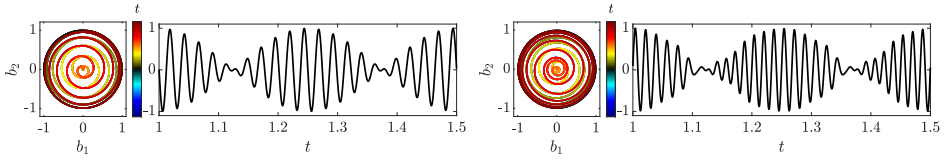
(c) $\tilde{\mathbf{c}}_i$: $i = 6$

(d) $\tilde{\mathbf{c}}_i$: $i = 8$



(e) $\tilde{\mathbf{c}}_i$: $i = 10$

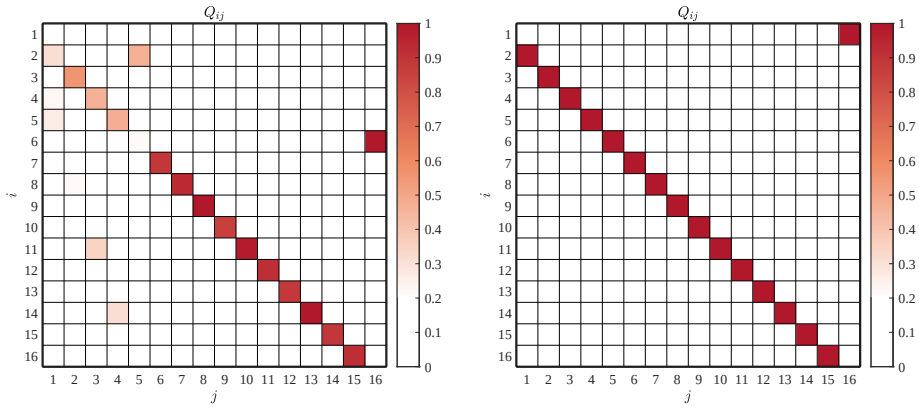
(f) $\tilde{\mathbf{c}}_i$: $i = 12$



(g) $\tilde{\mathbf{c}}_i$: $i = 14$

(h) $\tilde{\mathbf{c}}_i$: $i = 16$

Figure 4.39. Temporal dynamics and the corresponding phase space for even-numbered centroids $\tilde{\mathbf{c}}_i$. Panels from (a) to (h). The colorbar adjacent to the phase spaces indicates the progression of time.



(a) standard CNM **(b)** STFT-CNM

Figure 4.40. Direct transition probability matrix Q for standard CNM in panel (a) and STFT-CNM in panel (b).

a zoomed view targeting short time lags on the right. The visualizations represent the original data in black, the standard CNM in blue, and the STFT-CNM in red dashed lines. A close examination of this figure further emphasizes the superior ability of the trajectory-based CNM to effectively capture the complex dynamics inherent in the data.

Thus far, trajectory-based CNM has demonstrated its superiority in effectively capturing complex temporal behaviors exhibited by toy applications, like the Landau's problem and its associated amplitude and frequency modulation. In the subsequent sections, this innovative approach will be extended to a real-world fluid dynamic configuration, namely the fluidic pinball.

4.5.2 Fluidic Pinball

The fluidic pinball flow has recently emerged as an intriguing and efficient setup for investigating the application of machine learning in fluid flow control (Brunton et al. [185], Farzamnik et al. [186], Maceda et al.

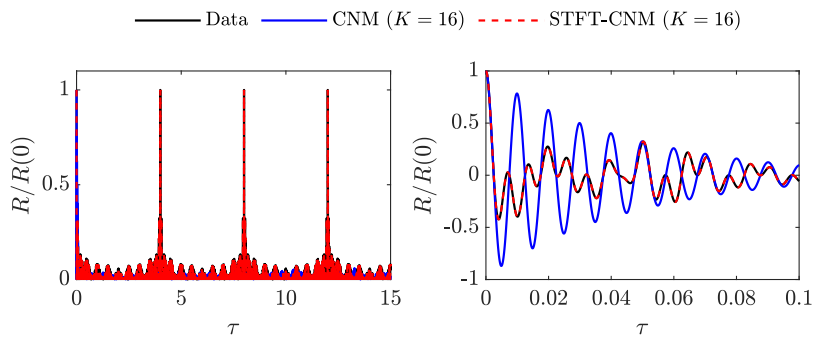
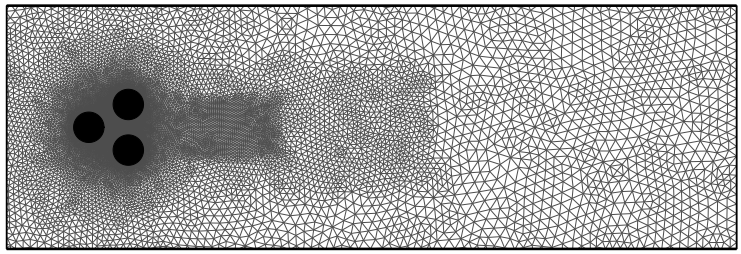


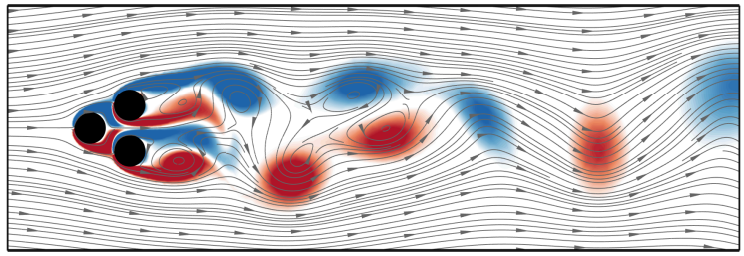
Figure 4.41. Comparison of the normalized unbiased autocorrelation function R: Original data in black, standard CNM in blue, and STFT-CNM in red dashed lines. The right panel provides a zoomed-in view for small time lags.

[187]). This system undergoes various flow patterns, starting with a stable flow and transitioning into a symmetric oscillatory flow characterized by post-Hopf bifurcation. It further evolves into asymmetric vortex shedding following the pitchfork bifurcation, eventually reaching a state of chaos, all by manipulating the Reynolds number (Deng et al. [183]). The system's geometric layout, as illustrated in Figure 4.42, comprises three cylinders with a diameter of D, positioned at the vertices of an equilateral triangle oriented upstream, with each side measuring 3D/2 in the (x, y) plane. The inlet of the domain is subjected to a uniform velocity U_∞ for the incoming flow, without any external forces acting on the cylinders throughout this study. We enforce a no-slip condition on the cylinders, and we assume that the velocity in the far wake region remains at U_∞. The Reynolds number, defined as $Re = U_\infty L/\nu$, where ν is the kinematic viscosity of the fluid. At the output region of the domain, a no-stress condition is applied.

The numerical data presented here has been generated using a solver based on a second-order finite-element discretization method of the Taylor–Hood type, as outlined by Taylor and Hood [188]. This solver operates on an unstructured grid consisting of 4225 triangles and 8633 vertices, employing implicit integration of the third order in time. The instantaneous

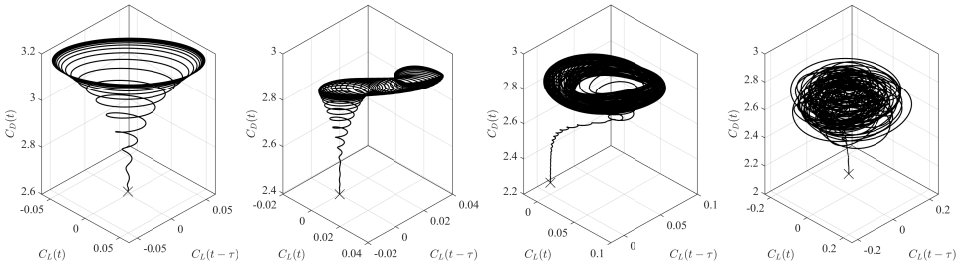


(a) Numerical grid



(b) Vorticity snapshot $Re = 160$

Figure 4.42. Numerical grid and a sample snapshot for fluidic pinball at $Re = 160$.



(a) $Re = 60$ **(b)** $Re = 90$ **(c)** $Re = 120$ **(d)** $Re = 160$

Figure 4.43. Phase portrait of force coefficients at different values of Re.

flow field is determined through a Newton–Raphson iteration process, persisting until the residual reaches a predefined, extremely small tolerance. A sketch of the grid and a sample snapshot are shown in Figure 4.42.

This approach is also utilized to compute the steady-state solution, which stems from the steady Navier–Stokes equations. The Navier–Stokes solver employed in this work has undergone validation in prior studies by Noack et al. [13] and Deng et al. [183], with comprehensive technical documentation available in Noack and Morzyński [189]. For further insights into the simulation methodology, refer to Deng et al. [183].

In subsequent sections the novel functional based CNM will be applied onto force coefficients and velocity fields of the fluidic pinball at different Re.

Force coefficients

The fluidic pinball system exhibits diverse behaviors depending on the Reynolds number (Re) value. In Figure 4.43, we present phase portraits illustrating the variations in force coefficients. The time delay for these representations is fixed at $\tau = 1/4f_c$, with f_c representing the characteristic frequency for each case. Panel (a) demonstrates the system's oscillatory behavior as it converges toward a limit cycle (Hopf bifurcation). In Panel (b), we observe the transition of the limit cycle into instability, marking a supercritical pitchfork bifurcation. Panels (c) and (d) display instances of

quasi-chaotic and fully chaotic motion, respectively (Deng et al. [184]).

Figure 4.44 presents a comparative analysis between the use of conventional CNM and functional-based CNM. In the case of STFT-CNM, a Hann window with a 75% overlap has been applied. The figure clearly illustrates the superior performance of functional-based CNM.

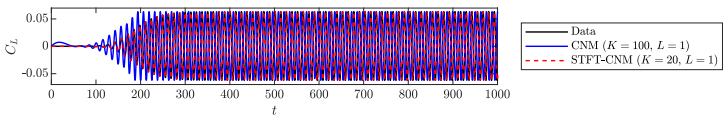
Flow fields

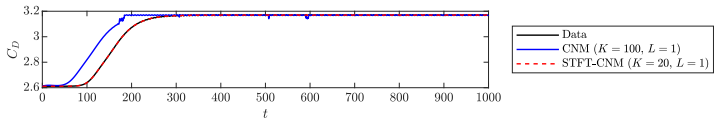
In this section, we utilize snapshots of both velocity components to construct both conventional CNM and functional-based CNM using Fourier basis. The comparison focuses on the configuration corresponding to $Re = 160$.

In Figure 4.45, we present spatial distributions of the leading POD modes in terms of vorticity. These visualizations allow us to identify both low-order shedding modes (modes 2-3) and higher-order structures in the wake region.

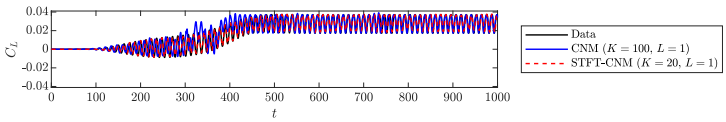
Figure 4.46 presents a side-by-side comparison of the results obtained using conventional CNM and functional-based CNM. When employing STFT-CNM, a Hann window with a 75% overlap is utilized. The figure demonstrates that functional-based CNM, even with a lower number of clusters, outperforms its conventional counterpart.

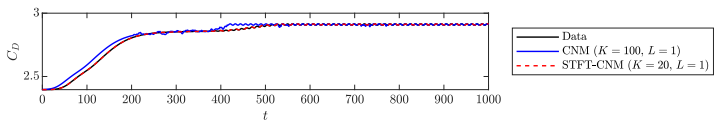
The new approach results mentioned in this section are encouraging for future developments of cluster-based reduced order modelling.



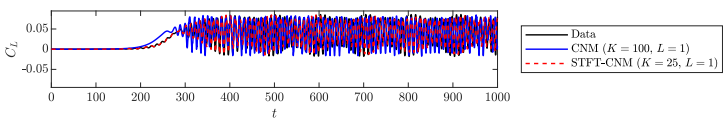


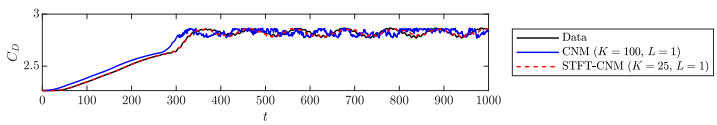
(a) $Re = 60$



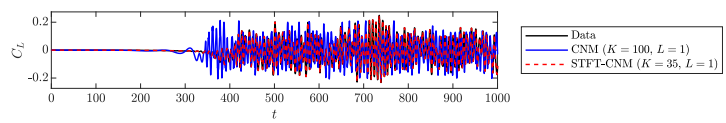


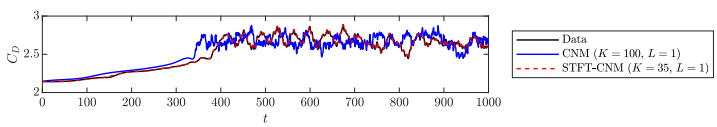
(b) $Re = 90$





(c) $Re = 120$





(d) $Re = 160$

Figure 4.44. Comparison between standard CNM and functional based CNM with Fourier basis.

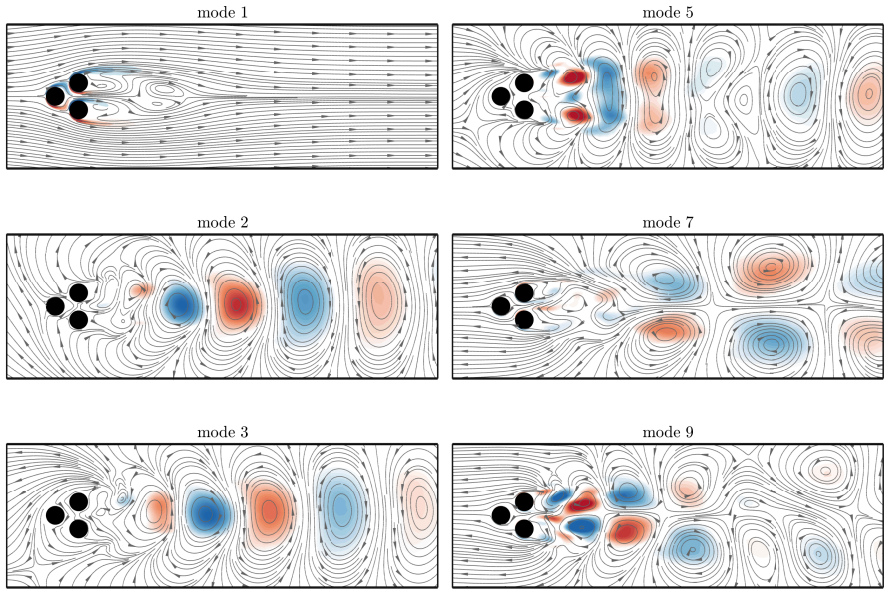


Figure 4.45. Leading POD modes in terms of vorticity for the post-transient solution of the fluidic pinball at $Re = 160$. All the field variables have been normalized with respect to its maximum.

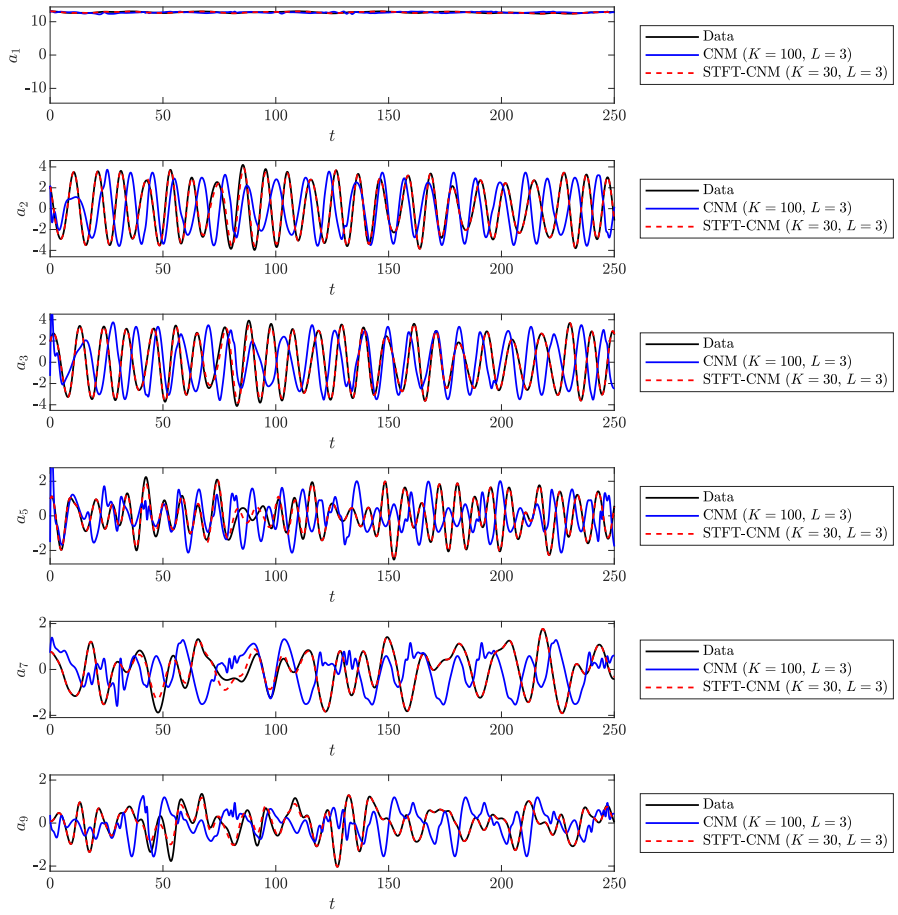


Figure 4.46. Comparison between the reconstructions of leading POD temporal coefficients using conventional CNM and functional-based CNM with Fourier basis.

Chapter 5

Conclusions

The main goals of the work developed in this thesis is to employ and develop data-driven approaches for modal analysis, stability analysis and reduced order modelling.

To accomplish these objectives, theoretical and numerical methodologies have been developed and employed over a wide range of fluid dynamics configurations.

This thesis has been developed within the Doctor Europaeus framework. Part of the work has been carried out in collaboration with prof. Francesco Grasso of Princeton University in the framework of modal analysis of two-phase flows. A novel approach for robust modal decomposition has been carried out in collaboration with prof. Oliver T. Schmidt of the University of California San Diego. The candidate has spent two months in Harbin Institute of Technology in Shenzhen (China), under the supervision of prof. Bernd R. Noack, developing a functional approach for cluster based network modelling. Furthermore, three months have been spent in Technische Universität Berlin (Berlin, Germany) under the supervision of prof. Kilian Oberleithner where the candidate has developed a novel clustering procedure that focuses on specific subspaces or subsets of the variables to extract the global behaviour of the system.

The main conclusions that can be drawn from this work are hereafter summarized, separately for the topics area, namely modal analysis, stability analysis and reduced order modelling. Recommendations for possible

future developments of the work are finally reported.

5.1 Modal analysis

In this work conventional modal analysis methods are employed to extract the main features of unsteady viscous liquid jet flows, both in 2D and 3D configurations. Innovative approaches for analysis of experimental data have been developed, namely the robust spectral proper orthogonal decomposition and the Gappy POD for two-phase PIV measurements.

A modal decomposition analysis, including both POD and SPOD techniques, has been performed on two-dimensional numerical data (obtained with the VOF code BASILISK) of vertical liquid sheets falling in still air, both in supercritical ($We > 1$) and subcritical ($We < 1$) regimes. The flow configurations analyzed in this study refer to unsteady sheet dynamics driven by a continuous harmonic perturbation in the lateral velocity component applied at the inlet section. POD and SPOD modes have been calculated for fluctuations of the streamwise and transverse velocity components and volume fraction. The investigation has been carried out at various Weber number, forcing frequencies and Reynolds numbers. In the supercritical regime, both POD and SPOD technique yield leading sinuous modes only and both exhibit a structure featuring a traveling perturbation. In the subcritical regime, when the system is forced at relatively high Reynolds number (approaching the inviscid limit), the analysis shows the excitation of a combined sinuous-varicose motion. The appearance of the varicose component is due to nonlinear coupling occurring in resonance conditions, i.e. when the sheet is driven by a frequency equal to the leading natural frequency and its emergence is favored by low Weber numbers.

A data-driven modal analysis of 3D gravitational liquid sheet direct numerical simulations has been carried out. The SPOD analysis of 3D gravitational liquid jet has highlighted the nucleation of three-dimensional asymmetric holes and their advection, characterizing the topology of the liquid sheet at low flow rate. The effects of the Weber number variation have been analyzed. It has been found that the leading frequency presents

a maximum in the region around $We \approx 0.75$ that is characterized by rapid holes formation and advection.

Data corruption is a significant obstacle to systems modeling and forecasting; noisy and gappy measurements can severely affect the learned models and lead to invalid conclusions. The ability to handle this type of data has a major role in the inference of physical insights from experimental investigations. These limitations are overcome by a novel approach, the robust spectral proper orthogonal decomposition which incorporates the robust principal component analysis within the SPOD algorithm. The investigation focused on evaluating the potential of this innovative method in improving the extraction of coherent structures from complex data-sets.

The effectiveness of robust SPOD was properly tested through two distinct fluid dynamics problems: the subsonic jet, investigated through numerical simulations, and the flow within an open cavity, studied by means of experimental measurements. Robust SPOD demonstrates superior performance by providing effectively converged and physically interpretable modes compared to those of classical SPOD. RSPOD was applied in the analysis of the open cavity flow, obtaining smooth modes, especially at high frequencies. This outcome is of substantial significance for applications where high-frequency information is critical. This new method not only addresses the challenges posed by corrupted data and outliers but also extends its utility to removing data noise, improving the reliability of results. The promising outcomes from both numerical and experimental data-set suggest that robust SPOD holds interesting potential for advancing the understanding of complex fluid dynamics phenomena and improving the robustness of modal analysis in various practical applications.

Finally, two-phase gappy PIV measurements of the two-phase mixing layer past a splitter plate have been analyzed by means of a tailored algorithm inspired by the Gappy POD, highlighting the leading structures and the characteristic frequencies of this experimental configuration.

5.2 Stability analysis

A data-driven approach, based on the Dynamic Mode Decomposition, has been employed to obtain 2D global (BiGlobal) spectra of gravitational viscous liquid sheet flows both in supercritical and subcritical regimes. The flow fields have been obtained by means of the BASILISK computer code, considering a one-fluid formulation and the volume-of-fluid approach. The procedure has been verified considering first simplified 1D equations, modeling the lateral displacement and the transversal velocity of the sheet. The estimated eigenmodes and eigenfunctions quickly convergence towards the standard ones of the linear stability theory, computed with the Chebyshev collocation method.

The BiGlobal spectrum for the supercritical case shows three different branches. The upper and the lower branches, already observed with the 1D model, are associated with the crossing time of slow and fast traveling waves, respectively; they exhibit a purely sinuous behavior with frequencies quite close to those predicted by the 1D model. The middle branch, instead, presents a predominant varicose component, increasing with the frequency, and a uniform energy distribution along the curtain. This last branch could not be predicted by the 1D model, which takes into account the curtain sinuous behavior only.

A parametric investigation has also been performed, to elucidate the impact of the Weber number and the density ratio between the two phases on the dynamics of the curtain. In the supercritical regime, for moderate values of We, the flow is stable and the leading eigenmodes are purely sinuous. When We increases, the varicose middle branch moves towards the upper one at high frequencies. The subcritical regime, instead, shows a leading mode with a frequency almost an order of magnitude higher than the previous ones, but still with a sinuous behavior; the second most unstable eigenvalue has a growth rate comparable with the leading one and a varicose behavior.

In the supercritical regime, increasing the density ratio, flow instabilities arise. The topology of the leading modes reveals how in the unstable configuration there is a superposition of both sinuous and vari-

cose regimes. By means of the inspection of the streamwise contribution to the total energy of the leading mode some insights on the disturbance propagation have been shown. In particular, moving downstream along the curtain there is a prominent shift from the sinuous shape to the varicose one, highlighting that the instability mechanism, which leads to the rupture of the sheet, is deeply related to the amplification of the varicose modes.

5.3 Reduced order modelling

In the framework of reduced order modelling several results and new methods have been carried out.

A novel approach, based on the Extended Cluster-based Network Modeling (eCNM) for analyzing complex fluid dynamics has been introduced. This procedure focuses the dynamics characterization on specific, potentially more informative, subspaces or subsets of variables. The applications of the eCNM within specific spatial subdomains and the challenges posed by temporally undersampled data have been explored. The novel approach has been applied to a swirl flame in unforced condition exibiting a precessing vortex core. Data consist in PIV measurements, UV-images filtered around the characteristic wavelength of OH* chemiluminescence and pressure signals from probes arranged around the inlet section of the jet.

The application of CNM on the velocity components has showcased the typical PVC flow field patterns and the characteristic frequency of the PVC mode has been estimated. A filtered distance, based on the filtered correlation matrix has been employed. This choice allowed us to enhance the separation of spatially and temporally coherent structures improving the interpretability of the phase portraits.

The construction of the extended cluster centroids has highlighted the existence of a rotating helical flame pattern, predominantly localized within regions influenced by PVC, and a symmetric motion of the flame from the central region of the combustor chamber, developing towards its

sides. The investigation then focused on the analysis of spatial subdomains, revealing the potential benefits of narrowing the focus to specific regions of interest within the fluid system, especially when a strong synchronization is present among the regions of the analyzed domain.

The eCNM approach gives the possibility to deal with different sampling frequencies among the dataset. Specifically, by leveraging high-resolution pressure measurements as the reference dataset and velocity components as the undersampled data, extended cluster centroids for velocity were successfully estimated. The analysis showcased the efficacy of this method in terms of mean reconstruction error even when the sampling frequency is artificially reduced below the Nyquist limit of the characteristic frequency.

This study has underscored the adaptability and robustness of Extended Cluster-based Network Modeling as a tool for comprehending and analyzing coherent structures in complex fluid flows.

Furthermore, a comparison between the Galerkin-POD-ROM and the conventional CNM has been carried out for NACA0012 airfoil flow fields at low Reynolds number. The results of this comparison have demonstrated the superior capability of the CNM-ROM in capturing the flow dynamics and statistics with a reduced computational effort. The CNM-ROM approach, based on direct transition matrices and cluster analysis, provided accurate predictions of the flow field and aerodynamic forces. Furthermore, the CNM-ROM model showcased its versatility by successfully predicting the effect of varying flow parameters, such as the angle of attack. By training the CNM-ROM with snapshots from different angle of attack settings, it accurately captured the amplitude and frequency of the lift and drag coefficients for unseen angle of attack values. This demonstrates the robustness and generalization capability of the CNM-ROM approach.

Finally, a functional approach for CNM, based on the analysis of piecewise trajectories has been proposed. This method aims at increasing the capability of CNM to capture high frequency behaviours and to reduce diffusion of the state vector to post transient attractors.

The trajectory-based CNM has demonstrated its superiority in effec-

tively capturing complex temporal behaviors exhibited the simple Landau oscillator and its variant with associated amplitude and frequency modulation. This innovative approach has been applied the fluidic pinball flow revealing its better performances compared with conventional CNM.

5.4 Recommendations for future work

In this work several data-driven approaches have been developed and applied on a variety of flow fields configurations.

The first step in future research should be the extension of the application of the data-driven methodologies here developed to diverse fluid dynamics problems, considering different flow regimes and physical phenomena; nevertheless to investigate their potential in complex industrial and environmental scenarios assessing their robustness in various contexts.

As regarding the future development in the modal analysis framework, further development should be the exploitation of the nonlinear modal reduction technique like Autoencoders, Neural Networks and Kernel POD. These approaches could further enhance the capability of the analysis in capturing complex nonlinear behaviours.

The development of an operator based Bi-Global Linear Stability Analysis code for two phase flows could be particularly useful in the characterization of mixing layers and jets. Furthermore, the nonlinear generalization the data-driven stability analysis approach proposed here, could be also effective for more complex flow configurations and applicable in industrial environment.

Step forward in reduced-order modeling techniques, exploring their integration with machine learning methods and expanding their application to high-dimensional and multiscale problems, may have the potential to significantly reduce computational costs in aerodynamic design, optimization processes and flow control. The development of Galerkin-ROMs and Neural Network ROMs with application to two-phase flows with Volume-of-Fluid formulation goes in this direction. Moreover, the control aimed

modal analysis of conventional Synthetic Jets (SJ) and Plasma Synthetic Jets (PSJ) actuators represents a challenging progress in this research topic.

The major findings from this work offer new insights into complex flow behaviors, by creating opportunities for further research and practical applications in various domains.

Appendix A

Insights into liquid curtains

Planar liquid sheets, falling under the effect of the gravity in still air, are often employed to deposit liquid layers over a solid surface in coating processes (e.g. Weinstein and Ruschak [50]).

Liquid sheet flows issuing a gaseous environment have been extensively investigated since the mid-20th century (Squire [190], Brown [191]). These studies hold significant scientific and industrial importance, including applications in coating deposition (Weinstein and Ruschak [192]), papermaking (Soderberg and Alfredsson [193]), the space industry (Chubb et al. [194]), dam safety, and waterfall noise reduction (Lodomez et al. [195]), among others. The literature encompasses a wide array of geometrical and physical configurations, spanning both planar (Barlow et al. [196], Weinstein et al. [197]) and round (Duke et al. [198]) geometries, as well as high-speed (Zandian et al. [199]) and low-speed (gravitational) flows (Lhuissier et al. [200], Girfoglio et al. [53]). In the case of low-speed flows, the term "curtain" is often used. These flows can occur in quiescent conditions (de Luca and Costa [201]) or in the presence of co-flowing air (Tammisola et al. [202]). The flow can either be confined by an enclosure on one side of the sheet (De Rosa et al. [203]) or unconfined on all sides (Barlow et al. [204]).

Despite extensive research efforts, which have been comprehensively reviewed in historical works by de Luca [205], Finnicum et al. [206], Lin and Jiang [207], Soderberg [51], and in a more recent contribution by

Torsey et al. [60], various aspects of the unsteady dynamics of gravitational liquid curtain flows remain the subject of ongoing studies and are presently unresolved. This work focuses on understanding the behavior of unconfined liquid sheets under different flow conditions, specifically in supercritical conditions ($We > 1$), during the supercritical-to-subcritical flow transition ($We = \mathcal{O}(1)$, transcritical regime), and in the subcritical regime ($We < 1$). The Weber number, denoted as We, serves as a primary governing parameter for the flow system. It is defined as the ratio between inertia and capillary forces within the liquid phase. Depending on the value of We, each of the three regimes mentioned above exhibits distinct characteristics and stability properties.

It is well-documented that the behavior of a liquid plane jet (often referred to as a sheet or curtain) interacting with an unconfined gaseous environment under the influence of gravity is not solely determined by the amplification of modal disturbances. Experimental observations indicate that the disintegration of the sheet occurs (especially when the flow rate is reduced) below the critical Weber number threshold ($We = 1$). This disintegration results from various phenomena, including the formation of holes characterized by two unsteady oscillating free-edges within the three-dimensional plane (Brown [191], de Luca and Meola [98], Roche et al. [208], Kacem et al. [209]).

The theoretical investigation of the dynamics of unconfined liquid sheets necessitates addressing a singularity that arises in the governing equations when employing low-order models (Finnicum et al. [206], Weinstein et al. [210], Girfoglio et al. [53]). This singularity emerges as the Weber number crosses the critical threshold of $We = 1$, that is, during the transition from supercritical ($We > 1$) to subcritical ($We < 1$) flow.

In the context of this research, the analysis of the singularity inherent in the equation governing the unsteady subcritical behavior of the liquid nappe has only recently been completed by Girfoglio et al. [53]. The liquid nappe represents a freely falling liquid sheet interacting with a one-sided closed air chamber. An illustrative example of this configuration is the flow of waterfalls, which occurs when water overflows over the long crests of dams or weirs. The noise generated by the oscillation of the waterfall can

propagate over considerable distances from the source, potentially causing vibrations in nearby building windows and even structural damage to the dam itself (Erpicum et al. [211], Lodomez et al. [54]).

A.1 Numerical simulation layout

Direct numerical simulations of thin liquid sheet flows have been obtained through the single-phase formulation and the Volume-of-Fluid (VOF) approach (Scardovelli and Zaleski [212]). In this framework the flow is incompressible, and the continuity and momentum equations (in index notation) for a two-dimensional formulation are respectively:

$$\frac{\partial u_i}{\partial x_i} = 0, \tag{A.1a}$$

$$\rho\left(\frac{\partial u_i}{\partial t} + u_j\frac{\partial u_i}{\partial x_j}\right) = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\right] + \sigma\kappa n_i\delta_S, \tag{A.1b}$$

in which $x_i = (x, y)$ are the spatial coordinates, $u_i = (u, v)$ are the longitudinal and transverse velocity components, p the pressure, σ the surface tension coefficient, κ the mean interface curvature and n_i the outward pointing unit normal vector $\mathbf{n}$ components. Following Popinet [95], $\mathbf{n}$ and κ are computed as: $\mathbf{n} = \boldsymbol{\nabla}\tilde{C}/|\boldsymbol{\nabla}\tilde{C}|$ and $\kappa = \boldsymbol{\nabla}\cdot\mathbf{n}$, where $\tilde{C}$ is a smoothed (i.e., diffused) version of the sharp volume fraction field. The effects of the gravitational field are contained in the pressure term p. The density (ρ) and dynamic viscosity (μ) are modeled as:

$$\rho = \rho_a + (\rho_l - \rho_a)C, \tag{A.2a}$$

$$\mu = \mu_a + (\mu_l - \mu_a)C, \tag{A.2b}$$

where subscripts a and l stand, respectively, for ambient and liquid, and $C(x_i, t)$ is the volume fraction, i.e. the ratio between the liquid volume to the total volume. In the liquid phase $C = 1$, whereas $C = 0$ in the ambient phase. Hence, from Eqs. (A.2a) and (A.2b) the density and the

viscosity are constant within each phase and vary across the separating interface, which is located where $0 < C < 1$ (throughout this work the interface is identified with the isoline $C = 0.5$). As shown by Hirt and Nichols [213], C satisfies the following advection equation:

$$\frac{\partial C}{\partial t} + \frac{\partial C u_i}{\partial x_i} = 0. \tag{A.3}$$

The governing equations are solved by means of the finite volume method using the open-source code BASILISK (http://basilisk.fr). For further details the reader is referred to the works of Popinet [94, 95].

In Fig. A.1 a schematic representation of the physical domain, the adopted numerical grid and a sample snapshot of the volume fraction C are reported. The right and left interfaces, respectively, are named $y^+(x,t)$ and $y^-(x,t)$ (panel (a)). Knowing the geometry of the interface, the centerline shape $l(x,t)$ and the liquid sheet thickness $h(x,t)$ are computed according to

$$\begin{bmatrix} l(x,t) \\ h(x,t) \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} y^+(x,t) \\ y^-(x,t) \end{bmatrix}. \tag{A.4}$$

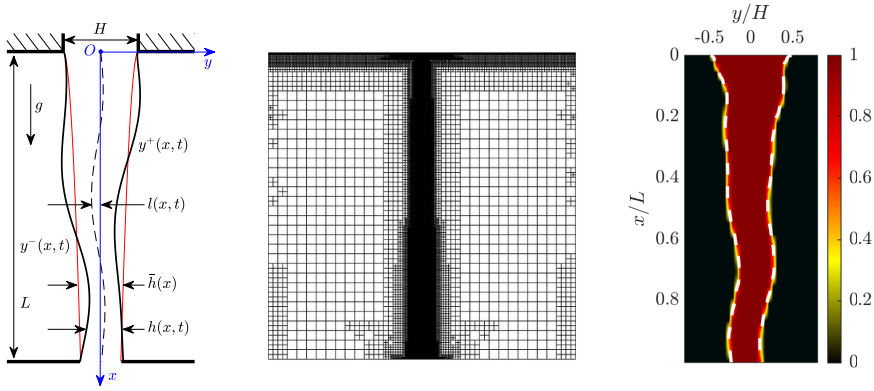
The configuration analyzed in section 2.1.1 consists in a forced liquid sheet flow, with a fully developed parabolic axial velocity profile at the inlet, which is perturbed by a sinusoidal forcing of variable amplitude in the lateral velocity component. Accordingly, Dirichlet boundary conditions are imposed at the inlet, for $y \in [-H/2; H/2]$:

$$u = \frac{3}{2} U \left[1 - \left(\frac{2y}{H} \right)^2 \right], \tag{A.5a}$$

$$v = \hat{A} U \, \sin\left(2\pi f_f (t - t_s)\right) \, \Theta\left(t - t_s\right), \tag{A.5b}$$

$$C = 1, \tag{A.5c}$$

where U and H are, respectively, the mean axial velocity and the thickness of the liquid sheet at the inlet, $\hat{A}$ is the dimensionless amplitude of



(a) Domain sketch. **(b)** Numerical grid. **(c)** Snapshot of C.

Figure A.1. Typical numerical arrangement and sample output of the supercritical case (a), Table A.1. (a) domain sketch, (b) numerical grid and (c) snapshot of C. The white dashed lines represent the interface between phases.

the forcing lateral velocity perturbation, f_f is the forcing frequency and $\Theta(t - t_s)$ the Heaviside function acting at the forcing starting time t_s. In the case of stability analysis carried out is Chapter 3, equation (A.5b) is replaced with homogeneous conditions. On the remaining part of the inlet side ($|y| > H/2$) no-slip conditions are imposed. At the right and left edges Neumann boundary conditions are imposed for all variables, whilst at the lower edge standard outflow conditions are set:

$$\frac{\partial u_i}{\partial x} = 0, \tag{A.6a}$$

$$\frac{\partial C}{\partial x} = 0, \tag{A.6b}$$

$$p = 0. \tag{A.6c}$$

Simulations have been carried out in both supercritical

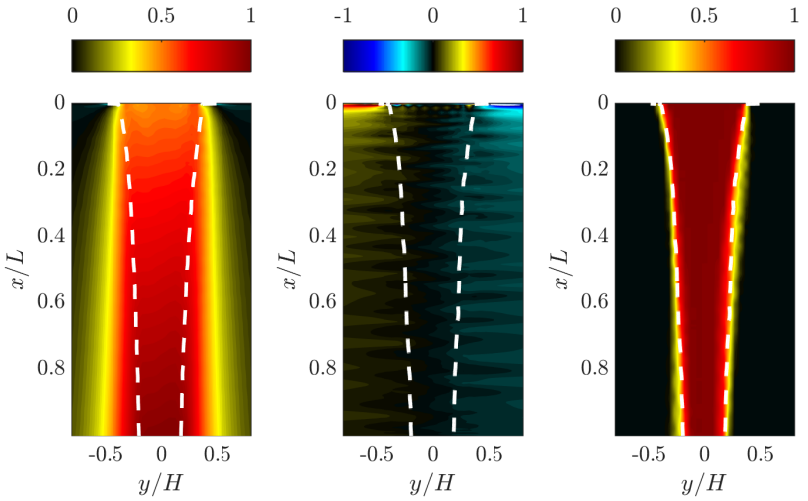
$$We = \frac{\rho_l U^2 H}{2\sigma} > 1,$$

where We is the Weber number based on half-thickness of the liquid sheet and subcritical ($We < 1$) regimes. The computational domain is a square of dimension $50H$. The adopted mesh is a quadtree-structured grid (see panel (b) of Fig. A.1) that retains its maximum level of refinement in the rectangular region $0 < x/L < 1$, $-1 < y/H < 1$, containing the entire liquid sheet, and dynamically refines the cells elsewhere according to user-defined adaptation criteria. For further details on the adaptive grid refinement strategy the reader is referred to van Hooft et al. [214]. The modal analysis focuses on the region $0 < x/L < 1$, $-1 < y/H < 1$, which is discretized considering 1025×41 uniform grid cells. Numerical data have been obtained every $5 \cdot 10^{-3}$ s over a period $T = 12$ s. For the subsequent analysis dimensionless parameters have been conveniently introduced, the Reynolds number ($Re = \rho_l U L/2\mu_l$), the Froude number ($Fr = U^2/(gL)$), the Strouhal number ($St = fH/U$) and the density ratio ($r_\rho = \rho_a/\rho_l$). The main flow parameters that have been investigated are reported in Table A.1, where St_f refers to the forcing frequency f_f. For all cases we have assumed $Fr = 0.33$ and $r_\rho = 0.01$.

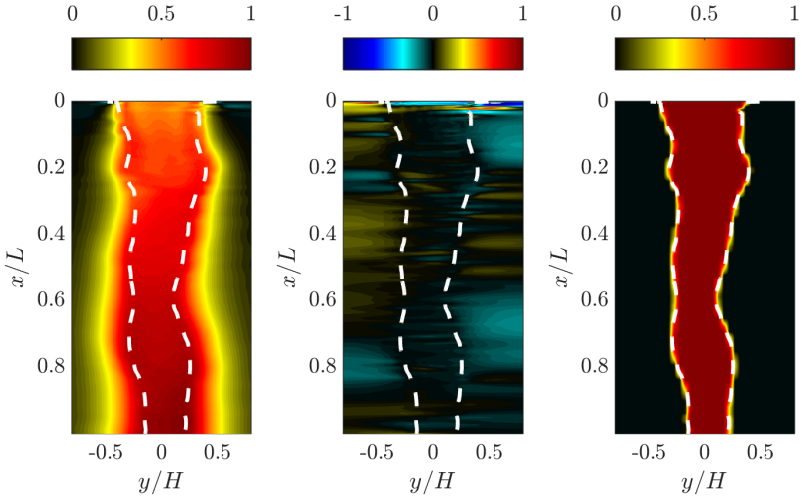
case	a	b	c	d	e	f	g	h	i	j	k
We	2.5	0.75	0.75	0.75	0.75	0.75	0.75	0.70	0.80	0.85	0.90
Re	400	20	400	1200	1600	1600	1600	1600	1600	1600	1600
St_f	0.076	0.076	0.076	0.076	0.076	0.137	0.182	0.076	0.076	0.076	0.076
$\hat{A}$	0.02	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Table A.1. Main flow parameters.

As an example, Fig. A.2 shows the mean flow (upper panels) and an instantaneous field (lower panels) corresponding to the subcritical case (e). The mean flow recovers the classic torricellian interface shape of the



(a) $\bar{u}(x,y)$ **(b)** $\bar{v}(x,y)$ **(c)** $\bar{C}(x,y)$



(d) Snapshot of u **(e)** Snapshot of v **(f)** Snapshot of C

Figure A.2. Supercritical conditions (case (e) of Table A.1). Mean flow ((a) $\bar{u}(x,y)$, (b) $\bar{v}(x,y)$ and (c) $\bar{C}(x,y)$) and generic snapshot of total field (snapshots of (d) u, (e) v and (f) C). In each panel the field variable is normalized by its maximum.

sheet; the snapshots show the emergence of varicose deformation. In fact, in the subcritical regime, as will be clarified in detail later on, a sinuous forcing in the v component at the inlet can excite a coupled sinuous-varicose response, highlighting the appearance of nonlinear resonance phenomena.

A.2 Linear stability analysis

To determine the asymptotic behaviour of a flow system, and to provide comparisons with results arising from direct numerical simulations and experiments, a boundary value eigenvalues problem can be formulated. For the case of a gravitational liquid sheet flow, this has been carried out starting from Eqs. (3.10)-(3.11) of the linear inviscid one-dimensional theoretical model presented in Section 3.1.1.

The classic global temporal modes position is then enforced assuming the following form of the disturbances:

$$\ell_s(x,t) = \hat{\ell}(x)\cdot e^{\lambda t}, \tag{A.7}$$
$$v_s(x,t) = \hat{v}(x)\cdot e^{\lambda t}, \tag{A.8}$$

where $\hat{\ell}$ and $\hat{v}$ are eigenfunctions and λ is the complex eigenvalue. This permits one to recast the governing equations into the following matrix form

$$\lambda\mathscr{M}\begin{pmatrix}\hat{v}\\ \hat{\ell}\end{pmatrix} = \mathscr{A}\begin{pmatrix}\hat{v}\\ \hat{\ell}\end{pmatrix}, \tag{A.9}$$

where the temporal operator is the block matrix

$$\mathscr{M} \equiv \begin{pmatrix} \mathscr{I} - RU\cdot\mathscr{I}\mathscr{N}\mathscr{T} & RU\cdot\mathscr{I}\mathscr{N}\mathscr{T}\cdot U\cdot\mathscr{D} \\ \mathscr{O} & \mathscr{I} \end{pmatrix},$$

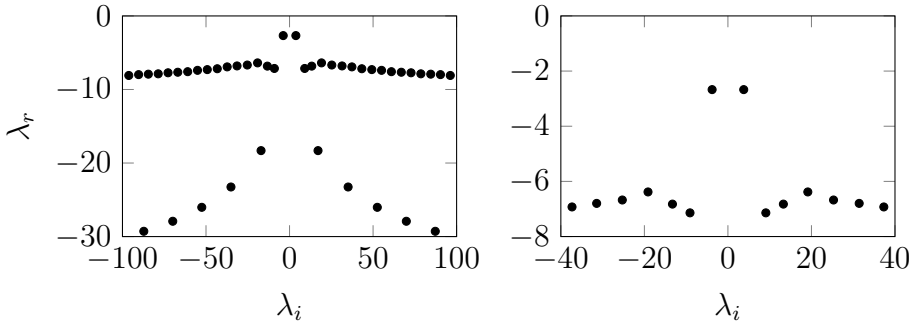


Figure A.3. Typical eigenvalue spectrum obtained from Eq. (A.9) (left panel) in supercritical conditions ($We > 1$), with zoom around the least stable part (right panel).

and the spatial operator is the block matrix

$$\mathscr{A} \equiv \begin{pmatrix} -U \cdot \mathscr{D} + \dfrac{\varepsilon}{2Re}\mathscr{D}^2 & U/We \cdot \mathscr{D}^2 \\ \mathscr{I} & -U \cdot \mathscr{D} \end{pmatrix},$$

with $\mathscr{I}$ denoting the identity operator, $\mathscr{O} = 0 \cdot \mathscr{I}$, $\mathscr{D}$ ($\mathscr{D}^2$) is the first (second) spatial derivative operator, and $\mathscr{INT}$ represents the integral $\int_0^1 (\cdot) \ln|x - \xi| \, d\xi$.

Eigenvalues and eigenfunctions are numerically computed by a Chebyshev collocation method in the MATLAB environment, with both differential and integral terms being spectrally accurate. A countable set of separated modes is found in the spectrum, which in supercritical conditions ($We > 1$) appears clustered in two separated branches, each one being characterized by eigenvalues with different real λ_r (i.e. growth rate) and evenly spaced imaginary λ_i (i.e. frequency) coefficients; an example is reported in Fig. A.3. In supercritical regime, two boundary conditions are required at the inlet section to close the system (A.9), i.e.

$$\hat{\ell}(0) = 0, \tag{A.10}$$

$$\left.\frac{\partial \hat{\ell}}{\partial x}\right|_0 = 0. \tag{A.11}$$

On the contrary, for subcritical cases, $We < 1$, due to the upstream directed wave characteristic velocity $U - \sqrt{U/We}$, the condition $\hat{\ell} = 0$ can only be retained at the inlet boundary. On the other hand, the system (A.9) becomes singular, and the condition removing the singularity constitutes the required second constraint to calculate the spectrum in subcritical regime ($We < 1$). This aspect is detailed in Della Pia et al. [61].

Appendix B

Functional data analysis

B.1 Functional principal component analysis

Functional principal component analysis (FPCA) is a method used in functional data analysis to reduce the dimensionality of the data by representing it in terms of a small number of principal components. FPCA provides a representation of the functional data as a linear combination of a small number of basis functions, where the coefficients corresponds to the principal components.

As shown in the Section 4.4, for the analysis, the raw original time series $\mathbf{u}(t)$ is split in N trajectories $\mathbf{u}_1(t'), \ldots, \mathbf{u}_n(t')$,with $t' = t - t_{0i}$ and defined in $t' \in [0, T_t]$. The functional data $\mathbf{u}_i(t')$ can be approximated as a linear combination of N basis functions, $\phi_1(t'), \ldots, \phi_p(t')$, such that

$$\mathbf{u}_i(t') \approx \boldsymbol{\mu}(t') + \sum_{k=1}^{N} c_{ik}\boldsymbol{\phi}_k(t'), \qquad i = 1, \ldots, N. \tag{B.1}$$

where $\boldsymbol{\mu}(t') = E(\mathbf{u}_i(t'))$ is the mean trajectory observed. To calculate the spatio-temporal modes $\boldsymbol{\phi}_k$ it is useful to define the covariance matrix :

$$\boldsymbol{C}(s,t) = E\big[(\mathbf{u}(s) - \boldsymbol{\mu}(s))(\mathbf{u}(t) - \boldsymbol{\mu}(t))^T\big], \tag{B.2}$$

that is estimated by means of:

$$\boldsymbol{C}(s,t) \approx \frac{1}{N-1}\sum_{i=1}^{N}((\mathbf{u}_i(s)-\boldsymbol{\mu}(s))(\mathbf{u}_i(t)-\boldsymbol{\mu}(t))^T). \tag{B.3}$$

where $\boldsymbol{C}(s,t)$ is a $d \times d$ matrix-valued function. The basis functions $\boldsymbol{\phi}_k$ are the eigenfunctions of the integral operator:

$$\int_0^{T_t} \boldsymbol{C}(s,t)\boldsymbol{\phi}_k(t)dt = \lambda_k \boldsymbol{\phi}_k(s), \tag{B.4}$$

where λ_k are the eigenvalues of the covariance operator and the first p eigenfunctions of the covariance operator, $\boldsymbol{\phi}_k(t)$, are the basis functions used in FPCA (4.47). The leading eigenfunctions capture most of the variation in the data, and the remaining eigenfunctions having small eigenvalues are considered as noise. Note that the $\boldsymbol{\phi}_k$ functions are orthonormal, as

$$\int_0^{T_t} \boldsymbol{\phi}_k(\tau)^T \boldsymbol{\phi}_l(\tau)d\tau = \delta_{kl}, \tag{B.5}$$

where δ_{kl} is the Kronecker delta. It is worth to notice that the discrete counterpart of (B.4) consists in an eigenvalue problem the dimensions of which being $(d \times n_i) \times (d \times n_i)$. This procedure can be done numerically using standard techniques, such as singular value decomposition. Notice that in Eq. (B.4) the covariance matrix $\boldsymbol{C}$ can be also substituted by nonlinear kernels $\boldsymbol{K}(s,t)$, as in [215].

Once the eigenfunctions are estimated, the functional data can be represented in terms of the principal component scores c_{ik}, which are projection of $\mathbf{u}_i(t') - \boldsymbol{\mu}(t')$ onto the $\boldsymbol{\phi}_k$ eigenfunctions:

$$c_{ik} = \int_0^{T_t} (\mathbf{u}_i(\tau) - \boldsymbol{\mu}(\tau))^T \, \boldsymbol{\phi}_k(\tau)d\tau. \tag{B.6}$$

The method involves the representation of the functional data as a linear combination of a small number of basis functions, which are chosen to

capture the essential features of the data. The principal component scores can be used to visualize the data or to perform further analysis, such as clustering.

In filtering methods, the matrices containing the STFT $\boldsymbol{\xi}_{ij}$, spline $\boldsymbol{s}_{ij}$ and wavelets $\boldsymbol{\beta}_{ijk}$ coefficients can be quite large. Despite the possibility of sparsity, these matrices can still be computationally expensive to cluster due to their size.

Consider a generic functional expansion basis for the signal $\mathbf{u}_i(t)$ as shown in Eq. (4.53). In this scenario, it's also possible to expand the function $\boldsymbol{\phi}_k(t')$ as follows:

$$\boldsymbol{\phi}_k(t') = \sum_{j=1}^{P} \boldsymbol{b}_{kj} f_j(t'). \tag{B.7}$$

By substituting Eq.s (4.53) and (B.7) into Eq. (B.4), it is possible to derive the coefficients $\boldsymbol{b}_{kj}$. Depending on the basis chosen, the coefficients c_{ik} can be obtained from Eq. (B.6).

In the work presented here, the number of expansion basis functions P has been set equal to the number of trajectories N, thus avoiding any loss in the approximation process.

B.2 Basis expansions

In this Appendix, several types of basis expansions that serve as crucial tools in signal processing are discussed. Section B.2.1 reports the Short Time Fourier Transform (STFT), an adaptable method used for balancing time and frequency information. Section B.2.2 introduces B-spline basis expansion, a versatile tool that is notable for its capability to create smooth and continuous functions. Lastly, Section B.2.3 explains the Wavelet Transform, a technique suited multi-resolution analysis.

B.2.1 Short time Fourier transform

The short-time Fourier transform (STFT) is a signal processing technique used to analyse the frequency content of a signal over time; it provides the spectral content of a signal in short overlapping time intervals, or windows. The STFT is computed by performing the Fourier transform of a windowed segment of the signal, and then repeating the process for each segment obtained sliding the window along the signal. This results in a time-frequency representation of the signal, reporting its frequency content as a function of time.

Considering a generic vectorial function $\mathbf{u}(t)$ it can be expressed as:

$$\hat{F}(\omega, t) = \int_{-\infty}^{\infty} w(t-\tau)\ \mathbf{u}(\tau) e^{-\mathrm{i}\omega\tau} d\tau, \tag{B.8}$$

where $\omega = 2\pi f$ represent the frequency and $w(t-\tau)$ is the chosen temporal window. The size and overlap of the window used for the STFT can be adjusted to control the trade-off between frequency and time resolution.

The discrete counterpart of (B.8) can be written as:

$$\hat{\alpha}_{ik} = \sum_{n=-\infty}^{\infty} \mathbf{u}(t_n) w(n - Ri) e^{-\mathrm{i}2\pi f_k n \Delta t} \tag{B.9}$$

where R is the sliding parameter equal to the difference between the window size n_w and overlapping n_{ov}. The inverse STFT is calculated as follows:

$$\mathbf{u}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{F}(\omega, \tau)\ e^{\mathrm{i}\omega t} d\tau d\omega, \tag{B.10}$$

that in discrete formulation reads:

$$\mathbf{u}(t_n) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \hat{\alpha}_{ik} e^{\mathrm{i}2\pi f_k n \Delta t}\ . \tag{B.11}$$

By substituting (B.9) into (4.54), it is possible to recover the Constant

Overlap-Add (COLA) Constraint:

$$\sum_{i=-\infty}^{\infty} w(n - Ri) = 1 \quad \forall n, \tag{B.12}$$

that assures a correct reconstruction.

B.2.2 B-splines decomposition

B-splines are a popular choice for modelling functional data, as they provide a flexible and computationally efficient way to represent functions/trajectories over a continuous domain $t \in [a, b]$. Each trajectory $\mathbf{u}_i$ is sampled over a set of knots $t_1, t_2, ..., t_{n_i}$, where $a = t_1 \leq t_2 \leq ... \leq t_{n_i} \leq b$, with a, b generally different between trajectories. By defining a set of basis functions $B_1(t), B_2(t), ..., B_p(t)$, where $p = n_i - o - 1$ (o being the degree of the polynomials) is the order of the B-spline basis, it is possible to represent each function u_i using such basis as follows:

$$\mathbf{u}_i(t) = \sum_{j=0}^{p} \mathbf{s}_{i,j} B_j(t), \tag{B.13}$$

where $\mathbf{s}_{i,j}$ are the coefficients of the B-spline expansion. The B-spline basis functions can be recursively defined as follows:

$$B_{i,0}(t) = \begin{cases} 1 \text{ if } t_i \leq t < t_{i+1} \\ 0 \text{ otherwise,} \end{cases} \tag{B.14}$$

$$B_{i,j}(t) = \frac{t - t_i}{t_{i+j} - t_i} B_{i,j}(t) + \frac{t_{i+j+1} - t}{t_{i+j+1} - t_{i+1}} B_{i+1,j-1}(t), \tag{B.15}$$

for $i = 1, 2, ..., n_i - p$ and $j = 2, 3, ..., p$. To fit the B-spline basis to each function $\mathbf{u}_i$, we can use a least squares regression to estimate the coefficients $\mathbf{s}_{i,j}$. It is worth to notice that $\sum_i B_{i,j}(t) = 1 \quad \forall t$ and the B-spline basis functions form a complete and orthonormal basis.

B.2.3 Wavelet Decomposition

Wavelets are mathematical functions with the property of being able to divide data into various components, each of which is studied with a resolution that matches its scale. This feature makes wavelets particularly suitable for analysing physical scenarios where the signal comprises discontinuities or sharp spikes.

The Continuous Wavelet Transform (CWT) of a signal, $\mathbf{u}(t)$, is mathematically expressed as:

$$CWT_{\mathbf{u}}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \mathbf{u}(t)\psi^* \left(\frac{t-b}{a}\right) dt. \tag{B.16}$$

In the above equation, a and b represent the scale factor and the translation factor respectively, while $\psi(t)$ denotes the mother wavelet. $(\cdot)^*$ represents the complex conjugate operator.

A mother wavelet $\psi(t)$ is a function in $L^2(\mathbb{R})$ (the space of square integrable functions), from which all other wavelet functions are generated through dilations and translations. Essentially, a wavelet is a waveform of effectively limited duration that has an average value of zero.

In practical applications, the Discrete Wavelet Transform (DWT) samples the CWT in a non-redundant manner, allowing for perfect reconstruction. The DWT of a signal $\mathbf{u}$ at a given scale level j and translation k is represented as:

$$DWT_{\mathbf{u}}[j,k] = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} \mathbf{u}(t)\psi \left(\frac{t-2^j k}{2^j}\right) dt, \tag{B.17}$$

In the equation above, j and k are integers representing the scale and the translation on the time axis, respectively. The 'db10' wavelet, also known as the Daubechies wavelet of order 10, is the wavelet chosen for this analysis. It has 10 vanishing moments, expressed as:

$$\int t^k \psi(t) dt = 0, \quad \text{for } k = 0, 1, ..., 9. \tag{B.18}$$

The Daubechies wavelet family, including the 'db10' wavelet, com-

prise orthogonal wavelets with varying numbers of vanishing moments. Figure B.1 showcases the 'db10' wavelets, illustrating how wavelet decomposition provides a multiresolution analysis that decomposes the signal into its constituent temporal scales.

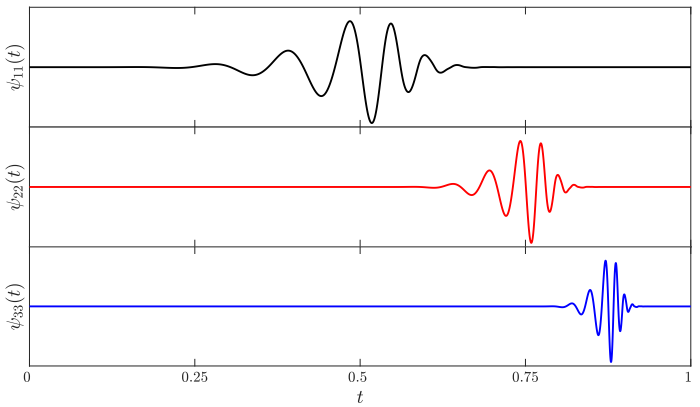


Figure B.1. The 'db10' wavelet and its decomposition of a time series signal.

The Inverse Continuous Wavelet Transform (ICWT) and Inverse Discrete Wavelet Transform (IDWT) are employed to reconstruct the signal from its continuous or discrete wavelet transform, respectively. In the continuous case, given $CWT_{\mathbf{u}}(a, b)$, the original signal can be retrieved by integrating across all scales (a) and translations (b):

$$\mathbf{u}(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} CWT_{\mathbf{u}}(a, b) \cdot \frac{1}{|a|} \psi^* \left(\frac{t-b}{a} \right) dbda, \qquad \text{(B.19)}$$

where C_ψ is a normalization constant that depends on the wavelet function.

For the discrete wavelet transform, given $DWT_{\mathbf{u}}[j, k]$, the original signal can be reconstructed by summing over all scales (j) and translations (k):

$$\mathbf{u}(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} DWT_{\mathbf{u}}[j,k] \cdot \psi \left(\frac{t - 2^j k}{2^j} \right). \tag{B.20}$$

For a perfect reconstruction, it is crucial that the wavelet functions are orthogonal, as is the case for the Daubechies family of wavelets. It is also essential to use the same wavelet function for both the wavelet transform and its inverse.

Bibliography

[1] AKM Fazle Hussain. Coherent structures and turbulence. *Journal of Fluid Mechanics*, 173:303–356, 1986.

[2] G. R. S. Assi, P. W. Bearman, and J. R. Meneghini. On the wake-induced vibration of tandem circular cylinders: the vortex interaction excitation mechanism. *Journal of Fluid Mechanics*, 661:365–401, aug 2010. doi: 10.1017/s0022112010003095.

[3] Bernhard C Bobusch, René Woszidlo, JM Bergada, C Navid Nayeri, and Christian Oliver Paschereit. Experimental study of the internal flow structures inside a fluidic oscillator. *Experiments in fluids*, 54:1–12, 2013.

[4] Thierry J. Poinsot, Arnaud C. Trouve, Denis P. Veynante, Sebastien M. Candel, and Emile J. Esposito. Vortex-driven acoustically coupled combustion instabilities. *Journal of Fluid Mechanics*, 177:265–292, 1987. doi: 10.1017/S0022112087000958.

[5] Philippe Ausoni, Mohamed Farhat, Xavier Escaler, Eduard Egusquiza, and François Avellan. Cavitation influence on von kármán vortex shedding and induced hydrofoil vibrations. 2007.

[6] Grégory Pujals, Sébastien Depardon, and Carlo Cossu. Drag reduction of a 3d bluff body using coherent streamwise streaks. *Experiments in fluids*, 49:1085–1094, 2010.

[7] F Bastin, P Lafon, and S Candel. Computation of jet mixing noise due to

coherent structures: the plane jet case. *Journal of fluid mechanics*, 335: 261–304, 1997.

[8] A. Nair, C.-A. Yeh, E. Kaiser, B. R. Noack, S. L. Brunton, and K. Taira. Cluster-based feedback control of turbulent post-stall separated flows. *J. Fluid Mech.*, 875:345–375, 2019.

[9] 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*, 55(12):4013–4041, 2017.

[10] Jan S Hesthaven, Gianluigi Rozza, and Benjamin Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer International Publishing, 2016. doi: 10.1007/978-3-319-22470-1.

[11] Steven L. Brunton and Bernd R. Noack. Closed-loop turbulence control: Progress and challenges. *Applied Mechanics Reviews*, 67(5), aug 2015. doi: 10.1115/1.4031175.

[12] V. Theofilis. Global linear instability. *Annual Review of Fluid Mechanics*, 43(1):319–352, 2011. doi: 10.1146/annurev-fluid-122109-160705.

[13] Bernd R. Noack, Konstantin Afanasiev, Marek Morzyński, Gilead Tadmor, and Frank Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497:335–363, 2003. doi: 10.1017/S0022112003006694.

[14] K. Taira, M. S. Hemati, S. L. Brunton, Y. Sun, K. Duraisamy, S. Bagheri, S. T. M. Dawson, and C. Yeh. Modal analysis of fluid flows: Applications and outlook. *AIAA Journal*, 58(3):998–1022, 2020. doi: 10.2514/1.J058462.

[15] O. Schmidt, A. Towne, G. Rigas, T. Colonius, and G. Brès. Spectral analysis of jet turbulence. *Journal of Fluid Mechanics*, 855, 11 2018. doi: 10.1017/jfm.2018.675.

[16] A. Towne, O. T. Schmidt, and T. Colonius. Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics*, 847:821–867, 2018.

[17] M. Chiatto, J. K. Shang, L. de Luca, and F. Grasso. Insights into low Reynolds flow past finite curved cylinders. *Physics of Fluids*, 33(3): 035150, 2021.

[18] S. Nidhan, K. Chongsiripinyo, O. T. Schmidt, and S. Sarkar. Spectral proper orthogonal decomposition analysis of the turbulent wake of a disk at re = 50 000. *Phys. Rev. Fluids*, 5:124606, Dec 2020. doi: 10.1103/PhysRevFluids.5.124606.

[19] Xiao He, Zhou Fang, Georgios Rigas, and Mehdi Vahdati. Spectral proper orthogonal decomposition of compressor tip leakage flow. *Physics of Fluids*, 33(10):105105, 10 2021. ISSN 1070-6631. doi: 10.1063/5.0065929.

[20] Shijun Chu, Chao Xia, Hanfeng Wang, Yajun Fan, and Zhigang Yang. Three-dimensional spectral proper orthogonal decomposition analyses of the turbulent flow around a seal-vibrissa-shaped cylinder. *Physics of Fluids*, 33(2):025106, 02 2021. ISSN 1070-6631. doi: 10.1063/5.0035789.

[21] Leandra I. Abreu, André V. G. Cavalieri, Philipp Schlatter, Ricardo Vinuesa, and Dan S. Henningson. Spectral proper orthogonal decomposition and resolvent analysis of near-wall coherent structures in turbulent pipe flows. *Journal of Fluid Mechanics*, 2020. doi: 10.1017/jfm.2020.445.

[22] M. Chiatto, L. de Luca, and F. Grasso. Modal analysis of actively controlled flow past a backward facing ramp. *Physical Review Fluids*, 6:064608, 2021.

[23] H. Huang, D. Dabiri, and M. Gharib. On errors of digital particle image velocimetry. *Measurement Science and Technology*, 8(12):1427 – 1440, 1997. doi: 10.1088/0957-0233/8/12/007.

[24] M.A. Mendez, M. Raiola, A. Masullo, S. Discetti, A. Ianiro, R. Theunissen, and J.-M. Buchlin. POD-based background removal for particle image velocimetry. *Experimental Thermal and Fluid Science*, 80:181–192, jan 2017. doi: 10.1016/j.expthermflusci.2016.08.021.

[25] Jerry Westerweel and Fulvio Scarano. Universal outlier detection for piv data. *Experiments in Fluids*, 39(6):1096 – 1100, 2005. doi: 10.1007/s00348-005-0016-6.

[26] M.A. Oliver and R. Webster. Kriging: A method of interpolation for geographical information systems. *International Journal of Geographical Information Systems*, 4(3):313 – 332, 1990. doi: 10.1080/02693799008941549.

[27] Richard M. Everson and Lawrence Sirovich. Karhunen–loève procedure for gappy data. *Journal of The Optical Society of America A-optics Image Science and Vision*, 12:1657–1664, 1995.

[28] Daniele Venturi and George Karniadakis. Gappy data and reconstruction procedures for flow past a cylinder. *Journal of Fluid Mechanics*, 519: 315–336, 2004. doi: 10.1017/S0022112004001338.

[29] Hasan Gunes, Sirod Sirisup, and George Karniadakis. Gappy data: To krig or not to krig? *Journal of Computational Physics*, 212(1):358–382, 2006. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2005.06.023. URL https://www.sciencedirect.com/science/article/pii/S0021999105003256.

[30] Akhil Nekkanti and Oliver T. Schmidt. Gappy spectral proper orthogonal decomposition. *Journal of Computational Physics*, 478: 111950, 2023. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2023.111950. URL https://www.sciencedirect.com/science/article/pii/S0021999123000451.

[31] Jakob G. R. von Saldern, Johann Moritz Reumschüssel, Thomas L. Kaiser, Moritz Sieber, and Kilian Oberleithner. Mean flow data assimilation based on physics-informed neural networks. *Physics of Fluids*, 115129, 2022. doi: 10.1063/5.0116218. URL http://arxiv.org/abs/2208.03109.

[32] M. Buzzicotti, F. Bonaccorso, P. Clark Di Leoni, and L. Biferale. Reconstruction of turbulent data with deep generative models for semantic inpainting from turb-rot database. *Physical Review Fluids*, 6(5), 2021. doi: 10.1103/PhysRevFluids.6.050503.

[33] Stefano Discetti and Yingzheng Liu. Machine learning for flow field measurements: a perspective. *Measurement Science and Technology*, 34 (2), 2023. doi: 10.1088/1361-6501/ac9991.

[34] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58, 12 2009. doi: 10.1145/1970392.1970395.

[35] X. Yuan and J. Yang. Sparse and low rank matrix decomposition via alternating direction method. *Pacific Journal of Optimization*, 9, 01 2009.

[36] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *Mathematical Programming*, 9, 09 2010.

[37] I. Scherl, B. Strom, J. Shang, O. Williams, B. Polagye, and S. Brunton. Robust principal component analysis for modal decomposition of corrupt fluid flows. *Physical Review Fluids*, 5, 05 2020. doi: 10.1103/PhysRevFluids.5.054401.

[38] P. Huerre and P. A. Monkewitz. Local and global instabilities in spatially developing flows. *Annual Review of Fluid Mechanics*, 22(1):473–537, 2005.

[39] Chang Hou, Nan Deng, and Bernd R. Noack. Trajectory-optimized cluster-based network model for the sphere wake. *Physics of Fluids*, 34(8), aug 2022. doi: 10.1063/5.0098655.

[40] Zhengxiao Ma, Jian Yu, and Ruoye Xiao. Data-driven reduced order modeling for parametrized time-dependent flow problems. *Physics of Fluids*, 34(7), 2022. doi: 10.1063/5.0098122.

[41] I. Mezic. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41:309–325, 08 2005. doi: 10.1007/s11071-005-2824-x.

[42] CLARENCE W. ROWLEY, IGOR MEZIĆ, SHERVIN BAGHERI, PHILIPP SCHLATTER, and DAN S. HENNINGSON. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, nov 2009. doi: 10.1017/s0022112009992059.

[43] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[44] J. Tu, C. Rowley, D. Luchtenburg, S. Brunton, and J. Kutz. On dynamic mode decomposition: theory and applications. *Journal of Computational Dynamics*, 1:391–421, 11 2013.

[45] P.J. Schmid. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*, 54(1):225–254, 2022. doi: 10.1146/annurev-fluid-030121-015835.

[46] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho. Detection of face spoofing using visual dynamics. *IEEE Transactions on Information Forensics and Security*, 10(4):762–777, 2015. doi: 10.1109/TIFS.2015.2406533.

[47] B.W. Brunton, L.A. Johnson, J.G. Ojemann, and J.N. Kutz. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, 2016. doi: 10.1016/j.jneumeth.2015.10.010.

[48] J. Proctor and P. Welkhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International health*, 7:139–45, 03 2015. doi: 10.1093/inthealth/ihv009.

[49] R. Ranjan, S. Unnikrishnan, and D. Gaitonde. A robust approach for stability analysis of complex flows using high-order navier-stokes solvers. *Journal of Computational Physics*, 403, 2020.

[50] Steven J. Weinstein and Kenneth J. Ruschak. Coating flows. *Annual Review of Fluid Mechanics*, 36(1):29–53, 2004. doi: 10.1146/annurev.fluid.36.050802.122049.

[51] L. D. Soderberg. Absolute and convective instability of a relaxational plane liquid jet. *Journal of Fluid Mechanics*, 493:89–119, 2003.

[52] A. M. Binnie. Resonating waterfalls. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 339(1619):435–449, 1974.

[53] M. Girfoglio, F. De Rosa, G. Coppola, and L. de Luca. Unsteady critical liquid sheet flows. *Journal of Fluid Mechanics*, 821:219–247, 2017.

[54] M. Lodomez, B. Tullis, P. Archambeau, V. Kitsikoudis, M. Pirotton, B. Dewals, and S. Erpicum. Nappe oscillations on free-overfall structures, data from laboratory experiments. *Scientific Data*, 7:180, 2020.

[55] S. Schmidt and K. Oberleithner. Instability of forced planar liquid jets: mean field analysis and non linear simulation. *Journal of Fluid Mechanics*, 883:1–38, 2020.

[56] A. Arote, M. Bade, and J. Banerjee. Numerical investigations on stability of the spatially oscillating planar two-phase liquid jet in a quiescent atmosphere. *Physics of Fluids*, 31:112103, 2019. doi: https://doi.org/10.1063/1.5123762.

[57] A. Arote, M. Bade, and J. Banerjee. On coherent structures of spatially oscillating planar liquid jet developing in a quiescent atmosphere. *Physics of Fluids*, 32:082111, 2020. doi: https://doi.org/10.1063/5.0016480.

[58] G. G. Agbaglah. Numerical study of hole formation in a thin flapping liquid sheet sheared by a fast gas stream. *Physics of Fluids*, 33(6), 2021. doi: 10.1063/5.0055167.

[59] A. Liu, D. Sun, B. Yu, J. Wei, and Z. Cao. An adaptive coupled volume-of-fluid and level set method based on unstructured grids. *Physics of Fluids*, 33:012102, 2021. doi: https://doi.org/10.1063/5.0031737.

[60] B. Torsey, S. J. Weinstein, D. Ross, and N. Barlow. The effect of pressure fluctuations on the shapes of thinning liquid curtains. *Journal of Fluid Mechanics*, 910:1–14, 2021.

[61] Alessandro Della Pia, Matteo Chiatto, and Luigi de Luca. Global eigenmodes of thin liquid sheets by means of volume-of-fluid simulations. *Physics of Fluids*, 32:082112, 2020. doi: 10.1063/5.0020559.

[62] A. Della Pia, A. Colanera, and M. Chiatto. Surface tension-induced instability in spatially developing subcritical liquid curtains. *Physics of Fluids*, 34:042122, 2022.

[63] A. Della Pia, A. Colanera, M. Chiatto, and L. de Luca. Energy insights into the unsteady dynamics of a viscous gravitational liquid sheet. *Physics of Fluids*, 33(9):092118, 2021. doi: 10.1063/5.0065590.

[64] A. Colanera, A. Della Pia, M. Chiatto, L. de Luca, and F. Grasso. Modal decomposition analysis of unsteady viscous liquid sheet flows. *Physics of Fluids*, 33:092117, 2021.

[65] Alessandro Della Pia, Matteo Chiatto, and Luigi de Luca. Receptivity to forcing disturbances in subcritical liquid sheet flows. *Physics of Fluids*, 33:032113, 2021. doi: 10.1063/5.0044322.

[66] Alfio Quarteroni, Gianluigi Rozza, and Andrea Manzoni. Certified reduced basis approximation for parametrized partial differential equations and applications. *Journal of Mathematics in Industry*, 1(1), jun 2011. doi: 10.1186/2190-5983-1-3.

[67] Panagiotis G. Papaioannou, Ronen Talmon, Ioannis G. Kevrekidis, and Constantinos Siettos. Time-series forecasting using manifold learning, radial basis function interpolation, and geometric harmonics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(8), aug 2022. doi: 10.1063/5.0094887.

[68] M. A. Mendez, M. Balabane, and J.-M. Buchlin. Multi-scale proper orthogonal decomposition of complex fluid flows. *Journal of Fluid Mechanics*, 870:988–1036, may 2019. doi: 10.1017/jfm.2019.212.

[69] Paola Cinnella, Pietro Marco Congedo, Valentino Pediroda, and Lucia Parussini. Sensitivity analysis of dense gas flow simulations to thermodynamic uncertainties. *Physics of Fluids*, 23(11), nov 2011. doi: 10.1063/1.3657080.

[70] W.N. Edeling, P. Cinnella, R.P. Dwight, and H. Bijl. Bayesian estimates of parameter variability in the k-ϵ turbulence model. *Journal of Computational Physics*, 258:73–94, feb 2014. doi: 10.1016/j.jcp.2013.10.027.

[71] Heng Xiao and Paola Cinnella. Quantification of model uncertainty in RANS simulations: A review. *Progress in Aerospace Sciences*, 108:1–31, jul 2019. doi: 10.1016/j.paerosci.2018.10.001.

[72] Clarence W. Rowley and Scott T.M. Dawson. Model reduction for flow analysis and control. *Annual Review of Fluid Mechanics*, 49(1):387–417, jan 2017. doi: 10.1146/annurev-fluid-010816-060042.

[73] Seungjoon Lee, Mahdi Kooshkbaghi, Konstantinos Spiliotis, Constantinos I. Siettos, and Ioannis G. Kevrekidis. Coarse-scale PDEs from fine-scale observations via machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(1), jan 2020. doi: 10.1063/1.5126869.

[74] Gianluca Fabiani, Francesco Calabrò, Lucia Russo, and Constantinos Siettos. Numerical solution and bifurcation analysis of nonlinear partial differential equations with extreme learning machines. *Journal of Scientific Computing*, 89(2), oct 2021. doi: 10.1007/s10915-021-01650-5.

[75] Hector Vargas Alvarez, Gianluca Fabiani, Nikolaos Kazantzis, Constantinos Siettos, and Ioannis G. Kevrekidis. Discrete-time nonlinear feedback linearization via physics-informed machine learning. *Journal of Computational Physics*, 492:112408, nov 2023. doi: 10.1016/j.jcp.2023.112408.

[76] Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367:1026–1030, 2020. ISSN 10959203. doi: 10.1126/science.aaw4741.

[77] D. Fernex, B. R. Noack, and R. Semaan. Cluster-based network modeling–from snapshots to complex dynamical systems. *Science Advances*, 7(25): eabf5006, 2021.

[78] Timo Vogel, Paul-Florian Kroll, Felix Dietrich, Jan Paul Beuth, Johann Moritz Reumschüssel, and Kilian Oberleithner. Data-driven modeling of flame dynamics using cluster-based network modeling. pages 27–28, 9 2023.

[79] Xin Wang, Nan Deng, Guy Y. Cornejo Maceda, and Bernd R. Noack. Cluster-based control for net drag reduction of the fluidic pinball. *Physics of Fluids*, 35(2), feb 2023. doi: 10.1063/5.0136499.

[80] N. Deng, B. R. Noack, M. Morzyński, and L. R. Pastur. Cluster-based hierarchical network model of the fluidic pinball–Cartographing transient

and post-transient, multi-frequency, multi-attractor behaviour. *J. Fluid Mech.*, 934:A24, 2022.

[81] J. Borée. Extended proper orthogonal decomposition: A tool to analyse correlated events in turbulent flows. *Experiments in Fluids*, 35:188–192, 8 2003. ISSN 07234864. doi: 10.1007/s00348-003-0656-3.

[82] Moritz Sieber, Christian Oliver Paschereit, and Kilian Oberleithner. Advanced Identification of Coherent Structures in Swirl-Stabilized Combustors. *Journal of Engineering for Gas Turbines and Power*, 139(2): 021503, 09 2016. ISSN 0742-4795. doi: 10.1115/1.4034261. URL https://doi.org/10.1115/1.4034261.

[83] Eurika Kaiser, Bernd R. Noack, Laurent Cordier, Andreas Spohn, Marc Segond, Markus Abel, Guillaume Daviller, Jan Östh, Siniša Krajnović, Robert K. Niven, and et al. Cluster-based reduced-order modelling of a mixing layer. *Journal of Fluid Mechanics*, 754:365–414, 2014. doi: 10.1017/jfm.2014.355.

[84] Rishabh Ishar, Eurika Kaiser, Marek Morzyński, Daniel Fernex, Richard Semaan, Marian Albers, Pascal S. Meysonnat, Wolfgang Schröder, and Bernd R. Noack. Metric for attractor overlap. *Journal of Fluid Mechanics*, 874:720–755, jul 2019. doi: 10.1017/jfm.2019.447.

[85] O. T. Schmidt and T. Colonius. Guide to spectral proper orthogonal decomposition. *AIAA Journal*, 58(3):1023–1033, 2020.

[86] C. W. Rowley and S. T. M. Dawson. Model reduction for flow analysis and control. *Annual Review of Fluid Mechanics*, 49(1):387–417, 2017.

[87] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25(1):539–575, 1993.

[88] P. A. Kadu, Y. Sakai, Y. Ito, K. Iwano, M. Sugino, T. Katagiri, T. Hayase, and K. Nagata. Application of spectral proper orthogonal decomposition to velocity and passive scalar fields in a swirling coaxial jet. *Physics of Fluids*, 32(1):015106, 2020.

[89] B. Zhang, R. Ooka, and H. Kikumoto. Identification of three-dimensional flow features around a square-section building model via spectral proper orthogonal decomposition. *Physics of Fluids*, 33(3), 2021.

[90] David Barreiro-Villaverde, Anne Gosset, and Miguel A. Mendez. On the dynamics of jet wiping: Numerical simulations and modal analysis. *Physics of Fluids*, 33(6):062114, 2021. doi: 10.1063/5.0051451.

[91] Sirovich Lawrence. Turbulence and the dynamics of coherent structures. i. coherent structures. *Quarterly of Applied Mathematics*, 45:561–571, 1987.

[92] P. Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[93] A. Nekkanti and O. T. Schmidt. Frequency-time analysis, low-rank reconstruction and denoising of turbulent flows using SPOD. *Journal of Fluid Mechanics*, 926(A26):1–35, 2021.

[94] S. Popinet. Gerris: a tree-based adaptive solver for the incompressible Euler equations in complex geometries. *Journal of Computational Physics*, 190(2):572–600, 2003.

[95] S. Popinet. An accurate adaptive solver for surface-tension-driven interfacial flows. *Journal of Computational Physics*, 228(16):5838–5866, 2009.

[96] C. Hayashi. Subharmonic oscillations in nonlinear systems. *Journal of Applied Physics*, 24(5):521–529, 1953. doi: 10.1063/1.1721322.

[97] O. Schmidt. Bispectral mode decomposition of nonlinear flows. *Nonlinear Dynamics*, 102:2479–2501, 2020.

[98] L. de Luca and C. Meola. Surfactant effects on the dynamics of a thin liquid sheet. *Journal of Fluid Mechanics*, 300:71–85, 1995.

[99] N. Le Grand-Piteira, P. Brunet, L. Lebon, and L. Limat. Propagating wave pattern on a falling liquid curtain. *Physical Review E*, 74:026305, 2006.

[100] A. Colanera, A. Della Pia, M. R. Acquaviva, M. Chiatto, and L. de Luca. Modal analysis of a three-dimensional gravitational liquid sheet. *European Congress on Computational Methods in Applied Sciences and Engineering*, Oslo (Norway), 5-9 June 2022.

[101] G.A. Brès, F.E. Ham, J.W. Nichols, and S.K. Lele. Unstructured large-eddy simulations of supersonic jets. *AIAA Journal*, 55(4):1164–1184, 2017. doi: 10.2514/1.J057012.

[102] Yang Zhang, Yiyang Sun, Nishul Arora, Louis N. Cattafesta III, Kunihiko Taira, and Lawrence S. Ukeiley. Suppression of cavity flow oscillations via three-dimensional steady blowing. *AIAA Journal*, 57(1):90–105, 2019. doi: 10.2514/1.J057012.

[103] Yiyang Sun, Kunihiko Taira, Louis N. Cattafesta, and Lawrence S. Ukeiley. Biglobal instabilities of compressible open-cavity flows. *Journal of Fluid Mechanics*, 826:270–301, 2017. doi: 10.1017/jfm.2017.416.

[104] Yiyang Sun, Qiong Liu, Louis N. Cattafesta, Lawrence S. Ukeiley, and Kunihiko Taira. Resolvent analysis of compressible laminar and turbulent cavity flows. *AIAA Journal*, 58(3):1046 – 1055, 2020. doi: 10.2514/1.J058633.

[105] J.E. Rossiter. *Wind tunnel experiments on the flow over rectangular cavities at subsonic and transonic speeds*. Number 3438 in ARC 26621. Ministry of Aviation, 1964.

[106] H. Heller and D. Bliss. *The physical mechanism of flow-induced pressure fluctuations in cavities and concepts for their suppression*. 1975. doi: 10.2514/6.1975-491.

[107] Michael Wilczek, Richard J. A. M. Stevens, and Charles Meneveau. Spatio-temporal spectra in the logarithmic layer of wall turbulence: large-eddy simulations and simple models. *Journal of Fluid Mechanics*, 769: R1, 2015. doi: 10.1017/jfm.2015.116.

[108] Carlos Henrique Grohmann, Mike J. Smith, and Claudio Riccomini. Multiscale analysis of topographic surface roughness in the midland valley,

scotland. *IEEE Transactions on Geoscience and Remote Sensing*, 49(4): 1200 – 1213, 2011. doi: 10.1109/TGRS.2010.2053546.

[109] D. Fuster, J. P. Matas, S. Marty, S. Popinet, J. Hoepffner, A. Cartellier, and S. Zaleski. Instability regimes in the primary breakup region of planar coflowing sheets. *Journal of Fluid Mechanics*, 736:150–176, 2013.

[110] P. G. Drazin and W. H. Reid. *Hydrodynamic Stability*. Cambridge Mathematical Library. Cambridge University Press, 2 edition, 2004. doi: 10.1017/CBO9780511616938.

[111] P.J. Schmid and D.S. Henningson. *Stability and Transition in Shear Flows*. Applied Mathematical Sciences. Springer New York, 2012. ISBN 9781461301851. URL https://books.google.it/books?id=PxHVBwAAQBAJ.

[112] W. O. Criminale, T. L. Jackson, and R. D. Joslin. *Theory and Computation of Hydrodynamic Stability*. Cambridge Monographs on Mechanics. Cambridge University Press, 2003. doi: 10.1017/CBO9780511550317.

[113] Xianyang Chen, Jiacai Lu, Stéphane Zaleski, and Grétar Tryggvason. Characterizing interface topology in multiphase flows using skeletons. *Physics of Fluids*, 34(9), 2022. doi: 10.1063/5.0109333.

[114] Peter J. Schmid. Nonmodal stability theory. *Annual Review of Fluid Mechanics*, 39(1):129–162, 2007. doi: 10.1146/annurev.fluid.38.050304.092139.

[115] E. Ferrer, J. de Vicente, and E. Valero. Low cost 3d global instability analysis and flow sensitivity based on dynamic mode decomposition and high-order numerical tools. *International Journal for Numerical Methods in Fluids*, 76(3):169–184, 2014. doi: 10.1002/fld.3930.

[116] B. Herrmann, P. J. Baddoo, R. Semaan, S. L. Brunton, and B. J. McKeon. Data-driven resolvent analysis. *Journal of Fluid Mechanics*, 918:A10, 2021.

[117] J.S. Kern, M. Beneitez, A. Hanifi, and D.S. Henningson. Transient linear stability of pulsating poiseuille flow using optimally time-dependent modes. *Journal of Fluid Mechanics*, 927:A6, 2021. doi: 10.1017/jfm.2021.743.

[118] V. Theofilis. Globally unstable basic flows in open cavities. In *6th Aeroacoustics Conference and Exhibit*. American Institute of Aeronautics and Astronautics, 2000.

[119] S. Bagheri, E. Akervik, L. Brandt, and D. S. Henningson. Matrix-free methods for the stability and control of boundary layers. *AIAA Journal*, 47(5):1057 – 1068, 2009.

[120] F. Gomez, R. Gomez, and V. Theofilis. On three-dimensional global linear instability analysis of flows with standard aerodynamics codes. *Aerospace Science and Technology*, 32(1):223 – 234, 2014.

[121] B. T. Chu. On the energy transfer to small disturbances in fluid flow (part i). *Acta Mechanica*, 1:215–234, 1965.

[122] A. Kornecki, E. H. Dowell, and J. O'Brien. On the aeroelastic instability of two-dimensional panels in uniform incompressible flow. *Journal of Sound and Vibration*, 47(2):163–178, 1976.

[123] P. J. Olsson and D. S. Henningson. Optimal disturbance growth in watertable flow. *Studies in Applied Mathematics*, 94:183–210, 1995.

[124] M. Chiatto and A. Della Pia. Natural frequency discontinuity of vertical liquid sheet flows at transcritical threshold. *Journal of Fluid Mechanics*, 945:A32, 2022.

[125] Giovanni Stabile, Saddam Hijazi, Andrea Mola, Stefano Lorenzi, and Gianluigi Rozza. POD-Galerkin reduced order methods for CFD using Finite Volume Discretisation: vortex shedding around a circular cylinder. *Communications in Applied and Industrial Mathematics*, 8(1):210–236, (2017). doi: 10.1515/caim-2017-0011.

[126] K. Carlberg, Y. Choi, and S. Sargsyan. Conservative model reduction for finite-volume models. *Journal of Computational Physics*, 371:280–314, 2018. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.05.

019. URL https://www.sciencedirect.com/science/article/pii/S002199911830319X.

[127] G. Stabile and G. Rozza. Finite volume pod-galerkin stabilised reduced order methods for the parametrised incompressible navier–stokes equations. *Computers & Fluids*, 173:273–284, 2018. ISSN 0045-7930. doi: https://doi.org/10.1016/j.compfluid.2018.01.035. URL https://www.sciencedirect.com/science/article/pii/S0045793018300422.

[128] B. Sanderse. Non-linearly stable reduced-order models for incompressible flow with energy-conserving finite volume methods. *Journal of Computational Physics*, 421:109736, 2020. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2020.109736. URL https://www.sciencedirect.com/science/article/pii/S0021999120305106.

[129] Aaron Towne. Space-time galerkin projection via spectral proper orthogonal decomposition and resolvent modes. In *AIAA Scitech 2021 Forum*. American Institute of Aeronautics and Astronautics, jan 2021. doi: 10.2514/6.2021-1676.

[130] Hugo Steinhaus et al. Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci*, 1(804):801, 1956.

[131] James B. MacQueen. Some methods for classification and analysis of multivariate observations. 1:281–297, 1967.

[132] Stuart P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129 – 137, 1982. doi: 10.1109/TIT.1982.1056489.

[133] Moritz Sieber, Christian Oliver Paschereit, and Kilian Oberleithner. Advanced Identification of Coherent Structures in Swirl-Stabilized Combustors. *Journal of Engineering for Gas Turbines and Power*, 139(2):021503, 09 2016. doi: 10.1115/1.4034261.

[134] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, and Steven L. Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, 38(3):63–86, 2018. doi: 10.1109/MCS.2018.2810460.

[135] H. Li, D. Fernex, R. Semaan, J. Tan, M. Morzyński, and B. R. Noack. Cluster-based network model. *J. Fluid Mech.*, 906, 2021.

[136] Jon Dattorro. *Convex optimization & Euclidean distance geometry*. Meboo, Palo Alto, Calif., version 2008.02.29 edition, 2008. ISBN 9780615193687.

[137] Moritz Sieber, C. Oliver Paschereit, and Kilian Oberleithner. Spectral proper orthogonal decomposition. *Journal of Fluid Mechanics*, 792: 798–828, 2016. doi: 10.1017/jfm.2016.103.

[138] Dan Pelleg and Andrew Moore. X-means: Extending k-means with efficient estimation of the number of clusters. *Machine Learning, p*, 01 2002.

[139] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[140] M.B. Priestley. *Spectral Analysis and Time Series*. Number v. 1-2 in Probability and mathematical statistics : A series of monographs and textbooks. Academic Press, 1981. URL https://books.google.it/books?id=RVTYvwEACAAJ.

[141] Sadanori Konishi. Information criteria and statistical modeling, 2008. Includes bibliographical references (p. [255]-267) and index.

[142] Ernst Wit, Edwin van den Heuvel, and Jan-Willem Romeijn. 'all models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236, jul 2012. doi: 10.1111/j.1467-9574.2012.00530.x.

[143] George Haller. Lagrangian coherent structures. *Annual Review of Fluid Mechanics*, 47(1):137–162, January 2015. ISSN 1545-4479. doi: 10.1146/annurev-fluid-010313-141322.

[144] G. Haller. Lagrangian structures and the rate of strain in a partition of two-dimensional turbulence. *Physics of Fluids*, 13(11):3365–3385, November 2001. ISSN 1089-7666. doi: 10.1063/1.1403336.

[145] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, jan 2000. doi: 10.1137/s0036144500371907.

[146] Elena Loli Piccolomini and Fabiana Zama. Monitoring italian COVID-19 spread by a forced SEIRD model. *PLOS ONE*, 15(8):e0237417, aug 2020. doi: 10.1371/journal.pone.0237417.

[147] C. Petz, H.-C. Hege, K. Oberleithner, M. Sieber, C. N. Nayeri, C. O. Paschereit, I. Wygnanski, and B. R. Noack. Global modes in a swirling jet undergoing vortex breakdown. *Physics of Fluids*, 23(9), sep 2011. doi: 10.1063/1.3640007.

[148] K. Oberleithner, M. Sieber, C. N. Nayeri, C. O. Paschereit, C. Petz, H.-C. Hege, B. R. Noack, and I. Wygnanski. Three-dimensional coherent structures in a swirling jet undergoing vortex breakdown: stability analysis and empirical mode construction. *Journal of Fluid Mechanics*, 679:383–414, 2011. doi: 10.1017/jfm.2011.141.

[149] Sébastien Candel, Daniel Durox, Thierry Schuller, Jean-François Bourgouin, and Jonas P. Moeck. Dynamics of swirling flames. *Annual Review of Fluid Mechanics*, 46(1):147–173, 2014. doi: 10.1146/annurev-fluid-010313-141300.

[150] Nicholas Syred. A review of oscillation mechanisms and the role of the precessing vortex core (pvc) in swirl combustion systems. *Progress in Energy and Combustion Science*, 32(2):93–161, 2006. ISSN 0360-1285. doi: https://doi.org/10.1016/j.pecs.2005.10.002. URL https://www.sciencedirect.com/science/article/pii/S0360128505000353.

[151] Kilian Oberleithner, Michael Stöhr, Seong Ho Im, Christoph M. Arndt, and Adam M. Steinberg. Formation and flame-induced suppression of the precessing vortex core in a swirl combustor: Experiments and linear stability analysis. *Combustion and Flame*, 162:3100–3114, 2015. ISSN 15562921. doi: 10.1016/j.combustflame.2015.02.015.

[152] Michael Stöhr, Kilian Oberleithner, Moritz Sieber, Zhiyao Yin, and Wolfgang Meier. Experimental study of transient mechanisms of bistable flame shape transitions in a swirl combustor. *Journal of Engineering for Gas Turbines and Power*, 140, 2018. ISSN 15288919. doi: 10.1115/1.4037724.

[153] Finn Lückoff, Moritz Sieber, Christian Oliver Paschereit, and Kilian Oberleithner. Impact of the Precessing Vortex Core on NOx Emissions in

Premixed Swirl-Stabilized Flames—An Experimental Study. *Journal of Engineering for Gas Turbines and Power*, 142(11):111010, 10 2020. ISSN 0742-4795. doi: 10.1115/1.4048603.

[154] Qiang An, Wing Yin Kwong, Benjamin D. Geraedts, and Adam M. Steinberg. Coupled dynamics of lift-off and precessing vortex core formation in swirl flames. *Combustion and Flame*, 168:228–239, jun 2016. doi: 10.1016/j.combustflame.2016.03.011.

[155] Finn Lückoff and Kilian Oberleithner. Excitation of the precessing vortex core by active flow control to suppress thermoacoustic instabilities in swirl flames. *International Journal of Spray and Combustion Dynamics*, 11:1756827719856237, 2019. doi: 10.1177/1756827719856237.

[156] J. S. Müller, F. Lückoff, P. Paredes, V. Theofilis, and K. Oberleithner. Receptivity of the turbulent precessing vortex core: synchronization experiments and global adjoint linear stability analysis. *Journal of Fluid Mechanics*, 888, feb 2020. doi: 10.1017/jfm.2019.1063.

[157] Finn Lückoff, Moritz Sieber, Christian Oliver Paschereit, and Kilian Oberleithner. Phase-opposition control of the precessing vortex core in turbulent swirl flames for investigation of mixing and flame stability. *Journal of Engineering for Gas Turbines and Power*, 141(11), sep 2019. doi: 10.1115/1.4044469.

[158] Bruno Schuermans, Felix Guethe, and Wolfgang Mohr. Optical transfer function measurements for technically premixed flames. *Journal of Engineering for Gas Turbines and Power*, 132(8), May 2010. ISSN 1528-8919. doi: 10.1115/1.3124663.

[159] Finn Lückoff, Thomas Ludwig Kaiser, Christian Oliver Paschereit, and Kilian Oberleithner. Mean field coupling mechanisms explaining the impact of the precessing vortex core on the flame transfer function. *Combustion and Flame*, 223:254–266, jan 2021. doi: 10.1016/j.combustflame.2020.09.019.

[160] Georg Pretzier. A new method for numerical abel-inversion. *Zeitschrift für Naturforschung A*, 46(7):639–641, jul 1991. doi: 10.1515/zna-1991-0715.

[161] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, jun 1967. doi: 10.1109/tau.1967.1161901.

[162] Bartosz Protas, Bernd R. Noack, and Jan Östh. Optimal nonlinear eddy viscosity in galerkin models of turbulent flows. *Journal of Fluid Mechanics*, 766:337–367, 2015. doi: 10.1017/jfm.2015.14.

[163] Giovanni Stabile and Gianluigi Rozza. Finite volume POD-Galerkin stabilised reduced order methods for the parametrised incompressible Navier-Stokes equations. *Computers & Fluids*, 2018. doi: 10.1016/j.compfluid.2018.01.035.

[164] Nil Couto and Josep M. Bergada. Aerodynamic efficiency improvement on a naca-8412 airfoil via active flow control implementation. *Applied Sciences (Switzerland)*, 12(9), 2022. doi: 10.3390/app12094269.

[165] M. Chiatto, F. Capuano, and L. de Luca. Numerical and experimental characterization of a double-orifice synthetic jet actuator. *Meccanica*, 53 (11-12):2883 – 2896, 2018. doi: 10.1007/s11012-018-0866-7.

[166] M. Chiatto, A. Palumbo, and L. de Luca. Design approach to predict synthetic jet formation and resonance amplifications. *Experimental Thermal and Fluid Science*, 107:79 – 87, 2019. doi: 10.1016/j.expthermflusci.2019.05.013.

[167] Giuseppe Ceglia, Matteo Chiatto, Carlo Salvatore Greco, Fabrizio De Gregorio, Gennaro Cardone, and Luigi de Luca. Active control of separated flow over 2d back-facing ramp by an array of finite-span slotted synthetic jets. *Experimental Thermal and Fluid Science*, 129:110475, 2021. ISSN 0894-1777. doi: https://doi.org/10.1016/j.expthermflusci.2021.110475.

[168] M. Girfoglio, C.S. Greco, M. Chiatto, and L. de Luca. Modelling of efficiency of synthetic jet actuators. *Sensors and Actuators, A: Physical*, 233:512 – 521, 2015. doi: 10.1016/j.sna.2015.07.030.

[169] L. de Luca, M. Girfoglio, M. Chiatto, and G. Coppola. Scaling properties of resonant cavities driven by piezo-electric actuators. *Sensors and Actuators, A: Physical*, 247:465 – 474, 2016. doi: 10.1016/j.sna.2016.06.016.

[170] N.M. Tousi, M. Coma, J.M. Bergadà, J. Pons-Prats, F. Mellibovsky, and G. Bugeda. Active flow control optimisation on sd7003 airfoil at pre and post-stall angles of attack using synthetic jets. *Applied Mathematical Modelling*, 98:435 – 464, 2021. doi: 10.1016/j.apm.2021.05.016.

[171] N.M. Tousi, J.M. Bergadà, and F. Mellibovsky. Large eddy simulation of optimal synthetic jet actuation on a sd7003 airfoil in post-stall conditions. *Aerospace Science and Technology*, 127, 2022. doi: 10.1016/j.ast.2022.107679.

[172] Saddam Hijazi, Giovanni Stabile, Andrea Mola, and Gianluigi Rozza. Data-driven pod-galerkin reduced order model for turbulent flows. *Journal of Computational Physics*, 416:109513, 2020. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2020.109513.

[173] Francesco Ballarin, Andrea Manzoni, Alfio Quarteroni, and Gianluigi Rozza. Supremizer stabilization of pod-galerkin approximation of parametrized steady incompressible navier-stokes equations. *International Journal for Numerical Methods in Engineering*, 102(5):1136 – 1161, 2015. doi: 10.1002/nme.4772.

[174] Julien Jacques and Cristian Preda. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3):231–255, nov 2013. doi: 10.1007/s11634-013-0158-y.

[175] Frank Nielsen. Hierarchical clustering. In *Introduction to HPC with MPI for Data Science*, pages 195–211. Springer International Publishing, 2016. doi: 10.1007/978-3-319-21903-5_8.

[176] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, aug 2007. doi: 10.1007/s11222-007-9033-z.

[177] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, page 1027–1035, USA, 2007. Society for Industrial and Applied Mathematics. ISBN 9780898716245.

[178] James O. Ramsay and Bernard W. Silverman. *Functional data analysis*. Springer series in statistics. Springer, New York, NY, 2. ed., [nachdr.] edition, 2006. ISBN 9780387400808.

[179] C. Abraham, P. A. Cornillon, E. Matzner-Løber, and N. Molinari. Unsupervised curve clustering using b-splines. *Scandinavian Journal of Statistics*, 30(3):581–595, aug 2003. doi: 10.1111/1467-9469.00350.

[180] M. Giacofci, S. Lambert-Lacroix, G. Marot, and F. Picard. Wavelet-based clustering for mixed-effects functional models in high dimension. *Biometrics*, 69(1):31–40, feb 2013. doi: 10.1111/j.1541-0420.2012.01828.x.

[181] David J. Hand. An introduction to statistical concepts, third edition by richard g. lomax, debbie l. hahs-vaughn. *International Statistical Review*, 80(3):489–491, nov 2012. doi: 10.1111/j.1751-5823.2012.00196_21.x.

[182] Dirk M. Luchtenburg, Bert Günther, Bernd R. Noack, Rudibert King, and Gilead Tadmor. A generalized mean-field model of the natural and high-frequency actuated flow around a high-lift configuration. *Journal of Fluid Mechanics*, 623:283–316, mar 2009. doi: 10.1017/s0022112008004965.

[183] Nan Deng, Bernd R. Noack, Marek Morzyński, and Luc R. Pastur. Low-order model for successive bifurcations of the fluidic pinball. *Journal of Fluid Mechanics*, 884, dec 2019. doi: 10.1017/jfm.2019.959.

[184] Nan Deng, Bernd R. Noack, Marek Morzyński, and Luc R. Pastur. Galerkin force model for transient and post-transient dynamics of the fluidic pinball. *Journal of Fluid Mechanics*, 918, may 2021. doi: 10.1017/jfm.2021.299.

[185] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52:477 – 508, 2020. doi: 10.1146/annurev-fluid-010719-060214.

[186] E. Farzamnik, A. Ianiro, S. Discetti, N. Deng, K. Oberleithner, B.R. Noack, and V. Guerrero. From snapshots to manifolds – a tale of shear flows. *Journal of Fluid Mechanics*, 955, jan 2023. doi: 10.1017/jfm.2022.1039.

[187] Guy Y. Cornejo Maceda, Yiqing Li, François Lusseyran, Marek Morzyński, and Bernd R. Noack. Stabilization of the fluidic pinball with gradient-enriched machine learning control. *Journal of Fluid Mechanics*, 917, apr 2021. doi: 10.1017/jfm.2021.301.

[188] C. Taylor and P. Hood. A numerical solution of the navier-stokes equations using the finite element technique. *Computers & Fluids*, 1(1):73–100, jan 1973. doi: 10.1016/0045-7930(73)90027-3.

[189] B. R. Noack and M. Morzyński. The fluidic pinball—a toolkit for multiple-input multiple-output flow control (version 1.0). Technical report, Tech. Rep. 02/2017. Chair of Virtual Engineering, Poznan University of Technology, Poland., 2017.

[190] H. B. Squire. Investigation of the instability of a moving liquid film. *British Journal of Applied Physics*, 4(6):167–169, 1953.

[191] D. R. Brown. A study of the behaviour of a thin sheet of moving liquid. *Journal of Fluid Mechanics*, 10:297–305, 1961.

[192] S. J. Weinstein and K. J. Ruschak. Coating flows. *Annual Review of Fluid Mechanics*, 36:29–53, 2004.

[193] L. D. Soderberg and P. H. Alfredsson. Experimental and theoretical stability investigations of plane liquid jets. *European Journal of Mechanics - B/Fluids*, 17:689–737, 1998.

[194] D. L. Chubb, F. D. Calfo, M. W. McConley, M. S. McMaster, and A. A. Afjeh. Geometry of thin liquid sheet flows. *AIAA Journal*, 32:1325–1328, 1994.

[195] M. Lodomez, B. M. Crookston, B. P. Tullis, and S. Erpicum. Mitigation techniques for nappe oscillations on free-overfall structures. *Journal of Hydraulic Engineering*, 145(2):04018086, 2019.

[196] N. S. Barlow, S. J. Weinstein, and B. T. Helenbrook. On the response of convectively unstable flows to oscillatory forcing with application to liquid sheets. *Journal of Fluid Mechanics*, 699:115–152, 2012.

[197] S. J. Weinstein, D. S. Ross, J. R. Kenneth, and N. S. Barlow. On oblique liquid curtains. *Journal of Fluid Mechanics*, 876:1–9, 2019.

[198] D. J. Duke, D. Honnery, and J. Soria. The growth of instabilities in annular liquid sheets. *Experimental Thermal and Fluid Science*, 68:89–99, 2015.

[199] A. Zandian, W. A. Sirignano, and F. Hussain. Planar liquid jet: early deformation and atomization cascades. *Physics of Fluids*, 29(6):062109, 2017.

[200] H. Lhuissier, P. Brunet, and S. Dorbolo. Blowing a liquid curtain. *Journal of Fluid Mechanics*, 795:784–807, 2016.

[201] L. de Luca and M. Costa. Instability of a spatially developing liquid sheet. *Journal of Fluid Mechanics*, 331:127–144, 1997.

[202] O. Tammisola, A. Sasaki, F. Lundell, M. Matsubara, and L.D. Soderberg. Stabilizing effect of surrounding gas flow on a plane liquid sheet. *Journal of Fluid Mechanics*, 672:5–32, 2011.

[203] F. De Rosa, M. Girfoglio, and L. de Luca. Global dynamics analysis of nappe oscillation. *Physics of Fluids*, 26(12):122109, 2014.

[204] N. S. Barlow, B. T. Helenbrook, and S. P. Lin. Transience to instability in a liquid sheet. *Journal of Fluid Mechanics*, 666:358–390, 2011.

[205] L. de Luca. Experimental investigation of the global instability of plane sheet flows. *Journal of Fluid Mechanics*, 399:355–376, 1999.

[206] D. S. Finnicum, S. J. Weinstein, and K. J. Rushak. The effect of applied pressure on the shape of a two-dimensional liquid curtain falling under the influence of gravity. *Journal of Fluid Mechanics*, 255:647–665, 1993.

[207] S. P. Lin and W. Y. Jiang. Absolute and convective instability of a radially expanding liquid sheet. *Physics of Fluids*, 15(6):1745–1754, 2003.

[208] J.S. Roche, N. Le Grand, P. Brunet, L. Lebon, and L. Limat. Perturbations on a liquid curtain near break-up: wakes and free edges. *Physics of Fluids*, 18(8):082101, 2006.

[209] A. Kacem, Y. Le Guer, K. El Omari, and P. Bruel. Experimental investigations of planar water sheets flowing under gravity. *WIT Transactions on Engineering Sciences*, 115:97–107, 2017.

[210] S. J. Weinstein, A. Clarke, A.G. Moon, and E.A. Simister. Time-dependent equations governing the shape of a two-dimensional liquid curtain, part 1: Theory. *Physics of Fluids*, 9(12):3625–3636, 1997.

[211] S. Erpicum, B. Dewals, P. Archambeau, M. Pirotton, and B. Blancher. Discussion of “Modeling and prototype testing of flows over flip-bucket aerators"; by Penghua Teng and James Yang. *Journal of Hydraulic Engineering*, 146(5):07020005, 2020.

[212] R. Scardovelli and S. Zaleski. Direct numerical simulation of free-surface and interfacial flow. *Annual Review of Fluid Mechanics*, 31(1):567–603, 1999.

[213] C. W. Hirt and B. D. Nichols. Volume of fluid (VOF) method for the dynamics of free boundaries. *Journal of Computational Physics*, 39(1): 201–225, 1981.

[214] J. A. van Hooft, S. Popinet, C. C. van Heerwaarden, S. J. A. van der Linden, S. R. de Roode, and B. J. H. van de Wiel. Towards adaptive grids for atmospheric boundary-layer simulations. *Boundary-Layer Meteorology*, 167(3):421–443, 2018.

[215] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, jul 1998. doi: 10.1162/089976698300017467.