33rd Symposium on Naval Hydrodynamics
Osaka, Japan, 18-23 October 2020

Observing PIV Measurements Through the Lens of Data Clustering

D. D'Agostino,[1,2] M. Andre,[3] P. Bardet,[3] A. Serani,[2] M. Felli[2] and M. Diez[2]

([1]Sapienza University of Rome, Rome, Italy,
[2]CNR–INM, Natl. Research Council–Institute of Marine Engineering, Italy,
[3]The George Washington University, Washington DC, USA)

ABSTRACT

Spatial and snapshot clustering approaches are presented and discussed for particle image velocimetry (PIV) data of high-Reynolds number uniform and buoyant jets and 4- and 7-bladed propeller wakes respectively. Data clustering is based on the k-means algorithm, along with the identification of the optimal number of clusters based on three metrics, namely the within-cluster sum of squares, average silhouette, and number of proper orthogonal decomposition (POD) modes required to resolve a desired variance. Spatial clustering for jets flow is based on three sets of clustering variables, namely cross-section velocity profiles, point-wise energy spectra, and point-wise Reynolds stress tensor components. Snapshot clustering of phase-locked propellers wake data is based on the vorticity with focus on tip vortices regions. POD and t-distributed stochastic neighbor embedding along with kernel density estimation are used to provide a two-dimensional visualization of data clusters for assessment and discussion. The objective of this work is to lay the ground for a systematic data-clustering analysis of PIV data. The examples discussed show how clustering methods can help in achieving physical insights of complex fluid dynamics problems.

INTRODUCTION

Data science has emerged as a cutting-edge research field that develops and applies rigorous methods and algorithms to gain knowledge from data. In recent years, data science methods have been applied in the context of more traditional disciplines to accelerate the experimental/computational analysis process and extract insights from experiments and simulations data.

In ship hydrodynamics, supervised machine learning methods (including, e.g., metamodeling, regression, multi-fidelity methods) have been used to integrate experiments and simulations data and reduce the computational cost of design performance assessment and optimization procedures. Unsupervised machine learning methods (such as proper orthogonal decomposition formulations, POD, along with linear/nonlinear principal component analysis, PCA) have been used to reduce the problem dimensionality, plan more efficient design- and/or operational-space explorations, and gain insights of complex physical phenomena. POD has been widely used for the identification of coherent structures in turbulent flows (Berkooz et al., 1993) and applied to steady/transient uniform/buoyant jets (Gordeyev and Thomas, 2002; Zhou and Hitt, 2004) and marine propeller wakes (Felli and Falchi, 2018a). In the discrete form, POD is equivalent to PCA and allows to decompose the flow into a linear combination of a subset of orthogonal eigenfunctions, capable of highlighting its spatial/temporal structure and providing a reduced-order/dimensionality model for the flow dynamics. Although POD is a widely used and has well-established global optimality properties, it is based on the linearity, stationarity, and ergodicity assumptions and may not be effective when nonlinear, transient, non-stationary, non-ergodic dynamics are investigated. For this reason nonlinear dimensionality reduction (NLDR) methods have been developed and applied to provide with a deeper understanding of data structures and physical phenomena. A straightforward approach to NLDR with POD/PCA is to use data clustering methods and perform POD/PCA within each cluster.

Cluster-based reduced-order and/or dimensionality modeling by POD/PCA can provide with physical insight of complex phenomena and is achieved by local PCA (LPCA), where the data set is divided into clusters and POD/PCA is applied to each cluster, assuming therefore an approximate linear structure within each cluster. The cluster centroids along with the associated modes are used to extract relevant flow feature in the spa-

tial/temporal domains. Applications of spatial clustering via the k-means method with POD/PCA have been presented in Serani et al. (2019) for a transient buoyant jet. POD/PCA approaches based on temporal and spatial k-means clustering have been presented in Barwey et al. (2020) for a swirl-stabilized combustor flow. In general, the number of clusters and the similarity metrics used for data clustering highly affect the quality of the resulting reduced-order/dimensionality model, and therefore the possibility to gain valuable physical knowledge from the clustering analysis. Rigorous data-clustering methods can assist in achieving a deeper understanding of experiments/simulations data and have been proposed in different fields, such as computer vision and speech recognition. A further step towards fully nonlinear dimensionality reduction of data sets has been proposed via t-distributed stochastic neighbor embedding (t-SNE) by van der Maaten and Hinton (2008). The method provides with the capability of embedding and visualizing high-dimensional data in a low-dimensional space and has been applied to turbulence data sets from simulations in Wu et al. (2017).

The objective of the present work is to lay the ground for a systematic data-clustering analysis of particle image velocimetry (PIV) data with the aim of achieving physical insights of complex fluid dynamics problems. Examples are provided for high-Reynolds number uniform/buoyant transient jets along with 4- and 7-bladed propeller wakes.

The velocity fields under investigation (both for jets and propellers) are obtained from experimental tests with large scale, time/phase-resolved, PIV measurements. GWU provided data for the jets, whereas data for the propeller wake were collected at CNR-INM. Clustering of PIV data is based on the k-means algorithm. Three clustering approaches are applied to the jet in the spatial domain to identify coherent/self-similar spatial regions using the following clustering variables: (a) cross-section velocity profiles, (b) point-wise energy spectra, and (c) point-wise Reynolds stress tensor components. The resulting clusters and centroids are representative of the local flow, in terms of cross-section profiles and turbulence variables. Data clustering for the propeller wake is applied to phase-locked snapshots to gain knowledge on the topology of wake-instability and its stochastic realizations. The vorticity is used as clustering variable. The resulting cluster centroids identify the topology of the instability, where two or more tip vortices interact and coalesce.

Three metrics are proposed for the identification and assessment of clustering methods, including the selection of the proper number of clusters, namely: (a) within-cluster sum of squares, (b) average silhouette, and (c) within-cluster number of POD modes required to resolve prescribed levels of total variance/energy. Additionally, embedding of data via POD/PCA and t-SNE is used to define and visualize data clusters in a reduced dimensionality space. Finally the kernel density estimation (KDE) is applied to POD/PCA and t-SNE representations to provide with continuous data distributions for assessment and discussion. A summary of test cases and clustering approaches used is presented in Table 1.

DATA ANALYSIS METHODS

Data analysis methods are outlined in the following. Specifically, the k-means clustering method is briefly recalled. POD/PCA implementation is described and used both as metric for identifying an optimal number of clusters and a visualization technique. For the same purpose, the t-SNE method is used as visualization technique and briefly discussed along with KDE.

k-Means Clustering

The k-means is a widely used clustering method (Jain, 2010), which allows to build partitions of the original data collected in an $[L \times S]$ matrix $\mathbf{U}$ in k different sets (clusters), defined by representative points (centroids). Here, the original data in $\mathbf{U}$ is rearranged as per the clustering approach and criterion used. In general, we refer to $\boldsymbol{\xi}_j$ as one realization (point) of the rearranged data. Note that, generally, $\boldsymbol{\xi}_j \in \mathbb{R}^Q$ with $j = 1, ..., H$, where $Q \neq L$ and $H \neq S$.

The Euclidean distance is used to measure both the similarity between data points $\boldsymbol{\xi}_j$ and evaluate the associated cluster centroids $\boldsymbol{\mu}_i$, the latter by averaging all data points within the i-th cluster $\mathbf{K}_i$. The assignment of data points to k clusters is achieved by the minimization of the squared Euclidean distance between $\boldsymbol{\xi}_j$ and $\boldsymbol{\mu}_i$ (within-cluster sum of squares, WCSS)

$$\mathrm{WCSS} = \sum_{i=1}^{k} \sum_{\boldsymbol{\xi}_j \in \mathbf{K}_i} \|\boldsymbol{\xi}_j - \boldsymbol{\mu}_i\|^2 \qquad (1)$$

The minimization of WCSS in Eq. 1 is a NP-hard problem (Drineas et al., 2004). For its solution, the heuristic approach presented in Lloyd (1982) is used and briefly presented in Alg. 1. Results are highly sensitive to centroids initialization. Here, the initialization strategy proposed in Arthur and Vassilvitskii (2007) is used.

Proper Orthogonal Decomposition/Principal Component Analysis

For the jet analysis, Reynolds decomposition of the velocity vector is used, with

$$u = \overline{u} + u' \qquad v = \overline{v} + v' \qquad (2)$$

Table 1: Summary of test cases and clustering approaches.

Test case	Uniform/buoyant jets	4/7-bladed propeller wakes
Data dimension	73,678 × 4,600 / 3,141	6,000 × 1,000 / 10,000 × 500
Clustering domain	Spatial	Snapshots
Clustering variables	Cross-section velocity profiles Point-wise energy spectra Point-wise Reynolds stress	Vorticity
Clustering metrics and assessment	Within cluster sum of squares Average silhouette Number of POD/PCA components	Within cluster sum of squares Average ilhouette POD/PCA embedding and KDE t-SNE embedding and KDE

where u and v indicate x- (axial/vertical) and y- (horizontal) components, respectively. Overbar and prime characters indicate time average and fluctuations, respectively.

The POD is implemented as follows. The $[L \times S]$ data matrix $\mathbf{U}$ is defined as

$$\mathbf{U} = \left[\begin{array}{c|c|c} \mathbf{u}^{(1)} & \ldots & \mathbf{u}^{(S)} \end{array} \right] \tag{3}$$

where $\mathbf{u} = \{u'(\mathbf{x}_1), \ldots, u'(\mathbf{x}_P), v'(\mathbf{x}_1), \ldots, v'(\mathbf{x}_P)\}^\mathsf{T}$ collects the discretized velocity fluctuations, $\mathbf{x}_i$ represents the i-th node of the spatial discretization, P is the spatial discretization size, superscript (i), with $i = 1, \ldots, S$, indicates the i-th time realization (snapshot), and finally $L = 2P$. The data matrix $\mathbf{U}$ is reduced in dimensionality through projection of the snapshots into a new linear subspace, formed by the eigenvectors of the $[L \times L]$ covariance matrix (Bishop, 2006)

$$\mathbf{C} = \frac{1}{S}\mathbf{U}\,\mathbf{U}^\mathsf{T} \tag{4}$$

evaluated by

$$\mathbf{C}\mathbf{Z} = \mathbf{Z}\mathbf{\Lambda} \tag{5}$$

where $\mathbf{Z}$ and $\mathbf{\Lambda}$ collect the L eigenvectors ($\mathbf{z}_i$) and eigenvalues (λ_i) of $\mathbf{C}$, respectively. This corresponds to performing the PCA of the matrix $\mathbf{U}$. The problem of Eq. 5 may be alternatively solved using the singular value decomposition, SVD (Golub and Reinsch, 1970). Furthermore, if $L > S$ the dual problem may be solved via the so-called snapshot-POD (equivalent) formulation, see e.g. Cizmas et al. (2003).

POD/PCA eigenvalues represent the variance (under proper assumptions this represents the turbulent kinetic energy) resolved along the corresponding eigenvectors. The linear subspace formed by the N eigenvectors (POD/PCA modes, collected in $\hat{\mathbf{Z}}$) associated to the largest N eigenvalues resolves (globally) the largest variance/energy, compared to any other linear subspace of dimension N (Bishop, 2006; Holmes et al., 2012). The cumulative sum of the eigenvalues is used to assess the variance resolved by the linear subspace of dimension N. Finally, the associated reconstruction of $\mathbf{U}$ is given by $\hat{\mathbf{U}} = \hat{\mathbf{Z}}\hat{\mathbf{Z}}^\mathsf{T}\mathbf{U}$, where by definition the coefficient $\boldsymbol{\alpha}_i = \hat{\mathbf{z}}_i^\mathsf{T}\mathbf{U}$ is the projection of the data matrix onto the i-th mode.

POD/PCA implementation for the propeller wake follows the same procedure, with

$$\omega = \overline{\omega} + \omega' \tag{6}$$

where $\omega = \partial v/\partial x - \partial u/\partial y$ is the vorticity z-component (out of plane) and x and y are the axial/horizontal and vertical coordinates respectively.

The $[L \times S]$ data matrix U is defined in this case as

$$\mathbf{\Omega} = \left[\begin{array}{c|c|c} \boldsymbol{\omega}^{(1)} & \ldots & \boldsymbol{\omega}^{(S)} \end{array} \right] \tag{7}$$

where $\boldsymbol{\omega} = \{\omega'(\mathbf{x}_1), \ldots, \omega'(\mathbf{x}_P)\}^\mathsf{T}$ collects the discretized vorticity fluctuations, $\mathbf{x}_i$ represents the i-th node of the spatial discretization, P is the spatial discretization size, superscript (i), with $i = 1, \ldots, S$, indicates the i-th time realization (snapshot), and finally $L = P$.

t-Distributed Stochastic Neighbor Embedding

The t-SNE is a machine learning algorithm proposed by van der Maaten and Hinton (2008), which is found very effective for embedding high-dimensional data for visualization in a low-dimensional space of two or three dimensions. The t-SNE first constructs joint probability densities p_{ij} that reflect pairwise similarty among datapoints $\mathbf{u}^{(i)}$ and $\mathbf{u}^{(j)}$ (or $\boldsymbol{\omega}^{(i)}$ and $\boldsymbol{\omega}^{(j)}$) parameterized by a Gaussian distribution

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2S} \tag{8}$$

with

$$p_{j|i} = \frac{\exp\left(-||\mathbf{u}^{(i)} - \mathbf{u}^{(j)}||^2/2\varsigma_i^2\right)}{\sum_{k\neq i}^{S} \exp\left(-||\mathbf{u}^{(i)} - \mathbf{u}^{(k)}||^2/2\varsigma_i^2\right)} \tag{9}$$

In a similar manner the joint probability densities q_{ij} are defined for the low-dimensional representations $\boldsymbol{\alpha}_i$ and

$\boldsymbol{\alpha}_j$, parameterized by a t-student distribution

$$q_{ij} = \frac{(1 + ||\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j||^2)^{-1}}{\sum_{k \neq i}^{S} (1 + ||\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_k||^2)^{-1}} \tag{10}$$

The points (or coefficients) $\boldsymbol{\alpha}_i$ are determined by minimizing the Kullback-Leibler divergence of the distribution Q from the distribution P as

$$\mathrm{KL}(P||Q) = \sum_{i \neq j}^{S} p_{ij} \log \left(\frac{p_{ij}}{q_{ij}} \right) \tag{11}$$

The minimization of Eq. 11 with respect to the points $\boldsymbol{\alpha}_i$ is performed using gradient descent. The parameter ς_i is set in such that the perplexity of the conditional distribution $P_i = \sum_j p_{ji}$ over all data points given $\mathbf{u}^{(i)}$ equals a predefined perplexity $\mathrm{Perp}(P_i)$,

$$\mathrm{Perp}(P_i) = 2^{\left(\sum_j^S p_{j|i} \log_2 p_{j|i}\right)} \tag{12}$$

which is solved with a bisection method. The perplexity can be interpreted as a smooth measure of the effective number of neighbors, with typical values ranging from 5 and 50 (van der Maaten and Hinton, 2008).

Multivariate Kernel Density Estimation

The kernel density estimation (KDE, Silverman 2018) is a non-parametric method to estimate the probability density function (PDF) of a random variable, introduced for univariate data. Extending the concept to multivariate data (Simonoff, 2012), let $\{\boldsymbol{\alpha}\}_{i=1}^{S}$ be a d-variate random vector whose PDF is estimated as

$$\mathrm{PDF}(\boldsymbol{\alpha}) = \frac{1}{S} \sum_{i=1}^{S} K_{\mathbf{H}} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}_i \right) \tag{13}$$

where $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_d\}^{\mathsf{T}}$, $\boldsymbol{\alpha}_i = \{\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{id}\}^{\mathsf{T}}$ with $i = 1, \ldots, S$, $\mathbf{H}$ is the bandwidth (or smoothing) $[d \times d]$ matrix which is symmetric and positive definite, and K is the kernel function which is a symmetric multivariate density defined as

$$K_{\mathbf{H}}(\boldsymbol{\alpha}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2} \boldsymbol{\alpha}) \tag{14}$$

DATA ANALYSIS METRICS

Three metrics are used for the assessment of clustering approaches and identification of the optimal number of clusters k, namely: (1) within-cluster sum of squares, (2) average silhouette, and (3) within-cluster number of POD modes required to resolve prescribed levels of total variance/energy. The latter is used only for buoyant jets. Their definition is included in the following.

Within-Cluster Sum of Squares

The WCSS in Eq. 1 is used as evaluation metrics to identify the optimal number of clusters k. Specifically, the elbow method (Ketchen and Shook, 1996) is used with the WCSS metrics.

Average Silhouette

The silhouette method provides a metrics of consistency of data within clusters (Rousseeuw, 1987). Assume a_i as the average Euclidean distance between $\boldsymbol{\xi}_i$ and any other data point within the cluster $\boldsymbol{\xi}_i$ belongs to. Assume then c_i as the smallest average Euclidean distance of $\boldsymbol{\xi}_i$ to all data points in any other cluster $\boldsymbol{\xi}_i$ does not belong to.

The silhouette associated to $\boldsymbol{\xi}_i$ is defined as

$$s_i = \frac{a_i - c_i}{\max[a_i, c_i]} \tag{15}$$

and is a measure of how similar the data point is to points in its own cluster as opposed to other clusters. It may be noted that s_i ranges from -1 to 1, where 1 indicates maximum similarity.

The average silhouette of all data points is used as a metrics for proper data clustering:

$$s_{\mathrm{avg}} = \sum_{i=1}^{H} s_i \tag{16}$$

Note that for $k = 1$ the silhouette is not defined. By convention, for $k = 1$ it is $s_{\mathrm{avg}} = 0$.

Variance/Energy of POD/PCA Representation

The relative variance/energy resolved by within-cluster POD representations of order N is also used to support WCSS and average silhouette in the identification of an optimal number of clusters k:

$$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{N} \lambda_j^{(i)}}{\sum_{j=1}^{L} \lambda_j^{(\mathrm{whole})}} \tag{17}$$

where $\lambda_j^{(k)}$ is the j-th POD eigenvalue associated to the i-th cluster, provided that $\lambda_j^{(i)} > \lambda_{j+1}^{(i)}$. Superscript "whole" refers to the whole field of view (no clustering applied). It is worth noting that, for the sake of simplicity, the same order N is assumed within each cluster in Eq. 17. This assumption will be removed in future studies, where an optimal order N will be investigated within each cluster.

EXPERIMENTAL METHODS

The following subsections present the experimental setup for PIV measurements of jets and propellers under investigation.

High-Reynolds Number Uniform and Buoyant Jets

The experiment is conducted at GWU and is the vertical discharge of high-Reynolds number uniform and buoyant transient jets. The latter is discharged in a linearly stratified environment. To enable optical diagnostic deployment, two refractive index matched solutions of different densities are employed; their density difference is 3.00%. Additionally, the dynamic viscosity of the solutions are within 0.7% of each other at 20°C. Details on the refractive index and dynamic viscosity matching, as well as on the linear stratification formation, are reported in Clément et al. (2018).

The facility produces a round vertical jet inside a clear acrylic tank (cube of $914\,\mathrm{mm}$-side), Figure 1. A linear motor drives a piston in a $203\,\mathrm{mm}$-diameter cylinder, which pushes the fluid through a first contraction section followed by a contoured nozzle with a $D = 6.35\,\mathrm{mm}$ exit diameter. The fluid in the cylinder is initially at rest and the jet has no initial disturbances. The piston-cylinder and contraction sections lead to a jet with a top-hat velocity profile. The jet Reynolds number is $Re_D = U_\circ D/\nu$ is 2.00×10^4. The run time, limited by the stroke of the piston, is $39\,\mathrm{s}$. The change in height in the tank from each run is $5.4\,\mathrm{mm}$ or about $0.8D$.

During the discharge of the buoyant jet, the stratified environment evolves continuously; the flow might not reach statistical stationarity. Therefore, the whole time history of the velocity fields is recorded in a time-resolved manner: initial circulations in the tank, the entire run, and the settling down after the jet ends. Additionally, the velocity field is recorded from the jet centerline to the wall of the tank. The recorded flow area is nearly 0.7 m horizontally by 0.5 m vertically. The spatial and temporal scales vary greatly over the field of view and to optimize the acquisition system, a multi-camera array is employed. Cameras 1 to 9 are 1.3 MPixel CMOS cameras and record the off-center and far field of the jet. They record either at 64 or 128 Hz depending on their radial location. Camera 10 is a CMOS camera with CoaXpress transfer protocol. It records on the centerline of the jet at 512 Hz at 4 MPixels. The spatial resolution for those cameras is on the order of the Taylor scale. Finally, two other cameras are recording at higher resolution, but the data are not treated here.

Three large laser sheets illuminate the fields of view the cameras. They are split from a single cavity of a dual cavity Nd:YLF laser (Photonics DM 527) operated at nearly 30 mJ/pulse. The intensity of each laser sheet is controlled individually by a set of beam splitters. Each laser is configured as a telescope, with a nearly constant $3\,\mathrm{mm}$ thickness.

Data are processed with Davis 8.4.0 from LaVision. The velocity fields recorded at 64 Hz are first up-sampled to 128 Hz using Davis super-time-sampling function, and data at 512 Hz are down-sampled at 128 Hz. Once all velocity fields are sampled at the same rate, they are spatially stitched together, applying a sliding average over the areas where cameras overlap.

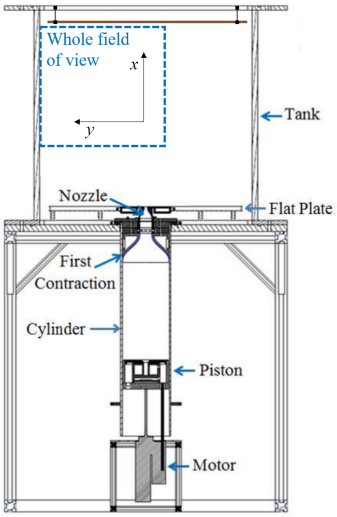


Figure 1: Experimental facility for the jets with the location of the PIV whole field of view.

Propeller Wake

The study is based on a comprehensive database of detailed flow measurements of the notional E1658 submarine propeller wake in open water using 2D-PIV (see Felli and Falchi 2018a; Posa et al. 2019). The database covers an extensive set of propeller conditions in terms of advance coefficients and blade number configurations providing a wide range of vortex instability and interaction mechanisms that are crucial for the objectives of the present study. In particular, the present study focuses on two propeller configurations with 4 and 7 blades and one value of the advance ratio, corresponding to a high propeller loading (i.e. $J = 0.56$).

The survey was carried out at the CNR-INM cavitation tunnel (i.e. 2.7 m long by 0.6 m width by 0.6 m height test section, 2% highest free-stream turbulence, mean velocity uniformity within % for the axial component and 3% for the vertical component), measuring the propeller wake flow at the vertical centerplane by a system of multiple, side-by-side, synchronous cameras, 2560×2160 pixels each, and two 200 mJ/pulse Nd-

YAG lasers. This arrangement, already adopted in other similar experiments (Felli et al., 2010; Felli and Falchi, 2018b,a) of the propeller wake, allowed the simultaneous reconstruction of a long portion of the wake flow (i.e. from the propeller plane to $3.3D$ downstream, where D is the propeller diameter) without jeopardizing the spatial resolution. More detailed information on the experimental set up are reported in Felli and Falchi (2018a). Camera acquisition was conditioned upon the passage of the propeller reference blade for a selected angular position. This was achieved by synchronizing the four cameras and the two lasers to a TTL OPR (i.e. Once Per Revolution) signal, supplied by a 3600 pulse/sec rotary incremental encoder mounted on the propeller dynamometer.

SPATIAL CLUSTERING APPROACH

Clustering approaches in the spatial domain are proposed for the turbulent transient jets, using as clustering variables (1) cross-section velocity profiles, (2) point-wise energy spectra, and (3) point-wise Reynolds stress tensor components, whereas propellers wakes are clusterized in the snapshot domain based on the vorticity only. The formulation underlying each approach is described in the following.

Cross-Section Velocity Profiles

Cross-sections are clustered together, based on their velocity profiles. Firstly, cross-section (x-constant) velocity profiles are scaled and secondly used as variables in the clustering process. Specifically, the profiles of the following variables are stitched together to form clustering arrays in Eq. 1:

$$\frac{\overline{u}}{u_c}, \qquad \frac{\left(\overline{u'u'}\right)^{1/2}}{u_{c,l}}, \qquad \frac{\left(\overline{v'v'}\right)^{1/2}}{u_{c,l}}, \qquad \frac{\overline{u'v'}}{u_{c,l}^2} \tag{18}$$

where u_c is the mean axial velocity at the center line; $u_{c,l}$ is the the mean axial velocity at the center line, assuming idealized profiles from the fully developed region where $1/u_c$ is linear (Serani et al., 2019).

Profile abscissa are scaled assuming that the velocity (positive mean axial component) profile follows a Gaussian distribution and using its standard deviation b, evaluated numerically as

$$b(x) = \sqrt{\frac{\int\limits_{y_{\min}}^{y_{\max}} (y - y_c)^2 \, \max[\overline{u}(x), 0] \, \mathrm{d}y}{\int\limits_{y_{\min}}^{y_{\max}} \max[\overline{u}(x), 0] \, \mathrm{d}y}} \tag{19}$$

where y_c is the horizontal coordinate of the center line. It may be noted that, under the Gaussian distribution assumption, the 95% of the (positive) flux is contained within $\pm 2b$. Similarly to u_c, abscissa scaling for turbulence variables in Eq. 18 is performed using idealized linearly increasing values of b from the fully developed region, referred to as b_l.

Accordingly, $\boldsymbol{\xi}_i$ (with $i = 1, ..., H$) arrays are formed as

$$\boldsymbol{\xi}_i = \left\{ \begin{array}{cc} \overline{u}(\hat{\mathbf{y}}, x_i) & [u_c(x_i)]^{-1} \\ \left[\overline{u'(\hat{\mathbf{y}}, x_i) \circ u'(\hat{\mathbf{y}}, x_i)}\right]^{1/2} & [u_{c,l}(x_i)]^{-1} \\ \left[\overline{v'(\hat{\mathbf{y}}, x_i) \circ v'(\hat{\mathbf{y}}, x_i)}\right]^{1/2} & [u_{c,l}(x_i)]^{-1} \\ \overline{u'(\hat{\mathbf{y}}, x_i) \circ v'(\hat{\mathbf{y}}, x_i)} & [u_{c,l}(x_i)]^{-2} \end{array} \right\} \in \mathbb{R}^Q \tag{20}$$

where $\hat{\mathbf{y}}$ collects discretized scaled abscissa and '$\circ$' indicates entry-wise product. Here, Q equals four times the size of $\hat{\mathbf{y}}$ and H equals the number of cross sections considered. Variables values at positions $\hat{\mathbf{y}}$ are evaluated by linear interpolation.

Point-Wise Spectra

Spatial points are clustered together based on their energy spectra. Each point is provided with the energy spectrum (Gibson, 1963; Benzi et al., 1990), namely $E(f)$, where f is the frequency. It may be noted that the use of energy spectra in the frequency domain is more convenient in the present context, where spatial clustering is sought after. Spectra are used in logarithmic scale to form clustering arrays in Eq. 1:

$$\boldsymbol{\xi}_i = \left\{ \log\left[E(\mathbf{f})\right] \right\} \in \mathbb{R}^Q \qquad i = 1, ..., H \tag{21}$$

where $\mathbf{f}$ is the vector collecting discretized frequencies, Q equals the size of $\mathbf{f}$, and $H = J$.

Point-Wise Reynolds Stress Tensor Components

Spatial points are clustered together based on their Reynolds stress tensor components. Specifically, each point is provided with Reynolds stress tensor components, which are combined together to form clustering arrays in Eq. 1:

$$\left(\overline{u'u'}\right)^{\beta}, \qquad \left(\overline{v'v'}\right)^{\beta}, \qquad \left|\overline{u'v'}\right|^{\beta} \tag{22}$$

where the exponent β is used as a tuning parameter for the clustering process, and assumed equal to $1/2$.

Accordingly, clustering variables collected in $\boldsymbol{\xi}_i$ are organized as

$$\boldsymbol{\xi}_i = \left\{ \begin{array}{c} \left[\overline{u'(\mathbf{x}_i)u'(\mathbf{x}_i)}\right]^{1/2} \\ \left[\overline{v'(\mathbf{x}_i)v'(\mathbf{x}_i)}\right]^{1/2} \\ \left|\overline{u'(\mathbf{x}_i)v'(\mathbf{x}_i)}\right|^{1/2} \end{array} \right\} \in \mathbb{R}^Q \qquad i = 1, ..., H \tag{23}$$

where $Q = 3$ and $H = P$.

SNAPSHOT CLUSTERING APPROACH

Snapshot clustering is performed for the propeller wake using vorticity snapshots following clustering variables as

$$\boldsymbol{\xi}_i = \boldsymbol{\omega}^{(i)} \qquad i = 1, ..., H \tag{24}$$

Data sets are organized in phase-locked snapshots where each phase is typically observed hundreds of times. In this case Q equals the size of the spatial discretization and $H = S$.

SPATIAL CLUSTERING: UNIFORM AND BUOYANT JETS RESULTS

The whole field of view is discretized with 197 $\times$ 187 points. Data rates of 128 and 64 Hz are used for the uniform and buoyant jet respectively. A total of 4,600 snapshots are used for the uniform jet, whereas 3,141 are used for the buoyant jet. For the purpose of current study, a spatial subsampling by a factor of 2 along each spatial direction is used. Coordinates x and y are relative to the jet virtual origin and reported as ratio with respect to the nozzle diameter D.

Mean Flow

Figure 2 shows the mean flow (left) and vorticity (right) for the uniform (top) and buoyant (bottom) jet respectively. It may be noted how impingement with the top and (to a minor extent) left walls occurs for the uniform jet, whereas the buoyant jet experiences a reverse flow still quite far from the top wall.

Clustering by Cross-Section Velocity Profiles

Figure 3a shows the mean axial velocity (in the form of $1/u_c$) at the center line and the jet width b for the uniform jet, showing a reasonable linear trend along the jet axis. Figures 3b–e show the scaled profiles of mean axial velocity (b) and root square of Reynolds stress tensor components (c–e) respectively. As expected for this type of flow, cross-Section velocity profiles scale very well with u_c and b. Profiles of the Reynolds stress tensor components scale reasonably well with $u_{c,l}$ and b_l especially if far from the top wall where impingement occurs.

Figure 4a shows the mean axial velocity (in the form of $1/u_c$) at the center line and the jet width b for the buoyant jet, showing a reasonable although quite limited linear trend in the bottom region of the jet. Figures **??**b–e show the scaled profiles of mean axial velocity (b) and root square of Reynolds stress tensor components (c–e) respectively. Velocity profiles scale reasonably well with u_c and b especially if far from the region where the flow starts reversing. Similarly, profiles of the Reynolds stress tensor components scale reasonably well with $v_{c,l}$ and b_l. It may be noted that the profiles that do not scale well are those corresponding to the jet dome and the near-wall (top) region.

Clustering of velocity profiles produces the metrics presented in Figures 5 and 6 for the uniform and buoyant jet respectively. Overall, three clusters may be identified for the uniform jet since (a) the WCSS presents a quite noticeable elbow point for $k = 3$, (b) the silhouette has a reasonably large value for $k = 3$, and finally (c) for $k = 3$ the number of POD/PCA components required to resolve the 75% or the total variance has the largest drop. It may be noted that the curve showing the 75% variance is only used as a trend. Larger (or smaller) variability thresholds can be applied similarly and will be investigated in future studies. Similar considerations apply for the buoyant jet, where five ($k = 5$) clusters are identified.

Figure 7 shows the uniform-jet cluster centroids for the mean axial velocity (a) and the profiles of Reynolds stress quantities (b–d). The corresponding cross-section clustering is presented in Figure 7e. Self-similarity of mean axial velocity profiles is reflected in the centroids (e). Significant differences emerge for the Reynolds stress, where the uu component (normal/axial) presents noticeable differences between the bottom and middle cluster (c), suggesting the turbulence fully develops transitioning from the bottom to the middle clusters. The vv component (normal/horizontal) has almost identical profiles for the bottom and middle clusters, whereas it presents a bi-modal shape for the top cluster where impingement occurs (d). Finally the uv component (shear) produces cluster centroids with significant larger values for the top cluster (e). We may conclude that three regions are identified via k-means clustering, namely bottom (developing), middle (fully developed), and top (impingement) regions.

Figure 8 shows the buoyant-jet cluster centroids for the mean axial velocity (a) and the profiles of Reynolds stress quantities (b–d). The corresponding cross-section clustering is presented in Figure 7e. Self-similarity of mean axial velocity profiles can be noticed

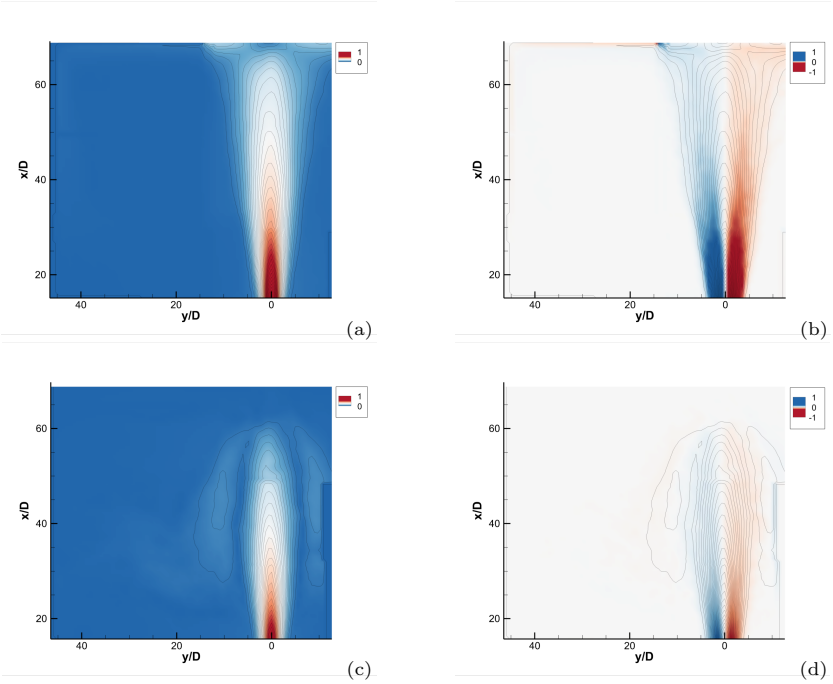


Figure 2: Mean flow for the uniform (a,b) and buoyant (c,d) jets – Velocity (a,c) and vorticity (b,d).

only for the first three clusters (from bottom to top), whereas the top two clusters present not only a remarkable reverse flow but also a different scaling properties with respect to b (a), which may be due to the presence of instabilities and very low velocity values. The uu component (normal/axial) presents noticeable differences between the bottom/middle clusters and the top three clusters (c). The vv component (normal/horizontal) has similar profiles for the bottom and middle clusters, whereas it presents a mild bi-modal shape for the top clusters, which also show reduced values (d). Finally the uv component (shear) produces cluster centroids with very small values for the two top clusters (e). We may conclude that five regions are identified via k-means clustering, namely first from bottom (developed, with almost no reverse flow), second (developed, with significant reverse flow and associated shear stress), third (flow reversing), fourth (jet dome), fifth (top wall) regions.

Clustering by Point-Wise Energy Spectra

Figure 9 shows the energy spectra associated to the whole field of view (in black) of the uniform (a) and buoyant (b) jets, where the inertial subrange is well visible with a slope equal to $-5/3$.

Clustering by point-wise energy spectra produces the metrics presented in Figures 10 and 11 for the uniform and buoyant jet respectively. The number of clusters selected for the uniform jet is 5 since for $k = 5$ (a) the WCSS shows a mild elbow point, (b) the silhouette is reasonably large, and especially (c) the number of POD/PCA components required to resolve the 75% of the variance experiences a significant drop. Similarly, the number of clusters for the buoyant jet is 4 since (a) the WCSS shows a mild elbow point, (b) the silhouette is reasonably large, and especially (c) the number of POD/PCA components required to resolve the 75% of the variance experiences a significant drop.

Cluster centroids are included in Figure 9, whereas spatial points associated to each cluster are shown in Figure 12 for the uniform (a) and buoyant (b) jets. For both jets (as expected) the cluster at the jet core retains the largest energy, showing a well visible inertial subrange with a $-5/3$ slope. Interestingly, moving from the jet core towards outer clusters identifies regions characterized by lower local Re and energy, bringing into light subregions with slope equal to -3 and -1, which may suggest the emergence of two-dimensional inertial and viscous convective subranges. Clustering results also emphasizes the differences between uniform and buoyant jets, where the former experiences a reverse flow with associated shear layer due to impingement with the top and left walls, whereas for the latest the reverse flow and shear layer are closer to the jet axis.

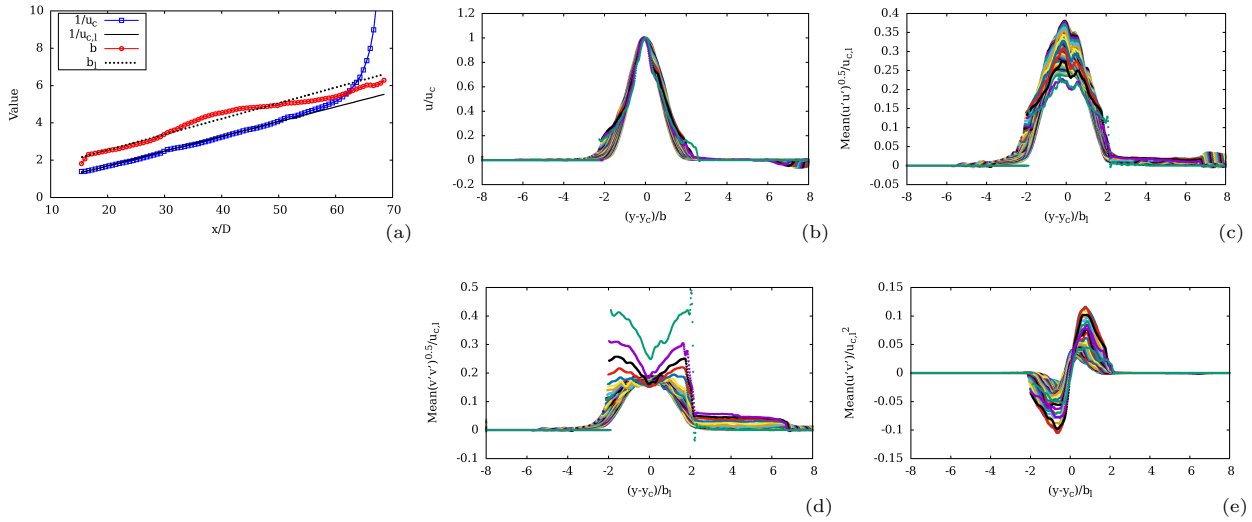


Figure 3: Clustering by cross-section velocity profiles – Uniform jet – Mean axial velocity at center line $(1/u_c)$ and jet width (b/D) along jet axis (a); scaled profiles of mean axial velocity (b) and Reynolds stress quantities, normal (c,d) and shear (e).

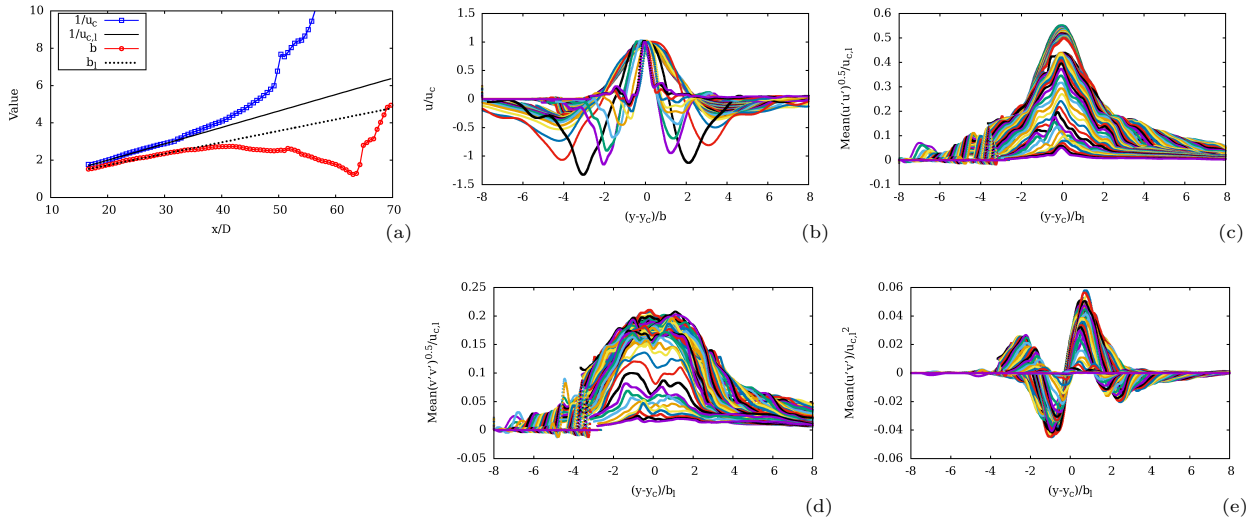


Figure 4: Clustering by cross-section velocity profiles – Buoyant jet – Mean axial velocity at center line $(1/u_c)$ and jet width (b/D) along with jet axis (a); scaled profiles of mean velocity (b) and Reynolds stress quantities, normal (c,d) and shear (e).

Clustering by Point-Wise Reynolds Stress Tensor Components

Clustering by point-wise Reynolds stress tensor components gives the metrics presented in Figures 13 and 14 for the uniform and buoyant jet respectively. A number of clusters equal to 3 is selected for both jets. Similarly to previously-shown clustering approaches, the selected number of clusters provides an elbow point for the WCSS (a), a reasonably large value for the silhouetee (b), and a visible drop in the number of POD/PCA components to resolve the 75% of the total variance (c).

Cluster centroids are shown in Figure 15, whereas spatial points associated to each cluster are shown in Figure 16. It may be observed how the clustering method defines spatial regions mainly by turbulence intensity, highlighting the differences between the two jets.

It may be noted that even if a number of clusters equal to 3 satisfies most criterion, by definition this number represents a compromise between physical resolution and complexity of the representation.

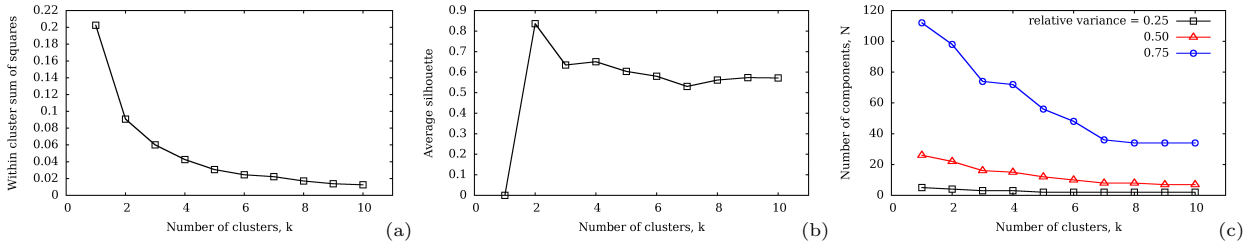


Figure 5: Clustering by cross-section velocity profiles – Uniform jet – WCSS (a), average silhouette (b), and number of POD/PCA components to resolve a prescribed relative variance (c).

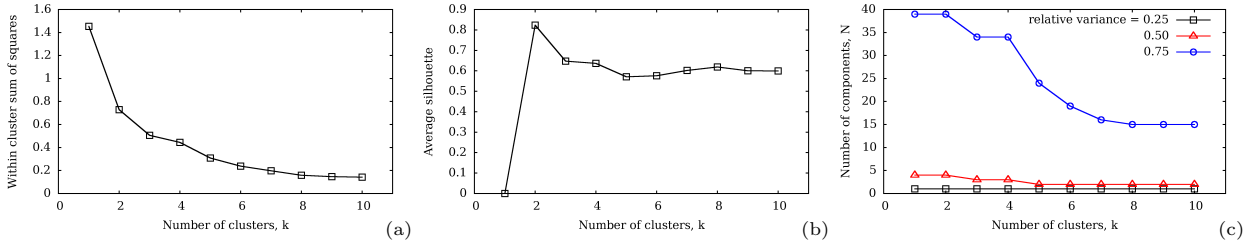


Figure 6: Clustering by cross-section velocity profiles – Buoyant jet – WCSS (a), average silhouette (b), and number of POD/PCA components to resolve a prescribed relative variance (c).

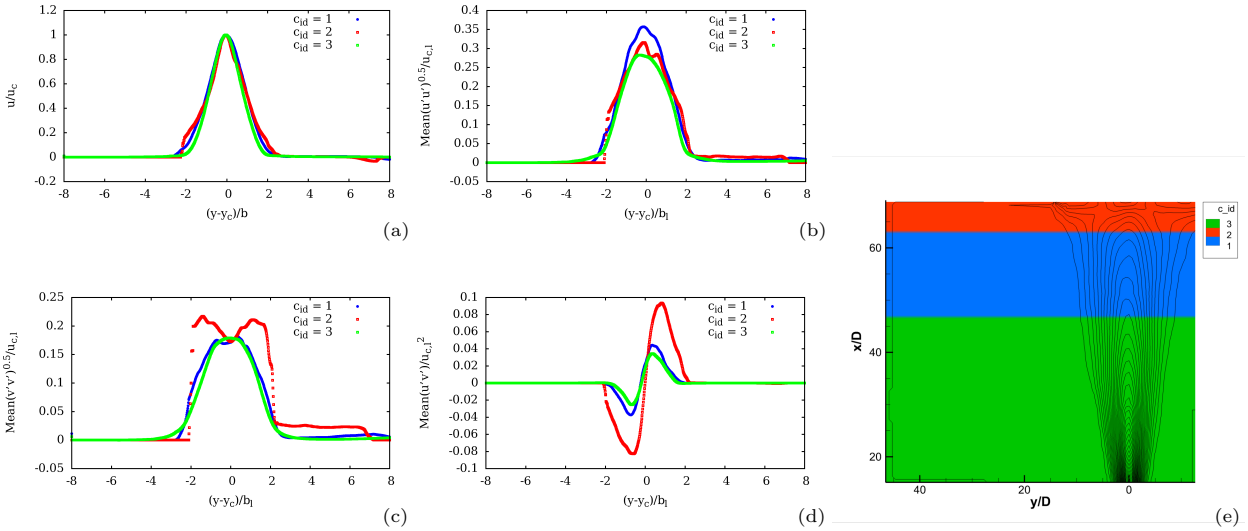


Figure 7: Clustering by cross-section velocity profiles – Uniform jet – Cluster centroids (C_{id}) for scaled mean velocity (a) Reynolds stress quantities (b,c,d); cross sections labeled by cluster (e).

SNAPSHOT CLUSTERING: PROPELLER WAKE RESULTS

4-Bladed Propeller

Four phase-locked vorticity data sets (0, 90, 180, and 270 deg) are used, where each phase is observed 250 times for a total of $1,000$ snapshots. The snapshots are organized (subsampled) in a 200×30 array, ranging axially from 0 to 3.3 D, and radially from 0.3 to 0.8 D, focusing on the tip vortex only (see Figure 17). The data matrix has a dimension equal to 6,000 $\times$ 1,000. It may be noted that once the data matrix is formed, the information on the phase is lost (as this information is not included in the data matrix).

First, the k-means is applied to the whole field of view (see Figure 18). Figure 19 shows that 4 clus-

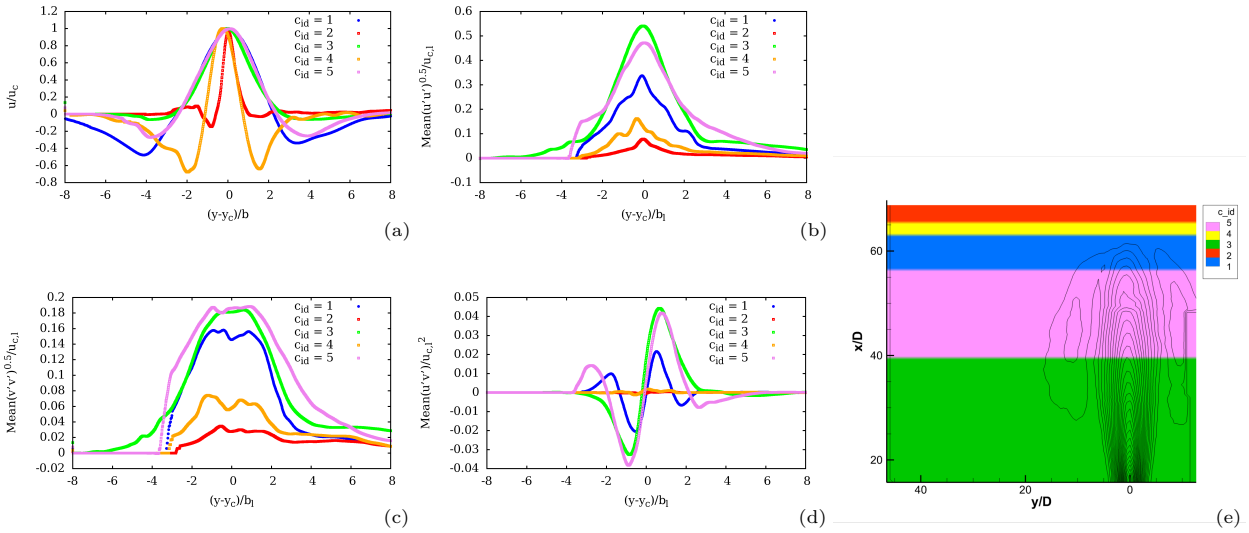


Figure 8: Clustering by cross-section velocity profiles – Buoyant jet – Cluster centroids (C_{id}) for scaled mean velocity (a) and Reynolds stress quantities (b,c,d); cross sections labeled by cluster (e).

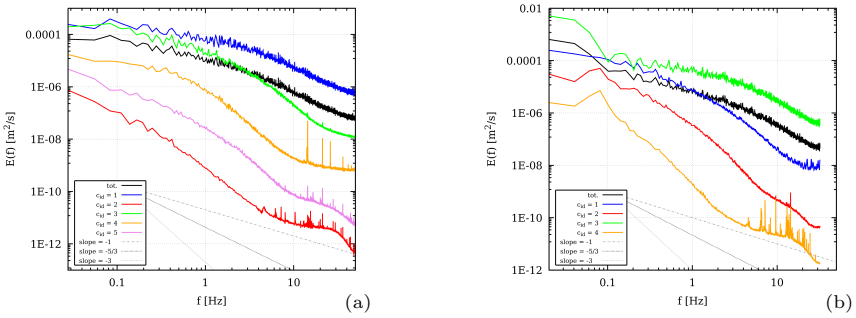


Figure 9: Clustering by point-wise energy spectra – Total energy spectrum and energy spectrum associated to cluster centroids (C_{id}) for uniform (a) and buoyant (b) jets.

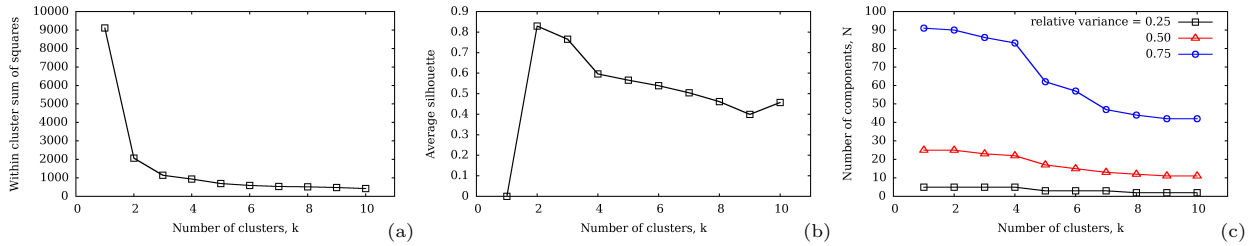


Figure 10: Clustering by point-wise energy spectra – Uniform jet – Within cluster sum of squares (a), average silhouette (b), and number of POD/PCA components to resolve a prescribed relative variance (c).

ters emerge from the data set. Specifically, WCSS shows a clear elbow corresponding to $k = 4$. The average silhouette exhibits a clear maximum corresponding to $k = 4$. Figure 20a shows the projection of the data set onto the first two POD/PCA modes. Data is labeled both by cluster and phase, showing that the method is able to recover phase information and the data set is clearly clustered by phase. A similar analysis and visualization is shown using t-SNE in Figure 20b, confirming the POD/PCA result. Finally, Figure 20c provides joint and marginal probability density functions of POD/PCA and t-SNE coefficients $\boldsymbol{\alpha}$ given by KDE, confirming the data has four clusters of equal size. The corresponding cluster centroids are presented in Figure 21, showing that the mechanism of vortex coupling and convection down-

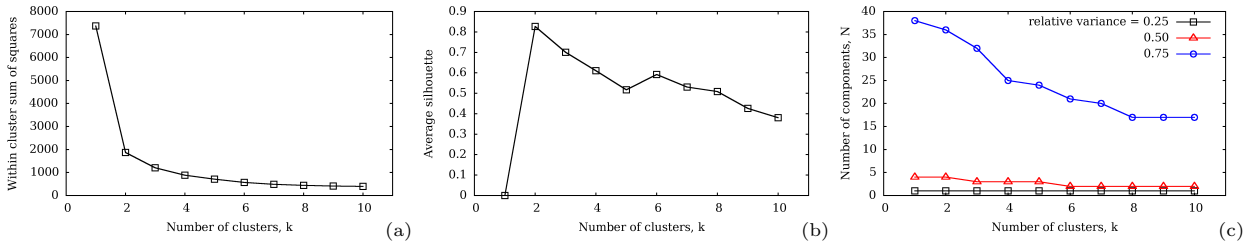


Figure 11: Clustering by point-wise energy spectra – Buoyant jet – Within cluster sum of squares (a), average silhouette (b), and number of POD/PCA components to resolve a prescribed relative variance (c).

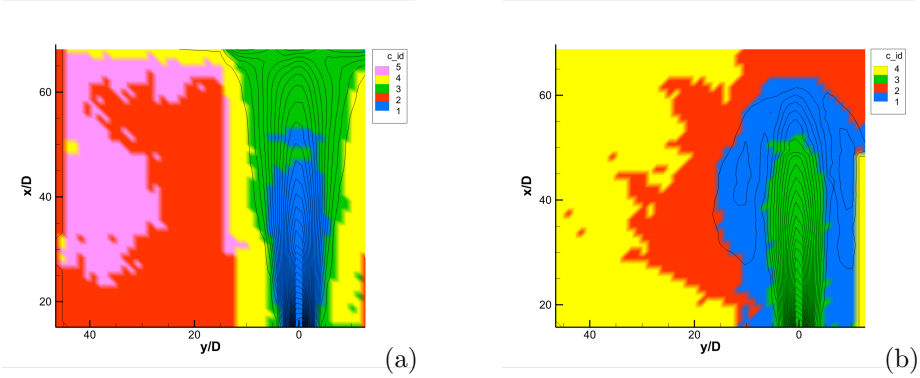


Figure 12: Clustering by point-wise energy spectra – Spatial points labeled by cluster (C_{id}) for uniform (a) and buoyant (b) jets.

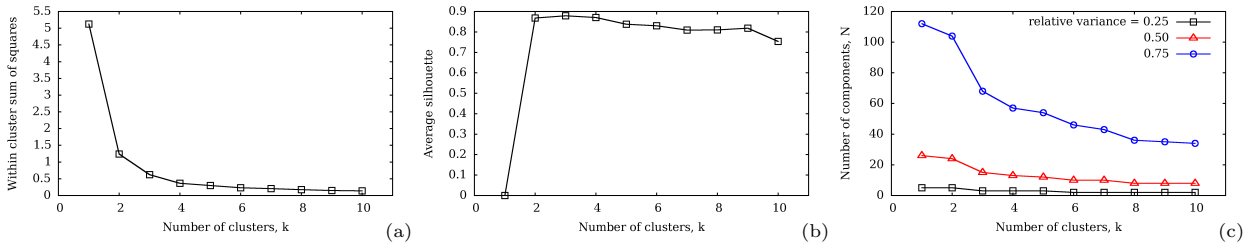


Figure 13: Clustering by point-wise Reynolds stress tensor components – Uniform jet – Within cluster sum of squares (a), average silhouette (b), number of POD/PCA components to resolve a prescribed relative variance (c).

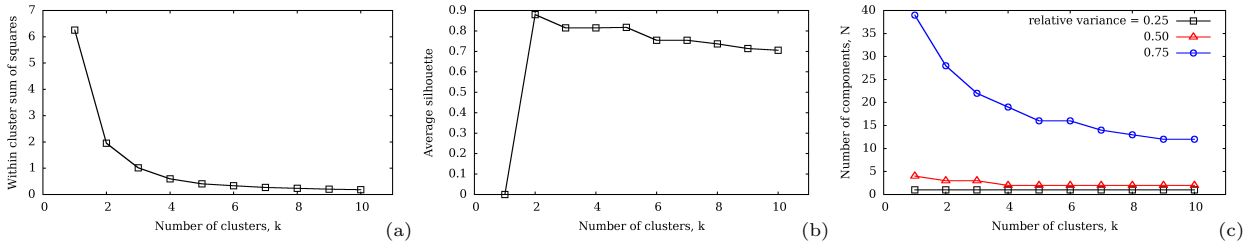


Figure 14: Clustering by point-wise Reynolds stress tensor components – Buoyant jet – Within cluster sum of squares (a), average silhouette (b), number of POD/PCA components to resolve a prescribed relative variance (c).

stream is globally (for the large scale) deterministic depending mainly on the phase.

A second analysis is performed, dividing the filed of view in several windows, based on the vorticity mean and variance associated to each cross section. Figure 22 shows the maximum mean and variance of cross sections along the propeller axis. Four windows are selected, as shown in Figure 17: (1) $0 \leq x/D < 0.4$,

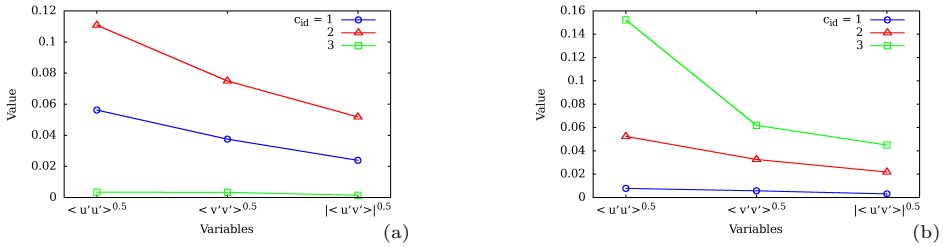


Figure 15: Clustering by point-wise Reynolds stress tensor components – Cluster centroids (C_{id}) for uniform (a) and buoyant (b) jets.

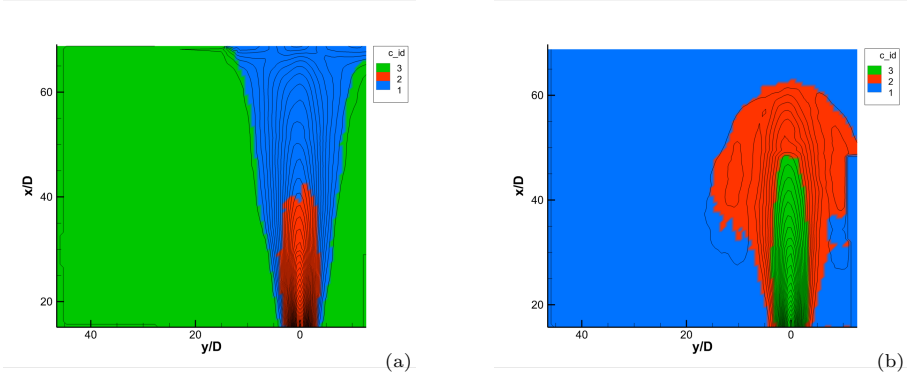


Figure 16: Clustering by point-wise Reynolds stress tensor components – Spatial points labeled by cluster (C_{id}) for transient uniform (a) and buoyant (b) jets.

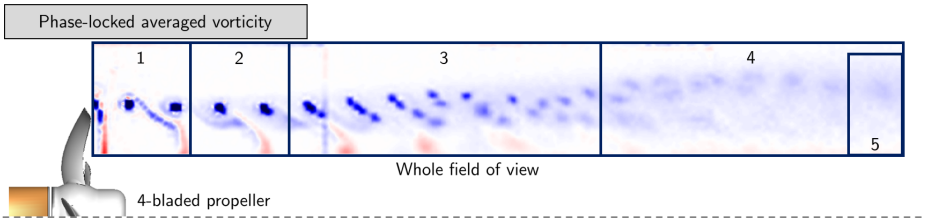


Figure 17: 4-bladed propeller – Windows used for clustering of vorticity snapshots.

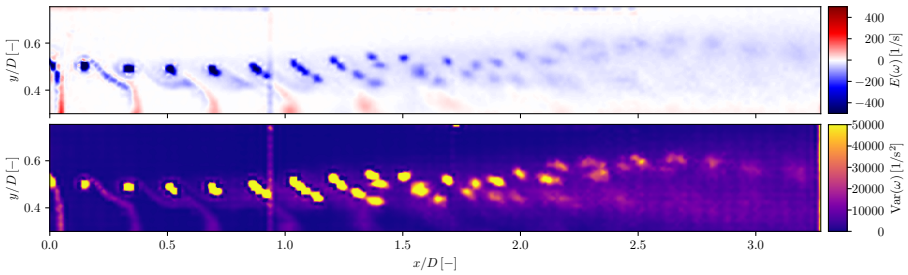


Figure 18: Vorticity mean value (top) and variance (bottom) for the 4-bladed propeller.

where the maximum variance is low and the wake is stable; (2) $0.4 \leq x/D < 0.8$, where the maximum variance starts increasing and the wake destabilizing; (3) $0.8 \leq x/D < 2$, where the maximum variance reaches its own maximum and starts decreasing along with the maximum mean and the wake experiences a fully developed tip vortex interaction; (4) $2 \leq x/D \leq 3.3$ where variance and mean are almost constant and a fully turbulent wake is observed.

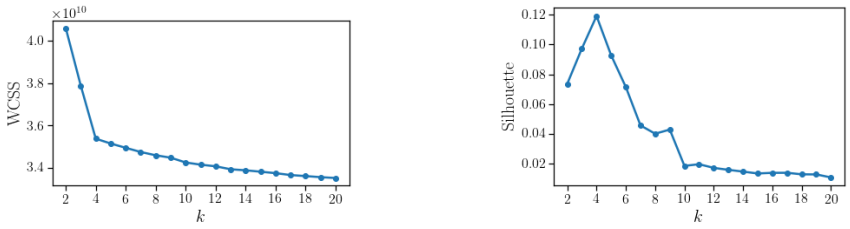
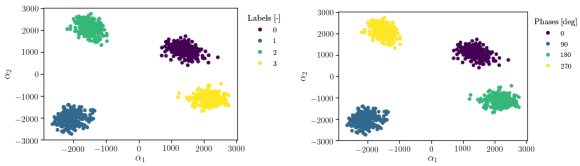
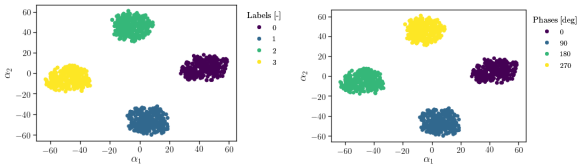


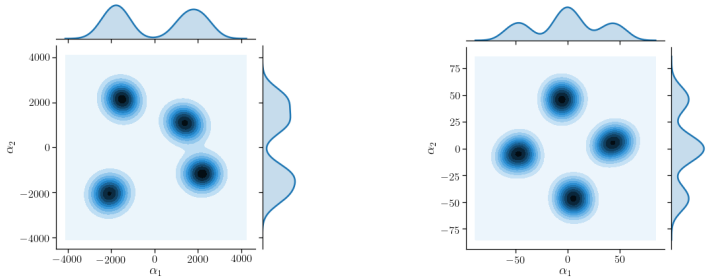
Figure 19: Clustering of vorticity snapshots – 4-bladed propeller, whole field of view – Within cluster sum of squares (left) and average silhouette (right).



(a) POD/PCA labeled by cluster (left) and phase (right)



(b) t-SNE labeled by cluster (left) and phase (right)



(c) KDE for POD/PCA (left) and t-SNE (right) coefficients

Figure 20: Clustering of vorticity snapshots – 4-bladed propeller, whole field of view – Data-projection on the first two POD/PCA modes (a) and embedding via t-SNE (b), along with joint and marginal probability density functions by KDE (c).

Figure 23, first column, shows the WCSS and silhouette for window 1. Figure 24, first column, shows the POD/PCA and t-SNE coefficients labeled by cluster and phase for the same window. Figure 25, fist column, shows the density functions of the coefficients. As expected, no significant structures are observed. The t-SNE highlights some structure and hints of data clustering. Nevertheless, these are not significant and the data can be interpreted as a single cluster.

Similarly, the second column of Figures 23, 24, and 25 provides the results for window 2. Clustering results are ambiguous since 4 and 3 clusters are identified by WCSS and silhouette, respectively. POD/PCA coefficients show 2 or 3 main clusters whereas the t-SNE clearly identifies 4 clusters associated to the propeller phases.

Results for window 3 and 4 are presented in the third and fourth column of Figures 23, 24, and 25, respectively where 4 clusters are clearly identified with a one-to-one association to the phase. It may be noted how an high degree of determinism is still present far downstream the propeller. It may be also noted how window 4 t-SNE analysis presents some hints of transition towards a different clustering structure.

Figure 26 shows the cluster centroids associated to windows 2, 3, and 4, where rows represent clusters and columns represent windows. As discussed earlier, these centroids also represent phase-locked averages. The unsupervised association of clusters to phases by k-means indicates that the destabilization (and coupling) of tip vortices progresses following mechanisms governed by deterministic chaos.

Finally, the same analysis is performed for window 5, which bounds more closely a single vortex (see Figure 17). Clustering results are ambiguous in this region (see Figure 27), even if some patterns are identified by both POD/PCA and t-SNE coefficients, where a pairwise mixture of phases $0 - 90$ and $180 - 270$ is present

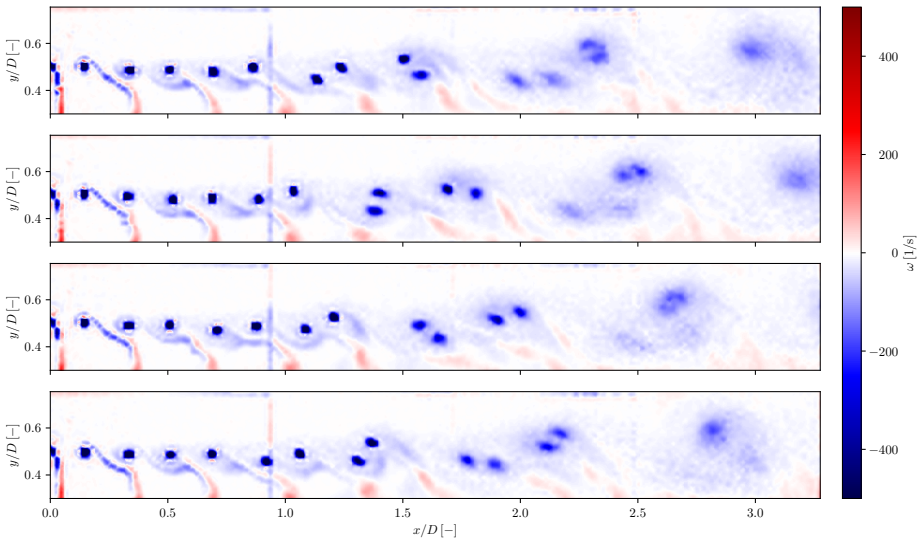


Figure 21: Clustering of vorticity snapshots – 4-bladed propeller, whole field of view – Cluster centroids corresponding to phases 0, 90, 180, 270 deg.

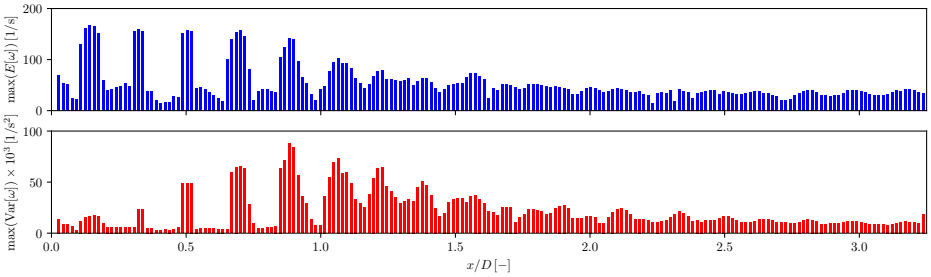


Figure 22: 4-bladed propeller, whole field of view – Maximum value of mean (top) and variance (bottom) of the vorticity along cross sections.

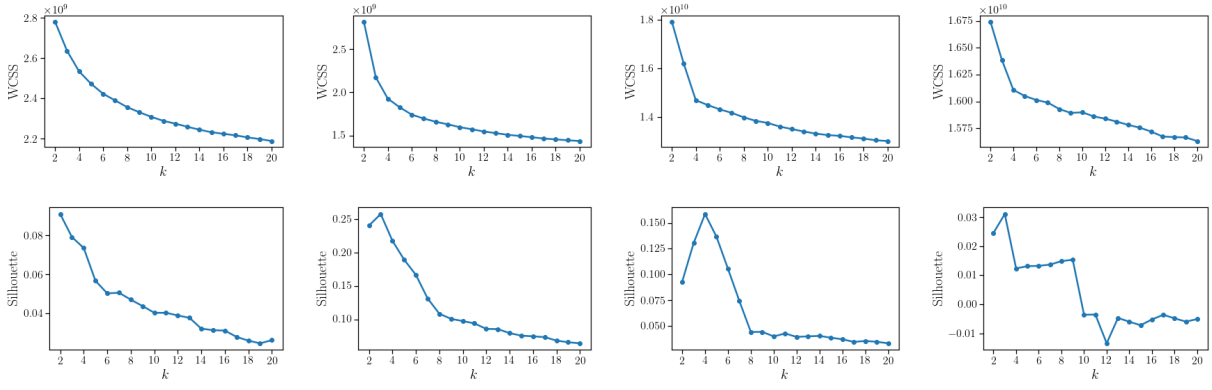
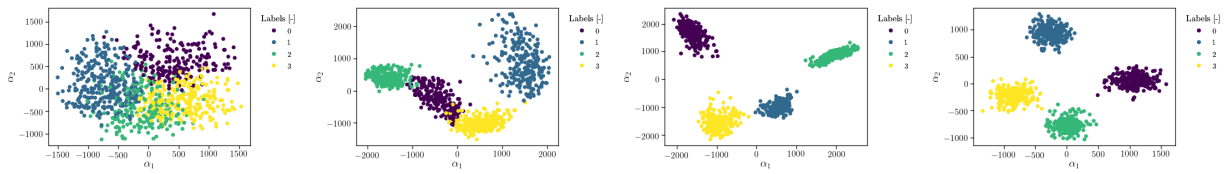


Figure 23: Clustering of vorticity snapshots – 4-bladed propeller – Within cluster sum of squares (top) and average silhouette (bottom). From left to right: windows 1, 2, 3, and 4.

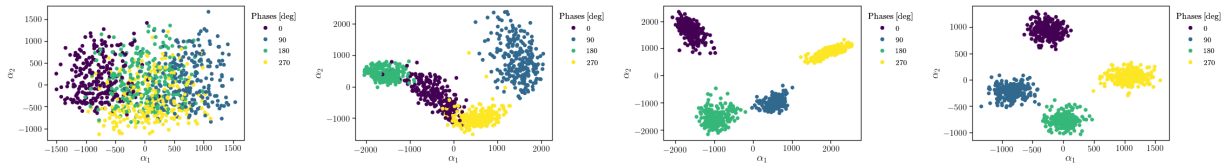
(see Figure 28). Cluster centroids are shown in Figure 29.

7-Bladed Propeller

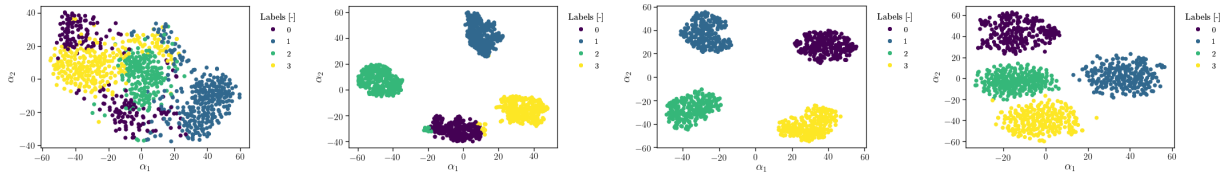
A similar analysis is performed for the 7-bladed propeller. In this case, the data set is composed by 500 snap-



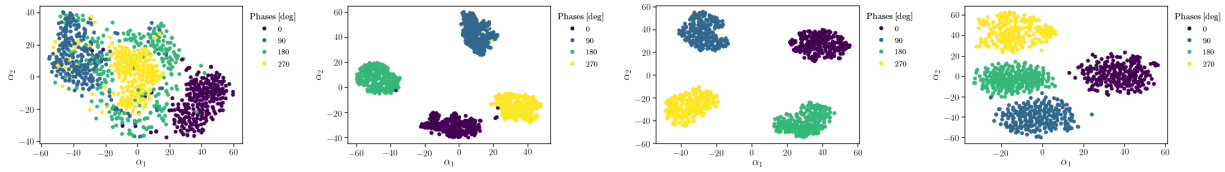
(a) POD/PCA labeled by cluster



(b) POD/PCA labeled by phase



(c) t-SNE labeled by cluster



(d) t-SNE labeled by phase

Figure 24: Clustering of vorticity snapshots – 4-bladed propeller – Data-projection on the first two POD/PCA modes (a,b) and embedding via t-SNE (c,d). From left to right: windows from 1 to 4.

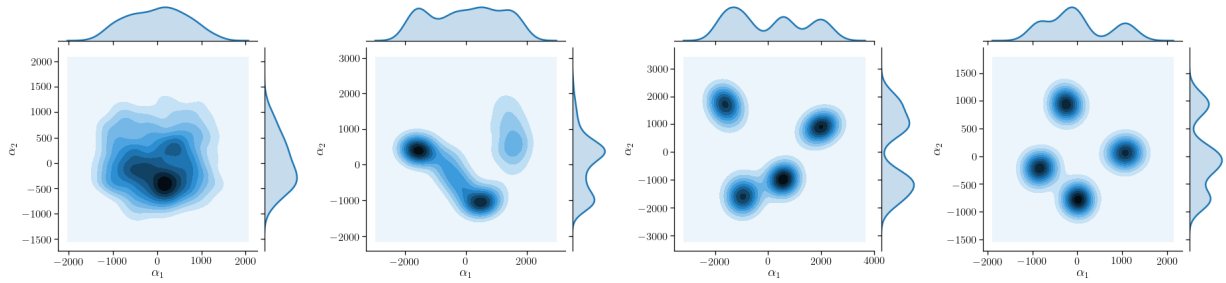
shots coming from a single phase (0 deg). The snapshots are organized (subsampled) in a 200×50 array, ranging axially from 0 to 3.3 D, and radially from 0 to 0.8 D (Figure 30). The data matrix has a dimension equal to $10{,}000 \times 500$.

Applying the clustering method to the whole field of view does not reveal any clusters, as shown by WCSS, silhouette (see Figure 31), and POD/PCA and t-SNE coefficients (Figure 32). This is due to the fact that the data set is composed by one phase only and it is consistent with what we found for the 4-bladed propeller, i.e., that overall the data clustering follows the phase. Also window 1 (bounding one vortex) does not show hints of clustering as Figures 33 and 34 show.

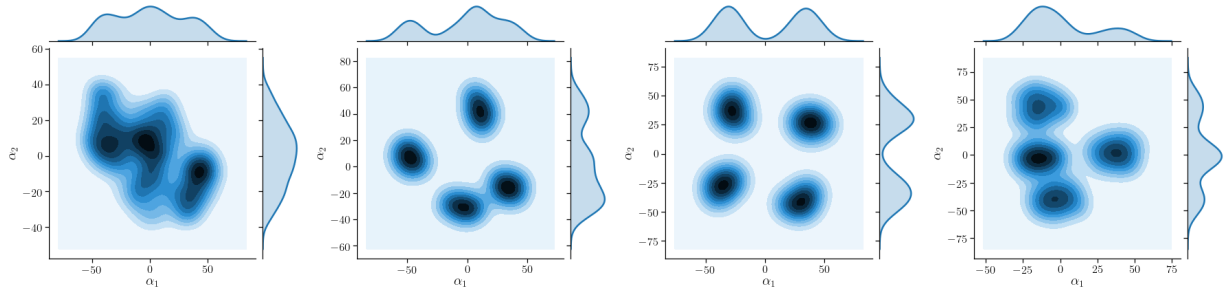
CONCLUSIONS AND FUTURE WORK

Spatial and snapshot clustering approaches have been presented and discussed for PIV data of high-Reynolds number uniform and buoyant jets and 4- and 7-bladed propeller wakes respectively. Data clustering was based on the k-means algorithm with the optimal number of clusters provided by the assessment of three metrics, namely the within-cluster sum of squares, average silhouette, and number of POD/PCA components required to resolve a prescribed level of variance. Spatial clustering for jet flows was based on three sets of clustering variables, namely the cross-section velocity profiles, point-wise energy spectra, and point-wise Reynolds stress tensor components. Snapshot clustering of phase-locked propellers wake data was based on the vorticity field with focus on the tip vortices. POD/PCA and t-SNE embedding along with KDE were used to provide a two-dimensional visualization of data clusters for assessment and discussion.

Clustering of jet cross-section velocity profiles helped identifying uniform and buoyant jet zones. The analysis of clustering variables allowed to propose new self-similarity laws for the jet, based on (i) actual center velocity and jet width for the velocity profiles and (ii)



(a) POD/PCA coefficients



(b) t-SNE coefficients

Figure 25: Clustering of vorticity snapshots – 4-bladed propeller – Joint and marginal probability density functions of POD/PCA (a) and t-SNE (b) coefficients by KDE. From left to right: windows from 1 to 4.

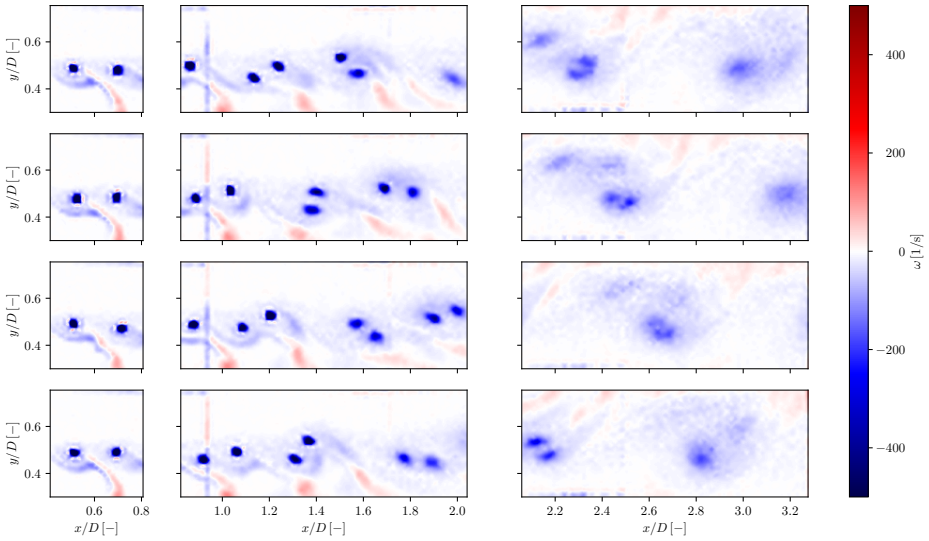


Figure 26: Clustering of vorticity snapshots – 4-bladed propeller – Cluster centroids corresponding to phases 0, 90, 180, 270 deg. From left to right: windows 2, 3, and 4.

their idealized linear representation for Reynolds-stress quantities. Three zones were identified for the uniform jet flow, namely bottom (developing), middle (fully developed), and top (impingement) regions. Five zones where identified for the buoyant jet, namely first from bottom (developed, with almost no reverse flow), second (developed, with significant reverse flow and associated shear stress), third (flow reversing), fourth (jet dome), fifth (top wall) regions. Clustering by point-wise energy spectra emphasized for both jets how jet core data retain the largest energy, with a well visible inertial subrange with slope equal to $-5/3$. Moving from the core

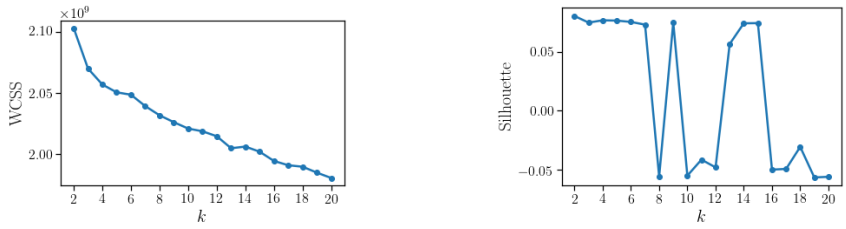
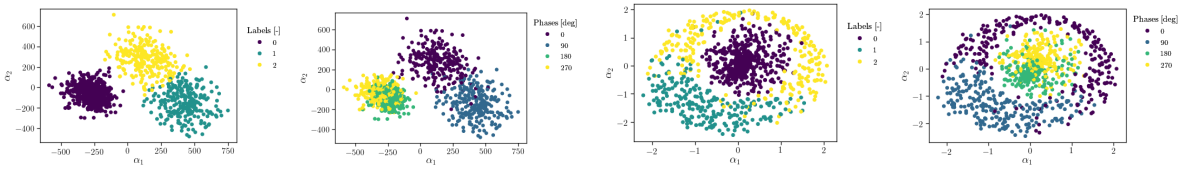
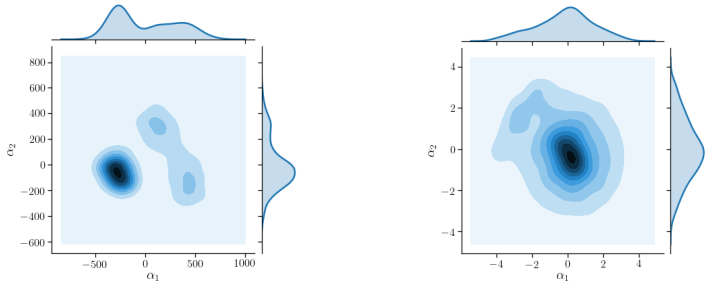


Figure 27: Clustering of vorticity snapshots – 4-bladed propeller, window 5 – Within cluster sum of squares (left) and average silhouette (right).



(a) POD/PCA labeled by cluster (left) and phase (right)

(b) t-SNE labeled by cluster (left) and phase (right)



(c) KDE for POD/PCA (left) and t-SNE (right) coefficients

Figure 28: Clustering of vorticity snapshots – 4-bladed propeller, window 5 – Data-projection on the first two POD/PCA modes (a) and embedding via t-SNE (b), along with joint and marginal probability density functions by KDE (c).

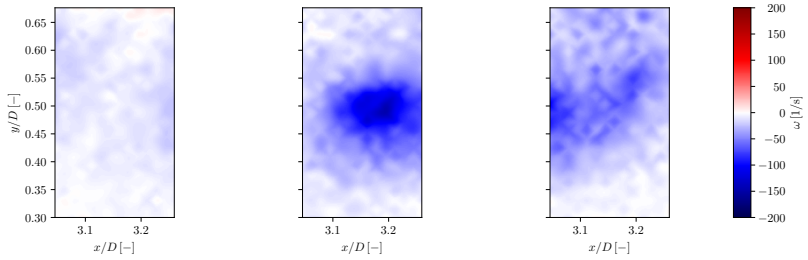


Figure 29: Clustering of vorticity snapshots – 4-bladed propeller, window 5 – Cluster centroids.

to outer regions and therefore to lower local Re, clusters brought into light subregions with slope equal to -3 and -1. Clustering results also emphasized the differences between uniform and buoyant jets, with significantly different clusters topologies. Finally, clustering by pointwise Reynolds stress tensor components produced spatial zones driven mainly by the turbulence intensity.

4-bladed propeller wake clustering of phase-locked snapshots produced no clusters (meaning only one cluster) for the near-field data window. Clusters with a one-to-one association to the phase were found for other data windows. Specifically, this was clearly observed for the whole field of view as well as for windows covering the region where the wake transitions to a unstable regimes and windows in the far field. Cluster-

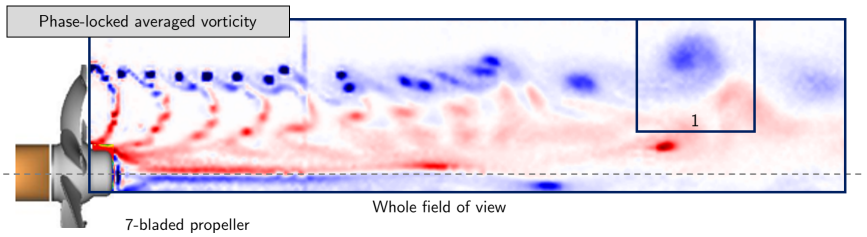


Figure 30: 7-bladed propeller – Windows used for clustering of vorticity snapshots.

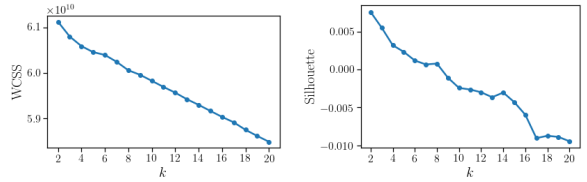


Figure 31: Clustering of vorticity snapshots – 7-bladed propeller, whole field of view – Within cluster sum of squares (left) and average silhouette (right).

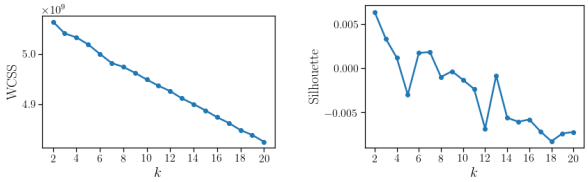
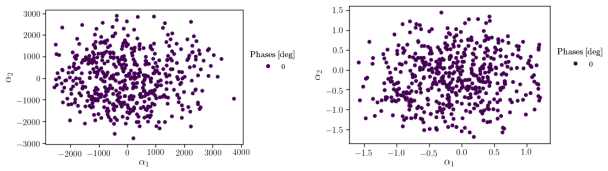
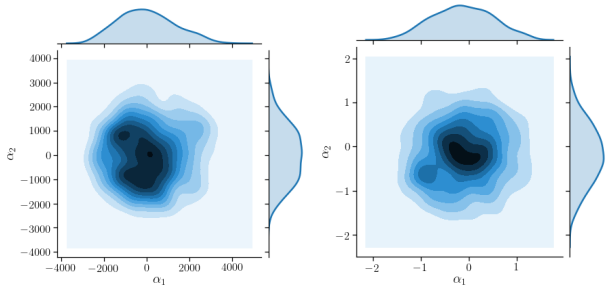


Figure 33: Clustering of vorticity snapshots – 7-bladed propeller, window 1 – Within cluster sum of squares (left) and average silhouette (right).

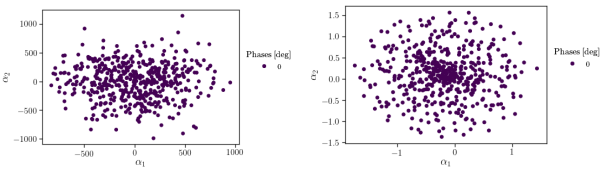


(a) POD/PCA (b) t-SNE

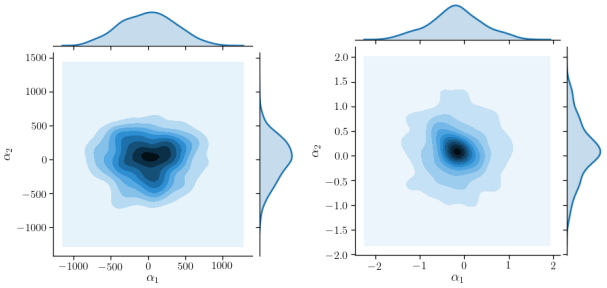


(c) KDE for POD/PCA (left) and t-SNE (right) coefficients

Figure 32: Clustering of vorticity snapshots – 7-bladed propeller, whole field of view – Data-projection on the first two POD/PCA modes (a) and embedding via t-SNE (b), along with joint and marginal probability density functions by KDE (c).



(a) POD/PCA (b) t-SNE



(c) KDE for POD/PCA (left) and t-SNE (right) coefficients

Figure 34: Clustering of vorticity snapshots – 7-bladed propeller, window 1 – Data-projection on the first two POD/PCA modes (a) and embedding via t-SNE (b), along with joint and marginal probability density functions by KDE (c).

ing results suggested that the wake instability and subsequent progression of tip vortices is characterized by mechanisms governed by deterministic chaos also in the far field. Results were confirmed by clustering of a single phase from the 7-bladed propeller data sets, where only a mild clusterization was found.

Ongoing and future work includes extending the analysis of clustering methods and results covering both spatial and temporal clustering for all jet and propeller wake cases with comparison and discussion of the results. The idea is to combine spatial/temporal clustering to fully exploits data reduction and visualization techniques to provide physical characterization of zones

and intervals in space and time domain respectively. Furthermore, a systematic analysis of POD/PCA modes for the whole and clusterized data set will be performed. Extensions to dynamic mode decomposition (DMD) are also part of the ongoing research.

ACKNOWLEDGMENTS

CNR-INM wishes to thank Dr. Woei-Min Lin, Dr. Elena McCarthy, and Dr. Salahuddin Ahmed of the ONR and ONR Global for their support through NICOP grants N62909-18-1-2033 and N62909-19-1-2001.

REFERENCES

Arthur, D. and Vassilvitskii, S., "k-means++: The advantages of careful seeding", Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp. 1027–1035. Society for Industrial and Applied Mathematics, 2007.

Barwey, S., Raman, V., and Steinberg, A. M., "Data-Driven Reduction and Decomposition via Time-Axis Clustering", AIAA Scitech 2020 Forum, pp. 0365, 2020.

Benzi, R., Paladin, G., and Vulpiani, A., "Power spectra in two-dimensional turbulence", Physical Review A 42(6), 1990, pp. 3654.

Berkooz, G., Holmes, P., and Lumley, J. L., "The proper orthogonal decomposition in the analysis of turbulent flows", Annual Review of Fluid Mechanics 25(1), 1993, pp. 539–575.

Bishop, C. M. Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

Cizmas, P. G., Palacios, A., O'Brien, T., and Syamlal, M., "Proper-orthogonal decomposition of spatio-temporal patterns in fluidized beds", Chemical Engineering Science 58(19), 2003, pp. 4417 – 4427.

Clément, S. A., Guillemain, A., McCleney, A. B., and Bardet, P. M., "Options for refractive index and viscosity matching to study variable density flows", Experiments in Fluids 59(32), 2018.

Drineas, P., Frieze, A., Kannan, R., Vempala, S., and Vinay, V., "Clustering large graphs via the singular value decomposition", Machine learning 56(1-3), 2004, pp. 9–33.

Felli, M. and Falchi, M., "A parametric survey of propeller wake instability mechanisms by detailed flow measurement and time resolved visualizations", Proceedings of the 32nd Symposium on Naval Hydrodynamics, Hamburg, Germany, 2018a.

Felli, M. and Falchi, M., "Propeller wake evolution mechanisms in oblique flow conditions", Journal of Fluid Mechanics, Vol. 845, 2018, pp. 520–559.

Felli, M., Falchi, M., and Pereira, F. J. A., "Distance effect on the behavior of an impinging swirling jet by PIV and flow visualizations", Experiments in Fluids 48(2), 2010, pp. 197–209.

Gibson, M., "Spectra of turbulence in a round jet", Journal of Fluid Mechanics 15(2), 1963, pp. 161–173.

Golub, G. H. and Reinsch, C., "Singular value decomposition and least squares solutions", Numerische mathematik 14(5), 1970, pp. 403–420.

Gordeyev, S. V. and Thomas, F. O., "Coherent structure in the turbulent planar jet. Part 2. Structural topology via POD eigenmode projection", Journal of Fluid Mechanics, Vol. 460, 2002, pp. 349–380.

Holmes, P., Lumley, J. L., Berkooz, G., and Rowley, C. W. Turbulence, coherent structures, dynamical systems and symmetry. Cambridge university press, 2012.

Jain, A. K., "Data clustering: 50 years beyond K-means", Pattern recognition letters 31(8), 2010, pp. 651–666.

Ketchen, D. J. and Shook, C. L., "The application of cluster analysis in strategic management research: an analysis and critique", Strategic Management Journal 17(6), 1996, pp. 441–458.

Lloyd, S., "Least squares quantization in PCM", IEEE transactions on information theory 28(2), 1982, pp. 129–137.

Posa, A., Broglia, R., Felli, M., Falchi, M., and Balaras, E., "Characterization of the wake of a submarine propeller via Large-Eddy simulation", Computers and Fluids, Vol. 184, 2019, pp. 138 – 152.

Rousseeuw, P. J., "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", Journal of Computational and Applied Mathematics, Vol. 20, 1987, pp. 53–65.

Serani, A., Durante, D., Diez, M., D'Agostino, D., Clement, S., Badra, J., Andre, M., Habukawa, M., and Bardet, P., "PIV Data Clustering of a Buoyant Jet in a Stratified Environment", 57th AIAA Aerospace Sciences Meeting, SciTech 2019, Manchester Grand Hyatt San Diego, San Diego, California, January 7-11, 2019.

Silverman, B. W. Density estimation for statistics and data analysis. Routledge, 2018.

Simonoff, J. S. Smoothing methods in statistics. Springer Science & Business Media, 2012.

van der Maaten, L. and Hinton, G., "Visualizing data using t-SNE", Journal of Machine Learning Research 9(Nov), 2008, pp. 2579–2605.

Wu, J., Wang, J., Xiao, H., and Ling, J., "Visualization of high dimensional turbulence simulation data using t-SNE", 19th AIAA Non-Deterministic Approaches Conference, pp. 1770, 2017.

Zhou, X. and Hitt, D. L., "Proper orthogonal decomposition analysis of coherent structures in a transient buoyant jet", Journal of Turbulence, Vol. 5, 2004, pp. N28.