Journal of *Marine Science and Engineering*

MDPI

Article

PIV Snapshot Clustering Reveals the Dual Deterministic and Chaotic Nature of Propeller Wakes at Macro- and Micro-Scales

Danny D'Agostino [1,2], Matteo Diez [1], Mario Felli [1] and Andrea Serani [1,*]

[1] CNR-INM, National Research Council-Institute of Marine Engineering, 00128 Rome, Italy
[2] Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 119077, Singapore
* Correspondence: andrea.serani@cnr.it

Abstract: This study investigates the underlying mechanisms governing the evolution of tip vortices in the far field of a naval propeller wake. To achieve this, a novel approach utilizing data clustering applied to particle image velocimetry snapshots is employed. The clustering of data is carried out using the k-means algorithm, with the optimal number of clusters determined by evaluating two metrics: the within-cluster sum of squares and the average silhouette. The clustering of phase-locked propeller wake data is focused on the vorticity associated with the regions containing tip vortices. Additionally, techniques such as proper orthogonal decomposition, t-distributed stochastic neighbor embedding, and kernel density estimation are employed to visually represent the data clusters in a two-dimensional space, facilitating their assessment and subsequent discussion. This paper shows how the application of data clustering enables a comprehensive understanding of the complex mechanisms driving the dynamics of propeller wake vortices in both the transitional and far fields. Specifically, it reveals the dual nature of the propeller wake flow, characterized by deterministic and chaotic behavior at macro- and micro-scales.

Keywords: snapshot clustering; *k*-means; proper orthogonal decomposition; principal component analysis; *t*-distributed stochastic neighbor embedding; propeller wake

Citation: D'Agostino, D.; Diez, M.; Felli, M.; Serani, A. PIV Snapshot Clustering Reveals the Dual Deterministic and Chaotic Nature of Propeller Wakes at Macro- and Micro-Scales. *J. Mar. Sci. Eng.* **2023**, *11*, 1220. https://doi.org/10.3390/jmse11061220

Academic Editor: Weicheng Cui

Received: 19 May 2023
Revised: 9 June 2023
Accepted: 11 June 2023
Published: 13 June 2023

1. Introduction

The knowledge of the underlying mechanisms of propeller/rotor tip vortex evolution in the far field is an intriguing and still open problem of fluid dynamic research, which plays an important role in many applications of marine, aerospace, and mechanical engineering because of its impact on performance, vibrations, noise, and structural problems. A prominent example is wind parks, where power production, radiated sound, and fatigue lifetime of interior turbines might be significantly affected by the dynamics of the wake vortices released from surrounding turbines, see, e.g., [1–3]. Previous studies [4,5] have shown that propeller wake instability is driven by mutual inductance mechanisms, which lead adjacent vortex filaments to pair analogously to the leapfrogging motion of two inviscid vortex rings [6,7]. Vortex pairing [4,8,9] manifests through a multi-step grouping process that depends on the blade number. This behavior has been highlighted by flow visualizations and phase-locked particle image velocimetry (PIV) measurements in the propeller transitional and far fields [10,11]. In spite of these results and to the authors' best knowledge, rigorous, in-depth analyses focused on classifying the nature of the propeller tip vortex dynamics in the far field are still extremely limited, likely due to the complexity associated with the inherently multi-scale, highly turbulent, strongly unsteady nature of the flow, which makes its survey extremely challenging even using the most advanced experimental and computational tools.

Data science, machine learning, and artificial intelligence have emerged as cutting-edge research fields where rigorous methods and algorithms are developed and applied

to gain knowledge from data. In recent years, these methods have been applied in the context of more traditional disciplines to accelerate the experimental/computational analysis process and extract insights from experiments and simulation data. In the field of ship hydrodynamics, supervised machine learning techniques, such as surrogate modeling, regression, and multi-fidelity methods, have been employed to integrate experimental and simulation data. These methods aim to reduce the computational burden associated with design performance assessment and optimization procedures along with facilitating the interpretation of the physics observed. Furthermore, unsupervised machine learning methods, including equivalent approaches to proper orthogonal decomposition (POD) and/or linear/non-linear principal component analysis (PCA), have been utilized to address challenges such as dimensionality reduction, efficient design, and operational space exploration, and gain insights into complex physical phenomena. POD, which is closely related to PCA, has found widespread application in the identification of coherent structures within turbulent flows [12] and has been employed in the study of various flow configurations, including steady and transient jets [13,14] and marine propeller wakes [5]. By decomposing the flow into a linear combination of orthogonal eigenfunctions, POD provides valuable information about the spatial and temporal structure of the flow, enabling the construction of reduced-order models. However, it is important to note that while POD enjoys established global optimality properties, its effectiveness may be limited when investigating non-linear, transient, non-stationary, and non-ergodic dynamics. To address these limitations, non-linear dimensionality reduction (NLDR) methods have been developed and applied to gain a deeper understanding of data structures and physical phenomena. One straightforward approach to NLDR in conjunction with POD/PCA is to employ data clustering methods and perform POD/PCA within each cluster. This approach allows for a more comprehensive analysis and characterization of the underlying dynamics.

Cluster-based reduced-order modeling through POD/PCA offers valuable insights into complex phenomena by employing local PCA (LPCA), which partitions the dataset into clusters and applies POD/PCA within each cluster, assuming an approximate linear structure within them. The cluster centroids, together with their associated modes, capture relevant flow features in the spatial and temporal domains. Previous studies have demonstrated the application of spatial clustering using the *k*-means method with POD/PCA for analyzing a transient buoyant jet [15] and for characterizing a swirl-stabilized combustor flow through temporal and spatial *k*-means clustering [16]. It is worth noting that the choice of the number of clusters and the similarity metrics used for data clustering significantly impact the quality of the resulting reduced-order/dimensionality model, thereby influencing the extraction of meaningful physical insights from the clustering analysis. Rigorous data-clustering methods have been proposed in various fields to facilitate a deeper understanding of experimental and simulation data. To achieve fully non-linear dimensionality reduction, a promising technique called *t*-distributed stochastic neighbor embedding (*t*-SNE) has been proposed [17]. This method allows for embedding and visualizing high-dimensional data in a low-dimensional space and has been successfully applied to turbulence datasets from simulations [18].

Examples of data-clustering analysis applied to PIV experimental data are still limited, see, e.g., [19], where *k*-means is applied to study the internal aerodynamics of a Miller cycle gasoline engine. Examples of data-clustering analyses applied to propeller performance and, in general, vortex dynamics are also limited. Calvet et al. [20] developed an unsupervised machine learning strategy to automatically cluster vortex wakes of bio-inspired propulsors into groups of similar propulsive thrust and efficiency metrics. Doijode et al. [21] introduced a clustering approach for optimizing propellers by directing the search for the optimal design towards design clusters with good performance, i.e., high propulsive efficiency and low cavitation. Finally, Sharma and Simula [22] considered computer-generated configurations of quantized vortices in planar superfluid Bose–Einstein condensates, showing how data-clustering approaches can be successfully used for classifying vortex configurations to identify prominent vortex phases of matter.

The objective of this study is to establish a systematic approach for conducting a data-clustering analysis of PIV data to gain valuable physical insights into the dynamics of propeller tip vortices. Phase-locked PIV snapshots recorded throughout the transitional and far fields of a propeller wake (see Figure 1) are used in the clustering analysis.

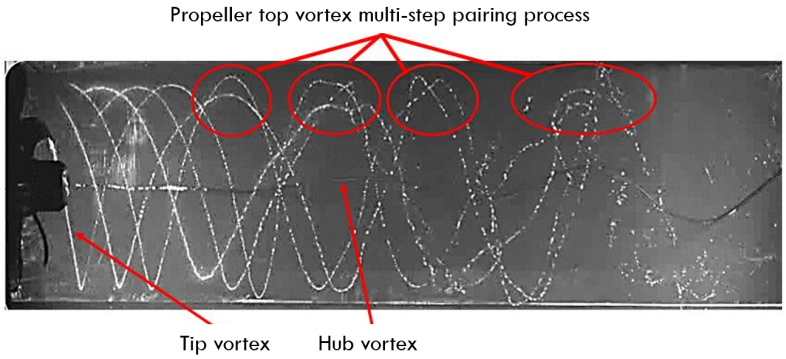

Figure 1. Snapshot of a cavitating propeller.

Specifically, the velocity fields under investigation are obtained from experimental tests with large-scale, phase-locked, PIV measurements. Data were collected in the cavitation tunnel of CNR-INM [23]. The clustering of PIV data is based on the *k*-means algorithm. Data clustering is applied to phase-locked snapshots to gain knowledge on the topology of wake instability and its stochastic realizations. The vorticity is used as a clustering variable. The resulting cluster centroids identify the topology of the instability, where two or more tip vortices interact and coalesce [4,11]. Two metrics are used for the identification and assessment of clustering methods, including the selection of the proper number of clusters, namely: (i) within-cluster sum of squares and (ii) average silhouette. Additionally, the embedding of data via POD/PCA and *t*-SNE is used to define and visualize data clusters in a reduced dimensionality space. Finally, kernel density estimation (KDE) is applied to POD/PCA and *t*-SNE representations to provide continuous data distributions for assessment and discussion. The overall workflow is synthesized in Figure 2.

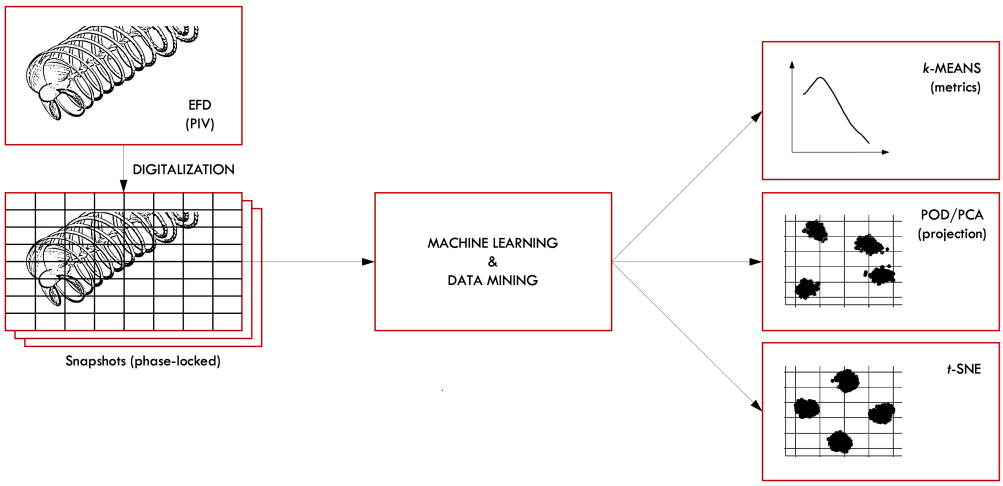

Figure 2. Overall machine learning and data mining workflow for propeller wake investigation.

The remainder of the paper is organized as follows: experimental details are provided in Section 2; Section 3 presents the data analysis methods and metrics; the numerical results are presented in Section 4; finally, conclusions and future perspectives are given in Section 5.

2. Experimental Methods

The present study is based on a comprehensive database of flow measurements of the E1658 submarine propeller wake in open water using 2D-PIV [5,24], see Figure 1. 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 a propeller configuration with four blades and one value of the advance ratio, corresponding to a high propeller loading (i.e., $J = 0.56$).

The experimental campaign was carried out at the CNR-INM cavitation tunnel, measuring the propeller wake flow at the vertical center-plane by a 2D-PIV system. The setup consisted of three 5Mpx sCMOS cameras by PCO and two 200 mJ Nd-Yag Lasers by QUANTEL. Either cameras and lasers were positioned side by side to simultaneously record/illuminate a long portion of the wake flow (i.e., from the propeller plane to 3.3D downstream, where D is the propeller diameter) at full resolution (4, 22, 23). This arrangement (see Figure 3), already adopted in other similar experiments [5,10,11], 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.

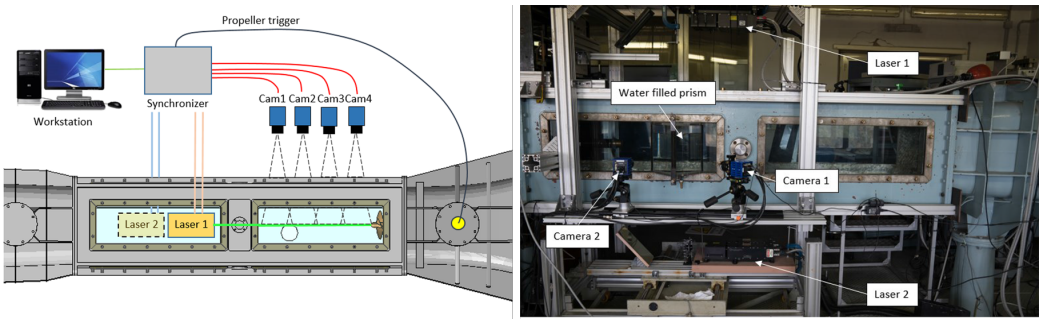

Figure 3. Sketch (**left**) and side (**right**) views of the PIV experimental setup.

The camera acquisition process was synchronized with the propeller reference blade at a specific angular position, facilitated by the coordination of four cameras and two lasers with a TTL OPR (Once Per Revolution) signal. This signal was generated by a rotary incremental encoder mounted on the propeller dynamometer, providing a frequency of 3600 pulses per second.

Hollow glass spheres, featuring a nominal diameter of 10 µm and a specific weight of 1.05 g/cm^3, were utilized as tracers in the experiment. The acquired images underwent pre-processing using a local background subtraction routine with a 4×4 px^2 kernel. This pre-processing step effectively eliminated unwanted reflections and improved the signal-to-noise ratio. The vector calculation was performed using the advanced iterative multi-pass, multi-grid, image deformation algorithm [25]. Leveraging the GPU architecture of the workstation and the PIV algorithm, a combination of direct cross-correlation and parallel processing was employed, enabling both accurate and efficient evaluation of the images. The interrogation windows were iteratively reduced to a final size of 32 × 32 px^2, with a 50% overlap between adjacent windows.

3. Numerical Methods

The subsequent section presents an overview of the data analysis methods and metrics employed in this study. Specifically, we briefly describe and discuss the k-means clustering and t-SNE methods as visualization techniques, along with KDE. Additionally, two metrics, namely the within-cluster sum of squares and the average silhouette, are utilized to assess the clustering approaches and determine the optimal number of clusters, denoted as k.

3.1. k-Means Clustering

The k-means clustering method [26] is commonly used to build partitions of data, represented by an $[L \times S]$ matrix $\mathbf{D}$ (see below Equation (2)), into k distinct sets or clusters. Each cluster is defined by a representative point known as a centroid. The rearrangement of the original data in $\mathbf{D}$ is performed according to the specific clustering approach and criterion utilized. Generally, we denote each realization or point of the rearranged data as $\mathbf{d}_j$. Note that, generally, $\mathbf{d}_j \in \mathbb{R}^Q$ with $j = 1, ..., H$, where $Q \neq L$ and $H \neq S$.

To measure the similarity between data points $\mathbf{d}_j$ and evaluate the cluster centroids $\boldsymbol{\mu}_i$, the Euclidean distance is utilized. The centroids $\boldsymbol{\mu}_i$ are obtained by averaging all data points within the i-th cluster $\mathbf{K}_i$. The assignment of data points to the k clusters is accomplished by minimizing the squared Euclidean distance between $\mathbf{d}_j$ and $\boldsymbol{\mu}_i$, known as the within-cluster sum of squares (WCSS):

$$\text{WCSS} = \sum_{i=1}^{k} \sum_{\mathbf{d}_j \in \mathbf{K}_i} \|\mathbf{d}_j - \boldsymbol{\mu}_i\|^2 \tag{1}$$

Solving the minimization problem in Equation (1) is known to be NP-hard [27]. To tackle this challenge, the heuristic approach proposed in [28] is employed. It is worth noting that the results obtained are highly sensitive to the initialization of the centroids. In this study, the initialization strategy suggested in [29] is utilized.

3.2. Proper Orthogonal Decomposition/Principal Component Analysis

The implementation of POD/PCA is carried out as follows. The data matrix $\mathbf{D}$ with dimensions $[L \times S]$ is defined as

$$\mathbf{D} = \left[\begin{array}{c|c|c} \mathbf{d}^{(1)} & \dots & \mathbf{d}^{(S)} \end{array} \right] \tag{2}$$

where $\mathbf{d} = \{d_1(\mathbf{x}_1), \dots, d_1(\mathbf{x}_P), \dots, d_M(\mathbf{x}_1), \dots, d_M(\mathbf{x}_P)\}^\mathsf{T}$ represents the collection of M discretized components of the quantity of interest $\mathbf{d}$. Each vector $\mathbf{d}^{(s)}$ corresponds to a snapshot at time realization s, while $\mathbf{x}_i$ denotes the i-th node of the spatial discretization. The size of the spatial discretization is denoted by P, resulting in $L = MP$.

The dimensionality of the data matrix $\mathbf{D}$ is reduced by projecting the snapshots onto a new linear subspace formed by the eigenvectors of the covariance matrix $\mathbf{C}$ of dimensions $[L \times L]$ [30]:

$$\mathbf{C} = \frac{1}{S} \mathbf{D}\, \mathbf{D}^\mathsf{T} \tag{3}$$

The eigenvectors ($\mathbf{z}_i$) and eigenvalues (λ_i) of $\mathbf{C}$ are obtained by solving the eigenvalue problem:

$$\mathbf{C}\mathbf{Z} = \mathbf{Z}\boldsymbol{\Lambda} \tag{4}$$

Alternatively, the problem in Equation (4) can be solved using the singular value decomposition (SVD) [31]. In cases where $L > S$, the dual problem can be solved using the snapshot-POD formulation [32].

The eigenvalues obtained from POD/PCA represent the variance explained by the corresponding eigenvectors. The linear subspace formed by the N eigenvectors (POD/PCA modes, collected in $\hat{\mathbf{Z}}$) associated with the largest N eigenvalues captures the highest amount of variance/energy compared to other linear subspaces of dimension N [30,33]. The cumulative sum of the eigenvalues is used to assess the variance resolved by the linear subspace of dimension N.

Finally, the reconstructed data matrix $\hat{\mathbf{D}}$ is given by:

$$\hat{\mathbf{D}} = \hat{\mathbf{Z}}\hat{\mathbf{Z}}^\mathsf{T}\mathbf{D} \tag{5}$$

In this equation, the coefficient $\boldsymbol{\alpha}_i = \hat{\mathbf{z}}_i^\mathsf{T}\mathbf{D}$ represents the projection of the data matrix onto the i-th mode.

3.3. t-Distributed Stochastic Neighbor Embedding

The t-SNE is a machine learning algorithm proposed by [17], 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 similarity among data-points $\mathbf{d}^{(i)}$ and $\mathbf{d}^{(j)}$ parameterized by a Gaussian distribution

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2S} \tag{6}$$

with

$$p_{j|i} = \frac{\exp\left(-||\mathbf{d}^{(i)} - \mathbf{d}^{(j)}||^2/2\varsigma_i^2\right)}{\sum_{k\neq i}^{S} \exp\left(-||\mathbf{d}^{(i)} - \mathbf{d}^{(k)}||^2/2\varsigma_i^2\right)} \tag{7}$$

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

$$q_{ij} = \frac{(1 + ||\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j||^2)^{-1}}{\sum_{k\neq i}^{S}(1 + ||\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_k||^2)^{-1}} \tag{8}$$

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

$$\mathrm{KL}(P||Q) = \sum_{i\neq j}^{S} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right) \tag{9}$$

The minimization of Equation (9) 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 $\mathbf{d}^{(i)}$ equals a predefined perplexity $\mathrm{Perp}(P_i)$,

$$\mathrm{Perp}(P_i) = 2^{\left(\sum_j^S p_{j|i} \log_2 p_{j|i}\right)} \tag{10}$$

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 [17].

3.4. Multivariate Kernel Density Estimation

The kernel density estimation (KDE, [34]) 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 [35], let $\{\boldsymbol{\alpha}\}_{i=1}^{S}$ be a d-variate random vector whose PDF is estimated as

$$\mathrm{PDF}(\boldsymbol{\alpha}) = \frac{1}{S}\sum_{i=1}^{S} K_{\mathbf{H}}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_i) \tag{11}$$

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

$$K_{\mathbf{H}}(\boldsymbol{\alpha}) = |\mathbf{H}|^{-\mathbf{1/2}}\mathbf{K}(\mathbf{H}^{-\mathbf{1/2}}\boldsymbol{\alpha}) \tag{12}$$

3.5. Metrics

The WCSS, as expressed in Equation (1), serves as an assessment metric for determining the optimal number of clusters, denoted as k. In this study, the WCSS metric is employed in conjunction with the elbow method [36].

To evaluate the consistency of data within clusters, the silhouette method is employed [37]. For a data point $\mathbf{d}_i$, the average Euclidean distance a_i is calculated as the

average distance between $\mathbf{d}_i$ and other data points within the same cluster. Additionally, the smallest average Euclidean distance c_i is computed as the minimum average distance of $\mathbf{d}_i$ to all data points in other clusters. The silhouette value s_i associated with $\mathbf{d}_i$ is defined as follows:

$$s_i = \frac{a_i - c_i}{\max[a_i, c_i]} \tag{13}$$

This value quantifies the similarity of the data point to points within its own cluster relative to points in other clusters. It should be noted that s_i ranges from -1 to 1, with a value of 1 indicating maximum similarity. The average silhouette, denoted as s_{avg}, is utilized as a metric for effective data clustering:

$$s_{\text{avg}} = \sum_{i=1}^{H} s_i \tag{14}$$

It is worth mentioning that the silhouette is not defined when $k = 1$. As a convention, the average silhouette is set to 0 for $k = 1$.

4. Results

POD/PCA implementation for the propeller wake uses the following decomposition of the vorticity

$$\omega = \overline{\omega} + \omega' \tag{15}$$

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. Overbar and prime characters indicate time average and fluctuations, respectively.

Snapshot clustering is performed using vorticity snapshots following clustering variables as

$$\mathbf{d}_i = \boldsymbol{\omega}^{(i)} \qquad i = 1, ..., H \tag{16}$$

Data sets are organized in phase-locked snapshots, where each phase is typically observed hundreds of times.

The $[L \times S]$ data matrix D is defined as

$$\boldsymbol{\Omega} = \left[\begin{array}{c|c|c} \boldsymbol{\omega}^{(1)} & \dots & \boldsymbol{\omega}^{(S)} \end{array} \right] \tag{17}$$

where $\boldsymbol{\omega} = \{\omega'(\mathbf{x}_1), \dots, \omega'(\mathbf{x}_P)\}^{\mathsf{T}}$ 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 $i = 1, \dots, S$, indicates the i-th time realization (snapshot), and finally, $L = P$. Data analysis is performed using in-house Python scripts based on the Scikit-learn library [38].

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 1000 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 4). The data matrix has a dimension equal to 6000×1000. 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).

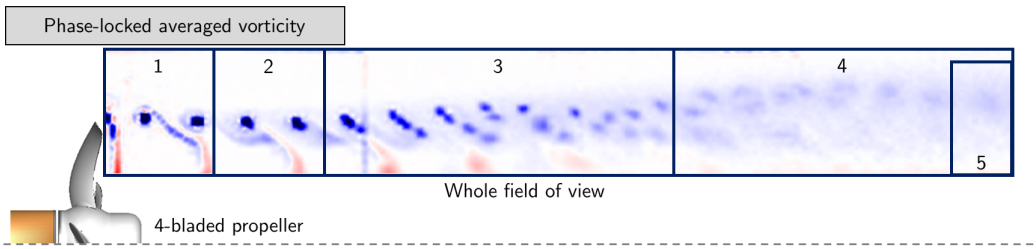

Figure 4. Windows used for clustering of vorticity snapshots.

First, the k-means is applied to the whole field of view (see Figure 4). Figure 5a shows that four clusters 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 5b shows the projection of the data set onto the first two POD/PCA modes. Data are 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 5c, confirming the POD/PCA result. Finally, Figure 5b,c (right side) provide joint and marginal probability density functions of POD/PCA and t-SNE coefficients α_1 and α_2 given by KDE, confirming the data have four clusters of equal size. The corresponding cluster centroids are presented in Figure 5d, showing that the mechanism of vortex coupling and convection downstream is globally (for the large scale) deterministic throughout the investigated wake region and depends mainly on the phase. This result is consistent with the literature (see, e.g., [8]) and further strengthens the conjecture that the far-field evolution of the propeller wake structures exhibits a deterministic behavior at the macro-scale and a chaotic nature at the micro-scale.

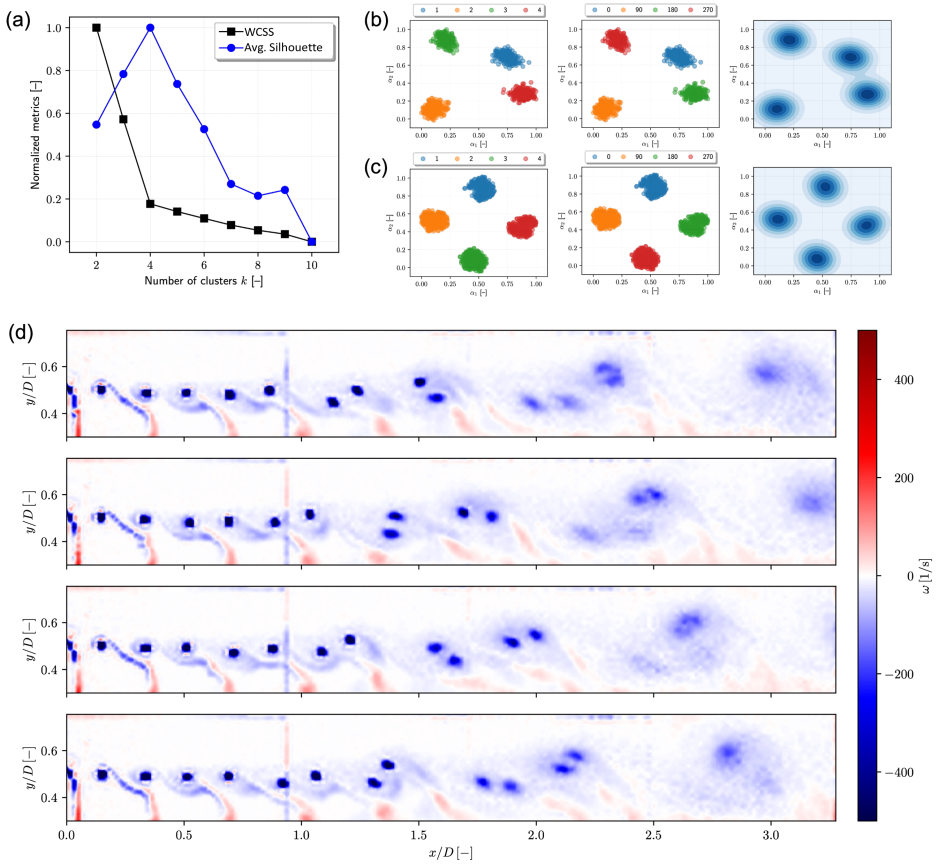

Figure 5. Whole field of view: (**a**) clustering metrics; (**b**) data-projection on the first two POD/PCA modes; and (**c**) embedding via t-SNE, labeled by cluster (left) and phase (center), along with joint probability density functions by KDE (right); (**d**) cluster centroids corresponding to phases 0, 90, 180, 270 deg (from top to bottom).

A second analysis is performed, dividing the field of view into several windows (local region of flow field), based on the vorticity mean and variance associated with each cross-section. Figure 6a shows the maximum mean and variance of cross-sections along the propeller axis. Four windows are selected, as shown in Figure 6a: (1) $0 \leq x/D < 0.4$, 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 destabilizes; (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. This second analysis is used to investigate if global wake behaviors are also present in local regions, or if the latter are dominated by some local behaviors of the wake.

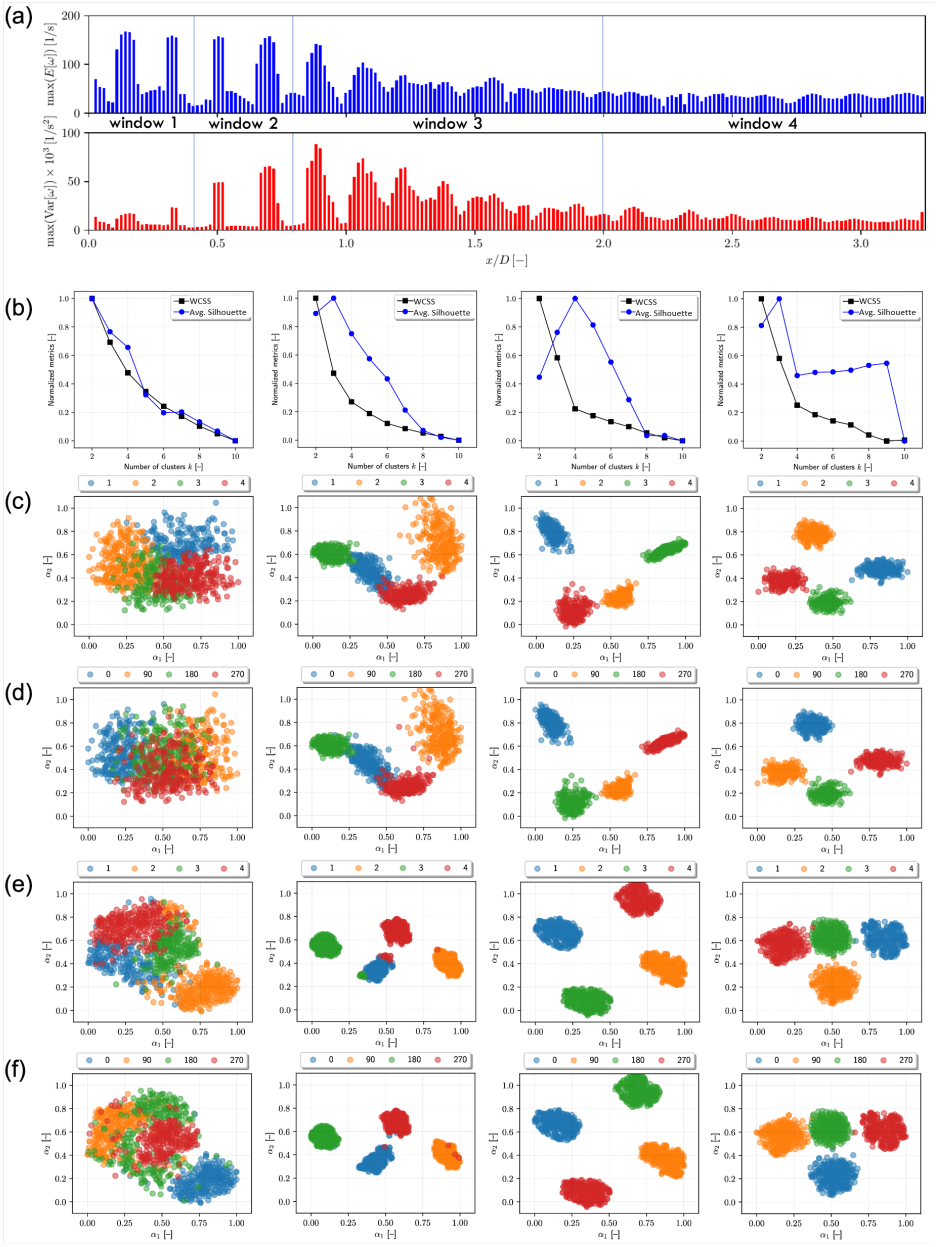

Figure 6. Whole field of view: (**a**) maximum value of mean (top) and variance (bottom) of the vorticity along cross sections; (**b**) clustering metrics by window; and (**c**,**d**) data-projection on the first two POD/PCA modes and (**e**,**f**) embedding via *t*-SNE, label by cluster and phase (from left to right, windows 1, 2, 3, 4).

In Figure 6b, the first column, shows the WCSS and silhouette for window 1. In Figure 6c–f, the first column shows the POD/PCA (c,d) and *t*-SNE (e, f) coefficients labeled by cluster (c, e) and phase (d, f) for the same window. In Figure 7a,b, the first column shows the density functions of the corresponding 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. Consequently, no cluster centroid is shown in Figure 7c. Similarly, the second column

of Figures 6b–f and 7a,b provide the results for window 2. The clustering results are ambiguous since four and three clusters are identified by WCSS and silhouette, respectively. POD/PCA coefficients show two or three main clusters whereas the *t*-SNE clearly identifies four clusters associated with the propeller phases.

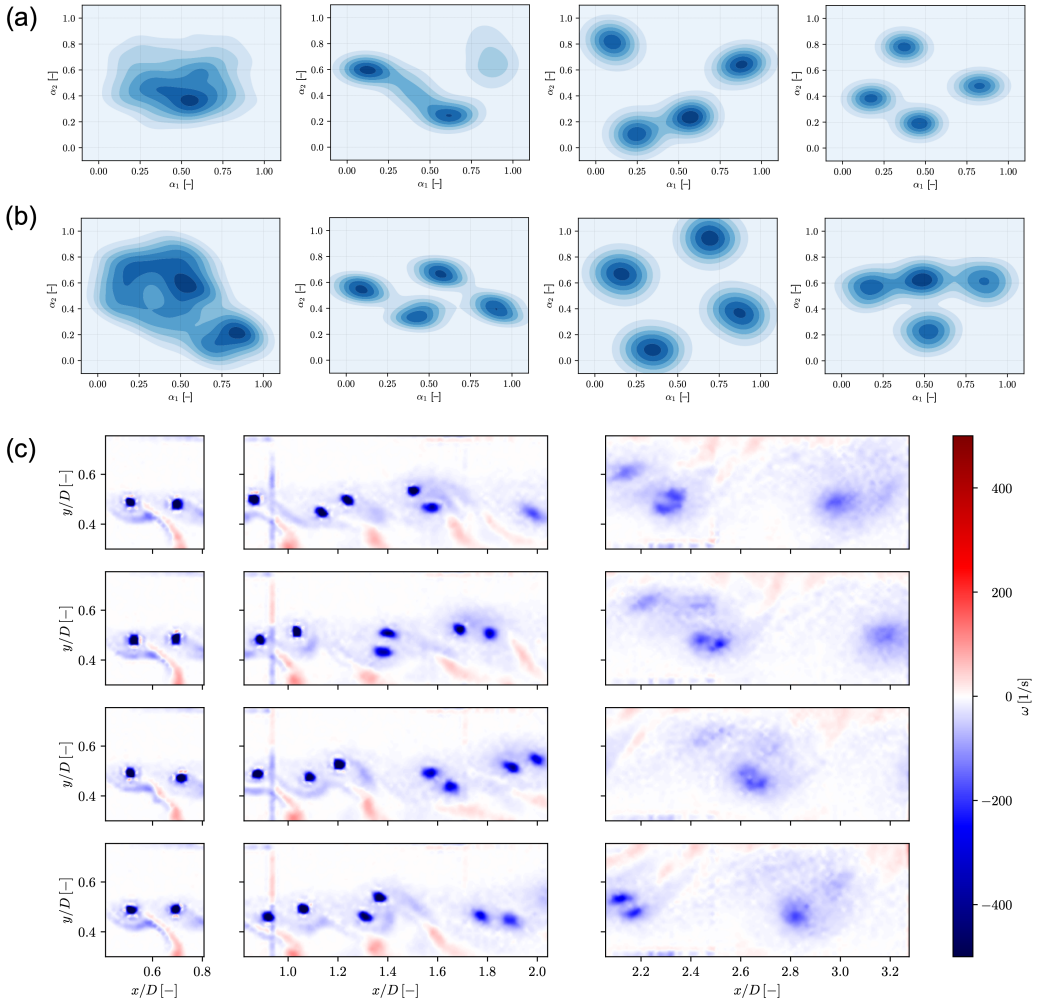

Figure 7. Joint probability density functions by window of (**a**) POD/PCA and (**b**) *t*-SNE coefficients by KDE (from left to right, windows 1, 2, 3, and 4.); (**c**) cluster centroids corresponding to phases 0, 90, 180, 270 deg, for windows 2, 3, and 4 (from left to right, windows).

The results for windows 3 and 4 are presented in the third and fourth column of Figures 6b–f and 7a,b, respectively, where four clusters are clearly identified with a one-to-one association to the phase. It may be noted how a high degree of determinism is still present far downstream of the propeller. It may be also noted how window 4 *t*-SNE analysis presents some hints of transition towards a different clustering structure. Figure 7c shows the cluster centroids associated with 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 to a single vortex (see Figure 8). The clustering results are ambiguous in this region (see Figure 8a), 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 (see Figure 8b,c). Cluster centroids are shown in Figure 8d, where each column represents a centroid. In this case,

the centroids cannot be associated with the phase-locked averages and their interpretation needs further investigation.

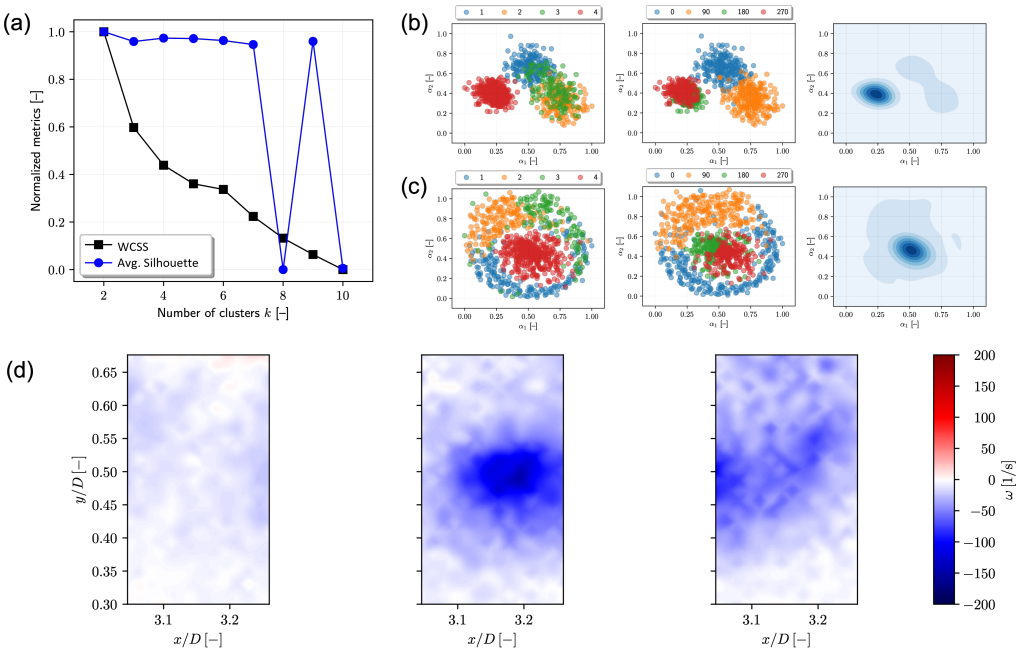

Figure 8. Window 5: (**a**) clustering metrics; (**b**) data-projection on the first two POD/PCA modes; and (**c**) embedding via *t*-SNE, labeled by cluster (left) and phase (center), along with joint probability density functions by KDE centroids (right); (**d**) cluster centroids.

5. Conclusions and Future Work

A snapshot clustering approach was presented and discussed for PIV data of a 4-bladed propeller wake. Data clustering was performed using the *k*-means algorithm, and the optimal number of clusters was determined by evaluating two metrics: the within-cluster sum of squares and average silhouette. The focus of the snapshot clustering was on the vorticity field of the phase-locked propeller's wake data, specifically targeting the tip vortices. To visually analyze and discuss the data clusters, a two-dimensional visualization was generated using POD/PCA, *t*-SNE embedding, and KDE.

The 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 unstable regimes and windows in the far field. Clustering results suggested that the wake instability and subsequent progression of tip vortices are characterized by mechanisms governed by deterministic chaos also in the far field. This result is consistent with the results of the experimental study conducted by [8], which highlights a deterministic behavior of the propeller tip vortices at the macro-scale and chaotic dynamics at the micro-scale. Finally, it can be highlighted that the use of phase-locked measurements, in combination with machine learning and data mining approaches, has allowed us to deeply investigate the propeller wake behaviors, without bothering the use of time-resolved measurements, which are definitely more challenging and expensive.

Ongoing and future work includes extending the analysis of clustering approaches to cover both spatial and temporal clustering 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 a 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 [39] are also part of the ongoing research.

Author Contributions: Conceptualization, M.D. and M.F.; methodology, D.D., M.D., M.F. and A.S.; software, D.D.; formal analysis, D.D., M.D., M.F. and A.S.; investigation, D.D., M.D., M.F. and A.S.; data curation, M.F.; writing—original draft preparation, D.D., M.D. and A.S.; writing—review and editing, M.D., M.F. and A.S.; visualization, D.D. and A.S.; supervision, M.D. and M.F.; funding acquisition, M.D. and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: Authors are grateful to the Office of Naval Research and Office of Naval Research Global for their support through grants N62909-18-1-2033, N62909-19-1-2001, and N00014-22-1-2827. CNR-INM is also grateful to the Italian Ministry of University and Research for the support through the project NAUSICA, n. ARS01_00334, under the Italian National Research Program PON "Ricerca e Innovazione" 2014–2020.

Data Availability Statement: The data that support the findings of this study might be made available from the authors only upon reasonable request and with the permission of ONR.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Nomenclature

EFD	Experimental fluid dynamics
KDE	Kernel density estimation
NLDR	Non-linear dimensionality reduction
PCA	Principal component analysis
PDF	Probability density function
PIV	Particle image velocimetry
POD	Proper orthogonal decomposition
t-SNE	*t*-distributed stochastic neighbor embedding
WCSS	Within-cluster sum of squares

References

1. Frandsen, S.; Thøgersen, M.L. Integrated fatigue loading for wind turbines in wind farms by combining ambient turbulence and wakes. *Wind. Eng.* **1999**, *23*, 327–339.
2. Leung, D.Y.; Yang, Y. Wind energy development and its environmental impact: A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1031–1039. [CrossRef]
3. Stevens, R.; Meneveau, C. Flow structure and turbulence in wind farms. *Annu. Rev. Fluid Mech* **2017**, *49*, 311–339. [CrossRef]
4. Felli, M.; Camussi, R.; Di Felice, F. Mechanisms of evolution of the propeller wake in the transition and far fields. *J. Fluid Mech.* **2011**, *682*, 5–53. [CrossRef]
5. Felli, M.; Falchi, M. A parametric survey of propeller wake instability mechanisms by detailed flow measurement and time resolved visualizations. In Proceedings of the 32nd Symposium on Naval Hydrodynamics, Hamburg, Germany, 5–10 August 2018.
6. Widnall, S.E. The stability of a helical vortex filament. *J. Fluid Mech.* **1972**, *54*, 641–663. [CrossRef]
7. Okulov, V. On the stability of multiple helical vortices. *J. Fluid Mech.* **2004**, *521*, 319–342. [CrossRef]
8. Felli, M.; Falchi, M. Propeller tip and hub vortex dynamic from the near to the very far field by time-resolved PIV measurements. In Proceedings of the the 33rd Symposium on Naval Hydrodynamics, Osaka, Japan, 18–23 October 2020.
9. Wang, L.; Martin, J.E.; Felli, M.; Carrica, P.M. Experiments and CFD for the propeller wake of a generic submarine operating near the surface. *Ocean Eng.* **2020**, *206*, 107304. [CrossRef]
10. Felli, M.; Falchi, M.; Pereira, F.J.A. Distance effect on the behavior of an impinging swirling jet by PIV and flow visualizations. *Exp. Fluids* **2010**, *48*, 197–209. [CrossRef]
11. Felli, M.; Falchi, M. Propeller wake evolution mechanisms in oblique flow conditions. *J. Fluid Mech.* **2018**, *845*, 520–559. [CrossRef]
12. Berkooz, G.; Holmes, P.; Lumley, J.L. The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **1993**, *25*, 539–575. [CrossRef]
13. Gordeyev, S.V.; Thomas, F.O. Coherent structure in the turbulent planar jet. Part 2. Structural topology via POD eigenmode projection. *J. Fluid Mech.* **2002**, *460*, 349–380. [CrossRef]
14. Zhou, X.; Hitt, D.L. Proper orthogonal decomposition analysis of coherent structures in a transient buoyant jet. *J. Turbul.* **2004**, *5*, N28. [CrossRef]
15. Serani, A.; Durante, D.; Diez, M.; D'Agostino, D.; Clement, S.; Badra, J.; Andre, M.; Habukawa, M.; Bardet, P. PIV Data Clustering of a Buoyant Jet in a Stratified Environment. In Proceedings of the 57th AIAA Aerospace Sciences Meeting, SciTech 2019, San Diego, CA, USA, 7–11 January 2019.
16. Barwey, S.; Raman, V.; Steinberg, A.M. Data-Driven Reduction and Decomposition via Time-Axis Clustering. In Proceedings of the AIAA Scitech 2020 Forum, Orlando, FL, USA, 6–10 January 2020; p. 0365.

17. van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
18. Wu, J.; Wang, J.; Xiao, H.; Ling, J. Visualization of high dimensional turbulence simulation data using t-SNE. In Proceedings of the 19th AIAA Non-Deterministic Approaches Conference, Grapevine, TX, USA, 9–13 January 2017; p. 1770.
19. Perceau, M.; Guibert, P.; Guilain, S. Analysis of a tumbling motion using a clustering algorithm on dual-PIV measurements: Application to the in-cylinder flow of a Miller cycle engine. *Exp. Fluids* **2022**, *63*, 54. [CrossRef]
20. Calvet, A.G.; Dave, M.; Franck, J.A. Unsupervised clustering and performance prediction of vortex wakes from bio-inspired propulsors. *Bioinspiration Biomim.* **2021**, *16*, 046015. [CrossRef]
21. Doijode, P.S.; Hickel, S.; van Terwisga, T.; Visser, K. A machine learning approach for propeller design and optimization: Part I. *Appl. Ocean Res.* **2022**, *124*, 103178. [CrossRef]
22. Sharma, R.; Simula, T.P. Machine-learning classification of two-dimensional vortex configurations. *Phys. Rev. A* **2022**, *105*, 033301. [CrossRef]
23. Felli, M.; Falchi, M.; Dubbioso, G. Tomographic-PIV survey of the near-field hydrodynamic and hydroacoustic characteristics of a marine propeller. *J. Ship Res.* **2015**, *59*, 201–208. [CrossRef]
24. Balaras, A.P.R.B.M.F.M.F.E. Characterization of the wake of a submarine propeller via Large-Eddy simulation. *Comput. Fluids* **2019**, *184*, 138–152. [CrossRef]
25. Scarano, F. Iterative image deformation methods in PIV. *Meas. Sci. Technol.* **2001**, *13*, R1. [CrossRef]
26. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [CrossRef]
27. Drineas, P.; Frieze, A.; Kannan, R.; Vempala, S.; Vinay, V. Clustering large graphs via the singular value decomposition. *Mach. Learn.* **2004**, *56*, 9–33. [CrossRef]
28. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]
29. Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, New Orleans, LA, USA, 7–9 January; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2007 ; pp. 1027–1035.
30. Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*; Springer: Secaucus, NJ, USA, 2006.
31. Golub, G.H.; Reinsch, C. Singular value decomposition and least squares solutions. *Numer. Math.* **1970**, *14*, 403–420. [CrossRef]
32. Cizmas, P.G.; Palacios, A.; O'Brien, T.; Syamlal, M. Proper-orthogonal decomposition of spatio-temporal patterns in fluidized beds. *Chem. Eng. Sci.* **2003**, *58*, 4417–4427. [CrossRef]
33. Holmes, P.; Lumley, J.L.; Berkooz, G.; Rowley, C.W. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*; Cambridge University Press: Cambridge, UK, 2012.
34. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Routledge: Oxfordshire, UK, 2018.
35. Simonoff, J.S. *Smoothing Methods in Statistics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
36. Ketchen, D.J.; Shook, C.L. The application of cluster analysis in strategic management research: An analysis and critique. *Strateg. Manag. J.* **1996**, *17*, 441–458. [CrossRef]
37. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
38. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
39. Kutz, J.N.; Brunton, S.L.; Brunton, B.W.; Proctor, J.L. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*; SIAM: Philadelphia, PA, USA, 2016.