

# PIV Data Clustering of a Buoyant Jet in a Stratified Environment

Andrea Serani<sup>\*</sup>, Danilo Durante,<sup>†</sup> and Matteo Diez<sup>‡</sup>  
*CNR–National Research Council, Rome, Italy*

Danny D’Agostino<sup>§</sup>  
*Sapienza University of Rome, Rome, Italy*

Simon Clement,<sup>¶</sup> Joseph Badra,<sup>||</sup> Matthieu Andre,<sup>\*\*</sup> Masayuki Habukawa,<sup>††</sup> and Philippe Bardet<sup>‡‡</sup>  
*The George Washington University, Washington, DC, USA*

**Three spatial clustering approaches of a high-Reynolds number transient buoyant jet in a linearly stratified environment are applied along with proper orthogonal decomposition to identify similar/consistent regions in the domain of interest. The velocity fields analyzed are obtained from an experimental test with large scale, time-resolved, particle image velocimetry (PIV) measurements. Clustering is performed by the  $k$ -means method considering: (a) cross-section velocity profiles, (b) point-wise energy spectra, and (c) point-wise Reynolds stress tensor components. Three metrics are used for the assessment of clustering approaches, 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. Results are promising and lay the foundation for an in depth analysis of local features of complex flows as well as the formulation of efficient reduced order models.**

## I. Introduction

The proper orthogonal decomposition (POD) is a statistical technique widely used for the identification of coherent structures in turbulent flows [1–3], and, among other applications, has been applied to the analysis of turbulent thermal convection problems [4, 5], study of jet pressure signals [6], modal identification of structures [7–9], and design-space dimensionality reduction in shape optimization problems [10, 11]. POD has been used to study coherent structures in a transient buoyant jet in [12], whereas examples of application to planar jets are provided in [3, 13]. Equivalent to the Karhunen-Loève expansion (KLE) and (in the discrete form) to the principal component analysis (PCA), POD 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 model (ROM) for the flow dynamics. Although POD is a widely used method with well-established global optimality properties [1, 2], it is based on the linearity, stationarity, and ergodicity assumptions and therefore may not be effective and efficient when strongly nonlinear interactions are investigated in transient/non-ergodic flows, and/or when an in depth analysis of local flow features is sought after.

The analysis and interpretation of high-dimensional complex phenomena are generally difficult to achieve, due to difficulties in understanding their underlying structure. Data clustering methods could help achieving a deeper understanding of experimental/numerical data, building hypotheses, identifying anomalies, and detecting dominant features. During the last decades, several clustering techniques have been proposed in different fields, such as computer vision, speech recognition, and bioinformatics. One of the most widely applied method is the  $k$ -means clustering [14, 15], closely related to the centroidal Voronoi tessellation [16]. Cluster-based reduced order modeling (CROM) has

---

<sup>\*</sup>Postdoctoral Research Fellow, INM–Institute of Marine Engineering, andrea.serani@insean.cnr.it

<sup>†</sup>Research Scientist, INM–Institute of Marine Engineering, danilo.durante@cnr.it

<sup>‡</sup>Research Scientist, INM–Institute of Marine Engineering, AIAA Member, matteo.diez@cnr.it

<sup>§</sup>PhD Student, Dept. of Computer, Control, and Management Engineering “A. Ruberti”, danny.dagostino@uniroma1.it

<sup>¶</sup>Engineer at PYTHEAS Technology, Meyreuil, France. Formerly Postdoctoral Research Fellow at The George Washington University, Dept. of Mechanical and Aerospace Engineering, simon.clement.pro@gmail.com

<sup>||</sup>Undergraduate Research Assistant, Dept. of Mechanical and Aerospace Engineering, jba@gwmail.gwu.edu

<sup>\*\*</sup>Research Professor, Dept. of Mechanical and Aerospace Engineering, matandre@gwu.edu

<sup>††</sup>Graduate Research Assistant, Dept. of Mechanical and Aerospace Engineering, masa19930523@gwmail.gwu.edu

<sup>‡‡</sup>Associate Professor, Dept. of Mechanical and Aerospace Engineering, AIAA Member, bardet@gwu.edu

been recently applied for the analysis of a mixing layer from direct numerical simulation [17] and a flow in the wake of a high speed train by large eddy simulations [18] in combination with classical POD analysis. Similar extensions of the PCA have been developed and investigated in several technological fields. Among others, local PCA (LPCA) [19] divides an initial data set into clusters and the PCA is then applied to each cluster, assuming an approximate linear structure within the cluster. LPCA has been applied, e.g., to structural damage diagnosis [20], image compression [19], and design-space dimensionality reduction in shape optimization [21]. In general, the number of clusters and distance function used as similarity metrics highly affect the quality of the clustering results and the resulting ROM.

The objective of the present work is the application and comparison of different spatial clustering approaches, based on the  $k$ -means algorithm, for a high-Reynolds number transient buoyant jet in a linearly stratified environment. The velocity fields analyzed are obtained from an experimental test with large scale, time-resolved, particle image velocimetry (PIV) measurements.

Three clustering approaches are applied in the spatial domain, by selecting as clustering variables: (a) cross-section velocity profiles, (b) point-wise energy spectra, and (c) point-wise Reynolds stress tensor components. Three metrics are proposed for their assessment, 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. The jet dynamics is observed, limited to the period corresponding to when the flow is on.

The manuscript is organized as follows. Section II describes the experimental facility and methods for the investigation of the buoyant jet. Section III briefly presents the POD notation and implementation used here, whereas section IV describes the clustering method. Section V and VI present the clustering approaches and their evaluation metrics, respectively. Finally, results are presented in Section VII, whereas conclusions are discussed in Section VIII.

## II. Experimental Method

The experiment is the vertical discharge of a high-Reynolds number buoyant jet 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 [22].

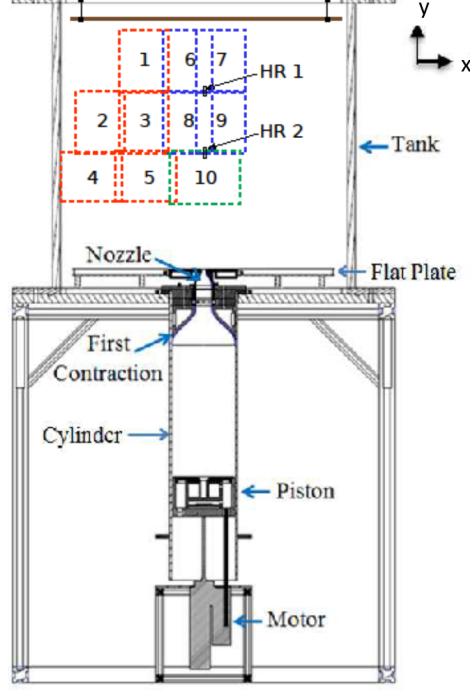
The facility produces a round vertical jet inside a clear acrylic tank (cube of 914 mm-side), Fig. 1. A linear motor drives a piston in a 203 mm-diameter cylinder, which pushes the fluid through a first contraction section followed by a contoured nozzle with a  $D = 6.35$  mm exit diameter. The fluid in the cylinder is initially at rest and the jet has no initial disturbances. The jet Reynolds number is  $Re_D = U_o D / \nu$  is  $2.00 \times 10^4$ . The run time, limited by the stroke of the piston, is 39 s. The change in height in the tank from each run is 5.4 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. A total of 10 cameras is employed.

Cameras 1 to 9 are 1.3 MPixel USB 3.0 and record the off-center and far field of the jet. They record either at 64 or 128 Hz depending on their radial location. They are integrated in groups of 4-5 into a single workstation and controlled with StreamPix software. Camera 10 is a CMOS cameras with CoaXpress transfer protocol. It records on the centerline of the jet at 512 Hz at 4 MPixels.

Three large laser sheets illuminate the fields of view of cameras 1 to 10. They are split from a single cavity of a dual cavity Nd:YLF laser (Photonics DM 527) operated at nearly 30 mJ/pulse. The performance of the beam splitters used to generate the laser sheets depends on polarization of the incoming light. To control the energy distribution between the planes (to accommodate the various sensitivities of the cameras) the polarization of the laser beam is adjusted with a waveplate. Each laser is configured as a telescope, with a nearly constant 3 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 with an in-house Matlab code, applying a sliding average over the areas where cameras overlap.



**Fig. 1** Experimental facility with the location of the PIV planes. Cameras 1 to 5 record at 64 Hz, cameras 6 to 9 at 128 Hz and camera 10 at 512 Hz.

### III. Proper Orthogonal Decomposition: Notation and Implementation

Reynolds decomposition of the velocity vector is used, with

$$u = \bar{u} + u' \quad v = \bar{v} + v' \quad (1)$$

where  $u$  and  $v$  indicate  $x$ - and  $y$ -components, respectively (see Fig. 1). 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

$$\mathbf{U} = \left[ \mathbf{u}^{(1)} \mid \dots \mid \mathbf{u}^{(S)} \right] \quad (2)$$

where  $\mathbf{u} = \{u'(\mathbf{x}_1), \dots, u'(\mathbf{x}_P), v'(\mathbf{x}_1), \dots, v'(\mathbf{x}_P)\}^T$  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  $i = 1, \dots, S$ , 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 [23]

$$\mathbf{C} = \frac{1}{S} \mathbf{U} \mathbf{U}^T \quad (3)$$

evaluated by

$$\mathbf{C} \mathbf{Z} = \mathbf{Z} \mathbf{\Lambda} \quad (4)$$

where  $\mathbf{Z}$  and  $\mathbf{\Lambda}$  collect the  $L$  eigenvectors ( $\mathbf{z}_i$ ) and eigenvalues ( $\lambda_i$ ) of  $\mathbf{C}$ , respectively. This corresponds to performing the PCA of the matrix  $\mathbf{U}$ . The problem of Eq. 4 may be alternatively solved using the singular value decomposition, SVD [24]. Furthermore, if  $L > S$  the dual problem may be solved via the so-called snapshot-POD (equivalent) formulation, see e.g. [25].

The eigenvalues (POD eigenvalues) represent the variance (under proper assumption this represents the turbulent kinetic energy) resolved along the corresponding eigenvectors. The linear subspace formed by the  $N$  eigenvectors (POD 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  $N$  [2, 23]. 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  $\hat{\mathbf{U}} = \hat{\mathbf{Z}} \hat{\mathbf{Z}}^T \mathbf{U}$ .

## IV. $k$ -Means Clustering Method

The  $k$ -means is a widely known clustering method [14, 15], which allows to build partitions of the original data  $\mathbf{U}$  in  $k$  different sets (clusters), defined by representative points (centroids). Here, the original data collected in  $\mathbf{U}$  is rearranged as per the clustering approach and criterion used. In general, we refer to  $\xi_i$  as one realization (point) of the rearranged data. Note that, generally,  $\xi_i \in \mathbb{R}^Q$  with  $i = 1, \dots, H$ , where  $Q \neq L$  and  $H \neq S$ .

A metrics (such as the Euclidean distance) is used to measure both the similarity between data points  $\xi_j$  and evaluate the associated cluster centroids  $\mu_i$ , the latter by averaging all data points within the  $i$ -th cluster  $C_i$ . The assignment of data points to clusters is achieved by the minimization of the squared Euclidean distance between  $\xi_j$  and  $\mu_i$  (within-cluster sum of squares, WCSS)

$$\text{WCSS} = \sum_{i=1}^k \sum_{\xi_j \in C_i} \|\xi_j - \mu_i\|^2 \quad (5)$$

where  $k$  is the number of clusters.

It may be noted that the minimization of WCSS in Eq. 5 is a NP-hard problem [26]. For its solution, the heuristic approach presented in [14] is used and briefly presented in Alg. 1. Results are highly sensitive to centroids initialization. Here, the initialization strategy proposed in [27] is used.

---

### Algorithm 1 $k$ -Means Clustering Algorithm [14]

---

**Require:**  $k$  centroids (selected randomly or using the strategy in [27]) as representative points of each cluster

- 1: **repeat**
  - 2:   Assign each point  $\xi_j$  to the cluster  $C_i$  of the nearest centroid  $\mu_i$  using the Euclidean distance as similarity measure
  - 3:   Update the centroids according to:  $\mu_i = \frac{1}{|C_i|} \sum_{\xi_j \in C_i} \xi_j$
  - 4: **until**  $\mu_i \forall i = 1, \dots, k$  remains unchanged
- 

## V. Data Clustering Approaches

Clustering approaches in the spatial domain are proposed, using as clustering variables: (a) cross-section velocity profiles, (b) point-wise energy spectra, and (c) point-wise Reynolds stress tensor components. The formulation underlying each approach is described in the following.

### A. Cross-Section Velocity Profiles

Cross-sections are clustered together, based on their velocity profiles. Firstly, cross-section ( $y$ -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. 5:

$$\frac{v'}{v_c}, \quad \frac{(\overline{u'u'})^{1/2}}{v_{c,l}}, \quad \frac{(\overline{v'v'})^{1/2}}{v_{c,l}}, \quad \frac{\overline{u'v'}}{v_{c,l}^2} \quad (6)$$

where  $v_c$  is the mean axial velocity at the center line;  $v_{c,l}$  is the the mean axial velocity at the center line, assuming idealized profiles from the fully developed region where  $1/v_c$  is linear.

Profile abscissa are scaled assuming that the velocity profile follows a Gaussian distribution and using its standard deviation  $b$ , evaluated numerically as

$$b = \sqrt{\frac{\int_{x_{\min}}^{x_{\max}} (x - x_c)^2 \max[\bar{v}(x), 0] dx}{\int_{x_{\min}}^{x_{\max}} \max[\bar{v}(x), 0] dx}} \quad (7)$$

It may be noted that, under the Gaussian distribution assumption, the 95% of the (positive) flux is contained within  $\pm 2b$ . Similarly to  $v_c$ , abscissa scaling for turbulence variables in Eq. 6 is performed using idealized linearly decreasing values of  $b$  from the fully developed region, referred to as  $b_l$ .



Accordingly,  $\xi_i$  arrays are formed as

$$\xi_i = \begin{pmatrix} v'(\hat{\mathbf{x}}, y_i) & [v_c(y_i)]^{-1} \\ \left[ \overline{u'(\hat{\mathbf{x}}, y_i) \circ u'(\hat{\mathbf{x}}, y_i)} \right]^{1/2} & [v_{c,l}(y_i)]^{-1} \\ \left[ \overline{v'(\hat{\mathbf{x}}, y_i) \circ v'(\hat{\mathbf{x}}, y_i)} \right]^{1/2} & [v_{c,l}(y_i)]^{-1} \\ \overline{u'(\hat{\mathbf{x}}, y_i) \circ v'(\hat{\mathbf{x}}, y_i)} & [v_{c,l}(y_i)]^{-2} \end{pmatrix} \in \mathbb{R}^Q \quad i = 1, \dots, H \quad (8)$$

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

### B. Point-Wise Spectra

Spatial points are clustered together based on their energy spectra. Each point is provided with the energy spectrum [28, 29], namely  $E(f)$ , where  $f$  is the frequency. This is used in logarithmic scale to form clustering arrays in Eq. 5:

$$\xi_i = \left\{ \log_{10} [E(\mathbf{f})] \right\} \in \mathbb{R}^Q \quad i = 1, \dots, H \quad (9)$$

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

### C. 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. 5:

$$\left( \overline{u'u'} \right)^p, \quad \left( \overline{v'v'} \right)^p, \quad \left| \overline{u'v'} \right|^p \quad (10)$$

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

Accordingly, clustering variables collected in  $\xi_i$  are organized as

$$\xi_i = \begin{pmatrix} \left[ \overline{u'(\mathbf{x}_i)u'(\mathbf{x}_i)} \right]^{1/2} \\ \left[ \overline{v'(\mathbf{x}_i)v'(\mathbf{x}_i)} \right]^{1/2} \\ \left| \overline{u'(\mathbf{x}_i)v'(\mathbf{x}_i)} \right|^{1/2} \end{pmatrix} \in \mathbb{R}^Q \quad i = 1, \dots, H \quad (11)$$

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

## VI. Evaluation Metrics

Three metrics are used for the assessment of the clustering approaches and identification of the optimal number of clusters  $k$ , 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. Their definition is included in the following.

### A. Within-Cluster Sum of Squares

The WCSS in Eq. 5 is used as evaluation metrics to identify the optimal number of clusters  $k$ . Specifically, the elbow method [30] is used with the WCSS metrics.

## B. Average Silhouette

The silhouette method provides a metrics of consistency of data within clusters [31]. Assume  $a_i$  as the average Euclidean distance between  $\xi_i$  and any other data point within the cluster  $\xi_i$  belongs to. Assume then  $c_i$  as the smallest average Euclidean distance of  $\xi_i$  to all data points in any other cluster  $\xi_i$  does not belong to.

The silhouette associated to  $\xi_i$  is defined as

$$s_i = \frac{b_i - c_i}{\max[b_i, c_i]} \quad (12)$$

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:

$$s_{\text{avg}} = \sum_{i=1}^H s_i \quad (13)$$

## C. Variance/Energy of POD Representation

The relative variance/energy resolved by within-cluster POD representations of order  $N$  is also used as a metrics to identify the optimal number of clusters  $k$ :

$$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^k \sum_{j=1}^N \lambda_j^{(i)}}{\sum_{j=1}^L \lambda_j^{(\text{global})}} \quad (14)$$

where  $\lambda_j^{(k)}$  is the  $j$ -th POD eigenvalue associated to the  $i$ -th cluster, provided that  $\lambda_j^{(i)} > \lambda_{j+1}^{(i)}$ . Subscript 'global' refers to the global 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. 14. This assumption will be removed in future studies, where an optimal order  $N$  will be investigated within each cluster.

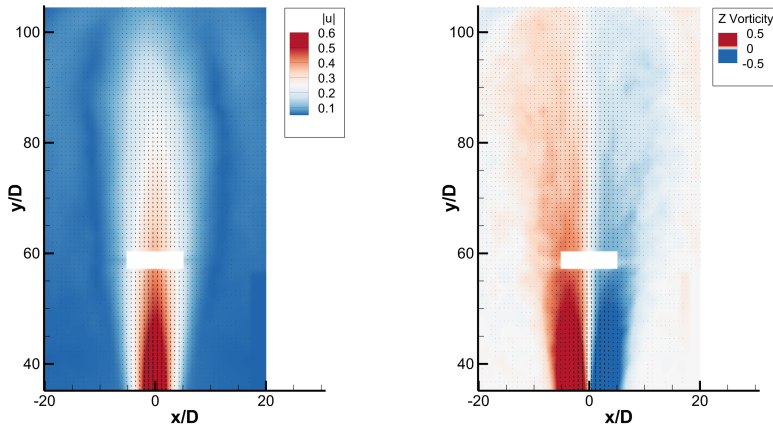
## VII. Results

The jet dynamics is studied, limited to the period corresponding to when the actuator is on. In the following, the (global) field of view is set in the range  $x/D \in [-20; 20]$  and  $y/D \in [35; 105]$ . The domain is discretized (sub-sampled) so as to have 52 nodes along  $x$  and 89 nodes along  $y$ . A total number of 4096 snapshots, sampled over time at 128 Hz, are considered. Accordingly, the data matrix  $\mathbf{U}$  has size equal to  $2P \times S = 9,256 \times 4,096$ . The results that follow present a blanked region across the center line at about  $y/D = 60$ . This corresponds to noisy data in the PIV data set that, therefore, were not processed.

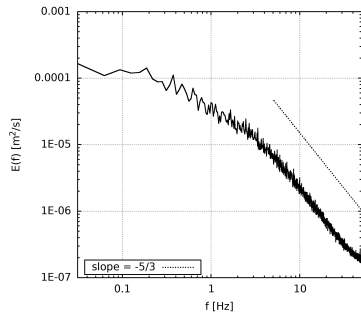
### A. Mean Flow, Energy Spectrum, and Proper Orthogonal Decomposition of the Global Field of View

Figure 2 shows the mean flow. The magnitude of velocity vector (left) and  $z$ -component of vorticity vector (right) are shown. Figure 3 shows the energy spectrum associated to the global field of view. It may be noted how, globally, the inertial subrange is well visible with a slope close to  $-5/3$ .

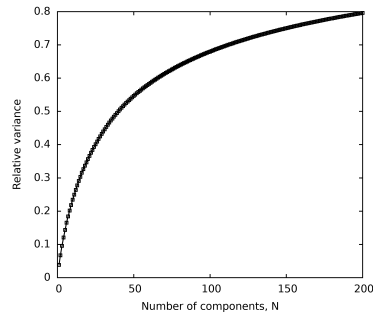
Results of the POD analysis are shown in Figs. 4 and 5. Specifically, Fig. 4 shows the relative variance/energy retained by a POD representation of order  $N$ . Figure 5 (a) shows the first POD mode, whereas Fig. 5 (b) shows the second modes. Velocity-vector magnitude and  $z$ -component of vorticity vector are presented on the left and right, respectively. It may be noted qualitatively how in the first mode there is a more significant participation of the top part of the domain, whereas in the second mode a more significant participation of the bottom part is observed. Finally, Fig. 5 (c) shows a higher-order mode, namely the fiftieth. The contribution of POD modes to the energy spectrum is part of ongoing studies and will be presented in future work.



**Fig. 2 Mean flow.**



**Fig. 3 Energy spectrum (global).**



**Fig. 4 Reconstructed variance/energy of (global) POD versus number of components/modes.**

## B. Clustering by Cross-Section Velocity Profiles

Scaling variables for cross-section velocity profiles are shown in Fig. 6. It may be noted how the jet-width parameter ( $b$ ) and the inverse axial velocity at the center line ( $1/v_c$ ) grow linearly with the distance from the jet outlet up to 60 and 75 diameters, respectively. Scaled profiles are depicted in Fig. 7. For current application, it is  $Q = 4,000$  and  $H = 89$ . It may be seen how the positive part of the mean axial velocity, Fig. 7 (a) scales quite well with the center-line mean axial velocity  $v_c$  (vertically) and with  $b$  provided by Eq. 7 (horizontally). Turbulence-related variables scale reasonably well with the idealized linear values  $v_{c,l}$  and  $b_l$ , as Figs. 7 (b)-(d) show.

The  $k$ -means method is applied to the scaled profiles, stitched together following Eq. 8. Evaluation metrics versus the number of clusters  $k$  are presented in Figs. 8 and 9 (a). Specifically, Fig. 8 (a) shows the normalized WCSS, where normalization is performed using WCSS value for  $k = 1$ . Figure 8 (b) shows the average silhouette of clustered data. Figure 8 (c) shows the within-cluster number of POD components  $N$  required to resolve the 50% of the total variance/energy. Finally, 9 (a) depicts the relative variance/energy resolved using  $k$  clusters, each with a POD representation of order  $N$ . We have identified two possible values of  $k$ , representing a good compromise among the different metrics, namely  $k = 2$  and 5. It may be noted that these two values correspond to gaps or jumps in the convergence curves of Figure 9 (a).

The following results pertain to  $k = 2$ , whereas we will leave  $k = 5$  for future studies. Figure 10 (a) shows how the cross sections are clustered together. Figure 11 shows the representative scaled profiles (centroids) of cluster #1 and 2, where cluster #1 is likely mainly momentum driven, whereas cluster #2 buoyancy driven. The energy spectra associated to clusters are presented in Fig. 12 (a). The inertial subrange is well visible for cluster #1, with a slope close to  $-5/3$ . Cluster #2, which is characterized by a lower local Re, brings into light also a slope close to  $-3$ . Finally, the first POD mode of clusters #1 and 2 is shown in Fig. 13, where a similar structure to Fig. 5 is observed.

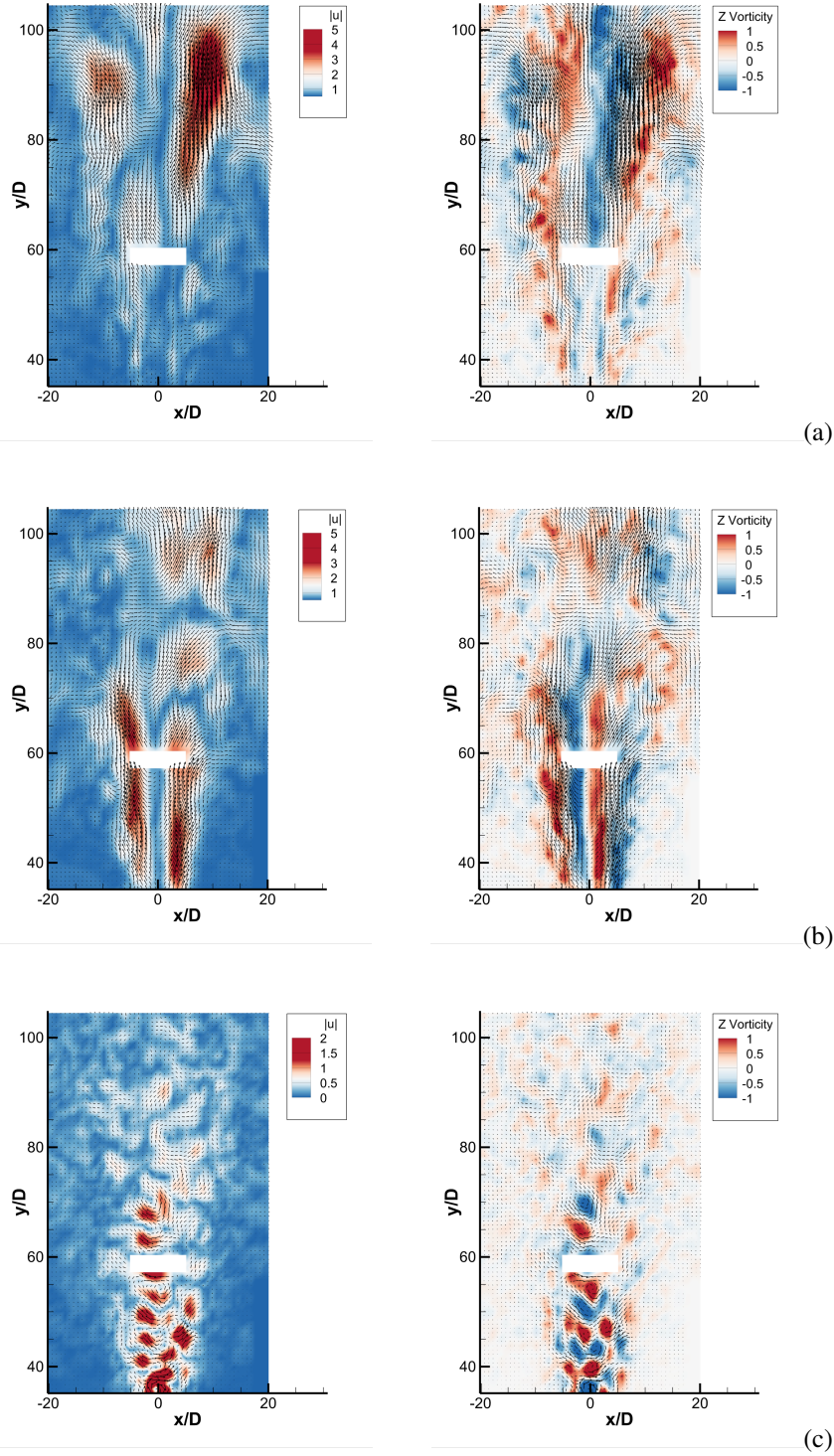
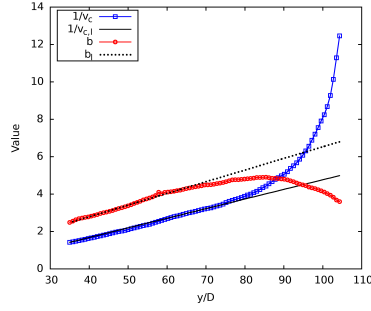


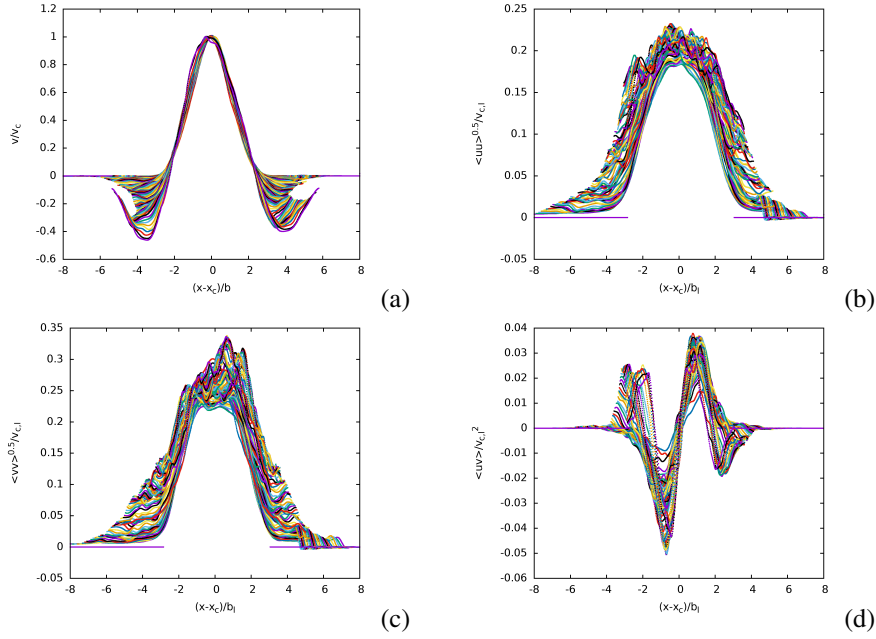
Fig. 5 Global field of view: POD mode #1 (a), #2 (b), and #50 (c).

### C. Clustering by Point-Wise Energy Spectra

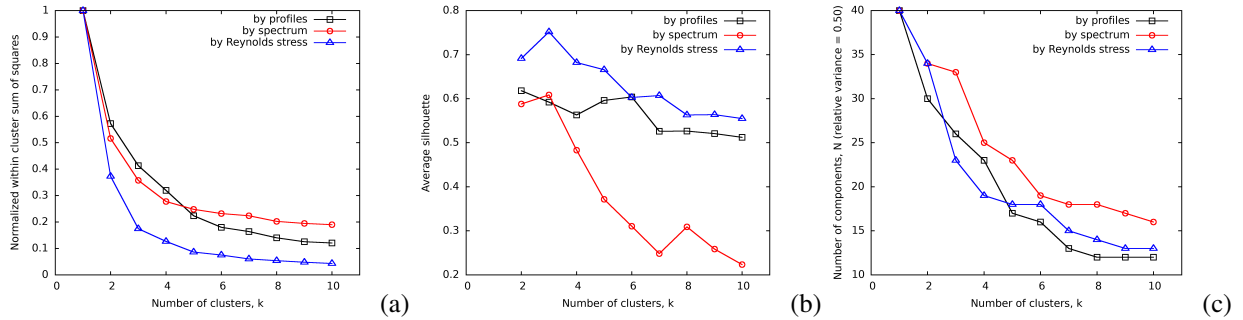
The  $k$ -means method is applied to point-wise energy spectra, following Eq. 9. Here,  $Q = 2,048$  and  $H = P = 4,628$ . Evaluation metrics versus the number of clusters  $k$  are presented in Figs. 8 and 9 (b). A number of clusters  $k = 4$  was



**Fig. 6** Scaling variables for cross-section velocity profiles.



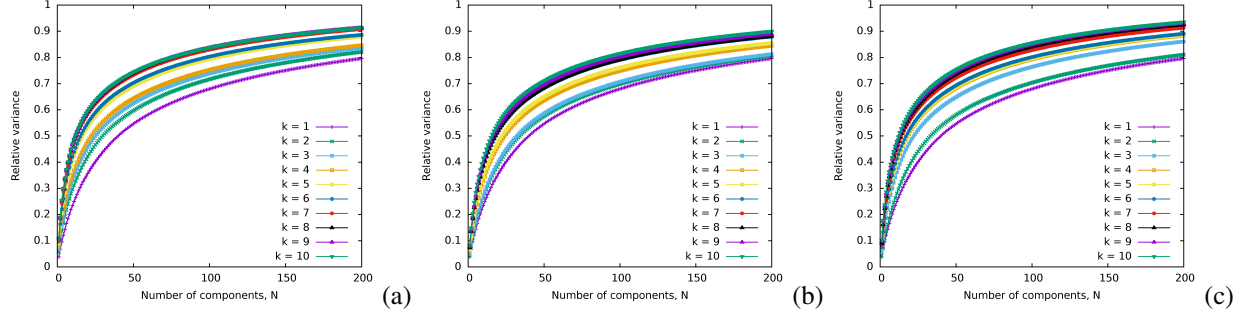
**Fig. 7** Cross-section velocity profiles (scaled): (a) mean axial velocity, (b)  $u'$  RMS, (c)  $v'$  RMS, and (d) Reynolds stress.



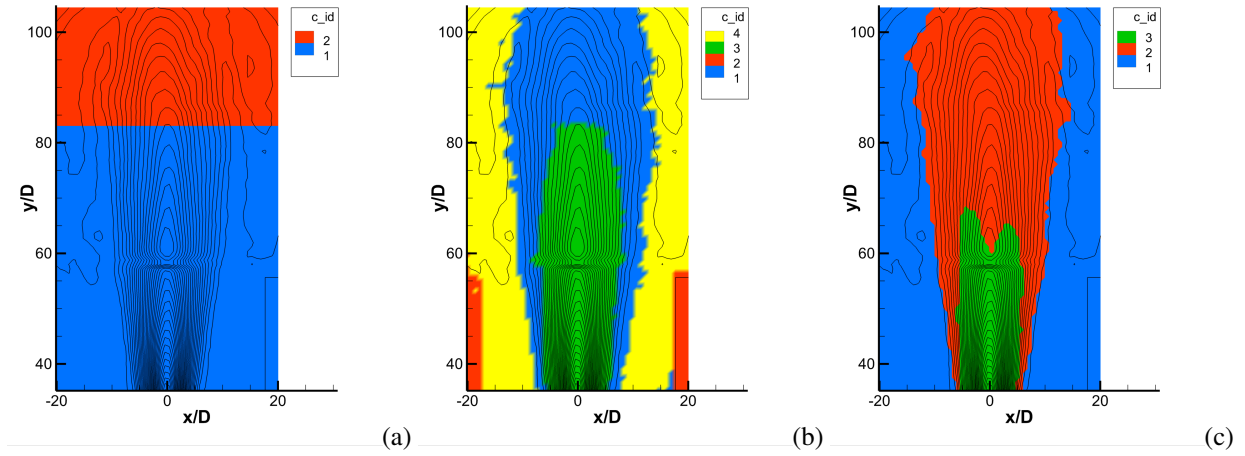
**Fig. 8** Evaluation metrics for optimal number of clusters: (a) normalized WCSS, (b) average silhouette, and (c) within-cluster number of POD modes required to resolve 50% of total variance/energy.

identified as a good compromise among all metrics. This corresponds to a quite visible jump in the convergence curves of Fig. 9 (b).

Figure 10 (b) shows how the the points are clustered together. The energy spectra associated to clusters are presented



**Fig. 9** Relative variance/energy of POD representation: clustering by (a) cross-section velocity profiles, (b) point-wise energy spectra, and (c) point-wise Reynolds stress components.



**Fig. 10** Data clustering in the spatial domain: (a) cross-section velocity profiles, (b) point-wise spectra, and (c) point-wise Reynolds stress tensor components.

in Fig. 12 (b). All spectra present a reasonable distribution of energy in the inertial subrange, with a slope close to  $-5/3$ . As expected, the energy content associated to each spectrum depends on the proximity of clustered points to the jet core. Clusters far from the core experience a smaller local Re and bring into light a slope close to  $-3$  in the high-frequency range.

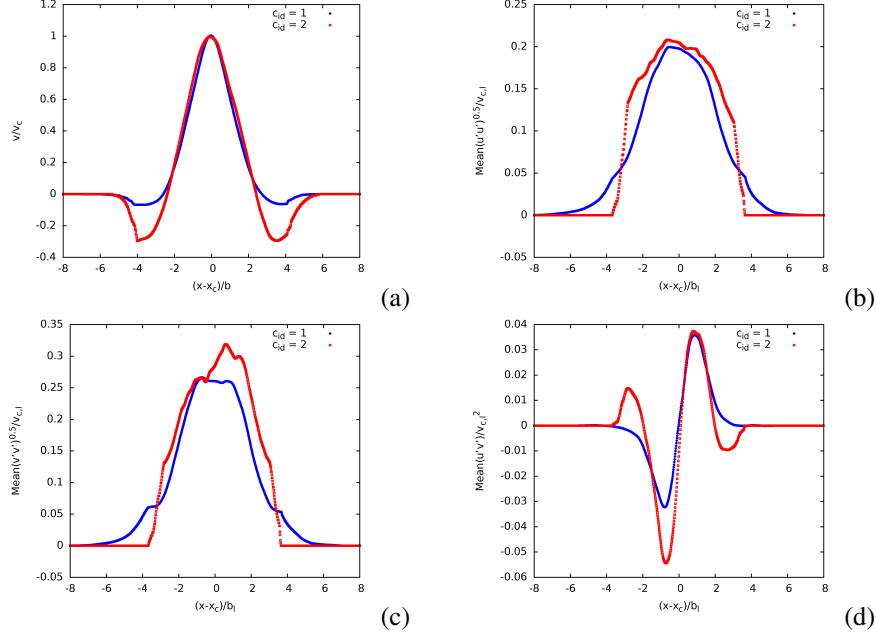
Finally, the first POD mode of clusters #1 and 3 is shown in Fig. 13, where, again, a similar structure to Fig. 5 is observed.

#### D. Clustering by Point-Wise Reynolds Stress Tensor Components

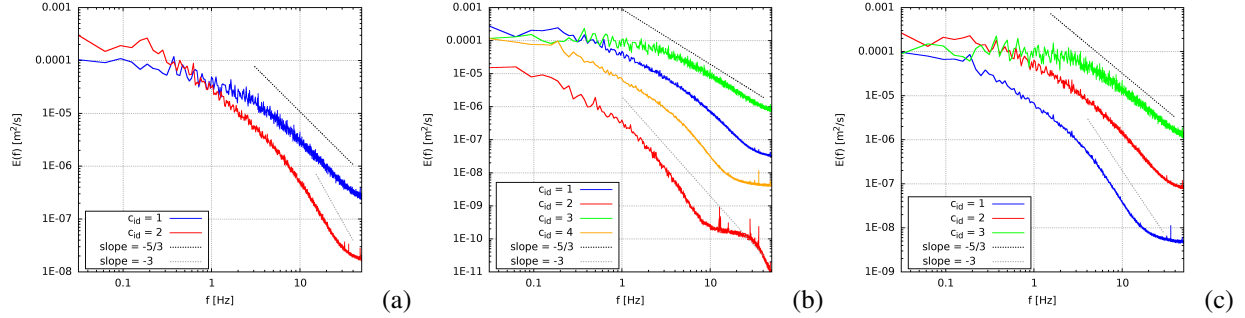
Point-wise Reynolds stress tensor components as used as clustering variables, as per Eq. 10. In this case,  $Q = 3$  and  $H = P = 4,628$ . Clustering metrics are presented in Figs. 8 and 9 (c). A number of cluster  $k = 3$  is selected as an optimal choice for all metrics. It may be noted the significant jump in the convergence curves of Fig. 9 (c), passing from  $k = 2$  to  $k = 3$ .

Figure 10 (c) shows how the the points are clustered together. Figure 15 shows the cluster centroids in terms of Reynolds stress tensor components. The energy spectra associated to clusters are presented in Fig. 12 (c). The same considerations made for the previous case apply. Here, it may be noted as all three clusters have a similar energy content at large scales. At smaller scales, they experience a similar direct energy cascade to the previous case.

Finally, the first POD mode of clusters #2 and 3 is shown in Fig. 13, where, also in this case, a similar structure to Fig. 5 is observed.



**Fig. 11** Cluster centroids using cross-section velocity profiles: (a) mean axial velocity, (b)  $u'$  RMS, (c)  $v'$  RMS, and (d) Reynolds stress.



**Fig. 12** Energy spectrum associated to clusters: (a) cross-section velocity profiles, (b) point-wise spectra, and (c) point-wise Reynolds stress tensor components.

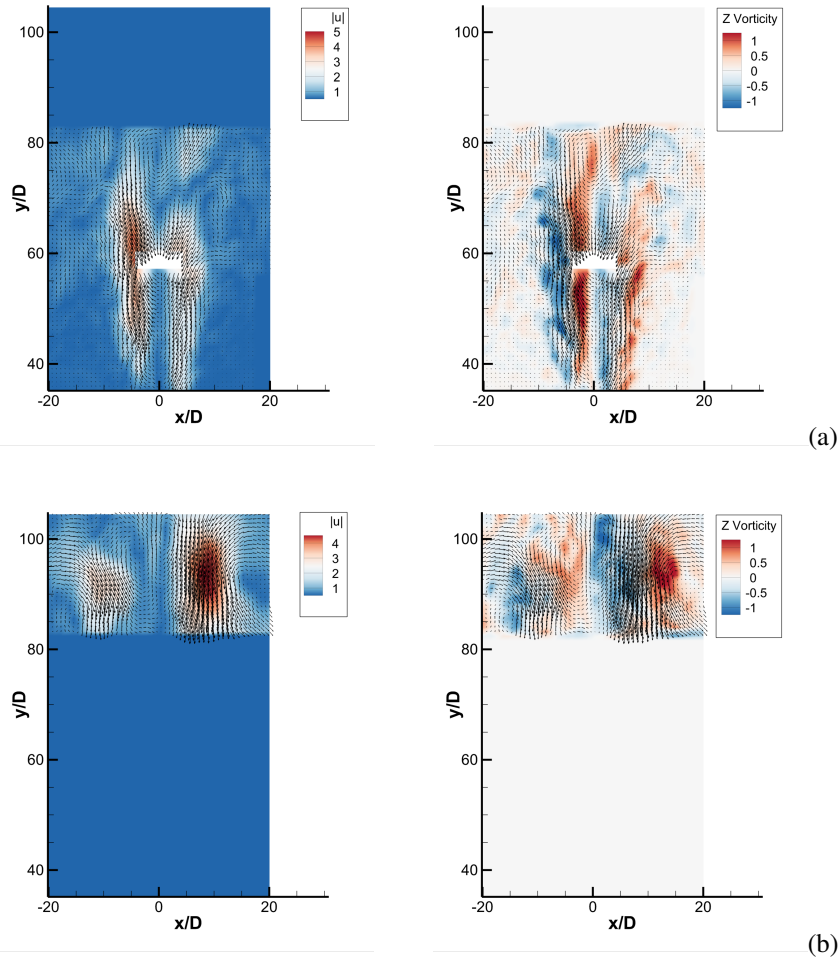
## VIII. Conclusions, Ongoing, and Future Work

Three spatial clustering approaches of a high Reynolds number transient buoyant jet in a linearly stratified environment have been applied along with POD analysis to identify similar/consistent regions in the domain of interest. Experimental data from large scale, time-resolved, PIV measurements have been used. Specifically, clustering has been performed by the  $k$ -means method considering: (a) cross-section velocity profiles, (b) point-wise energy spectra, and (c) point-wise Reynolds stress tensor components. Three metrics have been used for the assessment of clustering results, 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.

Clustering performance of different approaches can be compared in Fig. 8, as well as Fig. 17. Among current approaches and for the current application, using the Reynolds stress tensor components appears the most efficient and effective option, as a small number of clusters  $k$  are required to have a small normalized WCSS, see Fig. 8 (a), a large average silhouette, see Fig. 8 (b), and an efficient POD representation, see Fig. 8 (c) and 17. Furthermore, clustering arrays have a small dimensionality ( $Q = 3$ ), making the clustering process quite efficient. Overall, results are promising and lay the foundation for an in depth analysis of local features of complex flows as well as the formulation of efficient reduced order models.

Ongoing and future work include adding more clustering variables/approaches for comparison both in spatial





**Fig. 13 Spatial clustering by cross-section velocity profiles: first POD mode for cluster #1 (a) and #2 (b).**

and time domains, as well as a detailed study on scaling/normalization effects. Comparison to jets in homogeneous environments will be carried out as well as analyses with an extended field of view. Alternative methods [15] to the  $k$ -means clustering will be also considered for further studies and comparisons.

### Acknowledgments

The work is supported by the US Department of the Navy Office of Naval Research Global, NICOP grant N62909-18-1-2033, administered by Dr. Salahuddin Ahmed and Dr. Woei-Min Lin, and by the Italian Flagship Project RITMARE, funded by the Italian Ministry of Education.

### References

- [1] Berkooz, G., Holmes, P., and Lumley, J. L., “The proper orthogonal decomposition in the analysis of turbulent flows,” *Annual Review of Fluid Mechanics*, Vol. 25, No. 1, 1993, pp. 539–575.
- [2] Holmes, P., Lumley, J. L., Berkooz, G., and Rowley, C. W., *Turbulence, coherent structures, dynamical systems and symmetry*, Cambridge university press, 2012.
- [3] 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, p. 349–380. doi:10.1017/S0022112002008364.



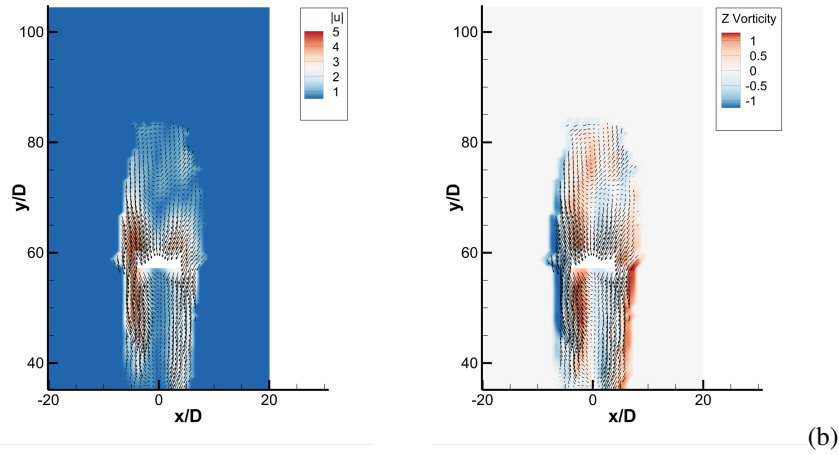
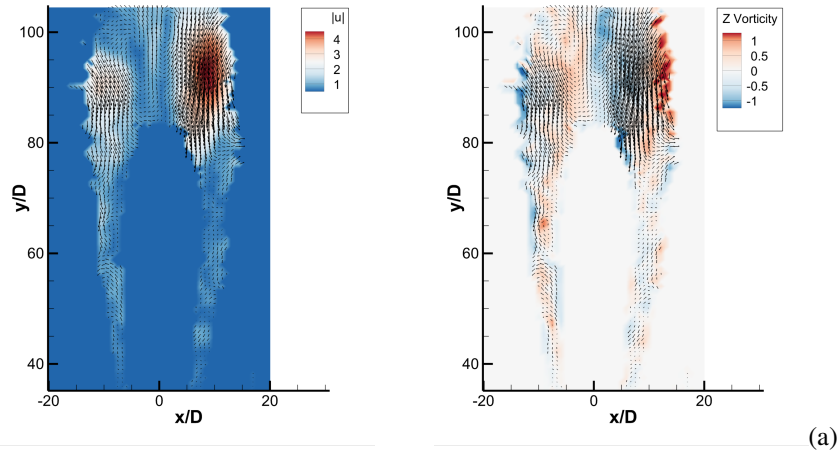


Fig. 14 Spatial clustering by point-wise spectra: first POD mode for cluster #1 (a) and #3 (b).

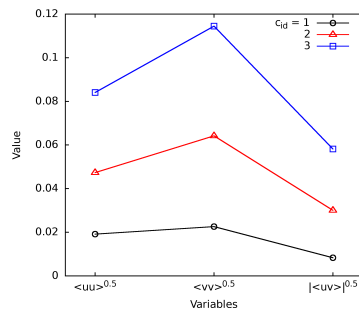


Fig. 15 Cluster centroids using point-wise Reynolds stress tensor components.

- [4] Tarman, H. I., “A Karhunen-Loève Analysis of Turbulent Thermal Convection,” *International Journal for Numerical Methods in Fluids*, Vol. 22, 1996.
- [5] Ly, H. V., and Tran, H. T., “Modeling and control of physical processes using proper orthogonal decomposition,” *Mathematical and Computer Modelling*, Vol. 33, No. 1, 2001, pp. 223 – 236. doi:[https://doi.org/10.1016/S0895-7177\(00\)00240-5](https://doi.org/10.1016/S0895-7177(00)00240-5), URL <http://www.sciencedirect.com/science/article/pii/S0895717700002405>, computation and control VI proceedings of the sixth Bozeman conference.

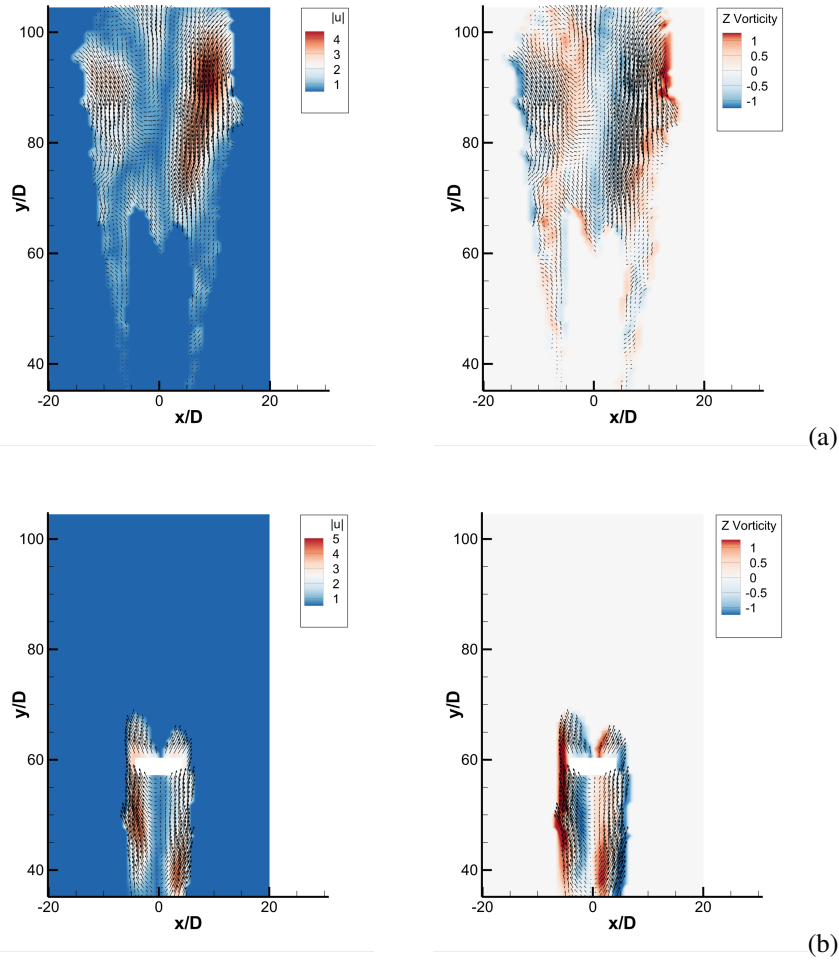


Fig. 16 Spatial clustering by Reynolds stress tensor components: first POD mode for cluster #2 (a) and #3 (b).

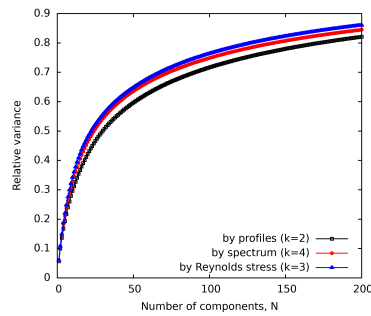


Fig. 17 Relative variance/energy of POD representation versus clustering method.

- [6] Mancinelli, M., Pagliaroli, T., Camussi, R., and Castelain, T., “On the interpretation of pressure POD modes in the near field of a subsonic jet in terms of hydrodynamic and acoustic pressures,” *2018 AIAA/CEAS Aeroacoustics Conference, AVIATION 2018*, Atlanta, GA, USA, June 25-29, 2018.
- [7] Iemma, U., Morino, L., and Diez, M., “Digital holography and Karhunen–Loève decomposition for the modal analysis of two-dimensional vibrating structures,” *Journal of sound and vibration*, Vol. 291, No. 1, 2006, pp. 107–131.

- [8] Iemma, U., Diez, M., and Morino, L., “An extended Karhunen–Loève decomposition for modal identification of inhomogeneous structures,” *Journal of vibration and acoustics*, Vol. 128, 2006, p. 357.
- [9] Mariani, R., and Dessi, D., “Analysis of the global bending modes of a floating structure using the proper orthogonal decomposition,” *Journal of Fluids and Structures*, Vol. 28, 2012, pp. 115–134.
- [10] Diez, M., Campana, E. F., and Stern, F., “Design-space dimensionality reduction in shape optimization by Karhunen–Loève expansion,” *Computer Methods in Applied Mechanics and Engineering*, Vol. 283, 2015, pp. 1525–1544.
- [11] Serani, A., and Diez, M., “Shape Optimization under Stochastic Conditions by Design-space Augmented Dimensionality Reduction,” *19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (MA&O), AVIATION 2018*, Atlanta, GA, USA, June 25-29, 2018.
- [12] Zhou, X., and Hitt, D., “Proper orthogonal decomposition analysis of coherent structures in a transient buoyant jet,” *Journal of Turbulence*, Vol. 5, No. 28, 2004, pp. 1–3.
- [13] Shim, Y.-M., Sharma, R., and Richards, P., “Proper orthogonal decomposition analysis of the flow field in a plane jet,” *Experimental Thermal and Fluid Science*, Vol. 51, 2013, pp. 37–55.
- [14] Lloyd, S., “Least squares quantization in PCM,” *IEEE transactions on information theory*, Vol. 28, No. 2, 1982, pp. 129–137.
- [15] Jain, A. K., “Data clustering: 50 years beyond K-means,” *Pattern recognition letters*, Vol. 31, No. 8, 2010, pp. 651–666.
- [16] Burkardt, J., Gunzburger, M., and Lee, H.-C., “POD and CVT-based reduced-order modeling of Navier–Stokes flows,” *Computer Methods in Applied Mechanics and Engineering*, Vol. 196, No. 1, 2006, pp. 337 – 355. doi:<https://doi.org/10.1016/j.cma.2006.04.004>, URL <http://www.sciencedirect.com/science/article/pii/S0045782506001575>.
- [17] Kaiser, E., Noack, B. R., Cordier, L., Spohn, A., Segond, M., Abel, M., Daviller, G., Östh, J., Krajnović, S., Niven, R. K., and et al., “Cluster-based reduced-order modelling of a mixing layer,” *Journal of Fluid Mechanics*, Vol. 754, 2014, p. 365–414. doi:10.1017/jfm.2014.355.
- [18] Östh, J., Kaiser, E., Krajnović, S., and Noack, B. R., “Cluster-based reduced-order modelling of the flow in the wake of a high speed train,” *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 145, 2015, pp. 327 – 338. doi:<https://doi.org/10.1016/j.jweia.2015.06.003>, URL <http://www.sciencedirect.com/science/article/pii/S0167610515001415>.
- [19] Kambhatla, N., and Leen, T. K., “Dimension reduction by local principal component analysis,” *Neural computation*, Vol. 9, No. 7, 1997, pp. 1493–1516.
- [20] Yan, A.-M., Kerschen, G., De Boe, P., and Golinval, J.-C., “Structural damage diagnosis under varying environmental conditions—part II: local PCA for non-linear cases,” *Mechanical Systems and Signal Processing*, Vol. 19, No. 4, 2005, pp. 865–880.
- [21] D’Agostino, D., Serani, A., Campana, E. F., and Diez, M., “Nonlinear Methods for Design-Space Dimensionality Reduction in Shape Optimization,” *3rd International Conference on Machine Learning, Optimization, and Big Data, MOD 2017, Volterra, Italy*, 2017, pp. 121–132.
- [22] Clément, S. A., Guillemain, A., McCleney, A. B., and Bardet, P. M., “Options for refractive index and viscosity matching to study variable density flows,” Vol. 59, No. 32, 2018.
- [23] Bishop, C. M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [24] Golub, G. H., and Reinsch, C., “Singular value decomposition and least squares solutions,” *Numerische mathematik*, Vol. 14, No. 5, 1970, pp. 403–420.
- [25] Cizmas, P. G., Palacios, A., O’Brien, T., and Syamlal, M., “Proper-orthogonal decomposition of spatio-temporal patterns in fluidized beds,” *Chemical Engineering Science*, Vol. 58, No. 19, 2003, pp. 4417 – 4427. doi:[https://doi.org/10.1016/S0009-2509\(03\)00323-3](https://doi.org/10.1016/S0009-2509(03)00323-3), URL <http://www.sciencedirect.com/science/article/pii/S0009250903003233>.
- [26] Drineas, P., Frieze, A., Kannan, R., Vempala, S., and Vinay, V., “Clustering large graphs via the singular value decomposition,” *Machine learning*, Vol. 56, No. 1-3, 2004, pp. 9–33.
- [27] Arthur, D., and Vassilvitskii, S., “k-means++: The advantages of careful seeding,” *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

- [28] Gibson, M., "Spectra of turbulence in a round jet," *Journal of Fluid Mechanics*, Vol. 15, No. 2, 1963, pp. 161–173.
- [29] Benzi, R., Paladin, G., and Vulpiani, A., "Power spectra in two-dimensional turbulence," *Physical Review A*, Vol. 42, No. 6, 1990, p. 3654.
- [30] Ketchen, D. J., and Shook, C. L., "The application of cluster analysis in strategic management research: an analysis and critique," *Strategic management journal*, Vol. 17, No. 6, 1996, pp. 441–458.
- [31] 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.