Flow Measurement and Instrumentation 67 (2019) 41–54

Contents lists available at ScienceDirect

Flow Measurement and Instrumentation

journal homepage: www.elsevier.com/locate/flowmeasinst

Machine learning approach to construct global phase-averaged flow field based on local flow features

Xin Wen[a,b], Ziyan Li[a,b], Jiajun Liu[a,b], Wenwu Zhou[a,b], Yingzheng Liu[a,b,*]

[a] Key Lab of Education Ministry for Power Machinery and Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China

[b] Gas Turbine Research Institute, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China

ARTICLE INFO

Keywords:
Data fusion
K-nearest neighbor
PIV
Phase averaging

ABSTRACT

This paper reports a new approach toward constructing a full-domain phase-averaged flow field; the approach applies particle image velocimetry (PIV) measurement without referring to time-resolved signals. This approach is a departure from the conventional phase-averaging method based on proper orthogonal decomposition (POD). The POD-based method requires the full flow field to be covered during the measurement, which can result in a low spatial resolution. The proposed method combines multiple local flow fields in different subdomains with high spatial resolution to construct a full phase-averaged flow field. The local flow fields are phase-identified using a machine learning approach, namely, k-nearest neighbor (KNN) classification. Prior to the classification, the full flow fields are first measured with low spatial resolution and then phase labeled by POD analysis. Then, the full flow fields are divided into multiple local flow fields. The major flow features of the phase-labeled local flow fields are extracted by POD again. And, the produced POD coefficients serve as training samples. Subsequently, the new local flow fields are measured with high spatial resolution and then phase-identified by comparing to the training samples using KNN classification. Finally, a full-domain flow field is constructed by combining the phase-averaged local flow fields with high spatial resolution. In the application of KNN, the training samples can be better differentiated in a high-dimensional space defined by the POD modes.

1. Introduction

As a commonly used optical measurement technique, particle image velocimetry (PIV) has facilitated convenient quantitative measurements over entire flow fields. For measuring periodic or quasi-periodic flow, phase averaging is an important step in the post-processing of PIV data. By averaging multiple instantaneous flows of the same phase, the dominant dynamics and consequently the statistical information of the flow field can be obtained. However, it is highly challenging to obtain the phase indicators of an instantaneous flow field owing to the low temporal resolutions of a standard PIV system [1]. In this regard, a PIV result can be phase-identified using additional time-resolved reference signals. However, the pointwise reference signal generally exhibits high fluctuations, phase jitter, and interruption to the flow field. Proper orthogonal decomposition (POD), which does not require time-resolved signals, is a popular method based on the major flow features of the global flow field. The phase indicators of instantaneous flow fields can be calculated in a space defined by the first two POD modes. However, it still exhibits a limitation and requires the full flow field to be covered during the PIV measurement for capturing the two dominating POD modes. This requirement generally results in a low spatial resolution of the measurement. Accordingly, a method to obtain a full phase-averaged flow field with high-spatial resolution without requiring time-resolved signals is highly desirable.

Various efforts have been undertaken to obtain time-resolved signals while conducting PIV measurement. For example, in a previous study on the vortex shedding process in bluff-body flow, the time-resolved velocity at a fixed position in the downstream wake was captured by hot wire anemometry to facilitate the phase-averaging process of PIV data [2]. Recently, this method has also been used in the PIV measurement of a self-oscillating sweeping jet [3]. By using pressure sensors to obtain time-resolved reference signals, the phase-averaged flow fields of a sweeping jet were effectively captured by PIV [4]. Although this method can be implemented conveniently, it exhibits several limitations. First, the pointwise reference signal generally exhibits high turbulent fluctuations and measurement errors. Notwithstanding

* Corresponding author. Key Lab of Education Ministry for Power Machinery and Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China.
E-mail address: yzliu@sjtu.edu.cn (Y. Liu).

https://doi.org/10.1016/j.flowmeasinst.2019.04.006
Received 22 October 2018; Received in revised form 10 April 2019; Accepted 14 April 2019
Available online 17 April 2019
0955-5986/

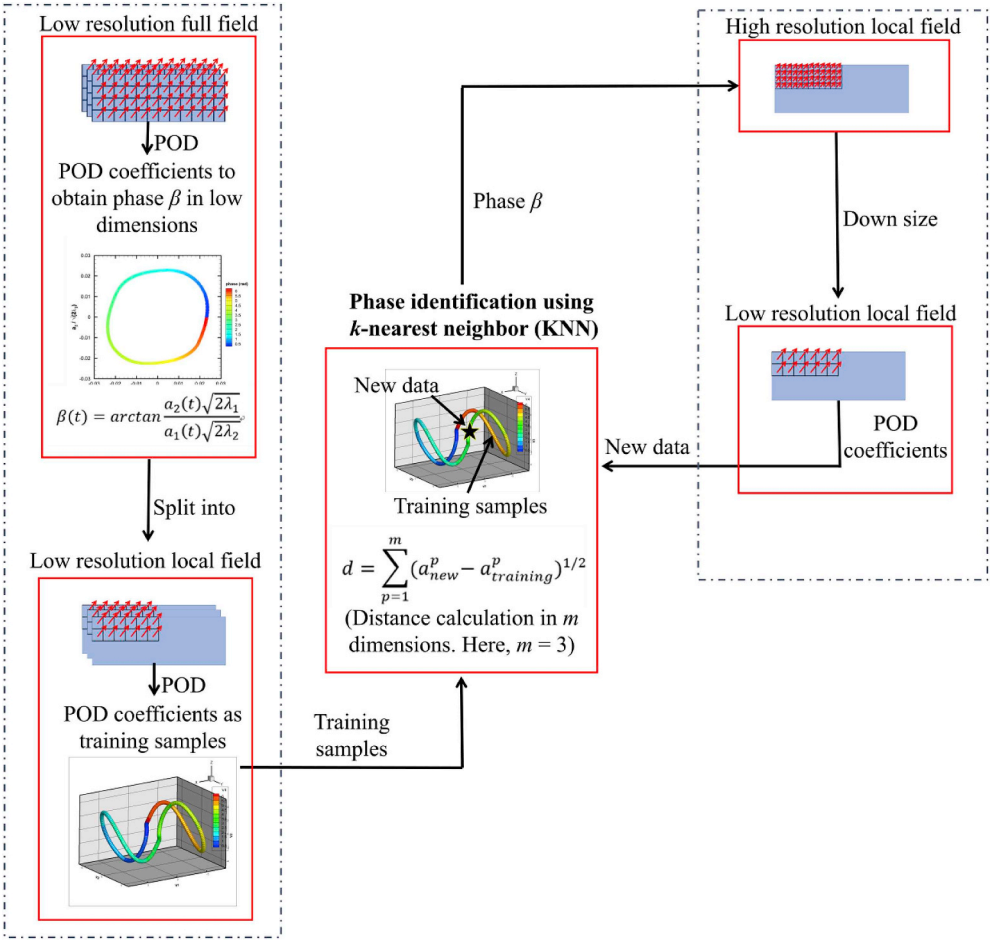


Fig. 1. Outline of phase identification of local flow field (right part) using KNN (middle part) by learning from full-domain flow field (left part).

Table 1
Parameters of fabricated pattern.

n	a	b	β	γ	f
1	0.03	0.05	2	1	0.25
2	0.0075	0.0125	2	1	0.25

the use of low-pass or band-pass filters, the filtered signal occasionally fail to provide suitable phase indicators owing to large fluctuations in the flow fields. Secondly, the reference signal extracted from a local location is required to capture the major dynamics of the periodic flow field. This requires the sensors to be placed in appropriate locations. However, this requirement generally cannot be satisfied owing to the constraints of the experimental setup or in order to prevent the interruption of the flow field. Therefore, it becomes less relevant as the distance from the most suitable location increases, resulting in the problem of phase jitter [5,6].

In order to overcome the above limitations, efforts have been made to accomplish phase identification based on the global flow field without requiring the time-resolved reference signal [7]. A commonly used method is based on POD analysis [8,9]. Using POD decomposition, the dominating flow features can be extracted in the first several POD modes [10–12]. The POD coefficients can be obtained by projecting the instantaneous flow fields on a POD space defined by the modes. Therefore, the temporally-changing coefficients reflect the oscillations of the flow field between the POD modes [6]. observed that for a quasi-periodic flow, the coefficients of the first two modes tend to form a limit cycle in the phase plane and used this feature to extract the vortex shedding phase in a square cylinder wake. Similarly, with regard to capturing the phase-averaged motion of the sweeping jets [3], observed that the POD method yielded results remarkably in agreement with those of the reference signal method. Although this POD method is a simple and accurate tool without requiring time-resolved data, it exhibits limitations. In order to identify the first two dominating POD modes, it is necessary to cover the full flow field within the measurement view. However, this requirement is generally satisfied at the cost of the spatial resolution of the PIV measurement. The resolution deteriorates further when the field of view cannot fit the camera sensor geometry. To obtain a high spatial resolution, the measurement is generally focused to the subdomains to capture the local flow fields. The multiple local flow fields can be phase-averaged and combined together to construct a full flow field. However, without time-resolved reference signals, it is highly challenging to phase-identify the local flow fields.

With the rapid development of data science, machine learning and data fusion techniques are attracting interest in the field of fluid dynamics [13–16]. In the post-processing of experimental data, data fusion has been used to combine different data sets to extract information that cannot be obtained from individual data sets. Recently [17], successfully recovered a clean and spatially resolved pressure field by fusing highly-noisy albeit spatially resolved data obtained from fast pressure sensitive paint and clean albeit scattered data obtained from microphones. While applying PIV measurement [18], constructed a full field of velocity data by fusing two sets of incomplete complementary measurement. Here, two types of PIV data sets can be fused together to obtain a full phase-averaged flow field with high spatial resolution. A data set is obtained by measurement covering the whole velocity field

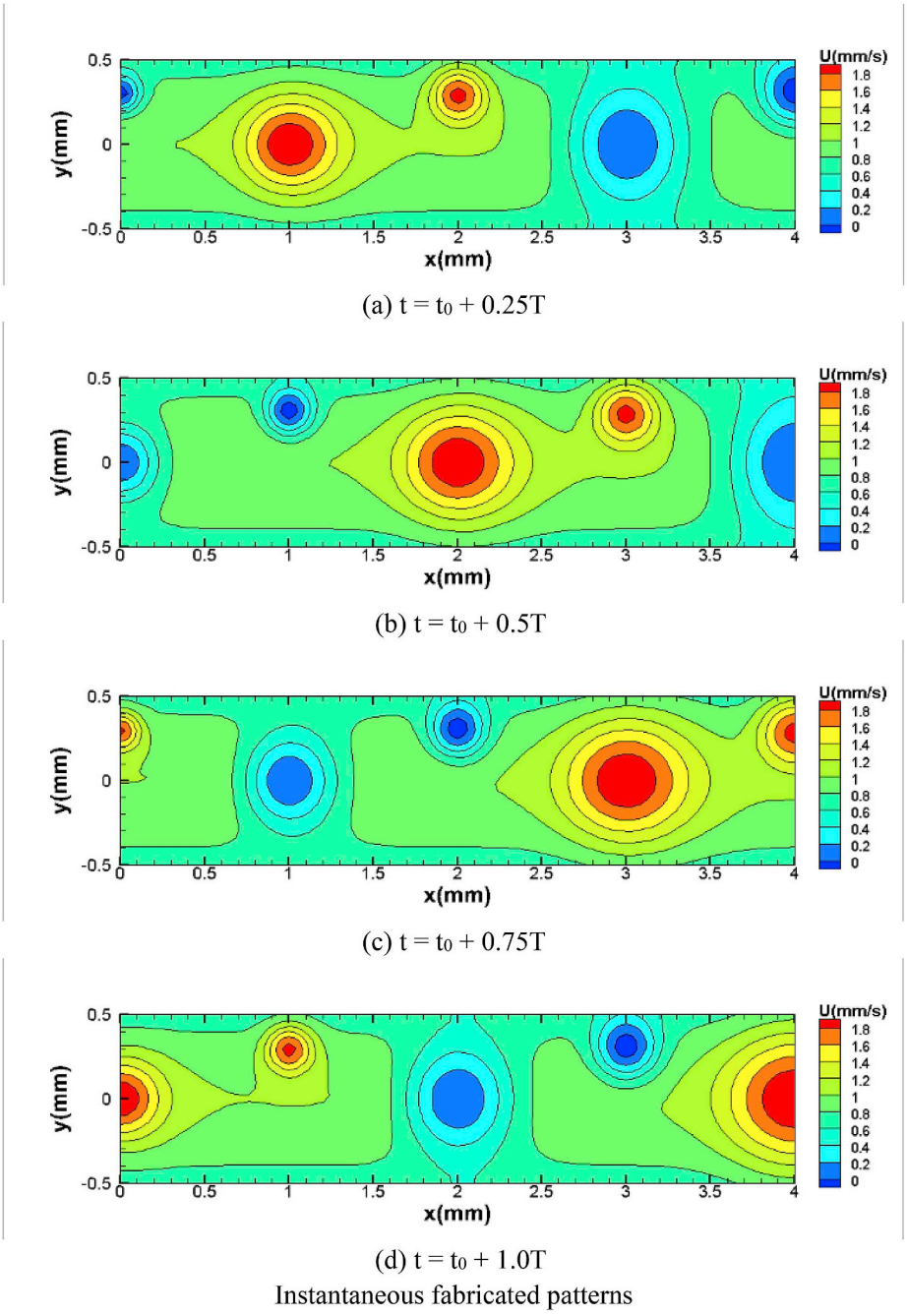


Fig. 2. Instantaneous fabricated patterns.

with a low spatial resolution. Therefore, the instantaneous full flow fields can be phase identified by POD analysis. The full flow fields are then divided into multiple local flow fields that produce training samples in a POD space. The other type of data sets is obtained by measuring multiple local flow fields in different sub domains with a high spatial resolution. The local flow field can be phase identified using a machine learning approach based on the training samples. Finally, the high-resolution local flow fields can be phase-averaged and combined together to construct a full flow field.

In this study, k-nearest neighbor (KNN) classification was used to phase label the local flow fields. Both synthetic convecting patterns and PIV data of a sweeping impinging jet were used to demonstrate the capability of this method.

2. Methodology

Fig. 1 presents the overall procedures of this KNN-based method. The procedure generally consists of two steps. In the first step, the full flow fields are measured with low-spatial resolution (left top of the figure). The POD analysis is then applied to obtain the coefficients of

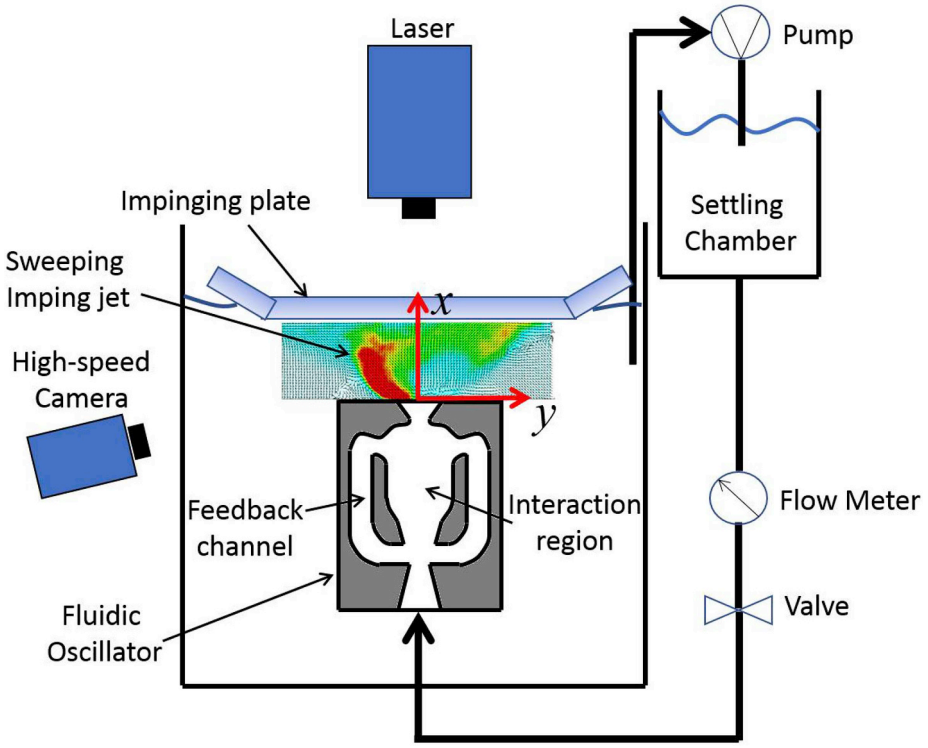


Fig. 3. Sketch of PIV measurement of sweeping impinging jet (not to scale).

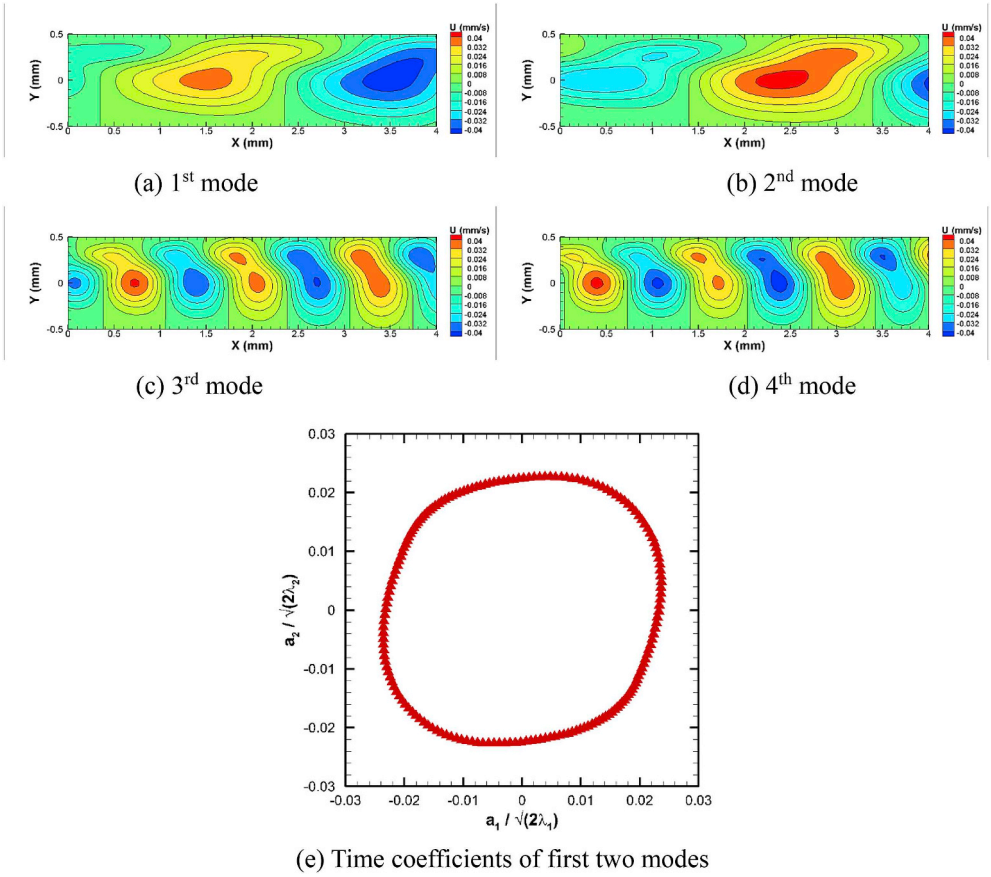
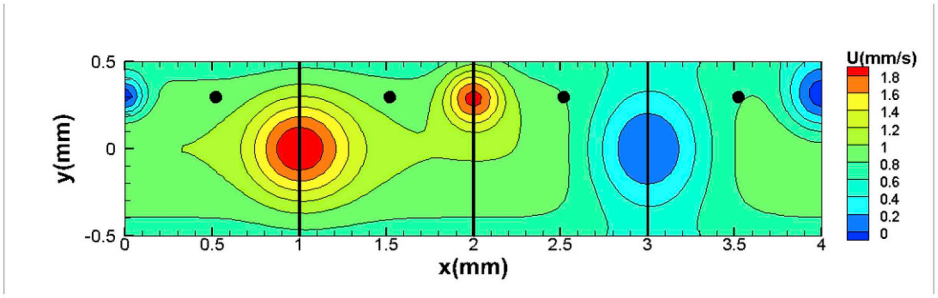
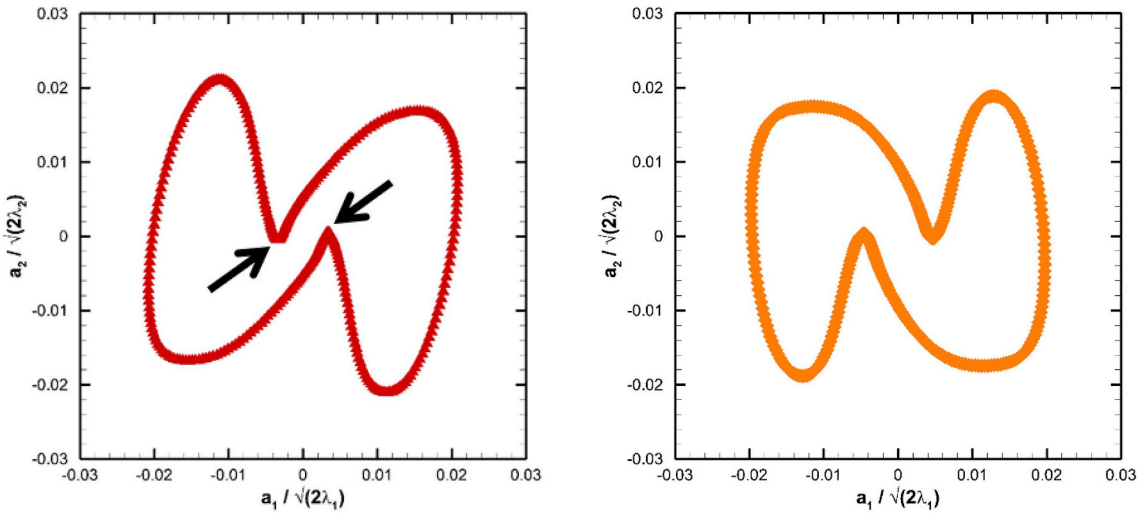


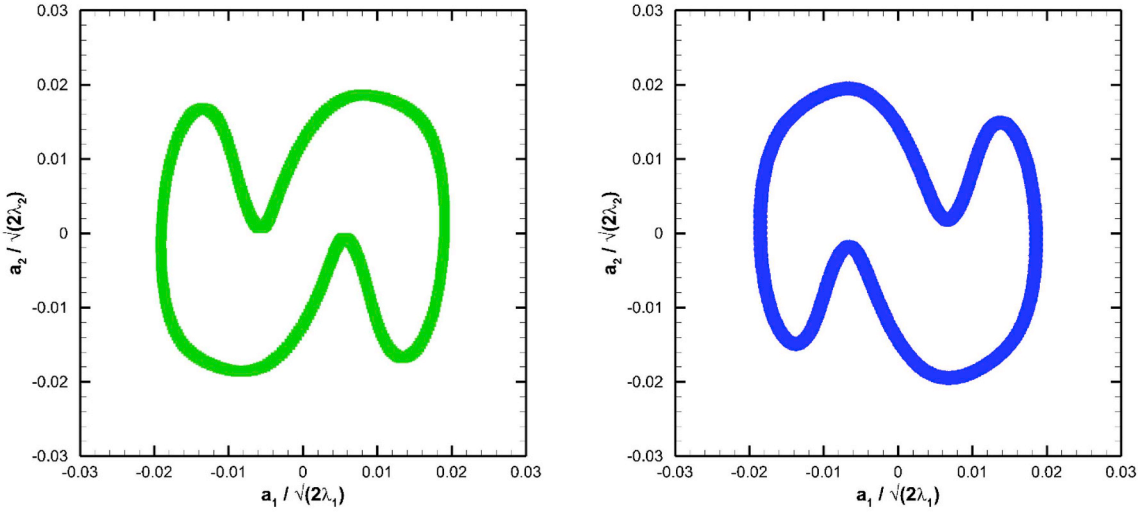
Fig. 4. POD analysis of full fields of fabricated patterns.



(a) Division of subdomains #1–4 from left to right separated by solid lines.



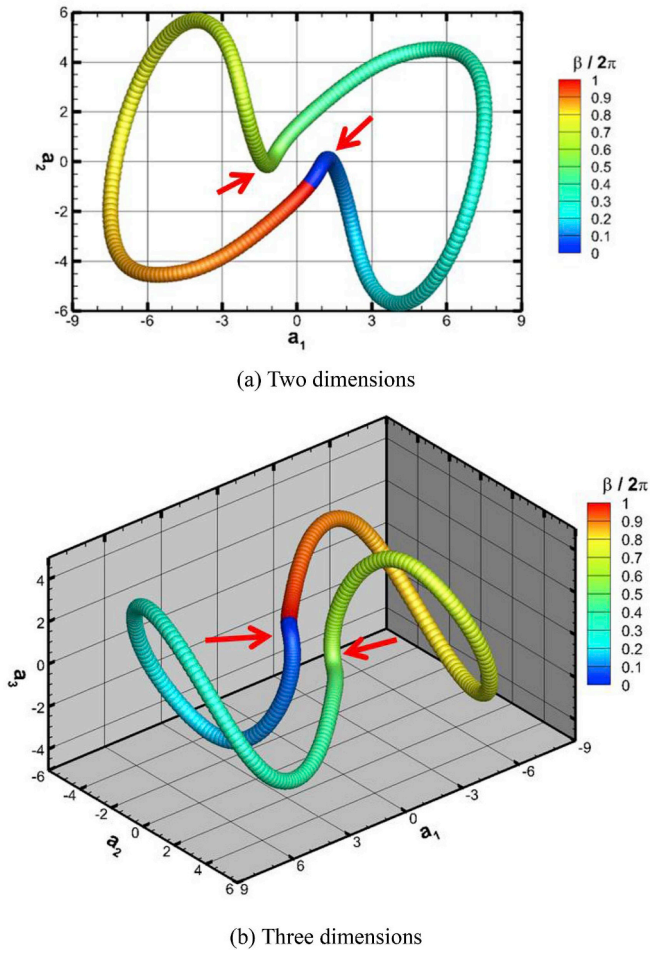
(b) Subdomain #1 (c) Subdomain #2



(d) Subdomain #3 (e) Subdomain #4

Fig. 5. POD analysis of local flow fields in subdomains.

the first two POD modes. The coefficients form a limit cycle in the phase plane and are used to calculate the phase identicators β. The phase-labeled full flow fields are then split into multiple local flow fields (left bottom of the figure). Similarly, the local flow fields are also analyzed by POD. In the local flow fields, only the first two POD modes cannot effectively capture the major flow features. Accordingly, the POD coefficients no longer form a limit cycle and cannot be used directly to identify the phase. Meanwhile, the phase labels of the local flow fields are transferred from the full flow fields. Subsequently, the POD coefficients of the phase-labeled local flow fields are used as training



(a) Two dimensions

(b) Three dimensions

Fig. 6. Training samples in different POD space. The scatters of the training samples are colored by the phase angle. The training samples are more conveniently differentiated in three dimensions than in two dimensions.

samples in the KNN. To better differentiate the training samples, the local flow fields are projected into a space with higher dimensions defined by the POD modes. In the other step, multiple measurements are conducted only in the local flow fields with a high-spatial resolution. The high-resolution PIV measurement can be temporally downsized to a low resolution similar to those in the first step. Then, the low-resolution local flow fields are projected on the same POD space. The obtained POD coefficients are used as new data. Based on the distance to the new data in the POD space, the closest k training samples are selected to identify the phase of the new data in KNN. After the phase identification, the phase labels β are transferred to the high-resolution local flow fields. This procedure is repeated in the other subdomains. The local flow fields with high-spatial resolution are then phase-averaged. Finally, the local flow fields in multiple subdomains are combined together under the same phase angle to construct a full flow field.

2.1. POD analysis

The POD analysis method [19] is a commonly-used tool to extract major flow features from unsteady flow fields [20,21]. In the present study, the "snapshot" POD method [22] is used. In this POD, an instantaneous flow field U(t), obtained from a data set $\boldsymbol{U}$, can be decomposed into the following form:

$$U(t) = \bar{\boldsymbol{U}} + \sum_{i=1}^{n} \boldsymbol{\Psi}_i a_i, \tag{1}$$

where $\bar{\boldsymbol{U}}$ is the long-time mean flow field, $\boldsymbol{\Psi}$ represents the spatial POD modes, $\boldsymbol{a}$ represents the corresponding POD time coefficients, and n represents the number of snapshots used in the calculation. In the POD analysis, $\boldsymbol{\Psi}$ and $\boldsymbol{a}$ can be calculated using single value decomposition as follows:

$$(\boldsymbol{\Psi}, \sqrt{\lambda}\ , \boldsymbol{a}) = \mathrm{svd}(\boldsymbol{U}), \tag{2}$$

The eigenvalue $\boldsymbol{\lambda}$ is the energy captured by the POD mode $\boldsymbol{\Psi}$. The POD modes are then sorted in the descending order of energy content. For a periodic or quasi-periodic flow, when the major flow dynamics are captured by the first two POD modes, the periodic flow behaviors can be considered as a harmonica oscillation between the first two modes [3]. Accordingly, the first two POD coefficients share the sinusoidal fluctuation, albeit with a phase difference of approximately $\pi/2$. Based on the first two coefficients, the phase angle of each instantaneous flow field can be calculated as follows:

$$\beta(t) = arctan\frac{a_2(t)\sqrt{2\lambda_1}}{a_1(t)\sqrt{2\lambda_2}} \tag{3}$$

The phase angle can be obtained without the time-resolved reference signal. However, the prior assumption of the POD-based method is that the major flow features can be captured by the first two POD modes. The mathematical background of this method is discussed in detail by Ref. [23].

2.2. KNN classification

KNN classification is the key component of this method for fusing the full flow field with low spatial resolution and local flow field with high resolution. Compared to other commonly used methods in the machine learning toolbox, KNN has the notable advantages of very rapid training and learning of complex functions. In a KNN for classifying a new input data, the k closest training samples are examined and the new data are classified based on the class of the selected training samples. To calculate the distance between the new data and the training samples, Euclidean distance is used in the current KNN as follows:

$$d = \sum_{p=1}^{m} (a_{new}^{p} - a_{training}^{p})^{1/2} \tag{4}$$

where m is the number of POD modes used (or the dimensions of the POD space); a_{new}^{p} is the coefficient of the new local flow field, which is projected on the pth POD mode. Similarly, $a_{training}^{p}$ is the coefficient of the training data. Note that because the used POD modes capture different energy as indicated by the different values of $\boldsymbol{\lambda}$, a^p accordingly exhibits distinct standard deviations along different dimensions. For example, the first POD mode exhibits the highest energy. Accordingly, a^1 exhibits the highest standard deviation. These dimensions can be standardized to equalize their importance. However, they are not standardized in the proposed method because the different standard deviations reflect the weight factors of the different dimensions. Similarly, the k closest training samples are also weighted. The weight factor is inversely proportional to the distance between the new data and the training sample. Finally, the phase angle of the new data can be determined by the k closest training samples using the weighted distance voting as

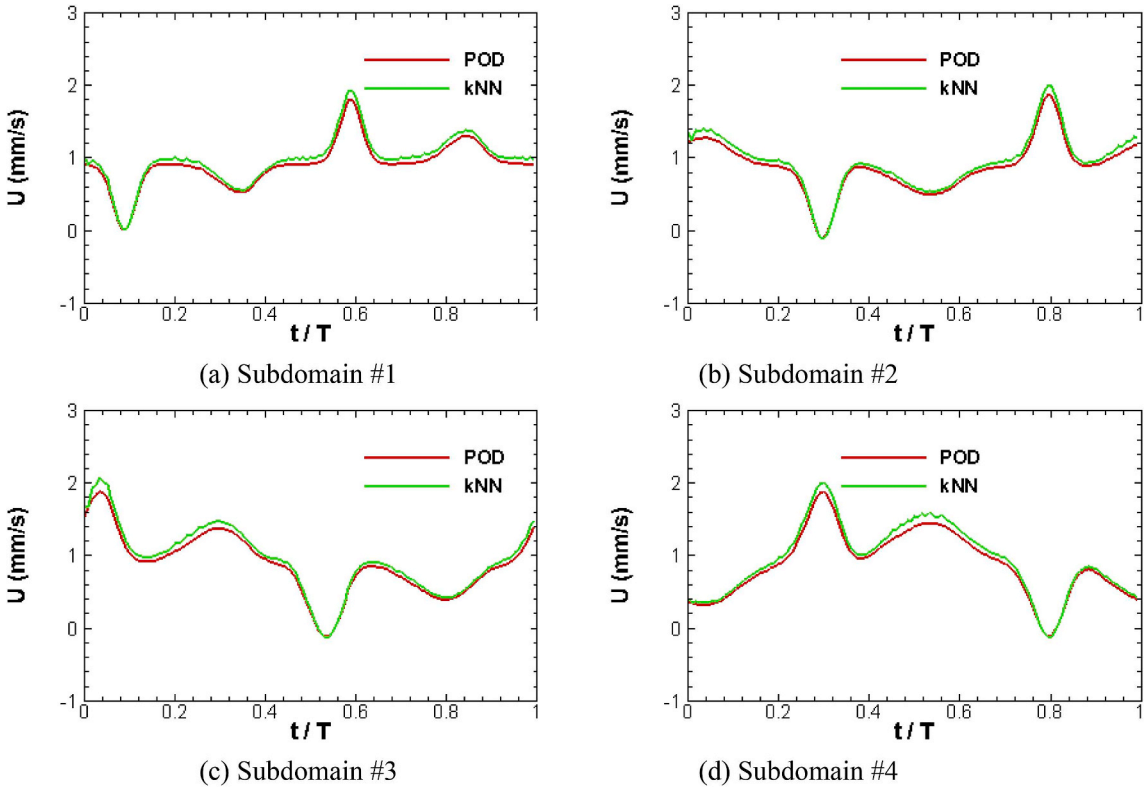


Fig. 7. Comparison of phase-resolved velocity based on KNN phase-identification of local flow fields and POD analysis of full flow fields. The velocities are extracted at the four subdomains indicated by the black dots in Fig. 5a.

$$\beta_{new} = \frac{\sum_{l=1}^{k} \frac{1}{d_l^2}\beta_p}{\sum_{l=1}^{k} \frac{1}{d_l^2}} \tag{5}$$

where β_{new} is the determined phase of the new data, d_l^2 is the square of the distance from the new data to the lth training sample, and β_l is the phase of the lth training sample. A fundamental requirement for KNN is a large number of training samples for achieving classifications of fine resolutions. This requirement is conveniently satisfied in the experiments in fluids. Compared to the time-consuming experimental setup and data post-processing, it is highly convenient to conduct multiple runs of experiment to collect adequate training samples.

3. Setup of simulation and PIV measurements

As proof-of-concept, this KNN-based approach is first applied to fabricated data. Subsequently, it is applied to the PIV measurement of a sweeping impinging jet. This section describes the setup of the simulation and the measurement.

3.1. Fabricated patterns

A series of fabricated patterns are constructed to simulate convecting patterns comprising multiple structures. Here, the method proposed by Ref. [24] is applied to generate integrated patterns as follows:

$$F = \sum_{j=0}^{j=2} q_l \tag{6}$$

$$q_0 = \exp(-y^2/0.7) \tag{7}$$

$$q_1 = \sum_{m=-\infty}^{m=\infty} (-1)^m \exp\left[-\left(\frac{(x-\beta_n m-\gamma_n t)^2}{d_n}\right) - \frac{y^2}{d_n}\right],\ n=1 \tag{8}$$

$$q_2 = \sum_{m=-\infty}^{m=\infty} (-1)^m \exp\left[-\left(\frac{(x-\beta_n m-\gamma_n t)^2}{d_n}\right) - \frac{(y-0.3)^2}{d_n}\right],\ n=2 \tag{9}$$

where x and y are the streamwise and spanwise dimensions, respectively. The diameters of the convecting structures are defined by $d = ax + b$, where a and b are constants. The dominating convecting structures in q_1 are designed to have a large diameter, whereas q_2 has a significantly smaller one. The integrated convecting patterns have multiple major dimensions in POD space, which will be illustrated later. The distance between the structures and the convection velocity are determined by β and γ, respectively. The two structures have identical frequency determined by $f = \gamma/(2\beta)$. The parameters of the fabricated patterns are listed in Table 1. Fig. 2 shows the evolution of the integrated fabricated patterns.

3.2. Experimental setup

To test this approach using real-world data, PIV measurements are performed on the unsteady flow fields of a sweeping imping jet in a water tank, as shown in Fig. 3. The sweeping jet is issued from a fluidic oscillator, which has an interaction region and two lateral feedback channels. Inside the fluidic oscillator, the main jet is attracted to a sidewall in the interaction region owing to the Coanda effect. A small

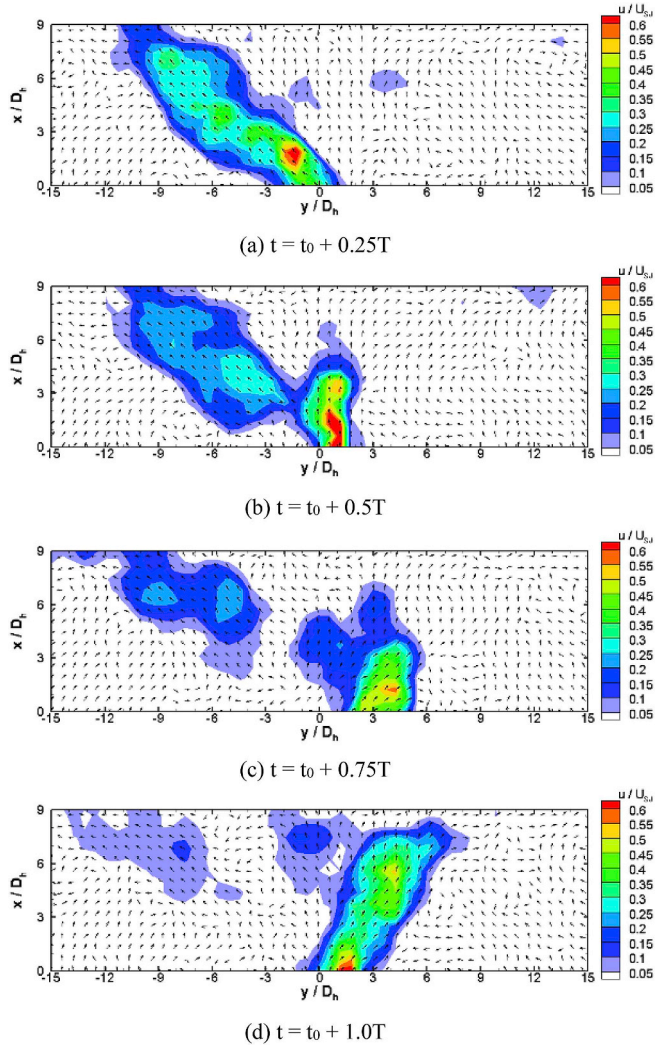


Fig. 8. Instantaneous flow fields of sweeping impinging jets.

portion of the main jet stream returns to the oscillator's inlet through the feedback channel and pushes the main jet to the opposite side. As a result, the jet issuing from the oscillator exhibits a periodic sweeping motion in the transverse direction. Owing to the self-sustained oscillation motion without requiring any moving part, sweeping jets are attracting interest in the application of impingement [25–28]. Owing to the self-sustained oscillation, it is also significantly challenging to obtain the phase-averaged flow fields without a time-resolved reference signal. Although the POD method can be used, it requires the whole flow field to be covered, which can result in a low spatial resolution. Therefore, the proposed approach is tested to obtain full phase-averaged flow fields of the sweeping impinging jet with high spatial resolution.

The experiment is performed in a water tank, as shown in Fig. 3. The fluidic oscillator has a square exit with a hydraulic diameter of $D_h = 10$ mm and is placed at the center of the tank. Water is channeled from an overhead settling chamber and impelled by gravity into the oscillator. The flow rate is regulated by a flowmeter monitor. The origin of the coordinate system is set at the center of the jet nozzle, with the x- and y-axes pointing in the jet's axial and transverse (corresponding to the jet's sweeping motion) directions. To facilitate the PIV measurements, glass beads ($\rho \approx 1050\,kg/m^3$, $d \approx 10\,\mu m$) are used as tracer particles in the entire water tunnel. The tracking particles are illuminated using an 8-W continuous-wave semiconductor 532-nm laser (MGL-N-532a-5w, CNI). A high-speed camera (dimax HS4, pco.) is applied to capture images in the area of interest, which spanned $y = \pm 15D_h$ and $x = 0–8D_h$ with an aspect ratio of approximately 3.8:1. Then, a multigrid cross-correlation technique in combination with subpixel recognition by Gaussian fitting is applied on the particle images. With a final interrogation window size of 32 × 32 pixels and 50% overlap, the PIV measurement yields a resolution of approximately $0.3D_h$. Details of the PIV measurement are available in the authors' previous reports [27]. The obtained 14,400 full flow fields with high spatial resolution are then used as the data pool to examine the KNN phase identification. Half of the flow fields are down-sized to a low resolution using an interrogation window of size 64 × 64 pixels. Then, they are divided into the local flow fields, yielding the training samples. Similarly, the remaining flow fields are used to generate the new data to be phase labeled.

4. Results

4.1. Fabricated patterns

Before addressing real PIV data, the KNN phase identification method is applied on the fabricated patterns. Note that the key step of KNN classification is focused upon here. The process of down-sizing high resolution flow fields to low resolution is excluded. Therefore, all the available data is treated as low-resolution data. As the first step in this method, POD analysis is applied on the full-field data. As shown in Fig. 4a and b, the first two modes capture similar flow features with a phase shift of approximately 1/4 of the wavelength. Therefore, the normalized coefficients of the first two modes form a limit cycle as shown in Fig. 4e. These coefficients are used to identify the phase angles of the full field based on Eq. (3). Here, it is noteworthy that the third and fourth POD modes also capture notable flow features, as shown in Fig. 4c and d; these can be used later in the phase identification of the local flow field.

The full flow fields are then divided into four subdomains as shown in Fig. 5a. Subsequent to the POD analysis on the local flow fields, the coefficients of the first two POD modes cease forming a cycle and exhibited distorted shapes, as shown in Fig. 5b–e. This is because the first two POD modes cannot effectively capture the major flow features of the local flow fields. For example, in the first subdomain, both the coefficients can have values approximately equal to zero as indicated by the black arrows in Fig. 5b. This occurs when only very weak flow structures are captured by the first two POD modes. Owing to the distorted shape, the first two coefficients cannot be used in conjunction with Eq. (3), to identify the phase of the local flow field. Rather, the phase labels of the full flow fields are transferred to the POD coefficients of the local flow fields, which serve as training samples in the subsequent KNN classification. In order to achieve a fine resolution of the phase classification, the size of the training samples need to be adequately large. However, the large number of training samples can be closely gathered and overlap with each other. This hinders the differentiation of the samples. We consider subdomain #1 as an example; as shown in Fig. 6a, each training sample consists of only two POD coefficients. The problem of overlapping is apparent as indicated by the red arrows in the two-dimensional space defined by the first two POD modes. This overlapping problem can be solved using a POD space with higher dimensions. As shown in Fig. 6b, each training sample consists of three POD coefficients. Therefore, the samples can be better differentiated in the POD space defined by the first three modes. Note that the POD space can have higher dimensions using more POD modes.

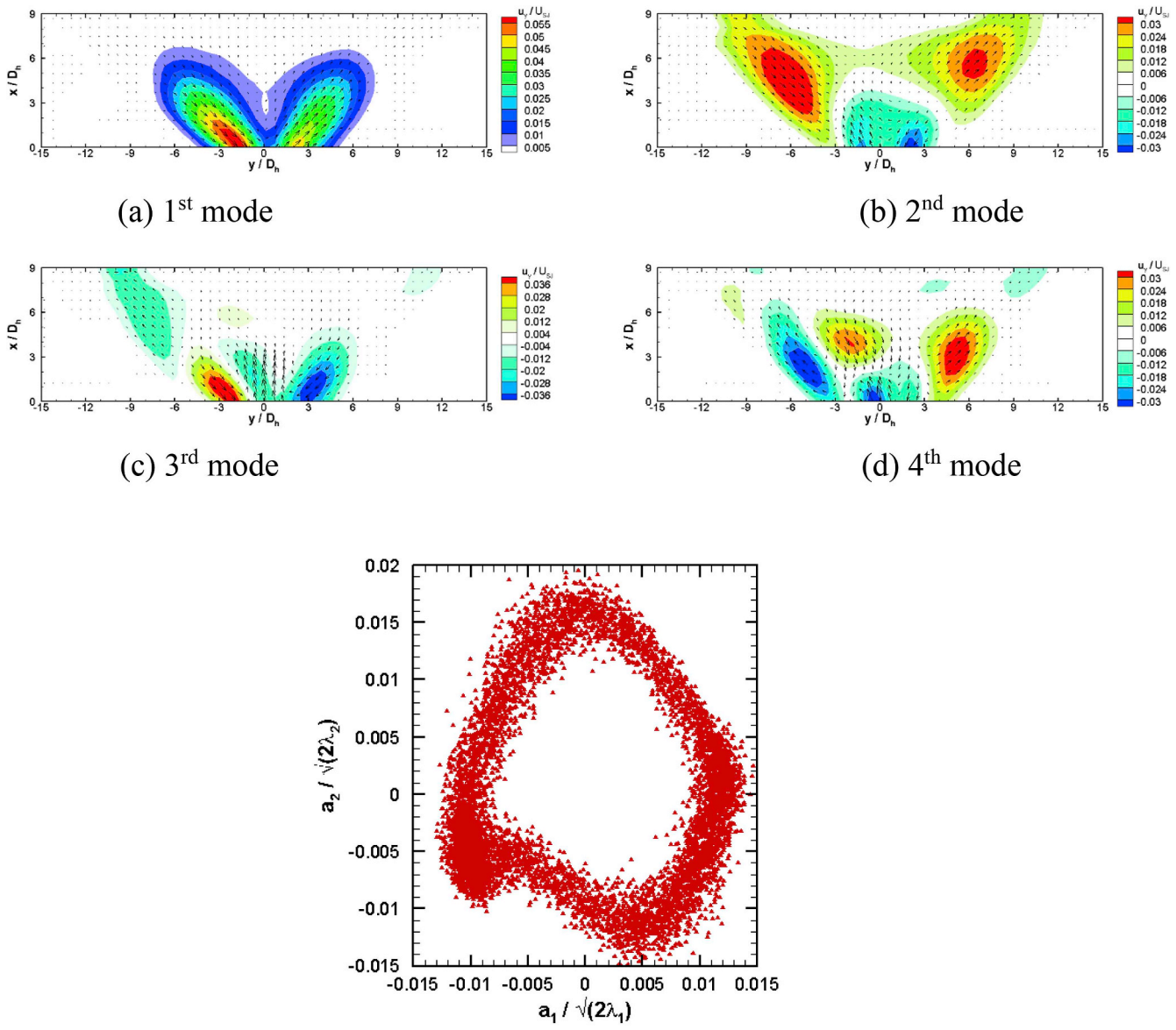


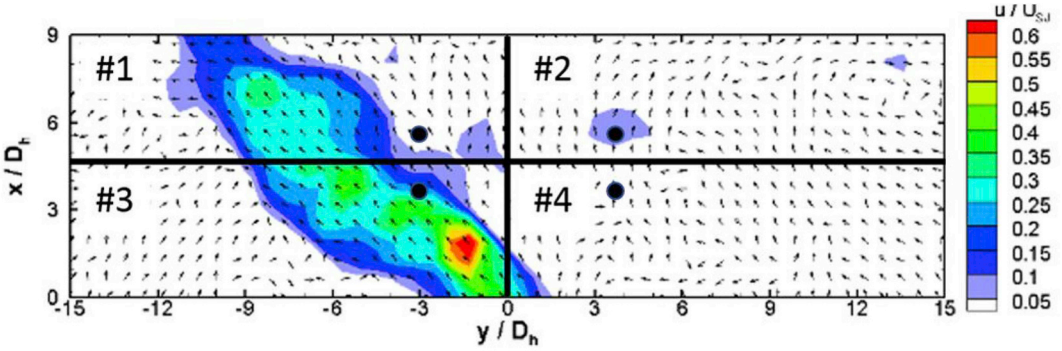
Fig. 9. POD analysis of full flow fields of sweeping impinging jet.

However, the higher dimensions cannot be effectively visualized.

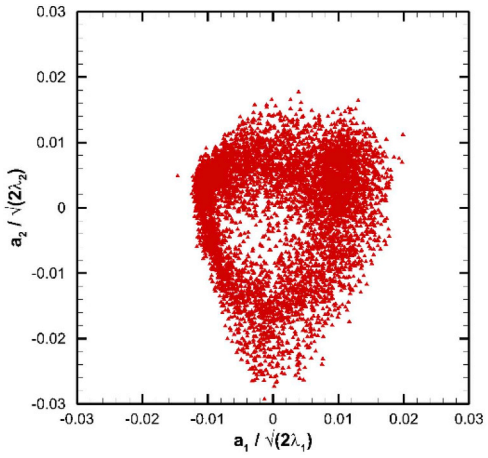
Then, the new local flow fields in the four subdomains are phase labeled using a KNN based on Eq. (5). In the KNN, 3600 instantaneous local flow fields with phase labels are used to produce training samples to achieve a fine resolution of phase identification (up to $2\pi/120$). There are 60 training samples under each phase interval. The training samples have 10 dimensions defined by the first ten POD modes. Therefore, each training sample consists of the first ten coefficients. Weighted by the distance reverse-proportionally, the five closest training samples are used to classify the new data. To evaluate the performance of the KNN classification, the new local flow fields are phase-averaged and compared with the phase-averaged velocity obtained based on the full-field POD method. The phase-resolved velocities are extracted from the four subdomains as shown by the black dots in Fig. 5a. As shown in Fig. 7, the phase-averaged velocity obtained from the KNN is highly consistent with that obtained from the POD analysis, indicating the good performance of this KNN classification method.

4.2. PIV data

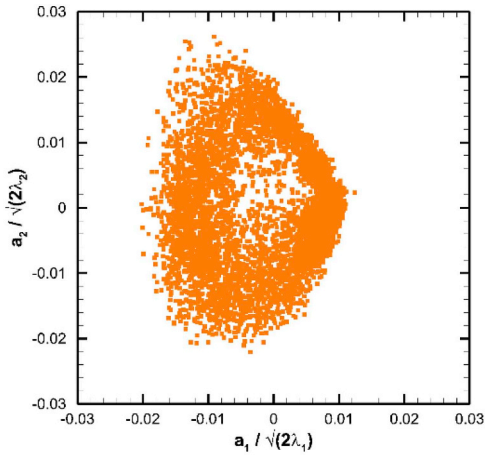
Subsequently, the KNN classification method is applied to the PIV data of sweeping impinging jets. In the first step of the KNN method, the flow fields are downsized to yield low-resolution full fields, as shown in Fig. 8. The sweeping jet evidently exhibits a periodic flow behavior albeit with high fluctuations. Then, POD analysis is applied on the full flow field. As shown in Fig. 9a and b, the first two POD modes capture the major flow features in the near exit region (in mode 1) and in the



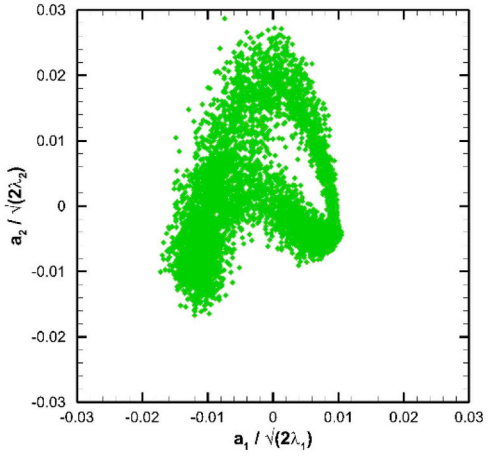
(a) Division of sub-domains #1–4 separated by solid lines.



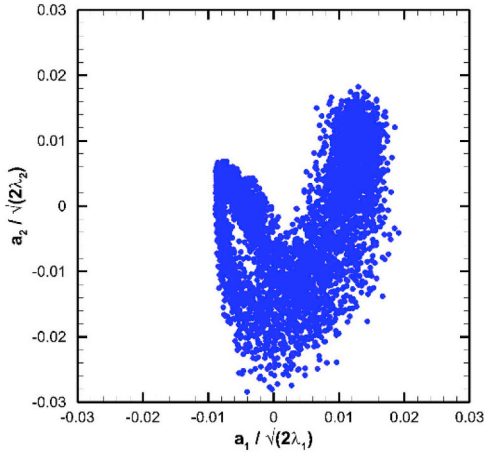
(b) Subdomain #1



(c) Subdomain #2



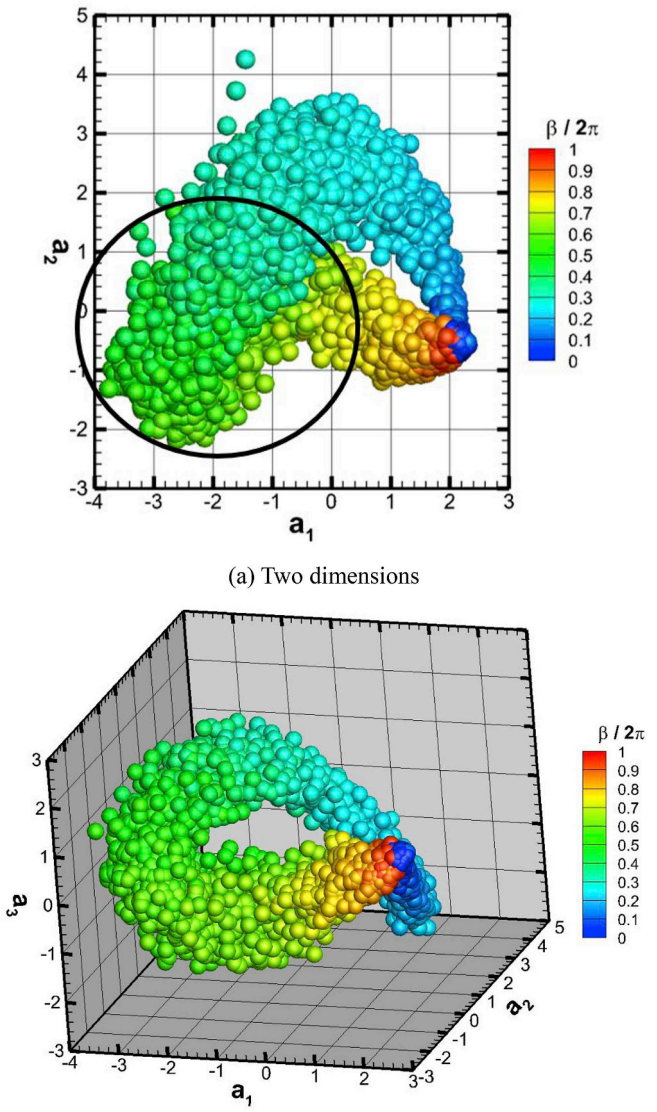
(d) Subdomain #3



(e) Subdomain #4

Fig. 10. POD analysis of local flow fields of sweeping impinging jets.

near wall region (in mode 2), respectively. The coefficients of the first two POD modes form a limited cycle, as shown in Fig. 9e. Note that this cycle is not as uniform as that obtained using fabricated data, due to the distortion of the periodic sweeping motion of the jet. Nevertheless, it still can be used to phase identify the instantaneous full flow field. Note that the third and fourth POD modes also capture notable flow



(a) Two dimensions

(b) Three dimensions

Fig. 11. Training samples in different POD space. The scatters of the training samples are colored by the phase angle. The overlapping problem in two dimensions can be mitigated in three dimensions.

structures, which can be used to define a higher order POD space subsequently. The phase-identified full flow fields are then divided into four local flow fields as shown in Fig. 10a. By conducting POD analysis on each local flow field, the first two POD coefficients cease forming a circle and exhibited distorted shapes, as shown in Fig. 10b–e. The phase of the full flow field is then transferred to the local flow fields. Then, the POD analysis is applied on the phase labeled local flow fields, producing coefficients as the training samples. Owing to the large number of training samples (7200 instantaneous local flow fields), the overlapping problem is more severe than that in the fabricated data. As shown in Fig. 11a, the training samples are colored by the phase angles. The training samples of green color overlap with those of yellow color in the two-dimensional space defined by the first two POD modes, as indicated by the black cycle. Similar to that for the fabricated data, this overlapping problem can be solved using a higher-order POD space. As shown in Fig. 11b, by projecting the local flow fields on the first three POD modes, each training sample consists of three POD coefficients. The samples can be differentiated significantly more effectively, as illustrated by the stereo view.

In the second step, the high-resolution local flow fields are downsized to resolutions similar to those in the first step. Then, the low-resolution local flow fields are phase-identified based on the KNN classification. After a convergence test, it is found that the POD mode number and value of k have converged effect after 10 modes and $k = 3$. Therefore, in the KNN, a POD space defined by 20 modes and $k = 5$ closet training samples are used. In the KNN, a POD space defined by 20 modes and $k = 10$ closet training samples are used. After the phase identification, the low-resolution local flow fields are phase-averaged. Then, the phase-averaged results are compared with those obtained by the full-field-POD-based method. The phase-resolved velocities are extracted in the four subdomains, as indicated by black dots in Fig. 10a. As shown in Fig. 12, the results are reasonably consistent in all the subdomains. Then, the difference between the phase-resolved velocities obtained using the KNN method and POD method at the selected positions is used to examine the influence of the two key parameters, i.e., the size of the training samples and the dimensions of the POD space. The mean squared error (MSE), which is commonly used to examine the accuracy of a machine learning model, is averaged during a cycle as illustrated below:

$$MSE = \frac{1}{120}\sum_{t=1}^{120}(\bar{u}_t^{KNN} - \bar{u}_t^{POD})^2 \tag{10}$$

where $\bar{u}_t^{KNN}$ is the phase-averaged velocity obtained using KNN, whereas $\bar{u}_t^{POD}$ using full-field-POD method. As shown in Fig. 13a, the error deceases with the increase in the size of the training samples. This is consistent with the fact that a large amount of data is required to resolve a fine classification in KNN. The impact of the dimension of the POD space is shown in Fig. 13b. The error generally decreases with an increase in the space dimensions, particularly at the beginning from two dimensions to three dimensions. This verifies the previous observation that the training samples can be more effectively differentiated in a POD space with a higher dimension.

Then, the new local flow fields with high-resolution are phase-averaged and combined together to construct a full flow field. The constructed flow fields are compared with those obtained based on the POD analysis of the full flow field in the data pool. As shown in Fig. 14, the constructed flow fields agree very well with the POD-obtained results in most of the selected phases; an exception is at t/T = 0.25. As shown in Fig. 14a, there is an evident jagged edge of the jet column around $y = 0$ and $x = 2D_h$. A comparison of the POD-obtained flow field illustrates this to be a result of the improper phase-alignment of the local flow field in subdomain #4 (right below part of the flow field) with the other flow fields. The physics behind this ill-classification is that the flow features are excessively weak in this subdomain in this phase. Therefore, it is highly challenging to accurately identify a very weak flow with limited information. To overcome this problem, the areas of the subdomains are enlarged in order to capture adequate flow features. As shown in Fig. 15a, the boundaries are extended by $3D_h$ in y direction for all the four subdomains. The constructed flow fields are significantly improved, which agrees very well with the POD-obtained flow field at t/T = 0.25.

5. Conclusion

In this study, a new algorithm based on KNN classification was developed to reconstruct a full phase-averaged flow field with high spatial resolution, without requiring time-resolved reference signals. In this method, two datasets are fused together: one covers the global flow

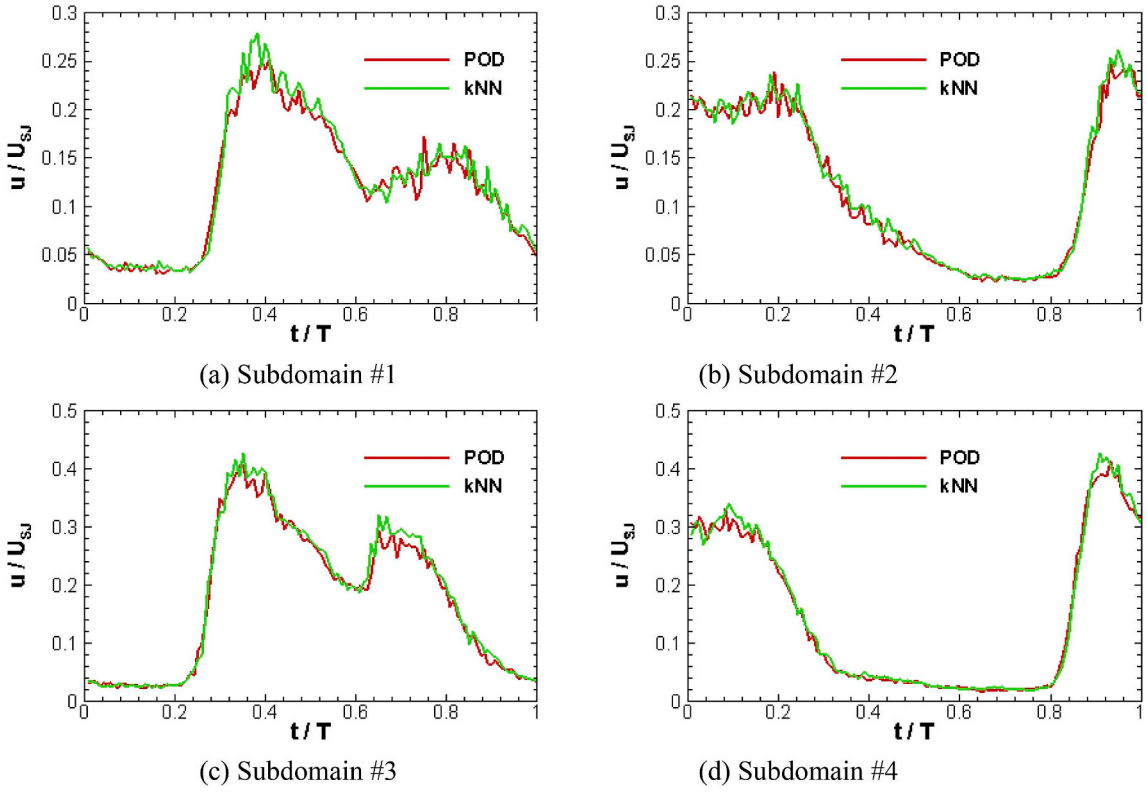


Fig. 12. Comparison of phase-resolved velocity based on KNN phase-identification of local flow fields and POD analysis of full flow fields. The velocities are extracted at the four subdomains indicated by the black dots in Fig. 10a.

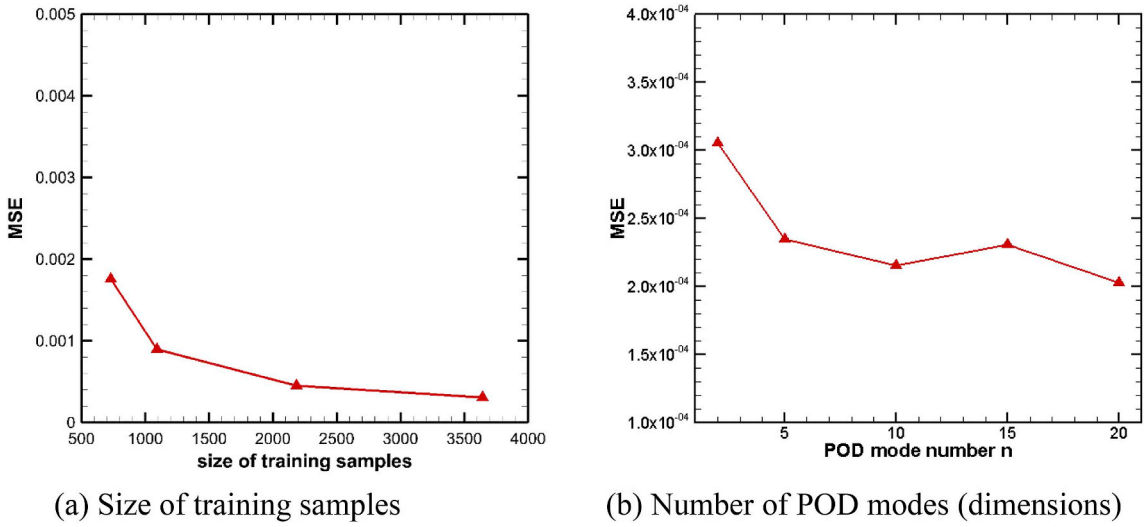


Fig. 13. Effect of key parameters in KNN phase-identification.

field albeit with a low spatial resolution, whereas the other covers multiple local flow fields in different subdomains with a high spatial resolution. The global flow fields can be phase-identified by POD analysis to produce the training data in a POD space. Then, the local flow fields are phase-identified by referring to the training data. Finally, the local flow fields in the different subdomains are phase-averaged and combined to construct a full phase-averaged flow field with a high spatial resolution.

Synthetic data and the PIV data of a sweeping impinging jet were used to demonstrate this approach. The application of KNN

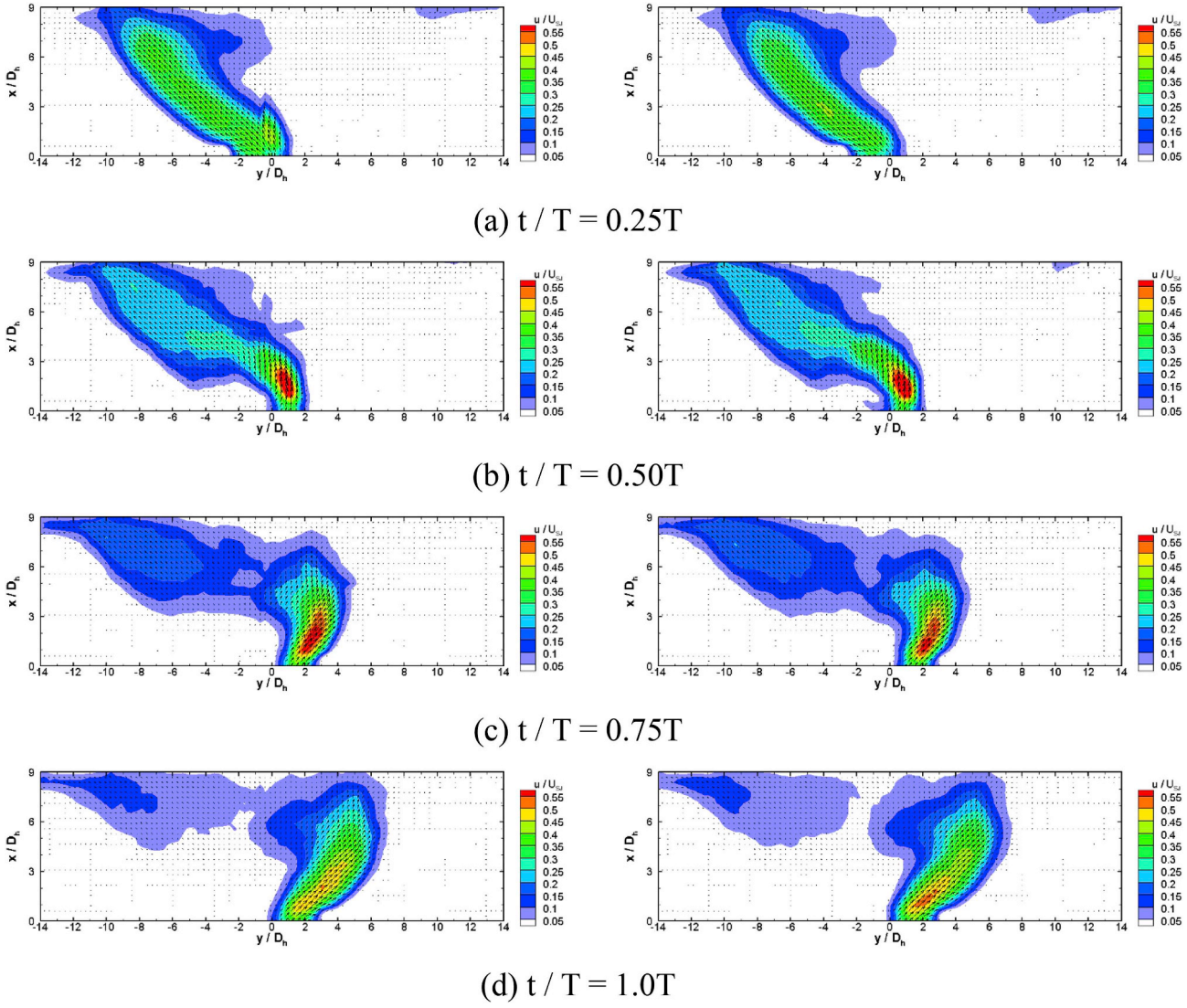
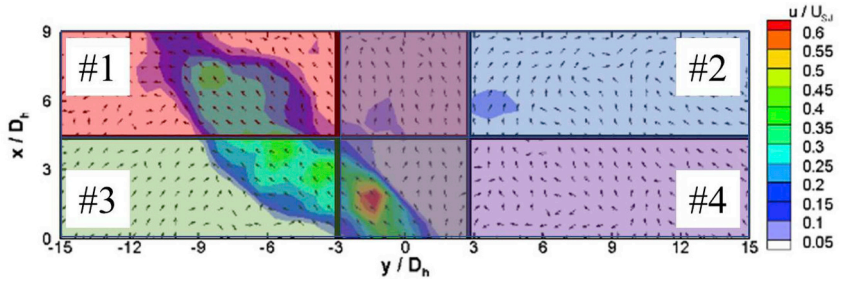


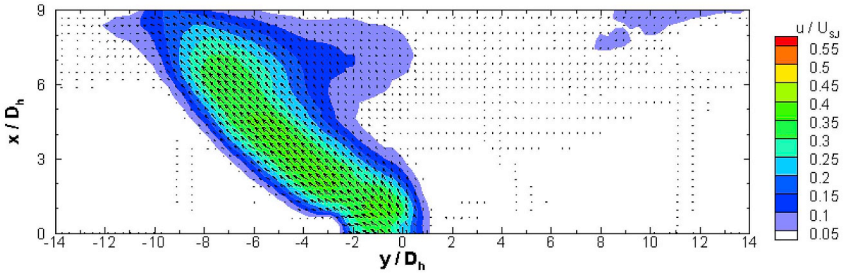
Fig. 14. Comparison of phase-averaged flow fields reconstructed based on KNN on local flow fields (left) and based on POD analysis of full flow fields (right).

classification requires a larger amount of training data to achieve a fine resolution of the phase angle (here, up to $2\pi/120$). The overlapping problem of the training samples is solved by using a high-dimensional space defined by the POD modes. By selecting appropriate parameters, i.e., the size of the training data and the dimensions of the POD space, the mean squared error can be reduced to below 0.001 by comparing the phase-resolved velocity obtained from the KNN classification and that from the traditional POD analysis.

In the present cases of demonstration, the full flow fields are equally divided into four local flow fields for convenience. In real-world applications, the division of the full flow field can be varied and optimized. As illustrated in the demonstration of PIV measurement, the size of the subdomains should be adequately large to capture sufficient local flow features. Therefore, the criterion value of the subdomain size is closely related to the flow behaviors and can change from case to case. However, as demonstrated in this study, the division of the subdomains can be adjusted through trial-and-error by referring to the low-resolution full flow fields. In addition, this method can find wide applications in other fields. For example, it can be used to fuse different numerical model output of ocean circulation together based on the large scale flow features of the ocean circulation.



(a) Division of larger subdomains.



(b) Phase-averaged flow fields by stitching larger local flow fields at t / T = 0.25T

Fig. 15. KNN phase-identification using larger subdomains and reconstruction of full domain flow field.

Acknowledgments

The authors gratefully acknowledge the financial support from the National Natural Science Foundation of China (Grant nos. 11702172 and 11725209).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.flowmeasinst.2019.04.006.

References

[1] C. Pan, H. Wang, J. Wang, Phase identification of quasi-periodic flow measured by particle image velocimetry with a low sampling rate, Meas. Sci. Technol. 24 (5) (2013) 055305.
[2] M. Brede, Measurement of turbulence production in the cylinder separated shear-layer using event-triggered laser-Doppler anemometry, Exp. Fluid 36 (2004) 860–866.
[3] F. Ostermann, R. Woszidlo, S. Gaertlein, C.N. Nayeri, C.O. Paschereit, Phase-averaging methods for the natural flow field of a fluidic oscillator, AIAA J. 53 (8) (2015) 2359–2368.
[4] R. Woszidlo, F. Ostermann, C.N. Nayeri, C.O. Paschereit, The time-resolved natural flow field of a fluidic oscillator, Exp. Fluid 56 (6) (2015) 125.
[5] D.A. Lyn, S. Einav, W. Rodi, J.H. Park, A laser-Doppler velocimetry study of ensemble-averaged characteristics of the turbulent near wake of a square cylinder, J. Fluid Mech. 304 (1995) 285–319.
[6] B.W. van Oudheusden, F. Scarano, N.P. van Hinsberg, D.W. Watt, Phase-resolved characterization of vortex shedding in the near wake of a square-section cylinder at incidence, Exp. Fluid 39 (1) (2005) 86–98.
[7] E. Konstantinidis, S. Balabani, M. Yianneskis, Conditional averaging of PIV plane wake data using a cross-correlation approach, Exp. Fluid 39 (1) (2005) 38–47.
[8] R. Perrin, M. Braza, E. Cid, S. Cazin, A. Barthet, A. Sevrain, C. Mockett, F. Thiele, Obtaining phase averaged turbulence properties in the near wake of a circular cylinder at high Reynolds number using POD, Exp. Fluid 43 (2–3) (2007) 341–355.
[9] C. Xia, H. Wang, D. Bao, Z. Yang, Unsteady flow structures in the wake of a high-speed train, Exp. Therm. Fluid Sci. 98 (2018) 381–396.
[10] A.M. Aliyu, H. Seo, H. Kim, K.C. Kim, Characteristics of bubble-induced liquid flows in a rectangular tank, Exp. Therm. Fluid Sci. 97 (2018) 21–35.
[11] L.H. Feng, J.J. Wang, C. Pan, Proper orthogonal decomposition analysis of vortex dynamics of a circular cylinder under synthetic jet control, Phys. Fluids 23 (1) (2011) 014106.
[12] Y. Xu, J.J. Wang, Digital particle image velocimetry study on parameter influence on the behavior of impinging synthetic jets, Exp. Therm. Fluid Sci. 100 (2019) 11–32.
[13] S. Discetti, M. Raiola, A. Ianiro, Estimation of time-resolved turbulent fields through correlation of non-time resolved field measurements and time-resolved point measurements, Exp. Therm. Fluid Sci. 93 (2018) 119–130.
[14] J.N. Kutz, Deep learning in fluid dynamics, J. Fluid Mech. 81 (2017) 155–166.
[15] J. Ling, K. Andrew, T. Jeremy, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, J. Fluid Mech. 807 (2016) 155–166.
[16] C.J. Ruscher, J.F. Dannenhoffer III, M.N. Glauser, Repairing occluded data for a mach 0.6 jet via data fusion, AIAA 55 (1) (2017) 255–264.
[17] X. Wen, Y. Liu, Z. Li, Y. Chen, D. Peng, Data mining of a clean signal from highly noisy data based on compressed data fusion: a fast-responding pressure-sensitive paint application, Phys. Fluids 30 (2018) 097103.
[18] X. Wen, Z. Li, D. Peng, W. Zhou, Y. Liu, Missing data recovery using data fusion of incomplete complementary data sets: a particle image velocimetry application, Phys. Fluids 31 (2) (2019) 25105.
[19] J.L. Lumley, The structure of inhomogeneous turbulent flow, In Atmospheric Turbulence and Radio Wave Propagation, Nauka, Moscow, 1967, pp. 166–178.
[20] X. Wen, H. Tang, F. Duan, Interaction of in-line twin synthetic jets with a separated flow, Phys. Fluids 28 (4) (2016) 209–216.
[21] Q.S. Zhang, Y.Z. Liu, S.F. Wang, The identification of coherent structures using proper orthogonal decomposition and dynamic mode decomposition, J. Fluids Struct. 49 (2014) 53–72.
[22] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, Journal of the Optical Society of America A-Optics Image Science and Vision 4 (3) (1987) 519–524.
[23] M. Legrand, J. Nogueira, A. Lecuona, Flow temporal reconstruction from non-time-resolved data, Part I: mathematic fundamentals, Exp. Fluid 51 (4) (2011) 1047–1055.
[24] A. Seena, H.J. Sung, Dynamic mode decomposition of turbulent cavity flows for self-sustained oscillations, Int. J. Heat Fluid Flow 32 (2011) 1098–1110.
[25] L. Agricola, M.A. Hossain, R. Prenter, R. Lundgreen, A. Ameri, J. Gregory, J.P. Bons, Impinging Sweeping Jet Heat Transfer, AIAA Paper, 2017, pp. 2017–4974.
[26] T. Park, K. Kara, D. Kim, Flow structure and heat transfer of a sweeping jet impinging on a flat wall, Int. J. Heat Mass Transf. 124 (2018) 920–928.
[27] X. Wen, Y. Liu, H. Tang, Unsteady behavior of a sweeping impinging jet: time-resolved particle image velocimetry measurements, Exp. Therm. Fluid Sci. 96 (2018) 111–127.
[28] X. Wen, Z. Li, W. Zhou, Y. Liu, Interaction of dual sweeping impinging jets at different Reynolds numbers, Phys. Fluids (2018), https://doi.org/10.1063/1.5054161 in press.