# Data mining of a clean signal from highly noisy data based on compressed data fusion: A fast-responding pressure-sensitive paint application

Xin Wen, Yingzheng Liu, Ziyan Li, Yujia Chen, and Di Peng

Citation: Physics of Fluids **30**, 097103 (2018); doi: 10.1063/1.5046681
View online: https://doi.org/10.1063/1.5046681
View Table of Contents: http://aip.scitation.org/toc/phf/30/9
Published by the American Institute of Physics

# Data mining of a clean signal from highly noisy data based on compressed data fusion: A fast-responding pressure-sensitive paint application

Xin Wen,[1,2] Yingzheng Liu,[1,2] Ziyan Li,[1,2] Yujia Chen,[1,2] and Di Peng[1,2,a)]
[1]*Key Lab of Education Ministry for Power Machinery and Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China*
[2]*Gas Turbine Research Institute, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China*

(Received 29 June 2018; accepted 5 September 2018; published online 24 September 2018)

A data mining approach based on compressed data fusion is developed to extract a clean signal from highly noisy data and it has been successfully applied to flow measurement using fast-responding pressure-sensitive paint (fast PSP). In this approach, spatially resolved but noisy full-field data are fused with clean but scattered data to reconstruct full-field clean data. The fusion process is accomplished based on a compressed sensing algorithm, which has shown significantly improved performance compared with low-dimensional analysis. This is because, in low-dimensional analysis such as proper orthogonal decomposition (POD), the selection criteria of proper POD modes for reconstruction are usually based on subjective observation and the mode coefficients can be severely distorted by noise, which restricts the applications of this method to complicated flow phenomena and leads to a low-quality reconstruction. The solutions to these two problems can be expressed via mathematical optimization by determining the optimal coefficients to reconstruct clean data using the most relevant POD modes. Here, compressed sensing is used as a suitable solution to explore the sparse representation of scattered clean data based on the POD modes obtained from noisy full-field data. A high-quality reconstruction can be obtained using the optimized coefficients. The new method is first demonstrated by using fabricated patterns, demonstrating a reduction of 75% in the reconstruction error compared with POD analysis. It is thereafter successfully applied to recover the unsteady pressure field induced by a cylinder wake flow at low speed. Fast PSP measurement and microphones are used to obtain full-field but noisy pressure field data and scattered but clean data, respectively. In the cases of single and step cylinders, the reconstruction errors are approximately 5% and 25%, respectively, and the accuracy of reconstruction depends on the low dimensionality of the flow phenomena and the total number of microphone sensors. The current technique provides a reliable method to recover clean signals from strong noise, with significant potential for applications to flow measurement, control, and monitoring. *Published by AIP Publishing.*
https://doi.org/10.1063/1.5046681

## I. INTRODUCTION

In the applications of flow measurement, monitoring, and control, a particularly challenging problem is that coherent flow patterns are often buried deeply in noise. During measurement of turbulent flows, it is critical to recover useful signals buried in strong noise. Also, to control or monitor a complicated flow behavior, it is often important to extract dominant coherent flow structures from highly turbulent background using limited information at the lowest time cost. As an optical method suitable for unsteady surface pressure measurement in complex flows, fast-responding pressure-sensitive paint (fast PSP) has undergone rapid development in recent years (Gregory *et al.*, 2014). Compared with traditional pointwise measurement, fast PSP has the advantages of high spatial resolution and has found wide applications in aerodynamic and acoustic research. However, its applications to low-speed flows ($U_\infty$ < 50 m/s) are usually restricted owing to limitations of the pressure sensitivity of paint and the capability of high-speed imagers (Pastuhoff *et al.*, 2013). The poor signal-to-noise ratio (SNR) in low-speed cases and complicated noise sources make it very challenging to mine a clean signal from PSP data (Peng *et al.*, 2016).

The signal obtained using fast PSP at low-speed flows can be simply overwhelmed by the photon shot noise of the camera and the error from image misalignment (Liu, 2003). The pressure resolution obtained using a 12-bit high-speed camera is estimated to be $\Delta P_{\min}$ = 900 Pa (Liu *et al.*, 2001). It is much higher than the intensity change of the PSP obtained with high-speed imaging with a small change in the pressure of approximately $\Delta P_{\min}$ = 100 Pa. The error from image misalignment is also difficult to remove as it can couple with the signal. To reduce camera noise in PSP measurement, spatial filtering and phase-averaging are commonly used, but at the expense of spatial and temporal resolution (Asai and Yorita, 2011).

a)Author to whom correspondence should be addressed: idgnep8651@sjtu.edu.cn

More advanced algorithms are desired to effectively identify and remove camera noise and errors from image misalignment. To this end, low-dimensional analyses, such as proper orthogonal decomposition (POD) and dynamic mode decomposition (DMD), have shown significant potential in extracting coherent flow patterns from background noise (Lumley, 1967; Zhang *et al.*, 2014; and Wen *et al.*, 2016). Previous studies by Gordeyev *et al.* (2014) and Ali *et al.* (2016) have shown that POD and DMD are very useful in extracting dominant pressure features from PSP data under a high-speed flow ($U_\infty > 100$ m/s) and in an acoustic cavity, respectively. In a flow with lower speed, special care is required in the selection of low-dimensional modes for the reconstruction of clean data, as the clean-signal modes can be buried in the noisy modes. Pastuhoff *et al.* (2013) conducted unsteady PSP measurements in a low-speed wind tunnel ($U_\infty < 50$ m/s). They applied singular-value decomposition (SVD) to PSP data which could filter out most of the camera noise by selecting the modes relevant to the vortex shedding and thereafter reconstructing the image. Peng *et al.* (2016) applied a POD-based method to identify the modes of image misalignment and remove them according to a clean reference signal obtained from microphone sensors. This method was effective at a flow speed of $U_\infty = 17$ m/s, but resulted in unsatisfactory reconstruction at a lower speed of $U_\infty = 10$ m/s. Therefore, an observation-based selection method is not only subjective but also difficult to apply in the cases of complicated flow phenomena. In addition, the mode coefficients used for reconstructions can be severely distorted owing to noise, eventually resulting in low-quality reconstruction (Peng *et al.*, 2016). Thus, two open questions remain to be addressed. First, is there a robust algorithm for mode selection to quantitatively assess the relevance of POD modes to a clean flow field? Second, can the POD mode coefficients be optimized to yield a better reconstruction?

In the era of big data, data science has attracted ever-increasing interest in the field of fluid dynamics (Ling *et al.*, 2016; Ruscher *et al.*, 2017; and Kutz, 2017). Data fusion refers to the combination of different data sets to extract information that cannot be obtained from any individual data set alone. Recently, Wen *et al.* (2018) successfully recovered missing velocity data by fusing two sets of incomplete complementary measurements using particle image velocimetry. The datasets obtained from not only similar measurement techniques but also different measurement techniques can be fused together. Here, data obtained from fast PSP measurement have the advantages of high-spatial resolution but suffer from strong noise. By contrast, data obtained from microphone sensors are scattered but are much cleaner. Therefore, data fusion provides a way to fuse the two different types of data together to obtain clean full-field pressure data. This is accomplished by addressing the two open questions mentioned above. In terms of mathematical expressions, the solution should optimize coefficients to reconstruct the clean data as much as possible while using POD modes as little as possible. As a breakthrough theory in machine-learning frameworks, compressed sensing allows the reconstruction of a signal using a small number of measurements based on the fact that the signal has a sparse representation in an appropriate basis (Candes *et al.*, 2006 and Donoho, 2006). Given the common existence of low-dimensional structures in the flow field, compressed sensing has been successfully used to characterize complex fluid flows based only on a scattered-sampled signal (Bright *et al.*, 2013; Brunton and Noack, 2015; and Manohar *et al.*, 2017). Therefore, compressed sensing is a very effective solution for the above optimization problem. To the best of our knowledge, this is the first time that the concept of integrating data fusion and compressed sensing to mine a clean signal from highly noisy data has been demonstrated. For the sake of brevity, this new method is hereafter referred to as compressed data fusion.

In this study, the effectiveness of the compressed data fusion approach for mining clean data is first demonstrated and validated using fabricated patterns and the results are compared with those obtained from the POD analysis. Subsequently, this method is applied to fuse PSP measurement and microphone data to obtain a clean pressure field. Here, fast PSP measurement is used to obtain full-field pressure data with strong noise, whereas microphones are used to collect scattered-sampled clean data. At a very low speed (10 m/s), this method can successfully mine clean pressure field data on a flat plate behind a single cylinder and step cylinders.

## II. METHODOLOGY

Figure 1 presents the overall procedures of this compressed data fusion method. The procedure generally consists of two steps. First, POD analysis is applied to the full-field noisy data $\boldsymbol{P}_{noise}$ (on the left side of the figure). The POD analysis yields a group of modes $\boldsymbol{\Psi}$. An individual mode can be either clean-signal dominated or noise dominated; thus, a method to select the clean modes for reconstruction is required. The original POD coefficients $\boldsymbol{a}^{noise}$ are also distorted and noisy owing to the strong noise and they should be optimized. The original variance spectrum $\boldsymbol{\lambda}^{\mathrm{noise}}$ calculated from the coefficients shows a slight difference among the modes. It is therefore difficult to determine which mode should be included in the reconstruction. In the other step (on the right side of the figure), several high-sensitivity sensors are scattered at locations defined by the measurement matrix $\Phi$ (indicated by the black dots) to obtain clean data $\boldsymbol{P}_{scatter}$. Similarly, scattered data are also extracted at the same location from POD modes obtained in the first step, as $\Psi_{scatter} = \Phi\Psi$. The scattered data from POD modes $\Psi_{scatter}$ and scattered clean data $\boldsymbol{P}_{scatter}$ are thereafter fused via compressed sensing. This yields the optimized coefficients $\boldsymbol{a}^{opt}$, which are modulated by the clean data. In addition, the optimized variance spectrum $\boldsymbol{\lambda}^{\mathrm{opt}}$ can also be calculated from the optimized coefficients and it reflects the sparse representation of the clean signal in the POD basis. A higher variance level indicates stronger correlation between the POD mode and the clean signal. Finally, full-field and much cleaner data can be reconstructed using the POD modes $\Psi$ and the optimized coefficients $\boldsymbol{a}^{opt}$.

### A. POD analysis

The POD analysis method (Lumley, 1967) is a useful tool to extract low-dimensional coherent flow structures (POD modes) from a complicated flow phenomenon (Zhang *et al.*,

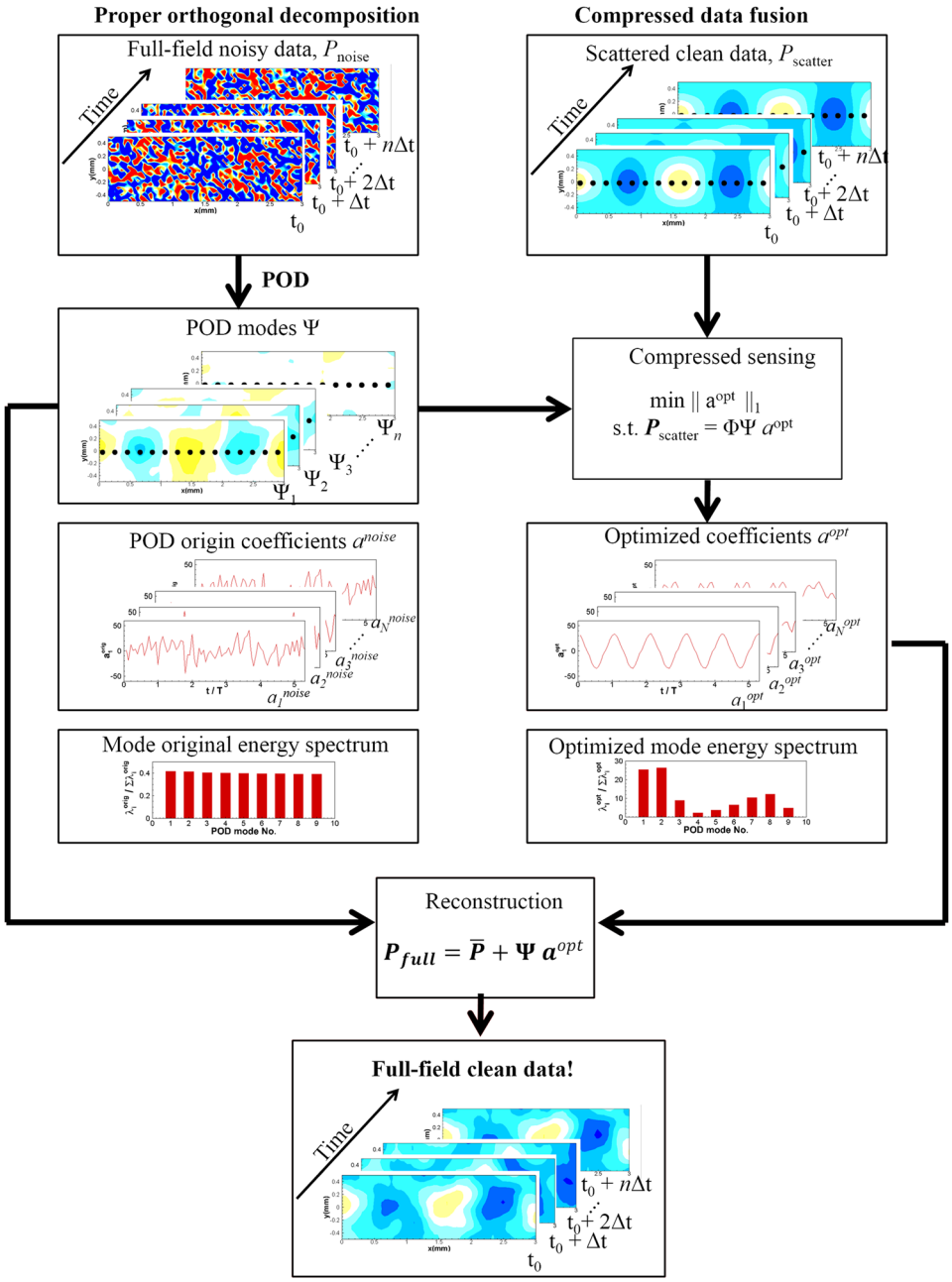



FIG. 1. Outline of the compressed data fusion approach.

2014 and Wen *et al.*, 2016). It can be used to extract a clean signal from noisy data as flow patterns are more coherent than background noise. In the present study, the "snapshot" POD method (Sirovich and Kirby, 1987) is applied to the full-field noisy data. In this POD, an unsteady field $P_k$, obtained from a data set $\boldsymbol{P}$, can be decomposed into the form of

$$P_k = \bar{\boldsymbol{P}} + \sum_{i=1}^{N} \boldsymbol{\Psi}_i a_i, \tag{1}$$

where $\bar{\boldsymbol{P}}$ is the long-time mean field, $\boldsymbol{\Psi}$ represents the spatial POD modes, $\boldsymbol{a}$ represents the corresponding POD coefficients, and $N$ represents the number of snapshots used in the calculation. The normalized temporal coefficients $\boldsymbol{a}_{\text{norm}}$ can firstly be calculated using single value decomposition as follows:

$$(\boldsymbol{a_{norm}}, \lambda) = \text{svd}(C). \tag{2}$$

Here, $C$ denotes the spatial correlation matrix calculated using inner products of the complete fluctuating parts of data $(\boldsymbol{P}, \boldsymbol{P'})$. The spatial POD modes $\boldsymbol{\Psi}$ are then calculated by projecting $\boldsymbol{P}$ onto the normalized coefficients before conducting normalization. The corresponding coefficients $\boldsymbol{a}$ used in Eq. (1) are calculated by projecting $\boldsymbol{P}$ on the normalized POD modes $\boldsymbol{\Psi}$. The eigenvalue $\boldsymbol{\lambda}$ is the variance of the corresponding POD mode $\boldsymbol{\Psi}$. The variance is usually normalized by the summation of total fluctuating variance as $\lambda_i/\sum\lambda_i$ to calculate the percentage. Higher variance level of $\lambda_i$ indicates stronger relevance of the POD mode to the reconstructed data in Eq. (1) in long-time averaged sense.

## B. Compressed data fusion

After the first step of POD analysis of the noisy data, a limited number of scattered sensors with high sensitivity

are placed to collect clean data. The measurement matrix $\Phi$ defines the sensor locations. Given the full-field data $\boldsymbol{P}_{clean}$, the scattered data are obtained as follows:

$$\boldsymbol{P}_{scatter} = \Phi \boldsymbol{P}_{clean}. \tag{3}$$

To fuse the scattered clean data and full-field noisy data, the measurement matrix $\Phi$ is also applied to the POD modes $\boldsymbol{\Psi}$ as follows:

$$\Psi_{scatter} = \Phi\Psi. \tag{4}$$

Now, the optimized POD coefficients $\boldsymbol{a}^{opt}$ should be obtained to address the two questions discussed above, regarding the quantitative assessment of the relevance of POD modes to a clean flow field and the optimization of the coefficients to yield a better reconstruction. The solutions can be expressed in a mathematical manner as follows: minimize the zero-norm of $\boldsymbol{a}^{opt}$

$$\left\|\boldsymbol{a}^{opt}\right\|_0 = card(\{i|a_i{}^{opt} \neq 0\}), \tag{5}$$

restricted to the linear constraint

$$\boldsymbol{P}_{scatter} = \Psi_{scatter}\boldsymbol{a}^{opt} = \Phi\Psi\boldsymbol{a}^{opt}. \tag{6}$$

Solving this problem is computationally intractable because it is non-deterministic polynomial (NP)-hard. Fortunately, it has been proven that $\boldsymbol{a}^{opt}$ can be obtained by minimizing the $L_1$ norm of $\boldsymbol{a}^{opt}$ rather than the $L_0$ norm (Candes *et al.*, 2006 and Donoho, 2006). This converts the problem into a computationally tractable convex optimization problem, or compressed sensing in other words, as follows:

$$\min\left\|\boldsymbol{a}^{opt}\right\|_1 \text{s.t. } \boldsymbol{P}_{scatter} = \Phi\Psi\boldsymbol{a}^{opt}. \tag{7}$$

By solving the above optimization problem, the two questions discussed above can be answered. The sparse representation of the clean signal in the POD basis is achieved by minimizing the $L_1$ norm of the optimized $\boldsymbol{a}^{opt}$. The optimized variance spectrum $\boldsymbol{\lambda}^{\text{opt}}$ can be calculated from the coefficients. The value of $\lambda_{\text{i}}{}^{\text{opt}}$ is proportional to the relevance of the corresponding POD mode $\Psi_i$ to the clean flow field. Statistically, the mode $\Psi_i$ will make a greater contribution to the reconstruction of the clean flow field if it has a higher value of $\lambda_{\text{i}}{}^{\text{opt}}$. Obviously, in order to better capture the dominating modes by POD in the first place, locations of clean-signal sensors should be optimized. For example, the sensors can be placed at the locations where the POD modes have strongest fluctuations (Bright *et al.*, 2013). As the coefficients $\boldsymbol{a}^{opt}$ are optimized according to the clean data, cleaner full-field data can be reconstructed by applying $\boldsymbol{a}^{opt}$ in Eq. (1) as follows:

$$\boldsymbol{P}_{full} = \bar{\boldsymbol{P}} + \sum_{i=1}^{k} \boldsymbol{\Psi}_i \boldsymbol{a_i}^{opt}, \tag{8}$$

where $k$ is the number of POD modes used in the reconstruction. The preferable value of $k$ can be estimated by examinations of the flow patterns captured by the modes with an aim of excluding the noise-dominated higher-order modes.

The main assumption when applying compressed data fusion is that the clean data have a sparse representation in a proper basis. The basis described in this paper is obtained using the POD method, which has been proven effective for extracting signal modes from noisy data (Pastuhoff *et al.*, 2013 and Peng *et al.*, 2016). If the reader is unfamiliar with compressed sensing, a set of lecture notes (Baraniuk, 2007) is recommended.

## III. SETUP OF SIMULATION AND FAST PSP MEASUREMENTS

As proof-of-concept, the compressed data fusion approach is first applied to fabricated data. Subsequently, it is applied to fast PSP measurement on a flat plate behind a single cylinder for a single-frequency flow and step cylinders for a more complicated flow. This section describes the setup of the simulation and measurement.

### A. Simulation of fabricated patterns

A series of fabricated patterns are constructed to simulate clean convecting patterns and those buried in noisy data. Here, the method proposed by Seena and Sung (2011) is applied to generate clean patterns as follows:

$$P_{clean} = q_0 + q_1, \tag{9}$$

$$q_0 = \exp\left(-\frac{y^2}{0.7}\right), \tag{10}$$

$$q_1 = \sum_{m=-\infty}^{m=\infty} (-1)^m \exp\left[-\left(\frac{(x-\beta m-\gamma t)^2}{d}\right) + \frac{y^2}{d}\right], \tag{11}$$

where $x$ and $y$ are the streamwise and spanwise dimensions, respectively, $d = ax + b$ is the diameter of the structure, $a$ and $b$ are constants, $\beta$ is a wavelength factor defined as the distance between two neighboring structures, and $\gamma$ is the convection velocity of the structure in the streamwise direction. The frequency of the structure $q_1$ is determined by $f = \gamma/(2\beta)$. The parameters of the fabricated patterns are listed in Table I. Figure 2 shows the evolution of clean fabricated patterns.

Subsequently, strong background Gaussian noise is added to the clean data as follows:

$$P_{noise} = P_{clean} + \left(\frac{\sigma^2{}_{clean}}{SNR}\right)^{1/2} * randn(M, 1), \tag{12}$$

where $P_{\text{noise}}$ is the combination data of clean patterns and background Gaussian noise generated by the function of MATLAB code *randn*. Here, $M$ is the size of the data $\text{U}_{\text{clean}}$. The variance of noise is calculated using the variance of the clean signal $\sigma^2{}_{clean}$ and the signal-to-noise ratio (SNR) is defined as

$$SNR = \frac{\sigma^2{}_{clean}}{\sigma^2{}_{noise}}. \tag{13}$$

Here, the value of SNR is set as 1/180 according to the SNR value obtained from the PSP data obtained later. Figure 3 presents the noisy data during one cycle of the clean patterns buried deeply in the strong background noise.

TABLE I. Parameters of the fabricated pattern.

| $a$ | $b$ | $\beta$ | $\gamma$ | $f$ |
|---|---|---|---|---|
| 0.03 | 0.05 | 0.5 | 0.8 | 0.5 |

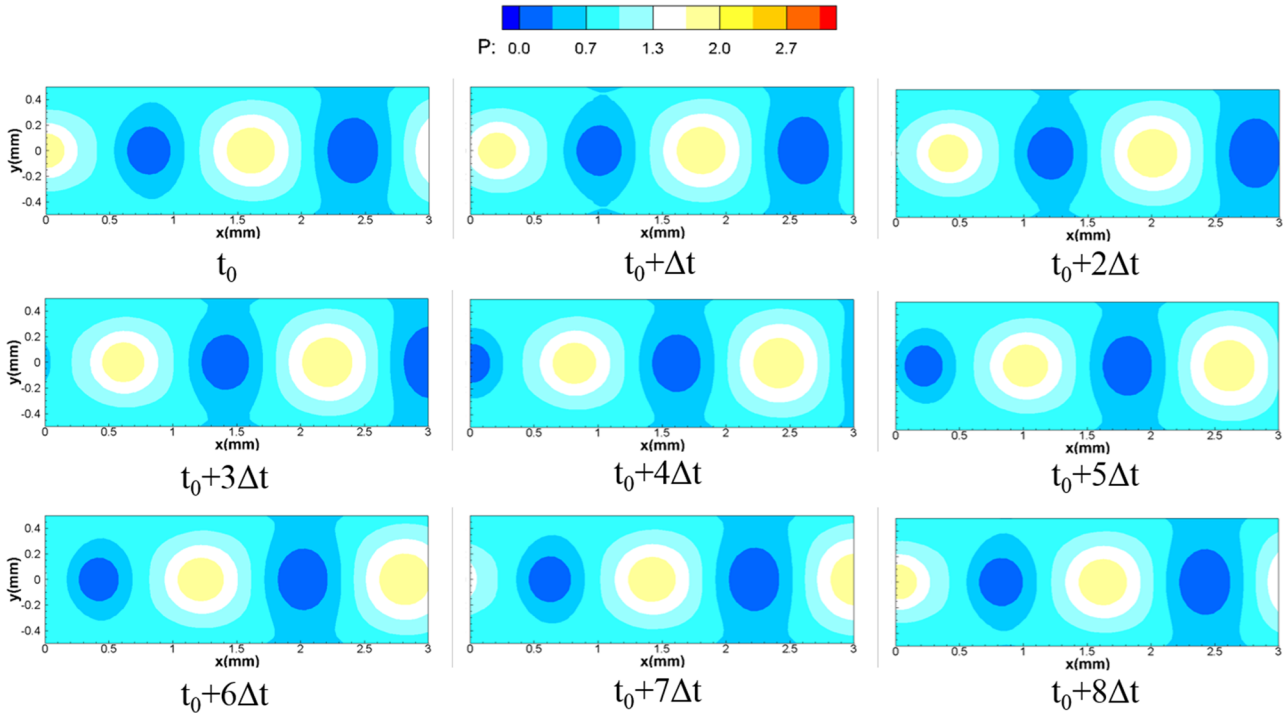



FIG. 2. Clean fabricated patterns (Δt = T/8 and T is the duration of an entire cycle).

### B. PSP experiment of a single cylinder

The experiments were performed in an open-circuit low-speed wind tunnel. The experimental setup is shown in Fig. 4(a). A circular cylinder with a diameter of 45 mm was mounted upstream of a flat plate with $c/d$ = 6 ($c$ = 3D, $d$ = 0.5D) along the centerline. Here, $c$ is the length of the flat plate, $d$ is the thickness of the flat plate, and $D$ is the diameter of the cylinder. The distance between the cylinder and the leading edge of the plate was fixed at 4D. As shown in Fig. 4(b), polymer ceramic PSP was employed to measure the pressure field, with a frequency response up

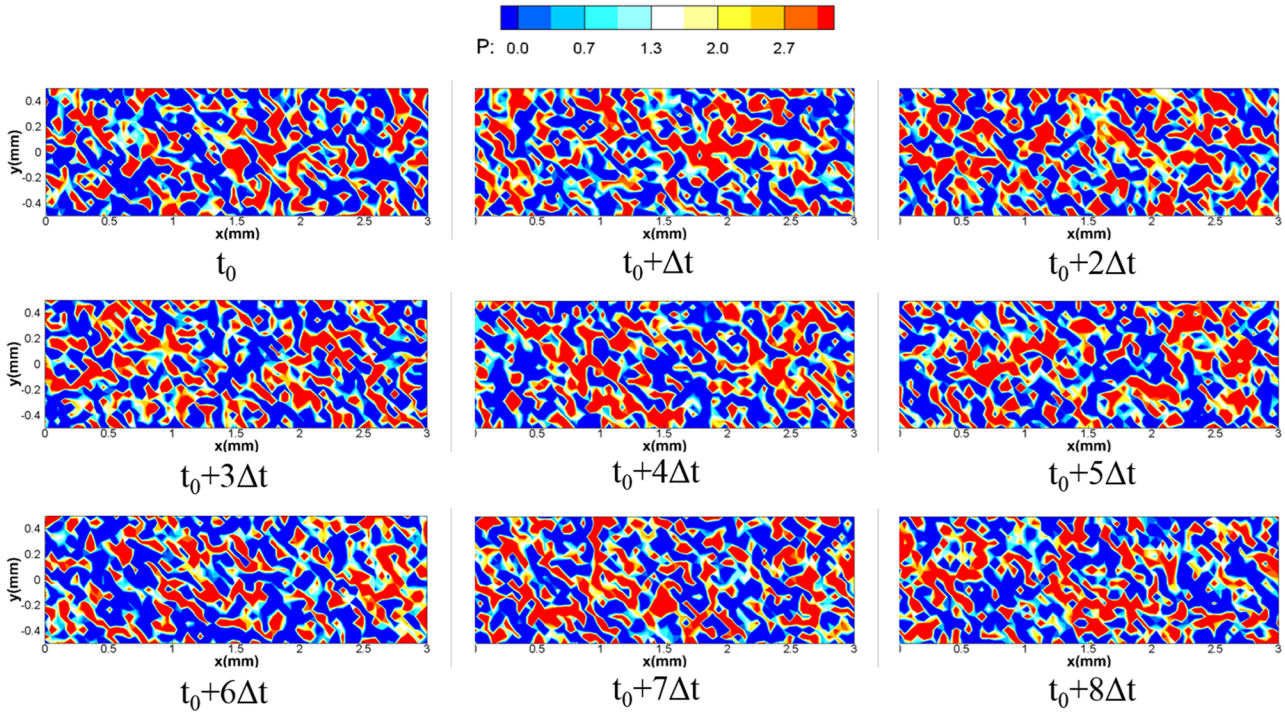



FIG. 3. Fabricated patterns with strong noise (SNR = 1/180).

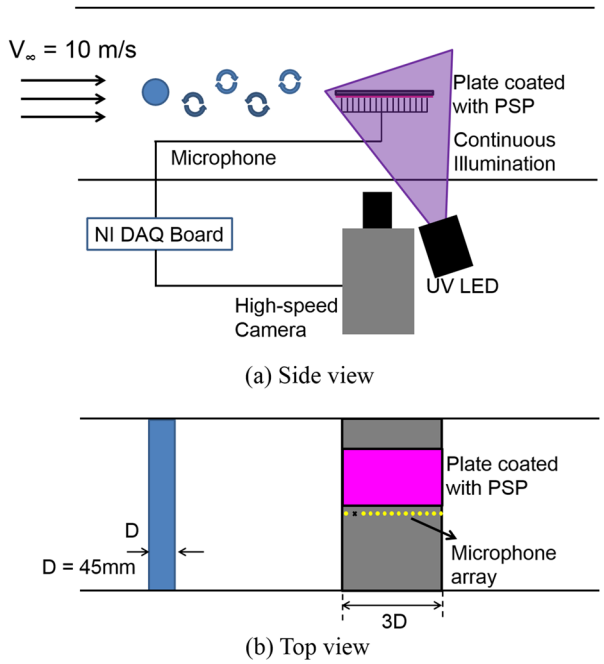



FIG. 4. Schematics of the experimental setup for unsteady PSP measurement and microphone measurement using a single cylinder: (a) side view and (b) top view of the plate. (The microphone indicated by the black cross is used for validation.)

to 6 kHz (Gregory *et al.*, 2008). The paint was excited continuously using a UV-light-emitting-diode, and the luminescent signal was captured using a 12-bit high-speed CMOS camera (dimax HS4, pco.) at 2k fps to collect 8000 images in total. The experiment was performed at a very low flow speed, i.e., $U_\infty$ = 10 m/s, which is far below the speed in typical PSP applications (>50 m/s). Therefore, it is very challenging to obtain a clean pressure field from strong background noise and the noise of image misalignment. Fourteen microphones (GRAS, 40PQ) were used as high-sensitivity sensors to acquire low-noise scattered pressure data. They were equally spaced along the centerline of the plate, as shown in Fig. 4(b). Typically, the sensitivity of the microphones was approximately 7.7–8.2 mV/Pa, resulting in a magnitude error within ±1 dB. Since the measurement error of the microphone was much lower than that of PSP, microphone data were treated as the clean signal. It should be noted that the accuracy of the microphone measurement can be further improved by applying frequency filtering techniques on the time-resolved data, which is out of the scope of the current work. Further details of the experiment can be found in the previous publication by Peng *et al.* (2016). One of the microphones was used for validation as shown in Fig. 4(b).

### C. PSP experiment of step cylinders

To evaluate the performance of the current method for a more complicated flow with relatively fewer high-sensitivity sensors, step cylinders were used instead of the single cylinder, as shown in Fig. 5. Twelve microphones were scattered across the measurement area of fast PSP. One of them was excluded from the fusion process and used for validation. Here,

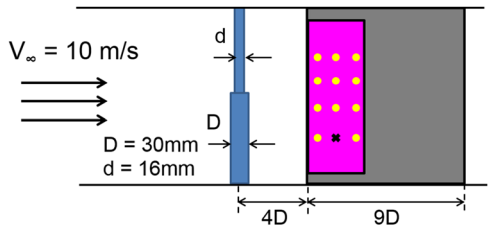



FIG. 5. Schematics of the experimental setup for unsteady PSP and microphone measurement using step cylinders. (The microphone indicated by the black cross is excluded from data fusion and used for validation.)

the microphones were installed in 1-mm-diameter holes. It should be noted that the basecoat of fast PSP could be directly sprayed onto the surface without clogging the holes as long as the nozzle of the airbrush was well controlled and a thin layer of paint was applied for each spray.

## IV. RESULTS

### A. Fabricated patterns

Before dealing with the noisy data, POD analysis was first applied to the full-field clean data for a benchmark reference. Eight-hundred successive snapshots of clean patterns with a sampling frequency of 8 Hz were analyzed using the POD algorithm. The variance spectrum, POD modes, and corresponding coefficients are presented in Fig. 6. From the variance spectrum in Fig. 6(a), there are only two dominating modes, i.e., $\Psi_1$ and $\Psi_2$, which consume approximately 99% of the total fluctuation variance. The spatial distributions of the first two POD modes are shown in Figs. 6(b) and 6(c). Similar to a previous study (Wen *et al.*, 2016), the first two modes couple in pairs and share similarities. The spatial correlation with the streamwise structures is approximately 1/4 of the wavelength of the convecting structures in the two modes. The phase correlation is more precisely presented by the coefficients. The plots of both coefficients show apparent periodic fluctuation, which indicates the convection movement of the patterns.

Subsequently, POD is applied to the highly noisy data. The results of POD analysis are also different from those of clean data. As shown in Fig. 7(a), the variance distribution among the POD modes is monotonous. The first 10 POD modes almost have equal variance levels. Therefore, it is difficult to distinguish the signal-dominated modes from the noise-dominated ones. By contrast, the spatial distribution of the POD modes can provide some useful clues. As shown in Figs. 7(b)–7(e), the first four modes generally capture the coherent patterns. As the mode number increases, the coherence dissipates fast. However, even though the modes can be selected based on observation, the mode coefficients are severely distorted, which could result in low-quality reconstruction. For example, the coefficients $a_1{}^{\text{noise}}$ and $a_2{}^{\text{noise}}$ hardly retain the periodic fluctuation of the convecting patterns.

Figure 8 presents the reconstruction using the first four POD modes and the original noisy coefficients. As expected, although the background noise is significantly reduced, the temporal revolution of the coherent patterns cannot be recovered. For an effective comparison, a horizontal line passing

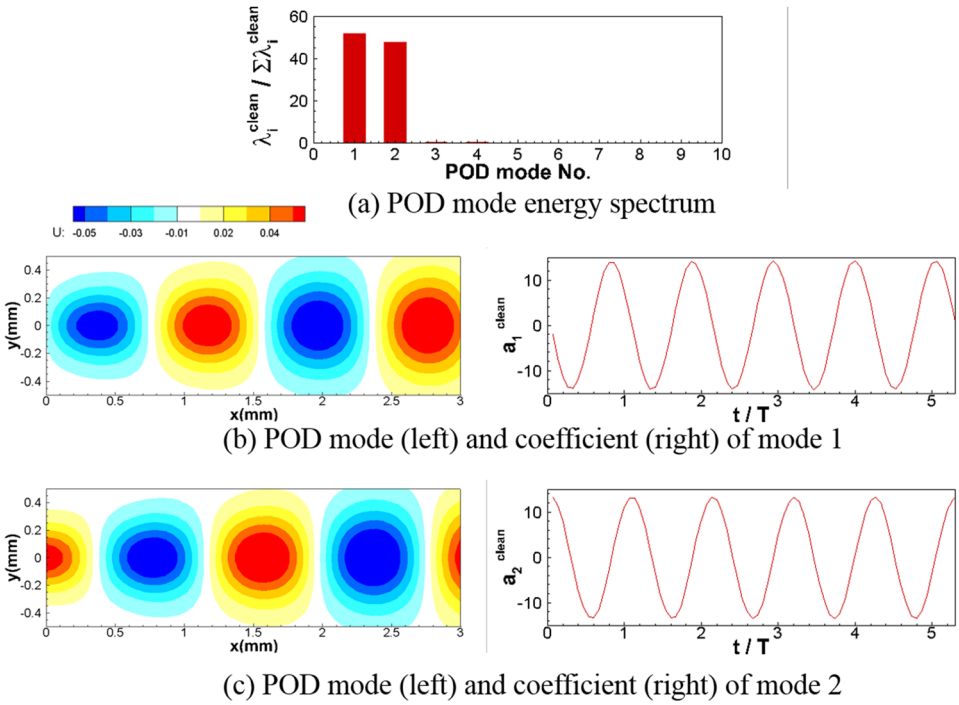



FIG. 6. POD results of clean patterns.

through the upper edge of the convecting patterns (as shown in Fig. 8) is used to compare the reconstructed data obtained using POD analysis, original noisy data, and the ground-truth clean data without any noise. As shown in Fig. 9, POD analysis is effective for reducing the background noise. However, the temporal evolution of the reconstructed signal is severely distorted. Within a cycle, the reconstructed data only match the clean data to a good extent at only two instances, i.e., $t_0 + 4\Delta t$ and $t_0 + 6\Delta t$, where $\Delta t = T/8$ and $T$ is the duration of an entire cycle.

To optimize the variance distribution and temporal coefficients of the POD modes, 15 imaginary sensors are scattered along the streamwise middle line to extract clean data, as shown in Fig. 10(a). The first nine POD modes are used to test the effectiveness of the current compressed data fusion method for mode selection. An office desktop takes several minutes to process the 800 data sets. As shown in Fig. 10(b), the optimized variance distribution of the POD modes is more informative than the original one. Similar to the results of clean data (Fig. 6), the first two POD modes are dominating ones, which consume approximately 50% of the total variance. The high variance level indicates that the two modes have the greatest relevance to the clean data. This is consistent with the observation of the spatial distribution of the modes. The remaining modes are relatively less related to the clean data, as indicated by the lower variance levels. With the variance ranking, the optimized number of POD modes can be determined in applications of real-time reconstruction. More importantly, the coefficients are also modulated according to the clean data. As shown in Figs. 10(c) and 10(d), the coefficients of the first two dominating modes return to a periodic shape with a phase difference of approximately 90°, similar to those of clean data (Fig. 6). The coefficients of the noise-dominated modes have much weaker fluctuations and will contribute less to the reconstruction of the data, as shown in Figs. 10(e) and 10(f).

With the optimized coefficients, the full-field clean data can be reconstructed using Eq. (8). Here, all nine modes are used as they are weighted by the corresponding coefficients already. As shown in Fig. 11, the reconstructed data not only capture the shapes of the coherent patterns but also resolve the evolution process. Similar to the previous comparison, Fig. 12 presents the comparison of ground-truth clean data and reconstructed data obtained using POD analysis and compressed data fusion along the same validation line. Within an entire cycle, the compressed data fusion performs much better than the POD analysis. The reconstructed data obtained from compressed data fusion are consistent with the clean data. The errors of reconstructed data $P_{recon}$ along the validation line are also estimated by comparing with the ground truth clean data $P_{clean}$ using the following equation:

$$E = \frac{abs(P_{recon} - P_{clean})}{\max(P_{clean}) - \min(P_{clean})} 100\%, \tag{14}$$

where $\max(P_{clean})$ is the maximum value and $\min(P_{clean})$ is the minimum value of the clean data. The maximum and minimum values are obtained over the entire actuation cycle. According to this definition, the POD analysis has a reconstruction error of more than 20%. This method is outperformed by the current compressed data fusion approach, which has an error less than 5%.

### B. PSP data of a single cylinder

Subsequently, the current compressed data fusion approach is applied to fast PSP measurement on a plate in the wake of a single cylinder. As discussed before, fast PSP

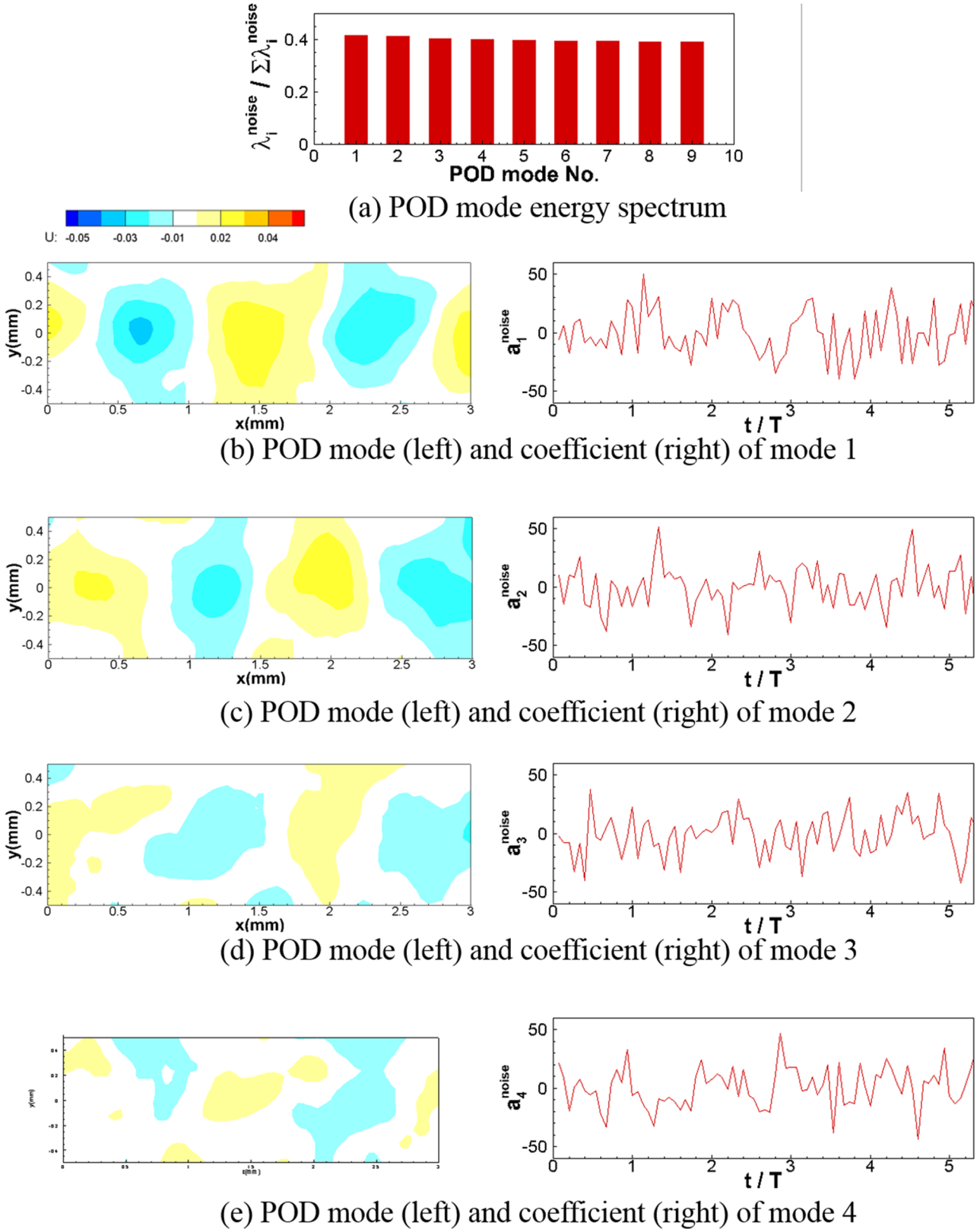



FIG. 7. POD results of noisy fabricated patterns.

measurement was used to obtain the full-field noisy pressure field, whereas microphones were used to collect scattered clean data.

First, the raw data obtained from PSP measurement are presented to provide an intuitive impression of the noisy data. As shown in Fig. 13, instantaneous pressure fields within a fluctuation cycle are presented. Here, the actuation period is determined by the fluctuation frequency of the microphone data, which is consistent with the vortex shedding frequency (Strouhal number $St$ = 0.2). From the raw

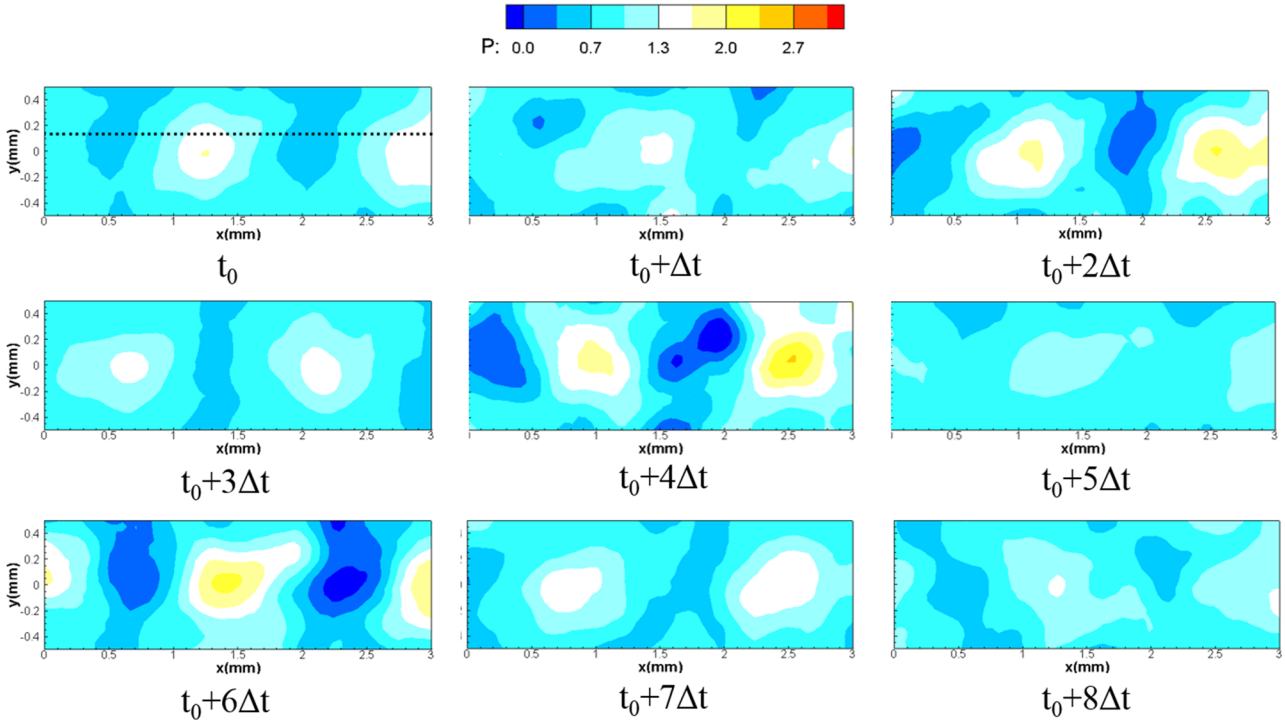



FIG. 8. Reconstruction using traditional POD analysis. (The dashed line in the figure at $t_0$ indicates the location for comparison.)

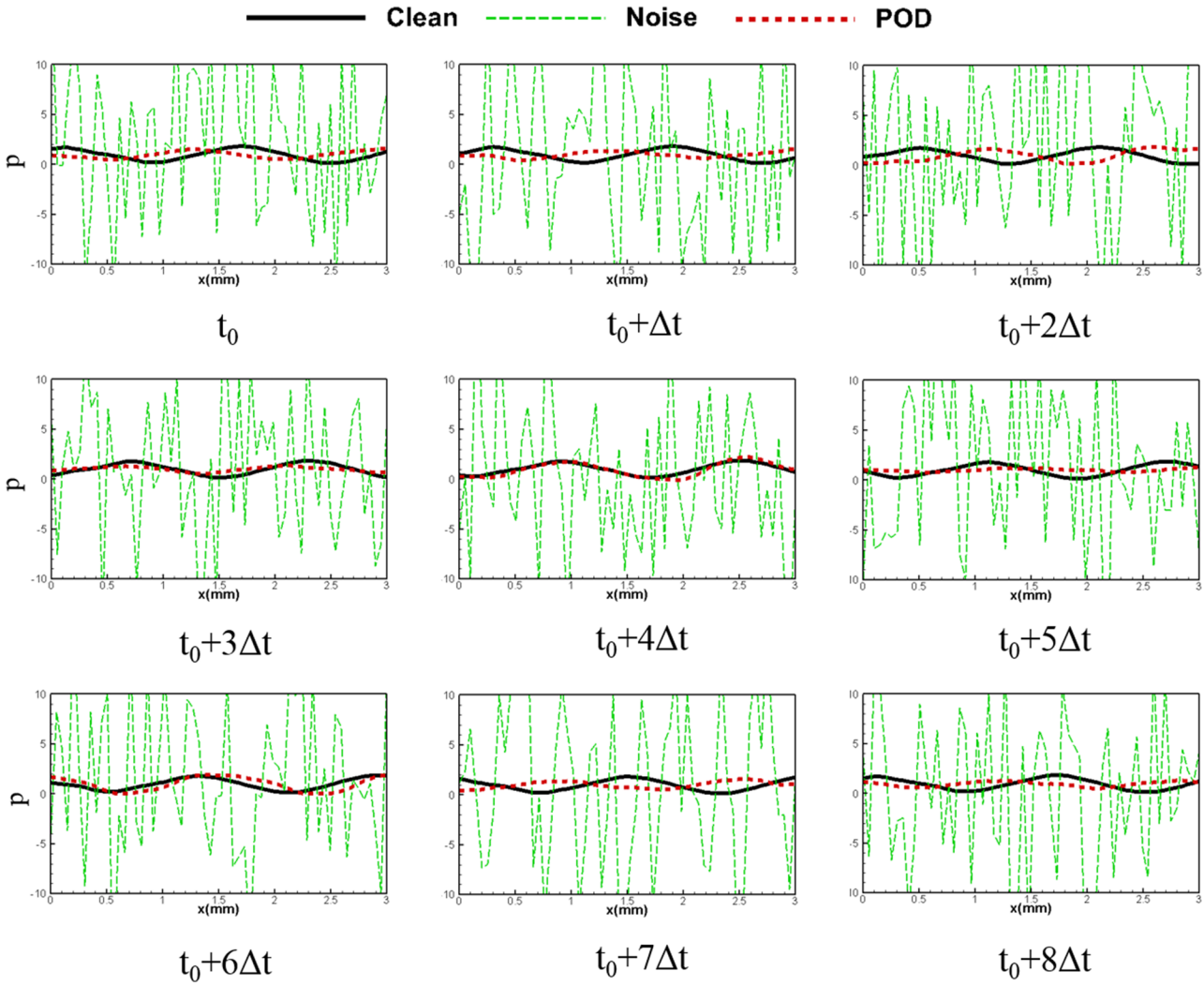



FIG. 9. Comparison of clean data, noisy data, and reconstructed data obtained using POD analysis. (The location for comparison is shown in Fig. 8.)

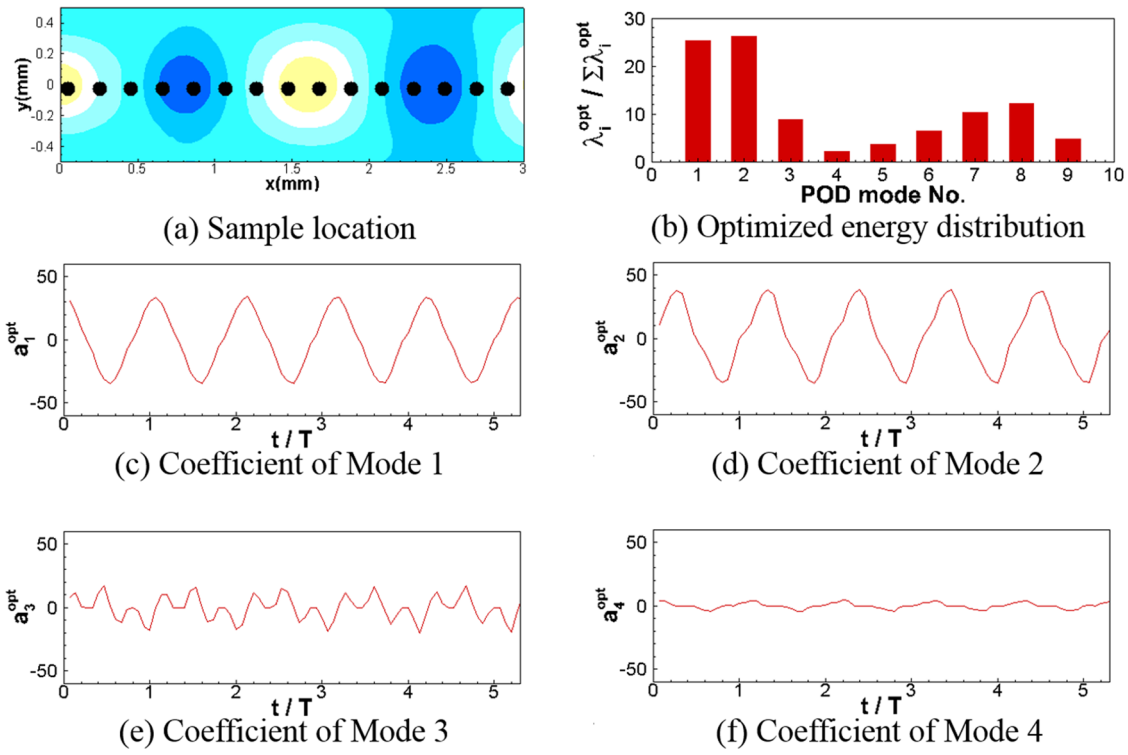



FIG. 10. Results of compressed data fusion: (a) sample locations (black dots) for clean data; (b) optimized variance distribution; [(c)–(f)] optimized coefficients.

data, even near the leading edge, it is difficult to extract meaningful pressure information owing to the overwhelming noise.

As the initial step of the current data fusion approach, POD analysis is applied to extract low-dimensional modes from the noisy data. As shown by the variance distribution in Fig. 14(a), the variance levels of the first four POD modes of a total number of 20 modes are clearly higher than the remaining modes. However, a close examination of the spatial distribution of the four modes reveals some interesting features.

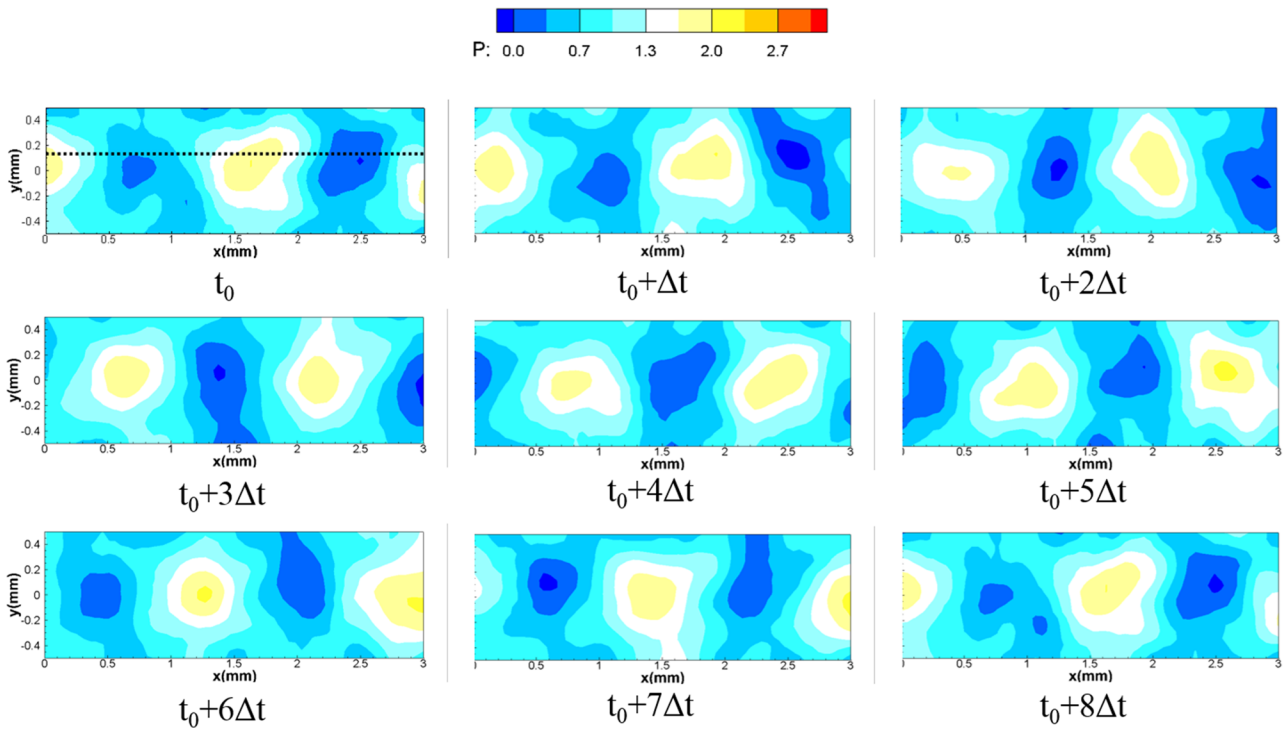



FIG. 11. Reconstruction using the compressed data fusion approach.

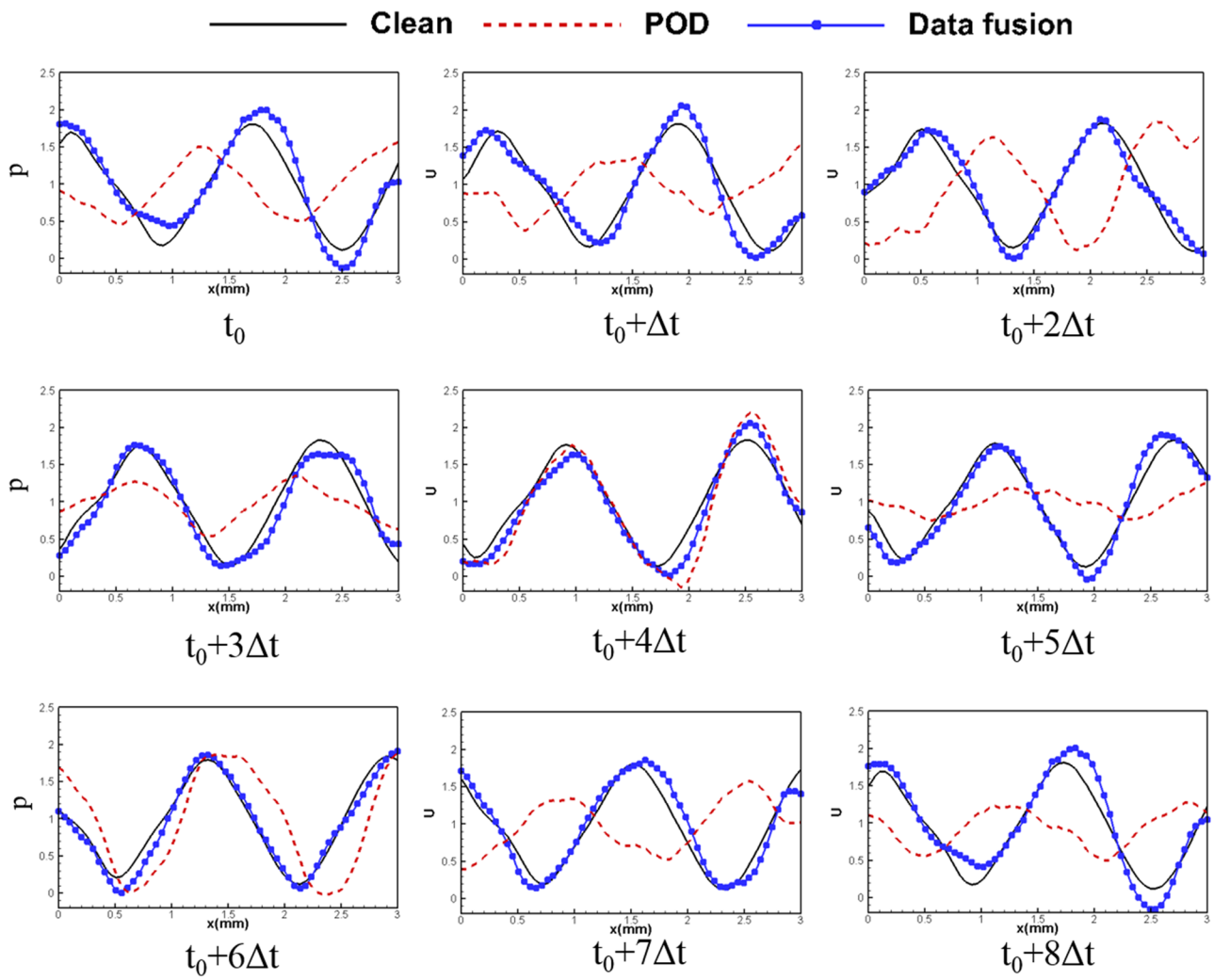



FIG. 12. Comparison of clean data and reconstructed data obtained using POD analysis and compressed data fusion. (The location for comparison is shown in Fig. 11.)

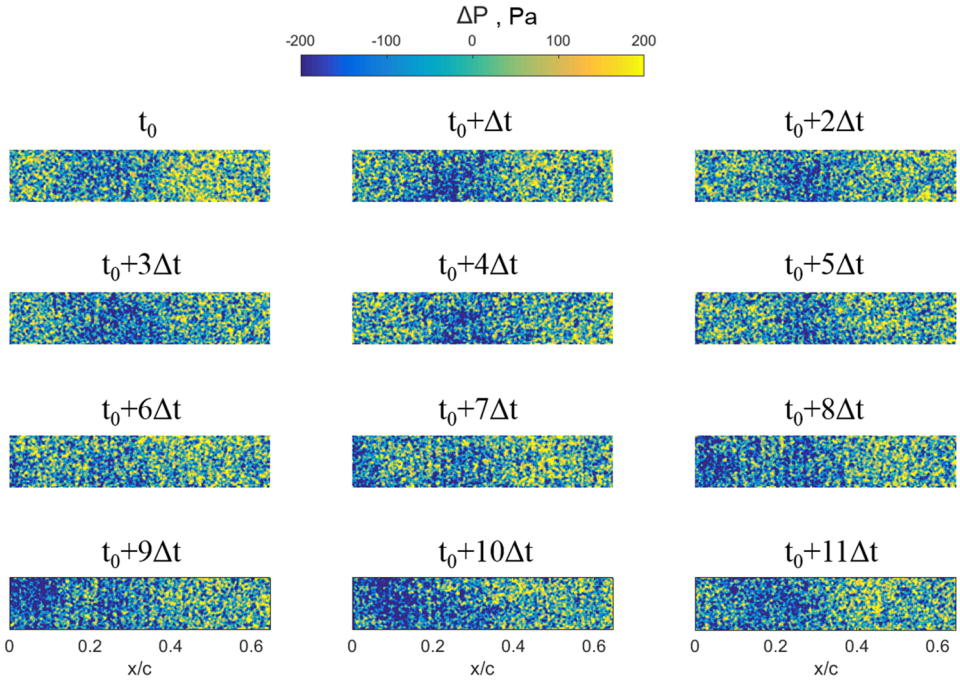
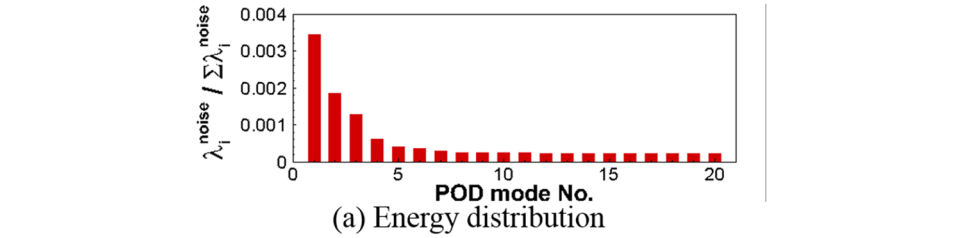
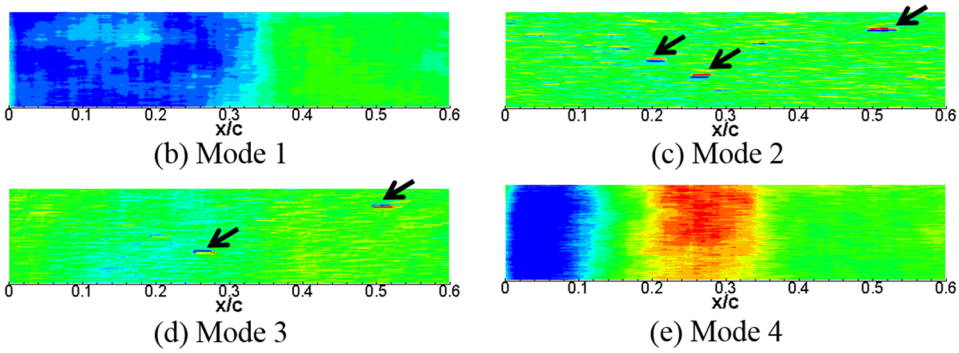



FIG. 13. Instantaneous pressure fluctuations (Δt = T/11 and T is the duration of an entire cycle determined using microphone data).



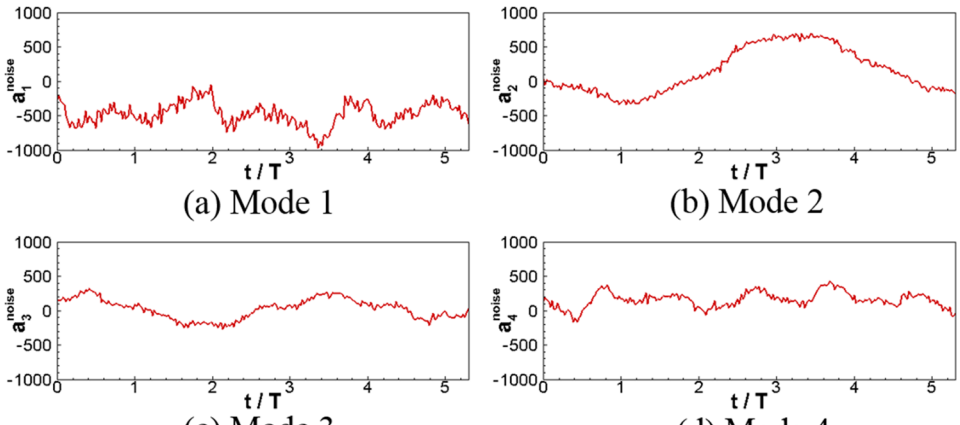

(a) Energy distribution




(b) Mode 1 (c) Mode 2

(d) Mode 3 (e) Mode 4

FIG. 14. POD results of noisy PSP measurement.




(a) Mode 1 (b) Mode 2

(c) Mode 3 (d) Mode 4

FIG. 15. Coefficients of the POD modes from noisy PSP measurement.

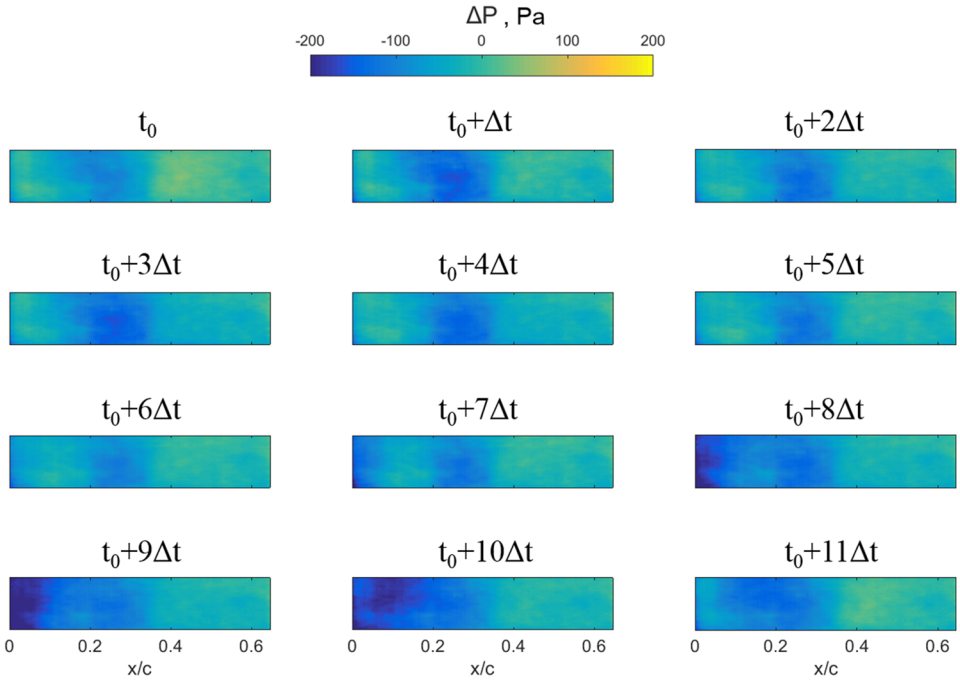



FIG. 16. Reconstruction of PSP data using traditional POD analysis.

As shown in Figs. 14(c) and 14(d), modes 2 and 3 are dominated by features related to defects of PSP paint, which is a result of image misalignment (owing to slight vibration), as indicated by the arrows. By contrast, modes 1 and 4 appear to be signal-dominated modes, as shown in Figs. 14(b) and 14(e). In addition, the coefficients are severely distorted as expected, as shown in Fig. 15. The coefficients of modes 1 and 4 cannot represent the periodic fluctuations very well. For example, the coefficient of mode 1 is well below zero. For the noisy modes (mode 2 and 3), the coefficients have much lower frequency. Based on these subjective observations, the first 10 POD modes excluding modes 2, 3, and 7 were used in the reconstruction of the pressure field as same as that in a previous study (Peng *et al.*, 2016).

Figure 16 presents the reconstructed instantaneous pressure field obtained from the above POD analysis. Compared with the raw data (Fig. 13), it is evident that the noise has been effectively removed. However, the pressure is generally below zero, which is not reasonable and indicates a large reconstruction error based on the microphone data (as discussed later). This is the result of the negative coefficient values of mode 1 as discussed above. Figure 17 plots the detailed comparisons between the original PSP data, reconstructed PSP data obtained using POD analysis, and microphone data. Here, the PSP data for comparison are obtained at an adjacent location with an area size similar to that of the microphone sensor. It can be observed that although the background noise of the original PSP data can be removed using POD analysis, the reconstructed data evidently deviate from the microphone data. Generally, the microphone data show apparent fluctuating behaviors in the range ±50 Pa, whereas the reconstructed data vibrate between −200 Pa and 0.

Therefore, the current compressed data fusion approach is applied using the scattered microphone data as a clean signal to optimize the coefficients of the POD modes. As shown

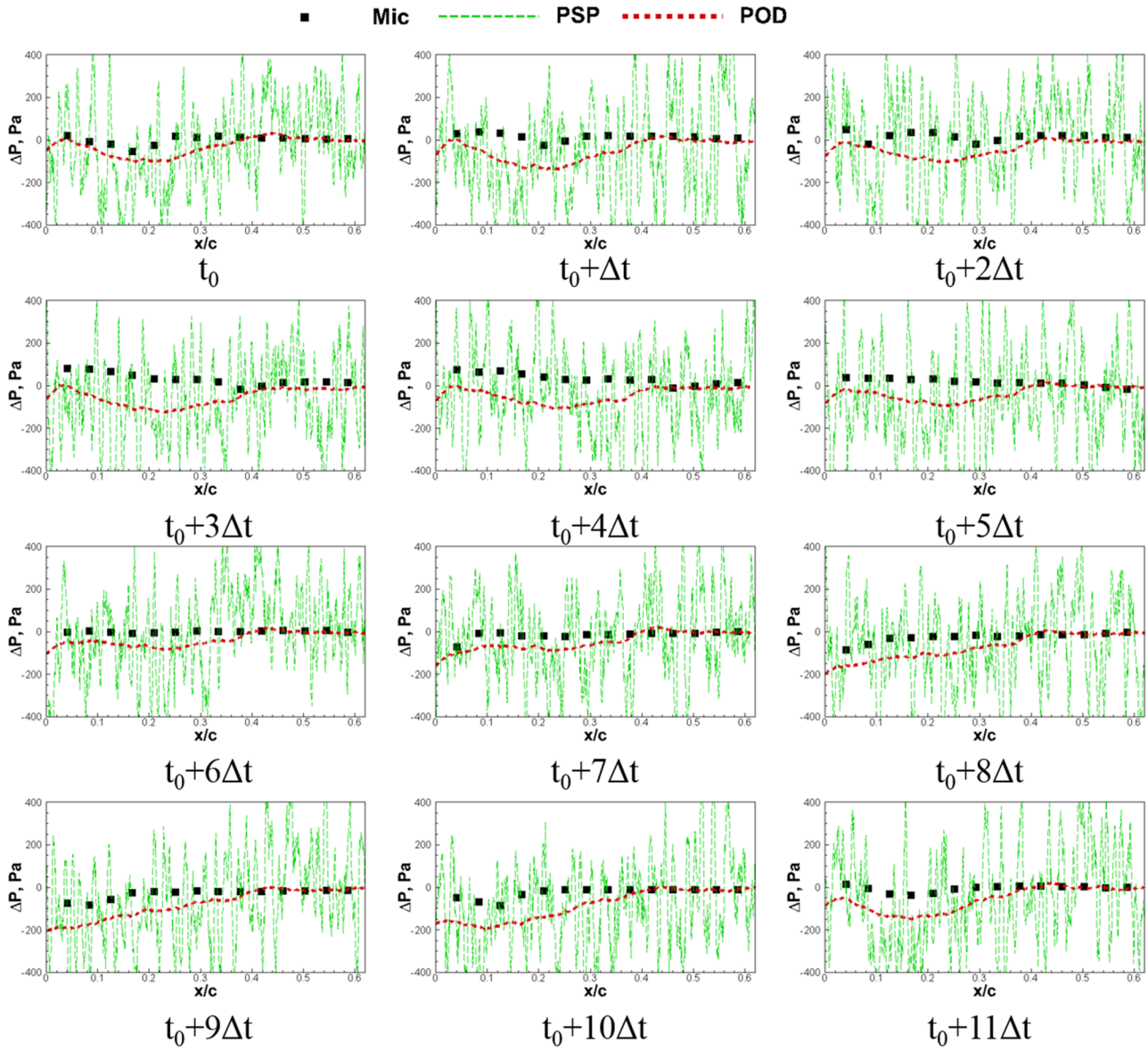



FIG. 17. Comparison of microphone data, raw PSP data, and reconstructed PSP data using POD analysis.

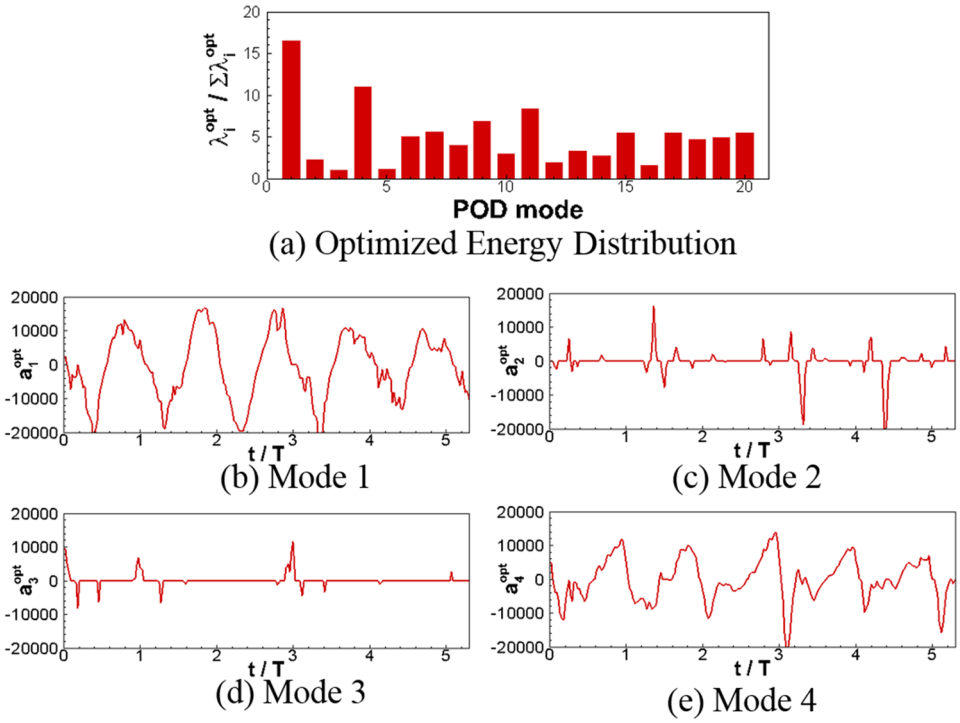



FIG. 18. Results of compressed data fusion: (a) optimized variance distribution; [(b)–(e)] optimized coefficients.

in Fig. 18(a), the optimized variance distribution shows that modes 1 and 4 are most relevant to the clean data, whereas the modes 2 and 3 have very weak relevance to the clean data. This is consistent with the above observations. More importantly, the optimized coefficients of modes 1 and 4 can well represent the periodic pressure fluctuations on the plate, as shown in Figs. 18(b) and 18(e). For example, the optimized coefficient of mode 1 fluctuates in the range ±10 000, which is much stronger than the original coefficient [Fig. 15(a)]. By contrast, the coefficients of the modes 2 and 3, which are noisy modes, are apparently suppressed, as shown in Figs. 18(c) and 18(d).

With the optimized coefficients, the pressure field on the plate can be reconstructed with significantly improved fidelity. As shown in Fig. 19, the reconstructed pressure field can better reveal the periodic fluctuation. For example, during the first half of the cycle, high pressure is induced in the leading edge of the plate with a peak value of approximately 50 Pa. During the second half, the pressure has the lowest value down to −50 Pa. Figure 20 plots the detailed comparison of the microphone data and reconstructed data obtained using POD analysis and compressed data fusion. The compressed data fusion approach shows significantly improved performance compared with

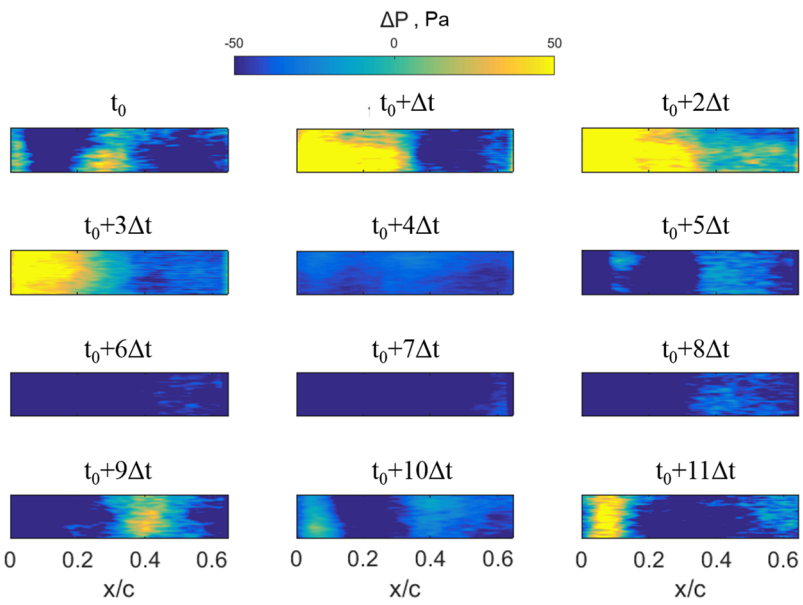



FIG. 19. Reconstruction of PSP data using the compressed data fusion approach.

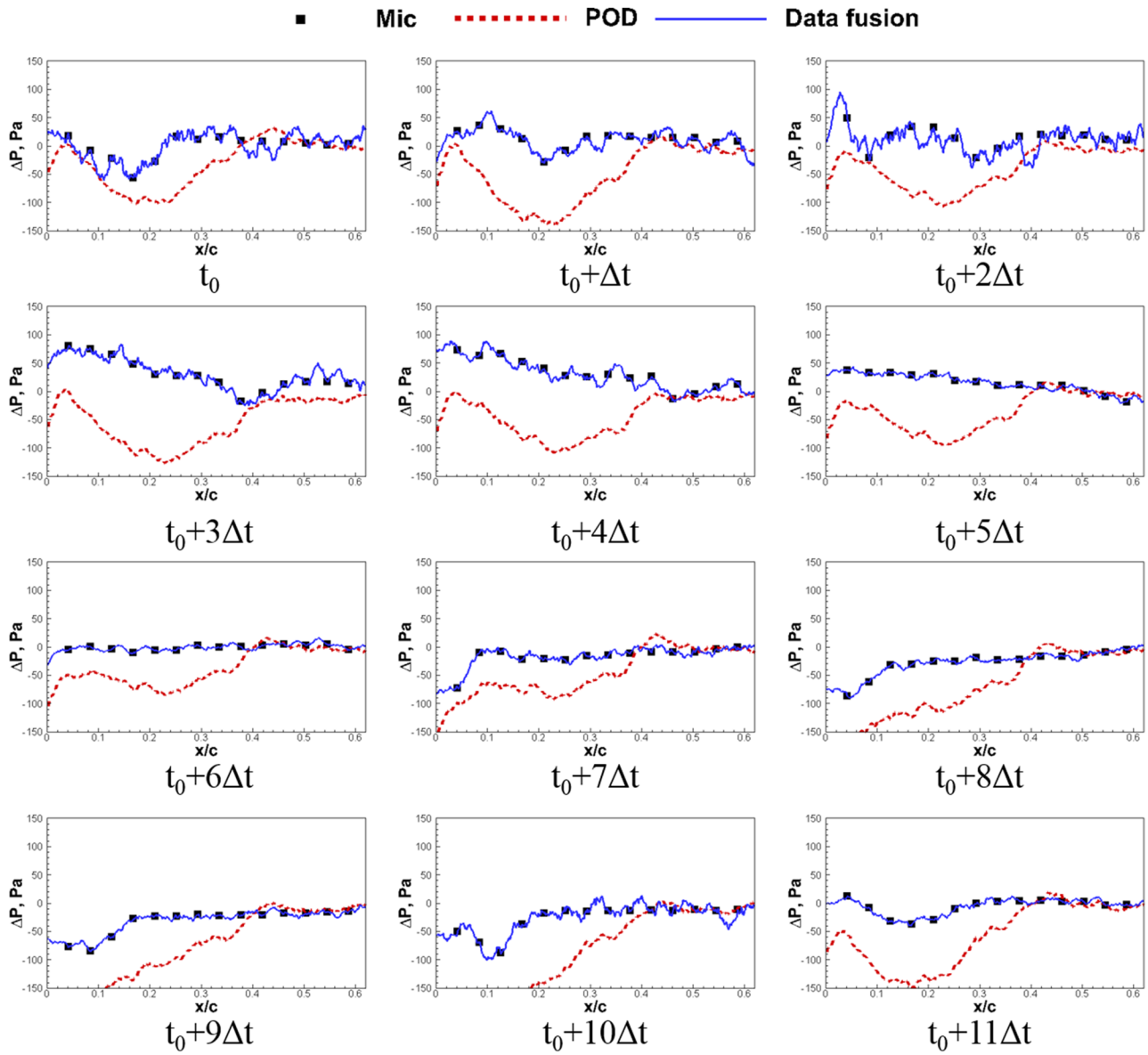



FIG. 20. Comparison of microphone data and reconstructed PSP data obtained using POD analysis and compressed data fusion.

POD analysis. Within an entire cycle, the profiles of reconstructed data obtained using compressed data fusion follow the microphone data very well, whereas the profiles obtained using POD analysis deviate from the microphone data significantly. To further evaluate the performance of compressed data fusion, a microphone is removed from the data fusion process

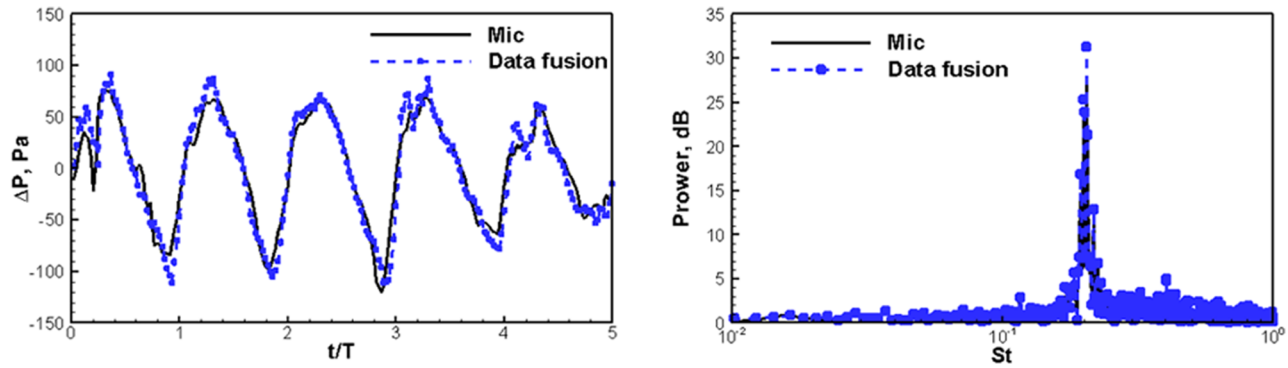



FIG. 21. Comparison of the missing microphone data and reconstructed PSP data obtained using compressed data fusion. (The microphone is excluded from data fusion and used for validation, as shown in Fig. 4.)

and it is used for validation instead (as shown in Fig. 4). Figure 21 plots the comparison between the data obtained from this microphone and the reconstructed data based on the other 13 microphones. The consistency is satisfactory in both temporal and frequency domains. The reconstruction error of the current data fusion method is estimated to be within 5% using Eq. (14).

### C. PSP experiment of step cylinders

To evaluate the performance of the current method for a more complicated flow with relatively fewer high-sensitivity sensors, step cylinders were used instead of the single cylinder. The instantaneous pressure fields obtained from fast PSP measurement are presented in Fig. 22. It can be observed that the entire field is dominated by strong noise. Subsequently, the pressure field is reconstructed using compressed data fusion. As shown in Fig. 23, the fluctuating pressure field owing to the wake flow of the step cylinders is recovered to a large extent, which resolves their three-dimensional behaviors. Figure 24 compares the original noisy PSP data, reconstructed data, and the microphone data for validation. In the temporal domain, it is confirmed that the current data fusion approach can successfully remove the strong background noise and recover the clean data. In the frequency domain, the reconstructed data are consistent with the microphone data at Strouhal number $St \approx 0.2$. Owing to the more complicated flow behaviors and relatively fewer microphone sensors, the reconstructed error increases to approximately 25%, compared with the case of a single cylinder. However, it is still impressive that a highly three-dimensional

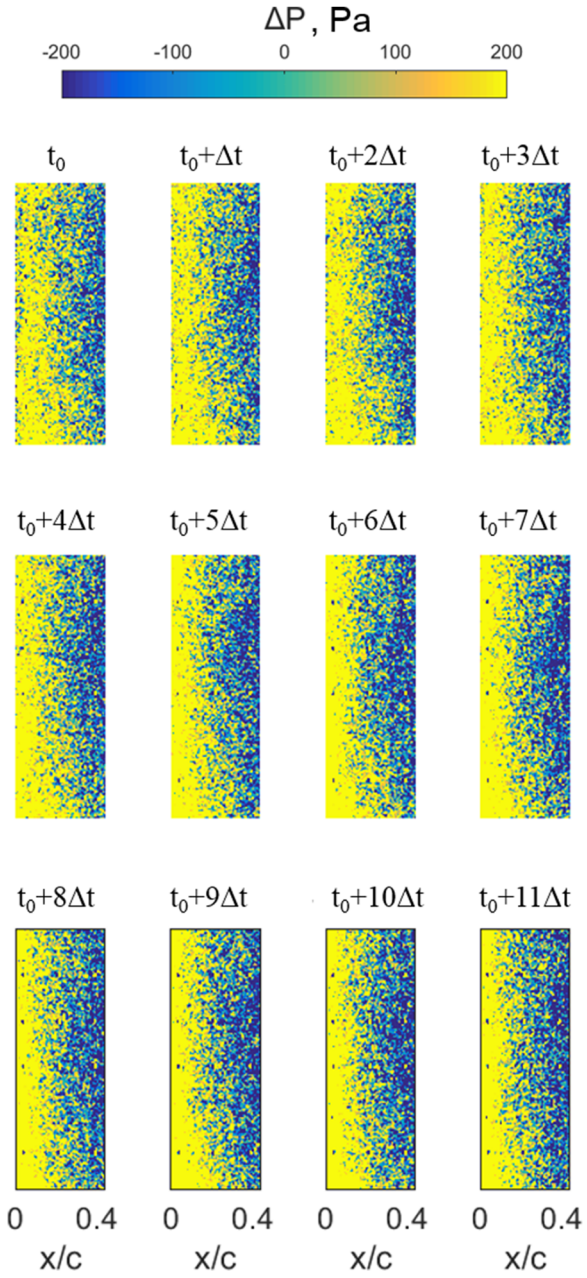



FIG. 22. Instantaneous pressure fluctuations (Δt = T/11 and T is the duration of an entire cycle determined using the validation microphone data).

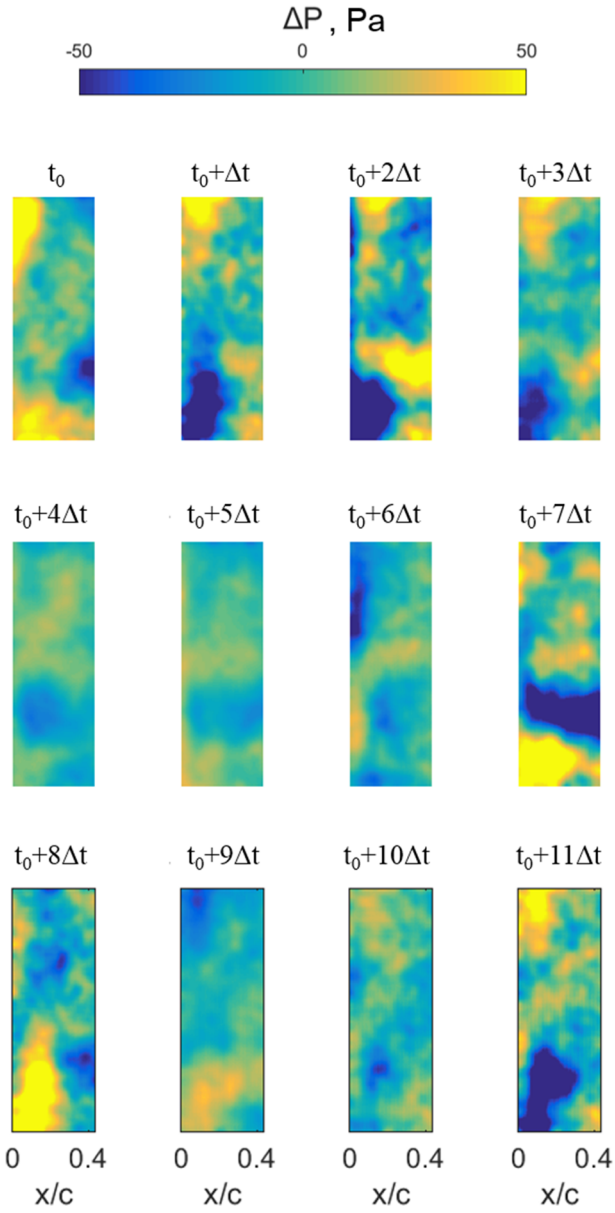



FIG. 23. Reconstruction of PSP data using the compressed data fusion approach.

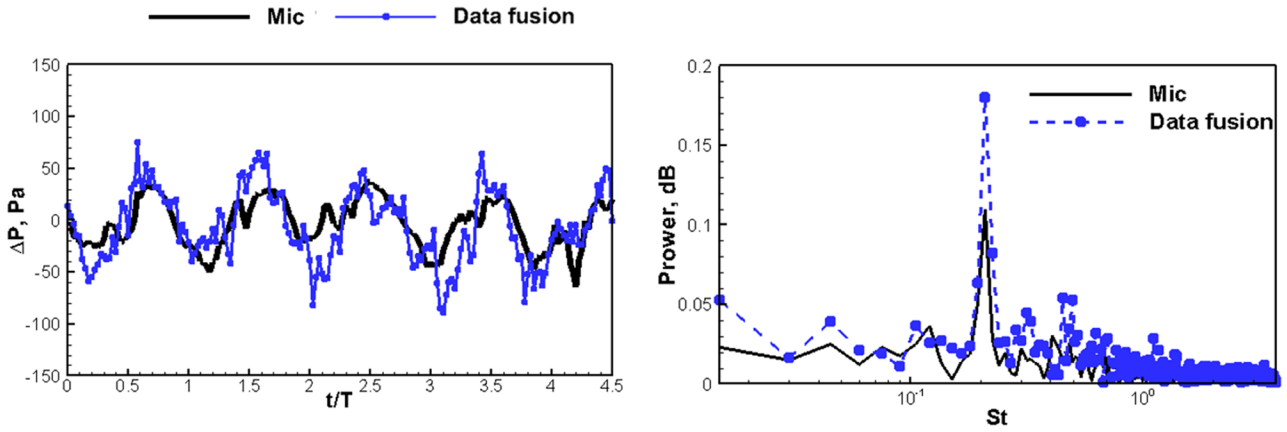



FIG. 24. Comparison of the microphone and reconstructed PSP data obtained using compressed data fusion. (The microphone is excluded from data fusion and used for validation, as shown in Fig. 5.)

pressure field can be recovered to a large extent using only 11 microphones.

## V. CONCLUSION

In this study, a new algorithm named compressed data fusion was developed for mining clean data from highly noisy data and it was successfully applied to fast PSP measurement. Scattered clean data were used to optimize the coefficients of low-dimensional (here, POD) modes obtained from full-field noisy data. Compressed sensing was used as a solution to this optimization problem. With the optimized coefficients, the relevance of the POD modes was quantitatively evaluated. More importantly, a better reconstruction of the clean full-field data could be obtained.

Fabricated patterns with strong noise were first used to validate this approach. The reconstruction error was estimated to be less than 5%, which is only 1/4 of that obtained using POD analysis. Subsequently, this approach was applied to a real fluid experiment, i.e., fast PSP measurement of pressure fields in the wake of a single cylinder and step cylinders with different diameters. It was confirmed again that the current compressed data fusion approach performs much better than POD analysis. The reconstruction error was approximately 5% and 25% for the cases of single and step cylinders, respectively. As demonstrated by the above two cases, the accuracy of the reconstruction is dependent on several key parameters, such as the low dimensionality of the flow phenomenon, total number of high-sensitivity sensors, and locations of the sensors. In general, fewer sensors will be required when the low-dimensional feature of the flow is more apparent. In addition, previous studies have determined that the sensor locations can be optimized based on the full-field data. For example, the sensors can be placed at the locations of maximum and minimum values of the POD modes (Bright *et al.*, 2013). Therefore, in order to evaluate the effectiveness and limitation of this approach in more general cases, a detailed examination on the performance of this approach under different conditions (number of sensors, sensor configurations, and flow fields) is highly desirable in the next phase. In addition, as a method of decomposing flow fields based on their frequency spectrum, dynamic mode decomposition should be a viable approach to obtain optimized modes which will be tested in future work.

Notably, this compressed data fusion approach can be applied to not only flow measurement but also flow control or monitoring. As the core optimization step of the current approach only deals with scattered data, it can be very fast and can be used in real-time applications. For example, to control or monitor a complicated flow phenomenon, it is often required to obtain full-field data from scattered sensors at the lowest time cost. Low-dimensional analyses are commonly used in this application given that the low-dimensional modes are *clean* (Mainini and Willcox, 2015). The current approach provides a solution when the basis of low-dimensional modes is contaminated, which is more common in real-world applications.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support from the National Natural Science Foundation of China (Grant Nos. 11702172, 11502144, and 11725209).

Ali, M. Y., Anshuman, P., and Gregory, J. W., "Dynamic mode decomposition of fast pressure sensitive paint data," Sensors **16**(6), 862 (2016).

Asai, K. and Yorita, D., "Unsteady PSP measurement in low-speed flow—Overview of recent advancement at Tohoku University," in *49th AIAA Aerospace Sciences Meeting* (AIAA, 2011), Vol. 847.

Baraniuk, R., "Compressive sensing," IEEE Signal Process. Mag. **24**(4), 118–124 (2007).

Bright, I., Lin, G., and Kutz, J. N., "Compressive sensing and machine learning strategies for characterizing the flow around a cylinder with limited pressure measurements," Phys. Fluids **25**(12), 127102 (2013).

Brunton, S. L. and Noack, B. R., "Closed-loop turbulence control: Progress and challenges," Appl. Mech. Rev. **67**(5), 050801–050848 (2015).

Candes, E. J., Romberg, J., and Tao, T., "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**(2), 489–509 (2006).

Donoho, D. L., "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).

Gordeyev, S., De Lucca, N., Jumper, E., Hird, K., Juliano, T. J., Gregory, J. W., Thordahl, J., and Wittich, D. J., "Comparison of unsteady pressure fields on turrets with different surface features using pressure-sensitive paint," Exp. Fluids **55**, 1661 (2014).

Gregory, J. W., Asai, K., Kameda, M., Liu, T., and Sullivan, J. P., "A review of pressure-sensitive paint for high-speed and unsteady aerodynamics," Proc. Inst. Mech. Eng. G. J. Aerosp. Eng. **222**, 249–290 (2008).

Gregory, J. W., Sakaue, H., Liu, T., and Sullivan, J. P., "Fast pressure-sensitive paint for flow and acoustic diagnostics," Annu. Rev. Fluid Mech. **56**, 303–330 (2014).

Kutz, J. N., "Deep learning in fluid dynamics," J. Fluid Mech. **814**, 1–4 (2017).

Ling, J., Andrew, K., and Jeremy, T., "Reynolds averaged turbulence modelling using deep neural networks with embedded invariance," J. Fluid Mech. **807**, 155–166 (2016).

Liu, T., "Pressure-correction method for low-speed pressure-sensitive paint measurements," AIAA J. **41**, 906–911 (2003).

Liu, T., Guille, M., and Sullivan, J. P., "Accuracy of pressure-sensitive paint," AIAA J. **39**, 103–112 (2001).

Lumley, J. L., "The structure of inhomogeneous turbulent flow," in *Atmospheric Turbulence and Radio Wave Propagation* (Nauka, Moscow, 1967), pp. 166–178.

Mainini, L. and Willcox, K., "A surrogate modeling approach to support real-time structural assessment and decision-making," AIAA J. **53**(6), 1612–1626 (2015).

Manohar, K., Brunton, S. L., and Kutz, J. N., "Environment identification in flight using sparse approximation of wing strain," J. Fluids Struct. **70**, 162 (2016).

Pastuhoff, M., Yorita, D., Asai, K., and Alfredsson, P. H., "Enhancing the signal-to-noise ratio of pressure sensitive paint data by singular value decomposition," Meas. Sci. Technol. **24**, 075301 (2013).

Peng, D., Wang, S., and Liu, Y., "Fast PSP measurements of wall-pressure fluctuation in low-speed flows: Improvements using proper orthogonal decomposition," Exp. Fluids **57**, 45 (2016).

Ruscher, C. J., Dannenhoffer III, J. F., and Glauser, M. N., "Repairing occluded data for a mach 0.6 jet via data fusion," AIAA J. **55**(1), 255–264 (2017).

Seena, A. and Sung, H. J., "Dynamic mode decomposition of turbulent cavity flows for self-sustained oscillations," Int. J. Heat Fluid Flow **32**, 1098–1110 (2011).

Sirovich, L. and Kirby, M., "Low-dimensional procedure for the characterization of human faces," J. Opt. Soc. Am. A **4**(3), 519–524 (1987).

Wen, X., Tang, H., and Duan, F., "Interaction of in-line twin synthetic jets with a separated flow," Phys. Fluids **28**(4), 043602 (2016).

Wen, X., Li, Z., Peng, D., Zhou, W., and Liu, Y., "Missing data recovery using data fusion of incomplete complementary measurement sets: A PIV application," Exp. Fluids (unpublished).

Zhang, Q. S., Liu, Y. Z., and Wang, S. F., "The identification of coherent structures using proper orthogonal decomposition and dynamic mode decomposition," J. Fluids Struct. **49**, 53–72 (2014).