RESEARCH ARTICLE | OCTOBER 18 2023

Information sharing-based multivariate proper orthogonal decomposition

Zihao Wang (王子豪) ; Guiyong Zhang (张桂勇) ; Teizhi Sun (孙铁志) ; Huakun Huang (黄华坤)

Physics of Fluids 35, 104108 (2023)
https://doi.org/10.1063/5.0169994

Physics of Fluids ARTICLE pubs.aip.org/aip/pof

Information sharing-based multivariate proper orthogonal decomposition

Cite as: Phys. Fluids **35**, 104108 (2023); doi: 10.1063/5.0169994
Submitted: 31 July 2023 · Accepted: 23 September 2023 ·
Published Online: 18 October 2023

Zihao Wang (王子豪),[1] Guiyong Zhang (张桂勇),[1,2,a)] Teizhi Sun (孙铁志),[1] and Huakun Huang (黄华坤)[3]

AFFILIATIONS

[1]State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, School of Naval Architecture Engineering, Dalian University of Technology, Dalian 116024, China
[2]Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration, Shanghai 200240, China
[3]Department of Aeronautical and Aviation Engineering, the Hong Kong Polytechnic University, Kowloon, Hong Kong

[a)]Author to whom correspondence should be addressed: gyzhang@dlut.edu.cn

ABSTRACT

This study explores challenges in multivariate modal decomposition for various flow scenarios, emphasizing the problem of inconsistent physical modes in Proper Orthogonal Decomposition (POD). This inconsistency arises due to POD's inability to capture inter-variable relationships and common flow patterns, resulting in a loss of phase information. To address this issue, the study introduces two novel data-driven modal analysis methods, collectively called Information Sharing-Based Multivariate POD (IMPOD). These methods, namely, Shared Space Information Multivariate POD (SIMPOD) and Shared Time Information Multivariate POD (TIMPOD), aim to regularize modal decomposition by promoting information sharing among variables. TIMPOD, which assumes shared time information, successfully aligns multivariate modes and corrects their phases without significantly affecting reconstruction error, making it a promising corrective technique for multivariate modal decomposition. In contrast, SIMPOD, which assumes shared space information, reorders modes and may lead to a loss of meaningful insight and reconstruction error.

I. INTRODUCTION

Analyzing high-dimensional spatiotemporal data can be considered a challenge in modern fluid dynamics.[1] The development of data-driven methods provides a pathway for tackling this problem comprehensively.[2,3] Dimensionality reduction methods, as an important task in machine learning, have emerged as promising approaches for analyzing fluid dynamics data and have been widely applied in the field. Representative methods include Proper Orthogonal Decomposition (POD) and Dynamic Mode Decomposition (DMD). Lumley[4] introduced POD to the field of fluid dynamics, where the POD modes, arranged by energy (eigenvalues) in descending order, can effectively describe and analyze the physical characteristics of the flow field by extracting the dominant modes with higher energy. DMD, proposed and applied to fluid dynamics by Schmid (based on Koopman operator theory),[5] can analyze the frequency, growth rate, and spatial structure of flow field features. It has gradually developed into a subfield of data processing known as data-driven modal analysis methods.[6] Modal analysis has played a significant role in both theoretical analysis and engineering applications.[7,8]

POD has been a cornerstone of modal decomposition techniques, mainly due to its orthogonality properties.[9,10] POD decomposes spatiotemporal data of the flow field into spatial modes and temporal coefficients. The spatial modes are visualized as coherent structures, where mutually orthogonal modes represent the fundamental dynamic features of the flow field that are unrelated to each other.[11,12] The mutually orthogonal temporal coefficients construct the state space of the flow field, and the point trajectories in the state space are regarded as low-dimensional representations of the temporal features of the flow field.[13] To address some limitations of standard POD analysis and expand its applicability, many POD-based variants have been studied. Snapshot POD overcomes the computational challenge when the spatial dimension is much larger than the temporal dimension, resulting in faster computation and improved efficiency.[14] Balanced POD (BPOD) is similar to balanced truncation techniques, finding a balance between controllability and observability and utilizing the obtained adjoint modes to capture highly observable small energy perturbations.[15] Spectral POD (SPOD) extends the acquisition of coherent structures in the time-scale.[16,17] Multi-scale POD (MPOD) combines multi-resolution techniques to differentiate potential modes on fast and slow scales.[18]

The state of a flow field is described using multivariate features such as velocity, vorticity, and pressure.[19] Performing modal

decomposition separately for each variable or selecting representative modes for each variable is commonly regarded as representing the fundamental patterns of the flow field in fluid dynamics.[20,21] Recent studies have shown that when multivariate variables are individually subjected to modal decomposition, the resulting flow field modes fail to collectively express the same physical behavior, and this issue becomes more pronounced in higher-order flow field patterns.[22,23] The reason behind this is that POD modes are computed optimally based on the L2 norm.[24] The POD decomposition is optimal for this univariate data, but it may not be the most suitable for multivariate physical spaces. In the individual modal decomposition process, stronger univariate features are introduced for each variable, which deviates from the complex physical characteristics of the multivariate flow field description.

To address the modal decomposition of multivariate features, an important approach is provided by Extended Proper Orthogonal Decomposition (EPOD).[25] It employs the same temporal basis to study the correlation between two variables and finds the modes of the second variable by solving a generalized eigenvalue problem with the first variable as the common variable. While valuable for investigating multivariate correlations, this method does not achieve a proper modal decomposition of multivariate flow fields. Joint Proper Orthogonal Decomposition (JPOD) presents a concept for simultaneously decomposing modes across multiple datasets. It involves the joint solution of POD for datasets from different operating conditions, examining their shared contributing characteristics through integral equations.[26] Brunton and Kutz suggested stacking multiple vector datasets together for modal decomposition when dealing with multiple datasets.[27] Csala *et al.* further explored various arrangements of datasets to address the issue of multivariate modal decomposition.[28]

The prior work of researchers has guided our approach, and this paper draws inspiration from the effect of data arrangement on multivariate modal decomposition. In this study, this is referred to as Information-Sharing-Based Multivariate Proper Orthogonal Decomposition (IMPOD). This method utilizes multivariate information to regularize the modal decomposition results of individual variables in the flow field. IMPOD is built upon the foundation of the POD method based on Singular Value Decomposition (SVD). SVD is a matrix decomposition technique used for rectangular matrices, and the technique of determining POD modes based on SVD is robust to rounding errors.[29] The spatial and temporal patterns of the POD modal decomposition are reflected in the left decomposition matrix (product of the left singular value matrix and the singular value matrix) and the right decomposition matrix (right singular value matrix) of the SVD. To facilitate the integration of multivariate information, two strategies are employed by sharing the left and right decomposition matrices, respectively.

These methods are referred to as Independent Modal POD with Shared Space Information (SIMPOD) and Independent Modal POD with Shared Time Information (TIMPOD). Three typical flow problems, namely, transient wake behind a cylinder, low Reynolds number pitching airfoil, and turbulent wake behind a cylinder, are employed to illustrate the common issue of multivariate POD methods and to evaluate the rationality of the two IMPOD methods in revealing transient and non-stationary dynamics of complex fluid flows.

The remainder of this paper is organized as follows: Section II introduces the definition of POD, emphasizes the relationship between IMPOD and POD, and presents the methods used for multivariate flow field modal decomposition, namely, SIMPOD and TIMPOD. Section III provides detailed explanations of the three test cases used for evaluating the algorithms. Section IV presents the results to reveal the explanations and universality of the defects in multivariate POD methods and evaluates the performance improvements of the two IMPOD methods. Finally, the conclusion in Sec. V summarizes the main findings and outlines future work.

II. METHODOLOGY

A. Singular value decomposition-based proper orthogonal decomposition

The core idea of the Proper Orthogonal Decomposition (POD) method is to find a set of optimal orthogonal basis functions that best represent a given set of spatial data samples in a least squares sense. This allows for a low-dimensional approximation of the original data using a reduced number of orthogonal basis functions. The mathematical principles of the POD method are as follows.

First, the original flow field is discretized by placing monitoring points, resulting in a data matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$,

$$\boldsymbol{A} = \begin{bmatrix} W_1^{(1)} & W_1^{(2)} & \dots & W_1^{(n)} \\ W_2^{(1)} & W_2^{(2)} & \dots & W_2^{(n)} \\ \vdots & \vdots & & \vdots \\ W_m^{(1)} & W_m^{(2)} & \dots & W_m^{(n)} \end{bmatrix}. \tag{1}$$

Here, n represents the number of flow field snapshots, m represents the number of spatial discretization points, and W represents the flow field variable values at the grid nodes.

It is worth noting that in some literature, the time average of the data at each spatial location is subtracted from each row as the target for singular value decomposition. Both approaches yield valid results,

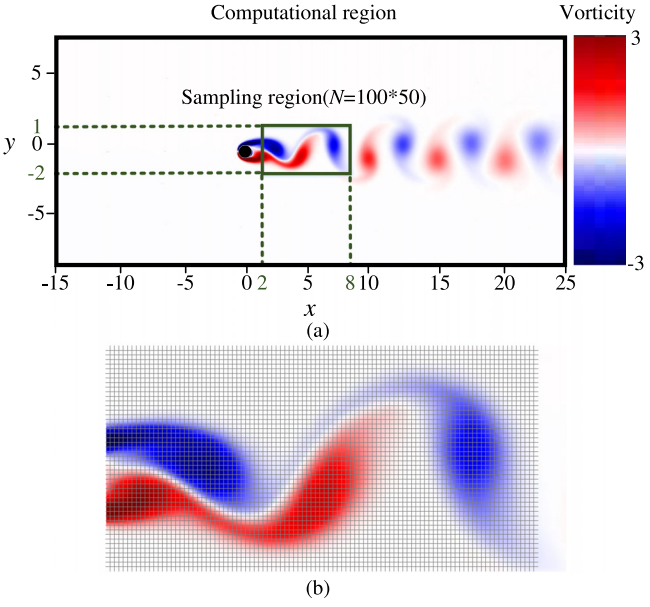

FIG. 1. The vorticity field displays (a) the distribution of the computational region and the sampling region, and (b) the grid details in the wake region.

but the interpretation of the results may differ.[24] When obtaining the state space of a flow field, the temporal mean is one of the fundamental components of the state space. The construction of the state space is based on the principle of orthogonality, which means that each mode should be orthogonal, i.e., they should have no linear correlation with each other. In order to incorporate the temporal mean into the orthogonal state space, we can consider it as one of the modes and include it in the analysis. Therefore, this study suggests using the original data for SVD. This technique allows us to decompose the data into orthogonal modes, including the temporal mean as one of the modes. By doing so, we can obtain a state space representation that captures both the spatial patterns and the temporal mean of the flow field.

In this study, the snapshot matrix $\boldsymbol{A}$ is subjected to singular value decomposition. The decomposition is expressed as

$$\boldsymbol{A} = \boldsymbol{U}^{*}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{\Phi}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{\Phi}\boldsymbol{a}. \tag{2}$$

Here, $\boldsymbol{U}^{*} \in \mathbb{R}^{m\times r}$ is the left singular value matrix, and $\boldsymbol{V} \in \mathbb{R}^{n\times r}$ is the right singular value matrix. They satisfy $\boldsymbol{U}^{*\mathrm{H}}\boldsymbol{U}^{*} = \boldsymbol{I}$ and $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ is the diagonal matrix of singular values, with the diagonal elements arranged in descending order. During the singular value decomposition process, one can choose to retain only the main singular values (truncating the rest) to achieve dimensionality reduction. In the context of the POD decomposition, the left decomposition matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times r}$ denotes the spatial modes, which capture the spatial distribution and patterns of the flow field. On the other hand, the right decomposition matrix $\boldsymbol{a} \in \mathbb{R}^{r\times n}$ represents the modal coefficients, which describe the temporal variations of the flow field at different time snapshots.

Thus, the snapshot matrix can be represented as a linear combination of the modal orders. The process of singular value decomposition can be equivalently formulated as an optimization problem of minimizing the L2 norm error,

$$\min_{\boldsymbol{\Phi a}} \| \boldsymbol{A} - \boldsymbol{\Phi a} \|_2^2 . \tag{3}$$

By retaining a subset of the POD modes, the original system can be effectively represented. Typically, the appropriate number of POD modes is selected based on the energy ratio criterion, which can be expressed as

$$\frac{\sum_{i=1}^{k}\sigma_i^2}{\sum_{i=1}^{r}\sigma_i^2} \geqslant \varepsilon. \tag{4}$$

Here, k represents the retained number of POD modes (the truncated value of SVD), and ε represents the chosen energy ratio.

The quality of the algorithm can be evaluated using the relative reconstruction error based on the Frobenius criterion. The relative reconstruction error based on the Frobenius criterion effectively eliminates the influence of data scale. Mathematically, it can be expressed as the ratio of the Frobenius norm of the difference between the original data matrix and the reconstructed data matrix to the Frobenius norm of the original data matrix. The specific formula is as follows:

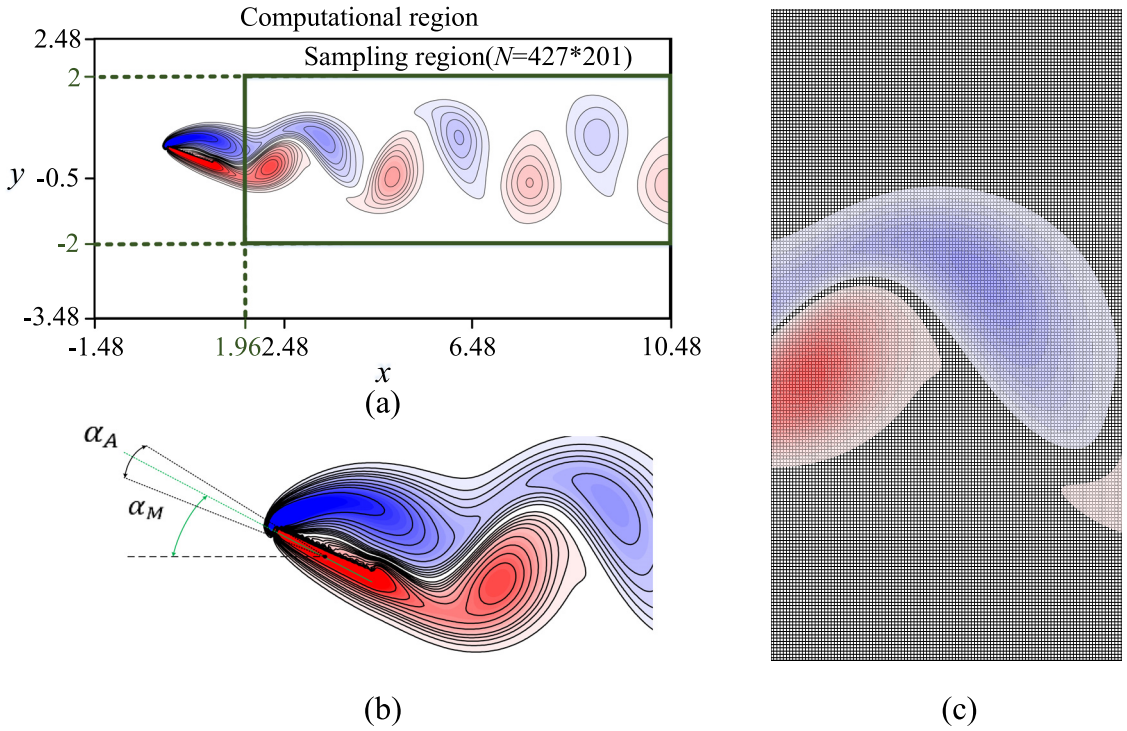

FIG. 2. The vorticity field displays (a) the distribution of the computational region and the sampling region, (b) the kinematics of the airfoil's pitching motion, and (c) the grid details in the wake region.

$$\varepsilon_r(\mathrm{X}, \mathbf{X}_r) = \frac{\|\mathrm{X} - \mathbf{X}_r\|_F}{\|\mathrm{X}\|_F}, \tag{5}$$

where $\mathbf{X}$ represents the original data, and $\mathbf{X}_r$ represents the reconstructed data using the DR algorithm.

B. Spatial information sharing-based multivariate proper orthogonal decomposition

The implementation of Spatial Information Sharing-Based Multivariate Proper Orthogonal Decomposition (SIMPOD) relies on sharing spatial information through the singular value decomposition of the left decomposition matrix. By leveraging shared spatial information, SIMPOD aims to achieve a modal decomposition process that yields temporally coherent modes with physical correlations. Assuming the presence of N variables' data matrices within the same flow field, the mathematical principle behind the shared left decomposition matrix's singular value decomposition is as follows:

$$\left[\mathbf{A}^1, \mathbf{A}^2, \ldots, \mathbf{A}^N\right] = \mathbf{\Phi}\left[\boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^N\right]. \tag{6}$$

This is achieved by performing SVD on the stacked column vectors to obtain a joint matrix. In this case, $\mathbf{\Phi}$ represents the shared left decomposition matrix (spatial modes), and $\boldsymbol{a}^n$, $n = 1, 2, \ldots, N$, denote the temporal modes for each variable.

The optimization problem for SIMPOD can be formulated as follows:

$$\min_{\mathbf{\Phi}\{\boldsymbol{a}^n\}_{n=1}^N} \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{A}^n - \mathbf{\Phi}\boldsymbol{a}^n\|_2^2. \tag{7}$$

In this formulation, we aim to minimize the average L2-norm error between the individual data matrices $\mathbf{A}^n$ (corresponding to each

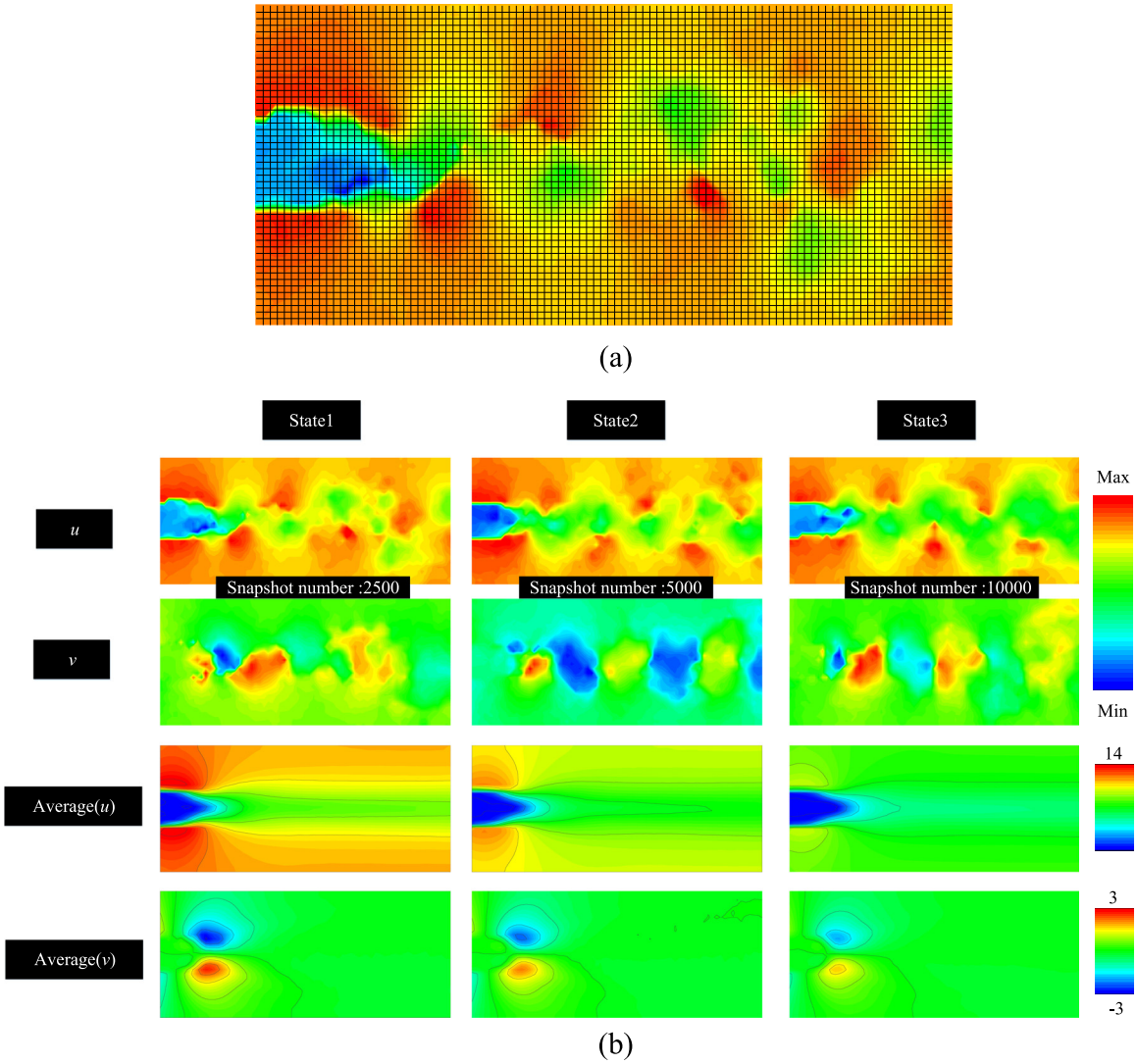

FIG. 3. (a) The grid details in the wake region and (b) the instantaneous snapshots of the three states along with their corresponding average flow fields.

variable) and their respective reconstructions using the shared left decomposition matrix $\mathbf{\Phi}$ and the corresponding temporal modes $\boldsymbol{a}^n$. By taking the average over all variables (N), we seek a solution that optimally captures the shared spatial information while preserving the specific temporal behavior of each variable. The objective is to find the optimal combination of $\mathbf{\Phi}$ and the temporal modes that best approximate the original data for all variables simultaneously.

C. Temporal information sharing-based multivariate proper orthogonal decomposition

The implementation of Time Information Sharing-based Multivariate POD (TIMPOD) relies on the shared singular value decomposition of the right decomposition matrix to incorporate time information sharing. By utilizing shared time information, TIMPOD aims to regularize the modal decomposition process of POD to obtain spatial modal features with physical correlations. The mathematical principle behind the shared singular value decomposition of the right decomposition matrix is as follows:

$$\begin{bmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \cdots \\ \mathbf{A}^N \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}^1 \\ \mathbf{\Phi}^2 \\ \cdots \\ \mathbf{\Phi}^N \end{bmatrix} \boldsymbol{a}. \tag{8}$$

This is achieved by performing SVD on the stacked row vectors of the joint matrix. Here, $\mathbf{\Phi}^n$, $n = 1, 2, \ldots, N$, represent the spatial modes of each variable, and a is the shared left decomposition matrix (temporal modes). The optimization problem of TIMPOD can be formulated as follows:

$$\min_{\{\mathbf{\Phi}^n\}_{n=1}^N, \boldsymbol{a}} \frac{1}{N} \sum_{n=1}^{N} \| \mathbf{A}^n - \mathbf{\Phi}^n \boldsymbol{a} \|_2^2. \tag{9}$$

This objective function minimizes the average L2 norm of the differences between each variable $\mathbf{A}^n$ and its corresponding spatial mode $\mathbf{\Phi}^n$ multiplied by the shared temporal modes a. The goal is to find the optimal spatial modes and shared temporal modes that best approximate the original multivariate data in terms of the L2 norm.

III. FLOW DATASETS

A. Transient cylinder wake direct numerical simulations

The flow dynamics of flow past a circular cylinder under low Reynolds number conditions exhibit periodic vortex shedding, making it a typical benchmark problem for unsteady fluid flow. It has been widely used for the development of low-dimensional models. The database used in this study is derived from Ref. 30, where the training data are obtained through direct numerical simulations (DNS) using high-order spectral/hp element method. The Reynolds number (based on the free-stream velocity $U = 1$, diameter of the cylinder $D = 1$, and kinematic viscosity $\mu = 0.01$) is set to 100 to ensure the onset of vortex shedding. The size of the sampling region is $L_x = 11.98$ ($-15 \le x \le 25$) and $L_y = 5.96$ ($-8 \le y \le 8$), which correspond to the streamwise (x) and transverse (y) directions, respectively. The center of the cylinder is

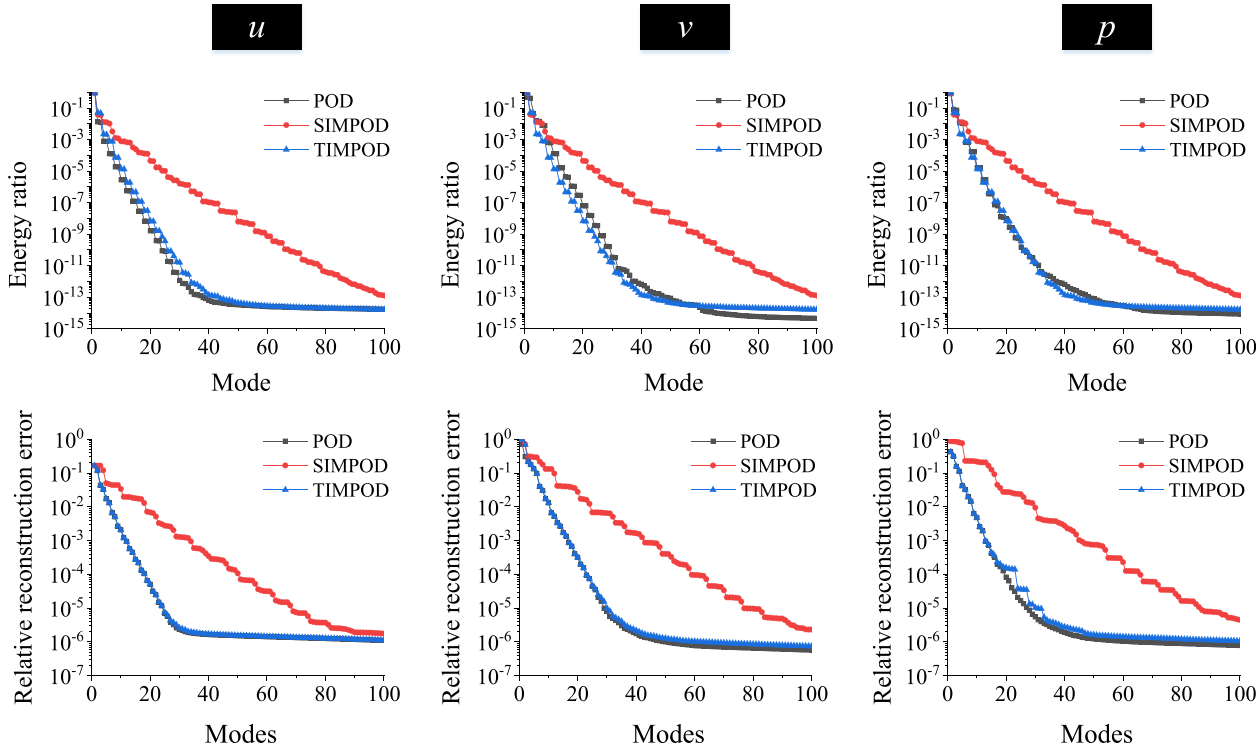

FIG. 4. Energy ratio and relative reconstruction error for the first 100 modes of transient cylinder wake.

located at $(x, y) = (0, 0)$. In this study, the focus is on the flow behind the cylinder. Therefore, a partial sampling region is extracted, specifically $2 \le x \le 8$ and $-2 \le y \le 1$, as depicted in Fig. 1(a). A Cartesian grid system is used with grid spacing $\Delta x = 0.07$ and $\Delta y = 0.08$ within the sampling region. The number of grid points in the sampling region is $(N_x, N_y) = (100, 50)$, as depicted in Fig. 1(b). The time interval between flow field data is $\Delta t = 0.01$. The dataset consists of 200 snapshots of the flow field. The velocity components, namely, the X-component (u), Y-component (v), and pressure (p), can be chosen as the multivariate variables for this dataset. These variables capture important characteristics of fluid flow fields and are commonly used in fluid dynamics analysis.

Multivariate data often exhibit differences in units or dimensions. Data preprocessing techniques, such as standardization, can address this issue. However, these data preprocessing methods may result in the equal weighting of flow field information represented by multivariate data. For fundamental variables like velocity and pressure, standardization can be applied because their importance is relatively consistent. For variable components, such as the Y-direction component of velocity and pressure, standardization is not suitable, as their contributions to the flow field vary significantly in importance. This paper recommends using dimensionless scaling to address the unit disparities. The specific approach to dimensionless scaling is outlined as follows:

$$\boldsymbol{D} = \frac{\boldsymbol{D}^*}{D^*}, \tag{10}$$

$$\boldsymbol{x} = \frac{\boldsymbol{x}^*}{D^*}, \tag{11}$$

$$\boldsymbol{y} = \frac{\boldsymbol{y}^*}{D^*}, \tag{12}$$

$$\boldsymbol{U} = \frac{\boldsymbol{U}^*}{U^*}, \tag{13}$$

$$\boldsymbol{u} = \frac{\boldsymbol{u}^*}{U^*}, \tag{14}$$

$$\boldsymbol{v} = \frac{\boldsymbol{v}^*}{U^*}, \tag{15}$$

$$\boldsymbol{p} = \frac{\boldsymbol{p}^* - p_\infty}{0.5\rho^* U^* U^*}, \tag{16}$$

$$t = \frac{t^* U^*}{D^*}. \tag{17}$$

Here, the superscript $*$ refers to the dimensional quantity.

B. Pitching airfoil direct numerical simulations

The sinusoidal pitching motion of a flat plate airfoil at low Reynolds numbers and high angles of attack serves as the fluid

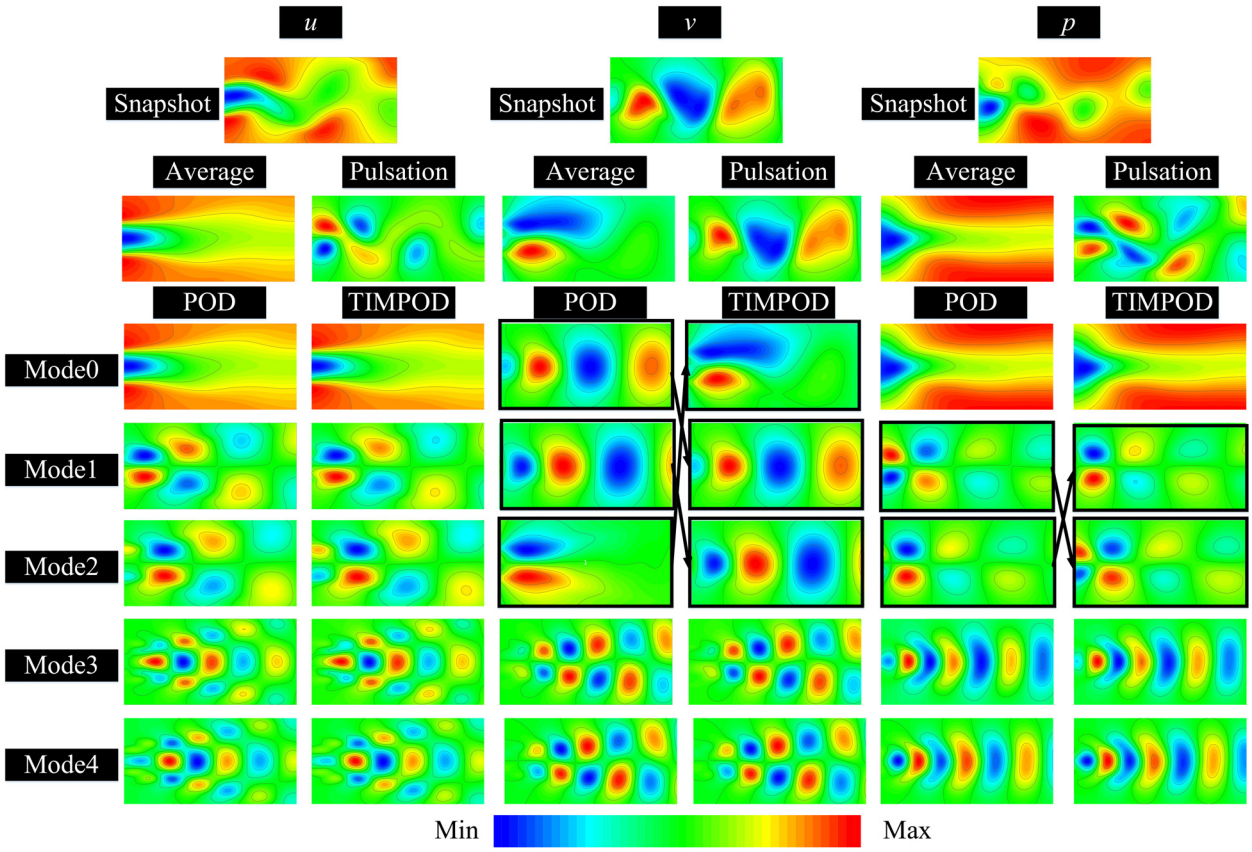

FIG. 5. Mode distribution characteristics of POD and TIMPOD, along with representative flow field features of transient cylinder wake.

dataset for this study. The interaction between the potential occurrence of periodic vortex shedding and the imposed sinusoidal pitching motion is a problem of interest in classical aerodynamics theory. The database used in this study is derived from Ref. 31 (for further details on the characterization of this flow, refer to Refs. 32 and 33). This dataset is obtained through two-dimensional DNS. The governing equations, which include the incompressible continuity and Navier–Stokes equations, are solved using the Immersed Boundary Projection method as described in Refs. 34 and 35,

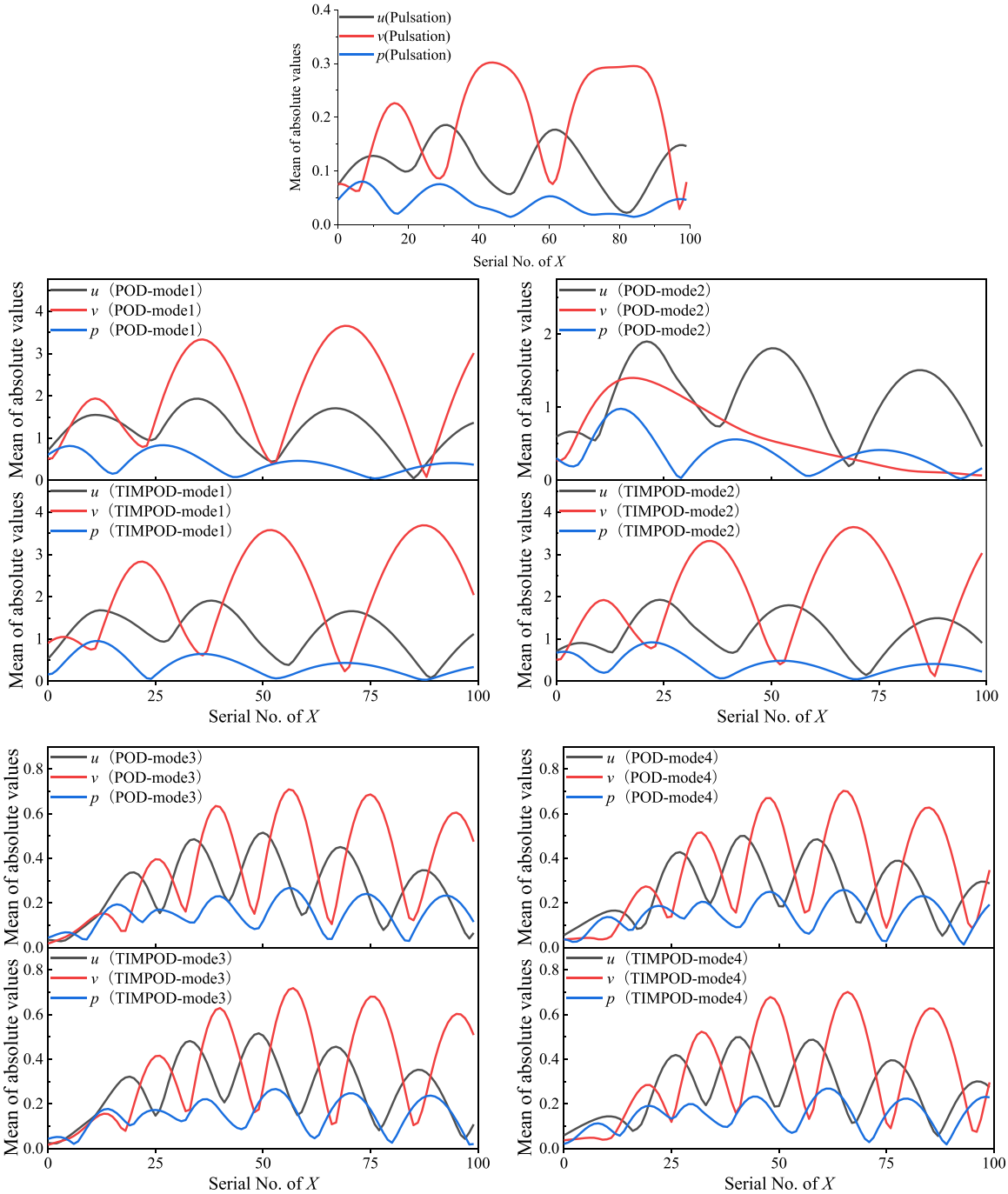

FIG. 6. Pulsation characteristics of different variable flow field snapshots of transient cylinder wake and the distribution of mean values of the X axis projections of the modes.

$$\nabla \cdot \boldsymbol{u} = 0, \tag{18}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} = -\nabla \cdot (\boldsymbol{u}\boldsymbol{u}) - \nabla p + \frac{1}{Re}\nabla^2 \boldsymbol{u}, \tag{19}$$

where $\boldsymbol{u}$ and p denote the velocity vector and pressure, respectively. The Reynolds number (based on the free-stream velocity U and chord length of the airfoil c) is set to $Re = 100$. According to the description in Ref. 29, flow separation occurs and remains stable on the upper surface of the airfoil at an angle of attack (α) of approximately 20°. Beyond the critical angle of attack (α_c) of approximately 27°, the steady-state solution becomes unstable, and periodic vortex shedding is observed.

In this study, the baseline angle of attack (α_M) of the airfoil is set to 25°. The kinematics of the airfoil are described by the following equation:

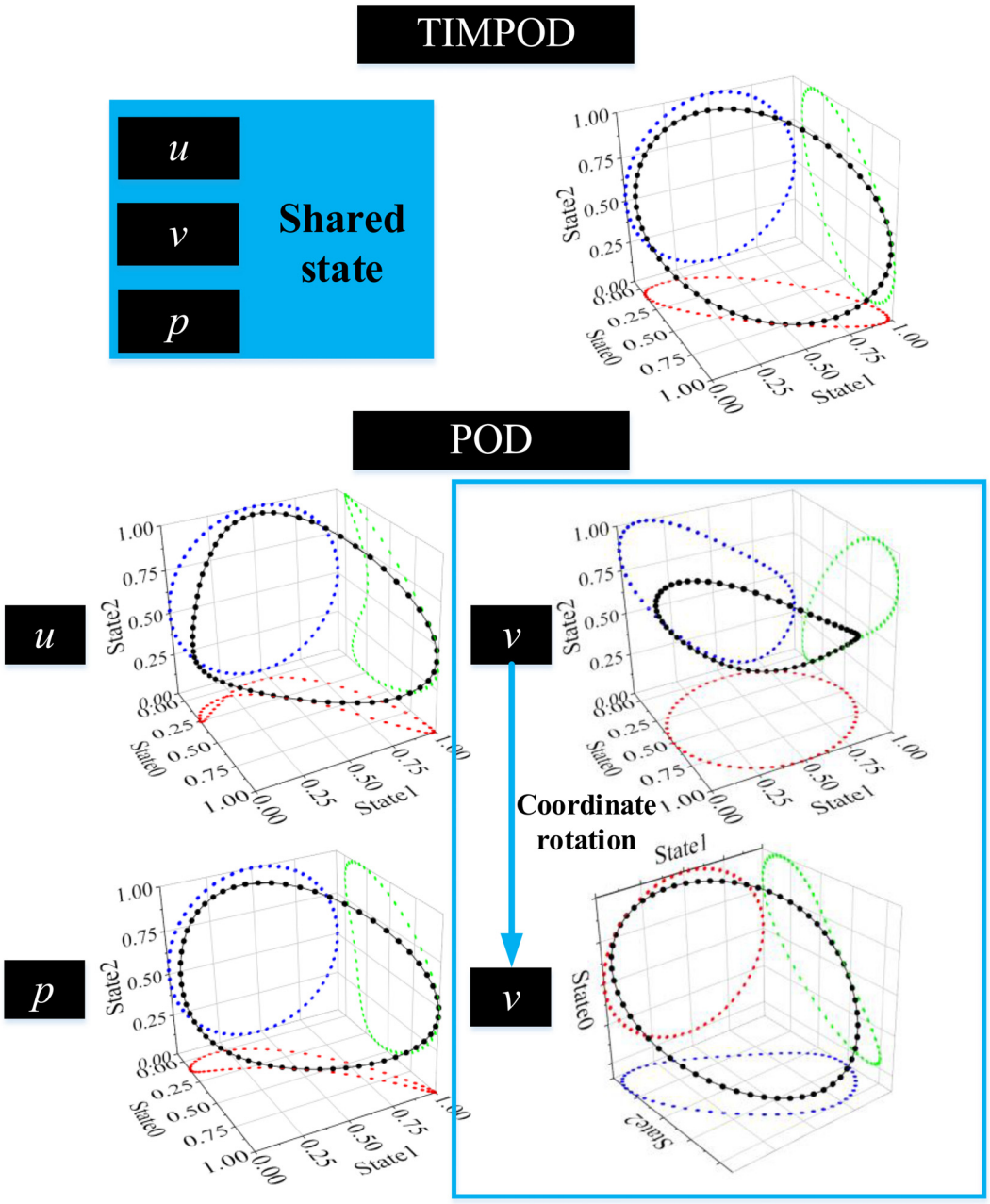

FIG. 7. Comparisons of POD with TIMPOD temporal trajectories within the state space at different variables of transient cylinder wake.

$$\alpha(t) = \alpha_M + \alpha_A \sin(2\pi f^* t), \tag{20}$$

where $\alpha(t)$ represents the time-varying angle of attack, α_M is the baseline angle of attack, α_A is the amplitude of the sinusoidal variation, where $f^* = fc/U$ is a dimensionless frequency, and t is the time. The boundary conditions for the airfoil are implemented by applying body forces along the moving surface of the airfoil. The size of the sampling region for the data is $L_x = 11.98$ $(-1.48 \le x \le 10.48)$ and $L_y = 5.96$ $(-3.48 \le y \le 2.48)$ in the streamwise (x) and transverse (y) directions, respectively. The midpoint of the airfoil's chord length is located at $(x, y) = (0.5,\ 0)$. A Cartesian grid system with a grid spacing of $\Delta x = \Delta y = 0.02$ is used within the sampling region. The number of grid points within the sampling region is $(N_x, N_y) = (599, 298)$. In this study, the focus is on the flow at the trailing edge of the flat plate airfoil. Therefore, a portion of the sampling region is extracted, specifically $1.96 \le x \le 10.48$ and $-2 \le y \le 2$, as depicted in Fig. 2(a). The time interval for the flow field data is $\Delta t = 0.01c/U$. The dataset contains 401 flow field snapshots, corresponding to 40 convective time units. The velocity components, namely, the X-component (u) and Y-component (v), can be chosen as the multivariate variables for this dataset.

This dataset considers a pitching amplitude of 5 deg. The pitching motion is performed about the mid-chord of the airfoil, as illustrated in Fig. 2(b). The dimensionless pitching frequency, f^*, takes on the values {0.3}. The number of grid points used for POD is $(N_x, N_y) = (427, 201)$, as depicted in Fig. 2(c). This dataset has undergone dimensionless scaling, and the specific method can be referred to in Eqs. (10)–(17), replacing the cylinder in the equations with the chord length.

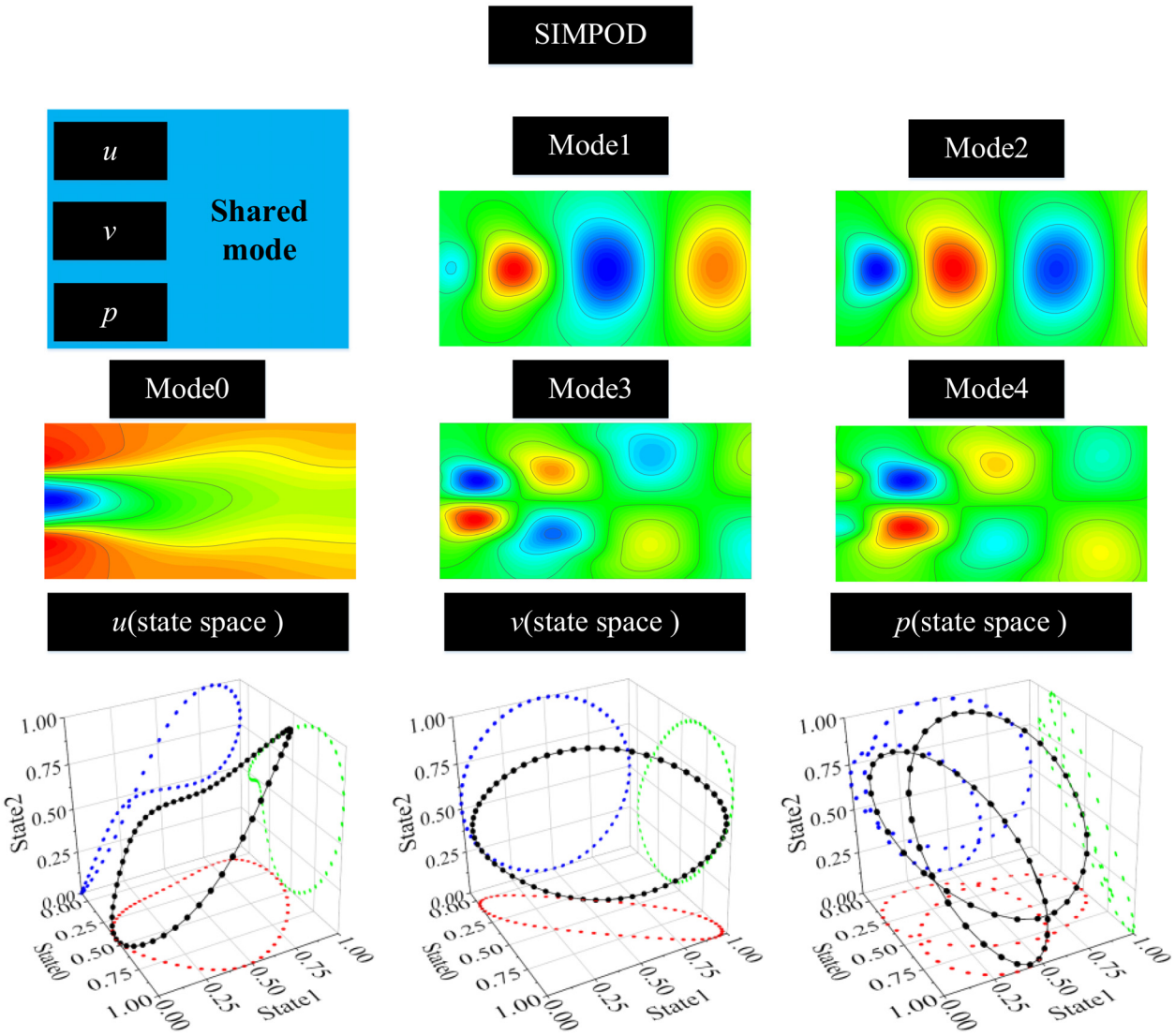

FIG. 8. Temporal trajectories of the shared flow field modes and the state space for different variables under SIMPOD of transient cylinder wake.

C. Turbulent flow experiment through a cylinder under multiple transient conditions

Let us consider a more complex real flow field obtained from experiments conducted in the von Karman Institute's low-speed wind tunnel. The data include numerical noise that occurred during the experimental process. The study focuses on the turbulent flow around a cylindrical object with a diameter (D) of 5 mm and a length (L) of 20 cm. The flow is inherently three-dimensional, but representative two-dimensional flow fields are obtained using Time-Resolved Particle Image Velocimetry (TR-PIV); a Cartesian grid system with a grid spacing of $\Delta x = \Delta y = 0.876$ mm is used within the sampling region. The number of grid points used for POD is $(N_x, N_y) = (71, 30)$, as depicted in Fig. 3(a).

The dataset is sourced from Ref. 36, with specific experimental setups and related processing detailed in Mendez *et al.*[37] The flow process encompasses three different states based on varying Reynolds numbers: two steady states ($Re = [2600, 4000]$) and the transitional state between them ($Re = 4000$ transitioning to $Re = 2600$). The change in the Reynolds number is achieved by adjusting the free-stream velocity. The dataset comprises a total of 13 200 snapshots, involving the velocity components of the X axis (u) and Y axis (v). The first 4000 snapshots cover the first steady state ($Re = 4000$). The subsequent 3000 snapshots capture the transitional flow state, while the remaining snapshots represent the second steady state ($Re = 2600$). Figure 3(b) illustrates representative instantaneous snapshots of the three states along with their corresponding average flow fields. The velocity transitions from a high speed in state 1 to a low speed in state 3. Both state 1 and state 3 exhibit stable states characterized by vortex shedding structures at a specific frequency behind the cylinder. State 2 showcases a more complex vortex shedding structure behind the cylinder. This dataset has not undergone dimensionless scaling.

IV. RESULTS AND DISCUSSION

A. The transient cylinder wake case

The energy ratio of each mode is often considered as *a priori* information representing the proportion of the original flow field information. It is used as a criterion for mode selection. The reconstruction error, on the other hand, serves as a posterior criterion to measure the proportion of the original flow field information truly represented by the reduced modes. Figure 4 displays the energy ratio and relative reconstruction error for the first 100 modes of flow over a cylinder, presented in a logarithmic scale. It is important to note that in the figures of this paper, "Modes" represents the first N modes, while Mode refers to the Nth mode. For each variable, the energy ratio tends to decrease with an increasing number of modes, and different algorithms exhibit variations in this behavior.

POD and TIMPOD show minor differences in the energy ratio trend. SIMPOD, on the other hand, exhibits significant differences compared to the other two algorithms, with a slower decay in the energy ratio. This suggests that TIMPOD has a relatively poor

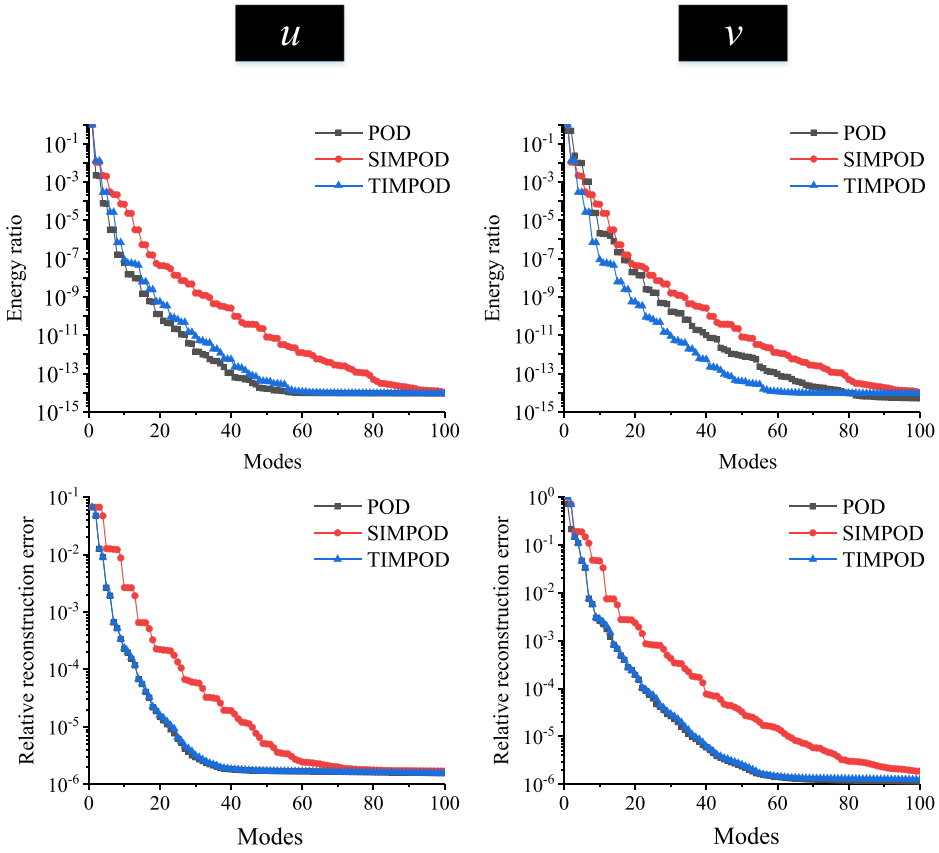

FIG. 9. Energy ratio and relative reconstruction error for the first 100 modes of flow over the pitching airfoil.

dimension reduction effect, as the data information is not concentrated in the lower-order modes. Further observation of the relative reconstruction error for each variable reveals that the error decreases with an increasing number of modes. POD and TIMPOD exhibit almost identical reconstruction errors, indicating that TIMPOD does not significantly affect the preservation of the original data. SIMPOD, however, significantly reduces the preservation of the original data. The consistent behavior between the energy ratio and relative reconstruction error indicates that the modes with higher energy ratios indeed capture a larger portion of the original flow field information. Therefore, using the energy ratio as a criterion for mode selection provides a reasonable approach to effectively represent the dominant flow features.

The left decomposition matrix obtained from singular value decomposition represents interpretable spatial patterns, which aid in understanding the underlying flow structures. Figure 5 illustrates the mode distribution characteristics of POD and TIMPOD, along with representative flow field features. For each method, five representative low-order modes are selected. Higher-order modes are deemed less important, so only the first five modes are visualized, with the sorting based on the energy ratio. The characteristics of modal distribution depend on the spatial distribution of positive and negative coherent structures.

Examining the modal distribution structure of POD, which has been extensively studied for this cylinder flow case, previous research

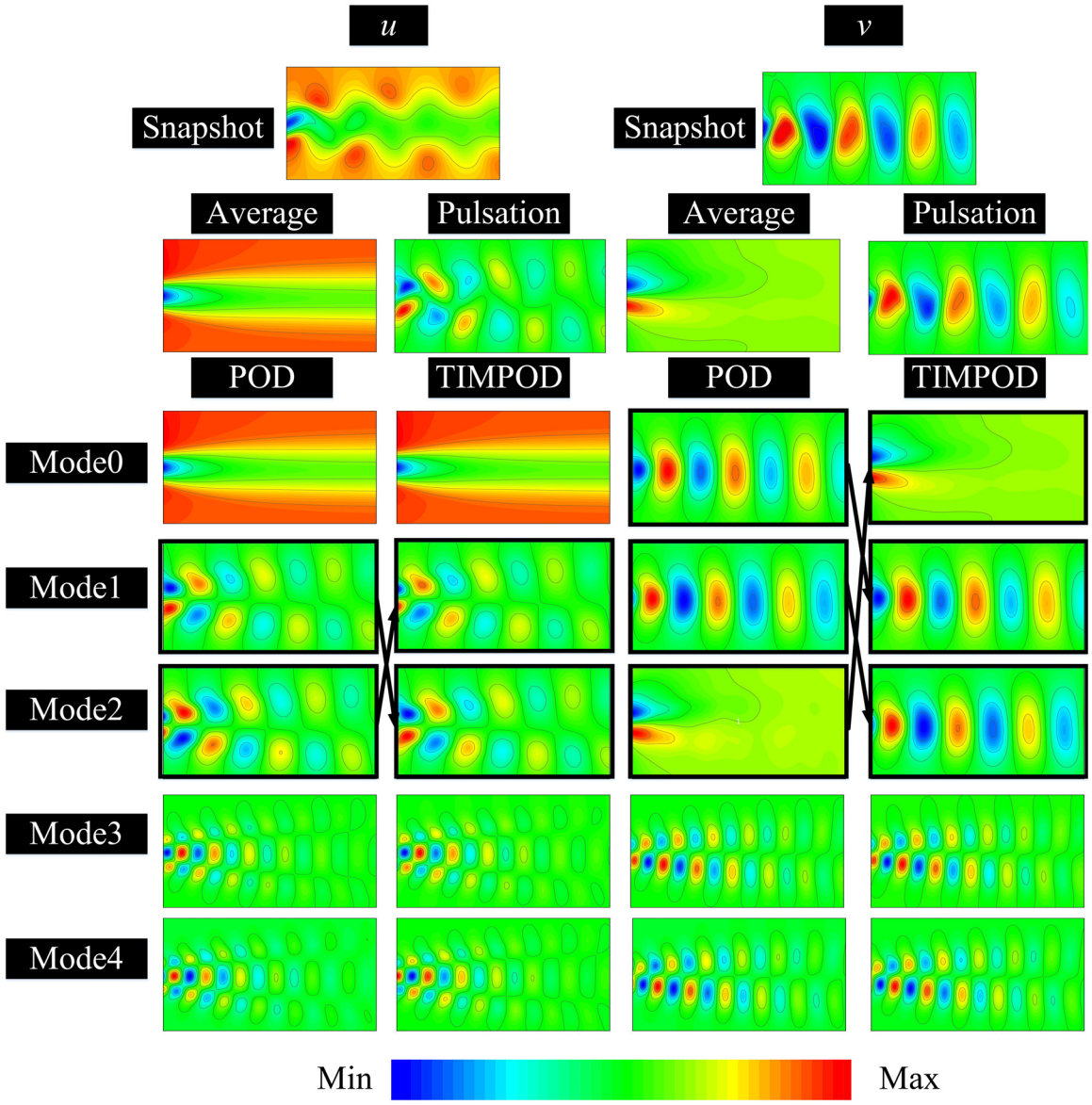

FIG. 10. Mode distribution characteristics of POD and TIMPOD, along with representative flow field features of the pitching airfoil.

has often summarized the fundamental patterns using the representative variable u as follows: Mode 0 represents the time-averaged flow field, which serves as the fundamental skeleton and the steady-state of the flow field. Modes 1 and 2 appear in pairs, exhibiting similar structures, representing the pulsation patterns associated with vortex shedding. Modes 3 and 4 also appear in pairs, demonstrating similar structures but with smaller-scale coherent structures, representing the pulsation patterns of lower energy.

Further comparing the modal distribution patterns across multiple variables in POD, differences from traditional single-variable

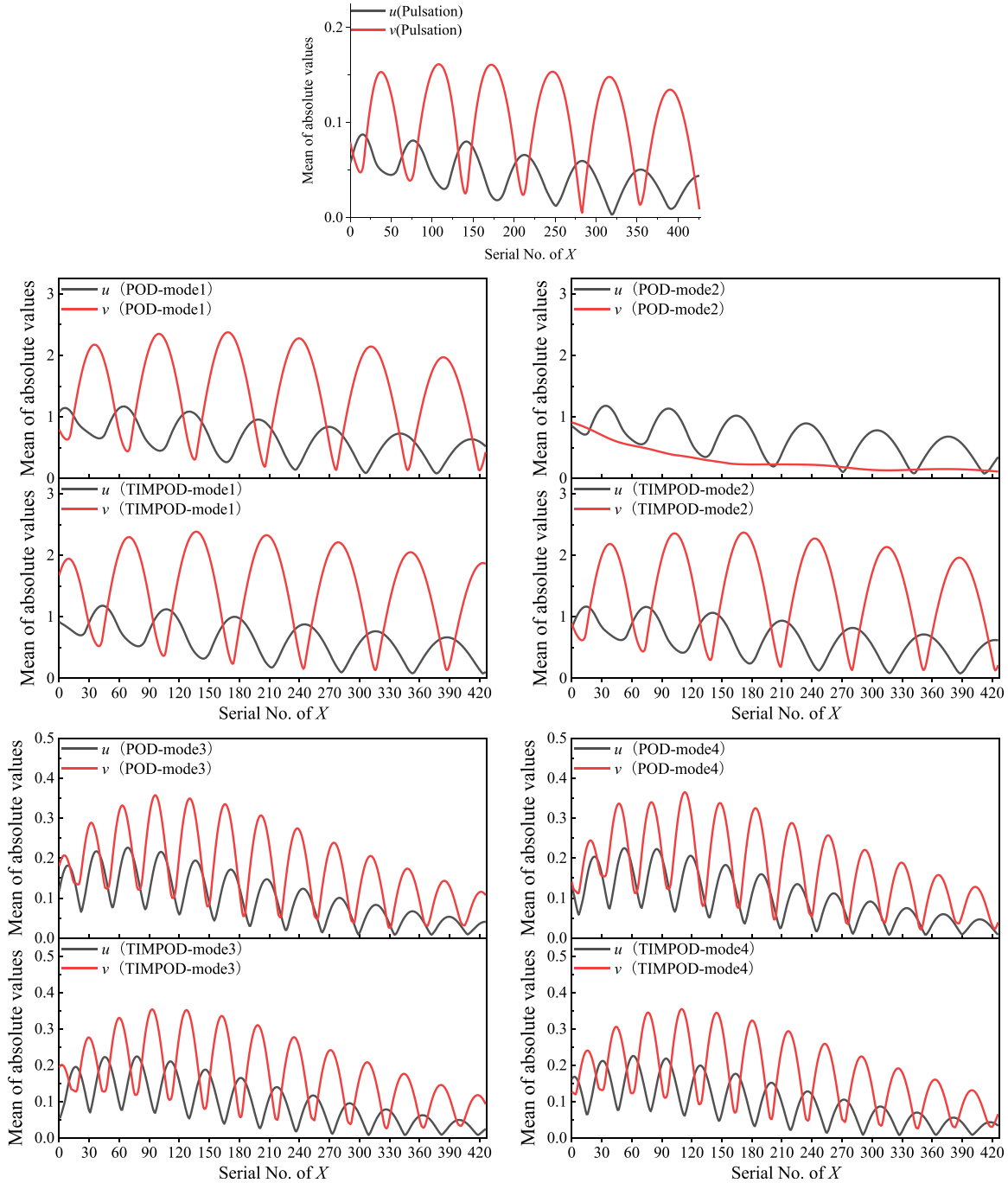

FIG. 11. Pulsation characteristics of different variable flow field snapshots and the distribution of mean values of the X axis projections of the modes of the pitching airfoil.

analysis emerge. Specifically, under the same mode number, modes of different variables do not represent the same physical patterns. This discrepancy manifests in two aspects: First, mode 0 of variable v does not represent the mean flow field; instead, the mean flow field is allocated to mode 3, deviating from the actual distribution of the mean flow field. Additionally, the coherent structure distribution of modes 1 and 2 of variable p does not correspond to the distribution of modes 1 and 2 of variable v.

Different variables should offer multiple perspectives on the same physical phenomenon. However, data-driven methods overlook the underlying connections between physical features. This paper addresses these issues by utilizing TIMPOD, as depicted in Fig. 5. TIMPOD achieves alignment of mode numbering across different variables, assigning modes with the same physical interpretation from different variables to the same mode number. For example, the ordering of modes 0, 1, and 2 for variable v is adjusted, and the ordering of modes 1 and 2 for variable p is swapped. Furthermore, TIMPOD corrects the distribution characteristics of the mode representing the mean flow field in variable v. Additionally, TIMPOD applies slight phase modulation in the X-direction to the modes of different variables, enabling spatial alignment between modes of different variables under the same mode number. TIMPOD accomplishes the modification and standardization of the mode distribution in POD, relying on the assumption that different variables possess the same temporal behavior.

To further visualize the spatial alignment capability of TIMPOD on modes, the pulsation features of flow field snapshots for different variables are projected onto a one-dimensional X axis and averaged, as shown in Fig. 6. Observing the projections of the pulsation features of different flow field snapshots, variables u and p exhibit the same phase. Their primary motion modes occur in the X-direction. On the other hand, the primary motion mode of variable v occurs in the Y-direction, resulting in a 180 deg phase difference between v and the other two variables, corresponding to the peaks and valleys. Apart from mode 0, the distributions of the other modes reflect the underlying patterns of flow field pulsations. Therefore, the phase characteristics of the mode distributions in different variables also correspond to the features of flow field pulsations. The phase characteristics of modes obtained from POD for different variables are chaotic. TIMPOD addresses this issue by achieving spatial alignment of modes, resulting in standardized and corrected phase characteristics of modes across different variables.

The right decomposition matrix obtained from singular value decomposition (time coefficient matrix) constructs an orthogonal state space that reflects the temporal characteristics of the flow field. Figure 7 displays the three-dimensional state space of the flow field using the time coefficients corresponding to the first three modes (where the state number corresponds to the mode number). The state space visualizes the temporal trajectories within the state space. By using the time coefficients, the state space captures the temporal evolution of the flow field and provides a visual representation of the flow dynamics. Each snapshot of the flow field at a given time is represented as a point in the state space, and the trajectories connecting these points represent the temporal states of the flow field. The temporal trajectories in state space are obtained through modal decomposition of the flow field data, mapping the temporal variations of the flow field onto coordinate axes in a lower-dimensional space. These axes represent different modes, where each mode corresponds to a specific

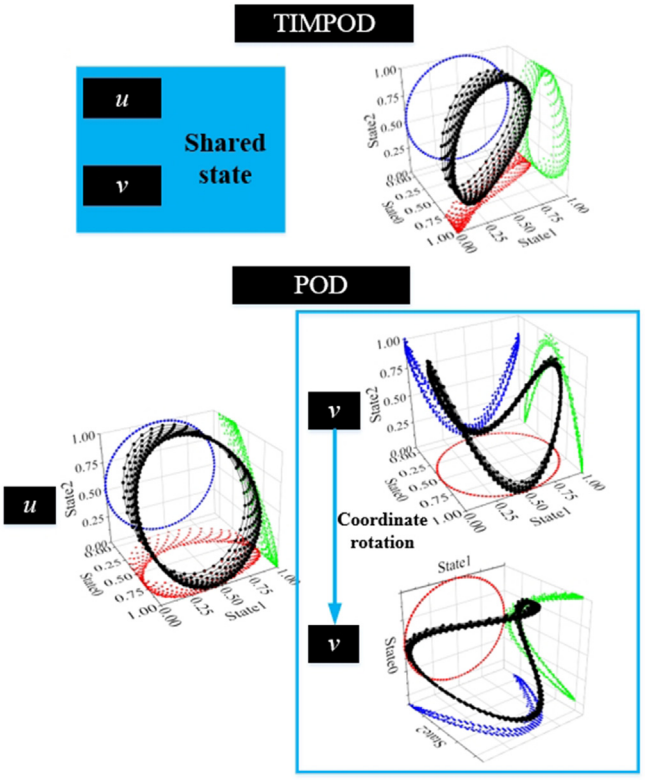

FIG. 12. Comparisons of POD with TIMPOD temporal trajectories within the state space at different variables of the pitching airfoil.

vibration or pulsation pattern within the flow field. The temporal trajectories describe the changing trends of these modes over time.

In summary, the basic characteristics of the state space for the flow field around a cylinder across multiple variables: The flow field snapshots contain multiple periodic flow states, and their temporal trajectories in the state space overlap, demonstrating stable periodicity. State 0 represents the average flow field state, representing the dominant mode as time progresses. State 1 and state 2 jointly reflect the pulsation patterns of the flow field, representing the temporal characteristics of the pulsations as time progresses. The combined state of state 1 and state 2 appears as a perfect circle, indicating that their temporal behaviors are conjugate. The combined state of state 0 with state 1 and state 2 exhibits elliptical shapes, indicating different temporal behaviors among these states.

It is worth noting that the state space of variable v under the POD state needs to be corrected to align with the basic characteristics of the state space. The alignment of mode numbering is reflected in the rotation of the coordinate axes in the state space. The temporal trajectories in the state space for the three variables under the POD state are similar, confirming the validity of the assumption made by TIMPOD. TIMPOD assumes that the three variables share a common state space, which encompasses the basic characteristics of the state spaces of the three variables under the POD state, reflecting the fundamental temporal patterns of the flow field.

The SIMPOD assumption states that three variables share the same spatial features. To verify the rationality of SIMPOD, Fig. 8 displays the temporal trajectories of the shared flow field modes and the state space for different variables under SIMPOD. The results indicate that the shared flow field modes are obtained by reordering the modes of the three variables under POD. The sorting criterion is based on the energy contribution of each mode. However, SIMPOD does not address the limitations of the POD modes. Under the shared modes, mode 0 represents the average velocity of variable u, while mode 1 and mode 2 correspond to the sub-modes of the POD modes for variable *v*. Similarly, mode 3 and mode 4 represent the sub-modes of the POD modes for variable *u*. As a result, the state space constructed based on the first three coefficient modes does not possess a specific physical interpretation. SIMPOD fails to address the limitations of the POD modes. The constructed state space under the shared modes lacks a definite physical meaning.

B. The pitching airfoil case

Unlike the distinct periodic temporal features of vortex shedding in the wake of a low Reynolds number cylinder, the wake state of a pitching airfoil at low Reynolds number emphasizes the sequential progression of temporal features. First, we observe the energy ratios and relative reconstruction errors of the first 100 flow modes decomposed through modal analysis on the pitching airfoil, as shown in Fig. 9. For each variable, both the energy ratios and relative reconstruction errors converge as the number of modes increases. Although there are differences in the energy ratio trends between POD and TIMPOD, the variation of relative reconstruction errors indicates that the data preservation performance of TIMPOD and POD is nearly identical. SIMPOD exhibits the poorest performance, as the assumption and approach of sharing the right decomposition matrix (spatial states) are not reasonable.

Figure 10 illustrates the mode distribution characteristics of POD and TIMPOD, as well as representative flow field features. For the wake of a pitching airfoil at low Reynolds number, the univariate modal decomposition of different variables still suffers from the problem of inconsistent physical modes. TIMPOD achieves alignment of modes and unification of potential physical modes across different variables by sharing temporal patterns among multiple variables. For example, the mode ordering of variables *u* is adjusted by swapping modes 1 and 2, while the mode ordering of variable *v* is adjusted for modes 0, 1, and 2.

Figure 11 depicts the pulsation characteristics of different variable flow field snapshots and the distribution of mean values of the X axis projections of the modes. Similar to the wake of a low Reynolds number cylinder, for the pitching airfoil at a low Reynolds number, variables *u* and *v* exhibit the correspondence between peaks and valleys, with a phase difference of half a cycle. This represents an inherent characteristic of the flow field. Under POD, the mode decomposition of the two variables disrupts the original relationship between the variables, resulting in mixed phase characteristics between them. TIMPOD addresses this issue by achieving spatial alignment of modes, thereby restoring the phase characteristics between different variables in the original flow field.

Figure 12 illustrates the three-dimensional state space of the flow field using time coefficients corresponding to the first three modes (with the number of states corresponding to the number of modes). State 0 represents the dominant mode evolving over time, with spatial trajectories spiraling along the axis of state 0. States 1 and 2 represent the pulsation patterns of the flow field and are displayed as perfect circles, indicating that their time behaviors are conjugate. Under POD, the state space of variable *v* requires a rotation of the coordinate axes to conform to the fundamental characteristics of the state space. TIMPOD achieves the sharing of a common state space for both variables, which is closer to the state space of variable *u*, suggesting that variable *u* plays a more significant role in the temporal progression.

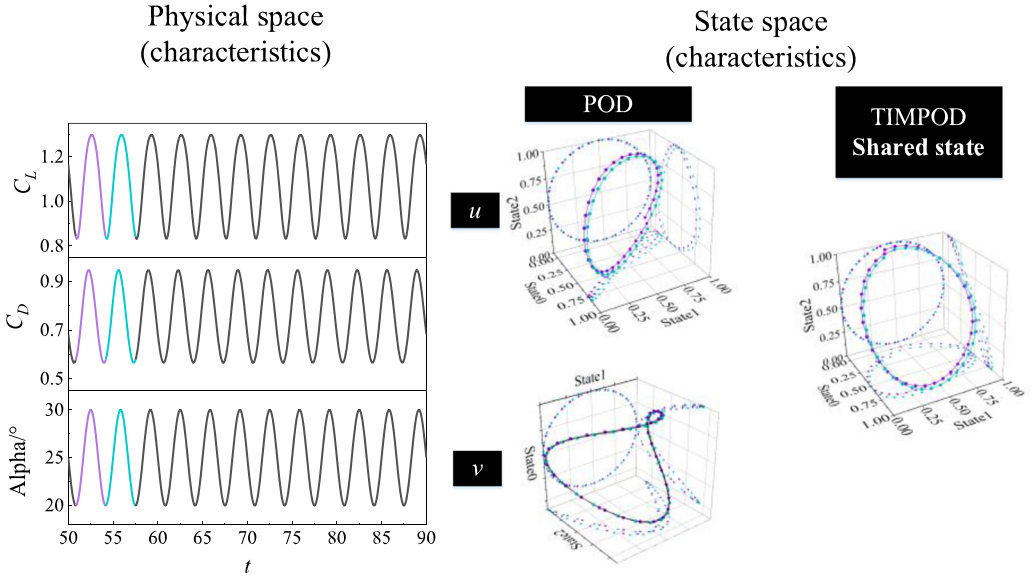

FIG. 13. The connection between temporal characteristics in physical space (pitch angle, lift coefficient, and drag coefficient) and temporal trajectories in state space.

Figure 13 reveals the temporal behavior and dynamic characteristics of the flow field in physical space by examining the temporal variations of lift and drag coefficients, as well as the corresponding pitch angle of the pitching airfoil. This depiction showcases the temporal features of trajectories in the state space. Different colors represent different periods, and there is a close connection between the temporal characteristics in physical space and the temporal trajectories in the state space. In the POD state space, the temporal trajectories of variable u under different periods spiral along the axis associated with state 0, capturing the temporal behavior of the pitching airfoil. State 1 and state 2 represent the temporal trajectories of pulsation features with a phase difference. The temporal behavior of variable v under different periods is not manifested in the state space, as its temporal trajectories exhibit a phase difference, reflecting the pulsation features. The shared state space obtained by TIMPOD, which closely resembles the POD state space of variable u, is reasonable because it captures both the temporal behavior and the pulsation features.

Figure 14 illustrates the temporal trajectories of the shared flow field modes obtained by SIMPOD for the pitching flat plate, as well as the state spaces of different variables. Under the shared modes, mode 0 represents the average velocity of variable u, while modes 1 and 2 correspond to sub-modes of the POD modes of variable v. Similarly, modes 3 and 4 represent sub-modes of the POD modes of variable u. The shared flow field modes are obtained by reordering the modes of the two variables under POD. However, SIMPOD does not address the limitations of the POD modes, and the reordering of modes lacks meaningful interpretation. As a result, the state space constructed based on the first three coefficient modes does not possess specific physical explanations.

C. The turbulent flow experiment through a cylinder

Using turbulent flow experiments with a cylinder under multiple transient conditions as an example, the applicability of the proposed

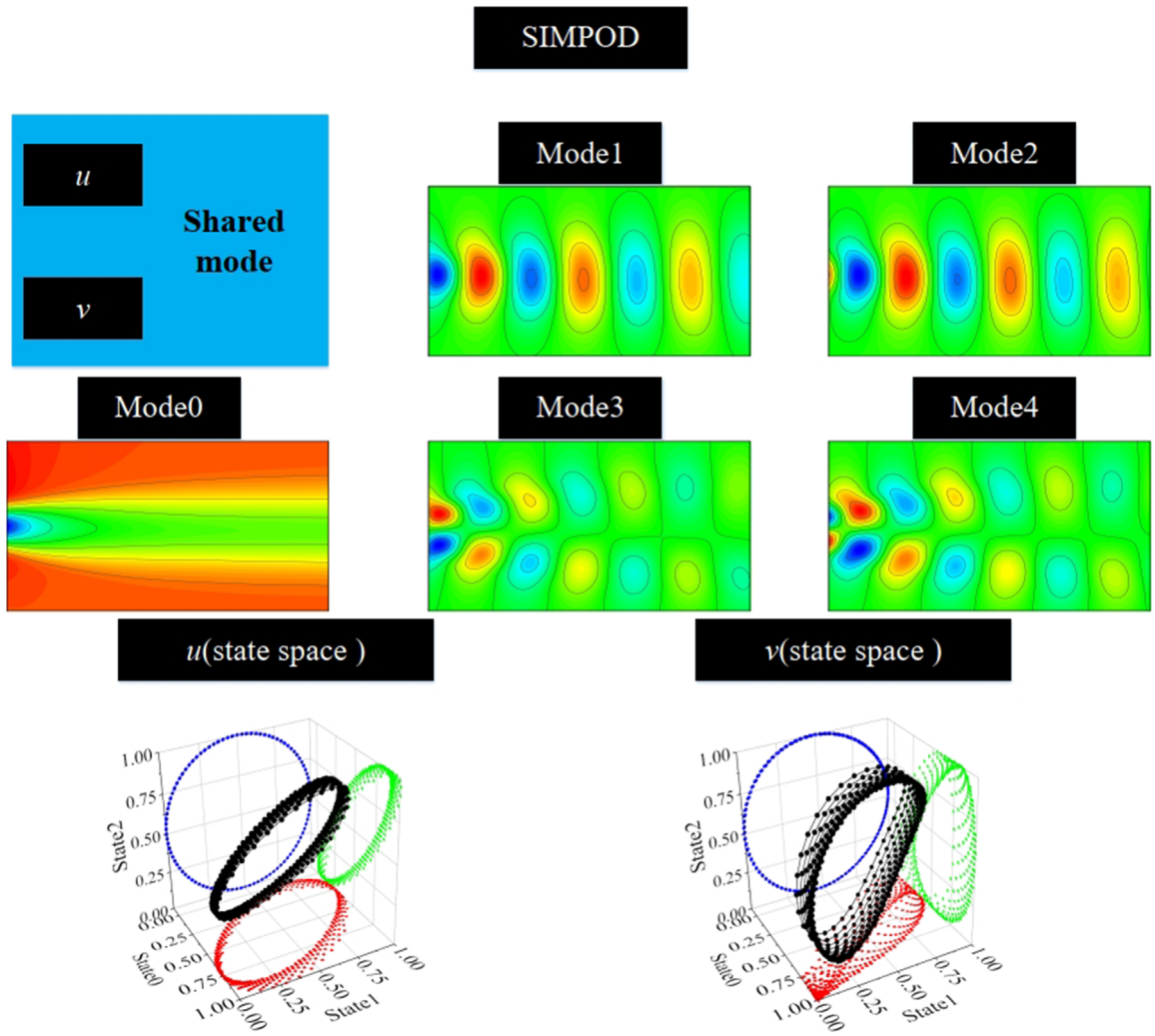

FIG. 14. Mode distribution characteristics of POD and TIMPOD, along with representative flow field features of the pitching airfoil.

method in real-world, multi-state turbulent flows is validated. Figure 15 displays the energy ratios and relative reconstruction errors of the first 100 flow modes after modal decomposition. For each variable, the energy ratios and relative reconstruction errors converge as the number of modes increases. However, the performance of the algorithms determined by the energy ratios and relative reconstruction errors is not consistent. This suggests that using energy ratios as prior information to represent the proportion of original flow field information in complex flow fields is not accurate. The reconstruction error, as a posterior criterion, provides more persuasive evaluation of the original flow field information. The reconstruction errors among the three algorithms are very close for low-order modes but diverge for high-order modes. Since low-order modes are more important as they represent more original flow field information, the three algorithms exhibit no significant difference in the reconstruction errors for the turbulent flow experiments with the cylinder in the low-order modes.

Figure 16 illustrates the mode distribution characteristics of POD and TIMPOD, as well as the time-averaged flow field features. For turbulent states, the univariate modal decomposition of different variables still suffers from the problem of inconsistent physical modes. This issue becomes more prominent for complex higher-order modes (modes 3 and 4). TIMPOD, however, is capable of achieving mode alignment between different variables and unification of potential physical modes. For instance, for the low-order variable u, the mode ordering is adjusted by swapping modes 1 and 2, while for variable v, the mode ordering is adjusted for modes 0, 1, and 2. For modes 3 and 4, TIMPOD undergoes more significant modifications, and the emergence of new mode distributions is obtained by exchanging from higher-order modes.

Figure 17 illustrates the distribution of the average X axis projections of different modes under turbulent flow conditions around a cylinder. TIMPOD achieves mode alignment and peak-valley correspondences for variables u and v. However, mode 3 lacks peak-valley correspondences due to its concentrated coherent structural distribution. Overall, whether it is for the steady state wake of a cylinder, the wake of a pitching airfoil with kinematic features, or the real multi-state turbulent flow around a cylinder, the issue of inconsistent physical modes in univariate POD decomposition results is a common problem. POD pursues the optimal representation of univariate features, thereby overlooking the inter-variable relationships, and the obtained mode features may not accurately represent the true underlying flow patterns. TIMPOD, as a modification algorithm for POD, assumes that variables share the same temporal transition state, which is reasonable. It achieves a flow decomposition process that better aligns with the physical characteristics by considering multiple variables.

Figure 18 reveals the temporal behavior and dynamic characteristics of the flow field in physical space by examining the time variation

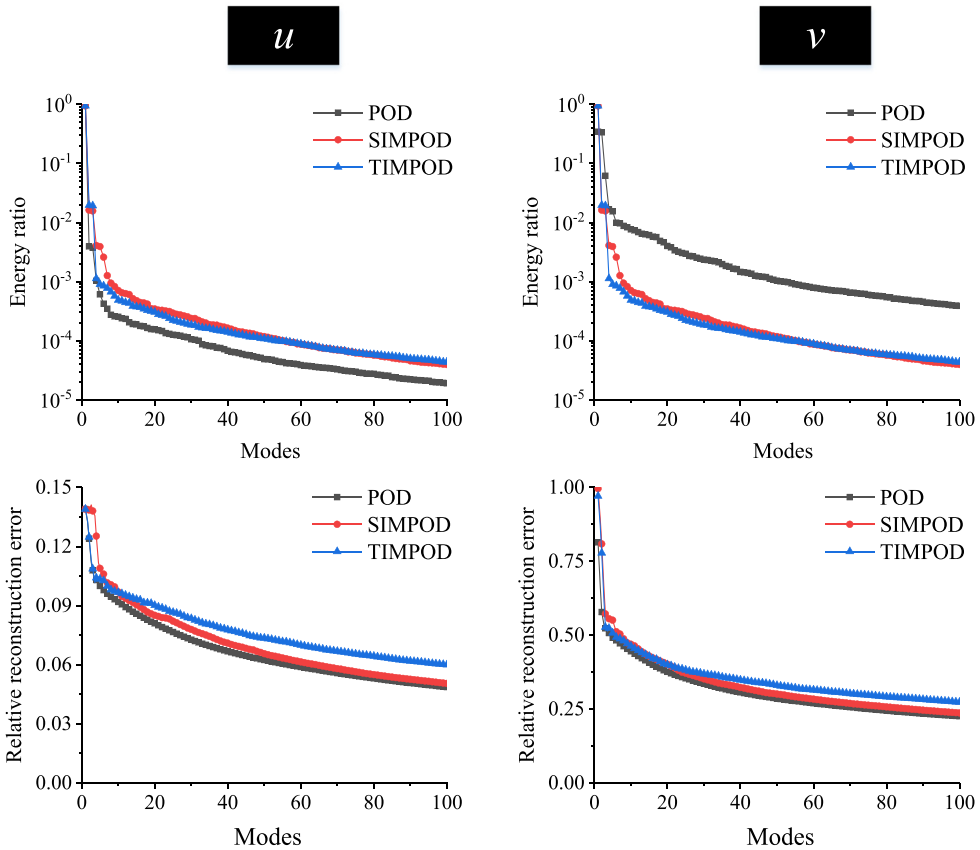

FIG. 15. Energy ratio and relative reconstruction error for the first 100 modes of flow over a cylinder of the turbulent flow experiment through a cylinder.

curves of velocity components at flow field monitoring points. In contrast to the smooth temporal trajectory distributions observed in the wake of a low Reynolds number cylinder and the pitching airfoil, the temporal trajectory distribution under real turbulent flow conditions exhibits a more unstable behavior over time. Variable u still plays a more significant role in the temporal progression, and the state space of variable v still requires a rotation of the coordinate axes to conform to the fundamental characteristics of the state space. For variable u, state 0 represents the dominant mode evolving over time, with its temporal behavior reflected in the direction of the axis represented by state 0. States 1 and 2 of variable u represent the flow's fluctuation modes and are displayed as complete circles. The radius of the circle varies under different temporal behaviors, but they exhibit conjugate temporal behavior. The temporal behavior of variable v is not as pronounced, but it is still observable as the temporal progression along the axis of state 2. States 0 and 1 of variable v construct circular orbits with varying radii under different temporal behaviors, similar to variable u. TIMPOD achieves the sharing of the state space between variables v and u, which is closer to the state space of variable u. This approach captures both the temporal behavior and the fluctuation features. The two different stable states are separated on distinct elliptical orbits, with the transitional state between them. The high-speed state exhibits larger radii in its circular orbit, while the low-speed state exhibits smaller radii.

Figure 19 displays the temporal trajectories of the shared flow field modes obtained by SIMPOD for transient turbulent flow around a cylinder, as well as the state spaces of different variables. Under the shared modes, mode 0 represents the average velocity of variable u, while mode 1 and mode 2 correspond to sub-modes of the POD modes of variable v. Similarly, modes 3 and 4 represent sub-modes of the POD modes of variable u. By combining the two previous cases, we can identify the characteristics of SIMPOD. The shared flow field modes obtained are the result of reordering the modes of the two variables under POD. The reordering is based on the energy ratios obtained from the univariate POD of each variable. The data of different variables are then projected into the newly composed state space. SIMPOD manages to retain the temporal trajectories and behaviors of variable u to some extent by preserving the average velocity of variable u as mode 0. However, the radii of the circular trajectories corresponding to the two different stable states of velocity are the same since

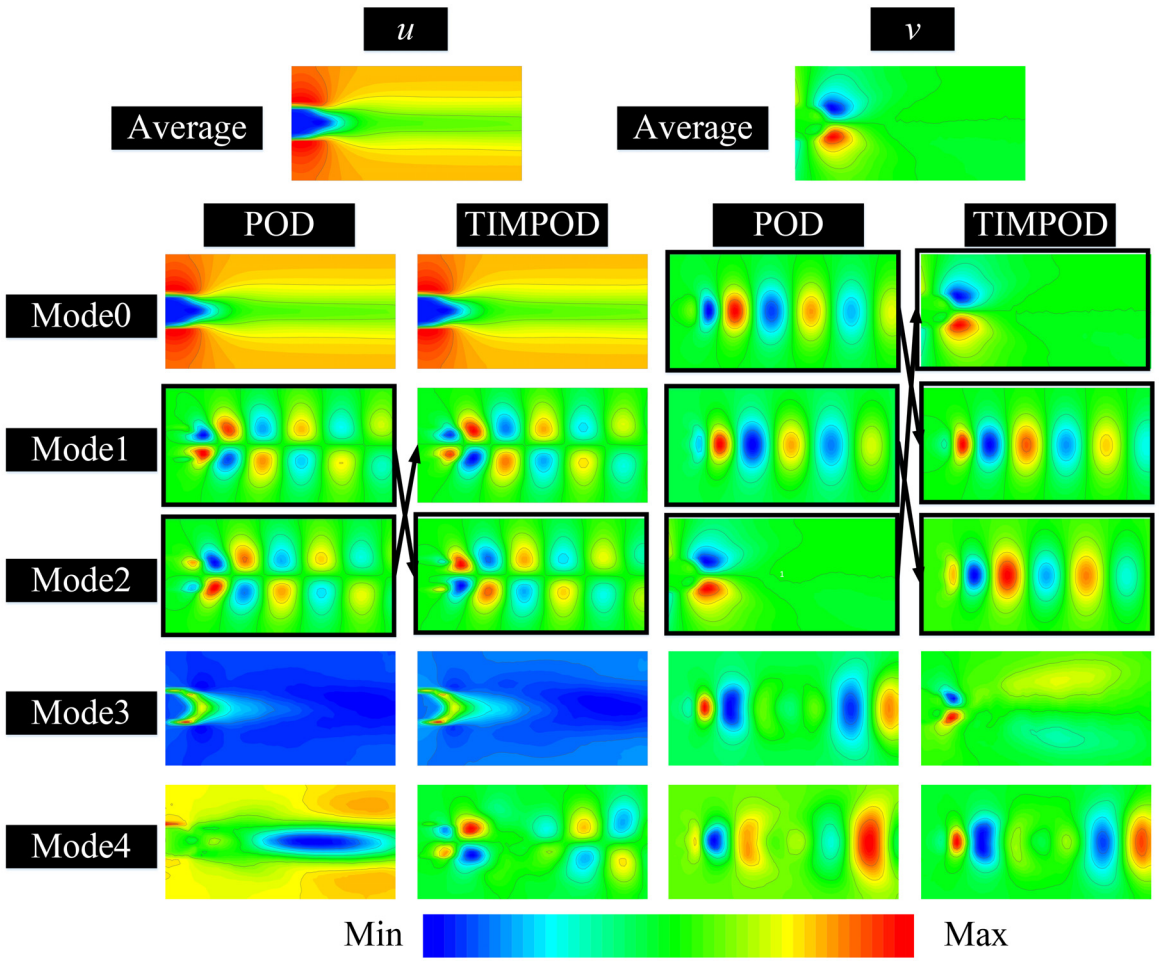

FIG. 16. Mode distribution characteristics of POD and TIMPOD, along with time-averaged flow field features.

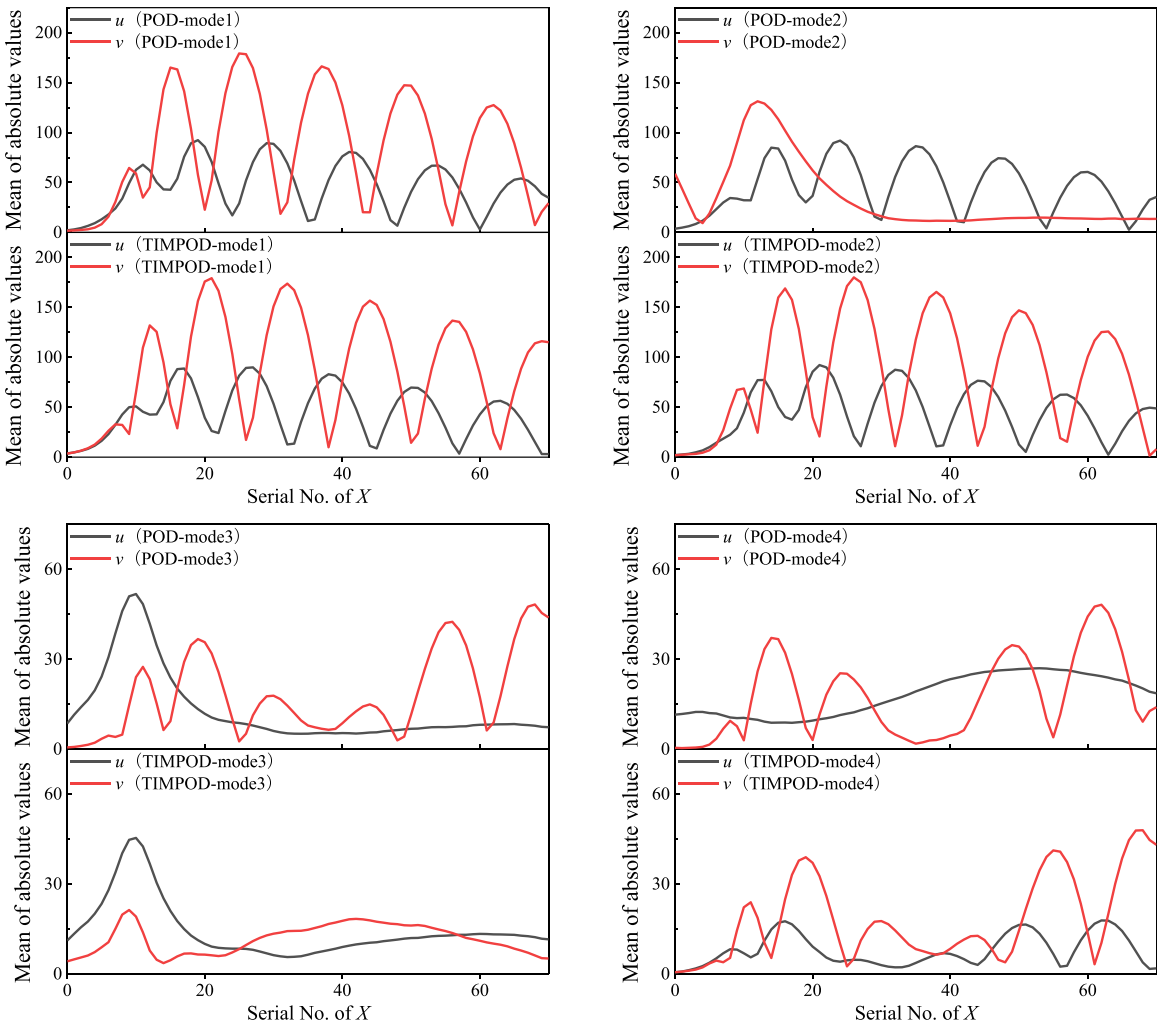

FIG. 17. Distribution of the average X axis projections of different modes under turbulent flow conditions around a cylinder.

modes 1 and 2 are sub-modes of variable *v*, leading to a weakening representation of the temporal behavior. Similarly, the temporal trajectories of variable *v* obtained by SIMPOD also encounter the same problem. Thus, the reordering of modes obtained by SIMPOD does not address the limitations of POD modes and can even degrade the performance of POD.

V. CONCLUSIONS

This study explores the challenges of multivariate modal decomposition in various flow scenarios using three typical datasets and presents a solution based on information sharing. The main conclusions are as follows:

(1) The inconsistency of physical modes in multi-dimensional POD is a common issue, specifically manifested as the loss of phase characteristics among variables in real physical space. This problem arises due to the neglect of relationships between physical features, leading to the inability of multi-dimensional POD to uncover the same flow patterns. Consequently, the results obtained from conventional POD are questionable.

(2) TIMPOD assumes shared time information between variables to align and correct the phases of multi-variable modes, allowing the modal distribution of different variables to reveal common physical features. Additionally, the shared time information represents the temporal state of the flow field, describing its temporal behavior. Importantly, TIMPOD has little impact on the reconstruction error of POD, making it an effective correction algorithm for POD and multi-dimensional modal decomposition.

(3) SIMPOD assumes shared spatial information among variables and obtains shared flow field patterns by reordering modes of multiple variables under the POD framework. However, this mode reordering affects the reconstruction error and fails to provide meaningful physical insight.

(4) In summary, leveraging shared physical information, especially time information, among variables significantly improves the

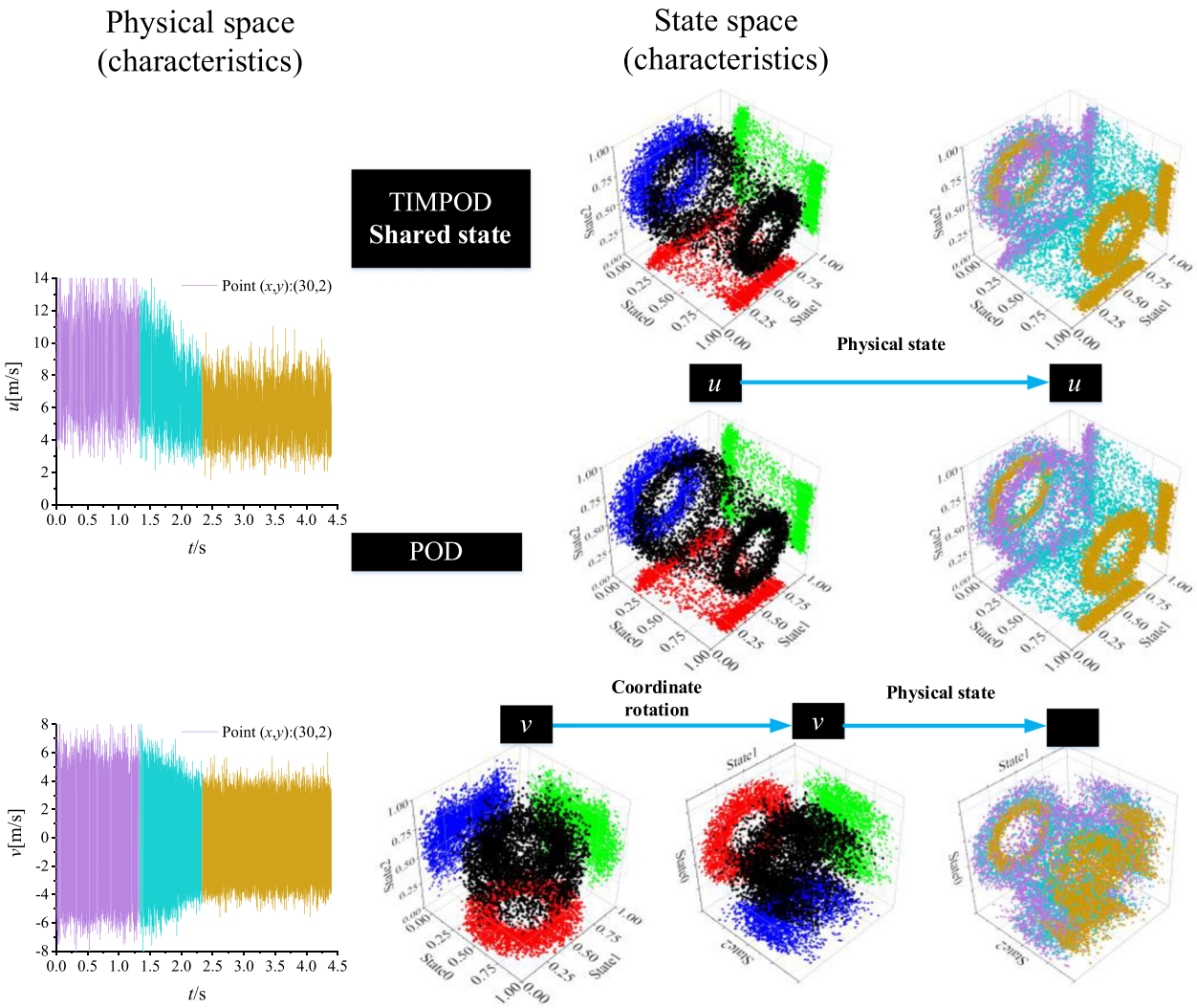

FIG. 18. The connection between temporal characteristics in physical space and temporal trajectories in state space.

limitations of the POD algorithm and enables multi-dimensional modal decomposition. Furthermore, TIMPOD is highly scalable as it is based on the joint SVD of multi-dimensional matrices. Hence, it can be combined with various existing modal decomposition algorithms (e.g., DMD) to extend their respective multi-dimensional modal decomposition and achieve more meaningful results, which represents a promising direction for future research.

ACKNOWLEDGMENTS

This work was partially funded by the National Natural Science Foundation of China (Nos. 52061135107 and 52071062), the Fundamental Research Fund for the Central Universities (No. DUT20TD108), and the computation support of the Supercomputing Center of Dalian University of Technology.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Zihao Wang: Conceptualization (lead); Methodology (lead); Software (equal); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **Guiyong Zhang:** Conceptualization (equal); Data curation (lead); Funding acquisition (lead); Project administration (lead); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal). **Tiezhi Sun:** Funding acquisition (equal); Investigation (lead); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Writing – original draft (equal). **Huakun Huang:** Formal analysis

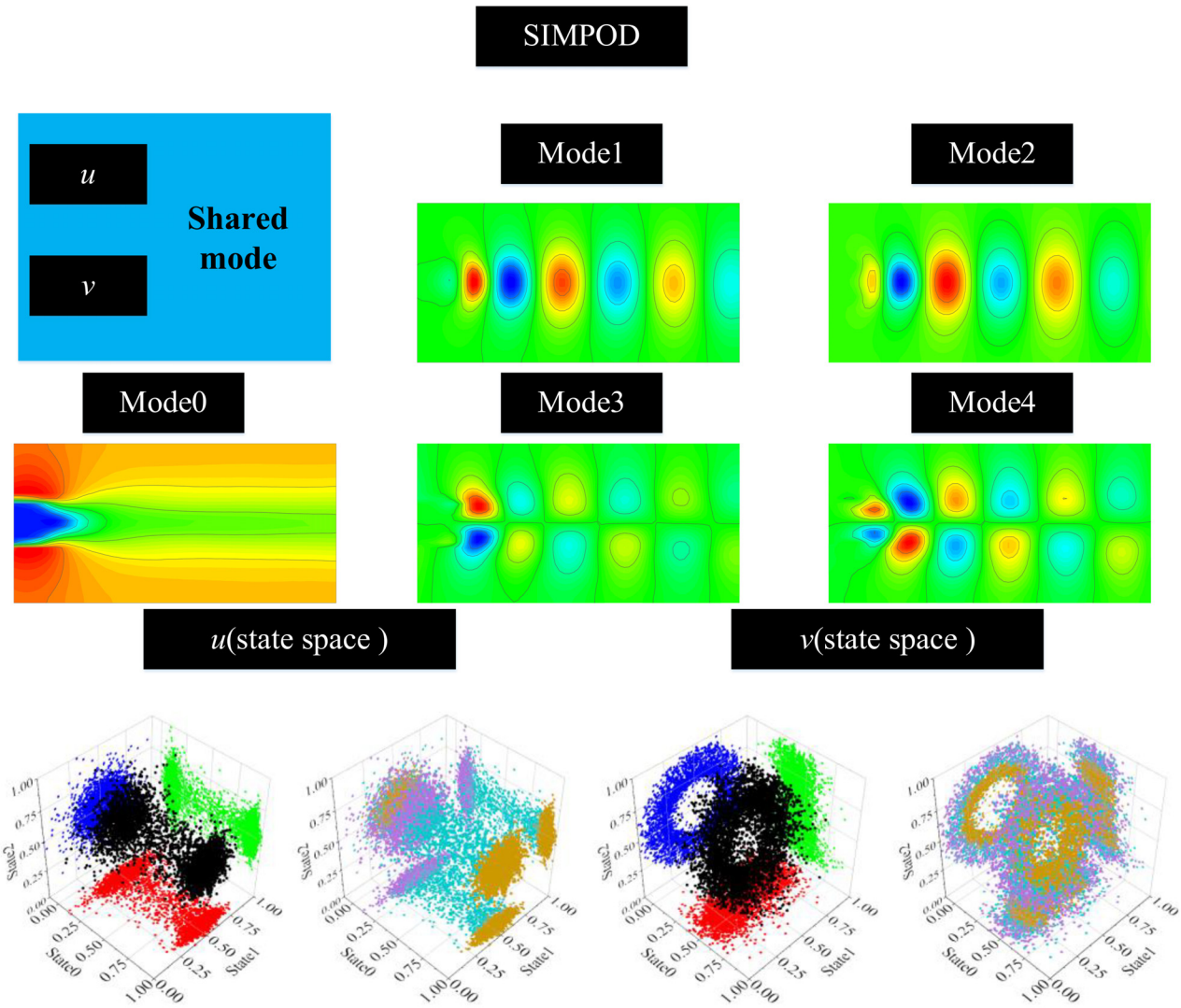

FIG. 19. Temporal trajectories of the shared flow field modes and the state space for different variables under SIMPOD of the turbulent flow experiment through a cylinder.

(equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

NOMENCLATURE

$\boldsymbol{A}$	Data matrix
$\boldsymbol{a}$	Right decomposition matrix
c	Chord length of the airfoil
D	Diameter of the cylinder
f	Frequency
f^*	Dimensionless frequency
$\boldsymbol{I}$	Identity matrix
L	Length
L_x	Length in x direction
L_y	Length in y direction
N_x	Number of grids in x direction
N_y	Number of grids in y direction
p	Pressure
Re	Reynolds numbers
t	Time
U	Free-stream velocity
$\boldsymbol{U}^*$	Left singular value matrix
u	Velocity of x-component

V	Right singular value matrix
v	Velocity of y-component
W	Variable value
$\mathbf{X}$	Original data
$\mathbf{X}_r$	Reconstructed data
x	x-coordinate; streamwise streamwise
y	y-coordinate; transverse streamwise
α	Angle of attack
α_c	Critical angle of attack
$\alpha(t)$	Time-varying angle of attack
α_A	Amplitude of the sinusoidal variation
α_M	Baseline angle of attack
Δ	Filter length
$\mathbf{\Sigma}$	Diagonal matrix of singular values
ε	Energy ratio
ε_r	Relative reconstruction error
μ	Kinematic viscosity
σ	Singular value
∇	Gradient operator

REFERENCES

[1]A. Pollard, L. Castillo, L. Danaila, and M. Glauser, "Challenges for large eddy simulation of engineering flows," in *Whither Turbulence and Big Data in the 21st Century?* (Springer, Cham, Switzerland, 2018).

[2]K. Duraisamy, G. Iaccarino, and H. Xiao, "Turbulence modeling in the age of data," Annu. Rev. Fluid Mech. **51**, 357–377 (2019).

[3]S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," Annu. Rev. Fluid Mech. **52**, 477–508 (2020).

[4]J. L. Lumley, "The structure of inhomogeneous turbulent flows," in *Atmospheric Turbulence and Radio Wave Propagation*, edited by A. M. Yaglom and V. I. Tatarski (Nauka, 1967), pp. 166–178.

[5]P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," J. Fluid Mech. **656**, 5–28 (2010).

[6]K. Taira, S. L. Brunton, S. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, "Modal analysis of fluid flows: An overview," AIAA J. **55**(12), 4013–4041 (2017).

[7]K. Taira, M. S. Hemati, S. L. Brunton, Y. Sun, and C. A. Yeh, "Modal analysis of fluid flows: Applications and outlook," AIAA J. **58**, 998–1022 (2020).

[8]Q. A. Huhn, M. E. Tano, J. C. Ragusa, and Y. Choi, "Parametric dynamic mode decomposition for reduced order modeling," J. Comput. Phys. **475**, 111852 (2023).

[9]K. Lu, Y. Jin, Y. Chen, Y. Yang, L. Hou, Z. Zhang, C. Fu, *et al.* "Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems," Mech. Syst. Signal Process. **123**, 264–297 (2019).

[10]P. Solán-Fustero, J. L. Gracia, A. Navas-Montilla, and P. García-Navarro, "A POD-based ROM strategy for the prediction in time of advection-dominated problems," J. Comput. Phys. **471**, 111672 (2022).

[11]E. H. Krath, F. L. Carpenter, P. G. Cizmas, and D. A. Johnston, "An efficient proper orthogonal decomposition based reduced-order model for compressible flows," J. Comput. Phys. **426**, 109959 (2021).

[12]G. Zhang, Z. Wang, H. Huang, H. Li, and T. Sun, "Comparison and evaluation of dimensionality reduction techniques for the numerical simulations of unsteady cavitation," Phys. Fluids **35**(7), 073322 (2023).

[13]Z. Wang, G. Zhang, T. Sun, C. Shi, and B. Zhou, "Data-driven methods for low-dimensional representation and state identification for the spatiotemporal structure of cavitation flow fields," Phys. Fluids **35**(3), 033318 (2023).

[14]L. Sirovich, "Turbulence and the dynamics of coherent structures—I: Coherent structures," Q. Appl. Math. **45**(3), 561–571 (1987).

[15]C. W. Rowley, "Model reduction for fluids, using balanced proper orthogonal decomposition," Int. J. Bifurcation Chaos **15**(3), 997–1013 (2005).

[16]C. Picard and J. Delville, "Pressure velocity coupling in a subsonic round jet," Int. J. Heat Fluid Flow **21**(3), 359–364 (2000).

[17]A. Towne, O. T. Schmidt, and T. Colonius, "Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis," J. Fluid Mech. **847**, 821–867 (2018).

[18]M. A. Mendez, M. Balabane, and J. M. Buchlin, "Multi-scale proper orthogonal decomposition of complex fluid flows," J. Fluid Mech. **870**, 988–1036 (2019).

[19]Z. Wang, X. Xing, T. Sun, and G. Zhang, "Segmentation of unsteady cavitation flow fields based on multivariate spatiotemporal hierarchical clustering," Phys. Fluids **35**(5), 053317 (2023).

[20]B. Begiashvili, N. Groun, J. Garicano-Mena, S. Le Clainche, and E. Valero, "Data-driven modal decomposition methods as feature detection techniques for flow problems: A critical assessment," Phys. Fluids **35**(4), 041301 (2023).

[21]R. Halder, K. Fidkowski, and K. Maki, "Local non-intrusive reduced order modeling using isomap," AIAA Paper No. 2022-0081, 2022.

[22]Z. Wang, G. Zhang, H. Huang, H. Xu, and T. Sun, "Joint proper orthogonal decomposition: A novel perspective for feature extraction from multivariate cavitation flow fields," Available at SSRN: https://ssrn.com/abstract=4477801 or http://dx.doi.org/10.2139/ssrn.4477801.

[23]J. Ren and X. Mao, "Uncertainty transmission of fluid data upon proper orthogonal decompositions," Phys. Fluids **35**(7), 071702 (2023).

[24]P. Holmes, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, 2012).

[25]J. Borée, "Extended proper orthogonal decomposition: A tool to analyse correlated events in turbulent flows," Exp. Fluids **35**(2), 188–192 (2003).

[26]S. Gordeyev, N. De Lucca, E. J. Jumper, K. Hird, T. J. Juliano, J. W. Gregory, D. J. Wittich, *et al.* "Comparison of unsteady pressure fields on turrets with different surface features using pressure-sensitive paint," Exp. Fluids **55**(1), 1661 (2014).

[27]S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, 2022).

[28]H. Csala, S. Dawson, and A. Arzani, "Comparing different nonlinear dimensionality reduction techniques for data-driven unsteady fluid flow modeling," Phys. Fluids **34**(11), 117119 (2022).

[29]J. N. Kutz, *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data* (OUP Oxford, 2013).

[30]M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. **378**, 686–707 (2019).

[31]S. T. Dawson, M. Hemati, D. C. Floryan, and C. W. Rowley, "Lift enhancement of high angle of attack airfoils using periodic pitching," AIAA Paper No. 2016-2069, 2016.

[32]S. T. Dawson, "Reduced-order modeling of fluids systems, with Applications in Unsteady Aerodynamics," Ph.D. dissertation (Princeton University, 2017).

[33]A. Towne, S. T. Dawson, G. A. Brès, A. Lozano-Durán, T. Saxton-Fox, A. Parthasarathy, K. Taira, *et al.* "A database for reduced-complexity modeling of fluid flows," AIAA J. **61**, 2867 (2023).

[34]K. Taira and T. Colonius, "The immersed boundary method: A projection approach," J. Comput. Phys. **225**(2), 2118–2137 (2007).

[35]T. Colonius and K. Taira, "A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions," Comput. Methods Appl. Mech. Eng. **197**(25–28), 2131–2146 (2008).

[36]M. A. Mendez, "Linear and nonlinear dimensionality reduction from fluid mechanics to machine learning," Meas. Sci. Technol. **34**(4), 042001 (2023).

[37]M. A. Mendez, D. Hess, B. B. Watz, and J. M. Buchlin, "Multiscale proper orthogonal decomposition (mPOD) of TR-PIV data—A case study on stationary and transient cylinder wake flows," Meas. Sci. Technol. **31**(9), 094014 (2020).