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

Uncertainty transmission of fluid data upon proper orthogonal decompositions

Cite as: Phys. Fluids **35**, 071702 (2023); doi: 10.1063/5.0157849
Submitted: 11 May 2023 · Accepted: 24 June 2023 ·
Published Online: 11 July 2023

Jie Ren (任杰),[1,2,3] and Xuerui Mao (毛雪瑞)[1,2,3,4,a)]

AFFILIATIONS

[1]School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China
[2]State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, Beijing 100081, China
[3]Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing 314003, China
[4]Advanced Research Institute of Multidisciplinary Sciences, Beijing Institute of Technology, Beijing 100081, China

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

ABSTRACT

Proper orthogonal decomposition (POD) serves as a principal approach for modal analysis and reduced-order modeling of complex flows. The method works robustly with most types of fluid data and is fundamentally trusted. While, in reality, one has to discern the input spatiotemporal data as passively contaminated globally or locally. To understand this problem, we formulate the relation for uncertainty transmission from input data to individual POD modes. We incorporate a statistical model of data contamination, which can be independently established based on experimental measurements or credible experiences. The contamination is not necessarily a Gaussian white noise, but a structural or gusty modification of the data. Through case studies, we observe a general decaying trend of uncertainty toward higher modes. The uncertainty originates from twofold: self-correlation and cross correlation of the contamination terms, where the latter could become less influential, given the narrow correlation width measured in experiments. Mathematically, the self-correlation is determined by the inner product of the data snapshot and the mode itself. Therefore, the similarity between the input data and the resulting POD modes becomes a critical and intuitive indicator when quantifying the uncertainty. A scaling law is shown to be applicable for self-correlation that promotes fast quantification on sparse grids.

I. INTRODUCTION

In fluid mechanics, proper orthogonal decomposition (POD)[1,2] is widely applied to investigate turbulent structures, model reduction, and flow control. Readers are referred to Taira *et al.*[3,4] for reviews. The input data of POD can be from experimental measurements or numerical simulations, which are inevitably subject to errors stemming from, e.g., environmental noise, truncation, and imperfection of models. Errors consequently become transferred into the POD modes, leading to uncertain results. The community has benefited from POD's excellent capability of modal analysis but has yet to make the transfer of uncertainty from input data to modes explored.

In terms of the truncation error incurred in POD, Singler[5] derived exact expressions for the error in two different Hilbert spaces and found that the error can be expressed solely as a function of the POD eigenvalues and eigenvectors. On top of the previous work, Fareed and Singler[6] performed error analysis for incremental POD in which a new data column acts as an input. The error bound for truncation and the convergence of reduced-order models is proved. García-Archilla *et al.*[7] considered three sources of errors (the spatial, temporal discretization, and truncation) and proved a rigorous error estimate for POD data assimilation schemes of Navier–Stokes equations.

In terms of noisy data, the mode corruption[8] and error estimation[9] have been comprehensively studied, showing that leading modes are robust. POD has also been applied to remove these noise. Brindise and Vlachos[10] proposed an entropy-line fit method for POD mode truncation and experimental data denoising. The method was able to keep modes with relevant flow structures whose energy was less than that of the noise, therefore improving POD truncation's effectiveness.

In other forms of modal decomposition techniques, such as the dynamic mode decomposition (DMD), Duke *et al.*[11] made a systematic error analysis of DMD for input data representative of shear flow instabilities, particularly sinusoidal, square, and sawtooth waves with tunable parameters. In the evaluation of growth rate, it was shown that the waveform, resolution of samples relative to characteristic wavelength, and signal-to-noise-ratio strongly affect the error. Therefore, special care must be taken for the quality of the input data when decoding the results.

On the quality of data generation, quantification of uncertainty in computational fluid dynamics has been reviewed by Roache.[12] Howell[13] comprehensively assessed error distribution and correlation from PIV measurement. The result indicates that the error, by and large, follows the Gaussian distribution and is influenced by the shear of the flow. The most interesting results are that errors are correlated in space. The correlation factor decays semi-linearly with the distance of two points. We term the distance at which the correlation gets zero the correlation width which is found to be 3–6 points in PIV.

Based on research to date, one can assess the error of POD-based reduced-order models even with an analytical formulation.[5] However, how does uncertainty transmit in the POD process when flow data are biased by structural (a single or several locations) or gusty (single or several snapshots contaminated) deviations whose mean is not necessarily zero? In this situation, the contamination is not of noise type but acts as a correction to the bulk data. How correct are the POD modes and their eigenvalue? Answering these questions forms the motivation for the present study. In Sec. II, we formulate the analytical expressions of the uncertainty. Section III presents results to reveal the interpretation, scaling, and generality of uncertainty. The study is concluded in Sec. IV.

II. UNCERTAINTY TRANSMISSION OF POD MODES FROM CONTAMINATED DATA

POD takes snapshots of physical variables, e.g., velocity and pressure, as the input data. We define this contaminated ensemble of snapshots $\boldsymbol{A_c}$, which is composed of the true value $\boldsymbol{A_o}$ and a latent contamination $\boldsymbol{A}'$, satisfying

$$\boldsymbol{A_o} = \boldsymbol{A_c} - \boldsymbol{A}' = \begin{pmatrix} a_{11} - a'_{11} & a_{12} - a'_{12} & \cdots & a_{1M} - a'_{1M} \\ a_{21} - a'_{21} & a_{22} - a'_{22} & \cdots & a_{2M} - a'_{2M} \\ a_{31} - a'_{31} & a_{32} - a'_{32} & \cdots & a_{3M} - a'_{3M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} - a'_{N1} & a_{N2} - a'_{N2} & \cdots & a_{NM} - a'_{NM} \end{pmatrix} \in \mathbb{R}^{N\times M}, \tag{1}$$

where M is the number of snapshots (in the direction to be decomposed, without loss of generality), and N stands for the rest of the degrees of freedom (spatial grid points multiplied by the number of independent variables). In this work, we consider structural (a single or several locations) or gusty (a single or several snapshots) contaminations. In practical applications, POD is solved with singular value decomposition of $\boldsymbol{A_c}$, equivalent to the eigenvalue decomposition of $\boldsymbol{A_c}\boldsymbol{A_c}^T$,

$$\left(\boldsymbol{A_c}\boldsymbol{A_c}^T - \lambda_i \boldsymbol{I}\right)\phi_{\boldsymbol{i}} = 0. \tag{2}$$

The question is how accurate λ_i and $\phi_{\boldsymbol{i}}$ are as the singular value/vector of $\boldsymbol{A_o}$ or eigenvalue/vector of $\boldsymbol{A_o}\boldsymbol{A_o}^T$. Substituting Eq. (1) into Eq. (2) gives

$$\left(\boldsymbol{A_o}\boldsymbol{A_o}^T - \lambda_i \boldsymbol{I}\right)\phi_{\boldsymbol{i}} = -\left(\boldsymbol{A_c}\boldsymbol{A}'^T + \boldsymbol{A}'\boldsymbol{A_c}^T - \boldsymbol{A}'\boldsymbol{A}'^T\right)\phi_{\boldsymbol{i}}. \tag{3}$$

By considering the contamination $\boldsymbol{A}'$ to be of small magnitude ($||\boldsymbol{A}'||_F \ll ||\boldsymbol{A_c}||_F$) and ignoring higher order terms (equivalent to ignoring higher order correlations), the accuracy of λ_i and $\phi_{\boldsymbol{i}}$ as eigensolution for $\boldsymbol{A_o}\boldsymbol{A_o}^T$ can be evaluated as

$$\begin{aligned} \varepsilon &= \frac{\mathrm{E}\left[\left(\boldsymbol{A_o}\boldsymbol{A_o}^{\mathrm{T}} - \lambda_i \boldsymbol{I}\right)\phi_{\boldsymbol{i}}\right]^2}{||\boldsymbol{A_c}||_F^2} \\ &\approx \frac{\mathrm{E}\left|\left(\boldsymbol{A_c}\boldsymbol{A}'^{\mathrm{T}} + \boldsymbol{A}'\boldsymbol{A_c}^{\mathrm{T}}\right)\phi_{\boldsymbol{i}}\right|^2}{||\boldsymbol{A_c}||_F^2} \\ &= \frac{E\left[\left(\phi_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{A}'\boldsymbol{A_c}^T + \phi_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{A_c}\boldsymbol{A}'^T\right)\left(\boldsymbol{A_c}\boldsymbol{A}'^T\phi_{\boldsymbol{i}} + \boldsymbol{A}'\boldsymbol{A_c}^T\phi_{\boldsymbol{i}}\right)\right]}{||\boldsymbol{A_c}||_F^2}. \end{aligned} \tag{4}$$

Here, E stands for the expectation of random variables. The uncertainty has been normalized by the square of the Frobenius norm as follows:

$$||\boldsymbol{A_c}||_F^2 = \sum_{n=1}^{N}\sum_{m=1}^{M} a_{nm}^2. \tag{5}$$

Furthermore, we substitute the identity

$$\boldsymbol{A_c}\boldsymbol{A}'^T\phi_{\mathbf{i}} + \boldsymbol{A}'\boldsymbol{A_c}^T\phi_i = \sum_{n=1}^{N}\sum_{m=1}^{M}\begin{pmatrix} a_{1m}a'_{nm}\phi_{in} + a'_{1m}a_{nm}\phi_{in} \\ a_{2m}a'_{nm}\phi_{in} + a'_{2m}a_{nm}\phi_{in} \\ a_{3m}a'_{nm}\phi_{in} + a'_{3m}a_{nm}\phi_{in} \\ \vdots \\ a_{Nm}a'_{nm}\phi_{in} + a'_{Nm}a_{nm}\phi_{in} \end{pmatrix} \in \mathbb{R}^{N\times 1} \tag{6}$$

into Eq. (4) and obtain the uncertainty evaluation for mode $\phi_{\boldsymbol{i}}$ as follows:

$$\varepsilon = \frac{\sum_{s=1}^{N} E\left[\sum_{n=1}^{N}\sum_{m=1}^{M}\left(a'_{nm}a_{sm}\phi_{in} + a'_{sm}a_{nm}\phi_{in}\right)\right]^2}{\sum_{n=1}^{N}\sum_{m=1}^{M} a_{nm}^2}. \tag{7}$$

The assumption so far is that the contamination's amplitude is much smaller than the fluid data. Therefore, Eq. (7) shall hold for general conditions, and it is clear that the transmission of uncertainty is determined by the POD modes, the contaminated data, and the correlation of the contamination, of which the last will be determined by experimental tests or empirical models. To show the idea, we make use of the following self-correlation and cross correlation models partly based on the experimental observation of error and correlation distributions:[13]

$$\begin{aligned} E\left[a'^2_{nm}\right] &= \sigma_{nm}(\text{self-correlation}), \\ E\left[a'_{n_1m_1}a'_{n_2m_2}\right] &= \sigma_x(n_1,n_2)\sigma_t(m_1,m_2)(\text{cross-correlation}). \end{aligned} \tag{8}$$

The limitation with Eq. (8) is that the self-correlation has the same degree of freedom as the input data, raising modeling burdens. Moreover, the cross correlation requires external knowledge. In a simplified case, the self-correlation is prescribed as a constant $\sigma_{nm} = \sigma_{\text{self}}$ (with σ_{self} being a constant value). The cross correlation follows a linear decaying property as demonstrated in Fig. 1, although more complex models can fit in this framework. Both σ_x and σ_t decay from the value of σ_{self} with the "distance" of two points, measured with $|n_1 - n_2|$ and $|m_1 - m_2|$. Once the values exceed a prescribed correlation

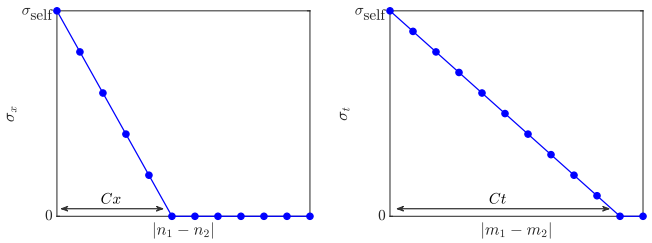

FIG. 1. Self-correlation and cross correlation models. The spatial and temporal correlation factors σ_x and σ_t linearly decay with the distance of two points. Both values drop from σ_{self} to zero.

width (C_x and C_t), the cross correlation becomes zero. In circumstances where the cross correlation ε_c (induced by σ_x and σ_t) plays a minor role, the uncertainty is determined by self-correlation, of which a physical interpretation presents

$$\varepsilon = \varepsilon_s + \varepsilon_c,$$

$$\varepsilon_s = \underbrace{\frac{\sum_{n=1}^{N}\sum_{m=1}^{M}\sigma_{\text{self}} a_{nm}^2 \phi_i^T \phi_i}{\sum_{n=1}^{N}\sum_{m=1}^{M} a_{nm}^2}}_{\varepsilon_{s1}} + \underbrace{\frac{(N+2)\sum_{m=1}^{M}\sigma_{\text{self}}\left(\mathbf{a}_m^T \phi_i\right)^2}{\sum_{n=1}^{N}\sum_{m=1}^{M} a_{nm}^2}}_{\varepsilon_{s2}}. \quad (9)$$

Here, $\mathbf{a}_m^T = (a_{1m}, a_{2m}, \ldots, a_{Nm})^T$ is a single snapshot of time step m. The first part ε_{s1} is invariant of the POD modes since $\phi_i^T \phi_i = 1$, and when σ_{self} is a constant, $\varepsilon_{s1} = \sigma_{\text{self}}$. ε_{s2} is dominated by the inner product of POD mode ϕ_i and the data snapshot $\mathbf{a}_m^T$, with the weighting parameter σ_{self}. Therefore, the similarity between the input data and the resulting POD modes becomes a critical and intuitive indicator to quantify the uncertainty.

III. RESULTS AND DISCUSSION

A. Interpretation of the uncertainty

To understand the composition of uncertainty, we consider a 1D traveling solitary wave,

$$a(x,t) = 100 \exp\left(-0.008(x - t - 50)^2\right). \quad (10)$$

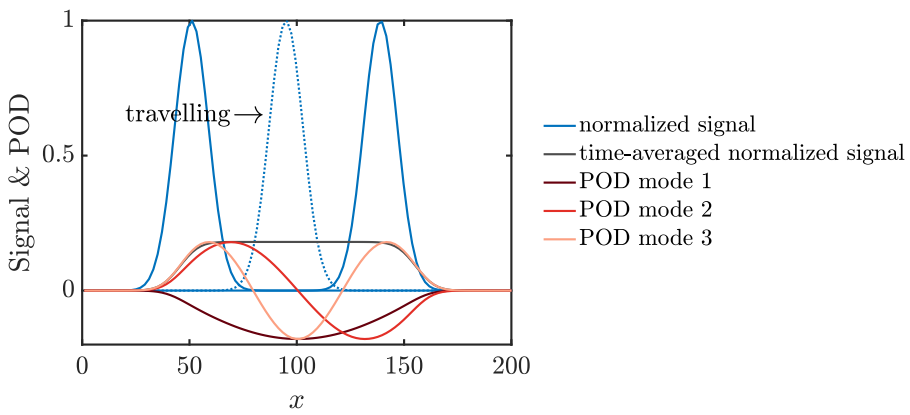

FIG. 2. The traveling Gaussian soliton and its first three POD modes.

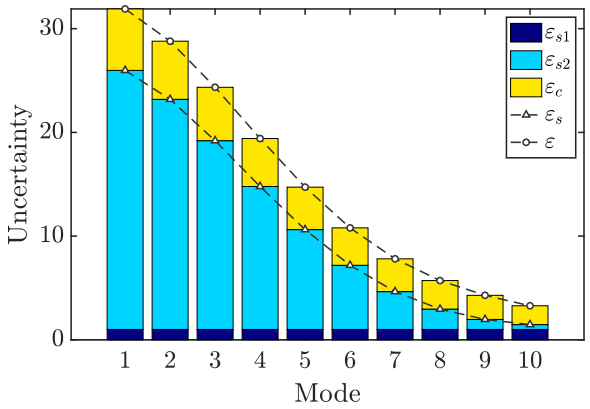

FIG. 3. Uncertainty of POD modes for the traveling Gaussian soliton. Terms shown are ε_{s1}, ε_{s2}, and ε_c of Eq. (9).

The problem is defined in a spatial and temporal domain of $x \in [1, 200]$ and $t \in [1, 100]$, where the grid points of $N = 100$ and $M = 10$ are linearly allocated. We present the signal (normalized by its maximum), the time-average, and the first three POD modes in Fig. 2. Due to the traveling nature of the input signal, its POD modes spontaneously approximate Fourier modes—the wavenumber increases with the order of the mode. Such properties create difficulty in effectively reducing the order of a model.[14]

To gain a first view of the uncertainty, we show results subject to a constant self-correlation factor $\sigma_{\text{self}} = 1$ (corresponding to 1% of the amplitude of the input signal) with cross correlation width $C_x = 5$ and $C_t = 1$ in Fig. 3. The form of C_x and C_t can be more complex and be estimated from the environment conditions of the experiment/simulation.[13] A decaying behavior of the uncertainty toward higher POD modes is found, the three parts of uncertainties (ε_{s1}, ε_{s2}, and ε_c) in Eqs. (7) and (9) are plotted for each POD mode. Comparison between the three terms shows this case is dominated by self-correlation, particularly ε_{s2}. Meanwhile, $\varepsilon_{s1} = \sigma_{\text{self}} = 1$. The monotonic decaying behavior is explained by Eq. (9): the inner product of the first POD mode with the signal (summation over time, equivalent to the mean square of the signal) is the largest and will decrease with mode number, as seen from the similarities of POD modes and time-averaged signal in Fig. 2.

Figure 3 shows that the cross correlation term followed a similar decaying trend. We investigate this term and the influence of C_x and C_t in Fig. 4. In either case, C_x or C_t is fixed to one, and the influence of the other parameter is revealed. The self-correlation is provided in the first row for comparison. As can be seen, the cross correlation increases with C_x or C_t and remains below self-correlation except for the first POD mode, indicating that the first mode (which contains the most energy) is the most affected regardless of the correlation width. The temporal and spatial correlation widths only positively contribute to the first few modes. For higher modes, a larger correlation width may decay the uncertainty. As shown in Sec. III C, the influence of correlation width is highly dependent on the correlation property of the input data. Thus, a general law for ε_c may not exist. It is worth commenting that the uncertainty formulation (7) holds universally as long as the contamination is small compared to the signal such that the formula (4) is valid. The constant self-correlation $\sigma_{nm} = \sigma_{\text{self}}$ used in Eq. (9) can also be relaxed to model more realistic conditions.

Figure 5 considers the circumstance when a single snapshot is contaminated, corresponding to problems that suffer from instantaneous perturbation at a particular time. Under this condition, the self-correlation is led by the inner product of a POD mode and the single contaminated data snapshot,

$$\varepsilon_s = \underbrace{\frac{\sum_{n=1}^{N}\sigma_{\text{self}}a_{nm_c}^2}{\sum_{n=1}^{N}\sum_{m=1}^{M}a_{nm}^2}}_{\varepsilon_{s1}} + \underbrace{\frac{(N+2)\sigma_{\text{self}}\left(\mathbf{a}_{m_c}^T\phi_i\right)^2}{\sum_{n=1}^{N}\sum_{m=1}^{M}a_{nm}^2}}_{\varepsilon_{s2}}. \tag{11}$$

Here, m_c is the time step that contamination happens. In contrast to the globally contaminated case reported in Fig. 4, the influence of individual snapshots that received the perturbation becomes essential. The amplitude of the uncertainty is thus determined by the inner product between $\mathbf{a}_{m_c}^T$ and ϕ_i, manifesting as the similarity of a POD mode and the contaminated snapshot. As shown in Fig. 5(a), the self-correlation is not monotonically decreasing with the mode number. The uncertainty of mode 1, for example, is maximum when the fifth and sixth snapshots are contaminated, at which the traveling wave locates near the center of the domain, bearing the most extensive similarity with mode 1. In other words, the median snapshot is the most critical in the traveling wave case as it is intuitively most similar to mode 1. Due to the linear nature of σ_{self} in Eq. (9), each column of Fig. 5(a) is actually a component of the self-correlation shown in Fig. 3. On the other hand, the cross correlation becomes solely spatial. We observe a similar trend as self-correlation.

B. Scaling of uncertainty

In the aspect of efficiency, one identifies that the computation amount scales with N^2M from Eq. (7). Thus, for problems with enormous degrees of freedom, the calculation becomes costly. This section seeks potential scaling properties of uncertainty such that the quantification can be performed on data interpolated to coarse grids.

As an example, we consider a 2D flow around a cylinder. The numerical details of the fluid data generation are not of direct relevance to the present research, and the data are publicly available online.[15] The instantaneous flow together with two of the POD modes are displayed in Fig. 6. The data consist of two velocity components: u and v on grid points of 129×129 ($N = 2 \times 129 \times 129$) with $M = 75$ timesteps.

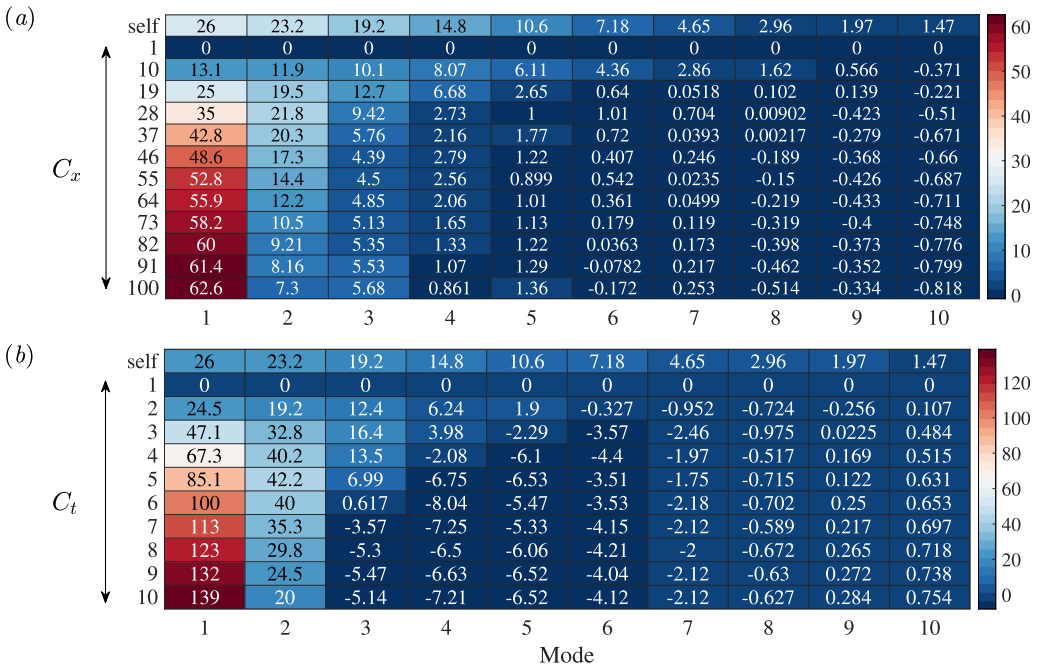

FIG. 4. Effects of the correlation width $C_x|_{C_t=1}$ (a) and $C_t|_{C_x=1}$ (b) on the cross correlation. The first row shows ε_s, while the rest is colored by ε_c.

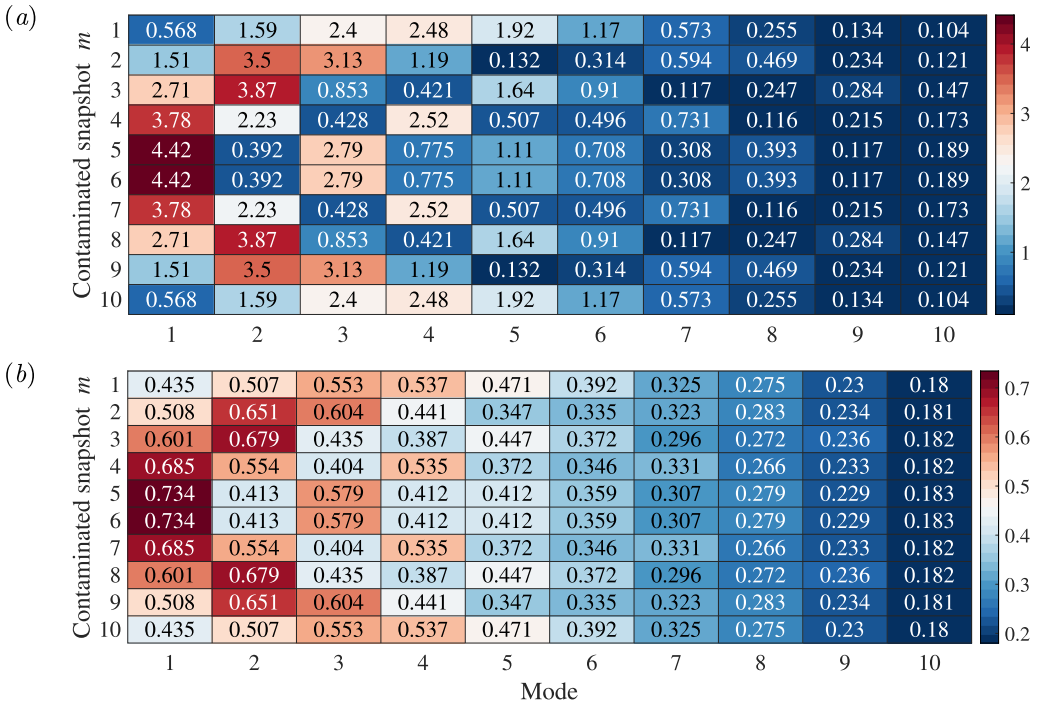

FIG. 5. Influence of a single snapshot contaminated on the uncertainty. Panels (a) and (b) are colored with ε_s and ε_c ($C_x = 5$), respectively.

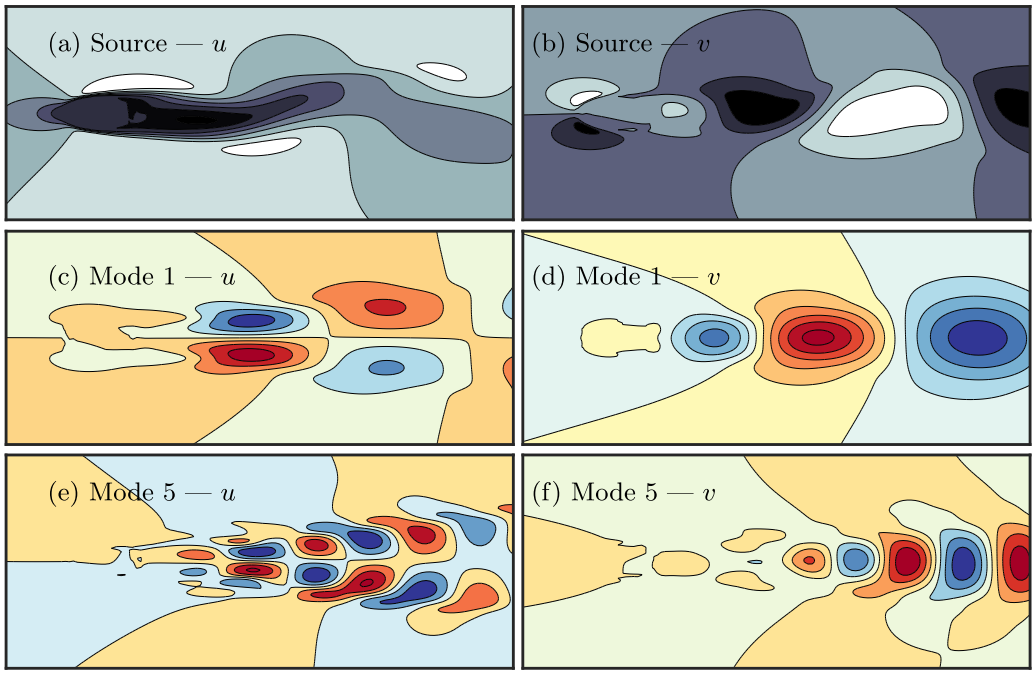

FIG. 6. 2D flow (a) and (b) past a cylinder together with its two POD modes (c)–(f).

We show the uncertainty in Fig. 7. The results are subject to global contamination and contamination only in a single snapshot, respectively. For this situation, it is clear that self-correlation will lead the uncertainty for lower modes regardless of the contamination form. These modes also have the most significant uncertainty that requires the most attention.

The scaling is observed by interpolating the data onto different meshes. We separately reduced the number of grids to 1/2, 1/3, and 1/4 in each direction. As can be seen in Fig. 8, the self-correlation term ε_{s1} remains the value of σ_{self} regardless of the grid used. ε_{s2} presents a scaling law with the number of grid points. This feature is explained by Eq. (9). From the expression, $\varepsilon_{s2} \propto (N+2)$. When the value of N is large and $\varepsilon_{s2} \gg \varepsilon_{s1}$, this scale approximately becomes $\varepsilon_s \propto N$. In this example, the cross correlation is much smaller, and the grid points do not notably influence ε_c of the first few modes. Therefore, it becomes feasible to quantify the uncertainty through coarse meshes to boost efficiency by exploiting the scaling property.

C. Generality of uncertainty

Previous examples have manifested a decaying trend of uncertainty toward higher modes, and the cross correlation is smaller than self-correlation. Is this property true for general data? Here, we evaluate the generality of uncertainty by considering two sets of random data. One is a random data matrix $\boldsymbol{A}_1 \in \mathbb{R}^{200\times 50}$ with the value of entries uniformly distributed between $[0, 1]$. The other is data matrix $\boldsymbol{A}_2 \in \mathbb{R}^{200\times 50}$ generated by a random combination of mathematical functions. These functions are chosen from a library of functions up to the fifth order, i.e., x, t, x^2, xt, t^2,…, t^5 as well as sine functions $\sin(x)$, $\sin(t)$, $\sin(xt)$,…. The generation of $\boldsymbol{A}_2$ follows the idea that a combination of these functions is capable of representing data created by common physical laws.[16] One, thus, treats $\boldsymbol{A}_1$ as intrinsically uncorrelated as opposed to $\boldsymbol{A}_2$. Both $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ have a sample size of 500.

Figure 9 investigates the dependence of self-correlation on various input data and the impact of wide-range correlation widths on cross correlation. Panels (a) and (c) indicate that the self-correlation follows the decaying trend regardless of the input data. This decay is profound for intrinsically correlated data, as seen from the log scale of (c). The self-correlation also appears insensitive to the random parameters that created the input data. Contrarily, the cross correlation behaves rather differently between $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. Panels (b) and (d) show results for $C_x = 10, 20, \ldots, 100$ and $C_t = 5, 10, 15, \ldots, 25$ for a random sample. The maximum correlation width in space and time each corresponds to half of the total grid points. For data group $\boldsymbol{A}_1$, cross correlation stays largely below the self-correlation. For each mode, the value does not necessarily increase with the correlation width. In $\boldsymbol{A}_2$, the cross correlation can dominate the uncertainty at a moderate correlation width since the data are intrinsically correlated. Thus, when considering the uncertainty of POD modes, the correlation property of the input data itself is important. For problems of intrinsically less correlated, e.g., traveling wave problems, the uncertainty is dominated by self-correlation, and a quick predicting scheme based on the scaling law is feasible.

Finally, we comment on the error for POD-based reduced-order models. As seen in Fig. 10, the error typically drops by keeping more

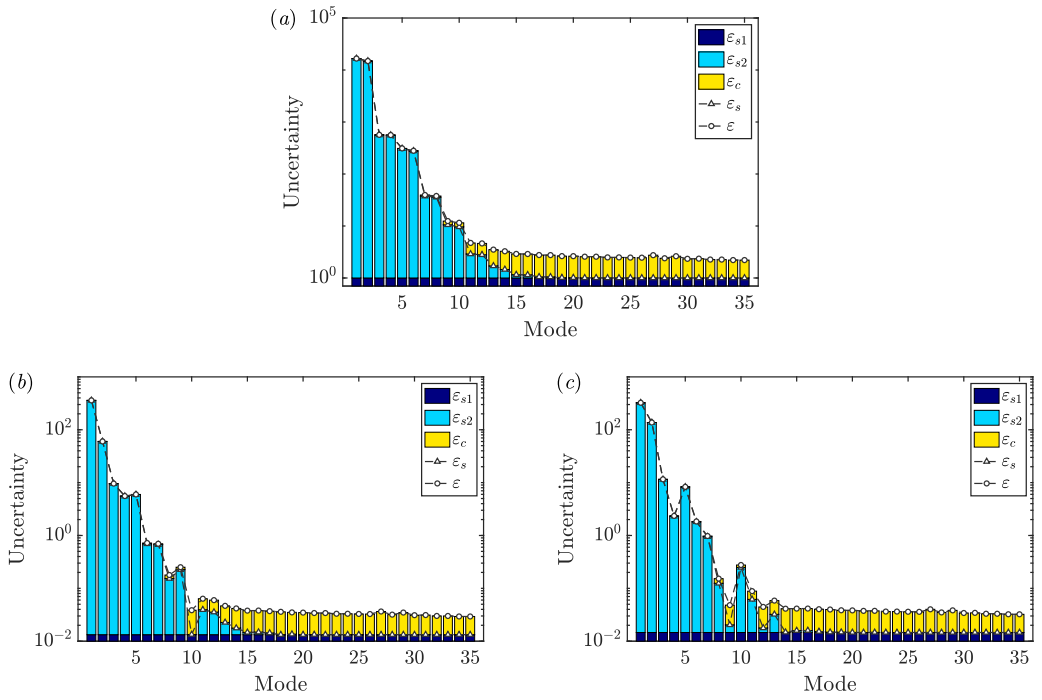

FIG. 7. Uncertainty transmission of the 2D cylinder flow with $C_x = 3$ and $C_t = 1$. The three panels show the results of global contamination (a), contamination at the first snapshot (b) and the tenth snapshot (c).

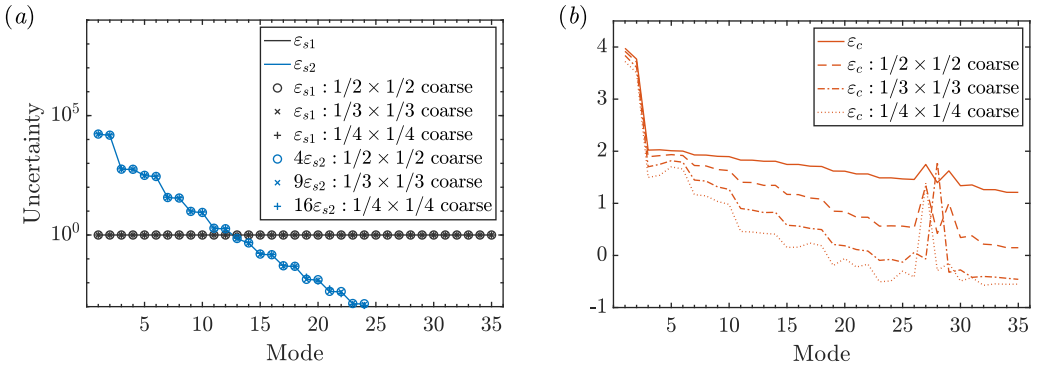

FIG. 8. Scaling of the uncertainty: (a) self-correlation and (b) cross correlation.

modes in building the reduced-order model. When uncertainty is considered, the error accumulated by the leading modes becomes considerable and not compensable. This leads to a cutoff point, above which including more modes will not reduce the error further. The higher the uncertainty, the lower the cutoff point shall be. Although the cutoff mode's position will also depend on the problem analyzed and the contamination behavior, understanding the uncertainty helps build the reduced-order models more effectively.

IV. CONCLUSIONS

This study formulates the uncertainty transmission of proper orthogonal decompositions. By evaluating the eigenvalue problem associated with POD, the uncertainty is expressed as the expectation of the residual term stemming from data contamination. Mathematically, the uncertainty contains two parts: self-correlation and cross correlation of the contamination terms. Based on experimental observations, we proposed a linear decay correlation model. Meanwhile, other models shall fit handily into this framework.

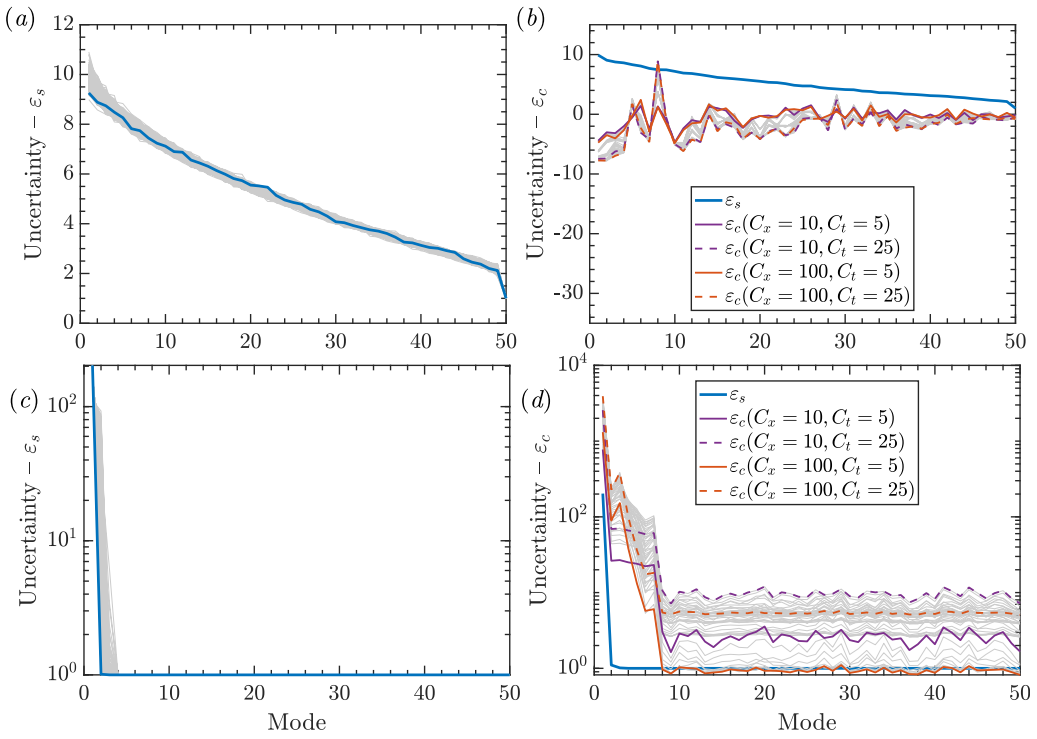

FIG. 9. Uncertainty for two sets of random data A_1 and A_2: (a) A_1, self-correlation; (b) A_1, cross correlation; (c) A_2, self-correlation; and (d) A_2, cross correlation. In panels (a) and (c), thick blue lines highlight the results for one sample, while the rest 499 are plotted in thin gray lines. In panels (b) and (d), cross correlations with four different correlation parameters are plotted in color as seen in the legend; the rest of the results are shown in gray. Self-correlation is plotted in blue for reference.

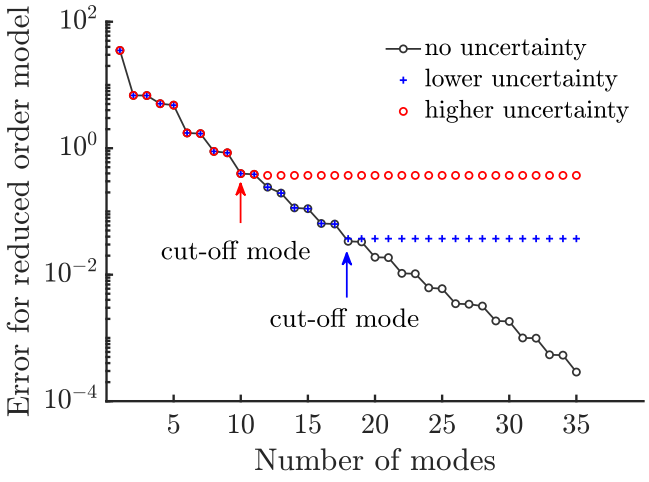

FIG. 10. Illustration of the error for reduced-order models considering uncertainty.

We examine the uncertainty transmission in three aspects. First, the self-correlation of the uncertainty has a clear physical definition: the inner product of the POD modes and the data matrix weighted by the self-correlation factor. Thus, the similarity between both terms, the input and output of POD, determines the amplitude of self-correlation. Cross-correlation, following the experimental observation of small spatial correlation width (3–6 points), may only have a minor influence in some cases. Second, by formulation, the self-correlation holds a scaling law with the degree of freedom of the data snapshot. This nature permits a fast prediction by evaluating the data on a coarse mesh. Third, we seek the generality of uncertainty distribution by investigating two sets of random data, with a total of 1000 samples. We observe a general decaying trend of uncertainty toward higher modes. This is due to the nature of POD that most energy resides in lower modes, whose shape thus matches more with the data bulk. The cross correlation has been examined for a wide range of correlation widths. For intrinsically correlated input data, the cross correlation can dominate the uncertainty at a moderate correlation width. On the contrary, the cross correlation can be neglected for less-correlated input, where a fast prediction can be realized utilizing the scaling properties.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from the National Natural Science Foundation of China (No. 92152109).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jie Ren: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (equal); Validation (lead); Writing – original draft (lead); Writing – review & editing (equal). **Xuerui Mao:** Conceptualization (equal); Funding acquisition (lead); Methodology (equal); Project administration (lead); Resources (lead); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

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

REFERENCES

[1]J. Lumley, "The structure of inhomogeneous turbulence," in *Atmospheric Turbulence and Wave Propagation*, edited by A. M. Yaglom and V. I. Tatarski (Nauka, 1967).
[2]G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," Annu. Rev. Fluid Mech. **25**, 539 (1993).
[3]K. Taira, S. L. Brunton, S. T. 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**, 4013 (2017).
[4]K. Taira, M. S. Hemati, S. L. Brunton, Y. Sun, K. Duraisamy, S. Bagheri, S. T. Dawson, and C.-A. Yeh, "Modal analysis of fluid flows: Applications and outlook," AIAA J. **58**, 998 (2020).
[5]J. R. Singler, "New POD error expressions, error bounds, and asymptotic results for reduced order models of parabolic PDEs," SIAM J. Numer. Anal. **52**, 852 (2014).
[6]H. Fareed and J. R. Singler, "Error analysis of an incremental proper orthogonal decomposition algorithm for PDE simulation data," J. Comput. Appl. Math. **368**, 112525 (2020).
[7]B. García-Archilla, J. Novo, and S. Rubino, "Error analysis of proper orthogonal decomposition data assimilation schemes with grad–div stabilization for the Navier–Stokes equations," J. Comput. Appl. Math. **411**, 114246 (2022).
[8]B. P. Epps and E. M. Krivitzky, "Singular value decomposition of noisy data: Mode corruption," Exp. Fluids **60**, 121 (2019).
[9]B. P. Epps and E. M. Krivitzky, "Singular value decomposition of noisy data: Noise filtering," Exp. Fluids **60**, 126 (2019).
[10]M. C. Brindise and P. P. Vlachos, "Proper orthogonal decomposition truncation method for data denoising and order reduction," Exp. Fluids **58**, 28 (2017).
[11]D. Duke, J. Soria, and D. Honnery, "An error analysis of the dynamic mode decomposition," Exp. Fluids **52**, 529 (2012).
[12]P. J. Roache, "Quantification of uncertainty in computational fluid dynamics," Annu. Rev. Fluid Mech. **29**, 123 (1997).
[13]J. A. Howell, "Distribution of particle image velocimetry (PIV) errors in a planar jet," Master's thesis (Utah State University, 2018).
[14]J. Ren, W. R. Wolf, and X. Mao, "Model reduction of traveling-wave problems via radon cumulative distribution transform," Phys. Rev. Fluids **6**, L082501 (2021).
[15]F. Zigunov, see https://www.mathworks.com/matlabcentral/fileexchange/72470-dynamic-mode-decomposition-dmd-wrapper for "Dynamic mode decomposition [DMD] – Wrapper"; accessed 5 June 2023.
[16]S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proc. Natl. Acad. Sci. **113**, 3932 (2016).