GAPPY SPECTRAL PROPER ORTHOGONAL DECOMPOSITION FOR RECONSTRUCTION OF TURBULENT FLOW DATA

Oliver T. Schmidt
Department of Mechanical and Aerospace Engineering
University of California San Diego
CA 92093, USA
oschmidt@eng.ucsd.edu

Akhil Nekkanti
Department of Mechanical and Aerospace Engineering
University of California San Diego
CA 92093, USA
aknekkan@eng.ucsd.edu

ABSTRACT

Flow fields obtained by time-resolved particle image velocimetry (PIV) and other experimental methods can contain gaps or other types of undesired artifacts. To reconstruct flow data in the compromised or missing regions, we propose a data completion method based on spectral proper orthogonal decomposition (SPOD). The proposed approach leverages the temporal correlation of the SPOD modes with preceding and succeeding snapshots, their spatial correlation with the surrounding data in the field of view. For each gap, the algorithm first computes the SPOD of the remaining, unaffected data. In the next step, the compromised data is projected onto the basis of the SPOD modes. This corresponds to a local inversion of the SPOD problem and yields expansion coefficients that permit the reconstruction of the missing data. This local reconstruction is successively applied to each gap. After all gaps are filled in, the procedure is repeated in an iterative manner until convergence. The method is demonstrated on time-resolved PIV data of turbulent cavity flow obtained by Zhang *et al.* (2019) with randomly added gaps that correspond to 1%, 5%, and 20% of data loss.

INTRODUCTION

Gappy data reconstruction techniques find wide use in the completion of partially missing or otherwise compromised experimental data. Data from particle image velocimetry (PIV), the focus of this work, for example, can exhibit artifacts or gaps due to the obstruction of the light path by objects, reflections of light from the surface of objects, the inaccessibility of regions for the imaging system, irregular seeding, and other sources. Standard mathematical tools for gappy data reconstruction include basic interpolation (Yates, 1933) and least-square estimation. Techniques that were devised specifically for the task of gappy data reconstruction include optimal interpolation (Reynolds & Smith, 1994) and Kriging (Oliver & Webster, 1990).

The use of proper orthogonal decomposition in conjunction with least-squares estimation for data reconstruction was proposed by Everson & Sirovich (1995). This original gappy POD algorithm was later extended by Venturi & Karniadakis (2004) and shown to outperforms Kriging for the reconstruction of cylinder flow with up to 50% of missing data. Gappy POD has since become an essential component of a number of model reduction methods that use POD modes as their basis (Chaturantabut & Sorensen, 2010; Benner *et al.*, 2015). In ocean sciences, a similar method was independently developed by Beckers & Rixen (2003).

Spectral proper orthogonal decomposition (SPOD) leverages the temporal homogeneity of statistically stationary processes to compute modes that are perfectly correlated in both space and time. The use of this frequency-domain version of POD, which has recently been popularized by Towne *et al.* (2018) and Schmidt *et al.* (2018), for gappy data reconstruction is at the center of this study. The proposed algorithm is fundamentally different from those of Everson & Sirovich (1995) and Venturi & Karniadakis (2004), and found capable of recovering large sections of missing data in a long time series of the turbulent flow over an open cavity.

Time-resolved particle image velocimetry (TR-PIV) was performed to obtain the streamwise velocity field in the center plane of the Mach 0.6, turbulent flow over an open cavity with a length-to-depth ratio of $L/D = 6$ and a width-to-depth ratio of $W/D = 3.85$. The sampling rate was 16 kHz and a total number of 16,000 image pairs were acquired to compute the velocity vector field. We refer to Zhang *et al.* (2019) and Zhang *et al.* (2017) for more details on this specific measurement campaign and the experimental setup, respectively.

SPECTRAL PROPER ORTHOGONAL DECOMPOSITION

A detailed theoretical discussions of SPOD theory and best practices for its applications can be found in Towne *et al.* (2018) and Schmidt & Colonius (2020), respectively. We provide an outline of a specific procedure of computing the SPOD based on Welch's method (Welch, 1967). Given a fluctuating flow field $\mathbf{q}_i = \mathbf{q}(t_i)$, where $i = 1, 2, \cdots, n_t$, that is obtained by subtracting the temporal mean $\bar{\mathbf{q}}$ from each snapshot of the data, the first step of the standard Welch approach is to seg-

ment the data into n_{blk} overlapping blocks, each containing n_{fft} snapshots. Next, we compute a windowed temporal discrete Fourier transform and arrange all the Fourier realizations at the l-th frequency, $\hat{\mathbf{q}}_l^{(j)}$, into a matrix,

$$\hat{\mathbf{Q}}_l = \left[\hat{\mathbf{q}}_l^{(1)}, \hat{\mathbf{q}}_l^{(2)}, \cdots, \hat{\mathbf{q}}_l^{(n_{\text{blk}})} \right]. \tag{1}$$

The SPOD modes, Φ_l, and associated energies, λ_l, can be computed as the eigenvectors and eigenvalues of the CSD matrix $\mathbf{S}_l = \frac{1}{n_{\text{blk}}} \hat{\mathbf{Q}}_l \hat{\mathbf{Q}}_l^* \mathbf{W}$, where $\mathbf{W}$ is a positive-definite Hermitian matrix that accounts for the component-wise and numerical quadrature weights. In practice, the number of spatial degrees of freedom is often much larger than the number of realizations. In that case, it is more economical to solve the analogous eigenvalue problem

$$\frac{1}{n_{\text{blk}}} \hat{\mathbf{Q}}_l^* \mathbf{W} \hat{\mathbf{Q}}_l \Psi_l = \Psi_l \Lambda_l \tag{2}$$

for the coefficients ψ that expand the SPOD modes in terms of the Fourier realizations. In terms of the column matrix $\Psi_l = [\psi_l^{(1)}, \psi_l^{(2)}, \cdots \psi_l^{(n_{\text{blk}})}]$, the SPOD modes at the l-th frequency are recovered as

$$\Phi_l = \frac{1}{\sqrt{n_{\text{blk}}}} \hat{\mathbf{Q}}_l \Psi_l \Lambda_l^{-1/2}. \tag{3}$$

The matrices $\Lambda_l = \text{diag}(\lambda_l^{(1)}, \lambda_l^{(2)}, \cdots, \lambda_l^{(n_{\text{blk}})})$, where by convention $\lambda_l^{(1)} \geq \lambda_l^{(2)} \geq \cdots \geq \lambda_l^{(n_{\text{blk}})}$, and $\Phi_l = [\phi_l^{(1)}, \phi_l^{(2)}, \cdots, \phi_l^{(n_{\text{blk}})}]$ contain the SPOD energies and modes, respectively. An important property of the SPOD modes is their orthogonality in their weighted inner product, $\langle \phi_l^{(i)}, \phi_l^{(j)} \rangle = \phi_l^{(i)} \mathbf{W} \phi_l^{(j)} = \delta_{ij}$. The associated norm is denoted by $\| \cdot \|_2$.

Data reconstruction

The cornerstone of the present work is the reconstruction of the original data from its SPOD. This inversion of the SPOD is discussed by Nekkanti & Schmidt (2021) in the context of frequency-time analysis, and different applications that involve partial reconstructions including filtering and denoising. The Fourier realizations at each frequency are reconstructed from the SPOD modes as $\hat{\mathbf{Q}}_l = \Phi_l \mathbf{A}_l$. Here, $\mathbf{A}_l$ is the matrix of the (scaled) expansion coefficients computed as

$$\mathbf{A}_l = \sqrt{n_{\text{blk}}} \Lambda_l^{1/2} \Psi_l^*, \quad \text{or} \tag{4}$$

$$\mathbf{A}_l = \Phi_l^* \mathbf{W} \hat{\mathbf{Q}}_l. \tag{5}$$

The expansion coefficients can be saved during the computation of SPOD using equation (4) or can be recovered later by projecting the Fourier realizations onto the modes using equation (5). Using the expansion coefficients a_{ik} contained in $\mathbf{A}$ at any given frequency, the k-th block can be reconstructed as

$$\mathbf{Q}^{(k)} = \mathscr{F}^{-1} \left[\left(\sum_i a_{ik} \phi^{(i)} \right)_{l=1}, \left(\sum_i a_{ik} \phi^{(i)} \right)_{l=2}, \cdots, \left(\sum_i a_{ik} \phi^{(i)} \right)_{l=n_{\text{fft}}} \right], \tag{6}$$

where $\mathscr{F}^{-1}$ is inverse windowed Fourier transform. Finally, the time series is reconstructed from the data segments by computing the average of the reconstructions from overlapping blocks, weighted by the relative value of their windowing function (Nekkanti & Schmidt, 2021).

ALGORITHM: GAPPY SPOD

(i) Segment the time series into overlapping blocks and compute the temporal Fourier transform of each block (if not computed in the previous iteration).

(ii) Proceed to the n-th gap and choose all the realizations of the Fourier transform that are *not* affected by this gap.

(iii) Compute the SPOD from all the Fourier realizations that are not affected by this gap (equations (2) and (3)) and save the corresponding expansion coefficients (equation (4)).

(iv) Compute the SPOD expansion coefficients for blocks *affected* by the n-th gap by projecting their Fourier transforms onto the SPOD basis (equation (5)).

(v) Reconstruct the affected blocks by inverting the SPOD (equation (6)) from the expansion coefficients computed in (iii) and (iv); replace the regions affected by the n-th gap.

(vi) (local loop) Go to (iv) to update the expansion coefficients now that the data is reconstructed in the affected regions until convergence criterion based on change of reconstruction of n-th gap is met.

(vii) (inner loop) Set $n \leftarrow n+1$ and go to (i) until all gaps are reconstructed.

(viii) (outer loop) Set $n = 1$ and go to (i); repeat until convergence criterion based on change of reconstruction between outer loop iterations is met.

Error and Convergence metrics

Define G as the index set corresponding to all gappy snapshots and $G_n \subset G$ as the subset of indices corresponding to the n-th gap. The following error and convergence metrics are used to evaluate the efficacy of our method:

$$e_i = \frac{\left\| \mathbf{q}_i - \tilde{\mathbf{q}}_i \right\|_2^2}{\left\| \mathbf{q}_i \right\|_2^2} \quad \text{(relative error of } i\text{-th snapshot)}, \tag{7}$$

$$e_n = \sum_{i \in G_n} e_i \quad \text{(relative error of } n\text{-th gap)}, \tag{8}$$

$$E = \sum_{i \in G_n} e_i \quad \text{(global relative error)}, \tag{9}$$

$$c_n = \sum_{i \in G_n} \frac{\left\| \tilde{\mathbf{q}}_i^{[j-1]} - \tilde{\mathbf{q}}_i^{[j]} \right\|_2^2}{\left\| \tilde{\mathbf{q}}_i^{[j-1]} \right\|_2^2} \quad \text{(convergence of } n\text{-th gap)}. \tag{10}$$

Here, $\mathbf{q}$ is the original data without gaps, $\tilde{\mathbf{q}}$ the reconstructed data, and superscript $[j]$ the iteration index. The calculation of the relative errors, equations (8) and (9), requires knowledge of the original data, $\mathbf{q}$. For demonstration purposes only, artificial gaps were introduced in this study, and hence the relative errors can be computed. The convergence metric defined in equation (10) does not require the true data and can be evaluated even for gappy datasets. Here, the threshold used for convergence criteria is $tol = 10^{-8}$.
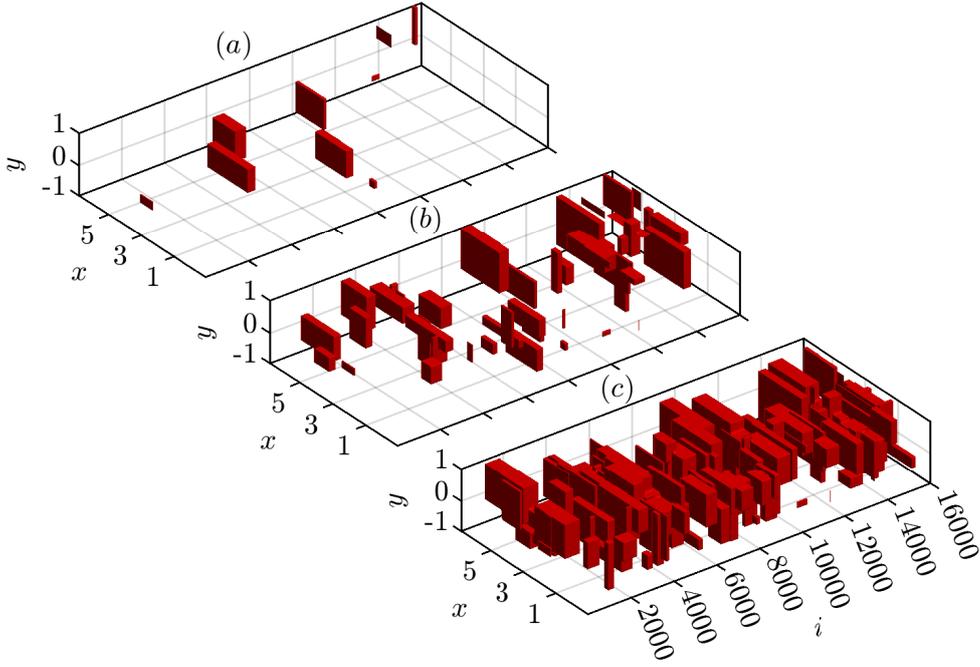
Figure 1.   Randomly selected gappy regions for the turbulent flow over an open cavity: (*a*) 1% gappyness; (*b*) 5% gappyness; (*c*) 20% gappyness. Red blocks indicate gaps. The streamwise (*x*-*y*) plane is plotted over the snapshot index, *i*.
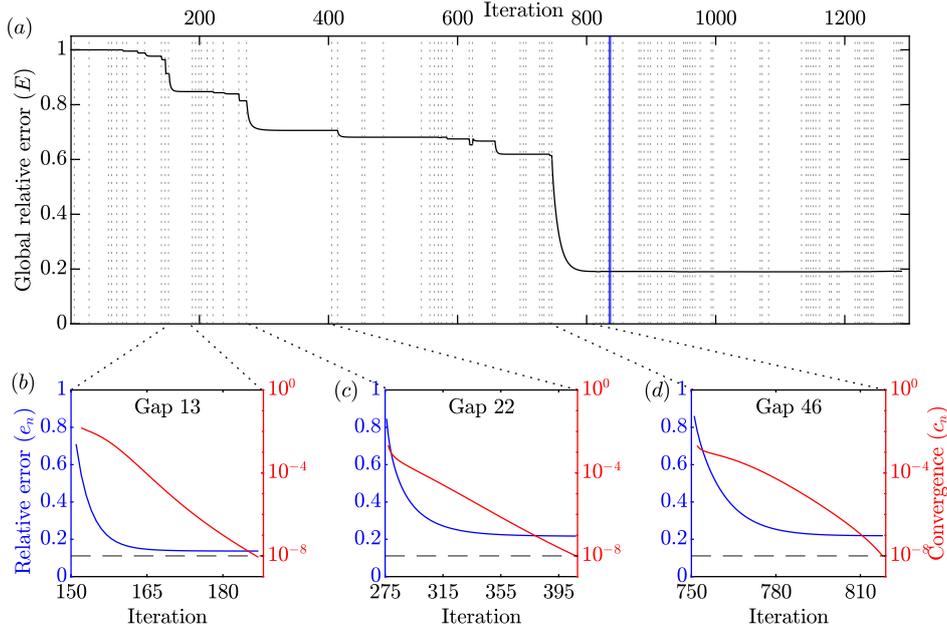
Figure 2.   Errors and convergence for 5% gappyness, see figure 1(*b*). The global relative error is shown in (*a*). Inner (gap-wise) iterations are denoted by grey dotted lines and outer iterations as blue solid lines. Panels (*b*-*d*) show the gap-wise error and convergence for three randomly selected gaps within the first outer loop.

RESULTS

Figure 1 illustrates the three test cases of randomly generated gaps accounting for 1%, 5%, and 20% of missing data, respectively. The gaps are randomly seeded in space. Similarly, the spatial extend and duration of each gaps is randomly sampled. Gaps are allowed to extend over the entire field-of-view and up to 600 snapshots. It is clear from figure **??**(c) that every block used for the spectral estimation of the SPOD contains missing data for the 20% case. Following best practices (Schmidt & Colonius, 2020), we segment the data into 124 blocks, each block spanning 256 snapshots, and with an

overlap of 50% between snapshots.

We start by exploring the 5% gappyness case, previously shown in figure 1(*b*). This case consists of $n_{\text{gaps}} = 50$ randomly seeded and sized gaps. Figure 2 illustrates the local and global errors and the convergence, of the algorithm. By *local*, we refer to the gap-wise iteration, that is, steps (iv)-(vi) of the algorithm, and by *global* to the outer iteration loop over steps (i)-(viii) of the algorithm. Figure 2(*a*) shows the global relative error as defined in equation (9) to converge to $c_n \leq tol = 1 \times 10^{-8}$ for all gaps. The gap-wise convergence, c_n, is defined by equation (10). The algorithm requires two

3
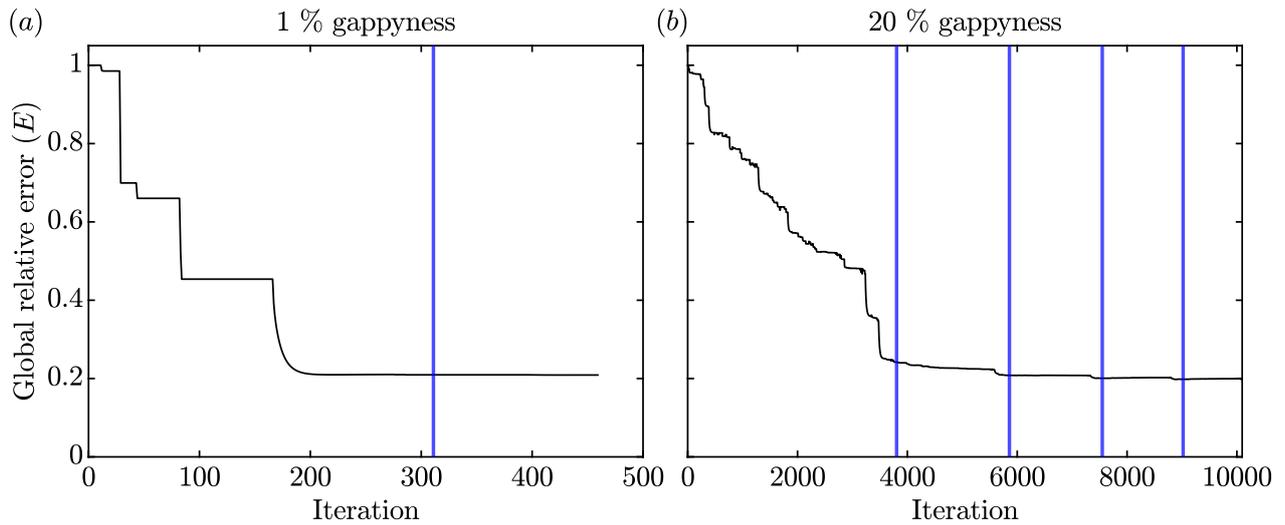
Figure 3. Global relative error of the gappy SPOD algorithm: (*a*) 1% gappyness; (*b*) 20% gappyness. The blue solid lines indicate the outer iterations.

outer loops, as denoted by the thick blue lines, to reach global convergence. Each of these outer iterations comprise n_{gaps} inner loops, indicated by grey dotted lines. The global relative error is normalized according to equation (9), and hence starts from unity. It can be observed that it decreases as the gaps are sequentially filled in. The amount by which each inner loop can reduce the global relative error is dependent on the spatial location and temporal extent of the corresponding gap. This explains the sharp drops at the beginning of some local iteration loops. At the end of the first outer loop, the global relative error reduced by 81%. The second outer loop does not further reduce the error by an appreciable amount and its final value hence is $E = 19\%$. Notably, most of the reconstruction error reduction is accomplished by the end of the first outer loop.

Figure 2(*b-d*) shows the local relative errors and convergence for three representative gaps during the first outer loop. The local relative error for all gaps decreases monotonically from unity to about 20%. An important observation is that the gap-wise relative error, in all cases, saturates before the convergence criterion of $tol = 10^{-8}$ is met.

To quantify the reconstruction for 1% and 20% of missing data cases, we report in figure 3 the global relative error of the reconstruction as a function of the iteration count. The error is measured in the same norm that is used for the SPOD and normalizing by the initial error. Solid blue lines indicate outer iterations and these two cases require two and five outer iterations, respectively. For the two cases, the relative errors are reduced to $E = 20.3\%$ and 19.4%, respectively. An intuitive feel for the quality of the reconstruction can be obtained by relating these numbers to the instantaneous flow fields shown and discussed in the context of figure 4.

The reconstructions of the data with 20% of missing data are compared to the original data in figure 4. Four randomly selected time instances are shown. An extreme example is the gap highlighted in figure 4(*a*). It is apparent that a large part of the field-of-view is missing and an animation of the snapshots in the vicinity of the gap further confirms that it persists over a long time. A direct comparison between the reconstructed field-of-view in figure 4(*e*) and the original data in 4(*i*) indicates that the gappy SPOD algorithm was able to fill in the missing regions with flow structures that to a large degree resemble the missing data. Note that the reconstruction has natu-

ral limits that are determined by the physical correlation length and time scales of the turbulent flow beyond which a reconstruction is impossible. Comparing the reconstructed to the original data for the three remaining time instances confirms that the algorithm is capable of estimating intricate details of the flow.

SUMMARY AND CONCLUSIONS

An algorithm that exploits the orthogonality and coherence, in space and time, of SPOD to reconstruct missing regions of a flow field is demonstrated on turbulent cavity flow PIV data. The algorithm is applicable to spatio-temporal data of stationary flows. It was able to reconstruct about 80% of the missing energy in randomly seeded and sized gaps that amount to 1%, 5%, and 20% of missing data. The algorithm converges almost monotonically for the data and type of gaps investigated. As user inputs, the gappy SPOD algorithm requires convergence tolerances for the local, gap-wise, and global iterations and the SPOD spectral estimation parameters. A visual inspection confirmed that the reconstructed flow fields resemble the original data closely.

Building on these promising results, we next plan to compare gappy SPOD to competing methods like gappy POD and Kriging, investigate the effect of the SPOD spectral estimation parameters, test the performance in the presence of missing snapshots, and study the dependence of the reconstruction on the convergence of the SPOD modes.

REFERENCES

Beckers, Jean-Marie & Rixen, M 2003 Eof calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and oceanic technology* **20** (12), 1839–1856.

Benner, P., Gugercin, S. & Willcox, K. 2015 A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review* **57** (4), 483–531.
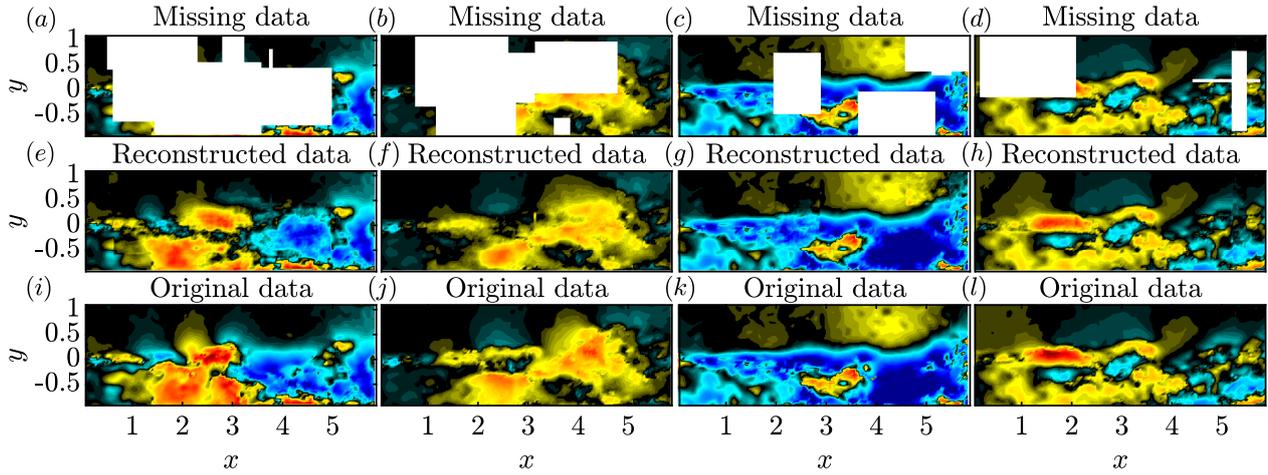
Figure 4. Reconstruction for the turbulent cavity flow with 20% gappyness at four time instances: (*a-d*) gappy data; (*e-h*) reconstructed data; (*i-l*) original data. False colors of the streamwise velocity fluctuations, *u*, are on the same scale in all plots.

Chaturantabut, S. & Sorensen, D. C. 2010 Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing* **32** (5), 2737–2764.

Everson, Richard & Sirovich, Lawrence 1995 Karhunen–loeve procedure for gappy data. *JOSA A* **12** (8), 1657–1664.

Nekkanti, A. & Schmidt, O. T. 2021 Frequency–time analysis, low-rank reconstruction and denoising of turbulent flows using spod. *Journal of Fluid Mechanics* **926**, A26.

Oliver, Margaret A & Webster, Richard 1990 Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System* **4** (3), 313–332.

Reynolds, Richard W & Smith, Thomas M 1994 Improved global sea surface temperature analyses using optimum interpolation. *Journal of climate* **7** (6), 929–948.

Schmidt, Oliver T & Colonius, Tim 2020 Guide to spectral proper orthogonal decomposition. *AIAA Journal* **58** (3), 1023–1033.

Schmidt, Oliver T, Towne, Aaron, Rigas, Georgios, Colonius, Tim & Brès, Guillaume A 2018 Spectral analysis of jet turbulence. *Journal of Fluid Mechanics* **855**, 953–982.

Towne, Aaron, Schmidt, Oliver T & Colonius, Tim 2018 Spec-

tral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics* **847**, 821–867.

Venturi, Daniele & Karniadakis, George Em 2004 Gappy data and reconstruction procedures for flow past a cylinder. *Journal of Fluid Mechanics* **519**, 315–336.

Welch, P. 1967 The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* **15** (2), 70–73.

Yates, Frank 1933 The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture* **1** (2), 129–142.

Zhang, Y., Cattafesta, L. & Ukeiley, L. 2017 Identification of coherent structures in cavity flows using stochastic estimation and dynamic mode decomposition. In *10th International Symposium on Turbulence and Shear Flow Phenomena*, p. 3.

Zhang, Y., Cattafesta, L. & Ukeiley, L. 2019 A spectral analysis modal method applied to cavity flow oscillations. In *11th International Symposium on Turbulence and Shear Flow Phenomena*.