15th International Symposium on Particle Image Velocimetry – ISPIV 2023
June 19–21, 2023, San Diego, California, USA

Gappy SPOD for TR-PIV data of stationary flows

O. T. Schmidt[1*], E. Brothers[1], A. Nekkanti[1], Y. Zhang[2], L. N. Cattafesta[2]

[1] University of California San Diego, Department of Mechanical and Aerospace Engineering, La Jolla, CA, USA

[2] Illinois Institute of Technology, Mechanical, Materials, and Aerospace Engineering Department, Chicago, IL, USA

[*] oschmidt@ucsd.edu

Abstract

Flow fields obtained by time-resolved particle image velocimetry (PIV) and other experimental methods can contain gaps or undesired artifacts. We extend and simplify the data completion method based on spectral proper orthogonal decomposition (SPOD) first proposed by Nekkanti and Schmidt (2023) to reconstruct flow data in the compromised or missing regions. The proposed approach leverages the temporal correlation of the SPOD modes with preceding and succeeding snapshots and their spatial correlation with the surrounding data in the field of view. The algorithm relies on the same data segmentation strategy used to compute the SPOD. For each block of the segmented data that contains gaps, it computes an SPOD basis from the subset of the data least affected by the same gaps. The compromised block is then projected onto the basis of 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 block. Successive sweeps over all blocks iteratively converge the reconstruction. In the final step, missing black zones are reconstructed from interpolated or extrapolated SPOD modes. The method is demonstrated on two time-resolved PIV datasets: turbulent open cavity flow Zhang et al. (2019) and flow between the slat and main airfoil of a deployed high-lift system (Zhang et al., 2020).

1 Introduction

Gappy data reconstruction techniques are widely used to complete 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 devised specifically for gappy data reconstruction include optimal interpolation (Reynolds and Smith, 1994) and Kriging (Oliver and Webster, 1990).

The use of proper orthogonal decomposition (POD) in conjunction with least-squares estimation for data reconstruction was proposed by Everson and Sirovich (1995). This original gappy POD (GPOD) algorithm was later extended by Venturi and Karniadakis (2004) and shown to outperform Kriging for the reconstruction of cylinder flow with up to 50% of missing data. GPOD has since become an essential component of a number of model reduction methods that use POD modes as their basis (Chaturantabut and Sorensen, 2010; Benner et al., 2015). In ocean sciences, a similar method was independently developed by Beckers and Rixen (2003). The current state-of-the-art in GSPOD is reflected in the median-filter GPOD (MF GPOD) algorithm by Saini et al. (2016), which determines whether to update a local missing data point after each iteration in an adaptive manner.

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. This frequency-domain version of POD 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 and Sirovich (1995) and Venturi and 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 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. The second PIV dataset is the flow between the deployed slat and the main airfoil of a high-lift system. Details for this case are reported in (Zhang et al., 2020).

This work extends and significantly simplifies the gappy SPOD algorithm developed by Nekkanti and Schmidt (2023). In particular, we propose a block-wise reconstruction strategy that largely renders the compute time independent of the number of gaps. This significantly speeds up the data reconstruction for realistic PIV data like the two datasets considered here.

2 Spectral Proper Orthogonal Decomposition

A detailed theoretical discussion of SPOD theory and best practices for its applications can be found in Towne et al. (2018) and Schmidt and Colonius (2020), respectively. We provide an outline of a specific procedure for computing the SPOD based on Welch's method (Welch, 1967). Unlike the original algorithm by Nekkanti and Schmidt (2023) that reconstructs missing data one gap at a time, the new algorithm uses the same segmentation strategy as the Welch spectral estimator. Given a flow field $\mathbf{q}_i = \mathbf{q}(t_i)$, where $i = 1, 2, \cdots, n_t$, the first step of the standard Welch approach is to segment the data into n_{blk} overlapping blocks, each containing n_{fft} snapshots. Usually, SPOD is computed from the fluctuating flow field obtained by subtracting the temporal mean $\bar{\mathbf{q}}$ from each snapshot to center the data. This is not required for the present application and leads to a simpler algorithm. After segmenting the data, 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, $\left\langle \phi_l^{(i)}, \phi_l^{(j)} \right\rangle = \phi_l^{(i)}\mathbf{W}\phi_l^{(j)} = \delta_{ij}$. The associated norm is denoted by $\|\cdot\|_2$.

2.1 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 and 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

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

where $\mathcal{F}^{-1}$ 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 and Schmidt, 2021).

New algorithm: Gappy SPOD (block-wise)

The new gappy SPOD algorithm differs from the original method introduced in Nekkanti and Schmidt (2023) in that it operates block-wise instead of gap-wise. This makes the algorithm more flexible, easier to implement and leads to a significant speed-up for data with many gaps. It also avoids the necessity to identify individual gaps. At a high level, the algorithm consists of the following steps:

(i) Segment the time series into overlapping blocks and compute the temporal Fourier transform of each block.

(ii) Sort FFT blocks by the percentage of gap overlap with the n-th block and keep those with fractions below $gap_{\text{ovlp,max}}$. Retain at least r_{min} blocks.

(iii) Compute the SPOD using equation (2) from the selected FFT blocks. Restrict the SPOD basis to r_{max} modes per frequency.

(iv) Compute the expansion coefficients for n-th block by projecting its FFT onto the new SPOD basis using equation (5).

(v) Reconstruct the n-th block by inverting the SPOD using the projection-based expansion coefficients in equation (6). Then replace the data in the missing regions with the reconstruction.

(vi) (Inner loop) Update the FFT of the n-th block and go to (iv) until the relative change between the current and previous inner iteration, in the L2 sense, falls below the tolerance, tol_{inner}, or until the maximum number of allowed inner iterations, $n_{\text{inner,max}}$, is reached.

(vii) Let $n \leftarrow n+1$ and go to (ii) until all blocks are reconstructed.

(viii) (Outer loop) Let $n \leftarrow 1$ and go to (i) for n_{outer} times.

This new version of the algorithm optionally extrapolates to black zones. Nearest neighbor interpolation is used to initialize the main algorithm and the optional extrapolation. If extrapolation to black zones is selected, the weight matrix $\mathbf{W}$ in equation (2) is set to zero in the corresponding regions. This ensures that only correlation information obtained from actual PIV data are used for the reconstruction. We have reconstructed the streamwise and wall-normal velocity components for the present study separately, but note that a synchronous reconstruction based on a compound solution vector, $[u\ v]^T$, is equally possible.

3 Results

The new version of the gappy SPOD algorithm is demonstrated on the two PIV datasets summarized in table 1. For the open cavity data, 4% of the field of view (FOV) are black zones, defined as regions in which no data are available at all times. These black zones occur close to the FOV borders. The overall gappyness, not including black zones, is somewhat below 2%. For the high-lift case, about 40% of the recorded FOV is usable; the remaining 60% are black zones. Within the usable FOV, the gappyness is about 5%. The open cavity and high-lift device results are presented in §3.1 and §3.2, respectively. Both discussions conclude with a standard SPOD analysis of the reconstructed data, highlighting each case's most prominent flow features.

case	n_t	n_x	n_y	Δt	n_{fft}	n_{ovlp}	n_{blk}	n_{outer}	tol_{inner}	$n_{\text{inner,max}}$	$gap_{\text{ovlp,max}}$	r_{min}	r_{max}
cavity	16,000	160	57	$6.25 \cdot 10^{-5}$s	128	64	249	2	10^{-2}	10	20%	20	10
high-lift	11,000	238	145	$9 \cdot 10^{-5}$s	128	64	170	2	10^{-2}	10	20%	20	10

Table 1: PIV and gappy SPOD parameters.

3.1 Open cavity flow

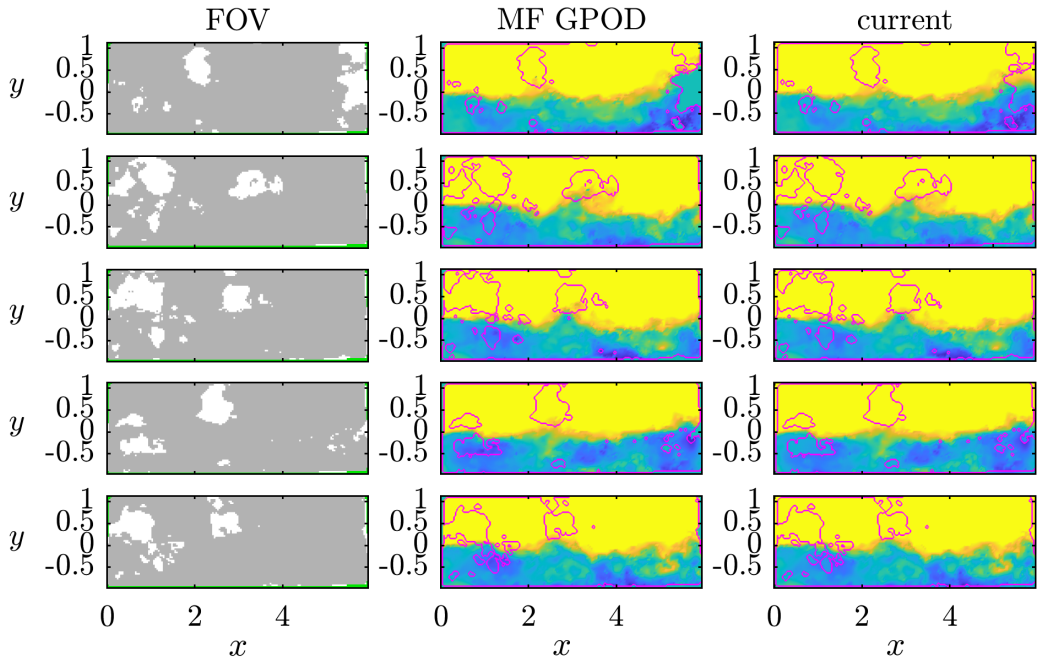

Figure 1: Streamwise velocity u at time instances with high gappyness in the shear-layer region. Most of the FOV (left column) is usable PIV data (■, gray), but it also contains gaps (□, white) and black zones (■, green). The middle and right columns show the MF GPOD and gappy SPOD reconstructions, respectively. The gaps in the original data (——, magenta) are outlined in both. The color map is saturated at ± 140 [m/s] (■ $u \geq 140$, ■ $u \leq -140$).

Figure 1 shows the FOV and compares the data reconstructions obtained using MF GPOD (Saini et al., 2016) and our new algorithm outlined above. The five snapshots with the highest gappyness in the shear-layer region, $-0.5 \leq y \leq 0.5$, are examples. This restriction focuses the analysis on the dynamically most interesting region (the largest gaps in the entire FOV occur outside the cavity near the upper edge of the FOV, and both algorithms perform well in this region.) Both algorithms provide reasonable reconstructions visually consistent with the surrounding flow in the interior region. The MF GPOD algorithm tends to predict local maxima in the missing regions, whereas the block-wise gappy SPOD algorithm tends to fill in gaps with values somewhat lower than those in the surrounding flow field. In the absence of the actual

flow field in those regions, it remains speculative which algorithm performs quantitatively better. In the narrow black zones near the borders of the FOV and for gaps that intersect with black zones, the MF GPOD algorithm, in many cases, fails to reconstruct the data, most prominently in the first time instant on the right boundary. The current method does not suffer from this problem and provides a believable reconstruction.

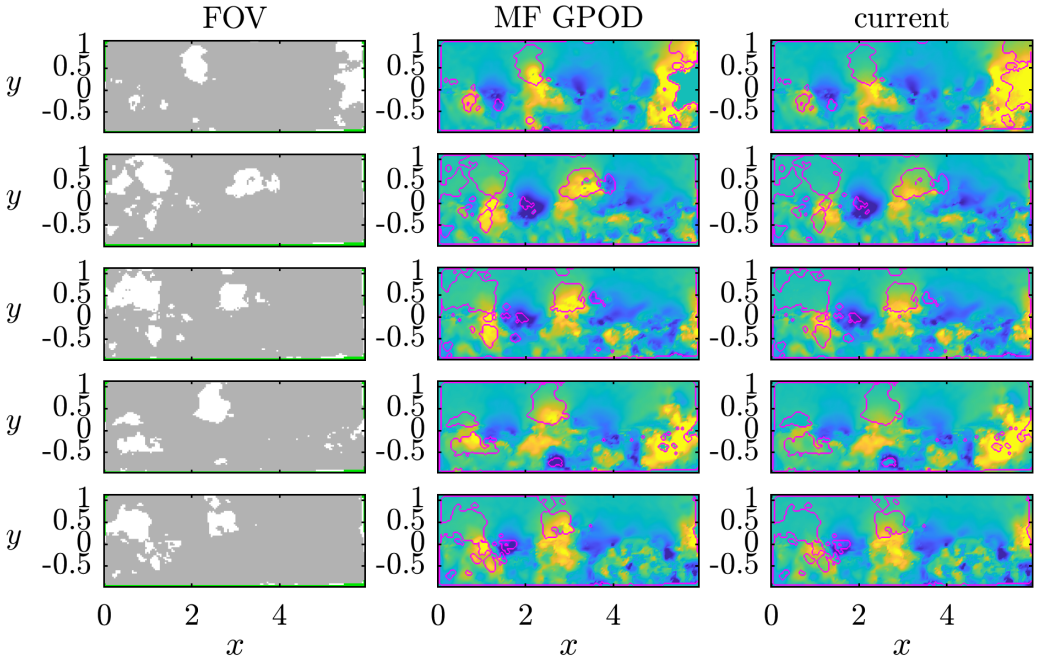

Figure 2: Same as figure 1 but for the wall-normal velocity, v. The color map is saturated at ± 70 [m/s] ($u \geq 70$, $u \leq -70$).

Similar observations can be made for the wall-normal velocity component shown in figure 2 at the same time instances. Both algorithms perform well for the interior gaps, with MF GPOD possibly overpredicting and gappy SPOD possibly underpredicting the flow field in the missing regions. Consistent with the observations for figure 1, gappy SPOD performs better at the fringes where gaps overlap with permanent black zones. For this study, the gappy SPOD reconstructions of the streamwise and wall-normal velocity components were computed separately, i.e., the results in figures 1 and 2 come from two runs of the code. In the context of the discovery of flow physics–arguably the method's main application–SPOD relies on a suitable inner product to enable the SPOD modes to be interpreted as the most energetic flow structures (see, e.g. Schmidt and Colonius, 2020). For incompressible, two-dimensional flows, the choice of the solution vector $[u\ v]^T$ yields modes that are optimal with respect to the turbulent kinetic energy. On the contrary, we found that treating the velocity components of the raw PIV data separately was advantageous in the present context of flow field reconstruction. This means that we are exclusively relying on the spatiotemporal autocorrelation information of each flow variable in this study.

case	n_t	n_x	n_y	Δt	n_{fft}	n_{ovlp}	n_{blk}	b_{win}	n_{tapers}
cavity	16,000	160	57	$6.25 \cdot 10^{-5}$s	512	256	61	2	3
high-lift	11,000	238	145	$9 \cdot 10^{-5}$s	512	256	41	2	3

Table 2: SPOD parameters for post-processing of the reconstructed data.

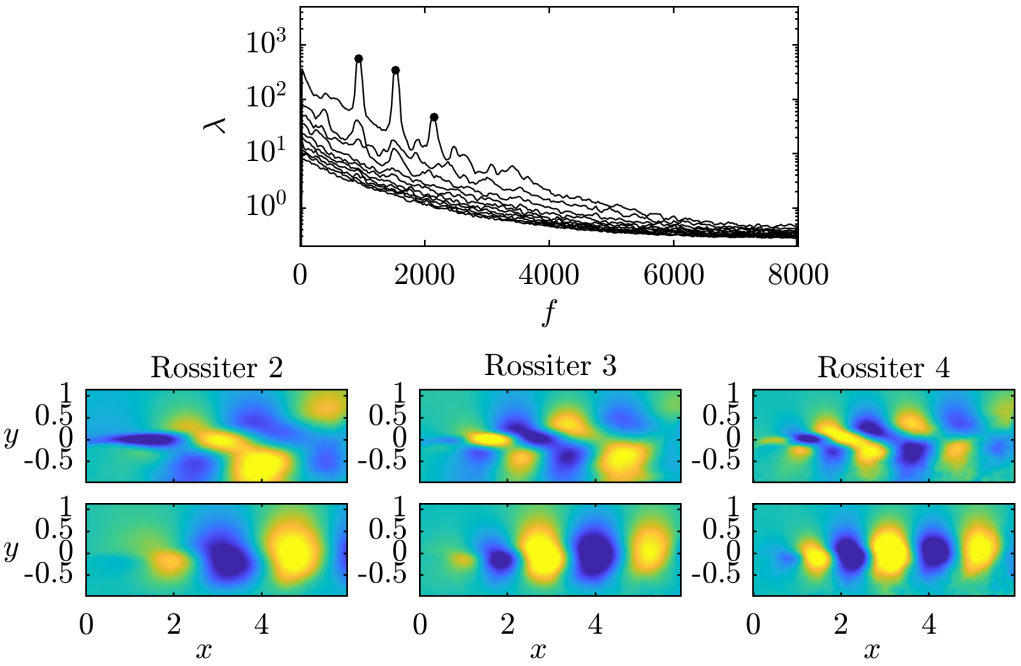

Figure 3: SPOD analysis of the reconstructed data. The eigenvalue spectra for the leading ten modes are shown at the top. The u (top row) and v (bottom row) components of the leading modes at the three marked peaks (•) are shown below. The three peaks correspond to the second, third, and fourth Rossiter modes. The real part of the modes is shown on a color map that is saturated for best interpretability.

Lastly, we use SPOD in the aforementioned context of physical discovery and briefly analyze the flow physics of the reconstructed data. The spectral estimation parameters of the SPOD analysis are independent of those of the gappy SPOD reconstruction. In particular, we use a larger number of 512 snapshots per block to obtain a higher frequency resolution and use multitapering (Schmidt, 2022) to decrease the variance of the spectral estimate. Specifically, we use a time-halfbandwidth product of $b_{\text{win}} = 2$, resulting in $n_{\text{tapers}} = 3$ discrete prolate spheroidal sequence (DPSS) tapers. The SPOD parameters used for the analysis of both of the reconstricted datasets are summarized in table 2. Figure 3 summarizes the SPOD post-processing results for the open cavity case. The eigenvalue spectrum of the leading SPOD mode exhibits a number of distinct peaks that are associated with the Rossiter tones. The three most prominent peaks are selected, and the corresponding SPOD modes' streamwise and wall-normal velocity components are visualized. The mode structures are consistent with those obtained by Zhang et al. (2019) for the same data. The flow physics of the Rossiter resonance mechanism are well-known and we refrain from their discussion. Instead, we note that the absence of any artifacts stemming from the missing data underlines the quality of the gappy SPOD data reconstruction.

3.2 High-lift device

Figure 4 shows the block-wise gappy SPOD reconstruction of the raw high-lift device PIV data. The parameters of the data and the reconstruction are summarized in table 1. The FOV with black zones and gaps indicated in the same way as in figures 1, 2) is shown in the left column. The remaining two columns show the streamwise and the normal velocity components, respectively. The three top rows show three consecutive representative snapshots from the middle of the time series. It can be seen that gaps mainly occur within the recirculation region on the underside of the slat and the shear layer that separates the recirculation zone from the bypassing flow that impinges on the main airfoil. As before, the gappy SPOD reconstruction of u and v were computed independently. The reconstructed fields of both components closely match the flow patterns surrounding the gaps. The bottom rows depict the three time instances with the largest percentages of missing data. In addition to smaller gaps within the recirculation and shear-layer regions similar to those seen earlier, large regions of missing data are located in the regions farthest from the sections of the high-lift device. In these large gaps, too, we observe that gappy SPOD produces smooth transitions between missing and existing data and that the reconstructed data closely resembles the flow features observed in previous

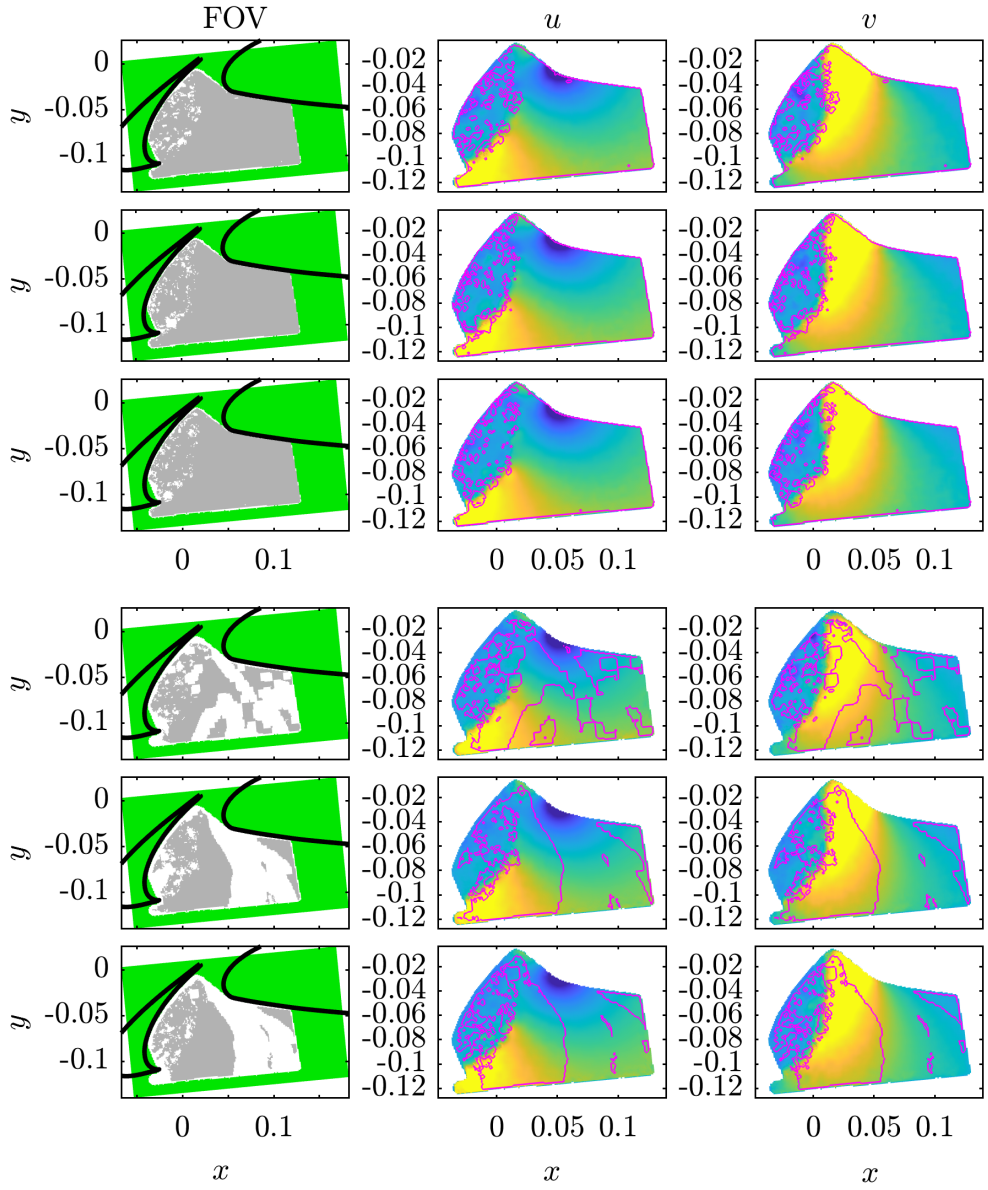

Figure 4: Reconstructed velocity fields at three consecutive representative time instances (top rows) and the three time instances with the largest overall gappyness (bottom rows). 40% of the FOV shown in the left column is usable PIV data (■, gray) containing gaps (□, white), while the remaining 60% belong to the black zone (■, green). The middle and right columns show u and v, respectively. The gaps in the original data (——, magenta) are outlined in both. The color map is saturated at -30 and 50 [m/s] (■ $u, v \geq 50$, ■ $u, v \leq -30$).

snapshots. As for the previous example, we proceed with an SPOD analysis of the completed database that uses the parameters reported in table 2, which differ from those for the gappy SPOD reconstruction.

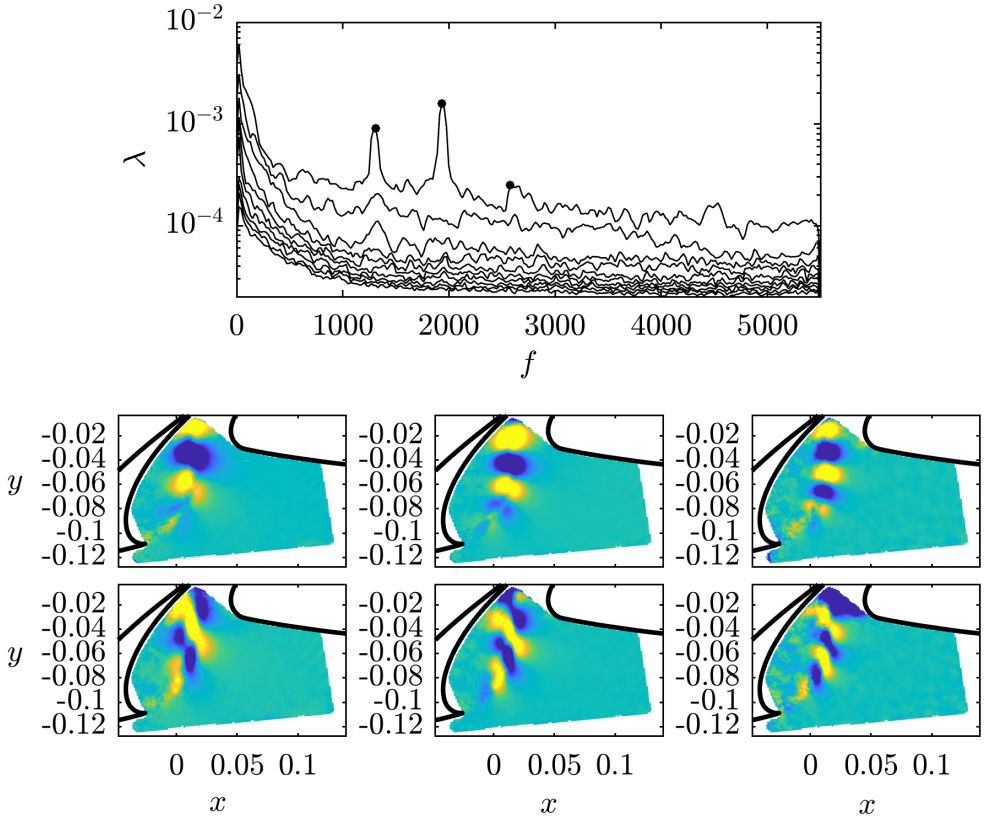

Figure 5: SPOD analysis of the reconstructed data. The eigenvalue spectra for the leading ten modes are shown at the top. The u (top row) and v (bottom row) components of the leading modes at the three marked peaks (•) are shown below. The real part of the modes is shown on a color map that is saturated for best interpretability.

Figure 5 reports the SPOD spectrum and three modes associated with two prominent and one minor peak in the leading-mode spectrum. The streamwise and normal velocity fields of the three modes reveal that the peaks are associated with Kelvin-Helmoltz-type instability waves that form in the shear layer between the recirculation region underneath the slat and the bypassing air stream. The shear layer reattachment location near the slat trailing edge on the cove surface establishes a turning point that facilitates resonance similar to the Rossiter mechanism in cavity flows. More importantly in the context of this work, the post-processing of the completed data yields clean flow structures that show no trace of the gaps in the raw data.

4 Discussion

Future work will address two primary limitations that were beyond the scope of this study. Firstly, the accuracy of the new algorithm's reconstruction has to be quantified. Here, we have applied the algorithm to raw PIV data, which precludes the computation of the actual reconstruction error since no true reference solution exists. This problem can be mitigated, as in Nekkanti and Schmidt (2023), by utilizing test data with artificial gaps for which the true solution is known. Second, extensive parameter studies on multiple datasets are required to identify optimal parameter sets and best practices.

Despite these limitations, this study demonstrates that the recently developed block-wise gappy SPOD algorithm is extremely effective at reconstructing missing data from raw PIV measurements. To compare the accuracy of the results, the present findings were qualitatively compared to the MF GPOD algorithm for the open cavity data. The comparison remained qualitative as the true velocity field in the missing regions

is unknown. The performance of both algorithms was satisfactory in the central region of the FOV, where they were capable of reconstructing large data gaps. However, the gappy SPOP algorithm performed better in the fringes and black zones by producing flow fields that are consistent with the interior domain.

Acknowledgements

The IIT authors gratefully acknowledge support from AFOSR Grant FA9550-22-1-0013, monitored by Dr. Gregg Abate, and NASA under contract 80NSSC18P3447, monitored by Dr. Meelan Choudhari. OTS gratefully acknowledges support from AFOSR Grant FA9550-22-1-0541, also monitored by Dr. Gregg Abate.

References

Beckers JM and Rixen M (2003) Eof calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and oceanic technology* 20:1839–1856

Benner P, Gugercin S, and Willcox K (2015) A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review* 57:483–531

Chaturantabut S and Sorensen DC (2010) Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing* 32:2737–2764

Everson R and Sirovich L (1995) Karhunen–loeve procedure for gappy data. *JOSA A* 12:1657–1664

Nekkanti A and Schmidt OT (2021) Frequency–time analysis, low-rank reconstruction and denoising of turbulent flows using spod. *Journal of Fluid Mechanics* 926:A26

Nekkanti A and Schmidt OT (2023) Gappy spectral proper orthogonal decomposition. *Journal of Computational Physics* 478:111950

Oliver MA and Webster R (1990) Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System* 4:313–332

Reynolds RW and Smith TM (1994) Improved global sea surface temperature analyses using optimum interpolation. *Journal of climate* 7:929–948

Saini P, Arndt CM, and Steinberg AM (2016) Development and evaluation of gappy-pod as a data reconstruction technique for noisy piv measurements in gas turbine combustors. *Experiments in Fluids* 57:122

Schmidt OT (2022) Spectral proper orthogonal decomposition using multitaper estimates. *Theoretical and Computational Fluid Dynamics* pages 1–14

Schmidt OT and Colonius T (2020) Guide to spectral proper orthogonal decomposition. *AIAA Journal* 58:1023–1033

Schmidt OT, Towne A, Rigas G, Colonius T, and Brès GA (2018) Spectral analysis of jet turbulence. *Journal of Fluid Mechanics* 855:953–982

Towne A, Schmidt OT, and Colonius T (2018) Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics* 847:821–867

Venturi D and Karniadakis GE (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* 15:70–73

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

Zhang Y, Cattafesta L, and 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*. page 3

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

Zhang Y, Cattafesta LN, Pascioni KA, Choudhari MM, Lockard DP, Khorrami MR, and Turner T (2020) Slat noise control using a slat gap filler. in *AIAA AVIATION 2020 FORUM*. American Institute of Aeronautics and Astronautics. AIAA Paper 2020-2553