BeBeC-2026-D06

JET COHERENT STRUCTURES PREDICTION USING THE MXARRAY AND NEURAL-NETWORK-BASED MAPPING

Ricardo Moreno[1], Luca Franceschelli[1], Luis A. Azpicueta-Ruiz[2],
Eduardo García-Portugués[3] and Marco Raiola[1]

[1] Department of Aerospace Engineering, Universidad Carlos III de Madrid
[2] Department of Signal Theory and Communications, Universidad Carlos III de Madrid
[3] Department of Statistics, Universidad Carlos III de Madrid
Av. Universidad 30, 28911, Leganés, Spain, jonmoren@ing.uc3m.es

May 20, 2026

Abstract

This study explores the non-intrusive reconstruction of turbulent flow dynamics from far-field acoustic measurements as an alternative to beamforming for jet noise source identification. The study exploits the MxArray, a 1024-microphone MEMS-based array, to capture bulk data of phase and coherence patterns from a scaled subsonic jet at $Re \approx O(10^4)$ in anechoic conditions. The research focuses on identifying wavepackets, large-scale coherent structures associated with jet-noise radiation, by correlating acoustic data sampled at 8kHz with synchronized, non-time-resolved 2D Particle Image Velocimetry (PIV) snapshots.

Considering the nonlinear correlation between both velocity and pressure fields, the study exploits Neural Network (NN) architectures combined with data compression to perform an aeroacoustic inversion. The network is trained to map temporal coefficients across the entire microphone array to the flow field reduced state, i.e., the spatio-temporal coefficients of the Hilbert Proper Orthogonal Decomposition modes.

The study explores a machine learning architecture, such as Multi-Layer Perceptron (MLP), which might exploit the temporal microphone array data and the spatial coherence of the flow field maps, treating the MxArray as a virtual sensor for real-time fluid-acoustic diagnostics. This approach provides a beamforming alternative to track wavepackets. As this is the first attempt of intricate high-frequency flow dynamics from acoustic data, refinement on inferring the temporal coefficients of the velocity field will be considered as future work.

1 INTRODUCTION

Aeroacoustic sources in subsonic turbulent flows remain an intricate noise source to be fully understood and characterized. The identification and localization of the acoustic energy emitted by chaotic flow motion is a common starting point of the majority of noise reduction strategies, employing acoustic cameras as the preferred instrument to diagnose this kind of noise.

Although phased microphone arrays provide high-frequency temporal data and facilitate localized acoustic map generation via beamforming techniques, this processing lacks of deterministic correlation between the back-projected acoustic features and the physical flow patterns responsible of the noise generation. Such a limitation arises because beamforming algorithms are intrinsically source-specific, requiring explicit *a priori* knowledge of the target source topology and its convective motion model to accurately solve the inverse problem.

On the side of imaging techniques, the Particle Image Velocimetry (PIV) one offers a spatial flow topology that is limited to the sampling of the set-up, often failing in describing those high-frequency dynamics of aeroacosutic sources.

To study the dynamics of turbulent flows, traditional modal decomposition techniques, such as *space-only* Proper Orthogonal Decomposition (POD), offer a clear description of coherent structures. However, POD is mathematically constrained when resolving coupled spatio-temporal dynamics due to it enforces an orthogonal projection optimized solely for spatial energy variance, and it is not capable of catching the continuous phase variations, particularly for structures advecting in the streamwise direction. In fact, one of the most prominent components of subsonic jet noise is attributed to large-scale coherent structures whose convective patterns, also known as wavepackets, are responsible for the most of sound emission compared to the fine-scale structures.

To counteract these limitations, we use the Hilbert Proper Orthogonal Decomposition (HPOD) to disentangle joint spatio-temporal dynamics [1]. In particular, the HPOD has been proved to be capable of delivering the same advective wavepacket when applied either in time or in the advective direction, thus proving ideal to decompose non-time-resolved velocity fields from PIV. By compressing the data in temporal coefficients and spatial modes, both the acoustic pressure field and the PIV velocity field are reduced in dimensionality, respectively. To capture the non-linearities among spatio-temporal coefficients, a Neural Network (NN) is proposed to infer from acoustic data to a description of velocity flow motion. To obtain the experimental bulk acoustic data for feeding the NN, the MxArray was employed. The MxArray is a massive MEMS-based microphone array, developed at Universidad Carlos III de Madrid (UC3M), controlled by an embedded Linux system, whose modularity was configured to be synchronized with PIV acquisition. The scalable 1024-channel instrument aims to obtain big data with low-cost sensors to infer those coherent structures responsible of the noise generation.

This study is documented in the following structure. Section 2 details the experimental setup, providing a comprehensive overview of the modular hardware architecture of the massive MEMS-based MxArray, its precision time protocol synchronization framework, and the technical specification of the PIV system utilized for concurrent cross-domain measurements. Section 3 establishes the mathematical methodology, focusing on the HPOD and its symmetric application for spatio-temporal modal extraction and data compression across both the acoustic pressure and hydrodynamic velocity fields. The architecture, input concatenation, and implementation details of the Multi-Layer Perceptron (MLP) network are also detailed here. Section 4 evaluates the predictive performance of the resulting aeroacoustic inversion framework,

presenting a comparative analysis through flow field modes that directly contrasts the velocity topology inferred from the acoustic time-series against the ground-truth spatial modes derived from the PIV HPOD. Finally, Section 5 summarizes the conclusions of this study, outlining the implications of this modal mapping for inferring spatial flow dynamics in non-intrusive aeroacoustic diagnostics and future lines regarding this technique.

2 EXPERIMENTAL SETUP

The aeroacoustic and flow-field diagnostic experiments are conducted inside the anechoic jet facility of the Aerospace Engineering Department at the UC3M, whose acoustic qualification and minor details can be found in [2]. In this controlled environment, we minimize sound reflections and ambient noise, ensuring measurement repeatability. The MxArray scalable system is used to capture acoustic data. This dense acoustic camera is configured with 16 modules, each of which is composed of 64 microphones and a BeagleBone Black (BBB), as shown in Fig. 1.

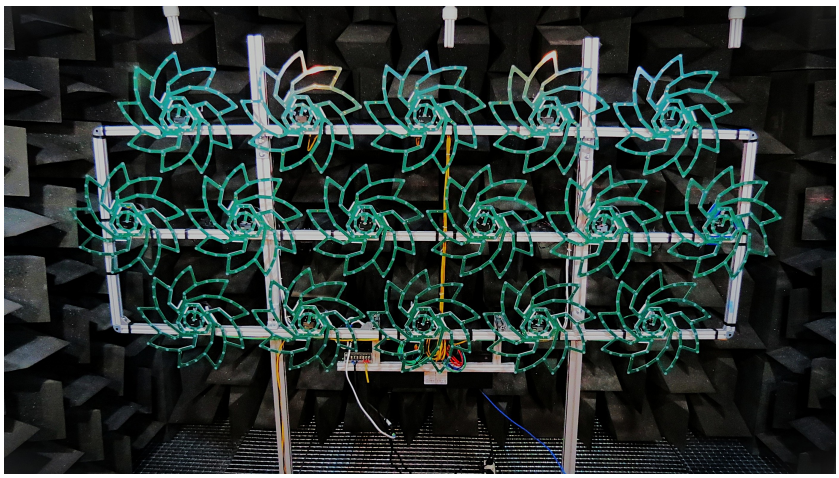

Figure 1: MxArray: A Modular, Multiplexed, and Massive MEMS-based Acoustic Array. Each module is composed of 64 microphones soldered in four printed circuit boards. The modular sensor distribution takes the form of a nine-armed Underbrink spiral, arranged in a honeycomb configuration. Each 64-microphone module is controlled by a BBB with a preempted Real-Time patch kernel.

Acoustic measurements were recorded at 8kHz. To achieve accurate synchronization, the MxArray utilizes a distributed architecture disciplined via Precision Time Protocol (PTP). The local hardware clocks of the 16 modules are synchronized over Ethernet by using a network switch operating as a PTP transparent clock, as depicted in Fig. 2.

In each module, a hardware pulse injection is triggered with a busy-wait loop to prevent local clock jitter, enabling signal alignment in post-processing. Furthermore, a National Instrument Data Acquisition (DAQ) system PCI-6040 concurrently logs the hardware pulse from an isolated, spare BBB node (BBB-117) alongside the TTL trigger pulses of the PIV system, which are actively disciplined by the master synchronizer. This hardware-level time-stamping guar-

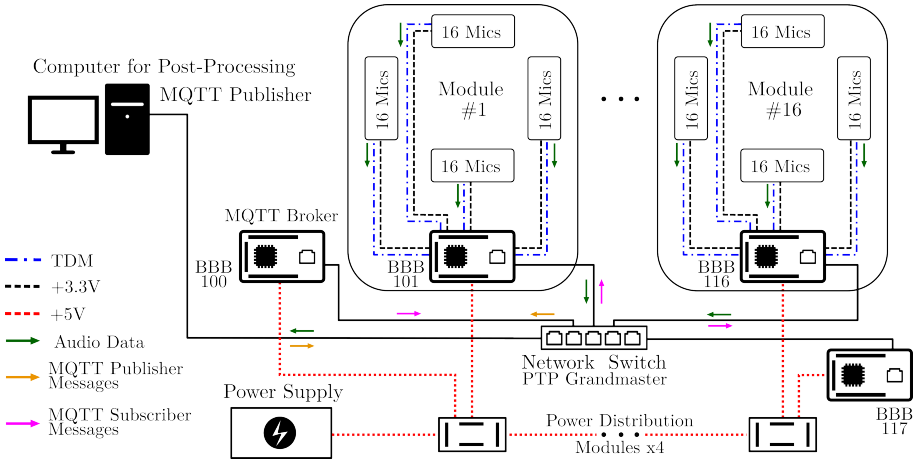

Figure 2: High-level system architecture of the scalable microphone beamforming array. The system consists of 16 acoustic modules, connecting 64 microphones through four TDM lines, all routed towards a dedicated BBB processing board. A central network switch serves as a PTP Grandmaster to ensure precise clock synchronization across all nodes. Control and audio data streaming are managed via an MQTT framework, which consists on: coordinating communication between the data collection nodes (BBB 101-117), centralizing messages to a broker (BBB 100), commanding the acquisition, and gathering audio files in a computer. Dual-rail DC power distribution (+3.3V and +5V) sustains the array infrastructure.

antees deterministic alignment between the modular acoustic system and the spatial flow field snapshots, mitigating any phase drift during testing.

The microphone array is oriented parallel to the axis of the jet plume to map acoustic source distribution (see Fig. 3). The jet flow is generated from a 10mm diameter nozzle discharging at a Mach number of 0.22, yielding a jet bulk velocity $U_j = 75\,\mathrm{m/s}$.

Velocity-field measurements are carried out using two-component PIV in the midfield region of the jet. The measurement plane (see Fig. 3) corresponds to the jet symmetry plane. The PIV acquisition system consists of a 5.5 Mpx Andor Zyla sCMOS camera mounted on a structural rig opposite to the MxArray, directly facing the illuminated measurement plane. The measurement plane is illuminated from below, with the laser sheet introduced through the floor of the anechoic chamber using a dual-cavity Nd:YAG laser operating at repetition rates at 8Hz, with a pulse energy of 200mJ. The flow is seeded with DEHS particles with a mean diameter of approximately $1\mu m$. The optical setup provides a spatial resolution of $13.5 px/mm$.

The particle images are processed using the PaIRS software developed at the University of Naples [3]. A multi-pass cross-correlation strategy is employed, with a final interrogation window size of 24×24 px and 75% overlap. Due to the presence of the microphone array in the background of the images, persistent optical reflections partially degrade the raw image quality. To reduce their influence, a background-removal pre-processing step is applied before the velocity-field calculation. This correction is based on a POD-based filtering procedure, following the approach described in [4].

The MxArray and PIV acquisitions are synchronized by means of reference pulses generated

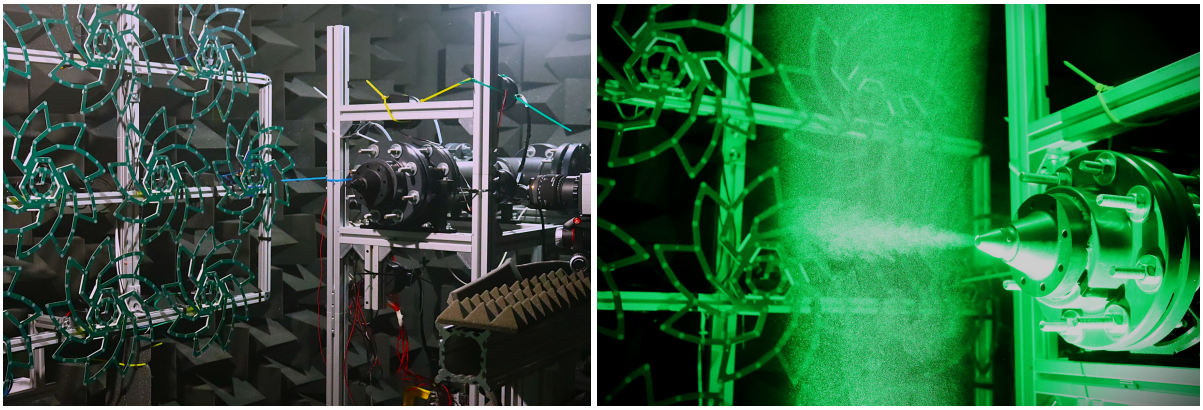

*Figure 3: Experimental setup inside the anechoic chamber. **(a)** MxArray positioned adjacent to the jet nozzle rig. **(b)** Flow visualization of the jet issuing from the nozzle.*

by a data acquisition system. The pulses are simultaneously recorded by the acoustic acquisition system and used as timing references for the PIV acquisition, enabling the microphone signals to be temporally aligned with the corresponding image pairs during post-processing.

3 METHODOLOGY

3.1 Hilbert Proper Orthogonal Decomposition

Advecting wavepackets, which are defined as local oscillations within a traveling envelope [5], can be characterized by means of a description of instantaneous dynamics (amplitude and phase). In modern signal processing, the Hilbert transform is a widely used tool for instantaneous envelope detection. By creating an analytic signal $\hat{s}(t)$ from a real-valued signal $s(t)$ by adding an imaginary component with the same frequency content of the original signal but shifted in phase by -90°, it is possible to express it according to a complex-valued polar representation

$$\hat{s}(t) = s(t) + i\mathscr{H}\{s(t)\} = A(t)e^{i\varphi(t)} \tag{1}$$

where $A(t)$ represents the instantaneous envelope or amplitude, and $\varphi(t)$ represents the instantaneous phase. The operator $\mathscr{H}\{\cdot\}$ represents the Hilbert transform, which allows a shift of -90° at all frequencies.

The HPOD distills spatio-temporal coherent structures of the wavepackets by applying a Hilbert transform to the multi-channel dataset before constructing a snapshot matrix. Subsequently, respective fields are decomposed into orthogonal spatio-temporal modes ranked by their energy content via Singular Value Decomposition (SVD) [1], according to

$$\hat{s}(x,t) = \sum_i \sigma_i \psi_i(t)\phi_i(x) = \sum_i \sigma_i A_{i,t}(t)e^{i\varphi_{i,t}(t)} A_{i,x}(x)e^{i\varphi_{i,x}(x)}. \tag{2}$$

where σ_i represents the i-th singular value, $\phi_i(x)$ is the i-th spatial mode depending only on x, and $\psi_i(t)$ contains the corresponding temporal coefficients depending only on t. Both $\phi_i(x)$

and $\psi_i(t)$ inherit the property from the decomposed snapshot matrix, thus allowing for a proper polar representation of the spatio-temporal wavepackets.

This dimensionality reduction framework is applicable to both the acoustic signals and the PIV measurements. For the PIV data, the snapshots are arranged as the columns of a snapshot matrix, concatenating the values of both u and v velocity field components over the 2D spatial grid covering the measured spatial domain. For the acoustic data, the HPOD is applied to a 3D snapshot tensor composed of *acoustic snapshots* synchronized to each flow-field snapshot. Each acoustic snapshot include a short sequence of 512 samples for each one of the MxArray channels. Given the lack of temporally resolved information in the PIV data, the Hilbert transform is applied in the axial direction in which the flow is advected, while for the acoustic data ,the Hilbert transform is applied in the temporal direction. For advective features, [1] demonstrates that the application of the Hilbert transform in space or time is equivalent.

3.2 Multi-Layer Perceptron Architecture

An artificial NN is established as a data-driven spatial transfer function to map the non-linear modal relationship between the synchronized acoustic and fluid velocity fields. Both fields are initially compressed into their most coherent spatial structures using HPOD, sorting the dominant modes by energy ranking. Because the HPOD input features are highly optimized and rank-ordered, they effectively allow non-linear mapping between the respective modal spaces.

To process HPOD complex values within the NN framework, the input and output are decoupled into their respective real and imaginary components. The acoustical input layer is a 64-dimensional feature vector constructed from the real and imaginary parts of the first 32 dominant acoustic modes (32 Real + 32 Imag).

A strict downward encoder progressively compresses the feature space through a sequence of fully connected layers, reducing the dimensionality from $64 \rightarrow 32 \rightarrow 16$. Each linear transformation within this compression funnel is followed by 1D Batch Normalization to stabilize training, a LeakyReLU activation function, and a Dropout layer to avoid overfitting to random noise.

The resulting 16-dimensional latent representation is subsequently branched into two parallel, task-specific prediction heads. The first head estimates the modal envelope amplitudes $A_{i,t}$ through a fully connected layer mapped via a ReLU activation function, allowing only positive values as physical boundary condition $A_{i,t} \geq 0$. Next, the second prediction head estimates the phase alignment parameters $\varphi_{i,t}$ utilizing a fully connected layer with a linear activation function. Finally, the outputs from both specialized branches are recombined through a complex-valued transformation to reconstruct the target velocity field parameters. The complete modal architecture is detailed in Table 1.

4 RESULTS

The predictive accuracy of the NN is evaluated through both the network's cross-modal training convergence history (Fig 4) and a direct structural comparison of the reconstructed velocity fields (Fig 5).

Figure 4 illustrates the MSE loss tracking over the optimization horizon. The training loss steadily drops from an initial value of 1.0406 to 0.9967, indicating that the network frame-

Table 1: Architecture of the Compressed Funnel Model

Layer / Operation	Type / Activation	Input Dimension	Output Dimension
Input Vector	Acoustic HPOD Coefficients		64 (32 Real + 32 Imag)
Strict Downward Encoder			
Linear 1	Fully Connected	64	32
Normalization 1	Batch Normalization 1D	32	32
Activation 1	LeakyReLU ($\alpha = 0.2$)	32	32
Regularization 1	Dropout ($p = 0.2$)	32	32
Linear 2	Fully Connected	32	16
Normalization 2	Batch Normalization 1D	16	16
Activation 2	LeakyReLU ($\alpha = 0.2$)	16	16
Regularization 2	Dropout ($p = 0.2$)	16	16
Parallel Head 1: Envelope			
Linear (A_i)	Fully Connected	16	8
Activation (A_i)	ReLU (Enforces $A_i \geq 0$)	8	8
Parallel Head 2: Alignment			
Linear (φ)	Fully Connected (Linear)	16	8 (φ)
Recombination Layer			
Transformation	$a + jb = A_i e^{j\varphi}$	8 + 8	16 (8 Real + 8 Imag)

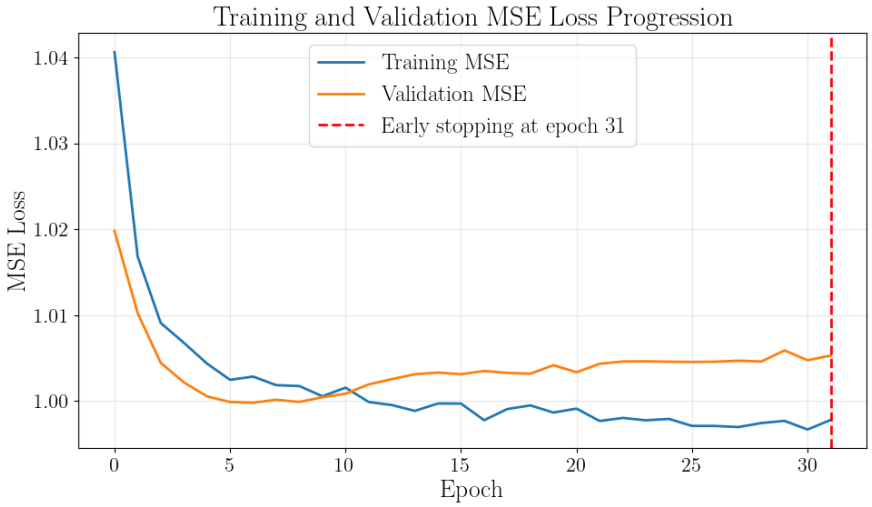

Figure 4: Model evaluation outcomes. MSE loss progression across training and validation splits illustrating early stopping enforcement at epoch 31.

work is able to capture the underlying cross-modal relationships. Notably, the validation loss reaches its global minimum of 0.9998 at epoch 6. This convergence means the high mutual information present within the energy-ranked HPOD input features, which effectively stabilizes

the optimization landscape early in the training process.

This NN rather than functioning as a continuous temporal coefficient predictor, it successfully operates as a non-linear spatial transfer function. Beyond epoch 6, the validation error slope increases while the training error continues to decrease, leading to overfitting to small-scale flow structures other than wavepackets. To preserve the generalized structural mapping capabilities and lock in optimal weights, early stopping is enforced at epoch 31, freezing the model state at the validation minimum and yielding a final scaled test MSE loss of 0.9807, as shown in Fig 4.

To verify the physical coherence of this modal inversion, Figure 5 provides a direct spatial comparison between the true spatial HPOD modes extracted from the experimental PIV measurements and those predicted by the NN using only the acoustic inputs. The velocity fields are obtained with the following equation

$$\mathbf{v}'_i(x,y) = \Re\left\{\phi_{y,i}(x,y)\cdot\psi_i\right\} \tag{3}$$

where $\mathbf{v}'_i(x,y)$ is the reconstructed 2D velocity fluctuation field for mode i; $\Re\{\cdot\}$ represents the real part operator; and $\phi_{y,i}(x,y)$ reads for the complex spatial mode. Last, ψ_i is the complex temporal coefficient for that snapshot, either the ground truth ($\psi_{\text{true},i}$) or the MLP's prediction ($\psi_{\text{inferred},i}$).

Since the neural network only predicts the marginal mean of the temporal coefficients, a fair comparison requires the field normalization described as

$$\mathbf{v}'_{\text{normalized},i}(x,y) = \frac{\mathbf{v}'_i(x,y)}{\max\left|\mathbf{v}'_i(x,y)\right|} \tag{4}$$

The MLP framework replicates the absolute polarity of the velocity wave-field, capturing the correct alternating positive and negative structure of the core fluctuations. Furthermore, the HPOD successfully shows the primary convective topologies and spatial wavelengths of the advecting wavepackets within the core jet stream region for the lower-order structures. However, beyond the third mode, the spatial reconstruction begins to exhibit aliasing effects. This physical limitation is the result of the spatial resolution boundaries imposed by the nozzle diameter (10mm) and the high bulk velocity (75m/s), which introduce high-frequency components that the microphone array is not able to capture at 8kHz.

5 SUMMARY

This aeroacoustic inversion involves treating audio signals, which have a superior temporal resolution than the PIV to infer velocity fields. In this sense, the HPOD allows to resolve in one-dimensional temporal direction to multiple-dimensional spatial fields by applying the Hilbert transform across the streamwise direction. This methodology resulted in a sort of non-linear spectral transfer function that relates acoustic modes with spatial modes of fluid velocity structures. It is worth highlighting that with measurements recorded at 8kHz, the network successfully reconstructs ($\varphi_{t,1}$, $\varphi_{t,2}$) without aliasing, while directly recovering the phase variations for modes 2 and 3 from far-field acoustic signatures. This demonstrates a strong, deterministic link between the acoustic field and the higher-order jet structures.

By integrating HPOD in the NN architecture, the model allows for wavepacket detection combining massive acoustic measurements and non-time resolved PIV. The proposed pipeline

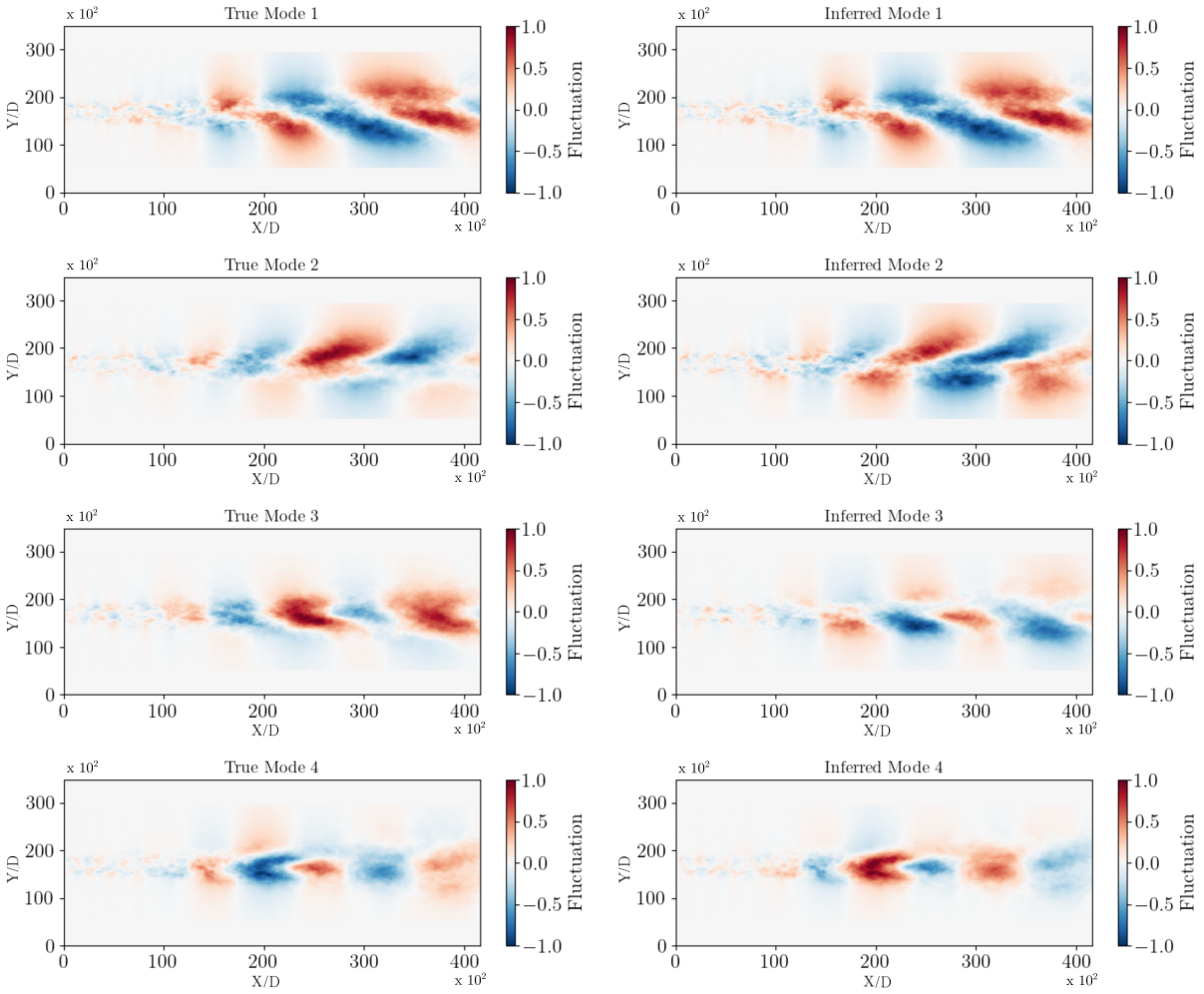

Figure 5: Spatial modal reconstruction. Side-by-side visualization comparing the true experimental HPOD spatial velocity modes against the NN predictions.

effectively functions as a non-linear spatial transfer function rather than a temporal coefficient predictor, due to the high mutual information contained within the energy-ranked, complex-valued HPOD features. Future work will investigate the generalization of this spatial transfer framework across varying jet Reynolds numbers and sheared flow profiles.

Acknowledgments

This work has been supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with UC3M (INFLUENTIA-CM-UC3M).

References

[1] Marco Raiola and Jochen Kriegseis. Hilbert proper orthogonal decomposition: A tool for educing advective wave packets from flow field data. *Physical Review Fluids*, 11(4):044905, 2026.

[2] Jonathan R Moreno, Luca Franceschelli, Daniel De la Prida, Luis A Azpicueta-Ruiz, and Marco Raiola. Implementation of a jet collector and dissipation cavity into a closed anechoic chamber for jet noise studies. In *30th AIAA/CEAS Aeroacoustics Conference (2024)*, page 3066, 2024.

[3] Gerardo Paolillo, Tommaso Astarita, et al. Pairs-unina: a robust and accurate free tool for digital particle image velocimetry and optical camera calibration. In *Proceedings of 21st International Symposium on Applications of Laser and Imaging Techniques to Fluid Mechanics*, 2024.

[4] Manuel Alejandro Mendez, Marco Raiola, Alessandro Masullo, Stefano Discetti, Andrea Ianiro, Raf Theunissen, and J-M Buchlin. Pod-based background removal for particle image velocimetry. *Experimental Thermal and Fluid Science*, 80:181–192, 2017.

[5] Peter Jordan and Tim Colonius. Wave Packets and Turbulent Jet Noise. *Annual Review of Fluid Mechanics*, 45(Volume 45, 2013):173–195, January 2013.