The 9th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2024
3–7 June 2024, Lisboa, Portugal

ANALYSIS OF CORRELATED FLOW FIELDS VIA EXTENDED CLUSTER-BASED NETWORK MODELS

Antonio Colanera[1], Moritz Reumschüssel[2], Jan Beuth[3], Matteo Chiatto[1], Luigi de Luca[1] and Kilian Oberleithner[3]

[1] Department of Industrial Engineering,
Università degli Studi di Napoli "Federico II",
Naples (80125), Italy
e-mail: antonio.colanera@unina.it

[2] Chair of Fluid Dynamics,
Technische Universität Berlin,
Berlin (100190), Germany

[3] Laboratory for Flow Instabilities and Dynamics,
Technische Universität Berlin,
Berlin (100190), Germany

Key words: Reduced Order Modelling, Clustering, Coherent Structures

Summary. This study introduces the Extended Cluster-based Network Modeling (eCNM), an innovative approach designed to enhance the understanding of coherent structures in turbulent flows. The eCNM focuses on characterizing the dynamics within specific subspaces or subsets of variables, providing valuable insights into complex flow phenomena. In the context of Proper Orthogonal Decomposition, several extended approaches have been proposed to tackle these challenges, such as Extended POD (EPOD) and Extended SPOD (ESPOD). One powerful method for data-driven modeling of complex nonlinear dynamics is the standard Cluster-based Network Modeling (CNM), consisting in an unsupervised machine learning procedure to reduce a dataset of snapshots to a few representative flow states. However, the presence of variable heterogeneity and measurement noise, both in time and space, can complicate interpretations and model training. The Extended Clustering approach offers enhanced control over the clustering process, can lead to significant computational savings, enables the extraction of dynamical features correlated with a specific subdomain or subset of variables, and facilitates the clustering of heterogeneous variables that are challenging to incorporate in a spatial norm. To demonstrate the effectiveness of the eCNM, it has been employed for the analysis of a swirl flame in unforced conditions, characterized by a precessing vortex core (PVC).

1 INTRODUCTION

Fluid dynamics involves the study of quantities that vary in space and time, often leading to complex interactions. Recognizing key flow features, or modes, is critical for simplifying complex flows [1]. Spectral Proper Orthogonal Decomposition (SPOD) is able to describe phenomena at different frequencies and energy levels [2] and is successfully employed in different flow configurations [3, 4].

Recently, the integration of Machine Learning, particularly through methods like Physics-Informed Neural Networks (PINNs) and Cluster-based Network Modeling (CNM) [5], provides powerful tools for reducing datasets and creating interpretable models, despite challenges like variable heterogeneity and measurement noise [6, 7].

Inherent variable heterogeneity and measurement noise, both in time and space, can complicate interpretations and model training. Within the framework of Proper Orthogonal Decomposition, extended approaches have been proposed to address these challenges, like the extended POD (EPOD) [8], extended spectral POD (ESPOD) [9] and the resolvent-based extended SPOD (RESPOD) [10].

This paper introduces an Extended Clustering approach [11] to focus on specific subspaces or variables, offering greater control and computational efficiency in handling complex, noisy data. The method is applied to a turbulent swirl flame case, demonstrating its utility in understanding the interplay between flow oscillations and flame behavior, particularly in the presence of a Precessing Vortex Core (PVC) [12].

The organization of the paper is as follows: the methodology is detailed in Section 2, with a focus on clustering approaches and the novel extended method. Section 3 illustrates the approach's effectiveness through its application to a swirl flame flow.

2 Methodology

2.1 Extended Clustering

Consider a field $\mathbf{u}(\mathbf{x}, t)$ defined within a steady domain Ω, which could originate from experimental data or computational simulations, encompassing variables like velocity and pressure.

The analysis is based on an ensemble of M time-resolved snapshots, statistically representative of the system. These snapshots are taken at uniform time intervals Δt, with $t^m = m\Delta t$ indicating the time of the mth snapshot. The field at this time is denoted by $\mathbf{u}^m(\mathbf{x}) := \mathbf{u}(\mathbf{x}, t^m)$ for $m = 1, \ldots, M$.

Cluster analysis, an unsupervised technique for data organization, groups similar instances into clusters without requiring prior labeling. For the set of M snapshots $\mathbf{u}^m(\mathbf{x})$, the data is coarsened into K clusters, with centroids identified using the k-means++ algorithm [13]. These centroids represent characteristic flow patterns or modes within each cluster.

In the Extended Clustering approach, akin to Extended Spectral Proper Orthogonal Decomposition (Extended SPOD) [14], the analysis is confined to a subset of variables from $\mathbf{u}$ or a spatial subspace $S \subseteq \Omega$. The snapshots $\mathbf{u}^m$ are restricted to a subset of variables or the subspace S and are denoted by $\mathbf{u}_s^m$, as illustrated in Figure 1.

The cluster-affiliation function assigns a specific velocity field $\mathbf{u}$ to the index of its nearest centroid:

$$k_s(\mathbf{u}) = \arg\min_i \|\mathbf{u} - \mathbf{c}_{si}\|_S, \tag{1}$$

where $\|\cdot\|_S$ denotes the norm in the spatial subspace S. Note that $k_s(\mathbf{u}) \equiv k_s(\mathbf{u}_s)$.

The characteristic function, indicating whether the mth snapshot corresponds to the ith centroid, is given by:

$$\xi_i^m = \begin{cases} 1, & \text{if } i = k_s(\mathbf{u}^m), \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

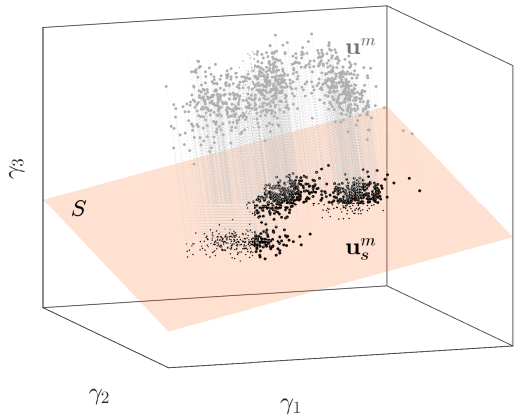


Figure 1: Illustration of the subspace S used for the reduced snapshots $\mathbf{u}_s^m(\mathbf{x})$. $(\gamma_1, \gamma_2, \gamma_3)$ represents a generic phase space.

The modified inner-cluster variance within the subspace S is expressed as:

$$J(\mathbf{c}_{s1}, \ldots, \mathbf{c}_{sK}) = \frac{1}{M} \sum_{m=1}^{M} \|\mathbf{u}_s^m - \mathbf{c}_{sk_s(m)}\|_S^2. \tag{3}$$

The goal is to find the optimal set of centroids $\{\mathbf{c}_{sk}^\star\}_{k=1}^K$ that minimize this inner-cluster variance (3). The centroid in the subspace is computed as:

$$\mathbf{c}_{sk} = \frac{1}{n_{sk}} \sum_{\mathbf{u}_s^m \in \mathcal{C}_{sk}} \mathbf{u}_s^m = \frac{1}{n_{sk}} \sum_{m=1}^{M} \xi_k^m \mathbf{u}_s^m, \tag{4}$$

where $n_{sk} = \sum_{m=1}^{M} \xi_k^m$ and $\mathcal{C}_{sk}$ are the clusters. The extended centroids are then obtained by:

$$\mathbf{c}_k^e = \frac{1}{n_{sk}} \sum_{m=1}^{M} \xi_k^m \mathbf{u}^m. \tag{5}$$

This Extended Clustering approach enhances the clustering process by focusing on a specific subset of variables, offering a more adaptable and robust method for managing complex, noisy, and high-dimensional data.

2.2 Cluster-based Network Model

The Cluster-based Network Model (CNM) approach represents the dynamics of a system using a directed network, where the nodes correspond to coarse-grained states of the system. The transition characteristics between these nodes are governed by high-order direct transition probabilities derived from the data [5].

The direct transition probability is calculated from the data as

$$Q_{k,j} = \frac{n_{k,j}}{n_j}, \tag{6}$$

where $Q_{k,j}$ represents the probability of transitioning from cluster $\mathcal{C}_j$ to $\mathcal{C}_k$, and $n_{k,j}$ denotes the number of such transitions. It is important to note that within the CNM framework, $n_{i,i} = Q_{i,i} = 0$.

The residence time $\tau^n = t^{n+1} - t^n$ refers to the duration spent in a cluster $\mathcal{C}_k$, where t^n and t^{n+1} indicate the time of the first and last snapshots when entering and leaving $\mathcal{C}_k$, respectively. The individual transition time from cluster j to cluster k is defined as half the residence time of both clusters:

$$\tau_{k,j}^n = \frac{\tau^{n-1} + \tau^n}{2} = \frac{t^{n-1} + t^n}{2}. \tag{7}$$

By averaging all $n_{k,j}$ individual transition times from cluster $\mathcal{C}_j$ to $\mathcal{C}_k$, the transition time $T_{k,j} = \frac{1}{n_{k,j}} \sum_{n=1}^{n_{k,j}} \tau^{n_{k,j}}$ is estimated.

For more detailed information, the reader is referred to [5] and [15].

2.3 Clustering distance metric and parameter choice

The choice of an appropriate distance metric is crucial for the effectiveness of the clustering algorithm. In this work, a filtered version of the Euclidean distance has been utilized to better separate spatially and temporally coherent structures in the phase space.

The squared standard Euclidean distance between two snapshots $\mathbf{u}^i$ and $\mathbf{u}^j$ is fundamentally related to the elements of the correlation matrix R_{ij}. This relationship can be expressed as

$$\|\mathbf{u}^i - \mathbf{u}^j\|_2^2 = R_{ii} + R_{jj} - 2R_{ij}, \tag{8}$$

where the elements R_{ij} are defined by

$$R_{ij} = \frac{\langle \mathbf{u}^i, \mathbf{u}^j \rangle}{M}, \tag{9}$$

and $\langle \cdot, \cdot \rangle$ represents the L_2 scalar inner product, defined as $\langle \mathbf{u}^i, \mathbf{u}^j \rangle = \mathbf{u}^{iT} \mathbf{W} \mathbf{u}^j$. Here, $\mathbf{W}$ is a weight matrix that accounts for grid non-uniformity and potential variable inhomogeneity, and $(\cdot)^T$ denotes the transpose operator.

Analogous to Spectral Proper Orthogonal Decomposition (SPOD), see [2], a filtered version of the correlation matrix can be introduced. The elements of this filtered matrix, denoted by S_{ij}, are given by:

$$S_{ij} = \sum_{k=-N_f}^{N_f} g_k R_{i+k,j+k}, \tag{10}$$

where g_k is a filter function, which can take various forms such as a box filter or a Gaussian filter, and N_f represents the filter span. Using the filtered correlation matrix, a filtered distance metric $\|\mathbf{u}^i - \mathbf{u}^j\|_{\text{filt}}^2$ can be defined, depending on the filtering parameters N_f and g_k:

$$\|\mathbf{u}^i - \mathbf{u}^j\|_{\text{filt}}^2 = S_{ii} + S_{jj} - 2S_{ij} = \sum_{k=-N_f}^{N_f} g_k \|\mathbf{u}^{i+k} - \mathbf{u}^{j+k}\|_2^2. \tag{11}$$

Notably, setting $N_f = 0$ and $g_k = 1$ reduces the metric to the standard Euclidean distance used in clustering. For further details on SPOD, the reader is referred to [2].

Selecting the appropriate number of clusters, K, and the optimal SPOD filter size, N_f, is critical for the accuracy and integrity of the clustering analysis. One approach is to use the Bayesian Information Criterion (BIC); see [16], which is quantified as:

$$\mathrm{BIC} = M\log(J) + K\log(M) - \frac{2}{d}\sum_{k=1}^{K} n_k \log\left(\frac{n_k}{M}\right), \tag{12}$$

where d (with $d \leq N$) is the dimension of the state vector. The BIC is a metric where a lower score indicates a more favorable model. When considering the number of snapshots M, this score penalizes a high value of the inner-cluster variance J and a large number of clusters K. A high J suggests that points within the cluster are spread out, which might indicate the need for additional clusters or a different value of N_f to better capture coherent spatio-temporal structures.

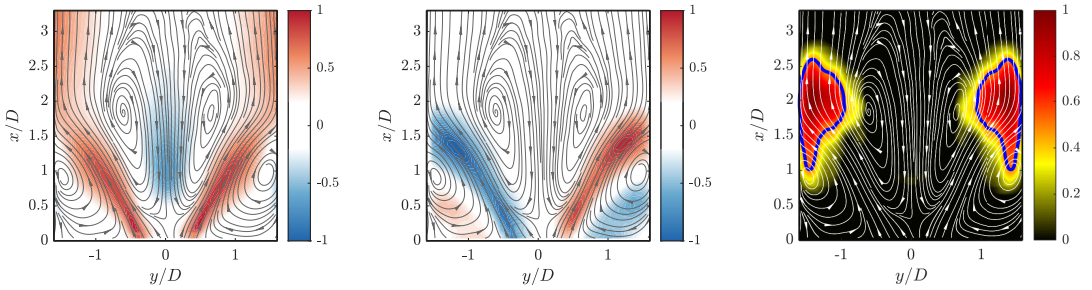
3 Results

The dataset analyzed here originates from a combustor with a swirl-stabilized natural gas flame. The flow within the combustion chamber features a precessing vortex core, a helically shaped coherent flow structure commonly observed in both reacting and non-reacting swirling flows when vortex breakdown occurs ([17]).

The dataset consists of Particle Image Velocimetry (PIV) measurements, which provide time-resolved data for both velocity components and synchronous flame images (data from [18]). The flame images are given as UV images filtered around the characteristic wavelength of OH* chemiluminescence. The chemiluminescence emission indicates the presence of excited OH molecules, a species produced in heat-releasing reactions within the flame. In particular, perfectly premixed flows exhibit a linear correlation between the spatial distributions of OH* chemiluminescence and the heat release rate. The combustion chamber consists of a cylindrical quartz glass tube with an inner diameter of 200 mm and a length of 300 mm. The diameter of the inflow section is $D = 55\,\mathrm{mm}$, and the Reynolds number based on D and the bulk velocity is $Re = 36000$. For further details on the experimental setup, the reader is referred to [18]. The sampling rate for the UV images and the PIV recordings is $f_s = 2\,\mathrm{kHz}$. The PIV grid consists of 73×77 elements, while the UV images are 251×251. By analyzing the velocity components, denoted by (u, v), along with the heat release, denoted by q, insights into the combustion dynamics in the presence of a PVC can be obtained.

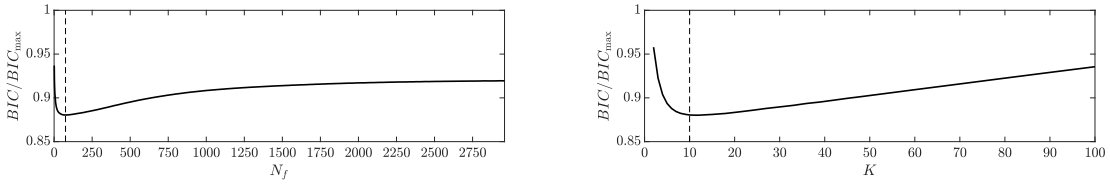
Figure 2 shows the time-averaged flow field within the combustor. Panels (a) and (b) display the temporal mean of the velocity components u and v, respectively. These panels reveal a radial expansion of the swirling jet flow, taking on a conical vortex breakdown shape. The measurement domain extends to the combustor side walls, located at $y/D = \pm 1.82$. Panel (c) illustrates the radial distribution of time-averaged heat release, $\bar{q}$, obtained from an Abel deconvolution of the UV images. This panel shows a detached, M-shaped flame configuration originating from the combustor inlet. In all panels, D denotes the diameter of the inlet section, x and y are the axial and transversal coordinates, respectively; every field variable has been normalized with respect to its maximum value. Here, the Strouhal number, St, is expressed as $St = fD/U$, where U denotes the reference velocity.

For subsequent analyses, each field variable ϕ has been decomposed into its temporal mean



(a) $\bar{u}$ (b) $\bar{v}$ (c) $\bar{q}$

Figure 2: Mean (time-averaged) flow. Panel (a)-(b): mean axial ($\bar{u}$) and transversal ($\bar{v}$) velocity components. Panel (c): Abel deconvoluted temporal mean of heat release distribution $\bar{q}$. Blue lines are the $\bar{q}/\bar{q}_{\max} = 0.5$ isolines. The Reynolds number based on D and the bulk velocity is $Re = 36000$. All variables are normalized with respect to their respective maxima.



(a) $K = 10$ (b) $N_f = 75$

Figure 3: Bayesian information criterion (BIC) for parameter selection. The minimum is observed at $K = 10$ and $N_f = 75$.

$\bar{\phi}$ and the fluctuation ϕ':

$$\phi(x, y, t) = \bar{\phi}(x, y) + \phi'(x, y, t). \tag{13}$$

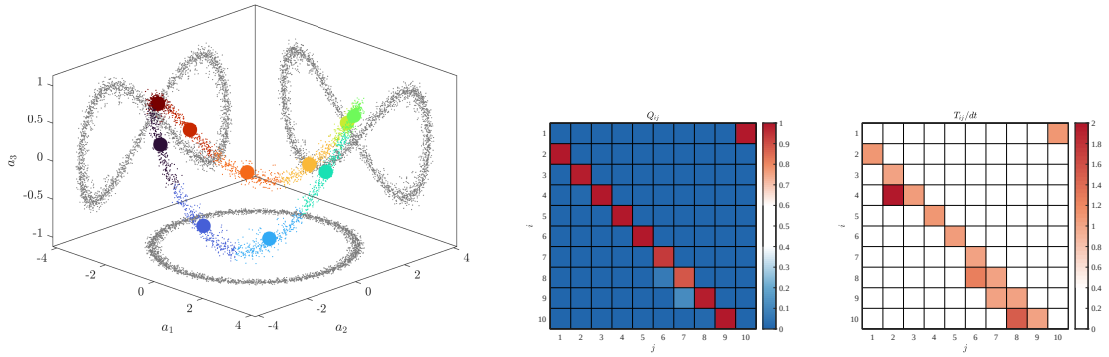
The eCNM has been applied to the fluctuating parts only.

3.1 The eCNM analysis

3.1.1 Optimal clustering parameter identification

The first step in the clustering approach is determining the appropriate size of the cluster-based model, specifically the optimal number of clusters K and the best filter dimension for the distance metric. The optimal values for N_f and K are determined using the Bayesian information criterion (BIC score) as described in equation (12). Figure 3 shows the evaluation of the BIC score.

A distinct minimum in the BIC score is found at $K = 10$ and $N_f = 75$, which corresponds to approximately 7.5 periods of PVC oscillations. These values are used for the subsequent analyses.



(a) (b) Q_{ij} (c) T_{ij}

Figure 4: Panel (a): Clusters of the a_i phase portrait. K-means++ algorithm has been employed for the clustering. $K = 10$ and $N_f = 75$ are used. Panels (b) and (c): CNM's direct transition probabilities (Q_{ij}) and times (T_{ij}) matrices.

3.1.2 CNM of velocity fluctuations

Once the optimal parameters are identified, the next step is partitioning the phase space. The Kmeans++ algorithm is used for this purpose. A visual representation of this clustering is shown in panel (a) of Figure 4.

The phase portrait differentiates the centroids of each cluster, represented by larger dots, from the individual snapshots, represented by smaller dots. Each snapshot is color-coded according to its associated centroid, highlighting the cluster boundaries and affiliations in the phase space. The phase portrait clearly displays the evolution corresponding to a limit cycle.

The limit cycle behavior is further emphasized through the construction of the CNM, as described in section 2.2. Specifically, the direct transition probabilities matrix Q_{ij} and the direct transition times matrix T_{ij} describe the system dynamics by indicating how often and how quickly the system transitions from one state (or cluster) to another. Panels (b) and (c) of Figure 4 visualize the matrices Q_{ij} and T_{ij}, respectively.

The most probable transitions correspond to the next cluster in the sequence, confirming the limit cycle nature of the dynamics. Additionally, the leading frequency can be estimated from the transition times matrix using:

$$St_{\mathrm{PVC}} \approx \frac{1}{T_{2,1} + T_{3,2} + \ldots + T_{1,10}} \frac{D}{U} \approx 0.14. \tag{14}$$

The cluster centroids for the velocity components u and v can be examined to reveal patterns and underlying dynamics. Using the most probable sequence of cluster centroids, Figure 5 presents the spatial distribution of FTLE for selected centroids. The backward integration time for the FTLE calculation was set to 1.5 times the period of the PVC oscillation. These distributions distinctly highlight the vortex structures formed by the PVC. The FTLE in the xy-plane reveals vortex roll-up, characteristic of Kelvin-Helmholtz instability. However, the PVC's helical nature results in alternating vortices forming a helix. Additionally, an overlay of the heat-release cluster centroids is included for further analysis in the subsequent section.

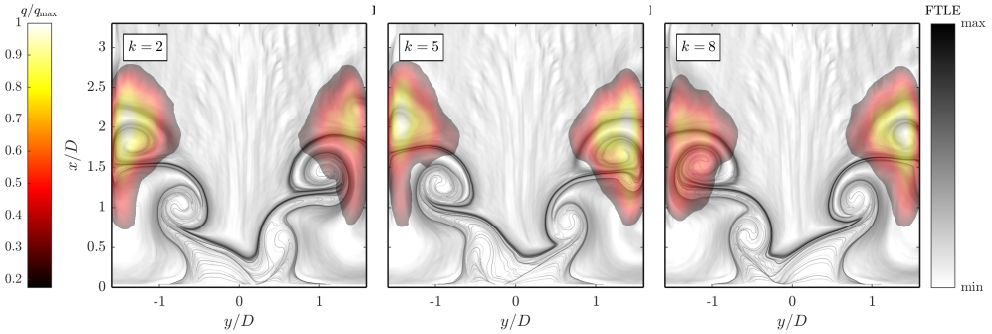


Figure 5: Spatial distribution of the backward finite-time Lyapunov exponent (FTLE) for selected velocity centroids. The figure also includes an overlay of heat-release cluster centroids. Each distribution has been normalized with respect to its maximum.

3.2 Extended CNM of the heat release

Having successfully clustered the velocity components and constructed the associated CNM, we can now proceed to the construction of the extended clusters. Utilizing the methodology introduced in section 2.1, the extended clusters $\mathbf{c}_k^e$ are derived using the function ξ_k^m, as detailed in Eq. (2). This approach allows the incorporation of heat release data into our clustering model, providing a comprehensive representation that combines both velocity and flame patterns.

For the analysis, the heat release has been decomposed into its temporal mean $\bar{q}$ and fluctuation q':

$$q = \bar{q} + q', \tag{15}$$

where the fluctuation term q' has been subjected to the clustering procedure.

The analysis of the heat release centroids can provide insights into the flame dynamics and the structures that form within the combustor. Figure 5 displays an overlay of the backward FTLE analysis of velocity centroids and the heat release reconstruction. The heat release reconstruction combines the Abel-deconvoluted symmetric mean flame with the fluctuating heat release rate centroids. This overlay reveals the predominant localization of heat release (q) within areas influenced by the PVC vortices roll-up. The progression of these vortices extends from the center to the sides of the combustor chamber, reflecting the PVC's impact on flame dynamics.

The efficacy of the extended CNM in predicting q dynamics is evaluated using an error measure, denoted by $\varepsilon(t)$. This metric is formulated as:

$$\varepsilon(t) = \sqrt{\frac{1}{A}\int_\Omega (q(x,y,t) - q_r(x,y,t))^2 dA}, \tag{16}$$

where A represents the measurement area, and q_r is the reconstructed distribution of the heat release rate. Given the inherent error introduced by the clustering procedure, it is insightful to consider the error in relation to its initial value, $\varepsilon_0 = \varepsilon(0)$.

Figure 6 displays the evolution of the error $\varepsilon(t)$. Notably, the error remains close to the initial value ε_0 ($\varepsilon/\varepsilon_0 \approx 1$), indicating the model's reliable predictive capability for the q temporal evolution. It should be noted that the initial error ε_0 is inherent to the clustering process and is unavoidable unless the number of clusters equals the number of snapshots. An effective model

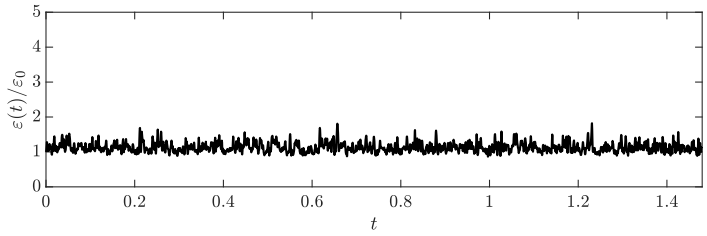


Figure 6: Reconstruction error $\varepsilon(t)$.

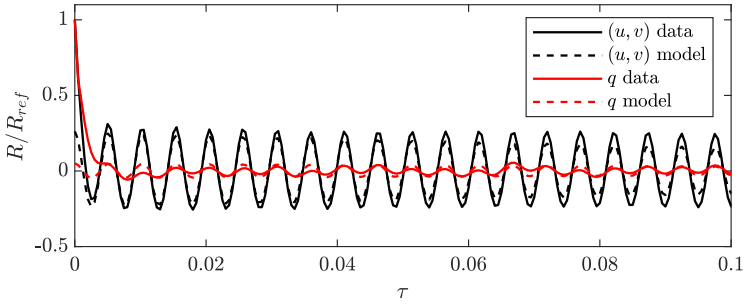


Figure 7: Autocorrelation function R for velocity components and q. Comparison between the real data and the model outputs. R_{ref} corresponds to the $R(0)$ of the actual dataset.

is characterized by its ability not to accumulate further errors arising from incorrect dynamics modeling.

The model's performance can also be assessed using the unbiased autocorrelation function R, as suggested by [19]. This function, defined by [20] as:

$$R(\tau) = \frac{1}{T_f - \tau} \int_{\tau}^{T_f} \mathbf{b}(t-\tau)^T \mathbf{b}(t) dt, \qquad \tau \in [0, T_f), \tag{17}$$

provides insights into the temporal relationships within the dataset. Here, T_f is the final time, τ stands for time delay, and $\mathbf{b}$ is a generic vector that contains the values of the field variable under examination at all spatial locations.

In Figure 7, the autocorrelation function for the velocity components and q is shown. Both real and model-generated data are included. The reference autocorrelation value, R_{ref}, used in these plots corresponds to the $R(0)$ of the actual dataset. By comparing the actual and model-derived R, it's evident that the model proficiently reproduces the temporal statistics of the flow.

The presented results underscore that the extended clustering method performs effectively when there's a pronounced correlation between the data incorporated into the clustering process and the data left unclustered. This correlation ensures that the method can accurately capture and represent underlying dynamics.

4 Conclusions

A novel approach based on Extended Cluster-based Network Modeling (eCNM) has been introduced for analyzing complex fluid dynamics. This procedure focuses on characterizing the dynamics within specific, potentially more informative, subspaces or subsets of variables. Several significant insights and findings have emerged from this investigation.

The first aspect of our analysis was the choice of the number of clusters characterizing the cluster-based network model and the distance metric for the coarse-graining process. The optimal cluster configuration, in terms of the number of clusters, was determined using the Bayesian information criterion. Minimizing this score provided the proper number of clusters, balancing the model's fit quality with its complexity. A filtered distance metric based on the filtered correlation matrix was employed, enhancing the separation of spatially and temporally coherent structures and improving the interpretability of the phase portraits.

The presented approach was applied to a swirl flame, exhibiting a precessing vortex core. The data consisted of PIV measurements and UV images filtered around the characteristic wavelength of OH* chemiluminescence, featuring the local heat release rate. The application of CNM to the velocity fluctuations revealed typical PVC flow field patterns, and the characteristic frequency of the PVC mode was estimated. The construction of extended cluster centroids highlighted the existence of a rotating flame pattern, predominantly localized within regions influenced by the PVC's vortex roll-up. The capability of eCNM in capturing flame dynamics was evaluated using the reconstruction error of the heat release field and the autocorrelation function comparison between the original dataset and the model. This confirmed the applicability of eCNM to field variables that exhibit strong correlations with those used to construct the model.

This study highlights the adaptability and robustness of Extended Cluster-based Network Modeling as a valuable tool for understanding and analyzing coherent structures within complex fluid flows. Particularly useful for extracting correlated flow field dynamics, this method could be successfully applied in various configurations, such as characterizing concentration distributions in multiphase flows, detecting particle motions in suspensions, and describing acoustic phenomena associated with velocity flow field subregions. Furthermore, it could be effective in the a-posteriori analysis of Reynolds stresses in turbulence modeling and assist in control design by coarse-graining the functional shape of the flow field.

The focus of this paper is to employ the "extended" ePOD approach within the CNM framework to combine the advantages of both methodologies. This integration allows for constructing models that can predict correlated flow fields based on dynamics inferred from only a subset of variables. This approach is particularly advantageous when applied to correlated flow fields or synchronized signals. However, it becomes a limitation when dealing with variables that are not correlated.

References

[1] C. W. Rowley, Igor Mezić, S. Bagheri, Philipp Schlatter, and D. S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, nov 2009.

[2] Moritz Sieber, C. Oliver Paschereit, and Kilian Oberleithner. Spectral proper orthogonal decomposition. *Journal of Fluid Mechanics*, 792:798–828, 2016.

[3] Jean Hélder Marques Ribeiro and William Roberto Wolf. Identification of coherent struc-

tures in the flow past a naca0012 airfoil via proper orthogonal decomposition. *Physics of Fluids*, 29(8), August 2017.

[4] A. Colanera, A. Della Pia, M. Chiatto, L. de Luca, and F. Grasso. Modal decomposition analysis of unsteady viscous liquid sheet flows. *Physics of Fluids*, 33:092117, 2021.

[5] D. Fernex, B. R. Noack, and R. Semaan. Cluster-based network modeling–from snapshots to complex dynamical systems. *Science Advances*, 7(25):eabf5006, 2021.

[6] Jakob G. R. von Saldern, Johann Moritz Reumschüssel, Thomas L. Kaiser, Moritz Sieber, and Kilian Oberleithner. Mean flow data assimilation based on physics-informed neural networks. *Physics of Fluids*, 115129, 2022.

[7] Antonio Colanera, Eduardo Di Costanzo, Matteo Chiatto, and Luigi de Luca. Reduced-order model approaches for predicting airfoil performance. *Actuators*, 13(3):88, February 2024.

[8] J. Borée. Extended proper orthogonal decomposition: A tool to analyse correlated events in turbulent flows. *Experiments in Fluids*, 35:188–192, 8 2003.

[9] Moritz Sieber, Christian Oliver Paschereit, and Kilian Oberleithner. Advanced Identification of Coherent Structures in Swirl-Stabilized Combustors. *Journal of Engineering for Gas Turbines and Power*, 139(2):021503, 09 2016.

[10] U. Karban, E. Martini, A.V.G. Cavalieri, L. Lesshafft, and P. Jordan. Self-similar mechanisms in wall turbulence studied using resolvent analysis. *Journal of Fluid Mechanics*, 939, April 2022.

[11] Antonio Colanera, Johann Moritz Reumschüssel, Jan Paul Beuth, Matteo Chiatto, Luigi de Luca, and Kilian Oberleithner. Extended cluster-based network modeling for coherent structures in turbulent flows. *Submitted to Theoretical and Computational Fluid Dynamics*, December 2023, doi: 10.21203/rs.3.rs-3788637/v1.

[12] Sébastien Candel, Daniel Durox, Thierry Schuller, Jean-François Bourgouin, and Jonas P. Moeck. Dynamics of swirling flames. *Annual Review of Fluid Mechanics*, 46(1):147–173, 2014.

[13] Hugo Steinhaus et al. Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci*, 1(804):801, 1956.

[14] Moritz Sieber, Christian Oliver Paschereit, and Kilian Oberleithner. Advanced Identification of Coherent Structures in Swirl-Stabilized Combustors. *Journal of Engineering for Gas Turbines and Power*, 139(2):021503, 09 2016.

[15] Eurika Kaiser, Bernd R. Noack, Laurent Cordier, Andreas Spohn, Marc Segond, Markus Abel, Guillaume Daviller, Jan Östh, Siniša Krajnović, Robert K. Niven, and et al. Cluster-based reduced-order modelling of a mixing layer. *Journal of Fluid Mechanics*, 754:365–414, 2014.

[16] Dan Pelleg and Andrew Moore. X-means: Extending k-means with efficient estimation of the number of clusters. *Machine Learning, p*, 01 2002.

[17] K. Oberleithner, M. Sieber, C. N. Nayeri, C. O. Paschereit, C. Petz, H.-C. Hege, B. R. Noack, and I. Wygnanski. Three-dimensional coherent structures in a swirling jet undergoing vortex breakdown: stability analysis and empirical mode construction. *Journal of Fluid Mechanics*, 679:383–414, 2011.

[18] Finn Lückoff, Moritz Sieber, Christian Oliver Paschereit, and Kilian Oberleithner. Phase-opposition control of the precessing vortex core in turbulent swirl flames for investigation of mixing and flame stability. *Journal of Engineering for Gas Turbines and Power*, 141(11), sep 2019.

[19] N. Deng, B. R. Noack, M. Morzyński, and L. R. Pastur. Cluster-based hierarchical network model of the fluidic pinball–Cartographing transient and post-transient, multi-frequency, multi-attractor behaviour. *J. Fluid Mech.*, 934:A24, 2022.

[20] Bartosz Protas, Bernd R. Noack, and Jan Östh. Optimal nonlinear eddy viscosity in galerkin models of turbulent flows. *Journal of Fluid Mechanics*, 766:337–367, 2015.