CA22151: Cyber-Physical systems and digital twins for the decarbonisation of energy-intensive industries

CYPHER

Deliverable D8

Unsupervised algorithms for the optimal local selection and simplification of modelling approaches

Authors: Yagiz Yalcinkaya (ULB), Nguyen Anh Khoa Doan (TU Delft), Katarzyna Bizon (PK), Alessandro Parente (ULB)

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by
the European Union

Table of Contents

1. Introduction

Computational Fluid Dynamics (CFD) has become a cornerstone of modern engineering, enabling the detailed analysis of fluid behavior for applications ranging from aerodynamics and propulsion to energy systems. Its value lies in the ability to provide predictive insights that guide design, optimization, and performance assessment without relying solely on costly experiments. However, when applied to reactive flows, CFD simulations encounter significant challenges. The need to account for multiple chemical species, complex reaction kinetics, and their tight coupling with turbulent mixing makes these simulations highly demanding in terms of both accuracy and computational cost. Approaches such as Large Eddy Simulation (LES) offer a practical compromise by resolving large-scale turbulent structures while modeling the impact of unresolved, small-scale dynamics. While methods like LES highlight the essential role of CFD in capturing complex flow–chemistry interactions, current approaches remain computationally intensive and leave room for further development and improvement. Furthermore, LES, and CFD in general, result in very large dataset given the spatial resolutions required, making it challenging to elucidate the key relevant physical phenomena contained therein.

Such limitations motivate the development of approaches that can extract essential patterns from large, complex datasets while reducing computational demands. In this context, unsupervised algorithms provide a versatile set of tools for addressing the challenges of high-dimensional CFD data. Clustering methods enable the automatic grouping of similar flow states and the identification of coherent structures [1-4], while dimensionality reduction and feature extraction techniques allow for compact yet informative representations of complex flow fields [4,5]. Building on these capabilities, adaptive model simplification strategies can leverage unsupervised insights to dynamically refine or select models based on local flow characteristics [5,6]. Together, these algorithms offer a systematic framework for improving the efficiency and interpretability of reactive flow simulations, supporting both reduced computational cost and enhanced physical fidelity in CFD analyses.

In the literature, various unsupervised algorithms have been applied to the analysis of CFD data. Clustering methods such as k-means have been utilized to classify flow states and identify coherent structures, while Local Principal Component Analysis (LPCA) has enabled the capture of localized variability in heterogeneous, multi-scale flows [1,3]. For dimensionality reduction and feature extraction, techniques like Principal Component Analysis (PCA) and Proper Orthogonal Decomposition (POD) are widely adopted to obtain compact yet physically meaningful representations of high-dimensional flow fields, reducing computational cost while retaining key dynamics [6,7]. In addition, methods such as Vector Quantization Principal Component Analysis (VQPCA) provide a classification-based framework for adaptive local model selection, in which different flow regions are dynamically assigned to the most appropriate reduced-order representation [1]. Collectively, these applications underscore the increasing importance of unsupervised algorithms in improving the interpretability and computational tractability of reactive flow simulations.

2. Clustering

Unsupervised learning has gained increasing importance in combustion and CFD research due to its ability to analyze large and complex datasets without requiring labeled information. The growing availability of high-fidelity simulations and experimental data has further accelerated its adoption, as traditional analysis tools often struggle to capture hidden structures in such high-dimensional spaces.

Among unsupervised methods, clustering has been widely applied to uncover coherent patterns in combustion data by grouping observations with similar thermo-chemical or fluid-dynamic characteristics. Various approaches have been explored. For example, Local Principal Component Analysis (LPCA) has been applied to identify low-dimensional manifolds in MILD combustion, enabling efficient chemistry tabulation and reduction, and facilitating the interpretation of DNS and LES datasets [2,4,5,8].

A particularly prominent technique is the k-means algorithm, arguably the most widely used clustering routine [9,28,29]. K-means partitions a dataset into user-defined clusters by minimizing the distance between data points and their centroids, thereby assigning each computational cell to a representative state. Prior studies have applied k-means to identify soot formation regions in Reactivity Controlled Compression Ignition (RCCI) engines or to classify combustion regimes in turbulent flames, demonstrating its versatility and effectiveness for combustion applications [9,10].

As an illustrative example, Savarese et al. presented an application of k-means clustering for the automatic generation of Chemical Reactor Networks (CRNs) from CFD simulations of a semi-industrial MILD furnace [11]. In their methodology, computational cells were grouped according to thermo-chemical variables such as temperature, reaction progress, velocity, and residence time, with each cluster centroid defining a characteristic reactor state. The outcome was a set of macro-zones that exhibit internal homogeneity and can be modeled as perfectly stirred reactors within a CRN framework. To address the limitation that k-means does not inherently ensure spatial connectivity, the authors introduced a graph-based reassignment procedure that guaranteed clusters corresponded to physically contiguous regions.

The resulting CRNs were then solved using detailed kinetic schemes, achieving accurate predictions of pollutant emissions, particularly NO, across different methane–hydrogen fuel mixtures. A notable strength of this approach is its ability to generalize networks derived from clustering in a single operating condition were successfully extrapolated to other conditions while maintaining predictive accuracy.

Compared to conventional CFD post-processing and manually designed CRNs, the clustering-based framework significantly reduced the need for expert intervention and cut computational costs from hours to minutes.

Taken together, this study demonstrates how unsupervised clustering, and k-means in particular, can be embedded into combustion modeling workflows to automate the design of reactor networks. By uncovering coherent structures in CFD data and translating them into reduced-order models, the method provides a systematic path toward efficient, accurate, and interpretable simulations of reactive flows.

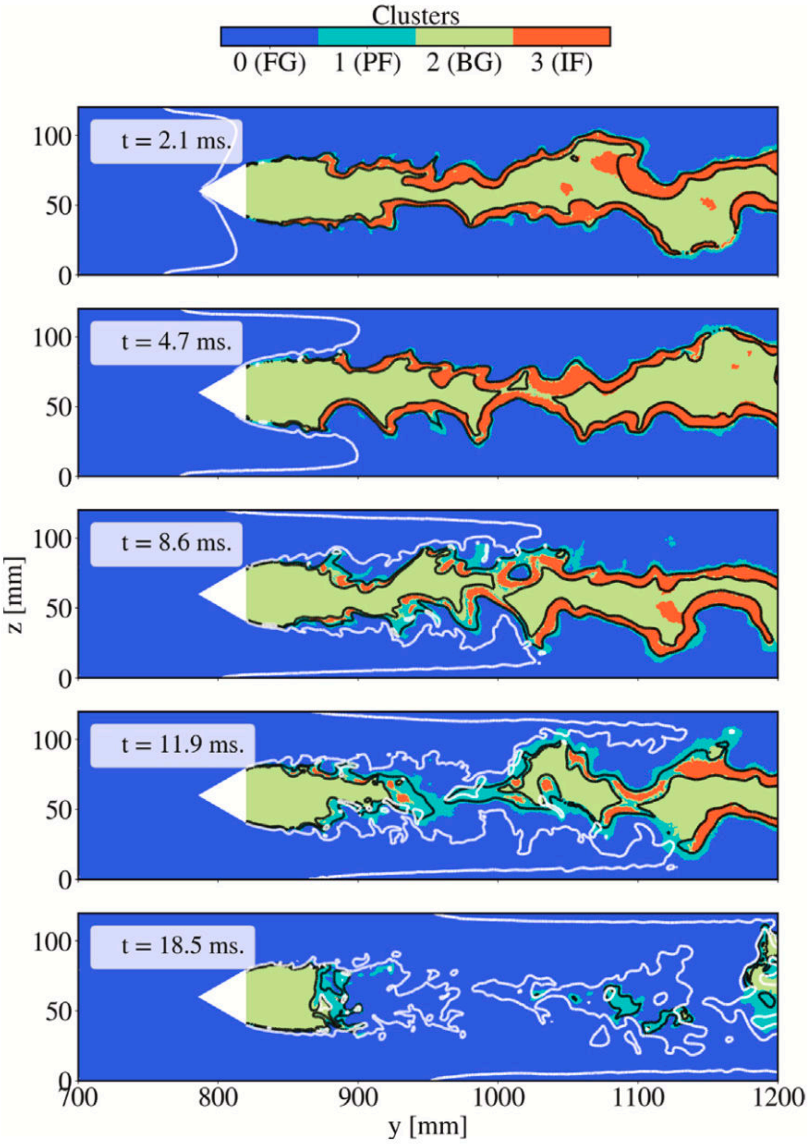

Figure 1 - Temporal evolution of the four k-means clusters (fresh gases (FG), preheat (PF), intense flame (IF), and burnt gases (BG)) in a longitudinal cut of the bluff-body combustor. The sequence illustrates how the intense flame cluster progressively diminish and disappears from the recirculation zone, providing a clear marker of global extinction. Figure adapted from [12].

A further example of unsupervised clustering in combustion research is provided by Lesaffre et al. [12], who investigated lean blow-off (LBO) dynamics in a bluff-body stabilized flame using a combination of Principal Component Analysis (PCA) and k-means clustering. Their approach began with PCA to reduce the dimensionality of high-fidelity LES data, ensuring that the most significant thermo-chemical and flow variables were retained. On this reduced dataset, k-means clustering revealed four physically distinct zones: fresh gases, burnt gases, preheat region, and intense flame. These clusters were not predefined but emerged naturally from the data, demonstrating the strength of unsupervised learning for unbiased zone identification.

Observing the temporal development of the clusters allowed the authors to define an extinction criterion, whereby global flame loss is marked by the disappearance of the intense flame cluster within the downstream recirculation zone. Figure 1 illustrates this process by showing the spatial distribution of the four clusters at successive instants in time. It becomes evident that the intense flame region gradually shrinks and is replaced by the preheat cluster until complete extinction occurs, whereas traditional global metrics such as mean temperature or integrated heat release fail to capture this localized event.

In addition, the clustering results were coupled with balance analyses of mass and energy fluxes, which revealed that blow-off was driven primarily by the weakening of the balance between chemical heat release and conductive heat flux, while convective transport remained nearly unchanged. This combined use of clustering and physical analysis offered a new framework to explain how local changes propagate to cause global flame extinction, highlighting the value of unsupervised learning in understanding complex unsteady combustion phenomena.

3. Dimensionality Reduction and Feature Extraction

Unsupervised learning techniques are increasingly being adopted not only for clustering but also for dimensionality reduction and feature extraction, which play a crucial role in handling the high-dimensional nature of combustion data. Detailed CFD simulations often involve hundreds of species and thermo-chemical variables, creating datasets that are too large and complex to be directly analyzed or efficiently integrated into combustion solvers. Traditional approaches that rely on manually selected features may overlook important correlations and tend to be problem-specific, limiting their general applicability.

Dimensionality reduction methods such as PCA provide a systematic, data-driven way to extract the most informative features while discarding redundancies [6,7]. By projecting high-dimensional states onto a lower-dimensional manifold that captures the dominant modes of variability, these methods enhance interpretability, reduce computational cost, and improve the robustness of subsequent unsupervised algorithms such as clustering.

High-fidelity simulations of reactive flows require detailed chemical kinetics to accurately capture species interactions and reaction dynamics. However, incorporating large kinetic mechanisms into multidimensional CFD models is computationally expensive, primarily due to the large number of species and the wide range of chemical timescales. In operator-splitting CFD solvers, this expense is most pronounced during the chemical integration step, where stiff, nonlinear ordinary differential equations (ODEs) are solved for each computational cell [13,14].

Several strategies have been developed to reduce this cost, including mechanism reduction techniques, adaptive chemistry methods, and efficient ODE solvers. Among these, Cell Agglomeration (CA) has proven effective by clustering cells with similar thermochemical states and performing chemistry calculations at the cluster level [15-18]. This reduces the number of ODE integrations while maintaining accuracy, and it can be combined with other acceleration techniques for further gains. The challenge in CA lies in defining appropriate similarity criteria, traditionally based on user-selected thermochemical variables, which requires prior knowledge and reduces adaptability.

To address this limitation, PCA is introduced into the CA framework. PCA provides a low-dimensional, uncorrelated representation of the thermochemical state by capturing the dominant variance in the data, removing redundancy, and preserving key combustion features. More importantly, PCA enables an unsupervised clustering process, since the extracted principal components do not depend on user-chosen thermochemical variables but are learned directly from the data. This unsupervised nature increases automation, reduces reliance on case-specific expert knowledge, and allows the method to generalize across different combustion configurations. In other words, PCA transforms CA into a more data-driven and adaptive clustering framework rather than one constrained by manually pre-selected features.

In this study [23], the conventional CA method and the PCA-enhanced CA approach are implemented within a turbulent combustion solver and evaluated in two benchmark configurations of the Adelaide Jet in Hot Coflow (AJHC) burner: (i) Reynolds-Averaged Navier-Stokes (RANS) simulation of an n-heptane flame, and (ii) LES of an equimolar methane-hydrogen flame [19,20].

The evaluation of these two benchmark cases highlights both the accuracy and efficiency of the proposed approach. For the *n*-heptane flame in the RANS framework, CA and CA-PCA can reproduce the reference temperature and major species profiles with good agreement, showing only minor deviations in OH. Figure 2 shows that both methods reproduce the reference temperature, OH, and n-C_7H_{16} profiles well, with only slight deviations in OH. Accuracy is also quantified using the normalized RMSD [21,22]:

$$\epsilon(\xi) = \frac{1}{\xi_{ref}} \sqrt{\frac{\sum_i^{N_{cells}} \left(\xi_i^{CA} - \xi_i^{org}\right)^2}{N_{cells}}}$$

where ξ_i^{CA} and ξ_i^{org} denote CA and detailed results, respectively. The quantitative error analysis shown in Table 1 confirms that the PCA-based clustering maintains acceptable accuracy, with errors arising mainly from variance truncation. Importantly, the use of PCA leads to a substantial gain in efficiency, while conventional CA reduces the overall runtime by more than 70%, the PCA-based method pushes this value beyond 80% with negligible overhead.

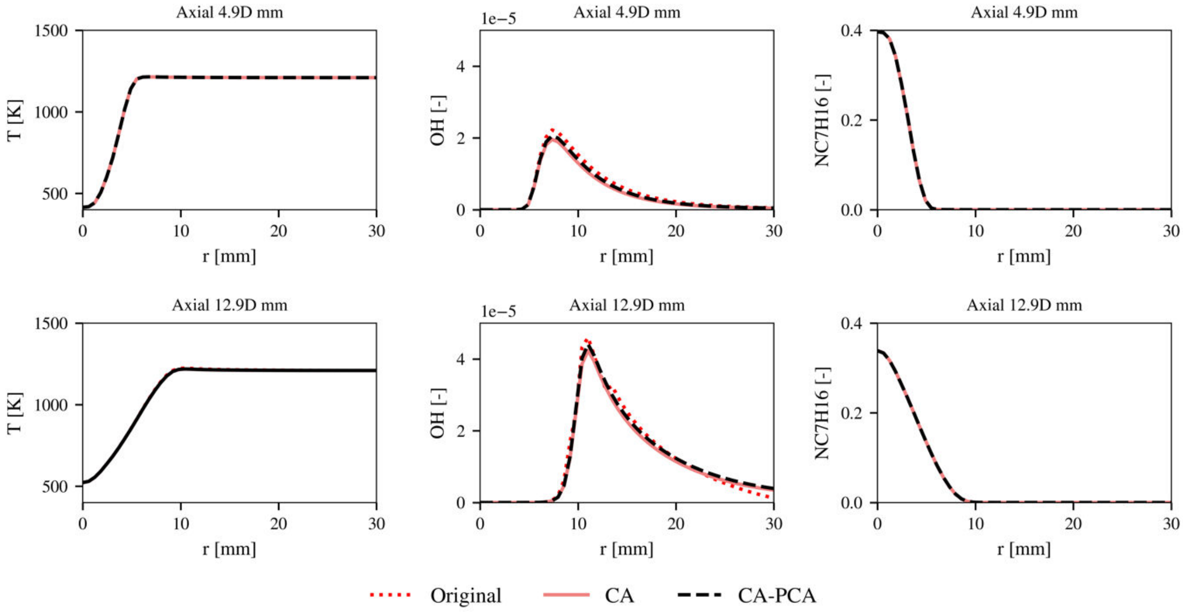

Figure 2 - Axial profiles of temperature, OH, and n-C_7H_{16} at 4.9D and 12.9D for the n-heptane flame. Figure adapted from [23].

Table 1 – Normalized RMSD between the original and CA simulations for the n-heptane flame.

ξ	CA (%)	CA-PCA (%)
T	0.0572	0.0650
OH	0.1776	0.2197
CO	0.1634	0.2015
CO_2	0.04456	0.04566
H_2O	0.07120	0.06923
O_2	0.03465	0.04441
n-C_7H_{16}	0.05425	0.05788

In the methane–hydrogen flame LES, both approaches capture the mean and RMS distributions of temperature and key species with high fidelity, as demonstrated in Figure 3 in terms of temperature, OH, and CO distributions along axial profiles. RMSD values remain below 0.6% across all scalars, as shown in Table 2, and the PCA-enhanced method exhibits marginal improvements for OH and CO. Although PCA introduces a modest overhead in this more complex configuration, the overall acceleration remains significant, and the method proves scalable through strategies such as reducing update frequency or limiting sampling. The CA reduces total time by 64.21% (87.15% chemistry), while CA-PCA achieves 60.10% (84.59% chemistry) with a 6.45% PCA cost. Overhead can be lowered via less frequent PCA updates or sampling fewer cells, making CA-PCA scalable for high-fidelity LES.

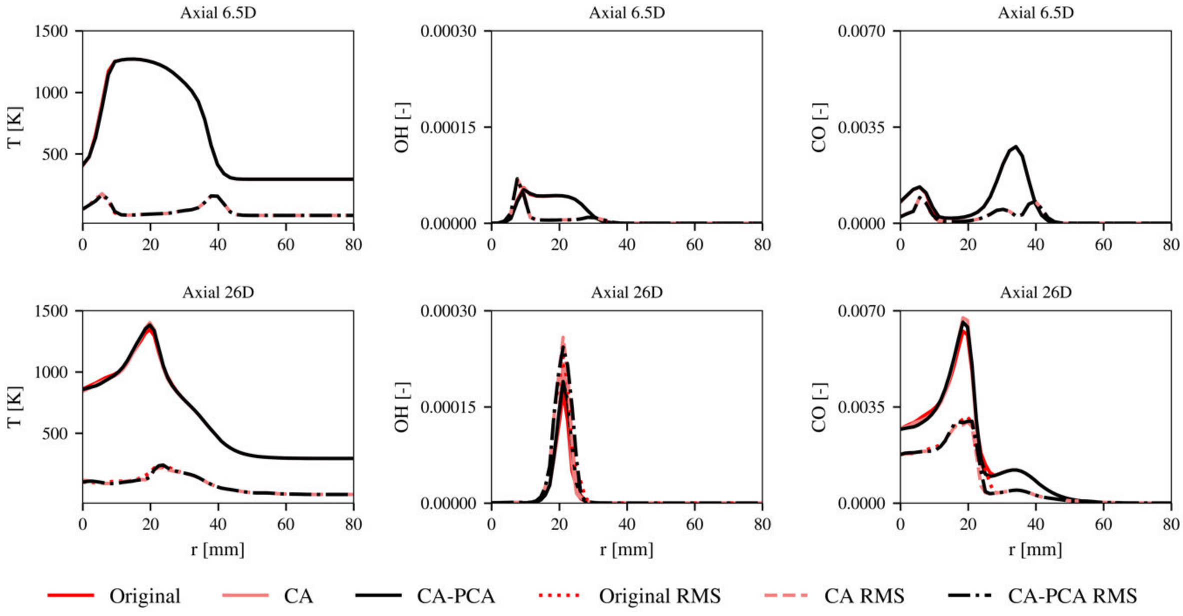

Figure 3 – Axial profiles of temperature, OH, and CO at 6.5D and 26D for the CH$_4$/H$_2$ flame. Figure adapted from [23].

Taken together, these results demonstrate that unsupervised algorithm integration through PCA not only preserves the accuracy of cell agglomeration but also enhances its adaptability and robustness across different combustion regimes by reducing the need for empirical tuning, while achieving a significant computational speedup. The method generalizes well beyond case-specific variable selection and offers a scalable pathway for accelerating chemistry in high-fidelity CFD simulations.

Table 2 – Normalized RMSD between detailed and CA simulations of the CH$_4$/H$_2$ flame.

ξ	CA (%)	CA-PCA (%)
T	0.5487	0.5107
OH	0.5209	0.4939
CO	0.4796	0.4434
CO$_2$	0.5413	0.5067
H$_2$O	0.4625	0.4315
O$_2$	0.2429	0.2424

Another related study by Rovira et al. [24] proposed an unsupervised workflow for extracting key features from high-dimensional reactive flow datasets. Their methodology consisted of three steps: dimensionality reduction, unsupervised clustering, and feature correlation. In the first step, two modern dimensionality reduction algorithms were applied. The first is t-distributed Stochastic Neighbor Embedding (t-SNE), a nonlinear method that maps high-dimensional data into a low-dimensional space by preserving local similarities between points. The second is Uniform Manifold Approximation and Projection (UMAP), which is also nonlinear but has stronger mathematical foundations, preserves both local and global structures, and is computationally more efficient.

For clustering, the study compared the classical k-means algorithm with Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Unlike k-means, which requires the number of clusters to be defined in advance and assumes spherical cluster shapes, HDBSCAN automatically determines the number of clusters, can detect arbitrarily shaped groups, and identifies noise points as outliers. In the final step, feature correlation analysis was carried out using mutual information, which identifies nonlinear dependencies and links each cluster to the most relevant thermochemical variables.

The evaluation on a counterflow reactor dataset demonstrated that both t-SNE and UMAP captured the main physical regions of the flow, including inlets, fast reaction zones, and slow reaction regions. However, UMAP offered several advantages: it preserved both local and global structures more effectively, produced tighter and more compact embeddings that improved clustering quality, and achieved up to 65–75% faster runtimes compared to t-SNE, particularly for larger meshes. In addition, UMAP embeddings enabled HDBSCAN to identify non-spherical clusters and correctly classify outliers, which t-SNE failed to separate reliably.

The effectiveness of this workflow is illustrated in Figure 4, which compares the two-dimensional embeddings produced by t-SNE and UMAP along with the corresponding clusters detected by HDBSCAN. While both methods distinguish the main regions, UMAP additionally identifies a distinct mixing zone beneath the jet nozzle and provides a clearer separation of outlier points, resulting in more physically meaningful clusters when mapped back to the reactor geometry. Finally, feature correlation analysis revealed the thermochemical variables most responsible for each cluster. For example, ozone concentration and velocity gradients were dominant in the fast reaction region, whereas N_2O_5 concentrations characterized the slow downstream regime. This step provided interpretability by linking the abstract clusters to physically relevant flow structures.

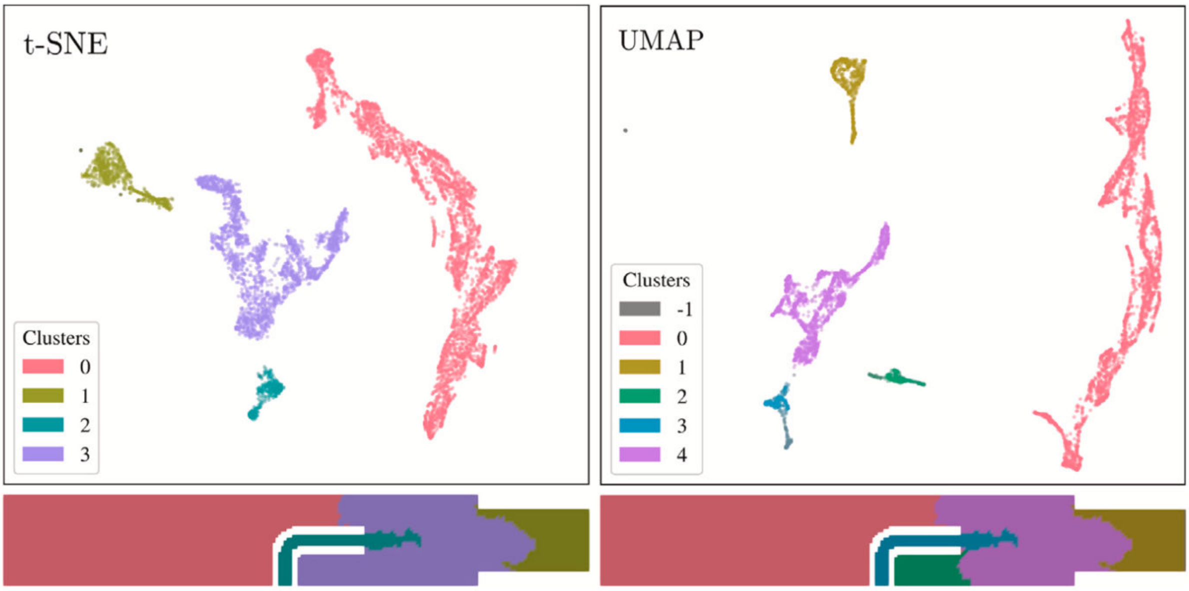

Figure 4 – Comparison of the region identification capabilities of t-SNE (left) and UMAP (right) when paired with the HDBSCAN clustering algorithm. The top row shows the two-dimensional mapping colored by the clusters found by HDBSCAN. It should be noted that the x and y axes have no labels as the exact values of the synthetic variables, which both t-SNE and UMAP reduce to have no intrinsic value. Hence, what is compared here are the shapes, distribution, and distances between clusters. The bottom row presents the location of the clusters within the reactor geometry. Figure adapted from [24].

The use of a combination of dimensionality reduction and clustering techniques was also investigated to identify precursors of intermittent hydrogen flashback [25], which were observed in LES of a reheat hydrogen combustor. In that work, a variant of PCA, called Co-Kurtosis PCA, was used to identify the thermochemical and hydrodynamic features which were most relevant in the onset of flashback. Then, from those identify features, modularity-based clustering was used to determine the specific combination of features acting as precursor to hydrogen flashback. Compared to previously mentioned clustering techniques, modularity-based clustering relies on a graph-interpretation of the evolution of the features, where nodes correspond to specific regions of the feature space and the edges represent the probability of transitioning between those regions. Modularity-based clustering then identifies clusters of nodes which have a strong intra-connectivity and weak extra-connectivity, resulting in isolated clusters of flashbacking states, clusters of normal states, and the precursors clusters which link these two groups of clusters. Relying on those precursors clusters, a warning system could be devised which warns of an impending flashback, as illustrated in Figure 5.

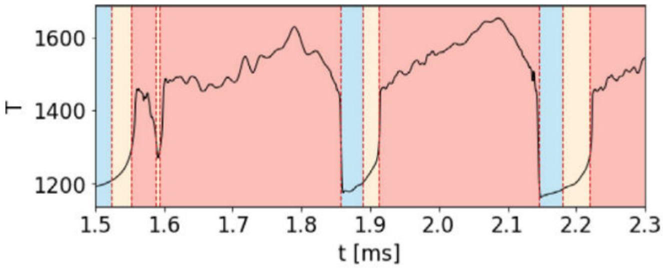

Figure 5 – Evolution of temperature during several flashback events. Background colors indicate the state the burner is identified to be in. Normal: blue, Precursor: orange and flashbacking: red. On average, the precursors provide ~34 μs of warning. Figure adapted from [25].

To conclude, the results of these studies [24,25] highlight that a combination of dimensionality reduction with clustering (UMAP with HDBSCAN or co-kurtosis PCA with modularity clustering) offer a robust and efficient unsupervised workflow. Compared with more traditional approaches such as PCA and k-means, these frameworks improve accuracy, scalability, and physical interpretability in the analysis of high-dimensional combustion data.

4. Adaptive Local Model Selection

Adaptive local model selection addresses the challenge of efficiently integrating detailed chemical kinetics into large-scale turbulent combustion simulations. In operator-splitting CFD solvers, the chemical step dominates the computational cost because each cell requires the integration of stiff nonlinear ODE systems involving many species. However, not all species or reactions are equally important in every region of the flame. Depending on the local thermo-chemical state, certain subsets of the mechanism may be inactive or redundant. This motivates a strategy in which different regions of the state space are associated with different reduced models, and the most suitable one is selected dynamically during runtime.

Traditional adaptive chemistry approaches, such as dynamic on-the-fly reduction, are limited by their high computational overhead, particularly when applied to large mechanisms or LES with millions of cells [26,27]. Pre-partitioning adaptive chemistry (PPAC) reduces this burden by constructing a library of reduced mechanisms beforehand, but its effectiveness strongly depends on how the training dataset is partitioned [28,29]. If the partitioning is not representative of the underlying dynamics, reduced mechanisms may become oversized or inaccurate, compromising both efficiency and reliability. Therefore, an automated and data-driven partitioning and classification strategy is essential.

In this context, Amaduzzi et al. proposed the Sample-Partitioning Adaptive Chemistry (SPARC) framework, which integrates unsupervised learning into both the clustering and classification stages [22]. First, the training dataset of one-dimensional flamelets is partitioned using LPCA. Unlike conventional clustering methods such as k-means or self-organizing maps, LPCA minimizes PCA reconstruction error and adapts the dimensionality of local manifolds, producing clusters that better capture the chemical variability of reacting flow data. Moreover, the number of clusters and local principal components are not fixed manually but optimized automatically using Bayesian optimization, thereby reducing reliance on expert intervention.

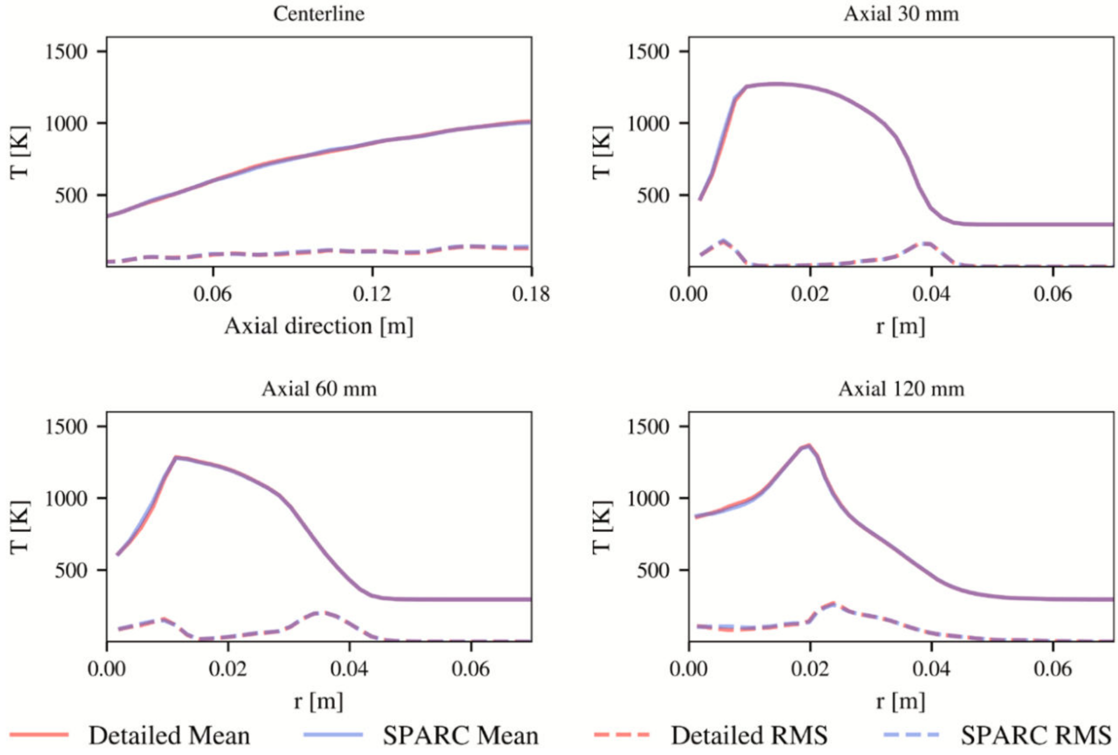

Figure 6 – Comparison of resolved mean and RMS of temperature of the detailed and SPARC LES. Figure adapted from [22].

During runtime, each computational cell is classified into one of the pre-computed clusters by the Vector Quantization Principal Component Analysis (VQPCA) algorithm [1]. This unsupervised classifier assigns each state to the cluster that yields the smallest reconstruction error, ensuring that the most appropriate reduced mechanism is selected at every timestep. The rationale for choosing VQPCA lies in its balance between accuracy and computational cost. It avoids heuristic similarity metrics or hand-crafted thresholds, while its reconstruction-error-based assignment directly reflects the quality of representation by each cluster.

The application to the AJHC flame demonstrated the benefits of this approach. Figure 6 illustrates this performance for temperature profiles, comparing mean and RMS values along radial and axial lines. The SPARC simulation with VQPCA closely follows the detailed LES, with no significant deviations observed across the flame. Similarly, Figure 7 indicates the comparison for OH and CO, two critical intermediate species in MILD combustion regimes. The adaptive approach captures both mean values and fluctuations with good fidelity, although minor discrepancies appear in the downstream region for OH. Importantly, these differences remain within the experimental uncertainty range and do not compromise the overall predictive accuracy. Together, these figures highlight that VQPCA-based adaptive model selection enables substantial computational savings while maintaining detailed-level agreement in both global scalars (temperature) and sensitive intermediates (OH and CO).

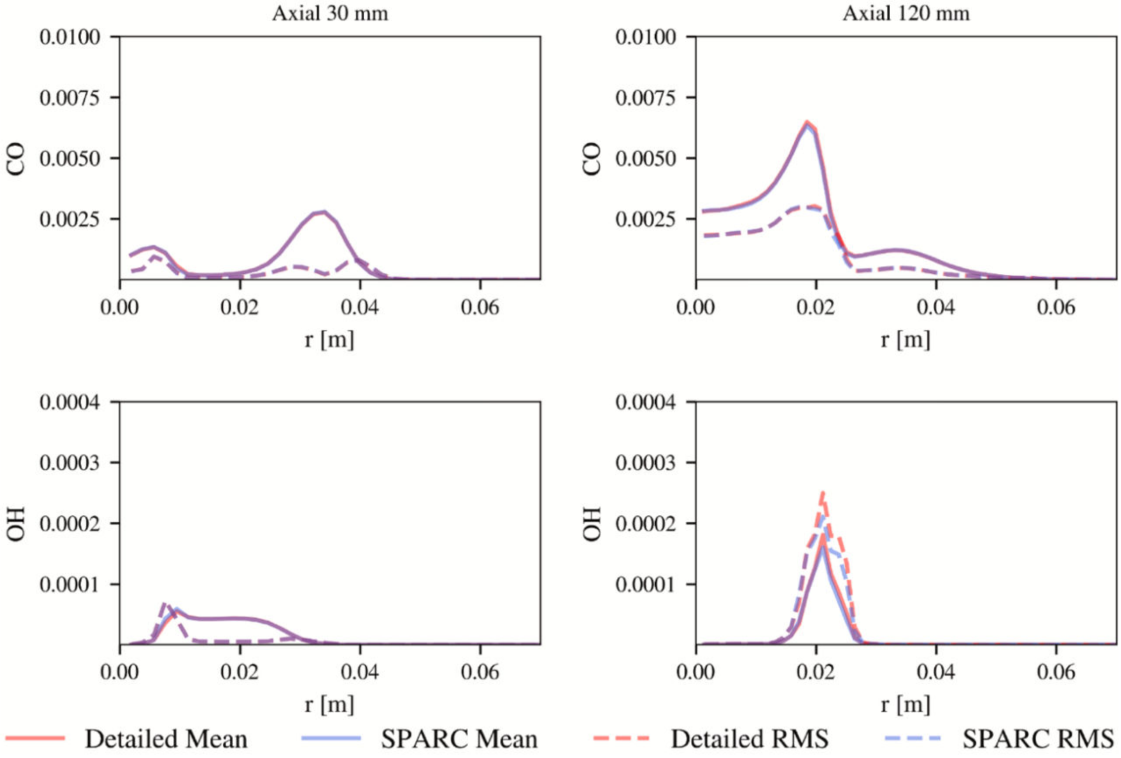

Figure 7 – Comparison of resolved mean and RMS of OH and CO mass fractions of the detailed and SPARC LES. Figure adapted from [22].

Additionally, quantitative results demonstrated that the adaptive method maintained accuracy, with normalized RMS deviations of less than 1.5% for temperature and major species compared to detailed LES, while reducing the average number of species per integration step from 36 to 24. This translated into a 2.2× speedup of the chemical integration.

Further insight into the role of adaptive local model selection is provided in Figure 8, which shows the instantaneous cluster assignment across the computational domain. The classifier clearly captures the three-stream mixing structure of the AJHC burner, with chemically complex regions assigned to larger skeletal mechanisms (up to 35 species) and diluted regions described with significantly reduced models (as low as 20 species). This spatial organization confirms that VQPCA not only reduces computational cost but also provides a physically consistent mapping of local chemical complexity. In particular, the algorithm identifies zones of intense turbulence–chemistry interaction, such as the air–fuel–coflow mixing layers, where more detailed models are required for accurate predictions.

Together, these results highlight how VQPCA-based adaptive model selection provides a scalable and automated way to accelerate chemistry while preserving detailed-level fidelity. Results demonstrate that the approach not only reproduces global flame characteristics but also adapts locally to the chemical activity of each region, ensuring both efficiency and accuracy in LES of turbulent combustion.

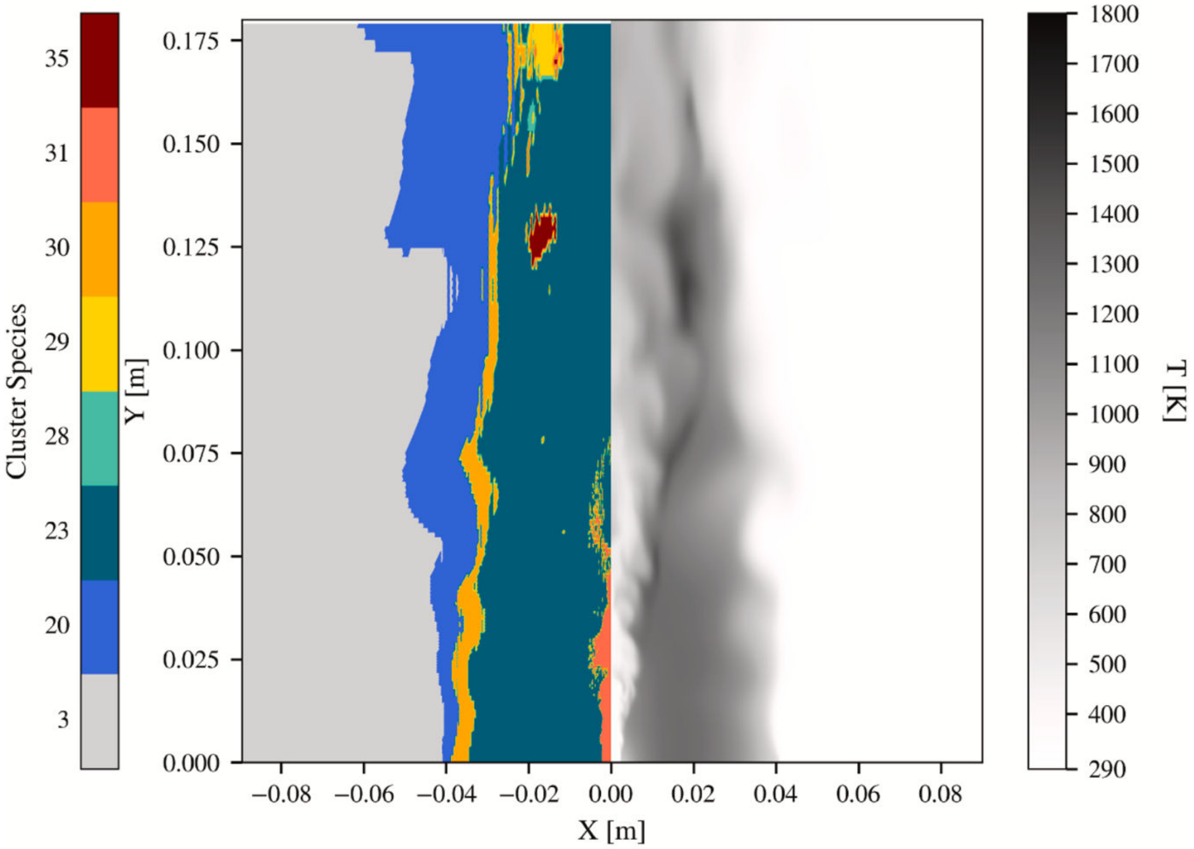

Figure 8 – Instantaneous 2D slice of the SPARC simulation, contours of the cluster number of species (left, colored) and instantaneous temperature (right, greyscale). Figure adapted from [22].

Overall, the adoption of VQPCA for adaptive local model selection reflects a broader motivation, to achieve robust, automated, and generalizable chemistry acceleration without extensive user tuning. By directly leveraging unsupervised learning, the method reduces computational cost while maintaining predictive fidelity, making high-fidelity LES with detailed kinetics feasible for more complex combustion systems.

5. Conclusion

This report has reviewed the integration of unsupervised learning techniques into computational fluid dynamics for reactive flow simulations, focusing on their role in clustering, dimensionality reduction, and adaptive model selection. The findings emphasize that unsupervised algorithms provide a systematic and data-driven approach to extract meaningful patterns from high-dimensional combustion data, enabling both computational acceleration and improved interpretability.

Clustering approaches such as k-means, LPCA, HDBSCAN and modularity clustering have proven effective for identifying coherent structures and physically important features, and automating the generation of chemical reactor networks, reducing the dependence on expert intervention while maintaining predictive accuracy. Dimensionality reduction methods, including PCA, co-kurtosis PCA and more recent nonlinear techniques like UMAP, have demonstrated their ability to reveal low-dimensional manifolds and support scalable clustering and feature analysis. Adaptive model selection strategies based on unsupervised classification, such as the use of VQPCA within SPARC, show particular promise by dynamically reducing chemical complexity while preserving fidelity in large-scale simulations.

Despite these advances, challenges remain regarding the sensitivity of results to algorithmic choices, hyperparameters, and similarity definitions, which can limit generalization across combustion regimes. Dimensionality reduction also introduces trade-offs, as variance truncation may reduce accuracy, while the interpretability of nonlinear embeddings remains a concern when applied to turbulent flow–chemistry interactions.

Looking ahead, several opportunities exist for further development. Hybrid learning strategies that combine unsupervised methods with physics-based constraints can help ensure consistency and interpretability. The design of scalable algorithms tailored to massive CFD datasets will be necessary to extend applicability to industrially relevant systems. Advances in representation learning, including self-supervised approaches, may uncover richer latent structures and improve adaptability across operating conditions. Embedding conservation laws and thermochemical constraints into unsupervised frameworks can also increase their physical reliability.

In conclusion, unsupervised learning represents a promising direction for the next generation of CFD frameworks. By integrating data-driven insights with physical modeling, these methods have the potential to achieve simulations that are not only faster and more efficient but also more adaptive and interpretable, paving the way for high-fidelity combustion modeling in increasingly complex applications.

Acknowledgments

This deliverable is based upon work from COST Action CA22151 – CYPHER, supported by COST (European Cooperation in Science and Technology). COST is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. For more information, visit www.cost.eu.

References

[1] N. Kambhatla and T. Leen, “Dimension reduction by local principal component analysis,” *Neural Computation*, vol. 9, pp. 1493–1516, 1997.

[2] H. Dave, N. Swaminathan, and A. Parente, “Interpretation and characterization of mild combustion data using unsupervised clustering informed by physics-based, domain expertise,” *Combust. Flame*, vol. 240, p. 111954, 2022.

[3] A. Coussement, O. Gicquel, and A. Parente, “Mg-local-pca method for reduced order combustion modeling,” Proc. Combust. Inst., vol. 34, no. 1, pp. 1117–1123, 2013.

[4] G. D’Alessio, A. Parente, A. Stagni, and A. Cuoci, “Adaptive chemistry via pre-partitioning of composition space and mechanism reduction,” *Combust. and Flame*, vol. 211, pp. 68–82, 2020.

[5] G. D’Alessio, A. Cuoci, G. Aversano, M. Bracconi, A. Stagni, and A. Parente, “Impact of the partitioning method on multidimensional adaptive-chemistry simulations,” *Energies*, vol. 13, no. 10, p. 2567, 2020.

[6] A. Parente, J. Sutherland, L. Tognotti, and P. Smith, “Identification of low-dimensional manifolds in turbulent flames,” Proc. Combust. Inst., vol. 32, no. 1, pp. 1579–1586, 2009.

[7] A. Parente, J. C. Sutherland, B. B. Dally, L. Tognotti, and P. J. Smith, “Investigation of the mild combustion regime via principal component analysis,” Proc. Combust. Inst., vol. 33, no. 2, pp. 3333–3341, 2011.

[8] Z. Li, S. Tomasch, Z. X. Chen, A. Parente, I. S. Ertesvåg, and N. Swaminathan, “Study of mild combustion using LES and advanced analysis tools,” Proc. Combust. Inst., vol. 38, no. 4, pp. 5423–5432, 2021.

[9] L. S. Lloyd, “Least squares quantization in PCM,” IEEE Trans. Inf. Theory, vol. 28, no. 2, pp. 129–137, 1982.

[10] W. Yu, F. Zhao, W. Yang, and H. Xu, “Integrated analysis of CFD simulation data with k-means clustering algorithm for soot formation under varied combustion conditions,” Appl. Therm. Eng., vol. 153, pp. 299–305, 2019.

[11] M. Savarese, A. Cuoci, W. D. Paepe, A. Parente, “Machine learning clustering algorithms for the automatic generation of chemical reactor networks from CFD simulations,” Fuel, vol. 343, 127945, 2023.

[12] T. Lesaffre, J. Wirtz, E. Riber, B. Cuenot, and Q. Douasbin, “Lean blowoff dynamics in bluff body stabilized flames: unsupervised classification and balance analysis,” Proc. Combust. Inst., vol. 40, no. 1–4, 105691, 2024.

[13] S. Pope, “Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation,” Combust. Theory Model., vol. 1, no. 1, pp. 41–63, 1997.

[14] Z. Ren and S. B. Pope, “Second-order splitting schemes for a class of reactive systems,” J. Comput. Phys., vol. 227, no. 17, pp. 8165–8176, 2008.

[15] Z. R. Graham, M. Goldin, and S. Zahirovic, “A cell agglomeration algorithm for accelerating detailed chemistry in CFD,” Combust. Theory Model., vol. 13, no. 4, pp. 721–739, 2009.

[16] S. Feng and H. Zhang, “Use of dynamic adaptive chemistry and dynamic cell clustering in computational fluid dynamics to accelerate calculation of combustion simulation of diesel engine,” Fuel, vol. 338, 127360, 2023.

[17] A. Stock, V. Moureau, J. Leparoux, and R. Mercier, "Low-cost Jacobian-free mapping for dynamic cell clustering in multi-regime reactive flows," Proc. Combust. Inst., vol. 40, no. 1, 105287, 2024.

[18] A. Cuoci, A. Nobili, A. Parente, T. Grenga, and H. Pitsch, "Tabulation-based sample-partitioning adaptive reduced chemistry and cell agglomeration," Proc. Combust. Inst., vol. 40, no. 1, 105386, 2024.

[19] B. Dally, A. Karpetis, and R. Barlow, "Structure of turbulent non-premixed jet flames in a diluted hot coflow," Proc. Combust. Inst., vol. 29, no. 1, pp. 1147–1154, 2002.

[20] J. Ye, P. R. Medwell, M. J. Evans, and B. B. Dally, "Characteristics of turbulent n-heptane jet flames in a hot and diluted coflow," Combust. Flame, vol. 183, pp. 330–342, 2017.

[21] A. S. Newale, P. Sharma, S. B. Pope, and P. Pepiot, "A feasibility study on the use of low-dimensional simulations for database generation in adaptive chemistry approaches," Combust. Theory Model., 2022, pp. 1–23.

[22] R. Amaduzzi, G. D'Alessio, P. Pagani, A. Cuoci, R. Malpica Galassi, and A. Parente, "Automated adaptive chemistry for large eddy simulations of turbulent reacting flows," Combust. Flame, vol. 259, 113136, 2024.

[23] Y. Yalcinkaya, R. Amaduzzi, A. Cuoci, and A. Parente, "Reducing Computational Costs in Turbulent Reacting Flow Simulations Using Cell Agglomeration," Appl. Energy Combust. Sci., (under review).

[24] M. Rovira, K. Engvall, and C. Duwig, "Identifying key features in reactive flows: A tutorial on combining dimensionality reduction, unsupervised clustering, and feature correlation," Chem. Eng. J., vol. 438, 135250, 2022.

[25] M. Floris, T. Shiva Sai, D. Nayak, I. Langella, K. Aditya, and N. A. K. Doan, "Data-driven identification of precursors of flashback in a lean hydrogen reheat combustor," Proc. Combust. Inst., vol. 40, no. 1-4, 105524, 2024.

[26] L. Liang, J. G. Stevens, and J. T. Farrell, "A dynamic adaptive chemistry scheme for reactive flow computations," Proc. Combust. Inst., vol. 32, no. 1, pp. 527–534, 2009.

[27] Z. Ren, C. Xu, T. Lu, and M. A. Singer, "Dynamic adaptive chemistry with operator splitting schemes for reactive flow simulations," J. Comput. Phys., vol. 263, pp. 19–36, 2014.

[28] Y. Liang, S. B. Pope, and P. Pepiot, "A pre-partitioned adaptive chemistry methodology for the efficient implementation of combustion chemistry in particle PDF methods," Combust. Flame, vol. 162, no. 9, pp. 3236–3253, 2015.

[29] A. S. Newale, Y. Liang, S. B. Pope, and P. Pepiot, "A combined PPAC-RCCE-ISAT methodology for efficient implementation of combustion chemistry," Combust. Theory Model., vol. 23, no. 6, pp. 1021–1053, 2019.