Higher order dynamic mode decomposition to model reacting flows

Adrián Corrochano\textsuperscript{a,}, Giuseppe D’Alessio\textsuperscript{b,c}, Alessandro Parente\textsuperscript{b}, Soledad Le Clainche\textsuperscript{a}

\textsuperscript{a}School of Aerospace Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain
\textsuperscript{b}Université Libre de Bruxelles, École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Bruxelles, Belgium
\textsuperscript{c}CRECK Modeling Lab, Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Abstract

In this work, the application of the multi-dimensional higher order dynamic mode decomposition (HODMD) is proposed for the first time to analyse combustion databases. In particular, HODMD has been adapted and combined with other pre-processing techniques (generally used in machine learning), in light of the multivariate nature of the data. A truncation step separate the main dynamics driving the flow from less relevant non-linear dynamics. The method is applied to analyse a database obtained from a Computational Fluid Dynamics (CFD) simulation of an axisymmetric, time varying, non-premixed, co-flow methane flame carried out by means of a detailed kinetic mechanism. Results show that HODMD can reconstruct the main jet dynamics with a reduced number of relevant modes, able to reproduce the system dynamics. These modes are found to be representative for the main flow physics with two main advantages: (i) they provide for the possibility to achieve a strong simplification with respect to the high-dimensional input data, and at the same time (ii) a small reconstruction error with respect to the original dataset is observed. In addition, the method was also validated considering a reduced matrix obtained using Principal Component Analysis (PCA) based feature selection and the Varimax rotation. This validation also reveals that it is not important to have all the variables in the dataset, just a group of them is necessary to obtain the main dynamics of the system. This has an impact on feature selection and on the cost these methodologies for very massive data.

Keywords: Combustion; Reacting flows; higher order dynamic mode decomposition; reduced order models

1. Introduction

Fighting the climate change is one of the major challenges of our society. Reducing the pollutants produced from the combustion of fossil fuels, improving the energy efficiency in industrial devices, or producing new smart energy carriers, are only few examples from the multiple research topics in this field. A strong effort is carried out by the research community to study in detail and to understand the flow physics of reacting flows, motivated by the need to improve the performance and the efficiency of the combustion process and with the aim at finding new eco-friendly alternatives to the current use of fossil fuels.

However, the high complexity associated to the combustion process, from both a physical and chemical point of view, makes it challenging to perform detailed experimental or numerical investigations. On the one hand, carrying out experimental measurements is very complicated because of the small characteristic time scales and the complex chemical reactions driving the dynamics in reacting flows, and also because of the high cost to the use of specialized and sophisticated infrastructures. On the other hand, the large number of species involved in combustion processes, the range of scales and the non-linear turbulence-chemistry impose the use of large computer facilities to obtain accurate and extensive databases allowing to study in detail the complex physics of the flow to understand turbulent combustion [1]. Moreover, the enormous computational cost (CPU time and memory) associated to solve industrial problems limits the advancement in the field. A good alternative to overcome such limitations is using reduced order models (ROMs). ROMs have lately spread within the field of reacting flows, as they offer the opportunity to

\*Corresponding author

Preprint
simplify the complex nature of the physical problem, as well as to create high fidelity lower-dimensional models which represent the physics with a low level of error [2–4]. Using ROMs, it is possible to model the main flow dynamics with a reduced computational cost.

Principal Component Analysis (PCA) [2, 5–7] has found widespread application in combustion. PCA is a statistical technique whose aim is to find lower-dimensional manifolds by identifying correlations between the input variables, which are then combined linearly to eventually reduce the problem dimension [8, 9]. The application of PCA has been proved suitable for dimensionality reduction and feature extraction [2, 3, 7, 10], clustering [11, 12], and data analysis and feature selection [13–17].

More recently, dynamic mode decomposition (DMD) [18] has also been introduced as a good tool to extract patterns and develop ROMs in reacting flows. The main advantage of this technique lies in its ability to identify the main frequencies and features, known as DMD modes, leading the flow dynamics of the problem studied. Such modes could be related to flow instabilities reflecting the underlying physical mechanisms driving the flow [19]. Hence, DMD can potentially complement PCA, while considering physical principles governing the flow to develop new efficient ROMs. Some examples of DMD applied to the analysis of reacting flows in the last decade are detailed as follows. Richcoeur et al. [20] applied DMD to study the main dynamics in turbulent combustion, studying in detail the performance of DMD when applied to the analysis of experimental databases. Souvick et al. [21] used DMD in experimental databases to identify the main patterns connected to flow stabilities in a swirl-stabilised dump combustor. In the same line, several authors have used DMD to identify the main flow patterns, linked with the main instabilities present in the flow. For instance, in the analysis of experimental databases, Quinlan et al. [22] developed a technique based on DMD to extract the main patterns and frequencies related to the resonant modes producing an induced transverse combustion instability; Huang et al. [23] used DMD to study of self-excited longitudinal combustion instabilities in laboratory-scaled single-element gas turbine and rocket combustors. Motheau et al. [24] used DMD for the analysis of a LES database, with the aim at developing a low-order model showing that one of the mechanism triggering combustion instability is related to convected and acoustic entropy waves. More applications of DMD to identify patterns and flow instabilities in reactive flows include the work by Abou-Taouk et al. [25], who studied a V-flame in an afterburner-type configuration; the work by Ghani et al. [26], who identified the main patterns in an industrial gas turbine combustion chamber; and the work by Grenga et al. [27, 28], who proposed a more efficient implementation of DMD, suitable to elucidate multiphysics requiring with a reduced memory usage, which was tested in a DNS database solving a turbulent premixed flame problem.

As explained in Ref. [29], the application of DMD is limited to some specific conditions of the database analysed and the physical mechanism underlying the problem under study. Hence, in the analysis of complex dynamical systems, noisy experimental databases, DMD suffers in identifying robust and accurate results modelling the flow dynamics, or may even fail. Higher Order Dynamic Mode Decomposition (HODMD) [30] is introduced as an alternative to overcome such limitations. HODMD is a more robust extension of DMD, which has been proven suitable to the analysis of non-reactive complex flows and noisy experimental data in several complex and industrial applications [31–34]. The performance of HODMD showed in the previous applications, makes it a suitable tool for the analysis of reactive flows and turbulent combustion. This article introduces for the first time, to the authors knowledge, the application of HODMD to analyse a combustion database. More specifically, HODMD algorithm is extended for the analysis of reacting flows, formed by a large number of variables and species, with different magnitudes. The method is able to reduce the data dimensionality and to identify the main frequencies and associated patterns driving the flow dynamics, with the aim at developing a ROM. This application of HODMD is illustrated in a numerical database of a laminar reacting flow representative for the oxidation of a methane jet in air. The results are validated with reduced dimensionality data obtained with PCA. Since HODMD is fully data-driven, this algorithm could be easily extended to the analysis of other type of flows.

The article is organised as follows. Section 2 introduces the methodology used to develop this work. The description of the database generated numerically is introduced in Section 3. Finally, the main results and conclusions are presented in Section 4 and Section 5, respectively.

2. Methodology

This section introduces the theoretical aspects of the two algorithms that are applied in this paper to identify the main patterns and frequencies driving the flow dynamics: the higher order dynamic mode decomposition [30] and
the principal component analysis (PCA) [5], both combined with pre-processing strategies generally used in machine learning. In particular, PCA will also be coupled with the Varimax rotation method [35] to identify a subset of main variables from a database containing a larger number of variables.

2.1. Higher Order Dynamic Mode Decomposition

Higher order dynamic mode decomposition (HODMD) [30] is a data-driven method introduced as a robust extension of dynamic mode decomposition (DMD) [18] for the analysis of complex non-linear dynamical systems. Similarly to DMD, HODMD decomposes spatio-temporal data \( v(x, y, z, t_k) \), collected at time instant \( t_k \) (for convenience expressed as \( v_k \)), as an expansion of \( M \) Fourier-like modes. For two-dimensional databases (as presented in this article), the decomposition reads

\[
v(x, y, t_k) \approx v_k^{\text{approx}} = \sum_{m=1}^{M} a_m u_m(x, y)e^{i \omega_m t_k + i \delta_m}, \quad k = 1, \ldots, K,
\]

where \( u_m \) are the DMD modes and \( a_m, \omega_m, \delta_m \) are their associated amplitudes, frequencies and growth rates. From a practical point of view, it is necessary to collect a group of data equi-distant in time with time interval \( \Delta t \), in the following snapshot matrix

\[
V_1^K = [v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_{K-1}, v_K].
\]

with dimension \( J \times K \), with \( J = N_{\text{variable}} \times N_x \times N_y \), where \( N_{\text{variable}} \) are the number of variables considered in the database, and \( N_x \) and \( N_y \) are the number of grid points along the streamwise and normal (or radial) directions. Hence, for a vector containing \( j \) variables, each variable, formed by vectors collected at time \( t_k \), is included in consecutive rows in the snapshot matrix (2). In what follows, when the snapshot matrix \( V_1^K \) formed by a single variable \( j \), it is referred as \( v_j \).

HODMD relies on the following high-order Koopman assumption, which relates \( d \) (tunable) subsequent snapshots as

\[
v_{k+d} = R_i v_k + R_{j} v_{k+1} + \ldots + R_d v_{k+d-1}, \quad k = 1, \ldots, K - d,
\]

where \( R \) (for \( i = 1, \ldots, d \)) are the (so-called) Koopman operators, which are linear and contain the dynamics of the system. This is the essence of HODMD, which provides high accurate results, even in the analysis of highly complex databases. When \( d = 1 \), the Koopman assumption approximates the standard DMD algorithm. Details about the HODMD algorithm can be found in [30, 36], and the Matlab codes can be found in [37].

In what follows, the accuracy of the approximation of the original database \( v_k^{\text{approx}} \) using the DMD expansion eq. (1) will be measured in terms of the relative root mean square (RRMS) error as

\[
\text{RRMS error} = \sqrt{\frac{\sum_{k=1}^{K} \|v_k^{\text{approx}} - v_k\|_2^2}{\sum_{k=1}^{K} \|v_k\|_2^2}}.
\]

2.2. HODMD in combustion flows

The robustness and accuracy of HODMD has been tested in a wide range of applications, including the analysis of complex flows (i.e., noisy experiments[36], transitional flows[38], turbulent flows[34]) to create reduced order models[39] or to extract patterns and to study the flow physics[29]. However, the analysis was limited to database including the three velocity components and pressure. In combustion systems, it is common to deal with a larger number of variables with different magnitudes. Hence, it is necessary carefully pre-processing the data before starting with the analysis, with the aim at extracting the maximum quantity of relevant information from the problem studied. For such aim, this article introduces a three-steps method based on HODMD, which summarizes the novel application of this algorithm for the analysis of combustion data. Moreover, different strategies to reduce the data dimensionality are introduced and discussed. The three-step method is defined as follows.

**Step 1: Centering and scaling.**

Combustion data usually consist of variables that have different units and ranges, such as temperature and chemical species, which therefore entails the need to pre-process them before applying any data-driven algorithm.
Centering consists on the subtraction of the mean value for a given variable to focus on the fluctuations. Scaling the variables to a uniform range is also important because it is common to have order of magnitudes of difference among the different variables. The aforementioned pre-processing operations are accomplished as follows:

$$\tilde{v}_j = \frac{v_j - \bar{v}_j}{c_j},$$  \hspace{1cm} (5)

where $\bar{v}_j$ represents the temporal mean of the variable $v_j$, and $c_j$ is the scaling factor. In this paper, two scaling methods are considered:

- Auto scaling: Uses the standard deviation of each variable, $\sigma_j$, as the scaling factor.
- Range scaling: Uses the difference between the maximum and minimum value of each variable as the scaling factor.

As observed by Parente and Sutherland [5], some scaling methods may emphasize the relevance of some variables because of their distribution. Auto scaling gives similar importance to all the different variables, while range scaling performs better for most of the major species.

**Step 2: Dimension reduction via HOSVD.**

A two-dimensional database can be re-organized into a fourth-order tensor $A_{i_1i_2i_3k}$ (with indexes associated to each one of the components, for $i_1 = 1, \ldots , N_{\text{variables}}$, $i_2 = 1, \ldots , N_x$, $i_3 = 1, \ldots , N_y$, and $k = 1, \ldots , K$), in which each one of the four dimensions correspond to the variables of the dataset, the two spatial components (namely, streamwise, $x$, and normal or radial, $y$, components) and the temporal component, $t$, respectively. The high order singular value decomposition (HOSVD) method [40] can then be used to reduce the data dimensionality in a highly efficient way. This algorithm applies standard SVD to each one of the fibers composing the tensor (see details and the Matlab code to perform this analysis in Ref. [37]), resulting in an expansion as

$$A_{i_1i_2i_3k} = \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} \sum_{p_3=1}^{P_3} S_{p_1p_2p_3} U_{i_1p_1} W^{(x)}_{i_2p_2} W^{(y)}_{i_3p_3} T_{kn},$$  \hspace{1cm} (6)

where $S_{p_1p_2p_3}$ is a fourth-order tensor (called the core tensor) and the columns of the matrices $U$, $W^{(x)}$, $W^{(y)}$, and $T$ are known as the SVD modes of the decomposition (related to the number of components, called as the component modes, and the streamwise and normal/radial spatial dimension, called as the spatial modes, respectively). Matrix $T$ corresponds to the SVD temporal modes matrix, the reduced snapshot matrix on which the HODMD algorithm is applied, as described in the next Step. The number of SVD modes, $N_e$, associated to the temporal matrix $T$ are calculated based on the tolerance (tunable) $\varepsilon$ as

$$\sigma_{N_e+1}/\sigma_1 \leq \varepsilon.$$  \hspace{1cm} (7)

The SVD modes, associated to the remaining matrices, $P_1$, $P_2$ and $P_3$, are calculated in a similar way for each one of the group of modes. Generally, in standard fluid dynamics applications, the first component of the tensor, $U$, is not truncated, since the number of variables is (at most) 4 (pressure and three velocity components). However, considering the typical size of combustion applications, we also truncate this first dimension, evaluating the influence of this truncation in the reconstruction error as in eq. (4). In this way, it is possible to identify what is the influence of the different variables in the system dynamics.

A new variable will quantify the compression carried out by the HOSVD application in the original database, the compression factor ($CF$), defined as

$$CF = \frac{N_e}{N_{\text{tot}}}$$  \hspace{1cm} (8)

where $N_e$ is the total number of elements of the tensor (namely, $N_e = N_{\text{variables}} \times N_x \times N_y \times K$), and $N_{\text{tot}}$ is the number of elements retained after reducing the data dimensionality with HOSVD.
Step 3: The DMD-d algorithm.

The DMD-d algorithm is applied to the matrix associated with the temporal modes $T$, referred as $\hat{T}_1^K$, of dimension $N \times K$, with $N$ the number of SVD temporal modes retained, also known as spatial complexity of the system, and $K$ the number of snapshots forming the original dataset. For consistency, to compare with the algorithm described in the literature [36], this matrix is called as the reduced snapshot matrix. The high order Koopman assumption eq. (3) is applied to this snapshot matrix as

$$\hat{T}_{d+1}^K \approx R_1 \hat{T}_1^K + R_2 \hat{T}_2^K + \ldots + R_d \hat{T}_d^K.$$  

This equation divides the snapshot matrix into $d$ blocks, each one containing time-delayed $K - d$ snapshots.

The previous equation is rewritten in terms of the modified Koopman matrix $\hat{R}$ and the modified snapshot matrix $\hat{T}_1^{K-d+1}$ as

$$\hat{T}_2^{K-d+1} = \hat{R} \hat{T}_1^{K-d},$$  

also written as

$$\begin{bmatrix}
\hat{T}_2^{K-d+1} \\
\hat{T}_3^{K-d+1} \\
\vdots \\
\hat{T}_d^{K-d+1}
\end{bmatrix} = 
\begin{bmatrix}
0 & 1 & 0 & \ldots & 0 & 0 \\
0 & 0 & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 1 & 0 \\
\hat{R}_1 & \hat{R}_2 & \hat{R}_3 & \ldots & \hat{R}_{d-1} & \hat{R}_d
\end{bmatrix}
\begin{bmatrix}
\hat{T}_1^{K-d} \\
\hat{T}_2^{K-d} \\
\vdots \\
\hat{T}_d^{K-d}
\end{bmatrix}.$$  

This new matrix, which organizes the snapshot blocks identified in the high order Koopman assumption in columns, increases the spatial complexity of the data to a value $N' > N$, increasing in this way the accuracy in the calculations of the DMD modes. This step justifies the better performance of HO-DMD compared to DMD and other variants, which are all based on a Koopman expansion eq. (3), with $d = 1$.

A new dimension-reduction is carried out into this matrix, to eliminate possible redundancies, via truncated SVD using again the tolerance $\varepsilon$ in the new calculated singular values as

$$\sigma_{N'+1} / \sigma_1 < \varepsilon,$$  

where $N' > N$ (new spatial complexity) is the number of retained SVD modes. Hence, the modified snapshot matrix is obtained as

$$\hat{T}_1^{K-d+1} = \hat{U} \hat{\Sigma} \hat{V}^T \approx \hat{U} \hat{\Sigma} \hat{V}^T,$$  

with $\hat{T}_1^{K-d+1} = \hat{U} \hat{\Sigma} \hat{V}^T$, where $\hat{U}^T \hat{U} = \hat{T}^T \hat{T}$ are the $N' \times N'$-unit matrices, and the diagonal of matrix $\hat{\Sigma}$ contains the singular values $\sigma_1, \ldots, \sigma_{N'}$. To complete this step, eq. (10) is pre-multiplied by $\hat{U}^T$, which invoking (13) results as

$$\hat{T}_2^{K-d+1} = \hat{R} \hat{T}_1^{K-d}.$$  

The new $N' \times N'$-Koopman matrix is related to $\hat{R}$ by $\hat{R} = \hat{U}^T \hat{R} \hat{U}$. Nevertheless, we use the pseudoinverse in eq.(14) (via non-truncated SVD) to the matrix $\hat{T}_1^{K-d}$, instead of computing the previous expresion, which yields

$$\hat{T}_1^{K-d} = \hat{U} \hat{\Lambda} \hat{V}^T,$$  

where $\hat{U} \hat{U}^T = \hat{U}^T \hat{U} = \hat{V}^T \hat{V}$ are the $N' \times N'$-unit matrices and the diagonal of $\hat{\Lambda}$ contains the $N'$ singular values. Substituting (15) into (14) and post-multiplying by $\hat{V} \hat{\Lambda}^{-1} \hat{U}^T$, it is possible obtaining the following equation

$$\hat{R} = \hat{T}_2^{K-d+1} \hat{V} \hat{\Lambda}^{-1} \hat{U}^T.$$  

The reduced DMD expansion for the reduced snapshots is then calculated using $N'$ eigenvectors $\hat{v}_m$ and eigenvalues $\mu_m$ solving the eigenvalue problem of $\hat{R}$, as

$$\hat{v}_k \approx \sum_{m=1}^M \hat{c}_{m} \hat{u}_m e^{(\delta_{m} + i \omega_{m}) t},$$  

where $\delta_{m}$ and $\omega_{m}$ are the real and imaginary parts of the $m$-th eigenvalue, respectively.
for \( k = 1, \ldots, K \). Retaining only the first \( N \) components of the vectors \( \tilde{q}_m = \tilde{U}_q \tilde{a}_m \), it is possible to approximate the reduced DMD modes \( \hat{u}_m \) (dimension \( Nd \)). The damping rates \( \delta_m \) and frequencies \( \omega_m \) are obtained as

\[
\delta_m + i\omega_m = \frac{\log(\mu_m)}{\Delta t}.
\]

The amplitudes \( \hat{a}_m \) are calculated as in optimized DMD [41], solving a least square fitting problem in eq.(17). Invoking (6) and rescaling with unit-norm the modes \( u_m \) and the amplitudes \( a_m \), it is possible to obtain the original DMD expansion (1). Finally, using a second tolerance \( \varepsilon_2 \) (tunable), it is possible to calculate the number of \( M \) DMD modes in eq. (1), called as the spectral complexity, as follows

\[
a_{M+1}/a_1 \leq \varepsilon_2.
\]

When the database analysed presents the transient state of a numerical simulation, an additional step is carried out to identify the leading modes driving the flow dynamics. This step consists on calculating the energy index \( I_m \), defined as

\[
I_m = K \sum_{k=1}^{K} |a_m e^{(\delta_m + i\omega_m)\Delta t}| ||u_m||^2_F \times \Delta t,
\]

where \( || \cdot ||^2_F \) is the square of the Frobenius norm. This index weights the modes taking into account also the damping or growth of the modes in time, which is reflected in \( \delta_m \). More details about this criterion, called as \( HODMD \) with criterion, are presented in [42].

### 2.3. Principal Component Analysis and Variable Selection

Principal Component Analysis (PCA) [7] reduces the dimensionality of an input matrix \( X \), consisting of \((K \times N_x \times N_y)\) rows representative for the statistical observations of a system (e.g., \( N_x \times N_y \) grid points for \( K \) snapshots of a CFD simulation) and \( N_{\text{variables}} \) columns for the associated observable variables, by projecting such matrix onto a lower-dimensional basis consisting of \( q \) eigenvectors.

The PCA algorithm is associated to the covariance matrix \( C \) of the input dataset \( X \), which is defined as follows

\[
C = \frac{1}{(K \times N_x \times N_y) - 1} X^T X.
\]

This covariance matrix can be approximated by solving an eigenvalue problem as

\[
C = ALA^T,
\]

where the matrices \( A \) and \( L \) contain the associated eigenvectors (in columns) and eigenvalues (in the diagonal), respectively. The columns of \( A \) are the Principal Components (PCs) and they represent geometrically the orthogonal coordinates of the lower-dimensional manifold, obtained as a linear combination of the original variables [8]. This matrix also represent the proper orthogonal decomposition (POD) modes, calculated by an SVD of the original dataset \( X \) (see details in [19]). The matrix \( L \) is instead a diagonal matrix, containing the eigenvalues \( \lambda_i \) (with \( i \in [1,N_{\text{variables}}] \)) which are found in descending order of magnitude. This last condition on the eigenvalues’ order has important effects from a mathematical point of view, and it represents a key point for the dimensionality reduction. The magnitude of each eigenvalue \( \lambda_i \), in fact, represents the amount of input data variance being explained by the associated eigenvector. Consequently, if the value of \( \lambda_i \) is large, it entails the associated PC having an important role in the lower-dimensional representation of the input data, as it is representative for a large amount of information. On the other hand, if \( \lambda_i \) is small, it means that the contribution of the associated PC to the orthogonal space representation is negligible, and it can eventually be considered as noise, spatial redundancies (or in turbulent flows, it can be related to small flow scales [34]). In light of what stated above, it follows that the last PCs can be removed so that the \( p \)-dimensional input matrix can be expressed only by \( q \) orthogonal components (with \( q < N_{\text{variables}} \)) associated to the highest eigenvalues, thus considering a truncated eigenvector matrix \( A_q \). The optimal dimensionality for the PCA manifold can be set by...
assessing the amount of input data variance \( t_q \) being explained by the truncated \( q \)-dimensional basis \([10]\), defined as the ratio between the cumulative sum of the first \( q \) eigenvalues and the sum of all eigenvalues

\[
t_q = \frac{\sum_{i=1}^{q} \lambda_i}{\sum_{j=1}^{N_{\text{variables}}} \lambda_j},
\]

and ensuring that the condition \( t_q \approx 1 \) is observed. The dimensionality reduction is finally accomplished by projecting the input data on the truncated basis

\[
Z_q = X A_q.
\]

After the projection, the original input can be then reconstructed leveraging the orthonormality of the eigenvectors matrix as

\[
X_r \approx Z_q A_q^T.
\]

The PCA algorithm shown above can be used not only to perform dimensionality reduction via feature extraction (i.e., to find a new, reduced-order, representation of the original space with a manifold whose components are obtained as a linear combination of the original ones), but also to perform feature selection (i.e., to reduce the dimensionality by selecting a subset of the original variables). In fact, PCA can effectively identify correlations among the variables defining the state space and can subsequently select the best reduced subset from the original group of input variables \([14, 15, 17]\). In particular, according to Joliffe \([17]\), it is possible to discard by means of a backward elimination algorithm (the B2 algorithm) the variables that are more aligned with the last PCs (i.e., the ones having the highest loadings on the last eigenvectors) for a higher correlation with the components whose associated eigenvalues are small.

Nevertheless, if high-dimensional problems are considered, the weights’ distribution on the PCs can be difficult to analyse or interpret, as it is not always possible to identify a clearly dominant weight on one or more eigenvectors. Thus, PCA can be coupled with a rotation technique to increase the eigenvectors’ physical interpretability, for the rotated PCs to be preferentially aligned with one variable, as shown in Fig. 1.

![Figure 1: Schematic of the different steps for a PCA reduction process when coupled with an orthogonal rotation of the eigenvectors.](image)

Different techniques are available in literature to accomplish this operation, and they can be grouped into two classes: orthogonal and oblique rotations. The main difference between the latter is that when the first class of rotations is employed the eigenvectors keep their reciprocal orthogonality (and, consequently, the amount of variance being explained by the reduced basis does not change), while this condition is lost with oblique rotations. As shown by Bellemans \textit{et al.} \([3]\), the Varimax rotation introduced by Kaiser \([35]\) can be considered the most accurate for application to reacting flows data. In particular, this technique belongs to the class of orthogonal rotations and its objective function is the maximization of the sum of variances of the squared coefficients within each eigenvector, i.e., the maximization of the following quantity

\[
\begin{align*}
V & = N_{\text{variables}} \sum_{i=1}^{N_{\text{variables}}} \left( b_{ij}^2 \right)^2 - \gamma \sum_{i=1}^{N_{\text{variables}}} \left( b_{ij}^2 \right)^2 \\
& = \frac{N_{\text{variables}}}{N_{\text{variables}}} \sum_{i=1}^{N_{\text{variables}}} \left( b_{ij}^2 \right)^2 - \gamma \sum_{i=1}^{N_{\text{variables}}} \left( b_{ij}^2 \right)^2,
\end{align*}
\]

where \( b \) is the PC loading, \( \gamma = 1, i \in [1, N_{\text{variables}}], \) and \( j = [1, q] \) with \( q \) being the dimensionality of the PCA manifold \([3]\). When a non-unitary \( \gamma \) is used, different rotation techniques are obtained (e.g., quartimax rotation when \( \gamma = 0 \).
When Varimax rotation is applied, a new representation of the eigenvectors \( \mathbf{A}_q \) is obtained that can be expressed in terms of original PCs when a rotation matrix \( \mathbf{T} \) is defined as in Ref. [3]

\[
\mathbf{A}_q' = \mathbf{T} \mathbf{A}_q,
\]  

(27)

which thus entail the possibility to also retrieve a rotated lower-dimensional representation of the input data as

\[
\mathbf{Z}_q' = \mathbf{X} \mathbf{T} \mathbf{A}_q.
\]  

(28)

In the current work, a rotated basis as in eq. (27) is utilized to perform feature selection via B2 algorithm (hereinafter referred to as B2r, when rotated PCs are taken into account). The impact of the rotation on feature selection will be then compared to the case when B2 is applied to the standard PCA (that is, when an unrotated truncated basis \( \mathbf{A}_q \) is used to discard the variables) in the following sections.

3. Numerical database and PCA

This section briefly introduces the numerical simulations carried out to generate the database of a reactive flow that will be analyzed using the algorithm introduced in Section 2.2. In particular, the considered numerical simulation is representative for the oxidation of a methane jet (65% methane, 35% nitrogen, on a molar basis) in air. The oxidizer is injected into the domain adopting a constant velocity of 35 cm/s, while the fuel velocity varies in space and time according to the following sinusoidal perturbation:

\[
v(r, t) = v_{\text{max}} \left( 1 - \frac{r^2}{R^2} \right) \left[ 1 + A \sin(2\pi f t) \right],
\]  

(29)

with \( R \) representing the internal radius of the nozzle, \( r \) is the radial coordinate, \( t \) is the time, \( v_{\text{max}} \) is the maximum velocity \( (v_{\text{max}} = 70 \text{ cm/s}) \) and \( f \) and \( A \) are the frequency \( (f = 10 \text{ Hz}) \) and the amplitude \( (A = 0.25) \) of the perturbation, respectively. With regard to the kinetic mechanism, the detailed mechanism POLIMI\textunderscore C1\textunderscore C3\textunderscore HT\textunderscore 1412 [43] (accounting for 82 species) was used for the numerical simulations, which were run with LaminarSMOKE code, an OpenFOAM-based operator-splitting solver proposed by Cuoci et al. [44] to simulate laminar reacting flows by means of detailed kinetic mechanism. Additional information regarding this numerical simulations and the numerical setup can be found in Refs. [11, 12].

The dataset representative for the numerical simulation considered for HODMD is a fourth-order tensor formed by 83 variables (the temperature and 82 different species), two spatial dimensions composed by a 150 × 150 grid points (along the streamwise and normal direction, respectively); and 117 snapshots equidistant in time with \( \Delta t = 10^{-3} \). Hence, in tensor form, the dimension of the original database is 83 × 150 × 150 × 117. The order of magnitude of the variables ranges from \( 10^2 \) (temperature) to \( 10^{-11} \) (RALD3G, the smallest radical computed). This means that there are 13 orders of magnitude difference, highlighting the importance of scaling the variables. The variables have been scaled using two different scaling methods, auto scaling and range scaling. In Fig. 2, a representative snapshot is plotted in order to visualise the aforementioned database.

For the PCA-based feature selection, the training matrix \( \mathbf{X} \) consists of 22 snapshots, also taking into account the 83-dimensional thermochemical space. As in the HODMD analysis, prior to the application of PCA for variable selection, the columns of \( \mathbf{X} \) were centered with their mean and scaled with their standard deviation (auto scaling), respectively, because of the different units and ranges of the variables being taken into account. Subsequently, the B2 backward elimination algorithm was applied considering a final number of selected variables ranging in the interval \([2, 81]\) (i.e., it was applied imposing each time a different number of variables to retain) and, in particular, first projecting on the standard PCA basis \( \mathbf{A}_q \), eq. (24), (i.e., applying the standard B2 algorithm), and then rotating the eigenvectors via Varimax (i.e., considering \( \mathbf{A}_q' \), eq. (27), and the B2r selection algorithm).

4. HODMD modelling reactive flows

The application of HODMD algorithm to develop a reduced order model (ROM) involves a three-step algorithm: the data are pre-processed using different scaling methods at step 1, then, a dimensionality reduction is carried out.
using HOSVD at step 2, finally the main dynamics associated to the patterns describing the flow are identified at step 3 using HODMD, and a ROM is developed using this information. Two different cases (obtained in step 2) are compared for each type of preprocessing technique used in step 1: (i) using all the components (retaining all the SVD component modes) and (ii) reducing the number of components as explained in eq. (7) in step 2 (only retaining some SVD component modes according to the tunable tolerance $\epsilon$). It is important to mention that, similarly to PCA, each one of the SVD component modes is related to the number of variables in the database. Therefore, an SVD component mode is composed by the linear combination of the variables studied. The differences found among the various SVD component modes lies in the values of the coefficients forming this linear combination of variables. In what follows, it is important not to confuse the number of components (linear combination of variables), which can be reduced using HOSVD (step 2), with the number of variables, which can be reduced using PCA (as explained in Section 2.3).

4.1. HOSVD and truncated variables

The dimensionality reduction carried out using HOSVD to develop a ROM is studied for two different values of the tolerance $\epsilon$ applied as in eq. (7). As explained before, this tolerance controls the reduction in the spatial and temporal dimensions of the database (SVD spatial and temporal modes), and will also reduce the number of components analysed (SVD component modes) based on their relevant contribution to the dynamical system. Two tolerances are selected, to compare coarse with fine results, these are: $\epsilon = 10^{-2}$ and $10^{-3}$. Figure 3 compares the number of component retained using the different tolerances using auto scaling and range scaling methods. The value of these tolerances is selected based on the magnitude of the singular values (SVD components). It can be observed that using larger tolerances, the number of components retained when using range scaling is too small (smaller than 5) to provide accurate results. The smaller tolerance preserves all variables but one (connected to noise) when using auto scaling. This indicates that lowering this limit will not bring useful information.

Table 1 shows the RRMSE eq. (4) calculated in the reconstruction of the database as in eq. (6): as function of the number of components (SVD component modes) and SVD spatial and temporal modes retained. The HOSVD algorithm has been applied to analyse the dataset with the variables scaled using the range and auto scaling methods, respectively. The compression factor, $CF$, as in eq. (8) for the previous dimensionality reduction is presented in Tab. 2.
Figure 3: Singular values corresponding to the SVD component modes calculated for the two different scaling methods analysed. The tolerances $\varepsilon$ set are highlighted as horizontal black lines.

Table 1: RRMSE eq. (4) (in percentage) comparing the original database to the reconstruction of the reduced tensor, retaining $P_1$, $P_2$, and $N$ spatial (in $x$ and $y$) and temporal SVD modes, respectively. The case All considers all the components ($P_1 = 83$), while the case Red considers the $P_1$ components (SVD component modes) presented in the table. The database is pre-processed as in eq. (5) using the range and auto scaling methods (denoted as Range and Auto, respectively).

Table 2: Compression factor (CF) eq. (8) of the original tensor after applying HOSVD in the databases presented in Tab. 1.

The RRMSE increases with the tolerance level, as expected. With higher tolerances, the reduction carried out by the HOSVD algorithm is larger, resulting in a larger $CF$. As expected, the RRMSE is larger when the number of components are reduced (Red), as well as the $CF$ increases in this case, especially using the case with the coarse tolerance and the range scaling method. In this case, the $CF$ increases by one order of magnitude while the RRMSE is still maintained below 3% (increasing only 0.3% compared to the case with all the components (All)).

Comparing the two variable scaling methodologies, the reconstruction of the tensor is slightly better performed when using auto scaling method compared to range scaling, because the former method provides a uniform weight to the variables. However, the compression factor is also much larger when the variables are scaled with range scaling method than with auto scaling. In other words, the number of SVD modes retained for a similar tolerance is larger with auto scaling, thus the compression factor is smaller. Similar results were identified when performing a PCA analysis, as described in Ref. [5].

The main goals of the proposed model is to provide the best reconstruction of the original dataset using the smallest amount of information possible. Based on this idea, the range scaling appears particularly promising because the order of magnitude or the RRMSE is one or two orders larger in the former case, while the rise in the RRMSE error is only reflected in the first decimal digit. More specifically, comparing the results obtained in both cases, the RRMSE when using range scaling is $\sim 0.3 - 0.4\%$ larger than when using auto scaling in all the cases (the RRMSE is always maintained smaller than 3% and 1% for the coarse and fine tolerances, respectively), while the $CF$ is $\sim 20 - 892\%$ larger when using range than auto scaling, as presented in Tab. 3. This result suggests that reducing the number of
components and using the range scaling method, it is possible to obtain an accurate ROM (with RRMSE smaller than 3% and 1% for the coarse and fine tolerances, respectively) with a high dimensionality reduction.

| DIFF | All | Red |
|------|-----|-----|
| $\epsilon = 10^{-2}$ | 187% | 892% |
| $\epsilon = 10^{-3}$ | 20% | 57% |

Table 3: Comparison of the CF obtained in the range and auto scaling methods as presented in Tab. 2. The rise in magnitude is calculated as $DIFF = (XX_{range} - XX_{auto})/XX_{auto} \times 100$, with $XX_{range}$ and $XX_{auto}$ as the CF calculated in the range or auto scaling methods, respectively.

Finally, to compare the benefit of each pre-processing method in the reconstruction of the flow (quantified with the RRMSE), when using range and auto scaling methods, it is necessary to adjust the tolerance to have the same compression factor in both cases. Hence, we select $\epsilon = 10^{-3}$ and $\epsilon = 7 \times 10^{-2}$ for range and auto scaling, respectively, to build a reduced-dimensionality database, which is considered as a ROM. The selection of these tolerances gives an RRMSE of 13.11% and 0.34% for auto and range scaling methods respectively, and the corresponding associated compression factor is 377 and 290 for the case considering all the components (All). The reason of this choice is based on maintaining the largest number of components for the range scaling method (fine tolerance), increasing the accuracy of the ROM. Hence, we adapt the number of components selected in the auto scaling method, as function of the previous choice. This reduced database will be used (as part of the HODMD algorithm) for the spectral analysis presented in the following section.

4.2. HODMD for the spectral analysis

HODMD is applied to identify the main patterns and associated frequencies driving the flow dynamics. To calibrate the algorithm, several values of $d$ have been used, with the aim at ensuring the robustness of the method, these are $d = 10, 12$ and 15, which are values close to 11, a value proportional to the number of snapshots that is calculated following the calibration tips presented in Ref. [37] as $K/10 = 117/10 \approx 11$. The tolerances used for developing the ROM have been selected in the previous section (to maintain a similar compression factor in both the range and auto scaling methods): $\epsilon = 10^{-3}$ for range scaling and $\epsilon = 7 \times 10^{-2}$ for auto scaling. Figure 4 shows the frequencies as a function of the amplitudes of the DMD modes identified with the different tolerances. The top figure shows the analysis for range scaling, and the bottom figure, for auto scaling. The spectrum calculated is similar in the analysis carried out in the two databases. The spectrum shows a clear periodic solution with a principal mode, the one with the biggest amplitude and frequency $\omega_p$, and the harmonics, modes with lower amplitude and frequency $\omega_n = n \times \omega_p$ for $n \in \mathbb{N} > 1$. The principal mode using range scaling is $\omega_p^{RANGE} \approx 60$, and the one using auto scaling is $\omega_p^{AUTO} \approx 50$. The reason for this difference lies in the different variables selected as dominant in each one of the scaling methods, which is explained in detail below. A similar spectrum (only one is shown for the sake of brevity) has been obtained for the case retaining all the components (All) and retaining a reduced number of components (Red), as previously presented in Tab. 1. This is because, as explained in section 2.2, the DMD-d algorithm is applied to the temporal matrix, which is not affected by the reduction of the components.

The RRMSE error eq. (4) is calculated for the reconstruction of the tensor using the DMD expansion (1) with the DMD modes presented in Fig. 4 (plus their complex conjugate), representing a purely periodic solution. The parameter $d = 12$ has been set as the reference to illustrate this behaviour (although similar results are obtained for $d = 10$ and 15). Table 4 presents this error for the cases All and Red (retaining all and a reduced number of components, respectively). The differences found in this error when HODMD is applied over the database containing all the components and the one with a reduced number of components is small, in good agreement with the results obtained in the previous section. This suggests that the influence of reducing the number of components on the flow dynamics is small. On the other hand, the error is much smaller when using range scaling (error smaller than 1%) than auto scaling (error $\sim 24$%), suggesting that the periodic behaviour identified by HODMD is driven by the main variables (identified in range scaling). This fact indicates that using HODMD to identify the flow dynamics in a database formed by a large number of variables with similar weight (as when using auto scaling) could mask the value of the leading flow frequency, or could alter the proper reconstruction of the original flow (due to the presence of some noise or some error assumed in the frequency calculations). This is in good agreement with the two different values of
leading frequencies identified in the analyses carried out in the database scaled with range and auto scaling methods \(\omega_{\text{AUTO}} \approx 50\) and \(\omega_{\text{RANGE}} \approx 60\).

Finally, Fig. 5 shows the reconstruction error for the 10 variables with minimum (top) and maximum (bottom) RRMSE identified with range scaling, and the results are compared with the RRMSE obtained for the same variables using auto scaling. It can be observed how range scaling reconstructs with smallest error the main variables, while the largest error is obtained in the reconstruction of the radicals. On the contrary, since auto scaling provides a uniform reconstruction (see details in Ref. [5]), the error for the main variables is larger than when using range scaling, while the inverse holds for radicals.

### 4.3. Robustness of the modelling tools

This section shows the robustness of the ROM previously developed, comparing the RRMSE in the reconstruction as function of number of variables analysed and the number of components. To select the number of variables, PCA is applied over the original database using two methods, B2 and B2 with rotation. The number of variables are selected progressively one by one, resulting in new databases. HODMD is then applied progressively over each one of these databases, calculating the RRMSE for the reconstruction of the database. Moreover, range and auto scaling methods are used in the first step of the HODMD analysis for each one of the new databases. Figure 6 shows the variation of the RRMSE as function of the number of variables (from PCA) and the number of components (from HODMD) retained. The analysis is carried out in the two databases presented before, the database with all the components (All), and the reduced database (Red). The variable selection method in PCA does not affect significantly the result, although for a small number of selected variables, B2 method decreases the RRMSE. However, the type of scaling method strongly affects the results. The RRMSE for the reconstruction of the flow fluctuates as function of the number of variables in both cases. The error varies from 20 – 65% when using auto scaling and from 0.6 – 1.5% for the range scaling. Additionally, the reconstruction error stabilises using a smaller number of variables with range scaling, suggesting once more the suitability of this scaling strategy to capture the main flow features. Regarding the number of components, which are marked in the points highlighted in grey in the figure, the results show that when increasing the number of variables, reducing the number of components does not increases the RRMSE to a large
extent. Hence, when the number of variables is increased, reducing the components may be a good strategy to reduce the data dimensionality of the model while maintaining a good reconstruction.

Finally, the main frequency calculated by HODMD as function of the number of variables is plotted in Fig. 7. Since the data analysed represent the transient regime of a numerical simulation (which also increases the difficulty of the analysis presented in this work), eq. (20) is used to set the leading modes in each case. The variable selection cases B2 and B2 with rotation are studied in the databases scaled using auto scaling and range scaling methods. Depending on the number of variables, the frequency changes between \(~40\), \(~60\) and \(~90\) when the database is scaled using range method in both selection methods. Nevertheless, this frequency value is stable for auto scaling, with \(\omega \sim 50\). This suggests once more that there is more than one dynamics driving the flow motion (actually, the method identifies three main dynamics) and the range scaling can prioritize them properly. When there is a small number of variables selected, the value of the main frequency oscillates for range scaling. This phenomenon stabilises when the number of variables is increased, hence the method identifies an averaged value for the frequency of the main variables selected. These results suggest once more, that the main dynamics of the main variables and radicals are driven by different frequencies, at least in the transient regime of the numerical database analysed.
5. Conclusions

This article introduces a new application of the multi-dimensional HODMD algorithm, adapted to develop ROMs in reactive flows. This application is based on a new methodology, fully data-driven, consisting on three main steps. The first step focuses on pre-processing the database analysed, using techniques (generally used in machine learning), which consist on centering and scaling each variable. On the second step, HOSVD is applied to reduce the data dimensionality of the database analysed. With this algorithm, the computational cost (time and memory requirements) of the next step thoroughly decreases. Finally, the third step uses the classical HODMD algorithm.

The effect of reducing the number of components is studied in detail measuring the accuracy of the ROM via the relative RMS error (RRMSE) from the reconstruction of the original database. Two different tolerances have been set (coarse and fine), to study two different cases: maintaining all the components in the model or reducing this number of components via SVD. In both cases, the reconstruction error is slightly smaller when using the auto scaling method, although the compression factor is much larger when using range scaling. The reason of this performance is that auto scaling method, gives all the variables the same importance (i.e., temperature and pressure will be considered as important as the radicals), retaining a bigger number of components for a similar tolerance. On the contrary, range scaling identifies the main variables of the flow, hence providing more accurate flow reconstructions using a smaller number of components. Additionally, reducing the number of components composing the model strongly increases the compression factor, while the RRMSE is slightly increased compared to the case with all the variables. This result suggests that using range scaling method and reducing the number of components can be considered as a good strategy to develop a ROM.

HODMD is applied to the previous reduced-dimensionality databases, finding that the main dynamics driving the flow motion is periodic. The main frequency identified using the range and scaling methods is different. The RRMSE for the reconstruction is smaller than 1% in the case scaled with range, while this error is \( \sim 24\% \) when using the auto scaling method. The variations in the reconstruction error is small for the cases reducing and maintaining all the components. The differences found between these errors suggests that using the range scaling, the dynamics driving the main variables are well captured by HODMD.

The robustness of the method to identify the main flow dynamics, has been tested by applying HODMD to different number of variables. These variables have been selected using PCA with two variable selection methods, B2 and B2 with rotation. The RRMSE of the reconstructed field fluctuates as function of the number of variables as \( \sim 0.6 \sim 1.5 \) and \( \sim 20 \sim 65 \) for the range and auto scaling methods, respectively. Moreover, this reconstruction error stabilises when using a smaller number of variables in the range scaling method. It is also found that reducing the number of components, does not increase the RRMSE of the reconstruction to a large extent when the error is stabilized. These results suggest once more that using the range scaling method and reducing the number of components is the best option to identify the main flow dynamics and to develop a ROM, with a high dimensionality reduction (compression factor).

Finally, the coupling of the HODMD and the variable selection using PCA suggests that a set of variables are adequate to model the main dynamics of the whole system. This has an impact on Feature selection and on the reduction of the computational cost associated with very massive data.
Acknowledgments

AC and SLC acknowledge the grant PID2020-114173RB-I00 funded by MCIN/AEI/10.13039/501100011033. AC also acknowledges the support of Universidad Politécnica de Madrid, under the programme ‘Programa Propio’. GDA acknowledges the support of the Fonds National de la Recherche Scientifique (FRS-FNRS) through a FRIA fellowship. AP acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 714605.

References

[1] Stewart Cant. High-performance computing in computational fluid dynamics: progress and challenges. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 360(1795):1211–1225, 2002.
[2] Alessandro Parente, JC Sutherland, Bassam B Dally, Leonardo Tognotti, and PJ Smith. Investigation of the mild combustion regime via principal component analysis. Proceedings of the Combustion Institute, 33(2):3333–3341, 2011.
[3] Aurélie Bellemans, Gianmarco Aversano, Axel Coussenemt, and Alessandro Parente. Feature extraction and reduced-order modelling of nitrogen plasma models using principal component analysis. Computers & chemical engineering, 115:504–514, 2018.
[4] Axel Coussenemt, Olivier Gicquel, and Alessandro Parente. Mg-local-pca method for reduced order combustion modeling. Proceedings of the Combustion Institute, 34(1):1117–1123, 2013.
[5] Alessandro Parente and James C Sutherland. Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. Combustion and flame, 160(2):340–350, 2013.
[6] Benjamin J Isaac, Axel Coussenemt, Olivier Gicquel, Philip J Smith, and Alessandro Parente. Reduced-order pca models for chemical reacting flows. Combustion and flame, 163(11):2785–2800, 2014.
[7] Alessandro Parente, James C Sutherland, Leonardo Tognotti, and Philip J Smith. Identification of low-dimensional manifolds in turbulent flames. Proceedings of the Combustion Institute, 32(1):1579–1586, 2009.
[8] Christopher M Bishop. Pattern recognition. Machine learning, 128(9), 2006.
[9] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065):20150202, 2016.
[10] Giuseppe D’Alessio, Alberto Cuoci, and Alessandro Parente. Feature extraction and artificial neural networks for the on-the-fly classification of high-dimensional thermochemical spaces in adaptive-chemistry simulations. Data-Centric Engineering, 2, 2021.
[11] Giuseppe D’Alessio, Alessandro Parente, Alessandro Stagni, and Alberto Cuoci. Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. Combustion and Flame, 211:68–82, 2020.
[12] Giuseppe D’Alessio, Alberto Cuoci, Gianmarco Aversano, Mauro Bracconi, Alessandro Stagni, and Alessandro Parente. Impact of the partitioning method on multidimensional adaptive-chemistry simulations. Energies, 13(10):2567, 2020.
[13] Giuseppe D’Alessio, Antonio Attili, Alberto Cuoci, Heinz Pitsch, and Alessandro Parente. Analysis of turbulent reacting jets via principal component analysis. In Data Analysis for Direct Numerical Simulations of Turbulent Combustion, pages 233–251. Springer, 2020.
[14] Ian T Jolliffe. Discarding variables in a principal component analysis. i: Artificial data. Journal of the Royal Statistical Society: Series C (Applied Statistics), 21(2):160–173, 1972.
[15] Ian T Jolliffe. Discarding variables in a principal component analysis. ii: Real data. Journal of the Royal Statistical Society: Series C (Applied Statistics), 22(1):21–31, 1973.
[16] Wojtek J Krzanowski. Selection of variables to preserve multivariate data structure, using principal components. Journal of the Royal Statistical Society: Series C (Applied Statistics), 36(1):22–33, 1987.
[17] Ian T Jolliffe. Choosing a subset of principal components or variables. Principal component analysis, pages 111–149, 2002.
[18] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. Journal of fluid mechanics, 656:5–28, 2010.
[19] Soledad Le Clainche. An introduction to some methods for soft computing in fluid dynamics. Advances in Intelligent Systems and Computing, 950:557–566, 2019.
[20] Franck Richcoeur, Layal Hakim, Antoine Renaud, and Laurent Zimmer. DMD algorithms for experimental data processing in combustion. In Proceedings of the 2012 Summer Program, pages 459–468. Center for Turbulence Research, Stanford University, December 2012.
[21] Souvik Chatterjee, Achintya Mukhopadhyay, and Swarnendu Sen. Stability study of laminar flame using proper orthogonal decomposition and dynamic mode decomposition. In n3l-Int’l Summer School and Workshop on Non-Normal and Nonlinear Effects In Aero-and Thermodynamics, page 13, 2013.
[22] John M. Quinlan and Ben T. Zinn. Transverse Combustion Instabilities: Modern Experimental Techniques and Analysis. 2014.
[23] Cheng Huang, William E. Anderson, Matthew E. Harvazinski, and Venkateswaran Sankaran. Analysis of self-excited combustion instabilities in a body-stabilized turbulent premixed flame. Combustion and Flame, 162:4075–4083, 09 2015.
[24] Abdulla Ghani, Thierry Poinset, Gicquel L.Y., and Gabriel Staffelbach. Les of longitudinal and transverse self-excited combustion instabilities in a bluff-body stabilized turbulent premixed flame. Combustion and Flame, 211:68–82, 2020.
[28] T Grenga and ME Mueller. Dynamic mode decomposition: A tool to extract structures hidden in massive datasets. In Data Analysis for Direct Numerical Simulations of Turbulent Combustion, pages 157–176. Springer, 2020.

[29] Soledad Le Clainche and José M Vega. Analyzing nonlinear dynamics via data-driven dynamic mode decomposition-like methods. Complexity, 2018, 2018.

[30] Soledad Le Clainche and José M Vega. Higher order dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems, 16(2):882–925, 2017.

[31] Adrián Corrochano, Donatella Xavier, Philipp Schlatter, Ricardo Vinuesa, and Soledad Le Clainche. Flow structures on a planar food and drug administration (fda) nozzle at low and intermediate reynolds number. Fluids, 6(1):4, 2021.

[32] Soledad Le Clainche Martínez, Francisco Sastre, José M. Vega, and Velázquez Angel. Higher order dynamic mode decomposition applied to post-process a limited amount of noisy PIV data.

[33] Soledad Le Clainche, Rubén Moreno-Ramos, Paul Taylor, and José M Vega. New robust method to study flight flutter testing. Journal of Aircraft, 56(1):336–343, 2019.

[34] S Le Clainche, Daulet Izbassarov, M Rosti, Luca Brandt, and Outi Tammisola. Coherent structures in the turbulent channel flow of an elastoviscoplastic fluid. Journal of Fluid Mechanics, 888, 2020.

[35] Henry F Kaiser. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23(3):187–200, 1958.

[36] Soledad Le Clainche, José M Vega, and Julio Soria. Higher order dynamic mode decomposition of noisy experimental data: The flow structure of a zero-net-mass-flux jet. Experimental Thermal and Fluid Science, 88:336–353, 2017.

[37] Jose Manuel Vega and Soledad Le Clainche. Higher order dynamic mode decomposition and its applications. Academic Press, 2020.

[38] Soledad Le Clainche and Esteban Ferrer. A reduced order model to predict transient flows around straight bladed vertical axis wind turbines. Energies, 11(3):566, 2018.

[39] Soledad Le Clainche and José M Vega. Higher order dynamic mode decomposition to identify and extrapolate flow patterns. Physics of Fluids, 29(8):084102, 2017.

[40] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. Psychometrika, 31(3):279–311, 1966.

[41] K K Chen, J H Tu, and C W Rowley. Variants of dynamic mode decomposition: Boundary condition, koopman, and fourier analyses. J. Nonlin. Sci., 22(6):887–915, 2012.

[42] Jiaqing Kou, Soledad Le Clainche, and Weiwei Zhang. A reduced-order model for compressible flows with buffeting condition using higher order dynamic mode decomposition with a mode selection criterion. Physics of Fluids, 30(1):016103, 2018.

[43] Eliseo Ranzi, Alessio Frassoldati, Roberto Grana, Alberto Cuoci, Tiziano Faravelli, Andrew P Kelley, and Chung K Law. Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels. Progress in Energy and Combustion Science, 38(4):468–501, 2012.

[44] Alberto Cuoci, Alessio Frassoldati, Tiziano Faravelli, and Eliseo Ranzi. Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method. Energy & Fuels, 27(12):7730–7753, 2013.