

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Aeronáutica y del Espacio



Innovative Data-Driven Methods for Clustering and Flow Control in Reacting Flows

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Adrián Corrochano Calcerrada
Máster en Matemática Industrial

Madrid, 2025



UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Aeronáutica y del
Espacio

Doctoral Degree in Aerospace Engineering

Innovative Data-Driven Methods for Clustering and Flow Control in Reacting Flows

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Adrián Corrochano Calcerrada

Máster en Matemática Industrial

Under the supervision of:

Dr. Soledad Le Clainche Martínez

Madrid, 2025

Title: Innovative Data-Driven Methods for Clustering and Flow Control in Reacting Flows

Author: Adrián Corrochano Calcerrada

Doctoral Programme: Aerospace Engineering

Thesis Supervision:

Dr. Soledad Le Clainche Martínez, Professor, Escuela Técnica Superior de Ingeniería Aeronáutica y del Espacio (Supervisor)

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been partially supported by Universidad Politécnica de Madrid under the program *Programa Propio*. A.C. acknowledges the grants *PID2020-114173RB-I00* and *TED2021-129774B-C21* funded by MCIN/AEI/10.13039/501100011033 for the support and funding of different conferences, visits to international research centers and open source publications.

*“Life before death,
strength before weakness,
journey before destination.”*

Brandon Sanderson – The way of kings

Acknowledgement

While the present Doctoral Thesis is fully written in English, some paragraphs will be written in different languages depending on the person for whom the message is intended.

My Doctoral Thesis has become one of my most challenging journeys, but also the most rewarding one. Thanks to the Ph.D., I have had the opportunity to meet a lot of wonderful people across different countries and continents.

Primero de todo me gustaría agradecer a mi directora de tesis, Prof. Soledad Le Clainche. Gracias por mostrarme el mundo de la investigación y tu visión de la ciencia desde hace más de seis años, cuando empecé contigo en el Trabajo de Fin de Grado. Gracias por las infinitas oportunidades que me has dado, estoy profundamente agradecido por los diferentes proyectos y colaboraciones en los que me has introducido. Gracias por haberme ayudado a convertirme en el científico que soy ahora, dándome ideas para continuar en el trabajo, pero también sabiendo escuchar y dando apoyo a mis propuestas. Gracias, Sole, por tu actitud ante mis problemas y agobios. Gracias por las reuniones "cortas" de 5 minutos que acababan siendo de horas porque hablábamos de cualquier cosa más allá del tema de la reunión. Pero lo más importante de todo, gracias por el trato personal y cercano que has tenido durante este tiempo.

I would like to thank Prof. Alessandro Parente for being an inspiration when hosting me during my doctoral stay. Thank you for the confidence you have shown in me in many different projects. Your trust and support have been invaluable. I would also like to acknowledge the support of Prof. Marco Rosti and Prof. Outi Tammisola for the invitation of researching in their labs and the inspiring scientific discussions throughout the Doctoral Thesis.

Vull donar les gràcies a totes aquelles persones que m'han acompanyat en aquest viatge. Especialment, m'agradaria agrair el suport a Eneko, Rubén i Charly, o Carlos segons el context. Contar-nos les miseries amb formatge pa i vi ens ha fet créixer junts. Hem viscut moments durs, però quan m'he fet pols, sempre m'ha fet millor estar amb vosaltres. Zerura begira, lagun artean goaz.

Gracias a Christian por darme la bienvenida y acogerme durante mi visita en Okinawa, así como por enseñarme la isla y nunca dejar que me aburriese. Gracias Belén por tu amistad y las risas compartidas desde que te conocí. Gracias a todos los amigos que he conocido gracias al balonmano, un deporte precioso que aprendí durante mi etapa universitaria y que tantas cosas me ha dado. Gracias a mis amigos del colegio por siempre saber escucharme.

Por supuesto, quiero agradecer a mi familia todo el apoyo que me ha dado desde pequeño, motivándome a siempre seguir creciendo y no rendirme nunca. Gracias a todos porque sin vosotros no estaría donde estoy ahora mismo. Mi más sincero y profundo agradecimiento por todo lo que habéis hecho.

Per últim, vull donar les gràcies a aquella persona que més m'ha aguantat i ajudat durant la redacció d'aquesta tesi doctoral. Jo mai mai no he pensat ...

Abstract

The present Doctoral Thesis develops new algorithms for developing Reduced Order Models (ROMs) in reacting flows. These novel algorithms pursue different objectives, which are the reduction of the dimensionality of big databases, the clustering of variables with similar dynamics and the control of the reacting flow to increase the efficiency and diminish the emission of pollutants.

These objectives are enclosed in the global goal for the European Union to become climate-neutral by 2050. With this purpose, the study of reacting flows has become very popular in the research community. However, modelling combustion is very challenging from both experimental and numerical insights. On the one hand, experiments are complicated due to the small time scales, the complex reaction driving their dynamics and the high cost of specialised materials, between other reasons. On the other hand, high-quality numerical simulations are also challenging. The numerous species involved, the turbulent-chemistry interaction and the wide range of magnitudes and time scales, make the numerical simulations highly expensive in terms of computational time and memory usage.

ROMs have raised to address this problem, as they can model the main flow dynamics at reduced computational cost. In the present Doctoral Thesis, the ROMs developed are based on Higher Order Dynamic Mode Decomposition (HODMD). This fully data-driven method, well-known in the fluid dynamics' community, decomposes a database into modes which move in time with a given frequency. These modes are patterns which describe the main flow dynamics. The first task is to adapt this algorithm to analyse reacting flows' databases. For this purpose, first, the database is preprocessed following the classical preprocessing techniques generally used in machine learning. The performance of the mentioned algorithm has been assessed with different quality indicators, measuring the error made on the calculations and the level of reduction of the dimensionality carried out. After that, the possibility to couple this algorithms with other typically used in combustion, as Principal Component Analysis (PCA) is also analysed with good results, demonstrating the robustness of the method for developing ROMs.

Once adapting the main algorithm used in the Doctoral Thesis for combustion databases, two different algorithms have been developed based on HODMD with two different and clear purposes. The first one, named hierarchical HODMD (h-HODMD), is a data-driven algorithm able to perform clustering of variables based on their dynamics. The algorithm iteratively performs HODMD and extracts a cluster of variables from the database, associating them a set of HODMD modes, frequencies and amplitudes. First, the algorithm has been used to reduce the reconstruction error, outperforming HODMD in the recreation of combustion databases. Secondly, the clusters have been analysed from the kinetics perspective, providing good results, as it made separated clusters for the nitrogen oxides and the oxy- and peroxy radicals, with different time scales for their dynamics.

The second algorithm developed aims to achieve flow control. With this purpose, the algorithm calculates the so-called non-linear structural sensitivity. This concept identifies the regions of the flow more prone to changes in the main dynamics. The algorithm was first developed for

laminar non-reactive flows, and its performance was then compared to the classical theory. Due to its potential, the algorithm was then applied to more complex cases. The algorithm has been found to be useful for non-Newtonian fluids and combustion. In this latter case, the control of the flow is translated into an increase in the combustion efficiency and the reduce of the nitrogen oxides emissions.

Resumen

La presente Tesis desarrolla nuevos algoritmos para la construcción de Modelos de Orden Reducido (MOR) en flujos reactivos. Estos algoritmos innovadores persiguen diversos objetivos, entre los que destacan la reducción de la dimensionalidad de grandes bases de datos, la agrupación de variables con dinámicas similares y el control de flujos reactivos para incrementar la eficiencia y reducir las emisiones de contaminantes.

Estos objetivos están alineados con la meta global de la Unión Europea de alcanzar la neutralidad climática para el año 2050. La modelización de la combustión plantea grandes desafíos tanto desde el punto de vista experimental como numérico. Por un lado, los experimentos son complejos debido a las pequeñas escalas de tiempo, las reacciones complejas que impulsan sus dinámicas y el alto coste de los materiales especializados, entre otras razones. Por otro lado, las simulaciones numéricas de alta calidad también presentan dificultades. La gran cantidad de especies involucradas, la interacción turbulencia-química y el amplio rango de magnitudes y escalas de tiempo hacen que estas simulaciones sean altamente costosas en términos de tiempo computacional y uso de memoria.

Los MORs han surgido como una solución a este problema, ya que permiten modelar las principales dinámicas a un coste computacional reducido. En la presente Tesis, los MORs desarrollados se basan en *Higher Order Dynamic Mode Decomposition* (HODMD). Este método, completamente basado en datos y ampliamente conocido en la comunidad de dinámica de fluidos, descompone una base de datos en modos que evolucionan en el tiempo con una frecuencia determinada. Estos modos son patrones que describen las principales dinámicas del flujo. La primera tarea consiste en adaptar este algoritmo para analizar bases de datos de flujos reactivos. Para ello, en primer lugar, la base de datos se preprocesa siguiendo técnicas clásicas de preprocesamiento. El rendimiento del algoritmo se ha evaluado mediante diversos indicadores de calidad, midiendo el error cometido en los cálculos y el nivel de reducción de la dimensionalidad alcanzado. Posteriormente, se ha analizado la posibilidad de acoplar este algoritmo con otros métodos comúnmente empleados en combustión, como el Análisis de Componentes Principales (PCA), obteniendo buenos resultados que demuestran la robustez del método para el desarrollo de ROMs.

Una vez adaptado el algoritmo principal utilizado en la Tesis Doctoral para bases de datos de combustión, se han desarrollado dos algoritmos diferentes basados en HODMD, cada uno con un propósito claro y definido. El primero, denominado *hierarchical HODMD*, es un algoritmo basado en datos capaz de realizar grupos de variables en función de su dinámica. El algoritmo aplica HODMD de manera iterativa, extrayendo un grupo de variables de la base de datos y asociando este grupo con un conjunto de modos, frecuencias y amplitudes. En primer lugar, el algoritmo se ha utilizado para reducir el error de reconstrucción, superando el rendimiento de HODMD. En segundo lugar, se han analizado los grupos desde una perspectiva cinética, obteniendo buenos resultados al separar, por ejemplo, los óxidos de nitrógeno de los radicales oxi- y peroxi, que presentan diferentes escalas de tiempo en sus dinámicas.

El segundo algoritmo desarrollado tiene como objetivo lograr el control del flujo. Con este propósito, el algoritmo calcula la llamada sensibilidad estructural no lineal. Este

concepto identifica las regiones del flujo más propensas a cambios en las dinámicas principales. El algoritmo se desarrolló inicialmente para flujos laminares no reactivos, comparando su rendimiento con la teoría clásica. Debido a su potencial, el algoritmo se aplicó posteriormente a casos más complejos. Se ha demostrado que el algoritmo es útil para fluidos no newtonianos y combustión. En este último caso, el control del flujo se traduce en un aumento de la eficiencia y en la reducción de las emisiones.

Table of Contents

Acknowledgement	v
Abstract	vi
Resumen	viii
List of Figures	xii
List of Tables	xiii
Abbreviations and acronyms	xvi
1 Introduction	1
1.1 Background in reacting flows	2
1.2 Governing equations	3
1.3 State of the art in Machine Learning applied to combustion	5
1.3.1 Principal Component Analysis	6
1.3.2 Proper Orthogonal Decomposition	7
1.3.3 Dynamic Mode Decomposition	8
1.3.4 Artificial Neural Networks	9
1.3.5 Flow Control Strategies	10
1.4 Objectives of the thesis	11
1.5 Achievements, Collaborations, and Publications	13
Peer-reviewed journal articles & book chapters	15
Conference presentations	17
1.6 Thesis Outline	18
2 Methodology	21
2.1 Preprocessing techniques	21
2.2 Higher Order Dynamic Mode Decomposition	22
2.2.1 Dimensionality reduction	23
2.2.2 The DMD-d algorithm	24
2.3 Multi dimensional HODMD	26
2.4 Hierarchical HODMD	28
2.5 HODMD for flow control	29
2.5.1 Linear stability	30
2.5.2 Structural sensitivity	31
2.5.3 Non-linear structural sensitivity	32

2.5.4	Use of data-driven methods for calculating the non-linear structural sensitivity	33
	Implementation of the algorithm	34
3	Higher order dynamic mode decomposition to model reacting flows	37
3.1	Preliminary results	37
3.1.1	Publication information	37
3.2	HODMD applied to a laminar flame	45
3.2.1	Publication information	45
4	Hierarchical higher-order dynamic mode decomposition for clustering and feature selection	59
4.1	Publication information	59
5	Structural sensitivity in laminar flames using data-driven methods	73
5.1	Structural sensitivity in non-linear flows using direct solutions	73
5.1.1	Publication information	73
5.2	Flow control with data driven approaches for laminar flames	84
5.2.1	Publication information	84
6	General Discussion	95
6.1	Adaptation of HODMD to model reacting flows	95
6.1.1	Using HODMD for the first time to model reacting flows	95
6.1.2	Thorough analysis of a flame using HODMD	96
6.2	New applications based on HODMD to model reacting flows	100
6.2.1	Hierarchical higher-order dynamic mode decomposition for clustering and feature selection	100
6.2.2	Structural sensitivity in laminar flames using data-driven methods . .	102
	Development of an algorithm to calculate the structural sensitivity in non-linear flows using direct solutions	103
	Passive flow control for laminar flames with data driven approaches .	107
6.3	Conclusions	110
6.3.1	Future work	113
	Referencias	115

List of Figures

1.1	A brief scheme of the combustion process of methane CH_4	2
1.2	Diagram summarising the main objectives of the present Doctoral Thesis. . .	12
2.1	On the left side, a sketch representing the snapshot matrix $\mathbf{V}_1^K$ (2.3). On the right, a sketch representing the sliding window process defined in Eq. (2.5). .	24
2.2	A visual representation of the transformation from the snapshot matrix to the snapshot tensor, in a database with two spatial dimensions and one temporal	27
2.3	A visual representation of the h-HODMD algorithm. Note that in each interaction, HODMD is applied to obtain the modes and the reconstruction error is calculated for each variable. If the reconstruction error is below a tolerance ε_v , the variable is selected to form a cluster.	29
6.1	Summary of the process of the HODMD analysis of a combustion database..	98
6.2	Summary of the cases analysed in Corrochano and Le Clainche, 2022. The geometry of the different cases is presented, as well as a representative snapshot and the Re conducted in the numerical simulation.	104
6.3	Summary of the process of passive flow control on a bidimensional EVP flow past a cylinder. a) Original unsteady simulation, represented with a representative snapshot and the plot of the streamwise velocity at a point $a = (x_a/D, y_a/D) = (9, 2)$ over time. b) HODMD analysis on the field, extracting the main frequencies and the modes associated to the dynamics of the simulation. The mode is normalised varying between -1 and 1 . c) Calculation of the non-linear structural sensitivity $\tilde{\mathbf{S}}^{NL}$ and zoom of the placement of little cylinders where $\tilde{\mathbf{S}}^{NL}$ is maximum. d) Controlled steady simulation, represented with a representative snapshot and the plot of the streamwise velocity at the same point a over time.	108
6.4	Diagram summarising the analysis of a combustion database, as proposed in the present Doctoral Thesis.	111

List of Tables

1.1	International collaborations during the Doctoral Thesis. The name of the main researcher appear with the topic of the collaboration and the publication published or conference presented (if any).	14
6.1	Frequency of the perturbation in Hertz, kinetic mechanism and the number of variables and reactions taken into account in each CFD simulation.	100
6.3	Efficiency of the flame and the production of NOx for the initial case and the three cases studied applying the flow control algorithm. For the three cases studied, in green if the efficiency increases or the NOx emissions decrease and in red, the opposite.	109

Abbreviations and acronyms

ANN	A rtificial N eural N etwork
CF	C ompression F actor
CFC	C loro F luo C arbons
CFD	C omputational F luid D ynamics
CNN	C onvolutional N eural N etwork
CO_2	C arbon D ioxide
CPU	C entral P rocessing U nit
DMD	D ynamic M ode D ecomposition
DNS	D irect N umerical S imulation
DRL	D ep R einforcement L earning
ERC	E uropean R esearch C ouncil
EU	E uropean U nion
EVP	E lasto V isco P lastic
FDA	F ood and D rug A dministration
FFT	F ourier F ast T ransform
FOM	F ull O rd M odel
H_2	M olecular H ydrogen
h-HODMD	h ierarchical- H igher O rd D ynamic M ode D ecomposition
HODMD	H igher O rd D ynamic M ode D ecomposition
HOSVD	H igher O rd S ingular V alue D ecomposition
IEA	I nternational E nergy A gency
IMFT	I nstitut de M écanique des F luides de T oulouse
KTH	K unliga T ekniska H ögskolan
LES	L arge E ddy S imulation

mdHODMD	m ulti d imensional H igher O rders D ynamic M ode D ecomposition
MILD	M oderate and I ntense L ow-oxygen D ilution
NO_x	N itrogen O xides
NS	N avier S tokes
ODE	O rdinary D ifferential E quation
ODS	O zone- D eplecting S ubstances
OpenFOAM	O pen S ource F ield O peration and M anipulation
OIST	O kinawan I nstitute of S cience and T echnology
PC	P rincipal C omponent
PCA	P rincipal C omponent A nalysis
PDE	P artial D ifferential E quation
PINN	P hysics- I nformed N eural N etwork
PhD	D octor of P hilosophy
POD	P roper O rthogonal D ecomposition
PSD	P ower S pectral D ensity
RANS	R eynolds- A veraged N avier S tokes
Re	R eynolds number
ROM	R educed O rders M odel
RRMSE	R elative R oot M ean S quare E rror
SPOD	S pectral P roper O rthogonal D ecomposition
St	S trouhal number
SVD	S ingular V alue D ecomposition
ULB	U niversit� L ibre de B ruelles
UPM	U niversidad P olit�cnica de M adrid
UPV	U niversitat P olit�cnica de V alencia
Vast	V ariable s tability
VQPCA	V ector Q uantization via P rincipal C omponent A nalysis

WMO World Meteorological Organization

Chapter 1

Introduction

During the research stay done in Brussels, Belgium, I had the opportunity to attend a lecture given by Professors Adel El Gammal and Romain Weikmans, who had just returned from the 27th Conference of the Parties (COP27), held in Sharm El Sheikh, Egypt. The lecture presented on the emergence of climate change (El Gammal et al., [2022](#)). One of the most striking conclusions was that it is necessary a disruption similar to a COVID-19 crisis, every year, to reach the carbon neutrality by 2050 (de Bruyn et al., [2020](#)). Despite the catastrophic sense of the statement, it is necessary to raise awareness on the difficulty of the challenge, but also give a sense of hope since there are episodes in history where these policies have worked.

The ozone hole and its recovery is an example of these public policies and the international cooperation. The ozone hole was primarily caused by human-made chemicals called chlorofluorocarbons (CFCs) and other ozone-depleting substances (ODS), widely used in common household aerosol sprays. Thanks to the scientific discoveries and to the social pressure exerted by those who voluntarily choose not to purchase this type of aerosol sprays, the different countries reached to agreements for banning the production of ozone-depleting chemicals, as CFCs. In this matter, the Vienna Convention, in 1985 (United Nations Environment Programme, [1985](#)) and the Montreal Protocol, in 1987 (United Nations Environment Programme, [1987](#)), stand out. The ban came into effect in 1989, and from that year the ozone levels stabilised in the 1990s and began to recover in the 2000s. Recent studies project that the ozone layer will fully regenerate by 2066 under the current regulations (World Meteorological Organization (WMO), [2022](#)).

The slow but steady recovery of the ozone layer demonstrates the effectiveness of international cooperation. In the context of climate change, the European Union (EU) aims for a 55% reduction in carbon dioxide (CO_2) emissions and the climate-neutrality by 2050 (European Commission and Directorate-General for Climate Action, [2019](#); European Commission, [2021](#)).

With the aim of reducing the emissions, a strong effort is currently carried out by the research community to study the combustion processes. Some of the lines pointing in this direction are the improvement of the efficiency in combustion systems (Z. Liu et al., [2020](#)), the reduction of pollutants produced from the combustion of fossil fuels (Gaba and Iordache, [2011](#)) and the creation of new eco-friendly alternatives to the current of fossil fuels. One of these new fuels

is hydrogen (H_2), whose production is expected to reach 20 Mt/year by 2030 (International Energy Agency (IEA), 2023).

The present Doctoral Thesis evaluates in this Chapter the modelling of combustion, reviewing the several approaches made to model reacting flows from both an experimental and a computational perspective to then examining the several machine learning (ML) techniques used for developing reduced order models (ROMs) in reacting flows.

1.1 Background in reacting flows

Before discussing the various experimental and numerical simulation approaches to model this problem, it is essential to briefly introduce the basic operating principles of a combustion process. A combustion process is a type of exothermic chemical reaction that leads to the rapid oxidation of combustible elements, mainly formed by carbon and hydrogen. Through this process, large amounts of thermal energy are released (Turns and Haworth, 2021). A 2D scheme of the combustion of methane is presented in Fig. 1.1. As seen in the figure, the combustion process is formed by three main different components:

- The fuel, which is methane in this case.
- The oxidizer, presented as oxygen, however most combustion systems use air as the oxidizer.
- The products resulted from the reaction of the fuel and the oxidizer. In Fig. 1.1, as it is supposed to be ideal, they are carbon dioxide (CO_2) and water (H_2O).

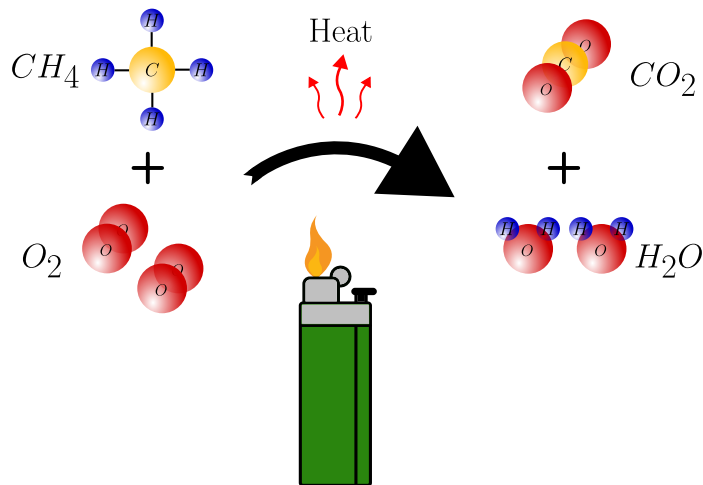


Figure 1.1: A brief scheme of the combustion process of methane CH_4 .

In the aforementioned scheme, the combustion is supposed to be ideal, as the fuel completely burns and produces exclusively water and CO_2 . Nevertheless, in real-life scenarios, a greater number of chemical species are formed due to combustion. Intermediate species are generated during the process, created and consumed within the combustion reaction, while additional products are formed by the completion of the process. These additional products include

unburnt fuel, hydrocarbons and nitrogen oxides, among others, which are generally expelled to the atmosphere.

The formation of the different chemical species and the combustion in general is function of multiple variables. Therefore, combustion is a field of knowledge in constant development and strong efforts are made by the research community to understand the dynamics of reacting flows.

Combustion can be modelled through experimental research. New fuels are being studied to reduce the emission of CO_2 to the atmosphere. In this context, molecular hydrogen (H_2) is studied under the base that ideally, the combustion produces only water, not CO_2 . Some experimental works with hydrogen include the research of Palomeque-Santiago et al., 2024, where they discovered different stable structures for ultra-lean hydrogen–air flames propagating in narrow gaps. Yahou et al., 2024, studied the flame propagation in lean hydrogen flames and compared them with methane flames and Gabana et al., 2025, analysed the combustion speed of an engine fuelled with gasoline and H_2 blends. All these studies have practical implications towards designing next-generation H_2 -combustion systems. However, there is still a need of more studies carried out with this fuel, as the transport and power generation sectors have generated valid concerns regarding the associated risks and potential threats (Salvi and Subramanian, 2015; Staffell et al., 2019).

The implementation of high-fidelity experimental measurements of combustion systems is complicated due to the small time scales, the complex reactions driving their dynamics, and also because of the high cost associated with the use of specialised and sophisticated infrastructures.

Similarly, high-precision measurements are available only for laboratory-scale systems. However, realistic combustion environments are typically very harsh, limiting the diagnostic options. Furthermore, even with the most advanced diagnostic techniques, only a limited number of variables of interest can be measured in industrial combustion systems, and these measurements are usually sparse in the spatial and temporal dimensions. Because of that, the computational approaches emerge to simulate the reality under specific conditions.

1.2 Governing equations

In this Section, the description of the main equations governing reacting flows is displayed. The solutions of the following equations will provide databases which will be used for developing ROMs. In what follows, the gradient of a vector variable ϕ (such as the velocity), will be defined as $\nabla\phi$. If the variable ϕ is a scalar, such as the pressure field, the gradient will be displayed as $\nabla\phi$. The divergence of the vector variable is denoted by $\nabla \cdot \phi$ and the material derivative is defined as $\frac{D\phi}{Dt} = \frac{\partial\phi}{\partial t} + \mathbf{v} \cdot \nabla\phi$, where $\mathbf{v}$ is the velocity of the mixture.

The governing equations leading the reacting flows dynamics are the following:

→ The conservation of mass:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (1.1)$$

where ρ is the density of the mixture.

→ The conservation of momentum:

$$\frac{D(\rho \mathbf{v})}{Dt} = -\nabla p + \nabla \cdot \boldsymbol{\tau} + \rho \mathbf{g}, \quad (1.2)$$

where p is the pressure, $\mathbf{g}$ is the gravity and $\boldsymbol{\tau}$ is the fluid stress tensor, calculated as

$$\boldsymbol{\tau} = -\mu \left(\nabla \mathbf{v} + (\nabla \mathbf{v})^\top \right) + \left(\frac{2}{3}\mu - \kappa \right) (\nabla \cdot \mathbf{v}) \mathbb{I}, \quad (1.3)$$

where μ is the viscosity of the mixture and κ is the mixture dilatational viscosity. The identity matrix is denoted by $\mathbb{I}$ and the superscript $()^\top$ refers to a matrix transpose.

→ The conservation of species:

$$\frac{D(\rho Y_k)}{Dt} = -\nabla \cdot (-\rho \mathbf{v}_k Y_k) + \dot{\Omega}_k, \quad k \in [1, N_{\text{species}} - 1], \quad (1.4)$$

where N_{species} is the number of species taken into account, Y_k is the mass fraction of a generic species k which is taken into account by the kinetic mechanism, $\mathbf{v}_k$ is the diffusion velocity corresponding to the species k and $\dot{\Omega}_k$ is its formation rate. Note that in Eq. (1.4), there are $N_{\text{species}} - 1$ equations. This is due to the fact that $\sum_{k=1}^{N_{\text{species}}} Y_k = 1$ and $\sum_{k=1}^{N_{\text{species}}} \dot{\Omega}_k = 0$. Therefore, the conservation of mass, Eq.(1.1), is the sum of Eq. (1.4) for all the N_{species} . The calculation of the mass fraction of the last chemical species can be computed from the constraint $\sum_{k=1}^{N_{\text{species}}} Y_k = 1$.

→ The conservation of energy:

$$\frac{D(\rho c_p T)}{Dt} = -\nabla \cdot \mathbf{q} - \rho \nabla T \sum_{k=1}^{N_{\text{species}}} c_{p,k} Y_k \mathbf{v}_k - \sum_{k=1}^{N_{\text{species}}} h_k \dot{\Omega}_k, \quad (1.5)$$

where T is the temperature of the mixture, c_p is the specific heat at constant pressure for the mixture, $\mathbf{q}$ is the heat flux vector, $c_{p,k}$ is the specific heat at constant pressure for the k -th chemical species and h_k is the individual-species enthalpy (Cuoci et al., 2013).

The set of Equations (1.1)–(1.5) need some closures via additional relations. The computation of the density of the mixture can be done via an equation of state. For an ideal gas, the closure equation is

$$p = \frac{\rho R_u T}{M}, \quad (1.6)$$

where R_u is the universal gas constant $\left(R_u \simeq 8,314 \left[\frac{J}{mol \cdot K}\right]\right)$ and $M = \left(\sum_{k=1}^{N_{\text{species}}} \frac{Y_k}{M_k}\right)^{-1}$ is the molar mass of the mixture, calculated using the mass fraction Y_k and the molar mass M_k of species k .

The heat flux $\mathbf{q}$ in Eq. (1.5) has to be modelled as it can include several means of heat transfer. One possibility is taking into account the conduction and the radiation as mechanisms of heat transfer, as

$$\mathbf{q} = \mathbf{q}_{diff} + \mathbf{q}_{rad} = -\lambda \nabla T - 4\sigma a_p (T^4 - T_{env}^4), \quad (1.7)$$

where λ is the thermal conductivity of the mixture, σ is the Stefan-Boltzmann constant, T_{env} is the environmental temperature, and a_p is the Planck mean absorption coefficient.

The model written above already has some simplifications, as it does not take into account turbulence and the interaction of it with the chemistry. Without taking this into account, what would complicate the calculations in numerical simulations, reacting flows simulations are challenging complex systems.

Several studies have been carried out simulating reacting flows with different purposes. To name a few, J. H. Chen et al., 2009, carried out terascale Direct Numerical Simulations (DNS) of turbulent jet flames, illustrating the chemistry-turbulence interactions and assessing the monitoring performance for science applications. Rieth et al., 2022 analysed DNS of fuel-lean turbulent premixed hydrogen-enriched flames for the understanding of the molecular diffusion of hydrogen species. Domínguez-González et al., 2023, two well-distinguished stable configurations of very lean hydrogen premixed flames are characterized by DNS. Suzuki et al., 2024 studied the detonation-turbulence interaction of H_2 jet flames, while the ignition of H_2 in the Argon Power Cycle is studied by Quan Reyes et al., 2024.

The numerous species involved in combustion processes, along with the wide range of time scales and magnitudes, necessitate large computational resources to produce accurate, extensive databases. These databases enable a detailed study of the complex flow physics needed to understand combustion (Cant, 2002). However, although the numerical simulations produce vast data in the spatial and temporal dimensions, the data are very sparse in parametric dimensions, as the fuel composition or the conditions of the flow. Hence, the generalization of the physics becomes very challenging.

Moreover, the computational costs in terms of memory usage and CPU time required for solving industrially relevant problems limit the advancement in the field. In light of what is stated above, Reduced Order Models (ROMs) represent a valuable alternative.

1.3 State of the art in Machine Learning applied to combustion

The use of different machine learning techniques for developing ROMs in reacting flows has gradually increased in recent years, as they provide an opportunity to simplify the complex

nature of a physical system and enable the formulation of high-fidelity, lower-dimensional models that accurately represent the physics with minimal error (Le Clainche et al., 2023; Parente and Swaminathan, 2024). ROMs enable the possibility of modelling the main flow dynamics at a reduced computational cost, what has applications in feature extraction and design optimization (Palomba et al., 2016). A comprehensive description of the application of data-driven machine learning techniques in combustion for developing ROMs is provided in Zhou et al., 2022. ROMs can be developed to serve diverse purposes, such as

- Dimensionality reduction and feature extraction. Bellemans et al., 2018 demonstrated how a reduced set of components, of dimensionality much lower than the original simulations, are enough to model the leading physics of nitrogen shock flows.
- Clustering. For example, D'Alessio et al., 2020 proposed a novel methodology for generating different reduced kinetic mechanisms for each cluster in CFD simulations of multi-dimensional flames.
- Accelerating numerical simulations. Mirgolbabaei and Echehki, 2013, where they proposed a parametrization of the thermo-chemical space in one dimension to solve a three-dimensional LES flow.
- Separate a complex system into simpler components. For example, C. Huang et al., 2022 created a component-based modelling framework, where each component can adopt either ROMs or full-order models (FOM) to finally reproduce the whole complex system.
- Flow control and design optimization. Torrez et al., 2011, implemented a ROM capable of modelling the turbulent reacting flow in ramjet and scramjet combustors, providing accurate results in seconds, useful for design optimization or control evaluation.
- Developing digital twins. Procacci et al., 2023, where they implemented a sparse sensing framework to build a hybrid numerical-experimental digital twin of a practical combustion system.

In the following Sections, the work done by the community using different machine learning algorithms is summarised. This algorithms are: Principal Component Analysis (PCA), Proper Orthogonal Decomposition (POD), Dynamic Mode Decomposition (DMD) and Artificial Neural Networks (ANN). Lastly, a brief explanation of the work done in flow control is elaborated in Sec. 1.3.5.

1.3.1 Principal Component Analysis

Principal Component Analysis (PCA, Jolliffe, 2002) is a statistical technique designed to identify lower-dimensional manifolds by analysing correlations among input variables. These variables are linearly combined to form principal components, ultimately reducing the dimensionality of the problem while preserving as much variability as possible (Bishop, 2006).

PCA has also found widespread applications to model flow combustion (Isaac et al., 2014; Zdybal et al., 2023). The application of PCA is suitable for speeding up numerical simulations. D'Alessio et al., 2020, developed a novel approach via clustering of variables to speed up CFD simulations. First, the clustering served for selecting proper reduced kinetic mechanisms

for each cluster. Then, as in each cluster the number of equations was lower, they achieved speed-up factors of ~ 4 . Malik et al., 2021, coupled PCA with Gaussian Process Regression (GPR) obtaining accurate results transporting two Principal Components instead of 35 species in the full model, which is a 3D LES simulation of the Sandia flames D–F. Other works have been carried out related to the speeding up of numerical simulations (Mirgolbabaei and Echekki, 2013; Isaac et al., 2014; Bellemans et al., 2017; D’Alessio et al., 2022).

PCA can also be employed for dimensionality reduction and feature extraction. Parente et al., 2009, identified lower-dimensional manifolds in turbulent flames which were able to reproduce the physical processes of the different state variables. Aversano et al., 2019 developed a ROM which extracted the physics-related information of an investigated combustion system and interpolated for parameter exploration with reduced computational cost. This work was paramount for the development of digital twins. The research group recently developed a multi-fidelity ROM of a combustion furnace operating under Moderate and Intense Low-oxygen Dilution (MILD) conditions (Özden et al., 2024). Other works related to feature extraction and dimensionality reduction can be found in the literature (Parente et al., 2011; Bellemans et al., 2018; D’Alessio et al., 2021).

Additional objectives can be pursued using PCA. Zdybał et al., 2023a, explored combustion datasets of varying complexity using PCA in order to find finding physically meaningful information from data. Tipler et al., 2022, applied PCA for selecting the most important variables of the combustion of complex fuels, prior to analysing them with artificial neural networks. The selection of the most important variables in a combustion dataset using PCA is a research topic of high interest, deeply studied in the literature (Jolliffe, 1972, 1973; Krzanowski, 1987).

In the present Doctoral Thesis, PCA has been applied in Sec. 6.1.2 to select subsets of variables from a combustion database, to then be analysed via HODMD (2.2).

1.3.2 Proper Orthogonal Decomposition

Proper Orthogonal Decomposition (POD) was introduced in the context of turbulence by Lumley, 1967. POD is a mathematical technique which decomposes a spatio-temporal database as a combination of POD modes, orthogonal in space and optimal in terms of preserved energy, with the temporal coefficients associated.

POD has been applied in the modelling of reacting flows, capturing the main flow dynamics and reducing computational cost. Swischuk et al., 2020, applied POD to a single-injector combustion process for feature extraction of the most relevant patterns of the CFD simulation. Following this work, McQuarrie et al., 2021, developed a predictive surrogate model based on POD via regularised Operator Inference to predict in time the main variables of the turbulent reactive flow. The main challenge in the prediction using POD-based model are the sharp gradients in the temperature and mass fractions of species (C. Huang et al., n.d.).

A variant of this method is widely used in the combustion community for the analysis of the temporal coefficients. This algorithm, named spectral POD (SPOD, Schmidt and Henningson, 2018), is essentially the frequency domain counterpart of POD. Bonuso et al., 2024, applied

SPOD to the chemiluminescence images of combustion of methane with and without H_2 blends. They found new frequencies related to the H_2 -enrichment.

Other studies address the characterisation of combustion instabilities. Choi et al., 2022, showed the effect of acoustic perturbations on lean blow off limits on the mesoscale burner array. Khalifehei et al., 2024, applied SPOD to bluff-body premixed turbulent combustion and showed the co-existence of sinuous and varicose oscillation modes. Brouzet et al., 2020, extracted the coherent structures that correlate with the radiated sound in lean and a stoichiometric methane-air turbulent premixed jet flame using SPOD. More studies where SPOD is applied to combustion databases can be found in the literature (Shoji et al., 2020; Casel et al., 2022).

1.3.3 Dynamic Mode Decomposition

Dynamic Mode Decomposition (DMD, Schmid, 2010) is a data-driven technique which decomposes a database into a expansion of DMD modes. In comparison to POD, the DMD modes are not orthogonal, although each DMD mode is associated to a single frequency.

In the context of reacting flows databases, some studies have been carried out using this algorithm in numerical simulations and experiments. Related to numerical simulations, Rodríguez-Pastor et al., 2024, applied DMD to two-dimensional laminar slit flames for inferring the underlying physics and the interaction between the flame and the flow. A. Sharma et al., 2024, investigated the combustion instabilities appearing in a methane-oxygen combustor using DMD, revealing the evolution of acoustic modes in the injector and the chamber for CVRC and multi-element combustor. J. Chen et al., 2024, employed DMD to investigate the effects of various methane fuel blending ratios and found high-speed shock waves occurring near the two side walls of the combustion chamber for certain values of the blending ratio. Additional applications of DMD for identifying patterns and flow instabilities in reacting flows include the work of Grenga et al., 2018 and Grenga and Mueller, 2020, who proposed a more efficient implementation of DMD, reducing the memory consumption, for the analysis of numerical simulations of turbulent reacting jets. Other studies analysing numerical databases can be found in the literature (Abou-Taouk et al., 2015; Ghani et al., 2015).

DMD can also be applied to experimental databases to identify the main instabilities leading changes in the flow. Richecoeur et al., 2012, employed DMD to investigate the main dynamics in experimental turbulent combustion, providing an in-depth evaluation of the algorithm's performance. Y. Liu et al., 2022, used DMD to study the flame fluctuation intensity when adding helium to a flame. Qiu et al., 2024, extracted the main structures appearing in turbulent flames from a hydrogen micromix burner for different inlet conditions. Gunaratne and Roy, 2021 analysed experiments of reacting flows behind a single bluff body. Depending on the equivalence ratio, they observed symmetric vortex shedding and asymmetric von Kármán vortices. Additional studies can be found in the literature related to the application of DMD to experimental combustion databases (Chatterjee et al., 2013; C. Huang et al., 2016).

In the present Doctoral Thesis, an extension of DMD, named Higher Order Dynamic Mode Decomposition (HODMD), is used as DMD is limited to specific conditions of the database

analysed. HODMD is more suitable for complex dynamical systems, and its use for combustion databases has been tested to the best of the author’s knowledge (Chapter 3). HODMD is the foundational framework of the developed algorithms in the present Doctoral Thesis.

1.3.4 Artificial Neural Networks

In the field of supervised learning, many studies have been carried out to create ROMs for combustion systems in several different applications.

Artificial Neural Networks (ANN) are computational models inspired by the way biological neural networks in the human brain process information. These machine learning algorithms are used for pattern recognition, classification, regression, and other tasks. They consist of layers of interconnected nodes (also called neurons) that process input data through weighted connections, and the output is typically transformed using an activation function (Rumelhart et al., 1986; Hecht-Nielsen, 1992; LeCun et al., 2015).

A basic structure of an ANN includes:

- Input Layer: Receives input data.
- Hidden Layers: One or more layers where data is processed through neurons. These layers apply weights and biases to the input and pass the result through an activation function.
- Output Layer: Produces the final output, depending on the task, like a classification label or a regression value.

The simplest form of an ANN is the autoencoder. In the context of reacting flows, autoencoders have been used with many different purposes. Zdybał et al., 2023b, developed a ROM based on autoencoders for dimensionality reduction in reacting flows. Li et al., 2020 used a single hidden-layer autoencoder to accelerate the design of new clean fuels.

More complex architectures have also been used with different applications. A. J. Sharma et al., 2020, constructed an ANN composed of 12 jointly-trained sub-networks to model a hydrogen combustion process. Other models have been used for the optimization of the performance of a combustion engine (Wong et al., 2010; Badra et al., 2020). Alternate applications use the limited information of experiments to predict some variables of interest for the combustion community. J. Huang et al., 2020, predicted the shape of a 3D flame feeding a convolutional neural network (CNN) with 2D projections of the flame. Furthermore, using a CNN, Maged and Nour, 2025, were able to predict the in-cylinder pressure in combustion engines from 2D flame propagation images. Bong et al., 2024, employed flame images captured with a low-cost monochrome camera to feed a CNN, predicting the fuel-air equivalence ratio and the total air flow rate.

ANN can incorporate physical laws, such as differential equations, into the training process. These neural networks are known as Physics-Informed Neural Networks (PINN). PINNs combine traditional machine learning with physics-based constraints, improving the accuracy of the predictions. Within the framework of reacting flows, Mariappan et al., 2024, developed a PINN to study the thermoacoustic interactions in a bluff body anchored flame combustor,

representative of ramjet and gas turbine combustors. Son and Lee, 2024, proposed a PINN-based methodology for computing the parameters of thermoacoustic oscillations in an annular combustor. Zhang et al., 2024, presented a PINN which can solve the chemical reaction kinetics, reconstruct the combustion process and infer the temperature and intermediate species. This PINN, applied on 0-D flames, 2-D laminar premixed flames and 3-D turbulent jet flames, therefore achieves high-fidelity predictions in tasks such as independent solving, reconstruction, inference, surrogate solving, and CFD acceleration.

During the present Doctoral Thesis, an ongoing project is currently working on the development of a new framework which combines POD (Sec. 1.3.2) with deep learning architectures for forecasting reacting flows. The new hybrid predictive model first segregates temporal and spatial information with POD. The POD modes contain the main flow characteristics, turning the model into a physics-based model. The deep learning block operates solely on the temporal domain with significantly reduced dimensionality. The results show that this hybrid architecture is able to predict accurately the evolution in time of reactive flows. The work has been presented in the *International Symposium on Unmanned Systems and the Defense Industry*, in 2022. From this conference an extended version was published in a conference proceedings (Corrochano et al., 2023f). A new paper extending the work examining different types of neural networks and an exhaustive analysis of the hyperparameters is under review, but a preprint can be found in Corrochano et al., 2023e.

1.3.5 Flow Control Strategies

Flow control can be categorized into active and passive approaches. Active flow control involves the use of external energy or active mechanisms to change the flow. In the context of combustion, Karagozian et al., 1996, explored the transverse gas jet injection to enhance the mixing in supersonic flows. Beér and Chigier, 1983, explored the use of atomizers for the same purpose. More recently, Palkin et al., 2022 studied the influence of a central jet in a reacting swirling flow in a model combustion chamber. They found that the central jet stabilised the flow and increased the mixing in the shear layer. Following this work, Kundashkin et al., 2022, explored the injection of gas in the swirling jet on the same geometry, suppressing the low-frequency oscillations.

Passive flow control depends exclusively on design features without introducing any energy source. Passive flow control techniques can be implemented through the optimal design of the industrial devices, adding structures or modifying the geometry. Karagozian, 1999, investigated the concept of a lobed fuel injector, which improved local mixing and reduced NO_x (nitrogen oxide) emissions by altering the geometry of the injector. What these techniques share is the absence of a clear guideline regarding the optimal location for implementing passive flow control.

ROMs can be used for designing flow control strategies. This is still a novel framework, with limited research conducted not only in reacting flows, but additionally in fluid dynamics in general. Recently, a few approaches in the field of flow control have been developed using data-driven methods, potentially leading to a significant reduction in computational costs. Deep Reinforcement Learning (DRL) has been applied as an active flow control strategy.

Varela et al., 2022 proposed an actuation using DRL to the wake past a cylinder with small jets attached to the surface of the cylinder. Guastoni et al., 2023 followed the same strategy for reducing the drag in turbulent channel flows. In the context of combustion systems, F. Chen et al., 2024, developed an energy management strategy based on DRL for ammonia-hydrogen hybrid electric vehicles, improving the operation efficiency.

In the present Doctoral Thesis, a novel methodology for passive flow control has been developed. This new method is based on HODMD and therefore, it is fully data driven and generalizable. The algorithm is explained in Section 2.5 and the application to several types of flow is shown in Chapter 5.

1.4 Objectives of the thesis

This Doctoral Thesis is focused on the adaptation of the current algorithms and the development of new ones for combustion databases. These combustion databases are usually high-dimensional and heavy in terms of memory. The algorithms created study the main physics of the flow and they are able to identify patterns in reactive flows. Therefore, they can be used to develop ROMs, which model the main dynamics of the system in a reduced dimensional space. The development of new algorithms and the modification of existing ones will lead to new ROMs to model reacting flows with accuracy at a reduced computational cost.

The reacting flows problem studied have laminar periodic dynamics, however, it is needed to understand what the algorithms are doing in the database. The complexity and the challenge in the creation of ROMs in laminar flames comes from the detailed kinetic mechanisms used. From a kinetic perspective, the CFD simulations conducted take into account a large number of species and thousands of reactions.

The complexity of the chemical systems considered arises not only from the high dimensionality of the thermochemical space but also from factors such as flame lift-off and the significant influence of differential diffusion, which generally contrasts with turbulent flames. Additionally, the intrinsic dimensionality of laminar reacting systems is greater than that of turbulent flames due to the lack of substantial interactions between fluid dynamics and system chemistry. The proposed methods effectively reduce data dimensionality while identifying the main frequencies and patterns driving flow dynamics, which are essential for developing a ROM. More details on the simulations used for testing the algorithms used can be found in the papers included in each chapter.

The objectives exposed in this Section can be interpreted in the diagram proposed in Fig. 1.2. In this diagram, the organization of the Doctoral Thesis is shown. Starting from a combustion database, we are going to apply HODMD (algorithm explained in Sec. 2.2) to model the main flows dynamics. Different purposes are pursued from the HODMD analysis, carried out in Chapter 3. An algorithm for variable clustering is explored in Chapter 4 and other for flow control is developed in Chapter 5.

The algorithms proposed are mainly based on HODMD (a fully data-driven algorithm further explained in Sec. 2.2), so they can readily be extended to analyse other types of flows. The

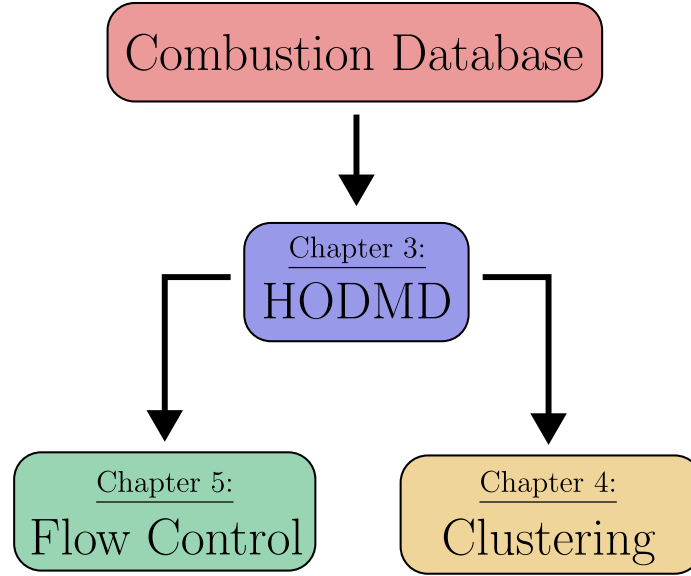


Figure 1.2: Diagram summarising the main objectives of the present Doctoral Thesis.

adjustments to HODMD for analysing combustion databases will be essential to develop ROMs used for clustering of variables and for flow control. These adjustments include:

- The use of **preprocessing techniques** to properly analyse the database, as it has variables with different magnitudes. The objective of this step is to extract the maximum amount of relevant information from the problem under study.
- The adaptation of the algorithms for **data dimensionality reduction** for the analysis of large amount of variables. The objective is to achieve maximum reduction of dimensionality, ensuring that the essential amount of information is fully preserved throughout the process.
- The **robustness** of the methodology and the potential for coupling this algorithm with state-of-the-art algorithms.

Once the main algorithm is adapted to the new type of databases desired to analyse, some modifications may be implemented with different purposes. Taking into account the large number of variables analysed at the same time, a clustering method for the variables may be desired. This clustering method will be fully data-driven, hence, it can be extended to other types of databases. It is based on the main dynamics of the system, so the clusters of variables will have their own main dynamics different from the other clusters. From a reduced-order modelling perspective, the algorithm should outperform established methods when organizing all variables into distinct clusters.

The last main objective of the present Doctoral Thesis is the creation of a novel data-driven algorithm for passive flow control. As it is based on the data, it is generalizable and useful for any type of flow (combustion, non-Newtonian, turbulence, etc.). This novel approach will encompass the following tasks:

- As it will be based on the calculation of the **structural sensitivity** (see Sec. 2.5), first

some test in laminar non-reactive flows will be made.

- The method should **provide the location** of the flow domain where to apply the passive flow control strategy.
- Different outputs can be obtained from this flow control, ranging from drag reduction and flow steadiness to, in reactive flows, **increased combustion efficiency** and **reduced pollutant emissions**.

1.5 Achievements, Collaborations, and Publications

Throughout the development of this Doctoral Thesis, the algorithms learned and created have been useful for different national and international collaborations. These international research collaborations have allowed me to broaden my scientific horizons beyond the main topic of this doctoral thesis. The knowledge acquired during the doctoral thesis is not confined to a single topic, the one presented in this manuscript, but rather spans across many different fields, thanks to the versatility of the algorithms presented in the methodology. I have closely worked with different research groups, enriching my perspective and knowledge in areas such as linear fluid dynamics, turbulence and non-Newtonian fluids, among others.

Data-driven algorithms have been applied to urban flows problems during the doctoral thesis. Urban flows are a research topic of high interest due to its implications in the pollutant dispersion, air quality and high propagation. A first collaboration with Universitat Politècnica de València (UPV) and Kungliga Tekniska Högskolan (KTH) provided useful information about the flow patterns forming around a simplified urban environment. From this first collaboration two research articles were published (Lazpita et al., 2022; Martínez-Sánchez et al., 2023), where the flow patterns appearing in this type of flow are examined. This is one of the topics considered in the European project MODELAIR, 2023–2026. In this European project, I am the representative of 12 Doctoral Candidates. In this context, I have been part of the committee in charge of organising the training courses for the Doctoral candidates.

Other collaborations and projects have been carried out where data-driven algorithms are applied to non-Newtonian fluids. In this type of flows, the inertial, viscous, elastic and plastic terms of the different equations induce the rise of complex flow patterns. Two types of non-Newtonian fluid have been studied. On the one side, the ElastoViscoPlastic (EVP) flow past a circular cylinder has been analysed with different conditions of the flow. In this configuration, transitions from periodic to chaotic when an elastic instability is triggered, i.e., when the elastic and plastic forces become sufficiently large. This work is in collaboration with the KTH, particularly with Prof. Outi Tammisola. Thanks to her invitation, I visited her research group at Stockholm and I worked there for two weeks in February 2024. This analysis is leading to a novel understanding of this type of flow, and a research article has already been submitted, assessing the flow patterns. Moreover one of this cases have been examined for flow control and the results have been presented in the 9th European Congress on Computational Methods in Applied Sciences and Engineering (Corrochano et al., 2024a).

On the other side, a regime called elastic turbulence was analysed in collaboration with the Okinawan Institute of Science and Technology (OIST). This regime occurs when the elasticity

of the fluid causes the turbulent motion in the flow, while the same case with a Newtonian fluid would be a laminar and steady solution. This project started in a collaboration in Okinawa during a visit of two months, from February to March 2023. The project continued and was presented in the "5th Madrid Turbulence Workshop", hosted by Prof. Javier Jiménez and funded by the CausT ERC advanced grant. From this Workshop, a conference proceedings has been published (Amor et al., 2024).

Table 1.1: International collaborations during the Doctoral Thesis. The name of the main researcher appear with the topic of the collaboration and the publication published or conference presented (if any).

Institution	Main researcher	Topic	Outcome
KTH	Ricardo Vinuesa	Urban flows	Martínez-Sánchez et al., 2023 Lazpita et al., 2022
		FDA nozzle	Corrochano et al., 2021
KTH	Outi Tammisola	EVP Flow past a circular cylinder	Ongoing
KTH	Luca Brandt	Gas-liquid planar jets	Corrochano et al., 2023b
Cranfield University	Nicholas Lawson	Transient aerodynamic flow field	Corrochano et al., 2022
UPV	Sergio Hoyas	Urban flows	Martínez-Sánchez et al., 2023 Lazpita et al., 2022
IMFT	David Fabre	Concentric jets	Corrochano et al., 2023a Martín et al., 2021
OIST	Marco Rosti	Elastic turbulence	Amor et al., 2024
Monash University	Julio Soria	Data repairing	Hetherington et al., 2024b
Università di Verona	Alessandro Marcon	Air pollution	Tagliaferro et al., 2024
ULB	Alessandro Parente	Combustion	
		Clustering	Corrochano et al., 2024c
		Time forecasting	Corrochano et al., 2023f
		Feature extraction	Corrochano et al., 2023d Corrochano et al., 2023c

During the doctoral thesis, multiple algorithms for data repairing, which can be also used for resolution enhancement, have been created and different collaborations have been carried out to test them. In the first project, different algorithms combining data-driven algorithms, such as SVD, and deep learning models were used to reconstruct aeronautical databases.

This collaboration resulted in a conference proceedings (Díaz et al., 2023) and in a research article (Díaz-Morales et al., 2024). Other collaboration, which resulted in a research paper (Hetherington et al., 2024b), compared different reconstruction methods between the hybrid deep learning models and the modal decomposition techniques developed.

The work done during the doctoral thesis is robust and the algorithms can be used in different fields of knowledge, since they have proven useful in pollen dispersion data, whose behaviour is very different from fluid databases. This collaboration with the University of Verona has resulted in a very enriching article on the reconstruction of aerobiological datasets with missing data, based on physical patterns (Tagliaferro et al., 2024). More information of the project can be found in the following link (<https://modelflows.github.io/modelflowsapp/airpollution/>).

All algorithms used during the present thesis have been collected into a single Python application, which contains also the algorithms proposed by other colleagues in our research group ModelFLOWS (<https://modelflowsmf.wixsite.com/modelflows>). The application, named as ModelFLOWS-app (Hetherington et al., 2024a), is available in GitHub and can be downloaded in the following link (<https://github.com/modelflows/ModelFLOWS-app>). Moreover, regarding this application, a patent has been granted in Spain. It has already been accepted and reviewed internally by our institution, UPM, as well as by external experts, receiving highly positive feedback.

Some other projects have been useful to prove the robustness of our algorithms and they have resulted in different publications, book chapters and conference talks and proceedings. The national and international collaborations can be summarised in Tab. 1.1, where the national projects where I have contributed also appear. All of the mentioned scientific outputs of the collaborations mentioned above and the ones which have not been mentioned are grouped in the two lists below.

Peer-reviewed journal articles & book chapters

The journal articles and conference proceedings which will enter on the different Sections of the Doctoral Thesis are the following:

- 2024 **Corrochano, A.**, Canals, J. L., Saavedra, L., & Le Clainche, S. (2024). Flow control with data-driven approaches. In *ECCOMAS Congress 2024 - 9th European Congress on Computational Methods in Applied Sciences and Engineering*.
 <https://doi.org/10.23967/eccomas.2024.085>
- 2024 **Corrochano, A.**, D'Alessio, G., Parente, A., & Le Clainche, S. (2024). Hierarchical higher-order dynamic mode decomposition for clustering and feature selection. *Computers & Mathematics with Applications*, 117, 25–50.
 <https://doi.org/10.1016/j.camwa.2024.01.003>
- 2023 **Corrochano, A.**, D'Alessio, G., Parente, A., & Le Clainche, S. (2023). Higher order dynamic mode decomposition to model reacting flows. *International Journal of Mechanical Sciences*, 249, 108219.
 <https://doi.org/10.1016/j.ijmecsci.2023.108219>

- 2023 **Corrochano, A.**, D'Alessio, G., Parente, A., & Le Clainche, S. (2023). Higher order dynamic mode decomposition to model reacting flows. In: Karakoc, T.H., Yilmaz, N., Dalkiran, A., Ercan, A.H. (eds) *New Achievements in Unmanned Systems. ISUDEF 2021. Sustainable Aviation*, pages 205–210, Springer, Cham.

📄 https://doi.org/10.1007/978-3-031-29933-9_22

- 2022 **Corrochano, A.**, & Le Clainche, S. (2022). Structural sensitivity in non-linear flows using direct solutions. *Computers & Mathematics with Applications*, 128, 69–78.

📄 <https://doi.org/10.1016/j.camwa.2022.10.006>

The journal articles, conference proceedings and book chapters coming from national and international collaborations which will not enter on the main topics of the Doctoral Thesis are the following:

- 2024 Tagliaferro, S., **Corrochano, A.**, Marchetti, P., Marcon, A., & Le Clainche, S. (2024). A new method based on physical patterns to impute aerobiological datasets. *PLoS ONE*, 19(11), e0314005.

📄 <https://doi.org/10.1371/journal.pone.0314005>

- 2024 Hetherington, A., Serfaty, D., **Corrochano, A.**, Soria, J., & Le Clainche, S. (2024). Data repairing and resolution enhancement using data-driven modal decomposition and deep learning. *Experimental Thermal and Fluid Science*, 157, 111241.

📄 <https://doi.org/10.1016/j.expthermflusci.2024.111241>

- 2024 Hetherington, A., **Corrochano, A.**, Abadía-Heredia, R., Lazpita, E., Muñoz, E., Díaz, P., Maiora, E., Martín, M. L., & Le Clainche, S. (2024). Modelflows-app: Data-driven post-processing and reduced order modelling tools. *Computer Physics Communications*, 301, 109217.

📄 <https://doi.org/10.1016/j.cpc.2024.109217>

- 2024 Amor, C., **Corrochano, A.**, Rota, G. F., Rosti, M. E., & Le Clainche, S. (2024). Coherent structures in elastic turbulent planar jets. *Journal of Physics: Conference Series*, 2753 (1), 012020.

📄 <https://doi.org/10.1088/1742-6596/2753/1/012020>

- 2024 Díaz-Morales, P., **Corrochano, A.**, López-Martín, M., & Le Clainche, S. (2024). Deep learning combined with singular value decomposition to reconstruct databases in fluid dynamics. *Expert Systems with Applications*, 238, 121924.

📄 <https://doi.org/10.1016/j.eswa.2023.121924>

- 2023 **Corrochano, A.**, Freitas, R. S. M., Parente, A., & Le Clainche, S. (2023). A Predictive Physics-Aware Machine Learning Model for Reacting Flows. In: Karakoc, T.H., Le Clainche, S., Chen, X., Dalkiran, A., Ercan, A.H. (eds) *New Technologies and Developments in Unmanned Systems. ISUDEF 2022. Sustainable Aviation*, pages 75–79, Springer, Cham.

📄 https://doi.org/10.1007/978-3-031-37160-8_12

- 2023 Díaz, P., **Corrochano, A.**, López-Martín, M., & Le Clainche, S. (2023). Machine Learning to Reconstruct Aeronautical Databases with Deep Neural Networks. In: Karakoc,

T.H., Le Clainche, S., Chen, X., Dalkiran, A., Ercan, A.H. (eds) *New Technologies and Developments in Unmanned Systems. ISUDEF 2022. Sustainable Aviation*, pages 125–129, Springer, Cham.

📄 https://doi.org/10.1007/978-3-031-37160-8_19

- 2023 **Corrochano, A.**, Sierra-Ausín, J., Martín, J. A., Fabre, D., & Le Clainche, S. (2023). Mode selection in concentric jets: the steady–steady 1 : 2 resonant mode interaction with O(2) symmetry. *Journal of Fluid Mechanics*, 971, A30.

📄 <https://doi.org/10.1017/jfm.2023.445>

- 2023 Martínez-Sánchez, Á., Lazpita, E., **Corrochano, A.**, Le Clainche, S., Hoyas, S., & Vinuesa, R. (2023). Data-driven assessment of arch vortices in simplified urban flows. *International Journal of Heat and Fluid Flow*, 100, 109101.

📄 <https://doi.org/10.1016/j.ijheatfluidflow.2022.109101>

- 2022 **Corrochano, A.**, Neves, A. F., Khanal, B., Le Clainche, S., & Lawson, N. J. (2022). DES of a Slingsby Firefly aircraft: Unsteady flow feature extraction using POD and HODMD. *Journal of Aerospace Engineering*, 35 (5), 04022063.

📄 [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0001457](https://doi.org/10.1061/(ASCE)AS.1943-5525.0001457)

- 2022 Lazpita, E., Martínez-Sánchez, Á., **Corrochano, A.**, Hoyas, S., Le Clainche, S., & Vinuesa, R. (2022). On the generation and destruction mechanisms of arch vortices in urban fluid flows. *Physics of Fluids*, 34 (5), 051702.

📄 <https://doi.org/10.1063/5.0088305>

- 2021 **Corrochano, A.**, Xavier, D., Schlatter, P., Vinuesa, R., & Le Clainche, S. (2021) Flow Structures on a Planar Food and Drug Administration (FDA) Nozzle at Low and Intermediate Reynolds Number. *Fluids*, 6(1):4.

📄 <https://doi.org/10.3390/fluids6010004>

Conference presentations

The underlined author is the speaker.

- 2025 Amor, C., **Corrochano, A.**, Soligo, G., Le Clainche, S., & Rosti, M. E. (May 2025). Sparse representation of elastic turbulence in planar jets. In *12th International Conference on Multiphase flow*, Toulouse, France (**upcoming**).

- 2024 **Corrochano, A.**, Canals, J. L., Saavedra, L., & Le Clainche, S. (October 2024). Flow control in laminar flames with data-driven approaches. In *2024 Meeting of the Spanish Section of The Combustion Institute*, Madrid, Spain.

- 2024 Amor, C., **Corrochano, A.**, Soligo, G., Le Clainche, S., & Rosti, M. E. (August 2024). Coherent structures in elastic turbulent planar jets. In *26th International Congress of Theoretical and Applied Mechanics*, Daegu, South Korea.

- 2024 **Corrochano, A.**, Canals, J. L., Saavedra, L., & Le Clainche, S. (June 2024). Flow control with data-driven approaches. In *9th European Congress on Computational Methods in Applied Sciences and Engineering*, Lisbon, Portugal.

- 2024 Barragán, G., **Corrochano, A.**, Abadía-Heredia, R., & Le Clainche, S. (March 2024). Improvement of coarse grid computational fluid dynamics simulations through higher order singular value decomposition and deep learning. In *Workshop on Machine Learning for Fluid Dynamics, ERCOFTAC*, Paris, France.
- 2024 Sengupta, A., Abadía-Heredia, R., **Corrochano, A.**, & Le Clainche, S. (March 2024). Forecasting in fluid dynamics through hybrid reduced order models. In *Workshop on Machine Learning for Fluid Dynamics, ERCOFTAC*, Paris, France.
- 2023 **Corrochano, A.**, Abadía-Heredia, R., Cialesi-Esposito, M., Brandt, L. & Le Clainche, S. (September 2023). Unsteady flow feature extraction of flapping events. In *18th European Turbulence Conference*, València, Spain.
- 2023 **Corrochano, A.**, Freitas, R. S. M., Parente, A. & Le Clainche, S. (July 2023). A hybrid predictive reduced order model for laminar flames. In *2nd Spanish Fluid Mechanics Conference*, Barcelona, Spain.
- 2023 **Corrochano, A.**, Amor, C., Foggi, G., Rosti, M. E. & Le Clainche, S. (June 2023). Unveiling elastic turbulence in free and wall-shear flows with data-analysis tools. In *Fifth Madrid Turbulence Workshop*, Madrid, Spain.
- 2022 **Corrochano, A.**, Vinuesa, R. & Le Clainche, S. (June 2022). Reduced order models in complex flows. In *1st Spanish Fluid Mechanics Conference*, Cádiz, Spain.
- 2022 **Corrochano, A.**, Freitas, R. S. M., Parente, A. & Le Clainche, S. (June 2022). A predictive physics-aware machine learning model for reacting flows. In *International Symposium on Unmanned Systems and the Defence Industry 2022*, Online.
- 2022 **Corrochano, A.** (January 2022). Modelling reactive flows with data-driven methods. In *Workshop Junior Interdisciplinar*, Online.
- 2021 **Corrochano, A.**, D'Alessio, G., Parente, A. & Le Clainche, S. (October 2021). Higher-order dynamic mode decomposition to model reacting flows. In *International Symposium on Unmanned Systems and the Defence Industry 2021*, Online.
- 2021 Abadía-Heredia, R., **Corrochano, A.**, Muñoz, E. & Le Clainche, S. (September 2021). Data-driven reduced order models in fluid dynamics. In *XIX Jacques-Louis Lions Spanish-French School on Numerical Simulation in Physics and Engineering*, Madrid, Spain.

1.6 Thesis Outline

The Doctoral Thesis follows the UPM format of the thesis by compendium. This first chapter has outlined the main aspects of the combustion processed and the state of art of the machine learning algorithms used by the community. In the second chapter, a description of the methodology used for the results is shown.

The third chapter focuses on the extraction of flow patterns using HODMD in combustion databases. In this chapter, a complementary conference proceeding is included, along with

the first main research article. The fourth chapter exposes the main results of the algorithm developed during the Doctoral Thesis for clustering of variables in combustion databases. As a result of this research, the second main research article was published.

The fifth chapter discusses the algorithm developed for flow control and its application to combustion databases. In this chapter, the third main publication is incorporated into the manuscript, describing the development of the algorithm and its comparison with the classical theory, as well as applications to laminar non-reactive flows. The application of this algorithm to combustion databases is shown in a complementary conference proceeding, featuring the control of a laminar flame.

Finally, in the sixth chapter, the main results and contributions of the thesis are discussed, with a Section reserved for future work and opportunities for further research.

Chapter 2

Methodology

This chapter summarizes the methodology developed and applied during this Doctoral Thesis. It provides an overview of the main algorithms utilized across the various articles presented in the Thesis, with a particular emphasis on the Higher Order Dynamic Mode Decomposition (HODMD) method as the foundational framework. Throughout the thesis, several variants of this method have been developed, expanding the scope of HODMD to include applications in reactive flows and flow control scenarios. The chapter begins by explaining the initial data processing required before introducing the HODMD algorithm and its extensions. This structured approach ensures a comprehensive understanding of the methodologies employed and highlights their significance in advancing the field.

2.1 Preprocessing techniques

Combustion data usually consists of the temperature and the chemical species. The scale of the temperature field is $\mathcal{O}(10^3)$, while some of the radicals presented in the chemical species ranges in $\mathcal{O}(10^{-6})$. Therefore, it is necessary to preprocess the data before analysing the database. In the present work, the technique used is centering and scaling. Centering means the subtraction of the temporal mean of the data, focusing on just the perturbations. Scaling the variables is motivated by the necessity to compare them on the same order of magnitude. These two operations can be joined into one single equation as

$$\tilde{\mathbf{x}}_j = \frac{\mathbf{x}_j - \bar{\mathbf{x}}_j}{c_j}, \quad (2.1)$$

where $\bar{\mathbf{x}}_j$ is the temporal mean of the variable $\mathbf{x}_j$ and c_j is the scaling factor. The scaling methods used in the present thesis are the following:

- **Auto scaling:** The standard deviation σ_j is used as the scaling factor ($c_j = \sigma_j$).
- **Pareto scaling:** The square root of the standard deviation is used as a scaling factor ($c_j = \sqrt{\sigma_j}$).
- **Range scaling:** The difference between the maximum and the minimum value is used

as a scaling factor ($c_j = \max(\mathbf{x}'_j) - \min(\mathbf{x}'_j)$, where $\mathbf{x}'_j = \mathbf{x}_j - \bar{\mathbf{x}}_j$).

→ **Vast scaling:** the product between the standard deviation and the so-called coefficient of variation ($\sigma_j/\bar{\mathbf{x}}_j$) is used as the scaling factor $\left(c_j = \frac{\sigma_j^2}{\bar{\mathbf{x}}_j}\right)$.

Different scaling methods emphasise the importance of some variables against others (Parente and Sutherland, 2013; Corrochano et al., 2023d). Auto scaling gives all the variables the same importance, as all the variables present a standard deviation equal to 1 after the scaling. Range scaling emphasise the largest variables. Vast scaling focus on the extremely stable variables. This effect is accentuated in Pareto scaling, which clearly focus on the main variables of the system.

The effect of the different scaling methods has been proved in other fields where the data analysed present same characteristics, as medicine (van den Berg et al., 2006) and molecular sciences (Yang et al., 2015). The importance of scaling the variables in a context where the variables range several orders of magnitude is clear. However, the most suitable scaling method depends on the problem under study and the objectives pursued. Therefore, it is important to test different scaling methods when testing an algorithm.

2.2 Higher Order Dynamic Mode Decomposition

Higher Order Dynamic Mode Decomposition (HODMD, Le Clainche and Vega, 2017a) is a widely used algorithm within the fluid mechanics community, with its robustness and accuracy validated across a broad range of applications. These applications include the creation of Reduced Order Models (ROMs) (Le Clainche and Ferrer, 2018; Le Clainche et al., 2020b; Beltrán et al., 2022) and the extraction of the main patterns in complex fluid flows problems (Amor et al., 2020; Lazpita et al., 2022; Muñoz and Le Clainche, 2022). HODMD analysis primarily focuses on complex fluid flows, encompassing turbulent flows (Le Clainche et al., 2020a; Corrochano et al., 2022), transitional flows (Le Clainche and Vega, 2017b), and data from noisy experiments (Le Clainche et al., 2017, 2019; Gutierrez-Castillo et al., 2022; X. Chen et al., 2024).

The applicability limits of this algorithm have also been successfully tested in other fields, such as:

- Disease detection using electrocardiograms (Groun et al., 2022).
- Classification of stars based on their light curve (Mekhaél et al., 2024).
- Feature extraction in plasma flows relevant to electric propulsion (Maddaloni et al., 2022; Perales-Díaz et al., 2023).
- Time prediction of electric power systems (Jones and Utyuzhnikov, 2024).
- Machinery health monitoring and prediction (Cheng et al., 2020).

HODMD is a data-driven algorithm used for the decomposition of spatio-temporal data $\mathbf{v}(x, y, t_k)$, for simplicity $\mathbf{v}_k$, collected at time instant t_k . In the present work, the spatio-

temporal data considered is bi-dimensional (x, y) ; however this equations can be extended for a third dimension in space. This algorithm is a robust extension of DMD (Schmid, 2010); therefore, the output is a DMD decomposition of M DMD modes as

$$\mathbf{v}(x, y, t_k) \simeq \mathbf{v}_k^{\text{approx}} = \sum_{m=1}^M a_m \mathbf{u}_m(x, y) e^{(\delta_m + i\omega_m)t_k}, \quad k = 1, \dots, K, \quad (2.2)$$

where $\mathbf{u}_m$ are the DMD modes, which are weighted by their amplitude a_m , they can grow or decay in time with their growth rate δ_m and oscillate in time with their frequency ω_m . For a practical implementation of the algorithm, it is necessary that the data is collected into a snapshot matrix $\mathbf{V}_1^K$. In this snapshot matrix, each snapshot is collected in a vector, and the snapshots are equidistant in time, with time interval Δt . The size of $\mathbf{V}_1^K$ is $J \times K$, where K is the number of temporal snapshots considered in the analysis and $J = N_x \times N_y$. N_x and N_y are the number of grid points in each one of the two spatial directions. The snapshot matrix can be described as follows:

$$\mathbf{V}_1^K = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_{K-1}, \mathbf{v}_K]. \quad (2.3)$$

On the left part of Fig. 2.1, the snapshot matrix $\mathbf{V}_1^K$ is shown. Using the previous notation, the standard DMD can be described as

$$\mathbf{V}_2^K = \mathbf{R} \mathbf{V}_1^{K-1}, \quad (2.4)$$

where $\mathbf{R}$ is the Koopman matrix. The Koopman matrix contains the dynamical information of the system, however, this correlation might be the simplest of Eq. (2.2). Le Clainche and Vega, 2017a extended the DMD method to analyse various types of complex flows, relating a snapshot with d time-delayed snapshots using higher-order Koopman assumption. This can be defined as

$$\mathbf{V}_{d+1}^K \simeq \mathbf{R}_1 \mathbf{V}_1^{K-d} + \mathbf{R}_2 \mathbf{V}_2^{K-d+1} + \dots + \mathbf{R}_d \mathbf{V}_d^{K-1}. \quad (2.5)$$

This equation divides the snapshot matrix into d blocks containing time-delayed $K - d$ snapshots, via a sliding window process. On the right part of Fig. 2.1, the sliding window process applied to $\mathbf{V}_1^K$ is presented. HODMD can be decomposed into two main steps, which will be described in the following sections.

2.2.1 Dimensionality reduction

The number of grid points is usually high when analysing complex data, so $J \sim \mathcal{O}(10^6)$ or more. Therefore, the computational cost of analysing directly this matrix is high. For this purpose, Singular Value Decomposition (SVD) (Sirovich, 1987) is applied to $\mathbf{V}_1^K$ (2.3). Apart from reducing the dimensionality of the problem, SVD reduces the noise and spatial redundancies and retain the large scales. Based on a tolerance ε_1 , it is possible to retain N linearly independent vectors, where $N \ll J$, as

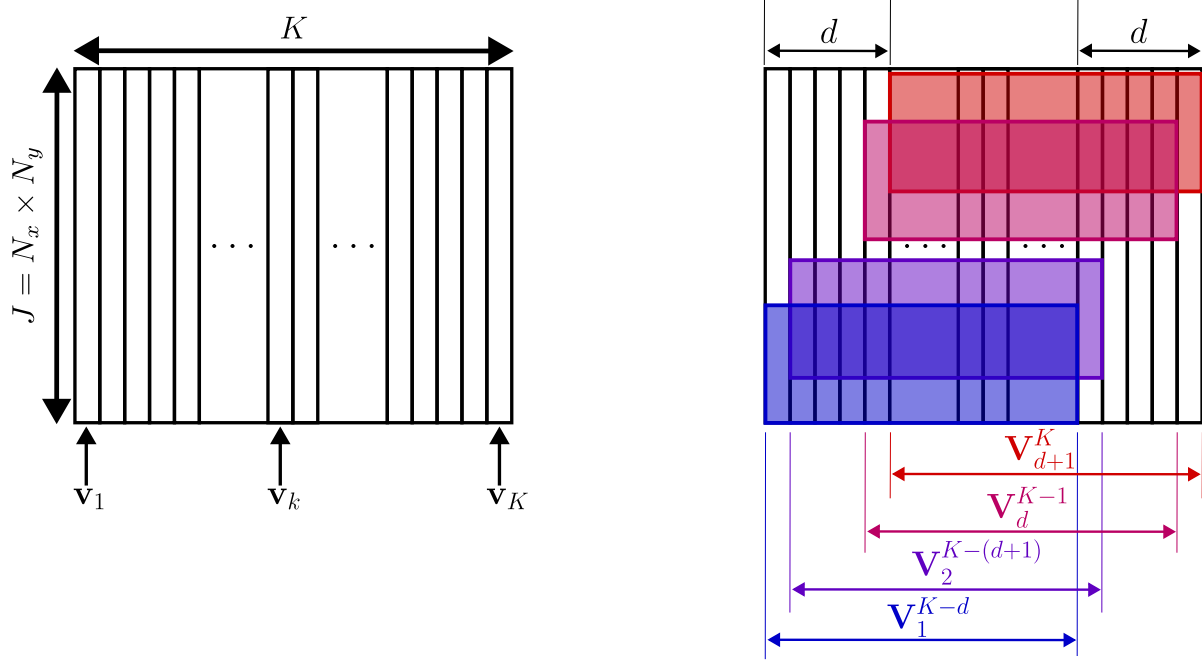


Figure 2.1: On the left side, a sketch representing the snapshot matrix $\mathbf{V}_1^K$ (2.3). On the right, a sketch representing the sliding window process defined in Eq. (2.5).

$$\mathbf{V}_1^K \simeq \mathbf{W}\mathbf{\Sigma}\mathbf{T}^T, \quad (2.6)$$

where $\mathbf{W}^T\mathbf{W} = \mathbf{T}^{T^T}\mathbf{T}$ are the identity matrix $\mathbb{I}_{N \times N}$ and $\mathbf{\Sigma}$ is a diagonal matrix which contains the singular values $\sigma_1, \dots, \sigma_N$ ordered by magnitude. The tunable tolerance ε_1 , which determine the number of SVD modes retained, is calculated as follows:

$$\sigma_{N+1}/\sigma_1 \leq \varepsilon_1. \quad (2.7)$$

Finally, the reduced snapshot matrix, which will be the input of the next step is calculated as

$$\hat{\mathbf{T}}_1^K = \mathbf{\Sigma}\mathbf{T}^T. \quad (2.8)$$

This matrix has dimensions $N \times K$, which is much smaller than the previous snapshot matrix.

2.2.2 The DMD-d algorithm

The reduced snapshot matrix has dimension $N \times K$, where N is the number of SVD modes retained, or the spatial complexity of the system. Recalling Eq. (2.5), the high order Koopman assumption can be applied to the reduced snapshot matrix as

$$\hat{\mathbf{T}}_{d+1}^K \simeq \hat{\mathbf{R}}_1 \hat{\mathbf{T}}_1^{K-d} + \hat{\mathbf{R}}_2 \hat{\mathbf{T}}_2^{K-d+1} + \dots + \hat{\mathbf{R}}_d \hat{\mathbf{T}}_d^{K-1}. \quad (2.9)$$

Moreover, the previous expression can be rewritten in matrix form as

$$\tilde{\mathbf{T}}_2^{K-d+1} = \tilde{\mathbf{R}} \tilde{\mathbf{T}}_1^{K-d}, \quad (2.10)$$

where $\tilde{\mathbf{R}}$ is the modified Koopman matrix and $\tilde{\mathbf{T}}_1^{K-d+1}$ is the modified snapshot matrix. Equation (2.10) can also be expressed as

$$\begin{bmatrix} \hat{\mathbf{T}}_2^{K-d+1} \\ \vdots \\ \hat{\mathbf{T}}_d^{K-1} \\ \hat{\mathbf{T}}_{d+1}^K \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbb{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{I} & \dots & \mathbf{0} & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbb{I} & \mathbf{0} \\ \widehat{\mathbf{R}}_1 & \widehat{\mathbf{R}}_2 & \widehat{\mathbf{R}}_3 & \dots & \widehat{\mathbf{R}}_{d-1} & \widehat{\mathbf{R}}_d \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{T}}_1^{K-d} \\ \hat{\mathbf{T}}_2^{K-d+1} \\ \vdots \\ \hat{\mathbf{T}}_d^{K-1} \end{bmatrix}. \quad (2.11)$$

This new matrix $\tilde{\mathbf{T}}_1^{K-d+1}$ is expected to exhibit redundancies. A new SVD is applied to the modified snapshot matrix as

$$\tilde{\mathbf{T}}_1^{K-d+1} \simeq \tilde{\mathbf{U}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{T}}^\top \simeq \tilde{\mathbf{U}} \overline{\mathbf{T}}_1^{K-d+1}, \quad (2.12)$$

where $\tilde{\mathbf{U}}^\top \tilde{\mathbf{U}} = \tilde{\mathbf{T}}^\top \tilde{\mathbf{T}}$ are the identity matrix $\mathbb{I}_{N' \times N'}$, and $\tilde{\mathbf{\Sigma}}$ is a diagonal matrix which contains the singular values $\tilde{\sigma}_1, \dots, \tilde{\sigma}_{N'}$, ordered by amplitude. The tolerance ε_1 is used to determine N' as

$$\tilde{\sigma}_{N'+1}/\tilde{\sigma}_1 < \varepsilon_1, \quad (2.13)$$

where $N' > N$. The modified snapshot matrix $\tilde{\mathbf{T}}_1^{K-d+1}$ organises the snapshot blocks of the high order Koopman assumption in columns. Therefore, the spatial complexity is increased from N to N' and the number of DMD modes M calculated, also known as spectral complexity, is increased to $M = \min(K, N')$. In cases where the number of snapshots K is larger than the spatial complexity N , HODMD outperforms the classical DMD method.

To complete the next step, Eq. (2.10) is multiplied by $\tilde{\mathbf{U}}^\top$, which invoking Eq. (2.12) results as

$$\overline{\mathbf{T}}_2^{K-d+1} = \overline{\mathbf{R}} \overline{\mathbf{T}}_1^{K-d}. \quad (2.14)$$

The new $N' \times N'$ -Koopman matrix is related to $\tilde{\mathbf{R}}$ by $\overline{\mathbf{R}} \simeq \tilde{\mathbf{U}}^\top \tilde{\mathbf{R}} \tilde{\mathbf{U}}$. Instead of computing this expression, the pseudoinverse of Eq. (2.14) is used. First, SVD is applied to $\overline{\mathbf{T}}_1^{K-d}$ as

$$\overline{\mathbf{T}}_1^{K-d} = \overline{\mathbf{U}} \mathbf{\Lambda} \overline{\mathbf{V}}^\top. \quad (2.15)$$

Substituting Eq. (2.15) into Eq. (2.14) the following equation is obtained:

$$\overline{\mathbf{R}} = \overline{\mathbf{T}}_2^{K-d+1} \overline{\mathbf{V}} \mathbf{\Lambda}^{-1} \overline{\mathbf{U}}^\top. \quad (2.16)$$

Calculating the eigenvalues μ_m and eigenvectors $\overline{\mathbf{q}}_m$ of the matrix $\overline{\mathbf{R}}$, it is possible to calculate the DMD modes $\mathbf{u}_m$ and the frequencies ω_m and growth rates δ_m from Eq. (2.2). Using N' eigenvectors $\overline{\mathbf{q}}_m$ and eigenvalues μ_m , it is also possible to calculate the reduced DMD expansion for the reduced snapshots, Eq. (2.8), as

$$\hat{\mathbf{t}}_k \simeq \sum_{m=1}^M \hat{a}_m \hat{\mathbf{q}}_m e^{\mu_m t_k}, \quad k = 1, \dots, K. \quad (2.17)$$

The reduced DMD modes $\hat{\mathbf{u}}_m$ are obtained by retaining the first N components of the vectors $\hat{\mathbf{q}}_m = \tilde{\mathbf{U}} \overline{\mathbf{q}}_m$. Then the DMD modes $\mathbf{u}_m$ are calculated, invoking Eq. (2.6), as $\mathbf{u}_m = \mathbf{W} \hat{\mathbf{u}}_m$.

From the eigenvalues μ_m obtained, the growth rates δ_m and frequencies ω_m are given by

$$\delta_m + i\omega_m = \frac{\log(\mu_m)}{\Delta t}. \quad (2.18)$$

Lastly, the amplitudes $\hat{a}_m$ from Eq. (2.17) are calculated via least square fitting (K. K. Chen et al., 2012). The amplitudes a_m from Eq. (2.2) are calculated establishing that the DMD modes $\mathbf{u}_m$ have unit norm. The number of retained DMD modes M is calculated with a second tolerance ε_2 , as

$$a_{M+1}/a_1 \leq \varepsilon_2. \quad (2.19)$$

The error in the reconstruction using HODMD is calculated using the Relative Root Mean Square Error (RRMSE), calculated as

$$RRMSE = \sqrt{\frac{\sum_{k=1}^K \|\mathbf{v}_k^{\text{approx}} - \mathbf{v}_k\|_2^2}{\sum_{k=1}^K \|\mathbf{v}_k\|_2^2}}. \quad (2.20)$$

A more detailed description and applications of this algorithm can be found in Vega and Le Clainche, 2020, including the implementation of this algorithm in Matlab. Recently, a version of the algorithm has been implemented in Python by Hetherington et al., 2024a.

2.3 Multi dimensional HODMD

In the explanation above, HODMD is applied to the snapshot matrix (2.3), where the snapshots are collected in columns. In fluid flows databases, the data is usually collected in a structured spatial mesh. Moreover, the data analysed, which is the velocity field in non-reactive flows and the temperature and chemical species in reactive flows, can be analysed at the same time.

In these cases, the data can be collected into a multi-dimensional array, with more than two indexes (as in the case of the snapshot matrix). The multidimensional array is referred to as the snapshot tensor, in order to maintain coherence in the names. In the case of two spatial dimensions and more than one variable to analyse, the snapshots can be organized in a fourth-order tensor $\mathbf{V}_{j_1 j_2 j_3 k}$, for $j_1 = 1, \dots, N_v$; $j_2 = 1, \dots, N_x$; $j_3 = 1, \dots, N_y$; and $K = 1, \dots, K$. K is the number of snapshots, as in the snapshot matrix, N_v is the number of variables analysed and N_x and N_y are the number of grid points in each one of the spatial directions. The snapshot tensor presents indexes associated with each one of the components. A visual representation of the transformation from the snapshot matrix to the snapshot tensor can be found in Fig. 2.2. For sake of clarity, the case illustrated would be a database with just one variable, two spatial dimensions and one temporal dimension, so when converting the snapshot matrix to the tensor matrix, the latter will have three dimensions.

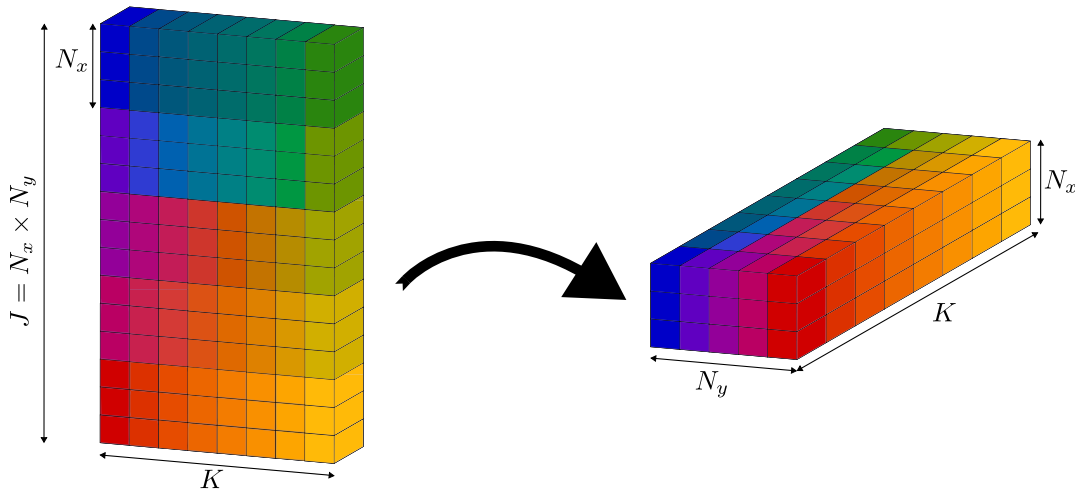


Figure 2.2: A visual representation of the transformation from the snapshot matrix to the snapshot tensor, in a database with two spatial dimensions and one temporal

For the analysis of the snapshot tensor, a newer version of HODMD is applied, named multi-dimensional HODMD (mdHODMD, Le Clainche et al., 2017). The tensor snapshot is introduced into the HODMD algorithm changing the first step from SVD to higher order SVD (HOSVD, Tucker, 1966). The HOSVD method Eq. (2.21) can be used to reduce the dimensionality of the data in an efficient way. The algorithm applies SVD in each one of the components of the tensor. Therefore, with HOSVD, there are modes related to each one of the dimensions of the tensor (in our case, variables), two spatial and the temporal dimension. The expansion of HOSVD results as

$$\mathbf{V}_{j_1 j_2 j_3 k} = \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} \sum_{p_3=1}^{P_3} \sum_{n=1}^N \mathbf{S}_{p_1 p_2 p_3 n} \mathbf{W}_{j_1 p_1}^{(v)} \mathbf{W}_{j_2 p_2}^{(x)} \mathbf{W}_{j_3 p_3}^{(y)} \mathbf{T}_{kn}, \quad (2.21)$$

where $\mathbf{S}_{p_1 p_2 p_3 n}$ is the core tensor and the columns of the matrices $\mathbf{W}^{(v)}$, $\mathbf{W}^{(x)}$, $\mathbf{W}^{(y)}$ and $\mathbf{T}$ are the SVD modes of the decomposition, one matrix for each one of the dimensions of the tensor. The temporal matrix $\mathbf{T}$, of dimensions $N \times K$, is the input of the DMD-d algorithm

(Section 2.2.2). As well as in the classical HODMD, the number of SVD modes N associated with the temporal matrix is calculated with the tunable tolerance ε_1 :

$$\sigma_{N+1}/\sigma_1 \leq \varepsilon_1. \quad (2.22)$$

The truncation of modes with the tolerance ε_1 is also applied to the matrices $\mathbf{W}^{(v)}$, $\mathbf{W}^{(x)}$ and $\mathbf{W}^{(y)}$. If these matrices are not truncated, just in the temporal one, HOSVD would provide the same results as the SVD algorithm. For the same level of truncation, HOSVD provides similar reconstruction error, while the compression of the data drastically increases. The compression increases as the order of the tensor increases (Lorente et al., 2010).

2.4 Hierarchical HODMD

Based on HODMD, new algorithms have been developed to analyse databases with different purposes. Hierarchical HODMD (h-HODMD) is a new algorithm, developed during the PhD, which is able to identify and create clusters of variables having the same dynamics behaviour (Corrochano et al., 2024c).

The aim of this algorithm is to target combustion databases, which contain several variables to be analysed. These variables, often the temperature field and the chemical species, present the same dynamics, as the system is coupled; however, they may present different amplitudes when analysing each one of them separately. The algorithm is able to analyse every variable at the same time, as in the mdHODMD, and then pick the variables which have similar dynamics. Other techniques have been used for clustering as K-Means (MacQueen, 1967; Lloyd, 1982); PCA (D'Alessio et al., 2020) or its variant, Vector Quantization via Principal Component Analysis (VQPCA, Parente et al., 2011). However, these techniques do not take into account the temporal variations. This novel algorithm is based on the DMD modes, thus it takes into account the temporal variation of the different variables and it is based on the physics of the system.

The algorithm is applied iteratively, extracting a cluster formed by some of the variables with their DMD modes, frequencies and amplitudes. h-HODMD is based on the error in the reconstruction of the data (2.20) and a tunable tolerance to retain the variables that verify

$$\sqrt{\frac{\sum_{k=1}^K \|\mathbf{x}_{j,k}^{\text{DMD}} - \mathbf{x}_{j,k}\|_2^2}{\sum_{k=1}^K \|\mathbf{x}_{j,k}\|_2^2}} < \varepsilon_v, \quad (2.23)$$

where $\mathbf{x}_j$ is the j -th variable. A summary of the algorithm can be found in Fig. 2.3.

As seen, first HODMD is applied to the full database and the DMD modes and frequencies are obtained. The database is reconstructed following the DMD expansion (2.2) and the error is calculated as in Eq. (2.23). The variables that satisfy the criterion are grouped into a cluster with their main dynamics, and removed from the initial database. The rest of the variables are analysed again with HODMD, and new clusters are created. The variables not grouped into clusters are progressively better reconstructed.

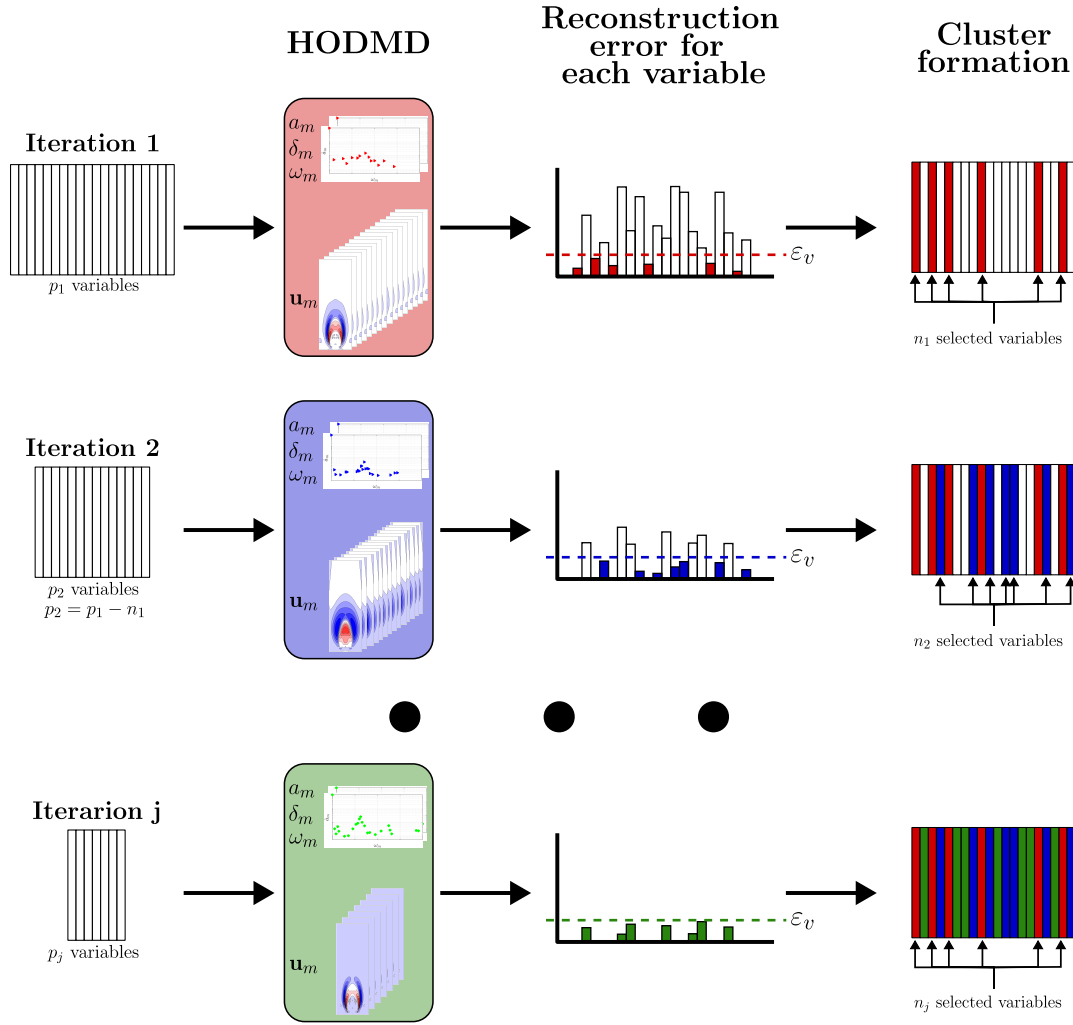


Figure 2.3: A visual representation of the h-HODMD algorithm. Note that in each iteration, HODMD is applied to obtain the modes and the reconstruction error is calculated for each variable. If the reconstruction error is below a tolerance ϵ_v , the variable is selected to form a cluster.

This algorithm is applied to combustion databases, and the main advantage of this algorithm is that the clusters created are based on the dynamics of the system, and have associated frequencies and modes.

2.5 HODMD for flow control

A novel algorithm to calculate the structural sensitivity in non-linear flows has been created during the Doctoral Thesis. Structural sensitivity defines the spatial regions of the flow most sensitive to changes. Therefore, it changes the leading flow driving the dynamics. While this concept is mostly used in fluid flows, it can be applied to solid mechanics (Martínez et al., 2020). Usually, the structural sensitivity is calculated using the adjoint equations, further explained in this Section, what makes the analysis expensive from a computational

perspective. Moreover, it is difficult to program these equations (especially in combustion) and the structural sensitivity cannot be directly applied in experimental settings.

The algorithm used for calculating the structural sensitivity is fully data-driven and it can be used to various types of flows. However, it is crucial first to understand the classical way to obtain these maps, via linearisation of the Navier Stokes equations.

2.5.1 Linear stability

For simplicity of the calculations, the problem studied is the two-dimensional unsteady incompressible fluid motion. This problem can be described with the following Navier Stokes (NS) equations:

$$\nabla \cdot \mathbf{v} = 0, \quad (2.24)$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla p + \frac{1}{Re} \Delta \mathbf{v}, \quad (2.25)$$

where $\mathbf{v}$ and p are the non-dimensional velocity and pressure fields, respectively, and Re is the Reynolds number. The Reynolds number measures the balance between the inertial forces and the viscous forces. It can be defined as $Re = VL/\nu$, where V and L are characteristic velocity and length of the problem under study, and ν is the kinematic viscosity of the fluid.

These equations are non-linear, and prior to apply the linear stability theory, they have to be linearised. Following the Reynolds decomposition, the solution of NS equations (2.24)–(2.25) can be decomposed as a base flow, invariant in time, plus a fluctuating field, as $\mathbf{v} = \bar{\mathbf{v}} + \mathbf{v}'$ and $p = \bar{p} + p'$. Introducing this decomposition in the previous equations, it yields

$$\nabla \cdot (\bar{\mathbf{v}} + \mathbf{v}') = 0, \quad (2.26)$$

$$\frac{\partial \mathbf{v}'}{\partial t} + (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}} + (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}' + (\mathbf{v}' \cdot \nabla) \bar{\mathbf{v}} + (\mathbf{v}' \cdot \nabla) \mathbf{v}' = -\nabla(\bar{p} + p') + \frac{1}{Re} \Delta(\bar{\mathbf{v}} + \mathbf{v}'). \quad (2.27)$$

The convective terms in Eq. (2.27) are renamed as $\mathbf{M} = (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}}$, $\mathbf{L} = (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}' + (\mathbf{v}' \cdot \nabla) \bar{\mathbf{v}}$ and $\mathbf{N} = (\mathbf{v}' \cdot \nabla) \mathbf{v}'$. The term $\mathbf{M}$ corresponds to a mean or averaged field, $\mathbf{L}$ to a linear term and $\mathbf{N}$ to a non-linear term.

Although the non-linear term is neglected in the linear stability theory, and the mean term is balanced out with the other terms of the NS equations, the two terms $\mathbf{M}$ and $\mathbf{N}$ will recover importance in the following sections, when a data-driven method is introduced to calculate these terms directly from the analysis of the databases. Assuming a two-dimensional Cartesian field, the velocity vector is defined as $\mathbf{v} = (v_1, v_2)$, corresponding to the streamwise and normal velocity components. The velocity field is connected with the spatial coordinates $\mathbf{x} = (x_1, x_2)$. Following this notation, the two terms $\mathbf{M}$ and $\mathbf{N}$ can be expressed as

$$\mathbf{M} = \bar{v}_j \frac{\partial \bar{v}_i}{\partial x_j}, \text{ for } i = 1, 2, \quad (2.28)$$

$$\mathbf{N} = v'_j \frac{\partial v'_i}{\partial x_j}, \text{ for } i = 1, 2. \quad (2.29)$$

Following the Reynolds decomposition the adjoint NS equations can be defined as

$$-\frac{\partial \mathbf{v}^*}{\partial t} + (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}} + (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}^* + (\mathbf{v}^* \cdot \nabla) \bar{\mathbf{v}} + (\mathbf{v}^* \cdot \nabla) \mathbf{v}^* = -\nabla(\bar{p} + p^*) + \frac{1}{\text{Re}} \Delta(\bar{\mathbf{v}} + \mathbf{v}^*), \quad (2.30)$$

where $\mathbf{v}^*$ and p^* denotes the adjoint counterpart of $\mathbf{v}'$ and p' . As in the direct equations, the convective terms are renamed as $\mathbf{M}^* = \mathbf{M} = (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}}$ (average field), $\mathbf{L}^* = (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}^* + (\mathbf{v}^* \cdot \nabla) \bar{\mathbf{v}}$ (linear term) and $\mathbf{N}^* = (\mathbf{v}^* \cdot \nabla) \mathbf{v}^*$ (non-linear term). Note that, as the average field is the same Eq. (2.27) and Eq. (2.30), the terms $\mathbf{M}$ and $\mathbf{M}^*$ are equal. In Cartesian notation, the non-linear term can be expressed as

$$\mathbf{N}^* = v_j^* \frac{\partial v_i^*}{\partial x_j}, \text{ for } i = 1, 2. \quad (2.31)$$

2.5.2 Structural sensitivity

The structural sensitivity identifies the regions of the flow where it is possible to produce the highest changes to the leading mode (Giannetti and Luchini, 2007). It can be determined introducing a forcing term $\mathbf{f}$ into the linearised NS equations (Eqs. (2.26)–(2.27)). The base flow, invariant in time, satisfies the NS equations, as

$$\nabla \cdot \bar{\mathbf{v}} = 0, \quad (2.32)$$

$$(\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}} = -\nabla \bar{p} + \frac{1}{\text{Re}} \Delta \bar{\mathbf{v}}. \quad (2.33)$$

Introducing Eqs. (2.32)–(2.33) into Eqs. (2.26)–(2.27), and considering that the non-linear term is negligible $(\mathbf{v}' \cdot \nabla) \mathbf{v}' \ll 1$, it is possible to obtain the following equations

$$\nabla \cdot \mathbf{v}' = 0, \quad (2.34)$$

$$\frac{\partial \mathbf{v}'}{\partial t} + (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}' + (\mathbf{v}' \cdot \nabla) \bar{\mathbf{v}} = -\nabla p' + \frac{1}{\text{Re}} \Delta \mathbf{v}' + \mathbf{f}. \quad (2.35)$$

From now on, it is supposed a solution of the fluctuations as $\mathbf{v}' = \mathbf{u} e^{\sigma t}$, where $\mathbf{u}$ is the global velocity mode and $\sigma = \delta + i\omega$ is the complex eigenvalue which determine the growthrate and

the frequency of the leading mode. The growthrate determines the stability of the solution. For values of $\delta < 0$, the solution of the system is stable and the flow recovers the baseflow solution when evolving in time. For values of $\delta > 0$, the solution is unstable and changes in the flow appear, such as symmetry breaking, flow unsteadiness and flow three-dimensionality. When the frequency of the mode $\omega = 0$, the changes on the flow are applied on the shape of the steady solution; while when $\omega \neq 0$, the solution becomes unsteady.

For a localized feedback process in (x_0, y_0) (in two-dimensional fields), the forcing term, proportional to the global velocity mode, can be mathematically described as

$$\mathbf{f} = \mathbf{C}_0 \delta(x - x_0, y - y_0) \mathbf{u}, \quad (2.36)$$

where $\delta(x - x_0, y - y_0)$ is the Dirac delta function and $\mathbf{C}_0$ is a matrix operator. Variations of the eigenvalues due to a small-amplitude modification of the force $\delta\sigma(x_0, y_0)$ can be represented as

$$|\delta\sigma(x_0, y_0)| \leq \|\mathbf{C}_0\| \cdot \|\mathbf{u}(x_0, y_0)\| \cdot \|\mathbf{u}^*(x_0, y_0)\|, \quad (2.37)$$

being $\mathbf{u}^*$ the adjoint global stability mode, $\|\cdot\|$ the Euclidean norm of a vector and $|\cdot|$ the modulus of a complex number. Equation (2.37) shows that the product between the direct and adjoint fields gives the maximum possible coupling among the velocity components (Giannetti and Luchini, 2007). Hence, the structural sensitivity calculated in linear flows is defined as

$$\mathbf{S}^{LIN} = \|\mathbf{u}\| \cdot \|\mathbf{u}^*\|. \quad (2.38)$$

For more details of this definition and a further mathematical explanation of the concept, it is recommended to read Ref. (Giannetti and Luchini, 2007, Marquet et al., 2008).

2.5.3 Non-linear structural sensitivity

In the previous sections, the linear approximation of the structural sensitivity has been explained. However, the objective of the present work is to analyse non-linear flows, and calculate the so-called non-linear structural sensitivity. In this type of flows, the non-linear term $\mathbf{N}$ plays a fundamental role.

As proposed by McKeon and Sharma, 2010, the non-linear term of the NS equations (2.24)–(2.25) can define the forcing term $\mathbf{f}$. Therefore, following the linear theory, this term applied to a localized feedback process in (x_0, y_0) can be mathematically described as

$$\mathbf{f} = \mathbf{C}_1 \delta(x - x_0, y - y_0) (\mathbf{u} \cdot \nabla) \mathbf{u}. \quad (2.39)$$

where $\mathbf{C}_1$ is a matrix operator. The variations of the eigenvalues due to small-amplitude modification of the force $\delta\sigma(x_0, y_0)$ can be defined similarly to the linear theory, as

$$|\delta\sigma(x_0, y_0)| \leq \|\mathbf{C}_1\| \cdot \|(\mathbf{u}(x_0, y_0) \cdot \nabla)\mathbf{u}(x_0, y_0)\| \cdot \|(\mathbf{u}^*(x_0, y_0) \cdot \nabla)\mathbf{u}^*(x_0, y_0)\| \quad (2.40)$$

Considering these terms and the direct and adjoint modes, the non-linear structural sensitivity can be calculated as

$$\mathbf{S}^{NL} = \|\mathbf{N}\| \cdot \|\mathbf{N}^*\|. \quad (2.41)$$

Note that the previous equation present the nonlinear terms from the direct and adjoint NS equations (2.27) and (2.30). There is still the need of calculation of the adjoint problem to calculate the structural sensitivity. In the present work, the novel algorithm identify the non-linear structural sensitivity without the need of calculating the adjoint solution.

2.5.4 Use of data-driven methods for calculating the non-linear structural sensitivity

In classical theory, the starting point is the base flow, which is perturbed, and the temporal evolution of the perturbation is analysed. The linear equations are solved as an initial value problem, where the ODE (ordinary differential equation) linear system is $\frac{d\mathbf{v}}{dt} = \mathbf{R}\mathbf{v}$. The solution of this problem provides the eigenvalues σ_m and eigenvectors $\mathbf{u}_m$, which are also known as the global velocity modes.

Data-driven methods, such as the classical DMD or HODMD (Section 2.2), directly consider the evolution of the non-linear NS equations in the form of temporal snapshots extracted from the simulation or experiment and collected in the snapshot matrix (2.3). The snapshot matrix is also the solution of a dynamical system which can be described as the state-transition equation (2.4) (or Eq. (2.5) in the case of analysing complex flows with HODMD). After the application of DMD (or HODMD), the snapshots are approximated as an expansion of DMD modes, as in Eq. (2.2).

Considering the same snapshot matrix (2.3), but descending in time, it is possible to obtain the so-called adjoint snapshot matrix, as

$$\mathbf{V}_K^1 = [\mathbf{v}_K, \mathbf{v}_{K-1}, \dots, \mathbf{v}_k, \mathbf{v}_{k-1}, \dots, \mathbf{v}_2, \mathbf{v}_1]. \quad (2.42)$$

The adjoint is not the same as time reversal. However, this approximation is correct when the advection term dominates the NS equations (see the Supplemental Material of (Luchini and Bottaro, 2014) for more information). In other cases, such as when the diffusive term dominates, further research must be carried out.

This new adjoint snapshot matrix can be the solution of a new state transition equation $\mathbf{v}_k = \mathbf{R}^T \mathbf{v}_{k+1}$. Applying DMD (or HODMD) to this snapshot matrix, the following DMD expansion is obtained:

$$\mathbf{v}^*(x, y, t_k) \simeq \tilde{\mathbf{v}}_k^* = \sum_{m=1}^M a_m \mathbf{u}_m^*(x, y) e^{(-\delta_m + i\omega_m)t_k}, \quad k = 1, \dots, K, \quad (2.43)$$

where $\mathbf{u}_m^*$ are the adjoint DMD modes. These adjoint modes are related to a direct DMD mode (from the expansion Eq. (2.2)) with similar frequency and opposite growth rate. The direct $\mathbf{u}_m$ and adjoint $\mathbf{u}_m^*$ modes are orthogonal between them, i.e. $\langle \mathbf{u}_m, \mathbf{u}_m^* \rangle = 0$, where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. It is worth noting that these adjoint modes are not the same as the eigenmodes obtained in the classical theory.

In a general case, when applying DMD to the direct snapshot matrix, the mode with highest amplitude $\mathbf{u}_1$ is steady ($\omega = 0$) and it corresponds to the mean flow. Hence, $\mathbf{u}_1 = \bar{\mathbf{v}}$. The second highest mode in periodic non-linear solutions $\mathbf{u}_2$ would correspond to the eigenmode calculated by the linear theory (Le Clainche et al., 2018). It moves at a certain frequency $\omega \neq 0$ and it approximates the flow fluctuations as $\mathbf{v}' \simeq \mathbf{u}_2 e^{(\delta_2 + i\omega_2)t} + \text{c.c.}$, where c.c. is the complex conjugate.

The calculation of the non-linear structural sensitivity using HODMD follows the steps presented in the next Section.

Implementation of the algorithm

- **Step 1: Simplification of the snapshot matrix**

The database analysed can be numerical or experimental, representing complex non-linear flows. Therefore, it is necessary to simplify the original data, removing the noise and retaining the main dynamics of the flow.

DMD is applied to the snapshot matrix $\mathbf{V}_1^K$ (2.2) and the DMD expansion retains just the first three modes, the ones with highest amplitude. The first mode retains the information related to the temporal mean. The second one has the information related to the leading frequency mode, which represents the velocity fluctuations. The third mode is the complex conjugate of the second mode, as the reconstructed solution has to be real. The leading frequency mode is in charge of producing the highest changes on the flow. Through this simple expansion, it is possible to obtain a simplified snapshot matrix $\tilde{\mathbf{V}}$. This approach aims to approximate the linear solution of the flow.

- **Step 2: Calculation of the direct DMD mode**

HODMD is applied to the simplified snapshot matrix $\tilde{\mathbf{V}}$, obtaining the DMD expansion (2.2) and the direct DMD modes $\mathbf{u}_m$.

- **Step 3: Calculation of the adjoint DMD mode**

The simplified snapshot matrix $\tilde{\mathbf{V}}$ is organised descending in time, as in Eq. (2.42). HODMD is applied to this new snapshot matrix $\tilde{\mathbf{V}}^*$, obtaining the DMD expansion (2.43) and the adjoint DMD modes $\mathbf{u}_m^*$. At this step, the temporal mean is also removed from the snapshot matrix to simplify the calculations, as the mean is the same in both direct and adjoint matrices.

- **Step 4: Calculation of the non-linear structural sensitivity**

The so-called non-linear structural sensitivity is calculated as follows:

$$\tilde{\mathbf{S}}^{NL} = \omega ||\mathbf{M}|| \cdot ||\mathbf{M}|| + 2||\mathbf{N}|| \cdot ||\mathbf{N}^*||. \quad (2.44)$$

The term $\mathbf{M}$ is calculated as in Eq. (2.28) with the first mode obtained in Step 2, the one with $\omega = 0$. The term $\mathbf{N}$ is calculated as in Eq. (2.29), using the second highest-amplitude mode from the DMD expansion at Step 2, the first with $\omega \neq 0$. Lastly, the term $\mathbf{N}^*$ is calculated using the highest amplitude mode obtained in Step 3. Note that at this Step, there is no steady mode, as the temporal mean was removed from the calculations.

For more details about the algorithm, examples and information regarding the theoretical base, please see Corrochano and Le Clainche, [2022](#).

Chapter 3

Higher order dynamic mode decomposition to model reacting flows

3.1 Preliminary results

3.1.1 Publication information

Title: Higher order dynamic mode decomposition to model reacting flows

DOI: https://doi.org/10.1007/978-3-031-29933-9_22

Journal: New Achievements in Unmanned Systems. ISUDEF 2021. Sustainable Aviation

Date: 28 June 2023

Publisher: Springer, Cham

Copyright: © 2023 The Author(s), under exclusive license to Springer Nature Switzerland AG

Higher Order Dynamic Mode Decomposition to Model Reacting Flows



Adrián Corrochano, Giuseppe D'Alessio, Alessandro Parente,
and Soledad Le Clainche

1 Introduction

Halting climate change is one of the greatest challenges facing the modern world. Improving the efficiency in combustion systems in aircraft, i.e., reducing the amount of fuel consumption, reducing the polluting fumes released from turbulent combustion or replacing fossil fuels by synthetic fuels are some of the strategies that could be addressed to reduce air pollution.

Modelling reacting flows is a highly complicated task, due to (i) the high complexity of the underlying physical problem and (ii) the need of using large computational resources, generally required to solve realistic problems. The alternative is to develop and use reduced order models (ROMs). ROMs model the main flow dynamics with relatively high accuracy at a reduced computational cost (in terms of memory and computational time).

A novel strategy has been carried out in this chapter to study the main dynamics of an axisymmetric, time varying, non-premixed laminar co-flow flame. The method identifies the main patterns describing the flow physics. Such patterns can be used to reconstruct the original solution with high accuracy, resulting in a ROM that can be used to model the problem studied. For such aim, we use higher order dynamic mode decomposition (HODMD) (Le Clainche & Vega, 2017), a data-driven method, suitable for the analysis of complex flows (see more details in Le Clainche et al., 2020a, b). Invoking the previous knowledge and based on the robustness of the

A. Corrochano (✉) · S. Le Clainche

School of Aerospace Engineering, Universidad Politécnica de Madrid, Madrid, Spain

e-mail: adrian.corrochanoc@upm.es; soledad.leclainche@upm.es

G. D'Alessio · A. Parente

École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Université Libre de Bruxelles, Bruxelles, Belgium

e-mail: Giuseppe.DAlessio@ulb.be; Alessandro.Parente@ulb.be

method presented in complex flows (i.e. multiscale, noisy, turbulent, et cetera), either numerically or experimentally (Vega & Le Clainche, 2020), a pioneering study of a combustion database has been carried out, showing that this algorithm is also suitable for this complex field of knowledge.

The chapter is organized as follows. In Sect. 2, the algorithm used in this work is introduced and the database where this method is going to be applied is presented in Sect. 3. The main results and the conclusions are explained Sects. 4 and 5, respectively.

2 Method

Higher order dynamic mode decomposition (HODMD) (Le Clainche & Vega, 2017) is a data-driven algorithm used to analyse complex nonlinear dynamical systems. This algorithm is an extension of the classical dynamic mode decomposition (DMD) (Schmid, 2010), which is suitable for the analysis of complex flows, such as turbulent flows (Le Clainche et al., 2020a, b), noisy experiments (Le Clainche et al., 2017, 2018), transitional flows (Le Clainche & Vega, 2017), etc., to develop ROMs (Le Clainche & Ferrer, 2018; Le Clainche et al., 2020a, b) or identify flow patterns and coherent structures (Le Clainche & Vega, 2018).

HODMD decomposes the original data v_k , collected at time instant t_k , which are the snapshots, as a Fourier expansion of M DMD modes u_m in the following way (in discrete form):

$$v_k \simeq \sum_{m=1}^M a_m u_m e^{(\delta_m + i\omega_m)t_k}, k = 1, \dots, K, \quad (1)$$

where a_m , δ_m , ω_m are their amplitudes, growth rates and frequencies, respectively. The temporal dimension can be referred to as K , which are the number of snapshots analysed by the method, and the number of modes M is called the spectral complexity. The number of DMD modes M depends on a tunable tolerance ε_1 , which can be adjusted to the desired accuracy to reconstruct the original data.

The algorithm can be summarised in two steps. Initially, the data are collected into a snapshot tensor $\mathbf{A}$. The first dimension of the tensor represents the variables studied. The next two or three dimensions are the spatial dimensions, depending on whether the problem is bi or tri dimensional. The last dimension will be the temporal dimension. The first step applies a higher order singular value decomposition (HOSVD) to the tensor, retaining the large size scales and reducing the noise and the spatial redundancies. A second tunable tolerance ε_2 will be used to define the HOSVD modes retained. In the second step, a DMD-like algorithm is applied to the previously processed data. This algorithm is called the DMD-d algorithm. Similarly to the sliding window process from the power spectral density (PSD) analysis, the DMD-d algorithm uses d time-delayed snapshots from the snapshot matrix. The

parameter d is tunable, as well as the tolerance ε_1 mentioned before, which will be applied now to define the number of DMD modes retained in the expansion Eq. 1.

Finally, the original data can be reconstructed as in Eq. (1), obtaining a new tensor $\mathbf{A}^{\text{DMD}}$. The error made in the calculations can be evaluated measuring the root mean square error, as follows

$$\text{RMSE} = \sqrt{\frac{|\mathbf{A} - \mathbf{A}^{\text{DMD}}|^2}{|\mathbf{A}|^2}}. \quad (2)$$

More details about the algorithm used in this chapter can be found in (Le Clainche et al., 2017). The Matlab codes are presented in (Vega & Le Clainche, 2020).

HODMD will be applied to identify the main frequencies leading the non-linear solution of a numerical simulation modelling a turbulent combustion. The high complexity of the data makes HODMD a suitable algorithm to analyse the data.

3 Numerical Simulation

Numerical simulations have been carried out using the numerical open source solver *laminarSMOKE*, a CFD solver based on *OpenFOAM* and conceived to simulate laminar flows by means of a detailed kinetic mechanism (Cuoci et al., 2013). This solver uses finite volumes as spatial discretization. Simulations model an axisymmetric, time varying, non-premixed laminar co-flow flame: the fuel consists of 65% of methane and 35% of nitrogen on molar basis, while the oxidizer is regular air. Both streams are fed at ambient temperature and atmospheric pressure. The transient behaviour is induced by a sinusoidal perturbation to the parabolic velocity profile of the fuel stream, with frequency f and amplitude A . Thus, the fuel velocity profile is as follows:

$$v(r, t) = v_{\max} * (1 - r^2/R^2) * (1 + A * \sin(2 * \pi * f * t)), \quad (3)$$

where r is the radial coordinate, R is the internal radius, t is the time, $v_{\max}$ is the velocity equal to 70 cm/s, $f = 10$ Hz and $A = 0.25$. The co-flow air is instead injected at constant velocity $v = 35$ cm/s. The current simulation arrives to ~ 0.12 s, and the time step is $\Delta t = 0.0001$ s. For the HODMD analysis, 1 of each 10 snapshots have been used, so the Δt between snapshots will be 0.001 s. A thorough description of the numerical setup can be found in (D'Alessio et al., 2020).

The kinetic mechanism used for the simulation is the *POLIMI_C1C3_HT_1412* (Ranzi et al., 2012), it accounts for 82 species and 1698 reactions.

4 Results

In this section, the algorithm mentioned before will be used to analyse the database corresponding to the numerical simulation explained above. The data base is a fourth order tensor, formed by 83 variables (temperature and 82 different species); two spatial dimensions composed by a 150×150 grid; and 117 snapshots equidistant in time. The order of magnitude of the variables ranges from 10^2 (temperature) to 10^{-11} (RALD3G, the smallest radical computed). To avoid that the small variables could be removed, the mean has been removed and all variables have been scaled with its range, as in Parente & Sutherland, 2013. This scaling method performs better for most of the main species, which contain the main flow dynamics.

The tolerances used in this case are $\varepsilon_1 = \varepsilon_2 = 10^{-3}$. This tolerance gives a good reconstruction in all the cases studied. Parameter d is calibrated performing several analyses using $d = 10, 12$ and 15 . Figure 1 shows the spectrum for all the cases studied. It shows a clearly periodic solution, with a dominant frequency (with highest amplitude) of $\omega_p = 60$ and its harmonics, which are the modes with lower amplitude and frequency $\omega_n = n * \omega_p$, being $n \in \mathbb{N} > 1$.

The case with $d = 12$ is selected as a representative case for studying the solution of this problem.

Figure 2 shows the error made in the reconstruction using the previous DMD modes in the DMD expansion Eq. (1). As expected, the best reconstructed variables are the main ones, i.e. the air components (N_2 and O_2), the fuel components (CH_4), the main products of the combustion (CO_2 and H_2O) and the temperature. These variables have a reconstruction error lower than 1% except the methane that has an error of 1.2%. Moreover, the nitrogen has a reconstruction error below 10^{-3} . On the other hand, the worst reconstructed variables are some of the smallest radicals. Their maximum value ranges from 10^{-5} to 10^{-11} . RALD3B presents the largest reconstruction error with 62%, while the order of magnitude of its maximum value is 10^{-9} .

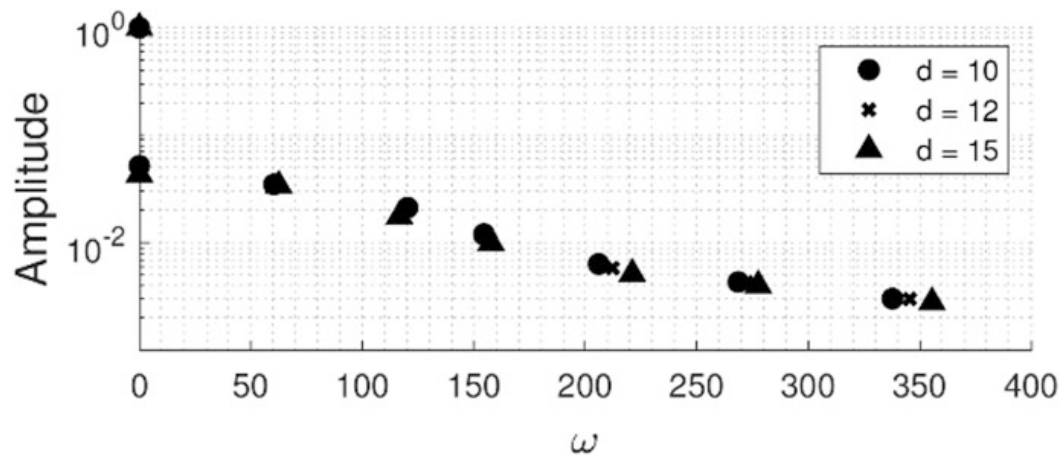


Fig. 1 Frequencies vs. amplitudes representing the DMD modes in the combustion database

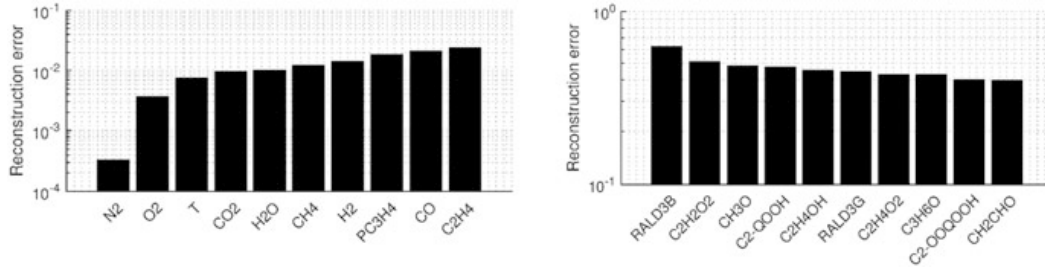


Fig. 2 Reconstruction error for the 10 best reconstructed variables (left) and the 10 worst reconstructed variables (right) using the DMD modes presented in Fig. 1 and the DMD expansion Eq. (1)

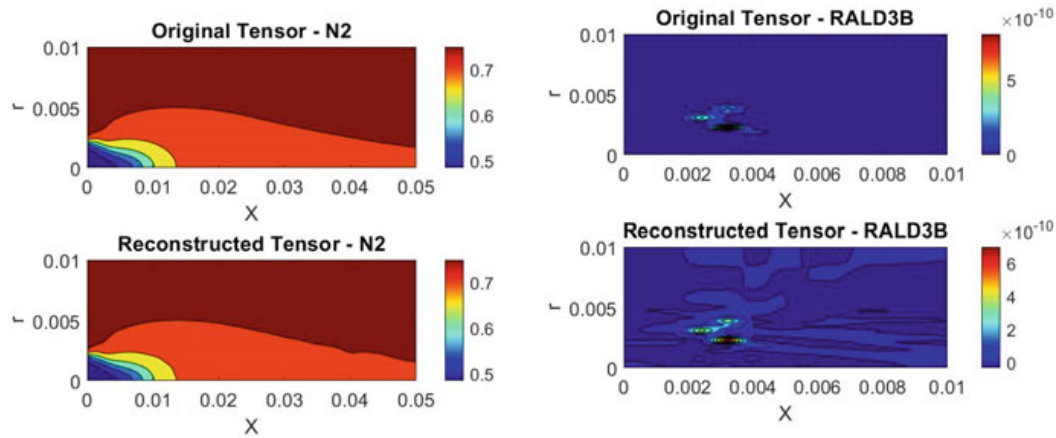


Fig. 3 Representative snapshot of the best reconstructed variable (N2). Right: Representative snapshot of the worst reconstructed variable (RALD3B)

Finally, a representative snapshot has been plotted with the best reconstructed variable and the worst one in Fig. 3. Nitrogen is very accurately reconstructed, as it shows the reconstruction error (below 10^{-3}). On the other hand, the RALD3B variable shows that, despite having high reconstruction error, the main structures can be identified, as seen in the reconstructed snapshot. This result suggests that the high error value could be connected to the fact that in most of the domain, the variable has a constant zero value. While the original data shows an order of magnitude of 10^{-23} on the zero part, on the reconstructed tensor reveals an order of magnitude of 10^{-12} on the zero part. This is between two and three orders of magnitude below the main structures. Therefore, with this method, the main structures of radical variables can be captured, although the whole snapshot may exhibit a large reconstruction error.

5 Conclusion

This work studies the main flow dynamics and the possibility of using HODMD as a reduced order model, to reconstruct transient data in turbulent combustion. The main dynamics of the flow have been found to be periodic, with a principal frequency of $\omega_p = 60$. The main species and the temperature are the variables better approximated by the method, while the worst reconstructed variables are represented by some radical species. Nevertheless, the method is also able to identify the main patterns in these poorly reconstructed variables, suggesting once more that HODMD is suitable to model turbulent combustion flows.

Acknowledgments A.C. and S.L.C. acknowledge the grant PID2020-114173RB-I00 funded by MCIN/AEI/ 10.13039/501100011033.

References

- Cuoci, A., Frassoldati, A., Faravelli, T., & Ranzi, E. (2013). Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method. *Energy and Fuels*, 27(12), 7730–7753.
- D'Alessio, G., Parente, A., Stagni, A., & Cuoci, A. (2020). Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. *Combustion and Flame*, 211, 68–82.
- Le Clainche, S., & Ferrer, E. (2018). A reduced order model to predict transient flows around straight bladed vertical Axis wind turbines. *Energies*, 11, 566.
- Le Clainche, S., & Vega, J. M. (2017). Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16, 882–925.
- Le Clainche, S., & Vega, J. M. (2018). Analyzing nonlinear dynamics via data-driven dynamic mode decomposition-like methods. *Complexity*, 2018, 6920783.
- Le Clainche, S., Vega, J. M., & Soria, J. (2017). Higher order dynamic mode decomposition for noisy experimental data: Flow structures on a zero-net-mass-flux jet. *Experimental Thermal and Fluid Science*, 88, 336–353.
- Le Clainche, S., Moreno, R., Taylor, P., & Vega, J. M. (2018). New robust method to study flight flutter testing. *Journal of Aircraft*, 56, 1–8.
- Le Clainche, S., Izbassarov, D., Rosti, M., Brandt, L., & Tammisola, O. (2020a). Coherent structures in the turbulent channel flow of an elastoviscoplastic fluid. *Journal of Fluid Mechanics*, 888.
- Le Clainche, S., Rosti, M., & Brandt, L. (2020b). Flow structures and shear-stress predictions in the turbulent channel flow over an anisotropic porous wall. *Journal of Physics: Conference Series*, 1522, 012016.
- Parente, A., & Sutherland, J. C. (2013). Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combustion and Flame*, 160(2), 340–350.
- Ranzi, E., Frassoldati, A., Grana, R., Cuoci, A., Faravelli, T., Kelley, A. P., & Law, C. K. (2012). Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels. *Progress in Energy and Combustion Science*, 38(4), 468–501.
- Schmid, P. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656, 5–28.
- Vega, J. M., & Le Clainche, S. (2020). *Higher order dynamic mode decomposition and its applications*. Elsevier.

3.2 HODMD applied to a laminar flame

3.2.1 Publication information

Title: Higher order dynamic mode decomposition to model reacting flows

DOI: <https://doi.org/10.1016/j.ijmecsci.2023.108219>

Journal: International Journal of Mechanical Sciences

Date: 1 July 2023

Publisher: Elsevier

Copyright: © 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license



Higher order dynamic mode decomposition to model reacting flows

Adrián Corrochano^{a,*}, Giuseppe D'Alessio^{b,c}, Alessandro Parente^b, Soledad Le Clainche^a

^a School of Aerospace Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain

^b Université Libre de Bruxelles, École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Bruxelles, Belgium

^c CRECK Modeling Lab, Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

ARTICLE INFO

Keywords:

Higher order dynamic mode decomposition
Reduced order models
Principal component analysis
Varimax rotation
Reacting flows

ABSTRACT

This work presents a new application of higher order dynamic mode decomposition (HODMD) for the analysis of reactive flows. Due to the high complexity of the data analysed, consisting of more than 80 variables (i.e., temperature and chemically reacting species) a new extension of HODMD has been developed combining the multi-dimensional HODMD algorithm with classical preprocessing techniques generally used in machine learning analyses, such as principal component analysis (PCA). This new methodology has proved to be suitable to identify the main patterns driving the main dynamics of the flow, as well as to develop reduced order models grounded in physical principles. The new algorithm was tested by means of a database obtained from a Computational Fluid Dynamics simulation of an axi-symmetric time varying non-premixed co-flow nitrogen-diluted methane flame, carried out by means of a detailed kinetic mechanism. Different dynamics were identified, and they were associated to the different variables of the reactive flow analysed. Also, this algorithm has been coupled with an additional feature selection step carried out via PCA and varimax rotation. The results that were obtained by coupling PCA with varimax rotation show the outstanding capabilities of this novel methodology to compress original databases as function of the different preprocessing techniques that are used.

1. Introduction

Fighting climate change is one of the major challenges facing our society today. Reducing the pollutants produced from the combustion of fossil fuels [1], improving the energy efficiency in industrial devices [2], or producing new smart energy carriers are just a few examples of the multiple research topics in this field. A strong effort is currently being carried out by the research community to study in detail and to understand the flow physics of reacting flows, to improve the performance and the efficiency of the combustion process, as well as to find new eco-friendly alternatives to the current use of fossil fuels.

However, the complexity associated to combustion, both from a physical and chemical point of view, entails high costs from both an experimental and a computational perspective. As in other challenging complex systems, the development of models often requires considerable efforts [3]. On the one hand, the implementation of experimental measurements of combustion systems is complicated due to the small time scales, the complex reactions driving their dynamics, and also because of the high cost associated with the use of specialised and sophisticated infrastructures. On the other hand, the large number of species involved in combustion processes, the range of scales, and the turbulence-chemistry interaction, require 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 [4]. Furthermore, the computational costs in terms of CPU time and memory associated with the solution of industrially relevant problems limit the advancement in the field. In light of what is stated above, a relevant alternative is represented by reduced order models (ROMs). The application of ROMs in reacting flows has progressively increased in recent years, as they offer the opportunity to simplify the complex nature of a physical system, as well as to achieve the formulation of high fidelity lower-dimensional models representing the physics with a low level of error [5–7]. Using ROMs, it is possible to model the main flow dynamics with a reduced computational cost, which can be useful for applications based on feature extraction and design optimisation [8].

In recent years, different types of ROMs have been developed to model reacting flows. To name a few, Torrez et al. [9] implemented a code capable of accurately modelling the turbulent reacting flow in a scramjet combustor. Huang et al. [10] created a model based on components, where each component was formed by a part of the domain modelled through a simulation or a ROM, to represent the whole domain. Other works show the good capabilities of Proper Orthogonal

* Corresponding author.

E-mail address: adrian.corrochanoc@upm.es (A. Corrochano).

Decomposition (POD) [11] to model single-injector combustion processes [12,13]. However, accurately predicting sharp gradients in the temperature and mass fractions of species is still a challenge when using POD-based models [14].

Principal Component Analysis (PCA) [5,15–17] has also found widespread applications to model flow combustion. More specifically, PCA is a statistical technique whose aim is to find lower-dimensional manifolds by identifying correlations between input variables, which are then combined linearly to eventually reduce the dimension of the problem [18,19]. The application of PCA, as well as its local formulation, has proved to be suitable for dimensionality reduction and feature extraction [5,6,17,20,21], clustering [22,23], as well as data analysis and feature selection [24–29].

More recently, dynamic mode decomposition (DMD) [30] has also been introduced as a tool to extract flow patterns and develop ROMs in reacting flows. The main advantage of this technique lies in its ability to identify the main frequencies and flow features (i.e., the DMD modes) leading the dynamics of a given physical system [31]. These modes can eventually be linked to flow instabilities that reflect the underlying physical mechanisms driving the flow dynamics [32]. Therefore, the results obtained when using DMD and PCA methodologies can be considered complementary in terms of reduction of data dimensionality and identification of the physical principles of the flow. Some examples of DMD for the analysis of reacting flows published in the last decade are the following. Richecoeur et al. [33] applied DMD to study the main dynamics in turbulent combustion, studying in detail the performance of this algorithm when applied to the analysis of experimental databases. Gunarate and Roy [34] employed this algorithm to analyse experiments of reacting flows behind a single bluff body, observing symmetric vortex shedding and asymmetric von Karman vortices, depending on their equivalence ratio. In addition, several authors used DMD to identify the main instabilities leading changes in the flow. For instance, in the context of data analysis for experimental databases in reacting and non-reacting fluid dynamics, Quinlan et al. [35] 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. [36] employed DMD to study self-excited longitudinal combustion instabilities in laboratory-scaled single-element gas turbine and rocket combustors. Souvick et al. [37] employed DMD in experimental databases to identify the main patterns connected to flow stabilities in a swirl-stabilised dump combustor. Motheau et al. [38] employed DMD for the analysis of a LES database, with the aim of developing a reduced order model to show that one of the mechanisms triggering combustion instabilities involves convected and acoustic entropy waves. Additional applications of DMD to identify patterns and flow instabilities in reacting flows include the work by Abou-Taouk et al. [39], who studied a V-flame in an afterburner-type configuration; the work by Ghani et al. [40], who identified the main patterns in an industrial gas turbine combustion chamber; and the work by Grenga et al. [41,42], who proposed a more efficient implementation of DMD, reducing the memory consumption, which was tested employing a dataset obtained from a Direct Numerical Simulation (DNS) of a turbulent reacting jet.

As shown in Ref. [43], the application of DMD is limited to very specific conditions of the database analysed, based on the physics of the problem. Consequently, when complex dynamical systems or noisy experimental databases are taken into account, DMD may encounter some difficulties in identifying the main patterns describing the flow dynamics. Thus, Higher Order Dynamic Mode Decomposition (HODMD) [44] was introduced as an alternative to overcome such limitations. This algorithm represents, in fact, a more robust extension of the standard DMD, and it has been proven to be suitable for the analysis of non-reactive complex flows, as well as noisy experimental data of complex dynamical systems, tested to solve several industrial applications in fluid dynamics [45–48]. The applications of HODMD to complex non-reacting dynamical systems that can be found in literature

show that this algorithm could also represent a valid tool for the analysis of any type of reacting flows.

The main goal of this work is to provide a new tool, highly robust and accurate, which is capable of modelling reacting flows with a high level of accuracy and, at the same time, which is capable to reveal the main physics of the flow. This last aspect is very important to properly develop ROMs based on physical principles, and this knowledge could also be used for future flow control applications to improve the system efficiency. Therefore, the current work introduces for the first time, to the best of the authors' knowledge, the application of HODMD to identify the main dynamics in a combustion database. A new extension of the HODMD algorithm is presented in the article, specifically developed to identify the main dynamics in reactive flows. In the database analysed, the unsteady behaviour was obtained by imposing a sinusoidal perturbation in the parabolic velocity profile of the fuel stream. From a kinetic point of view, the CFD simulation was carried out utilising a very detailed mechanism which takes into account a large number of species and thousands of reactions. The complexity of the considered chemical system is due not only to the high dimensionality of the thermochemical space, but also to the flame liftoff, and to the non-negligible contribution of the differential diffusion (as generally opposed to any turbulent flame). Moreover, the intrinsic dimensionality of laminar reacting systems is higher with respect to turbulent flames because of the absence of any major interaction between the fluid dynamics and the chemistry of the system. The proposed method is capable of reducing the data dimensionality as well as identifying the main frequencies and patterns driving the flow dynamics, which are both important requirements to develop a ROM. 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 types of flows.

The article is organised as follows. Section 2 introduces the methodology used to develop this work. The description of the database that is analysed is introduced in Section 3. Finally, the main results and conclusions are presented in Sections 4 and 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: HODMD [44] and PCA [15], both combined with preprocessing strategies generally used in machine learning. In particular, PCA will also be coupled with the Varimax rotation method [49] to identify a subset of main variables from a database containing a large number of variables.

2.1. Higher Order Dynamic Mode Decomposition (HODMD)

HODMD [44] is a data-driven method introduced as a robust extension of DMD [30] for the analysis of complex non-linear dynamical systems. Similarly to DMD, HODMD decomposes spatio-temporal data $\mathbf{v}(x, y, z, t_k)$, collected at time instant t_k (for convenience expressed as $\mathbf{v}_k$), as an expansion of M Fourier-like modes. For three-dimensional databases, the decomposition reads as

$$\mathbf{v}(x, y, z, t_k) \simeq \mathbf{v}_k^{\text{approx}} = \sum_{m=1}^M a_m \mathbf{u}_m(x, y) e^{(\delta_m + i\omega_m)t_k}, \quad k = 1, \dots, K, \quad (1)$$

where $\mathbf{u}_m$ are the DMD modes and a_m , ω_m , δ_m are their associated amplitudes, frequencies and growth rates. From a practical point of view, it is necessary to organise a database as a group of vectors equidistant in time, with time interval Δt , in the following snapshot matrix

$$\mathbf{V}_1^K = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_{K-1}, \mathbf{v}_K]. \quad (2)$$

with dimension $J \times K$, with $J = N_{\text{variables}} \times N_x \times N_y \times N_z$, where $N_{\text{variables}}$ are the number of variables considered in the database, and

N_x , N_y and N_z are the number of grid points along the streamwise, normal and spanwise 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

$$v_{k+d} = \mathbf{R}_1 v_k + \mathbf{R}_2 v_{k+1} + \dots + \mathbf{R}_d v_{k+d-1}, \quad k = 1, \dots, K-d, \quad (3)$$

where $\mathbf{R}_i$ (for $i = 1, \dots, d$) 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 Refs. [44,50], and the Matlab codes can be found in Ref. [51].

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

$$\text{RRMS Error} = \sqrt{\frac{\sum_{k=1}^K \|v_k^{\text{approx}} - v_k\|_2^2}{\sum_{k=1}^K \|v_k\|_2^2}}. \quad (4)$$

This metric has been used to measure the uncertainty of the reconstruction with HODMD. There are other techniques, based on Uncertainty Quantification methods [52–54], to measure the uncertainty of the reconstruction, but we have followed the metrics used in the multiple publications available in the literature where HODMD was applied to solve a wide range of academic and industrial problems [45–47,55,56].

2.2. HODMD in combustion flows

HODMD is a widely used algorithm in the fluid mechanics community, and its robustness and accuracy has been tested in a wide range of applications, including the analysis of complex flows (i.e., noisy experiments [47,50], transitional flows [57], turbulent flows [48,56]) to create reduced order models (ROMs) [55,58,59], to extract patterns and study flow physics [43,60–62], or to calculate areas of structural sensitivity for applications in flow control [63]. The limits of applicability for this algorithm have also been successfully tested in other fields such as electrocardiograms for disease detection [64]. However, in all the studies carried out in the literature, the analysis was limited to databases including the three components of the velocity and, in some cases, the pressure field. In combustion systems, it is common to deal with a larger number of variables with different magnitudes. Hence, it is necessary to carefully preprocessing the data before starting the analysis, with the aim to extract the maximum quantity of relevant information from the problem under study. To this end, this article introduces a three-step method based on HODMD, which summarises 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: Centring 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 preprocess them before applying any data-driven algorithm. *Centring* consists of subtracting the mean value of 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

mentioned preprocessing operations are accomplished as follows:

$$\bar{v}_j = \frac{v_j - \bar{v}_j}{c_j}, \quad (5)$$

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, σ_j , as the scaling factor.
- Range scaling: Uses the difference between the maximum and minimum values of each variable as the scaling factor.

As observed by Parente and Sutherland [15], some scaling methods may emphasise the relevance of some variables as function of their distribution. Auto scaling gives similar importance to all the different variables, while range scaling performs better for most of the major species. However, to the authors' knowledge, there is no quantitative way to determine which scaling method is better, as it depends on the variables that the user may want to emphasise, thus it has to be assessed case by case.

Step 2: Dimension reduction via HOSVD.

A three-dimensional database can be reorganised into a fifth-order tensor $\mathbf{A}_{i_1 i_2 i_3 i_4 k}$ (with indexes associated with each one of the components, for $i_1 = 1, \dots, N_{\text{variables}}$, $i_2 = 1, \dots, N_x$, $i_3 = 1, \dots, N_y$, $i_4 = 1, \dots, N_z$ and $k = 1, \dots, K$), in which each of the five dimensions corresponds to the variables of the dataset, the three spatial components (namely, streamwise, x , normal, y , and spanwise, z , components) and the temporal component, t , respectively. The high order singular value decomposition (HOSVD) method [65] can then be used to reduce the data dimensionality in a highly efficient way. This algorithm applies standard SVD to each one of the fibres composing the tensor (see details and the Matlab code to perform this analysis in Ref. [51]), resulting in an expansion of modes as

$$\mathbf{A}_{i_1 i_2 i_3 k} = \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} \sum_{p_3=1}^{P_3} \sum_{p_4=1}^{P_4} \sum_{n=1}^N S_{p_1 p_2 p_3 p_4 n} U_{i_1 p_1} W_{i_2 p_2}^{(x)} W_{i_3 p_3}^{(y)} W_{i_4 p_4}^{(z)} T_{kn} \quad (6)$$

where $S_{p_1 p_2 p_3 p_4 n}$ is a fourth-order tensor (called the *core tensor*) and the columns of the matrices U , $W^{(x)}$, $W^{(y)}$, $W^{(z)}$, and T are known as the *SVD modes* of the decomposition (related to the number of components, called the *variable modes*, and the streamwise, normal and spanwise spatial dimension, called 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 , associated with the temporal matrix T is calculated based on the tolerance (tunable) ε as

$$\sigma_{N+1}/\sigma_1 \leq \varepsilon. \quad (7)$$

The SVD modes, associated with the remaining matrices, are calculated in a similar way for each one of the groups 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 databases, we also truncate this first dimension, evaluating the influence of this truncation on the reconstruction error as in Eq. (4). In this way, it is possible to identify the influence of the different variables on the dynamics of the system.

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

$$CF = \frac{N_e}{N_{er}} \quad (8)$$

where N_e is the total number of elements of the tensor (namely, $N_e = N_{variables} \times N_x \times N_y \times N_z \times K$), and N_{er} is the number of elements retained after reducing the data dimensionality with HOSVD. High values of the CF means high data compression, or strong dimensionality reductions.

Step 3: The DMD-d algorithm.

The DMD-d algorithm is applied to the matrix associated with the temporal modes T , known 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 that forms the original dataset. For consistency, to compare with the algorithm described in the literature [50], this matrix is called as the *reduced snapshot matrix*. The high order Koopman assumption Eq. (3) is applied to this snapshot matrix as

$$\hat{T}_{d+1}^K \simeq \hat{R}_1 \hat{T}_1^{K-d} + \hat{R}_2 \hat{T}_2^{K-d+1} + \dots + \hat{R}_d \hat{T}_d^{K-1}. \quad (9)$$

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

$$\tilde{T}_2^{K-d+1} = \tilde{R} \tilde{T}_1^{K-d}, \quad (10)$$

also written as

$$\begin{bmatrix} \hat{T}_2^{K-d+1} \\ \dots \\ \hat{T}_d^{K-1} \\ \hat{T}_{d+1}^K \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \\ \hat{R}_1 & \hat{R}_2 & \hat{R}_3 & \dots & \hat{R}_{d-1} & \hat{R}_d \end{bmatrix} \cdot \begin{bmatrix} \hat{T}_1^{K-d} \\ \hat{T}_2^{K-d+1} \\ \dots \\ \hat{T}_d^{K-1} \end{bmatrix}. \quad (11)$$

This new matrix, which organises 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 HODMD 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 in this matrix, to eliminate possible redundancies, via truncated SVD, again using the tolerance ε_1 in the new calculated singular values as

$$\tilde{\sigma}_{N'+1}/\tilde{\sigma}_1 < \varepsilon_1, \quad (12)$$

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

$$\tilde{T}_1^{K-d+1} \simeq \tilde{U} \tilde{\Sigma} \tilde{T}^\top \simeq \tilde{U} \tilde{T}_1^{K-d+1}, \quad (13)$$

with $\tilde{T}_1^{K-d+1} = \tilde{\Sigma} \tilde{T}^\top$, where $\tilde{U}^\top \tilde{U} = \tilde{T}^\top \tilde{T}$ are the $N' \times N'$ -unit matrices, and the diagonal of the matrix $\tilde{\Sigma}$ contains the singular values $\tilde{\sigma}_1, \dots, \tilde{\sigma}_{N'}$. To complete this step, Eq. (10) is premultiplied by $\tilde{U}^\top$, which invoking (13) results as

$$\bar{T}_2^{K-d+1} = \bar{R} \bar{T}_1^{K-d}. \quad (14)$$

The new $N' \times N'$ -Koopman matrix is related to $\hat{R}$ by $\bar{R} \simeq \tilde{U}^\top \hat{R} \tilde{U}$. However, we use the pseudoinverse in Eq. (14) (via

non-truncated SVD) to the matrix $\bar{T}_1^{K-d}$, instead of computing the previous expression, which yields

$$\bar{T}_1^{K-d} = \bar{U} \bar{\Lambda} \bar{V}^\top, \quad (15)$$

where $\bar{U} \bar{U}^\top = \bar{V}^\top \bar{V} = \mathbf{I}$ are the $N' \times N'$ -unit matrices and the diagonal of $\bar{\Lambda}$ contains the N' singular values. Substituting (15) into (14) and after multiplying by $\bar{V} \bar{\Lambda}^{-1} \bar{U}^\top$, it is possible to obtain the following equation

$$\bar{R} = \bar{T}_2^{K-d+1} \bar{V} \bar{\Lambda}^{-1} \bar{U}^\top. \quad (16)$$

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

$$\hat{t}_k \simeq \sum_{m=1}^M \hat{a}_m \hat{q}_m e^{\mu_m t_k}, \quad (17)$$

for $k = 1, \dots, K$. Retaining only the first N components of the vectors $\hat{q}_m = \bar{U} \bar{q}_m$, it is possible to approximate the reduced DMD modes $\hat{u}_m$ (dimension Nd). The damping rates δ_m and frequencies ω_m are obtained as

$$\delta_m + i\omega_m = \log(\mu_m)/\Delta t. \quad (18)$$

The amplitudes $\hat{a}_m$ are calculated by solving a least square fitting problem in Eq. (17) [66]. 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 ε_2 (tunable), it is possible to calculate the number of M DMD modes in Eq. (1), called as the *spectral complexity*, as follows

$$a_{M+1}/a_1 \leq \varepsilon_2. \quad (19)$$

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 of calculating the energy index I_m , defined as

$$I_m = \sum_{k=1}^K |a_m e^{(\delta_m + i\omega_m)t_k}| \|u_m\|_F^2 \times \Delta t, \quad (20)$$

where $\|\cdot\|_F^2$ is the square of the Frobenius norm. This index weights the modes taking into account also the damping or growth of the modes over time, which is reflected in δ_m . More details about this criterion, called *HODMD with criterion*, are presented in Ref. [67].

2.3. Principal Component Analysis (PCA) and variable selection

PCA [17] reduces the dimensionality of an input matrix $\mathbf{X}$, consisting of $(K \times N_x \times N_y \times N_z)$ rows representative for the statistical observations of a system (e.g., $N_x \times N_y \times N_z$ grid points for K snapshots of a CFD simulation) and $N_{variables}$ columns for the associated observable variables, by projecting such matrix onto a lower-dimensional basis consisting of q eigenvectors.

The PCA algorithm is associated with the covariance matrix $\mathbf{C}$ of the input dataset $\mathbf{X}$, which is defined as follows

$$\mathbf{C} = \frac{1}{(K \times N_x \times N_y \times N_z) - 1} \mathbf{X}^T \mathbf{X}. \quad (21)$$

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

$$\mathbf{C} = \mathbf{A} \mathbf{\Lambda} \mathbf{A}^T, \quad (22)$$

where the matrices $\mathbf{A}$ and $\mathbf{\Lambda}$ contain the associated eigenvectors (in columns) and eigenvalues (in the diagonal), respectively. The columns of $\mathbf{A}$ are the Principal Components (PCs) and they geometrically represent the orthogonal coordinates of the lower-dimensional manifold,

obtained as a linear combination of the original variables [18]. This matrix also represents the proper orthogonal decomposition (POD) modes, calculated by an SVD of the original dataset $\mathbf{X}$ (see details in [32]). Instead, the matrix $\mathbf{L}$ is a diagonal matrix containing the eigenvalues λ_i (with $i \in [1, N_{variables}]$) 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 λ_i , in fact, represents the amount of input data variance being explained by the associated eigenvector. Consequently, if the value of λ_i is large, it means that the associated PC plays an important role in the lower-dimensional representation of the input data, as it is representative of a large amount of information. On the other hand, if λ_i is small, it means that the contribution of the associated PC to the orthogonal space representation is negligible, and can eventually be considered as noise or spatial redundancies (or in turbulent flows, it can be related to small flow scales [48]). In light of what is stated above, it follows that the last PCs can be removed so that the p -dimensional input matrix can only be expressed by q orthogonal components (with $q < N_{variables}$) associated with the highest eigenvalues, thus considering a truncated eigenvector matrix $\mathbf{A}_q$. The optimal dimensionality for the PCA manifold can be established by assessing the amount of input data variance (t_q) being explained by the truncated q -dimensional basis [20], defined as the ratio between the cumulative sum of the first q eigenvalues and the sum of all eigenvalues

$$t_q = \frac{\sum_{i=1}^q \lambda_i}{\sum_{j=1}^{N_{variables}} \lambda_j}, \quad (23)$$

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

$$\mathbf{Z}_q = \mathbf{X}\mathbf{A}_q. \quad (24)$$

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

$$\mathbf{X}_r \approx \mathbf{Z}_q \mathbf{A}_q^T. \quad (25)$$

The PCA algorithm discussed 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 features that define the state space and can subsequently select the best reduced subset from the original group of input variables [25,26,68]. In particular, according to Jolliffe [68], 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.

However, if high-dimensional problems are considered, the weight 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 [69]. Thus, PCA can be coupled with a rotation technique to increase the physical interpretability of eigenvectors, so that rotated PCs are preferentially aligned with one variable, as shown in Fig. 1.

Different techniques are available in literature to accomplish this operation, and can be grouped into two classes: orthogonal and oblique rotations. The main difference is that with the first class of rotations, 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 et al. [6], the Varimax rotation introduced by

Kaiser [49] can be considered to be 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 maximisation of the sum of variances of the squared coefficients within each eigenvector, i.e., the maximisation of the following quantity

$$\mathbf{V} = \frac{N_{variables} \sum_{i=1}^{N_{variables}} (b_{ij}^2)^2 - \gamma \sum_{i=1}^{N_{variables}} (b_{ij}^2)^2}{N_{variables}^2}, \quad (26)$$

where b is the PC loading, $\gamma = 1$, $i \in [1, N_{variables}]$, and $j = [1, q]$ with q being the dimensionality of the PCA manifold [6]. When a non-unitary γ 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 ($\mathbf{A}_q^r$) is obtained that can be expressed in terms of original PCs when a rotation matrix $\mathbf{T}$ is defined as in Ref. [6]

$$\mathbf{A}_q^r = \mathbf{T}\mathbf{A}_q, \quad (27)$$

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

$$\mathbf{Z}_q^r = \mathbf{X}\mathbf{T}\mathbf{A}_q. \quad (28)$$

In the current work, a rotated basis as in Eq. (27) is utilised 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 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

This section describes the numerical simulations performed to generate the database of a reactive flow that will be analysed 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 oxidiser is injected into the domain (whose dimensions are 54 mm in the radial direction and 120 mm in the axial direction, discretised by means of a Cartesian mesh whose number of cells is 25,000 selected after a mesh sensitivity analysis) adopting a constant velocity of 35 cm/s, while the fuel velocity varies in space and time according to the following sinusoidal perturbation:

$$v(r, t) = v_{max} \left(1 - \frac{r^2}{R^2} \right) [1 + A \sin(2\pi f t)], \quad (29)$$

with R representing the internal radius of the nozzle, r is the radial coordinate, t is the time, v_{max} is the maximum velocity ($v_{max} = 70$ cm/s) and f and A are the frequency ($f = 10$ Hz) and the amplitude ($A = 0.25$) of the perturbation, respectively. The internal diameter of the fuel nozzle is 4 mm with a thickness of 0.38 mm, while the annular region's diameter for the oxidiser injection is 50 mm. Regarding the kinetic mechanism, the detailed mechanism POLIMI_C1_C3_HT_1412 [70] (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. [71] to simulate laminar reacting flows by means of detailed kinetic mechanisms. The governing equations solved in LaminarSMOKE for the conservation of mass, momentum, species, and temperature are the following

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (30)$$

$$\frac{D(\rho \mathbf{v})}{Dt} = -\nabla p + \nabla \cdot \boldsymbol{\tau} + \rho \mathbf{g} \quad (31)$$

$$\frac{D(\rho Y_k)}{Dt} = -\nabla \cdot (-\rho \mathbf{v}_k Y_k) + \dot{\Omega}_k, \quad k \in [1, N_{species}] \quad (32)$$

$$\frac{D(\rho c_p T)}{Dt} = -\nabla \cdot q - \rho \nabla T \sum_{k=1}^{N_{species}} c_{p,k} Y_k \mathbf{v}_k - \sum_{k=1}^{N_{species}} h_k \dot{\Omega}_k \quad (33)$$

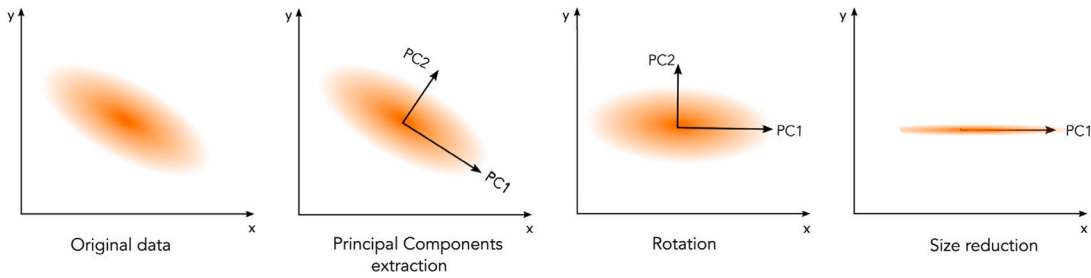


Fig. 1. Schematic of the different steps for a PCA reduction process of two variables when coupled with an orthogonal rotation of the eigenvectors.

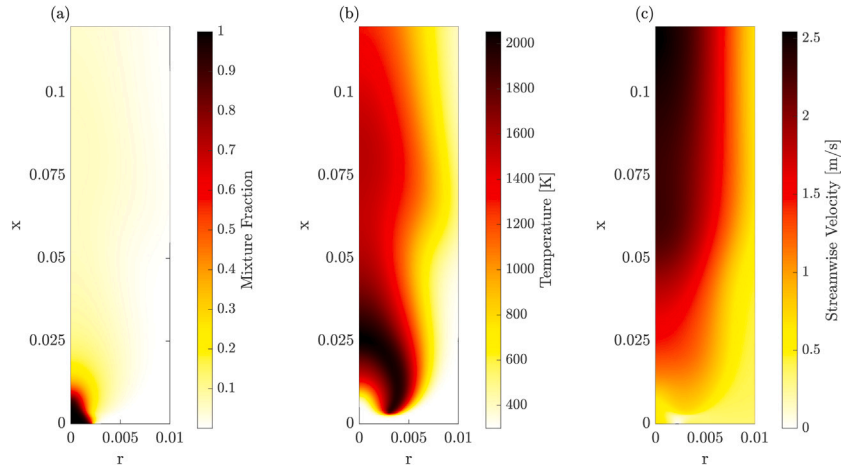


Fig. 2. Representative snapshot of the instantaneous field in the database analysed. Variables plotted: (a) mixture fraction, (b) temperature, and (c) streamwise velocity.

In Eqs. (30)–(33), the operator $\frac{D}{Dt}$ represents a substantial derivative, t is the time, ρ is the density of the mixture, $\mathbf{v}$ is the velocity of the mixture, p is the pressure, $\boldsymbol{\tau}$ is the fluid stress tensor, $\mathbf{g}$ is the gravity, Y_k is the mass fraction of a generic species k which is taken into account by the kinetic mechanism, $\mathbf{v}_k$ is the diffusion velocity corresponding to the species Y_k , $\dot{\Omega}_k$ is the formation rate associated to Y_k , c_p is the specific heat for the mixture (at constant pressure), T is the temperature, $\mathbf{q}$ is the heat flux vector, $c_{p,k}$ is the specific heat at constant pressure for the k th chemical species, h_k is the sum of the sensible and formation enthalpy associated with species Y_k [71]. The diffusion velocities for each species, $\mathbf{v}_k$, are computed considering both mass (Fick's diffusion) and thermal diffusion (Soret effect) as follows

$$\mathbf{v}_k = -\frac{D_k}{Y_k} \nabla Y_k - \frac{D_k \Theta_k}{X_k} \frac{1}{T} \nabla T, \quad (34)$$

where X_k is the mole fraction of the k th species, Θ_k is the thermal diffusion for the same chemical species, and D_k is the individual species diffusion coefficient computed from the binary diffusion coefficient Γ_{jk} defined as

$$D_k = \frac{1 - Y_k}{\sum_{j \neq k}^{N_{\text{species}}} \frac{X_j}{\Gamma_{jk}}} \quad (35)$$

To compute the density of the mixture, the ideal gas equation of state was employed. With regard to heat flux $\mathbf{q}$ in Eq. (33), conduction and radiation were taken into account according to

$$\mathbf{q} = \mathbf{q}_{\text{diff}} + \mathbf{q}_{\text{rad}} = -\lambda \nabla T - 4\sigma a_p (T^4 - T_{\text{env}}^4), \quad (36)$$

where λ is the thermal conductivity of the mixture, σ is the Stefan-Boltzmann constant, T_{env} is the environmental temperature, and a_p is the Planck mean absorption coefficient computed considering the partial pressures of water, carbon dioxide, carbon monoxide, and methane.

Additional information regarding this numerical simulations and the numerical setup can be found in Refs. [22,23].

The representative dataset for the numerical simulation considered for the HODMD analysis is a fourth-order tensor consisting of 83 variables (temperature and the 82 species), two spatial dimensions composed of a 150×150 grid points (along the streamwise and radial directions, respectively); and 117 snapshots equidistant in time, with a time step of $\Delta t = 10^{-3}$. Hence, the time interval covered by the dataset is $T = 117 \times \Delta t = 0.117$ s, approximately 1 cycle. The reduced number of snapshots forming the dataset could be considered a limitation to properly identify the main flow dynamics when algorithms such as POD and DMD are used. Nevertheless, if HODMD is employed, even a limited amount of data, as in this case, is enough to properly identify the main frequencies of the flow and to develop accurate ROMs.

In tensor form, the dimension of the original database is $83 \times 150 \times 150 \times 117$. The order of magnitude of the variables ranges from 10^3 (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, as stated on Section 2.2. The variables were scaled considering 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 centred with their mean and scaled with their standard deviation (auto scaling), respectively, because of the different units and ranges of the variables that were 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).

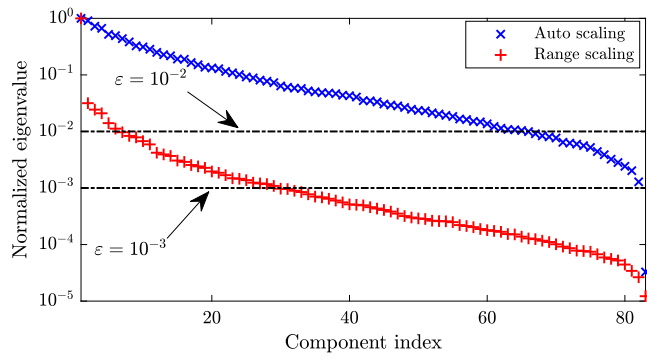


Fig. 3. Singular values corresponding to the SVD variable modes calculated for the two different scaling methods analysed: auto scaling (blue crosses) and range scaling (red pluses). The tolerances ε_1 set are highlighted as horizontal black lines.

4. HODMD modelling reactive flows

As explained in Section 2.2, the application of the HODMD algorithm to develop a ROM involves a three-step algorithm: the data are preprocessed using different scaling methods in step 1, after which a dimensionality reduction is carried out using HOSVD in step 2, and finally the main dynamics associated with the flow patterns are identified in step 3 using HODMD. In this section, a ROM is developed using this information.

In Section 4.1, two different cases (obtained in step 2) were compared for each type of preprocessing technique used in step 1: (i) using all components (retaining all the SVD variable modes) and (ii) reducing the number of components as explained in Eq. (7) (only retaining some SVD variable modes according to the tunable tolerance ε_1). It is important to note that, similarly to PCA, each one of the SVD variable modes is related to the number of variables in the database. Therefore, an SVD variable mode consists of the linear combination of the variables which are taken into account. The differences found among the various SVD variable modes lie in the values of the coefficients forming this linear combination of variables. In the following section, 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).

In Section 4.2, the performance of HODMD is evaluated for both scaling methods and for the cases where taking all components and a reduced set of them. For that, the RRMSE has been calculated for the reconstructed tensor, once HODMD has been applied. To study the influence of the scaling method, the main dynamics have been compared, showing the main frequencies and the leading DMD modes. Finally, in Section 4.3, the robustness of the method is tested applying HODMD to different sets of variables.

4.1. HOSVD and truncated variables

The dimensionality reduction carried out using HOSVD to develop a ROM is studied for two different values of tolerance ε_1 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 variable modes) based on their relevant contribution to the dynamical system. Two tolerances are selected to compare coarse with fine results, and these are: $\varepsilon_1 = 10^{-2}$ and 10^{-3} . Fig. 3 compares the number of components retained using the different tolerances and the auto scaling and range scaling methods. The value of these tolerances is selected on the basis of the magnitude of the singular values (SVD components). It can be observed that using larger tolerances,

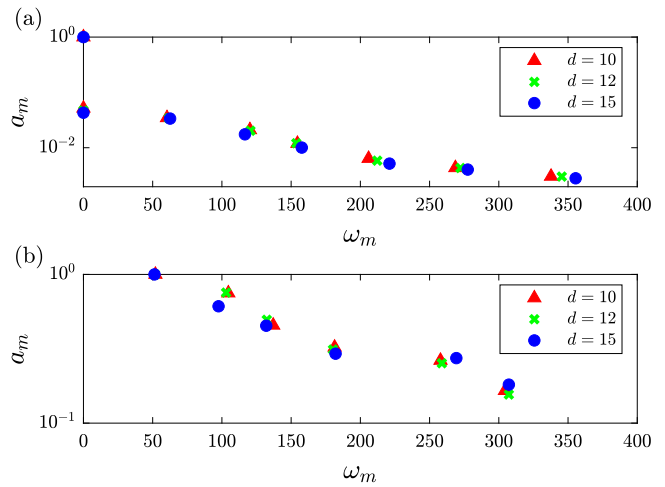


Fig. 4. Frequencies vs. amplitudes corresponding to the DMD modes calculated in the combustion database. Data scaled with: (a) Range scaling and (b) auto scaling methods. Three values of d have been used for the calibration: $d = 10$ (red triangles), 12 (green crosses) and 15 (blue circles).

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. Consequently, lowering this limit will not bring representative information related to the dynamics of the system.

Table 1 shows the RRMSE Eq. (4) calculated in the reconstruction of the database as in Eq. (6): as a function of the number of components (SVD variable 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 Table 2.

The RRMSE increases with the tolerance level, as expected. With coarser tolerances, the reduction carried out by the HOSVD algorithm is larger, resulting in a larger CF . As expected, the RRMSE is higher when the number of components is 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 components (All)).

Comparing the two variable scaling methodologies, the reconstruction of the tensor is slightly better when auto scaling is employed, as it provides for a uniform weight distribution to all 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. [15].

The main goal of the proposed model is to provide the best reconstruction of the original dataset using the smallest amount of information possible. For this reason, the range scaling appears to be particularly promising. In fact, the order of magnitude of the CF is one to two orders larger, and, at the same time, the increase in the RRMSE error is limited to the first decimal digit. More specifically, comparing the results obtained in both cases, the RRMSE when using range scaling is ~ 0.3 – 0.4% larger than auto scaling in all cases (the RRMSE is always maintained smaller than 3% and 1% for coarse and fine tolerances, respectively), while the CF is ~ 20 – 892% larger when using range than auto scaling, as presented in Table 3. This result suggests that when reduction is achieved using range scaling, it is possible to obtain an accurate ROM (with an RRMSE smaller than 3% and 1% for coarse and fine tolerances, respectively) with a high dimensionality reduction.

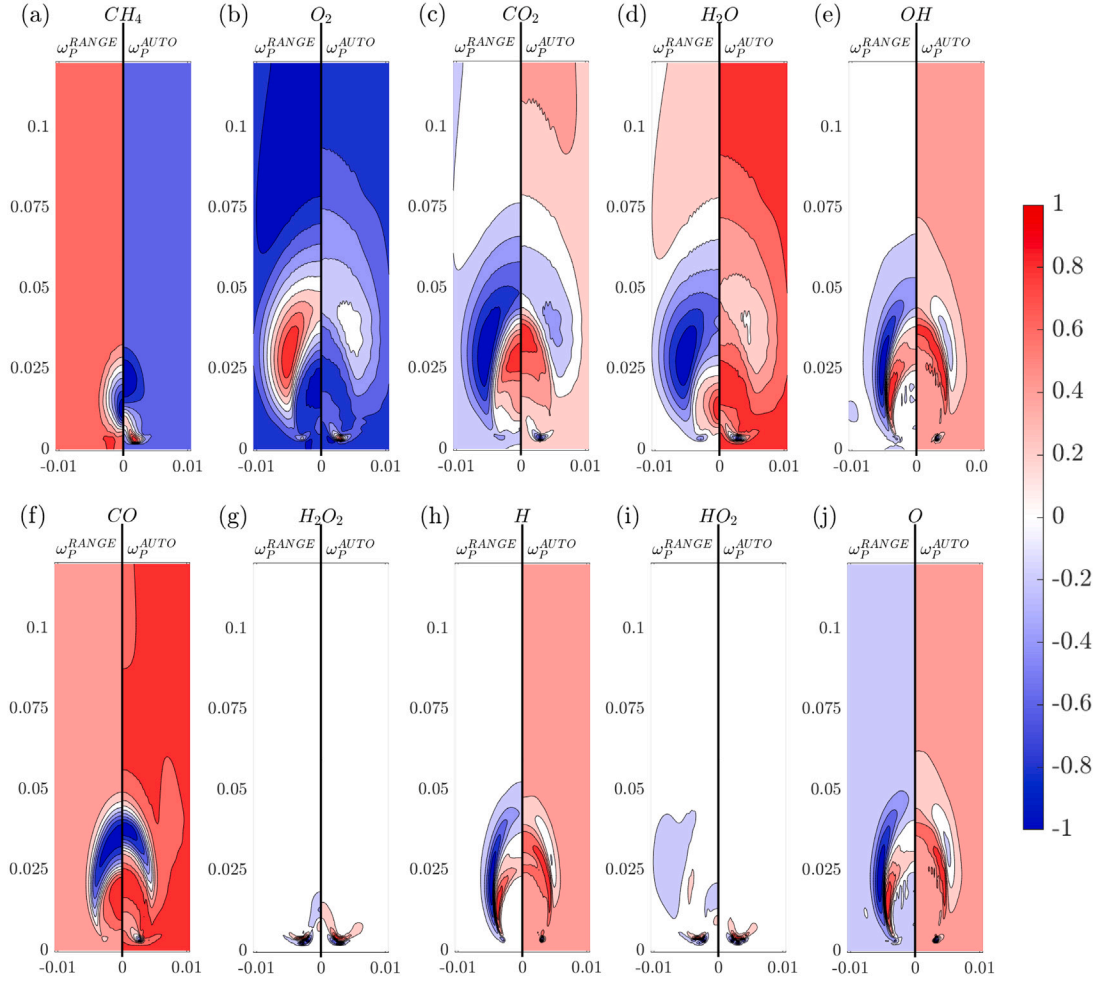


Fig. 5. Real part of the leading DMD mode for different variables. The right part represents the mode when using auto scaling ($\omega_p^{AUTO} \approx 50$) and the left part, the mode when using range scaling ($\omega_p^{RANGE} \approx 60$). Each mode have been normalised between values -1 and 1 .

Table 1

RRMSE Eq. (4) (in percentage) comparing the original database to the reconstruction of the reduced tensor, retaining P_2 , P_3 , and N spatial (in x and y) and temporal SVD modes, respectively. The case *All* considers all components ($P_1 = 83$), while the case *Red* considers the P_1 components (SVD variable modes) presented in the table. The database is preprocessed as in Eq. (5) using the range and auto scaling methods (denoted as range and auto, respectively).

	Range						Auto					
	$RRMSE_{All}$	$RRMSE_{Red}$	P_1	P_2	P_3	N	$RRMSE_{All}$	$RRMSE_{Red}$	P_1	P_2	P_3	N
$\epsilon = 10^{-2}$	2.51%	2.82%	6	7	5	5	2.03%	2.34%	65	40	32	42
$\epsilon = 10^{-3}$	0.34%	0.43%	29	26	19	18	0.18%	0.18%	82	59	48	67

Table 2

Compression factor (CF), calculated as in Eq. (8), of the original tensor after applying HOSVD in the databases presented in Table 1.

	Range		Auto	
	All	Red	All	Red
$\epsilon = 10^{-2}$	9.18×10^3	5.55×10^4	48.72	62.15
$\epsilon = 10^{-3}$	2.9×10^2	8.11×10^2	13.84	14.01

Finally, to compare the benefit of each preprocessing 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 $\epsilon_1 = 10^{-3}$ and $\epsilon_1 = 7 \times 10^{-2}$ 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

Table 3

Comparison of the CF obtained in the range and auto scaling methods as presented in Table 2. The rise in CF value 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.

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

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 a function of the previous choice. This reduced database will be used

Table 4

RRMSE of the reconstructed tensor (in percentage), calculated using HODMD. Range and auto scaling methods have been used.

	Range			Auto		
	$d = 10$	$d = 12$	$d = 15$	$d = 10$	$d = 12$	$d = 15$
$RRMSE_{All}$	0.58%	0.71%	1%	23.51%	24.71%	29.78%
$RRMSE_{Red}$	0.63%	0.75%	1.02%	24.94%	26.06%	30.82%

(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 of 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, calculated following the calibration tips presented in Ref. [51] as $K/10 = 117/10 \approx 11$, with K being the total number of snapshots. The tolerances used to develop the ROM have been selected in the previous section (to maintain a similar compression factor in both the range and auto scaling methods): $\epsilon_1 = 10^{-3}$ for range scaling and $\epsilon_1 = 7 \times 10^{-2}$ for auto scaling. Fig. 4 shows the frequencies as a function of the amplitudes of the DMD modes identified with different tolerances. The top figure shows the analysis for range scaling, and the bottom figure shows the analysis for auto scaling. The calculated spectrum is similar to the analysis carried out in the two databases. The spectrum shows a clear periodic solution with a principal mode, the one with the highest amplitude and frequency ω_p , and the harmonics, the modes with lower amplitude and frequency $\omega_n = n \times \omega_p$, being $n \in \mathbb{N} > 1$. 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 in the following. Similar spectrums (only one is shown to be concise) are obtained when all the components (All) are retained and when a reduced number of components (Red) is retained, as previously presented in Table 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 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. Table 4 presents the reconstruction 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. However, the error is much smaller when using range scaling (error $\sim 1\%$) than auto scaling (error $\sim 25\%$), 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 the leading frequencies identified in the analyses carried out in the database scaled with range and auto scaling methods ($\omega_p^{AUTO} \approx 50$ and $\omega_p^{RANGE} \approx 60$).

For the next steps, $d = 12$ has been selected as the reference to illustrate the reconstruction (although similar results are obtained with $d = 10$ and 15). The shape of the leading mode when using range and auto scaling is compared in Fig. 5 for different variables. As seen, when

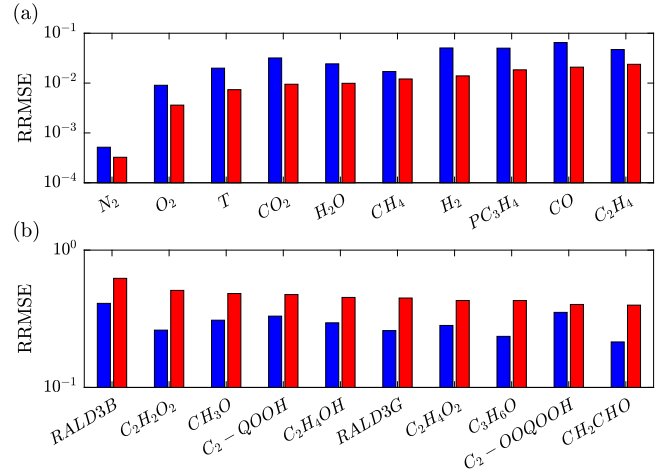


Fig. 6. Reconstruction error for (a) the 10 best reconstructed variables when applying HODMD with range scaling and (b) the 10 worst reconstructed variables when applying HODMD with range scaling. Two different scaling methods have been applied: Auto scaling (red bars) and Range scaling (blue bars).

range scaling is selected for the analysis, the highest intensity area of the mode focuses on the movement of the wake downstream the two jets, the main dynamic of the flow. However, when auto scaling is used, this high intensity region is more focused on the part of the domain where the oxidiser and the fuel merge, where the oxidation reaction begins. For the species that are concentrated in this region, H₂O₂ and HO₂, the shape of the mode is similar in both scaling methods. When analysing the imaginary part of the modes (not shown for the sake of brevity), the modes show a similar shape, highlighting the same intensity areas as their corresponding real part for both range and auto scaling.

Fig. 6 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 smaller error the main variables than the radicals. The same results are obtained with auto scaling, however, it provides a more uniform reconstruction (see details in Ref. [15]). The error for the main variables is larger than when using range scaling, while the inverse holds for radicals.

Figs. 7 and 8 show the reconstruction of a representative snapshot and the comparison with the original database for both range and auto scaling. The plots reported in the two aforementioned figures clearly show that the reconstruction error which is observed when HODMD is applied with the procedure described in the previous sections is negligible. This is true not only for the main species and reactants (i.e., fuel, oxidiser, main products such as water, carbon monoxide, and dioxide), but also for radicals (such as OH, O, etc.) whose dynamics is very fast as they are involved in propagation and branching reactions. Moreover, the displacement due to the flame lift-off is correctly predicted for all the considered species, despite it represents the most complex region from a physical point of view because of the high gradients which are observed. This result suggests the good properties of HODMD to develop ROMs based on physical principles. The HODMD modes and

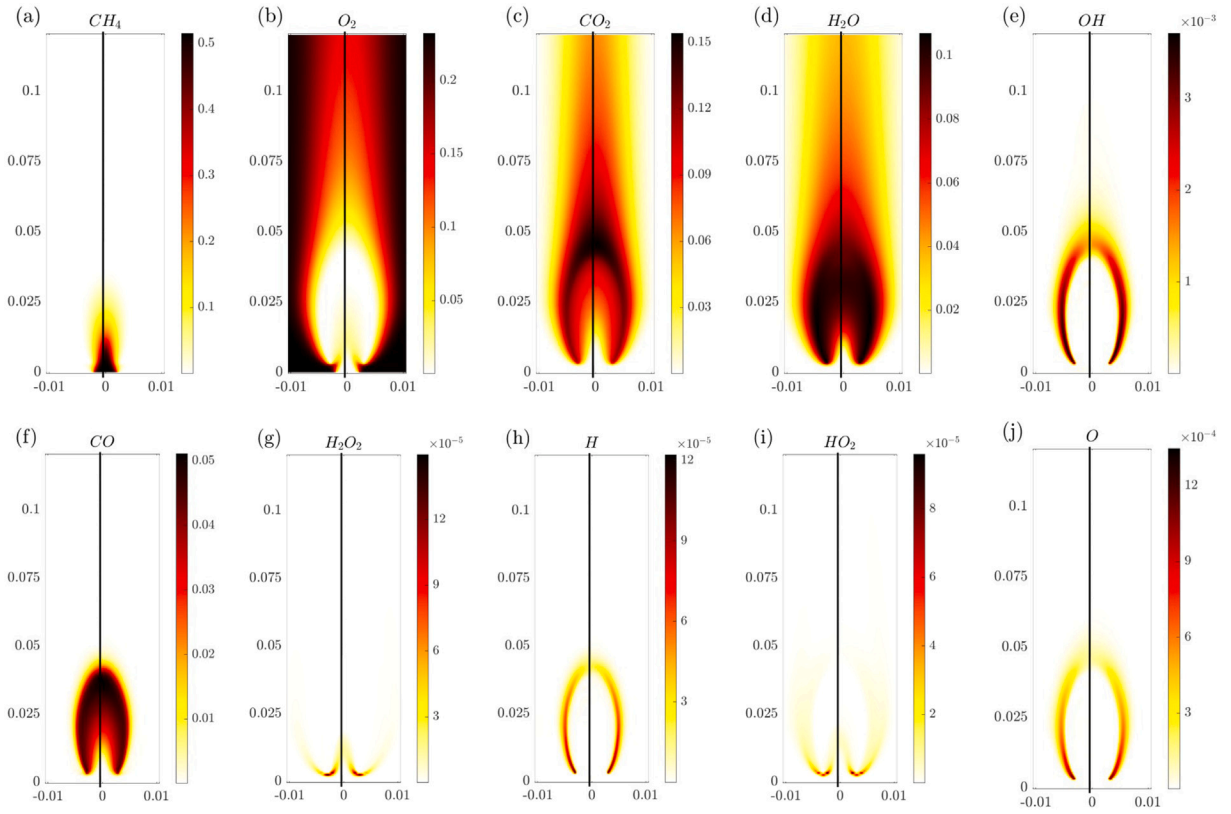


Fig. 7. Reconstruction of a representative snapshot for different variables using HODMD and Range scaling. The right side of each contour represents the original snapshot, while the left part the reconstruction obtained by means of HODMD.

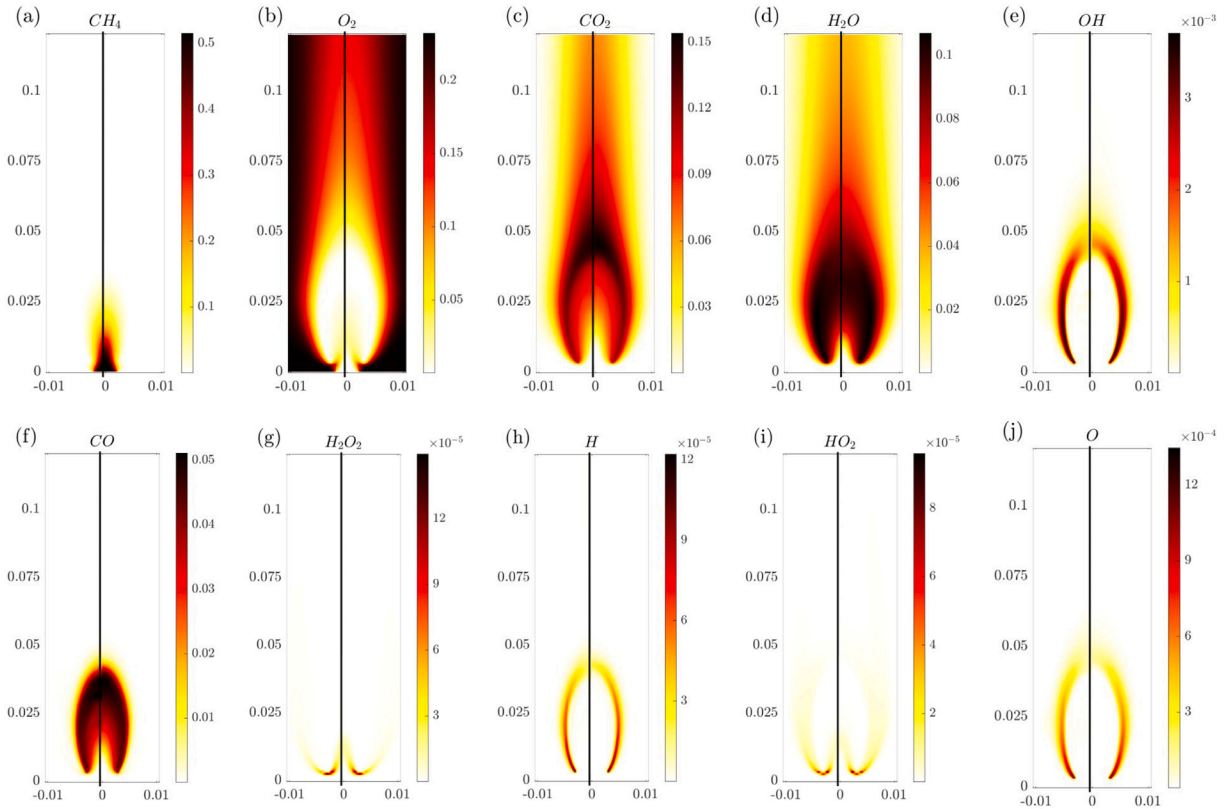


Fig. 8. Reconstruction of a representative snapshot for different variables using HODMD and Auto scaling. The right side of each contour represents the original snapshot, while the left part the reconstruction obtained by means of HODMD.

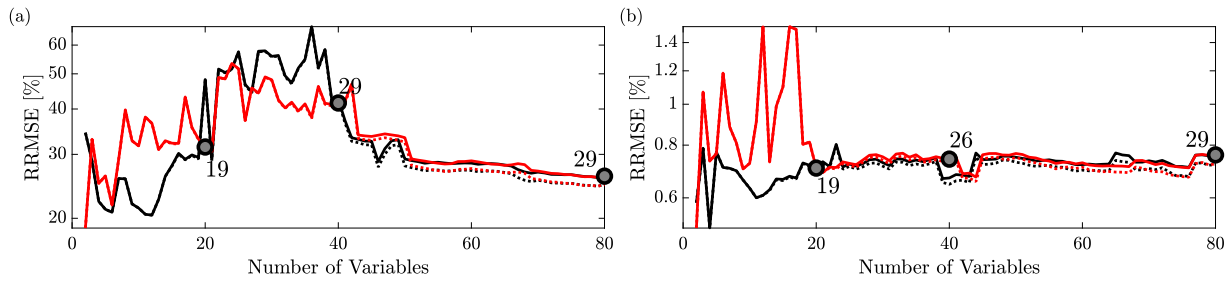


Fig. 9. Reconstruction error for different variable selection. The selected variables have been calculated using: B2 Variable selection (black line) and B2 with rotation Variable selection (red line). HODMD has been applied using all components (dashed line) and a reduced number of components (solid line). The number marked in the grey points show the number of components selected in HODMD for the specific number of variables presented in the x-label. Two different scaling methods have been applied for the analysis: (a) Auto scaling and (b) Range scaling.

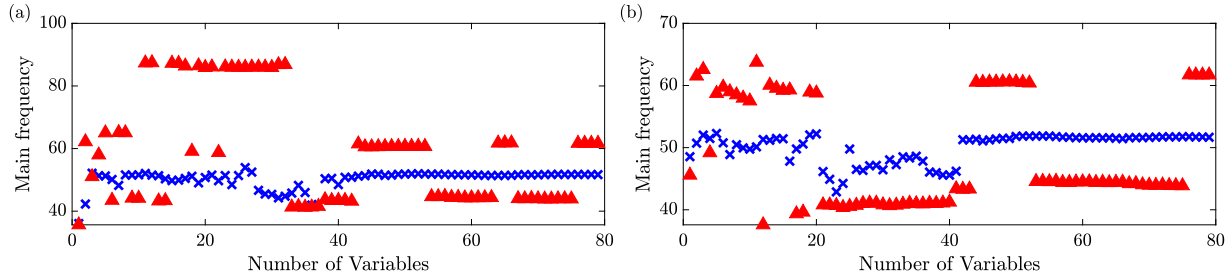


Fig. 10. Main frequency, calculated with Eq. (20), vs. number of variables analysed via HODMD. The selected variables have been calculated using: (a) B2 Variable selection and (b) B2 with rotation Variable selection. Two different scaling methods have been applied: Auto scaling (blue crosses) and Range scaling (red triangles).

the corresponding frequencies represent the main dynamics driving the flow; hence, the data dimensionality is reduced based on the properties that better describe the evolution of the dynamical system represented by the reactive flow. This new application of HODMD presents an advance in the field of the development of ROMs grounded in physics to model combustion dynamics.

4.3. Robustness of the modelling tools

This section shows the robustness of the previously developed ROM, comparing the RRMSE in the reconstruction as a function of the number of variables analysed and the number of components. To select the number of variables, PCA is applied to 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 to 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. Fig. 9 shows the variation of the RRMSE as a function of the number of variables (from PCA) and the number of components (from HODMD) retained. The analysis is carried out on the two databases presented before, the database with all components (All), and the reduced database (Red). The variable selection method in PCA does not significantly affect the result, although for a small number of selected variables, the 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 a 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 range scaling. Additionally, the reconstruction error reaches a lower stable value 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 at 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 increase 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 a function of the number of variables is plotted in Fig. 10. 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 the 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 prioritise them properly. When there are 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 again 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. In particular, a modification to the original algorithm was introduced to develop ROMs for reacting flows. This application is based on a new methodology, fully data-driven, consisting of three main steps. The first step focuses on preprocessing the database analysed, using techniques generally used in machine learning and PCA, which consist on centring and scaling each variable. In the second step, HOSVD is applied to reduce the data dimensionality in time, as well as in each one of the spatial dimensions and variables, separately. In this way, different dimensionality reductions are carried out for each dimension of the tensor. This local application of SVD improves the performance of the method, enhancing the compression capabilities

of the system. Finally, the third step consists of the application of the classical HODMD algorithm.

The effect obtained from the reduction of the number of components is studied in detail, measuring the accuracy of the ROM through the RRMSE from the reconstruction of the original database. Two different tolerances (coarse and fine) were set to study two different cases: (i) keeping all components in the model and (ii) reducing them 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 range scaling is used. The reason for these results is that auto scaling gives to all the variables the same importance. Range scaling, instead, identifies the main flow variables, thus providing more accurate flow reconstructions using a smaller number of components.

HODMD is applied to the previous reduced-dimensionality databases, finding that the main dynamics driving the flow motion is periodic. The main frequency identified when using the range and scaling methods is different. The analysis of the leading modes in both cases suggests that each scaling method also focuses on different parts of the domain. While range scaling is focused on the far field, where the flow motion is lead by the velocity perturbation of the velocity profile, auto scaling is focused on the part of the domain where the oxidation reaction starts. The RRMSE for the reconstruction is smaller than 1% in the case scaled with range scaling, while this error is ~24% when using auto scaling. The differences found between these errors suggest that using the range scaling, the dynamics driving the main variables are well captured by HODMD.

The robustness of the method for identifying the main flow dynamics was tested by applying HODMD to different sets of variables from the same numerical simulation. The variables in each set have been selected using PCA with two variable selection methods, B2 and B2 with rotation. The RRMSE of the reconstructed field fluctuates as a function of the number of variables as ~0.6%–1.5% and ~20%–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 stabilises. These results suggest once again 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 between HODMD and the variable selection using PCA suggests that using a subset of the original variables is 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. HODMD is suitable to construct physics-aware ROMs based on the main dynamics of the system. Scaling of the variables in the preprocessing step is the key point in this methodology, which is capable of reconstructing the original database with precision with a strong reduction in data dimensionality, especially when using the range scaling methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

AC and SLC acknowledge the grant PID2020-114173RB-I00 funded by MCIN/AEI/10.13039/501100011033. SLC and AC also acknowledge the support of Comunidad de Madrid through the call Research Grants for Young Investigators from Universidad Politécnica de Madrid, Spain. AC also acknowledges the support of Universidad Politécnica de Madrid, Spain, 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] Gaba Aurel, Iordache Stefania. Reduction of air pollution by combustion processes. 2011.
- [2] Liu Zhe, Braun James, Paniagua Guillermo. Thermal power plant upgrade via a rotating detonation combustor and retrofitted turbine with optimized endwalls. *Int J Mech Sci* 2020;188:105918.
- [3] Ordaz-Hernandez Keny, Fischer Xavier. Fast reduced model of non-linear dynamic Euler–Bernoulli beam behaviour. *Int J Mech Sci* 2008;50(8):1237–46.
- [4] Cant Stewart. High-performance computing in computational fluid dynamics: progress and challenges. *Phil Trans R Soc A* 2002;360(1795):1211–25.
- [5] Parente Alessandro, Sutherland JC, Dally Bassam B, Tognotti Leonardo, Smith PJ. Investigation of the MILD combustion regime via principal component analysis. *Proc Combust Inst* 2011;33(2):3333–41.
- [6] Bellemans Aurélie, Aversano Gianmarco, Coussement Axel, Parente Alessandro. Feature extraction and reduced-order modelling of nitrogen plasma models using principal component analysis. *Comput Chem Eng* 2018;115:504–14.
- [7] Coussement Axel, Gicquel Olivier, Parente Alessandro. MG-local-PCA method for reduced order combustion modeling. *Proc Combust Inst* 2013;34(1):1117–23.
- [8] Palomba Ilaria, Richiedei Dario, Trevisani Alberto. Mode selection for reduced order modeling of mechanical systems excited at resonance. *Int J Mech Sci* 2016;114:268–76.
- [9] Torrez Sean M, Driscoll James F, Ihme Matthias, Fotia Matthew L. Reduced-order modeling of turbulent reacting flows with application to ramjets and scramjets. *J Propuls Power* 2011;27(2):371–82.
- [10] Huang Cheng, Duraisamy Karthik, Merkle Charles. Component-based reduced order modeling of large-scale complex systems. *Front Phys* 2022;10:900064.
- [11] Lumley JL. The structure of inhomogeneous turbulent flows. In: Yaglam AM, Tatarsky VI, editors. *Proceedings of the international colloquium on the fine scale structure of the atmosphere and its influence on radio wave propagation*. Moscow: Doklady Akademii Nauk SSSR, Nauka; 1967.
- [12] Swischuk Renee, Kramer Boris, Huang Cheng, Willcox Karen. Learning physics-based reduced-order models for a single-injector combustion process. *AIAA J* 2020;58(6):2658–72.
- [13] McQuarrie Shane A, Huang Cheng, Willcox Karen E. Data-driven reduced-order models via regularised Operator Inference for a single-injector combustion process. *J R Soc N Z* 2021;51(2):194–211.
- [14] Huang Cheng, Duraisamy Karthik, Merkle Charles. Challenges in reduced order modeling of reacting flows. In: 2018 joint propulsion conference.
- [15] Parente Alessandro, Sutherland James C. Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combust Flame* 2013;160(2):340–50.
- [16] Isaac Benjamin J, Coussement Axel, Gicquel Olivier, Smith Philip J, Parente Alessandro. Reduced-order PCA models for chemical reacting flows. *Combust Flame* 2014;161(11):2785–800.
- [17] Parente Alessandro, Sutherland James C, Tognotti Leonardo, Smith Philip J. Identification of low-dimensional manifolds in turbulent flames. *Proc Combust Inst* 2009;32(1):1579–86.
- [18] Bishop Christopher M. Pattern recognition. *Mach Learn* 2006;128(9).
- [19] Jolliffe Ian T, Cadima Jorge. Principal component analysis: a review and recent developments. *Phil Trans R Soc A* 2016;374(2065):20150202.
- [20] D'Alessio Giuseppe, Cuoci Alberto, Parente Alessandro. Feature extraction and artificial neural networks for the on-the-fly classification of high-dimensional thermochemical spaces in adaptive-chemistry simulations. *Data-Centric Eng* 2021;2.
- [21] Zdybał Kamila, D'Alessio Giuseppe, Aversano Gianmarco, Malik Mohammad Rafi, Coussement Axel, Sutherland James C, Parente Alessandro. Advancing reacting flow simulations with data-driven models. 2022.
- [22] D'Alessio Giuseppe, Parente Alessandro, Stagni Alessandro, Cuoci Alberto. Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. *Combust Flame* 2020;211:68–82.

- [23] D'Alessio Giuseppe, Cuoci Alberto, Aversano Gianmarco, Bracconi Mauro, Stagni Alessandro, Parente Alessandro. Impact of the partitioning method on multidimensional adaptive-chemistry simulations. *Energies* 2020;13(10):2567.
- [24] D'Alessio Giuseppe, Attili Antonio, Cuoci Alberto, Pitsch Heinz, Parente Alessandro. Analysis of turbulent reacting jets via principal component analysis. In: *Data analysis for direct numerical simulations of turbulent combustion*. Springer; 2020, p. 233–51.
- [25] Jolliffe Ian T. Discarding variables in a principal component analysis. I: Artificial data. *J R Stat Soc Ser C Appl Stat* 1972;21(2):160–73.
- [26] Jolliffe Ian T. Discarding variables in a principal component analysis. II: Real data. *J R Stat Soc Ser C Appl Stat* 1973;22(1):21–31.
- [27] Krzanowski Wojtek J. Selection of variables to preserve multivariate data structure, using principal components. *J R Stat Soc Ser C Appl Stat* 1987;36(1):22–33.
- [28] Tipler Steven, D'Alessio Giuseppe, Van Haute Quentin, Parente Alessandro, Contino Francesco, Coussement Axel. Predicting octane numbers relying on principal component analysis and artificial neural network. *Comput Chem Eng* 2022;161:107784.
- [29] D'Alessio Giuseppe, Attili Antonio, Cuoci Alberto, Pitsch Heinz, Parente Alessandro. Unsupervised data analysis of direct numerical simulation of a turbulent flame via local principal component analysis and procustes analysis. In: *International workshop on soft computing models in industrial and environmental applications*. Springer; 2020, p. 460–9.
- [30] Schmid Peter J. Dynamic mode decomposition of numerical and experimental data. *J Fluid Mech* 2010;656:5–28.
- [31] Yang Ce, Fu Li, Hu Chenxing, Shi Xin. Modelling and dynamic mode analysis of compressor impeller spike-type stall with global stability approach. *Int J Mech Sci* 2021;201:106486.
- [32] Le Clainche Soledad. An introduction to some methods for soft computing in fluid dynamics. In: *Advances in intelligent systems and computing*, Vol. 950. Cham: Springer; 2019, p. 557–66.
- [33] Richecoeur Franck, Hakim Layal, Renaud Antoine, Zimmer Laurent. DMD algorithms for experimental data processing in combustion. In: *Proceeding of the 2012 summer program*. Center for Turbulence Research, Stanford University; 2012, p. 459–68.
- [34] Gunaratne Gemunu, Roy Sukesh. Robust mode analysis. *Mathematics* 2021;9:1057.
- [35] Quinlan John M, Zinn Ben T. Transverse combustion instabilities: Modern experimental techniques and analysis. In: *50th AIAA/ASME/SAE/ASEE joint propulsion conference*. 2014.
- [36] Huang Cheng, Anderson William E, Harvazinski Matthew E, Sankaran Venkateswaran. Analysis of self-excited combustion instabilities using decomposition techniques. *AIAA J* 2016;54(9):2791–807.
- [37] Chatterjee Souvik, Mukhopadhyay Achintya, Sen Swarnendu. Stability study of laminar flame using proper orthogonal decomposition and dynamic mode decomposition. In: *N3I-Int'l summer school and workshop on non-normal and nonlinear effects in aero-and thermoacoustics*. 2013, p. 13.
- [38] Motheau Emmanuel, Nicoud Franck, Poinot Thierry. Mixed acoustic-entropy combustion instabilities in gas turbines. *J Fluid Mech* 2014;749:542–76.
- [39] Abou-Taouk Abdallah, Sadasivuni SK, Lörstard Daniel, Bulat Ghenadie, Eriksson Lars-Erik. CFD analysis and application of dynamic mode decomposition for resonant-mode identification and damping in an SGT-100 DLE combustion system. In: *Proceedings of the 7th European combustion meeting*. 2015.
- [40] Ghani Abdulla, Poinot Thierry, Gicquel LYM, Staffelbach Gabriel. LES of longitudinal and transverse self-excited combustion instabilities in a bluff-body stabilized turbulent premixed flame. *Combust Flame* 2015;162:4075–83.
- [41] Grenga Temistocle, MacArt Jonathan F, Mueller Michael E. Dynamic mode decomposition of a direct numerical simulation of a turbulent premixed planar jet flame: convergence of the modes. *Combust Theory Model* 2018;22(4):795–811.
- [42] Grenga T, Mueller ME. Dynamic mode decomposition: A tool to extract structures hidden in massive datasets. In: *Data analysis for direct numerical simulations of turbulent combustion*. Springer; 2020, p. 157–76.
- [43] Le Clainche Soledad, Vega José M. Analyzing nonlinear dynamics via data-driven dynamic mode decomposition-like methods. *Complexity* 2018;2018.
- [44] Le Clainche Soledad, Vega José M. Higher order dynamic mode decomposition. *SIAM J Appl Dyn Syst* 2017;16(2):882–925.
- [45] Corrochano Adrián, Xavier Donnatella, Schlatter Philipp, Vinuesa Ricardo, Le Clainche Soledad. Flow structures on a planar food and drug administration (FDA) nozzle at low and intermediate Reynolds number. *Fluids* 2021;6(1):4.
- [46] Martínez Soledad Le Clainche, Sastre Francisco, Vega José M, Angel Velazquez. Higher order dynamic mode decomposition applied to post-process a limited amount of noisy PIV data. In: *47th AIAA fluid dynamics conference*.
- [47] Le Clainche Soledad, Moreno-Ramos Rubén, Taylor Paul, Vega José M. New robust method to study flight flutter testing. *J Aircr* 2019;56(1):336–43.
- [48] Le Clainche S, Izbassarov Daulet, Rosti M, Brandt Luca, Tammisola Outi. Coherent structures in the turbulent channel flow of an elastoviscoplastic fluid. *J Fluid Mech* 2020;888.
- [49] Kaiser Henry F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958;23(3):187–200.
- [50] Le Clainche Soledad, Vega José M, Soria Julio. Higher order dynamic mode decomposition of noisy experimental data: The flow structure of a zero-net-mass-flux jet. *Exp Therm Fluid Sci* 2017;88:336–53.
- [51] Vega Jose Manuel, Le Clainche Soledad. Higher order dynamic mode decomposition and its applications. Academic Press; 2020.
- [52] Fürst Magnus, Parente Alessandro. Uncertainty quantification of chemical kinetics for MILD combustion. 2016.
- [53] Widuch Agata, Nowak Marcin, Sukienik Dawid, Myöhänen Kari, Nikku Markku, Shah Srujal, Parente Alessandro, Adamczyk Wojciech. Towards application of uncertainty quantification procedure combined with experimental procedure for assessment of the accuracy of the DEM approach dedicated for granular flow modeling. *Powder Technol* 2022;399:117182.
- [54] Cafiero Marianna, Dias Véronique, Iavarone Salvatore, Coussement Axel, Jeanmart Hervé, Parente Alessandro. Investigation of temperature correction methods for fine wire thermocouple losses in low-pressure flat premixed laminar flames. *Combust Flame* 2022;244:112248.
- [55] Le Clainche Soledad, Ferrer Esteban. A reduced order model to predict transient flows around straight bladed vertical axis wind turbines. *Energies* 2018;11(3):566.
- [56] Corrochano Adrián, Neves Ana F, Khanal Bidur, Clainche Soledad Le, Lawson Nicholas J. DES of a slingsby firefly aircraft: Unsteady flow feature extraction using POD and HODMD. *J Aerosp Eng* 2022;35(5):04022063.
- [57] Le Clainche Soledad, Vega José M. Higher order dynamic mode decomposition to identify and extrapolate flow patterns. *Phys Fluids* 2017;29(8):084102.
- [58] Beltrán Víctor, Le Clainche Soledad, Vega José. An adaptive data-driven reduced order model based on higher order dynamic mode decomposition. *J Sci Comput* 2022;92.
- [59] Clainche Soledad Le, Rosti Marco, Brandt Luca. Flow structures and shear-stress predictions in the turbulent channel flow over an anisotropic porous wall. *J Phys Conf Ser* 2020;1522(1):012016.
- [60] Muñoz Eva, Le Clainche Soledad. On the topology patterns and symmetry breaking in two planar synthetic jets. *Phys Fluids* 2022;34(2):024103.
- [61] Lazpita Eneko, Martínez-Sánchez Álvaro, Corrochano Adrián, Hoyas Sergio, Le Clainche Soledad, Vinuesa Ricardo. On the generation and destruction mechanisms of arch vortices in urban fluid flows. *Phys Fluids* 2022;34(5):051702.
- [62] Amor Christian, Pérez José M, Schlatter Philipp, Vinuesa Ricardo, Le Clainche Soledad. Modeling the turbulent wake behind a wall-mounted square cylinder. *Logic J IGPL* 2020;30(2):263–76.
- [63] Corrochano Adrián, Le Clainche Soledad. Structural sensitivity in non-linear flows using direct solutions. *Comput Math Appl* 2022;128:69–78.
- [64] Groun Nourelhouda, Villalba-Orero María, Lara-Pezzi Enrique, Valero Eusebio, Garicano-Mena Jesús, Le Clainche Soledad. Higher order dynamic mode decomposition: From fluid dynamics to heart disease analysis. *Comput Biol Med* 2022;144:105384.
- [65] Tucker Ledyard R. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966;31(3):279–311.
- [66] Chen KK, Tu JH, Rowley CW. Variants of dynamic mode decomposition: Boundary condition, koopman, and fourier analyses. *J Nonlinear Sci* 2012;22(6):887–915.
- [67] Kou Jiaqing, Le Clainche Soledad, Zhang Weiwei. A reduced-order model for compressible flows with buffeting condition using higher order dynamic mode decomposition with a mode selection criterion. *Phys Fluids* 2018;30(1):016103.
- [68] Jolliffe Ian T. Choosing a subset of principal components or variables. In: *Principal component analysis*. Springer; 2002, p. 111–49.
- [69] Zdybał Kamila, D'Alessio Giuseppe, Attili Antonio, Coussement Axel, Sutherland James C, Parente Alessandro. Local manifold learning and its link to domain-based physics knowledge. 2022.
- [70] Ranzi Eliseo, Frassoldati Alessio, Grana Roberto, Cuoci Alberto, Faravelli Tiziano, Kelley Andrew P, Law Chung K. Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels. *Prog Energy Combust Sci* 2012;38(4):468–501.
- [71] Cuoci Alberto, Frassoldati Alessio, Faravelli Tiziano, Ranzi Eliseo. Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method. *Energy Fuels* 2013;27(12):7730–53.

Chapter 4

Hierarchical higher-order dynamic mode decomposition for clustering and feature selection

4.1 Publication information

Title: Hierarchical higher-order dynamic mode decomposition for clustering and feature selection

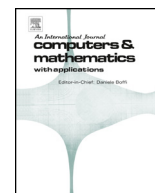
DOI: <https://doi.org/10.1016/j.camwa.2024.01.003>

Journal: Computers & Mathematics with Applications

Date: 15 March 2024

Publisher: Elsevier

Copyright: © 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license



Hierarchical higher-order dynamic mode decomposition for clustering and feature selection

Adrián Corrochano^{a,*}, Giuseppe D'Alessio^{b,c}, Alessandro Parente^b, Soledad Le Clainche^a

^a School of Aerospace Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain

^b Université Libre de Bruxelles, École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Bruxelles, Belgium

^c CRECK Modeling Lab, Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

ARTICLE INFO

Keywords:

Reacting flows

Higher-order dynamic mode decomposition

Reduced order models

ABSTRACT

This article introduces a novel, fully data-driven method for forming reduced order models (ROMs) in complex flow databases that consist of a large number of variables. The algorithm utilizes higher order dynamic mode decomposition (HODMD), a modal decomposition method, to identify the main frequencies and associated patterns that govern the flow dynamics. By incorporating various normalization techniques into an iterative process, clusters of variables with similar dynamics are identified, allowing the classification of different instabilities and patterns present in the flow. This method, known as hierarchical HODMD (h-HODMD), has been thoroughly tested in the development of ROMs using three different databases obtained from numerical simulations of a nonpremixed coflow methane flame. The effectiveness of h-HODMD has been demonstrated as it consistently outperforms HODMD in terms of modeling and reconstructing flow dynamics using a reduced number of modes. Additionally, the clusters of variables identified by h-HODMD reveal the algorithm's ability to group chemical species whose behavior is consistent from a kinetic perspective. h-HODMD allows for the construction of inexpensive reduced dynamical models that can predict flame liftoff, identify the occurrence of local extinction and blowout conditions, and facilitate control purposes.

1. Introduction

Understanding the complex physical and chemical phenomena occurring in reacting systems is crucial because of the urgent need to reduce the environmental impact of next generation combustion systems. However, performing detailed experimental and/or numerical investigations of combustion systems (especially when large-scale systems are taken into account) is not always straightforward because of the associated high computational cost [1].

There is an increase of general interest in using reduced order models (ROMs) for complex systems, as they can model the main dynamics at a reduced computational cost. Baiges et al. [2] created a ROM combining proper orthogonal decomposition (POD) and a Galerkin projection for the incompressible Navier-Stokes equations. This combination of POD and Galerkin has also been used for creating ROMs for the Reynold-Averaged Navier-Stokes (RANS) equations [3] and for integrating the Shifted Boundary Method [4]. More recently, the Galerkin projection has been used for control problems for high-dimensional linear systems [5]. Thus, the adoption of ROMs represents a viable option

in describing complex reacting systems with low errors and minimal computational costs [6–8].

One of the most employed algorithms to construct ROMs for reacting flows is represented by the principal component analysis (PCA) [9], also known as proper orthogonal decomposition (POD) [10]. The latter is a dimensionality reduction algorithm based on the projection of the original system onto a lower-dimensional manifold whose directions (the principal components, PCs) are orthogonal and obtained as a linear combination of the original variables. In literature, it is possible to find many applications exploiting PCA to speed-up computational fluid dynamics (CFD) simulations of reacting systems by reducing the number of the differential equations to be solved for combustion [11–14] or atmospheric re-entry applications [15]. Furthermore, the algorithm and its local formulation were recently validated for clustering in chemical kinetics applications [16,17], for the data analysis of Direct Numerical Simulation (DNS) of reacting jets [18,19], and for feature extraction and selection to enhance the classification and regression accuracy of Artificial Neural Networks (ANN) [20,21].

* Corresponding author.

E-mail address: adrian.corrochanoc@upm.es (A. Corrochano).

Recently, dynamic mode decomposition (DMD) [22] was proposed for the analysis and development of ROMs of unsteady systems. The main advantage of this algorithm over alternative modal decomposition techniques, such as PCA, is represented by the possibility of retrieving the frequency associated with the modes to analyze the system. In addition, DMD finds patterns that can be related to flow instabilities [23–26], and it can also be used for dimensionality reduction, as fewer DMD modes are usually capable of reproducing the main flow dynamics with a reduced computational cost. Nevertheless, its applicability is usually limited to specific conditions [27]. Thus, higher order dynamic mode decomposition (HODMD) [28] was found to be a more robust alternative to standard DMD for nonreactive complex flows and noisy experimental data, and has been widely used in the analysis of several complex flows in industrial applications, either to create ROMs or to study flow patterns [29–32]. Recently, HODMD was also proposed and validated for the development of reduced order models for reacting flows [33].

In the current work, a modified formulation for HODMD is proposed to accomplish feature selection and clustering of variables when high-dimensional reacting systems are considered. More specifically, the hierarchical HODMD (hereinafter referred to as h-HODMD), iteratively identifies the best reconstructed variables and associates them with the corresponding DMD modes and frequencies. The performance of the proposed algorithm is assessed using three databases obtained from CFD simulations of a laminar co-flow flame taking into account different kinetic mechanisms, as well as different boundary conditions. The two kinetic mechanisms used to describe methane oxidation consist of 84 and 53 chemical species, respectively. A wide range of chemical time scales is also considered since the nitrogen oxides' chemistry is also included in one of the two mechanisms. The considered flame, despite being laminar, exhibits complex physics because of local extinction phenomena. Finally, the results are compared with the ones obtained by using a standard HODMD in terms of the achievable degree of reduction as well as in terms of reconstruction error.

The current work is organized as follows. In Section 2, the original HODMD is explained and the modifications to obtain the h-HODMD are discussed. In Section 3, the numerical setup for the CFD simulations that were considered for the training datasets is shown and discussed. Finally, the main results and conclusions are presented in Sections 4 and 5, respectively.

2. Higher order dynamic mode decomposition

Higher Order Dynamic Mode Decomposition (HODMD) [28] is a data driven method used in fluid dynamics to identify flow patterns. The latter represents an extension of Dynamic Mode Decomposition (DMD) [22] used for the analysis of complex flows [34], noisy experiments [35], or turbulent flows [36]. HODMD decomposes data equispaced in time as an expansion of Fourier-like DMD modes, as

$$v(x, y, t_k) \simeq \sum_{m=1}^M a_m \mathbf{u}_m(x, y) e^{(\delta_m + i\omega_m)t_k}, \quad k = 1, \dots, K, \quad (1)$$

where $\mathbf{u}_m$ are the DMD modes and a_m , ω_m , δ_m are their associated amplitudes, frequencies and growth rates. The number of DMD modes (M) is referred to as *spectral complexity*, whereas K in Equation (1) represents the number of snapshots available for the analyzed data, which is referred to as the *temporal dimension*. A thorough description of the algorithm can be found in Ref. [28], while the associated code is available in Ref. [37].

2.1. Main algorithm

Firstly, equidistant snapshots in time for the considered dynamical system (with time interval Δt) are organized as the following snapshot matrix

$$\mathbf{V}_1^K = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_{K-1}, \mathbf{v}_K]. \quad (2)$$

From the definition in Equation (2), it follows that the dimensions of the input matrix are thus $J \times K$, where K is the number of snapshots analyzed and $J = N_x \times N_y$ is the number of grid points considered for the CFD simulations. The HODMD algorithm can be summarized into two main steps: first, a spatial reduction via singular value decomposition (SVD) is accomplished and then DMD algorithm with a sliding window process (DMD-d) is applied.

Within the first step, SVD is applied to the snapshot matrix to reduce noise and spatial redundancies to only retain the large scales:

$$\mathbf{V}_1^K = \mathbf{U} \mathbf{\Sigma} \mathbf{T}^T = \mathbf{U} \hat{\mathbf{V}}_1^K, \quad (3)$$

where $\hat{\mathbf{V}}_1^K$ is the *reduced snapshot matrix* whose dimensions are $N \times K$. The number of SVD modes to be retained (N) can be determined setting a tolerance ε_1 as

$$\frac{s_{N+1}}{s_1} \leq \varepsilon_1, \quad (4)$$

where s_i are the singular values. By means of the operations discussed above, the spatial complexity can be reduced from J grid points to N linearly independent vectors. When high-dimensional data are analyzed, such as those obtained from complex reacting systems [38], SVD is replaced with Higher Order Singular Value Decomposition (HOSVD) [39]. To perform this decomposition, the snapshot matrix $\mathbf{V}_1^K$ is organized in tensor form $\mathbf{V}_{ijkl}$ having dimensions $N_v \times N_x \times N_y \times K$, where N_v is the number of variables taken into account on the analysis. HOSVD can be used to reduce the dimensionality in an efficient way. The algorithm filters better small scales and noise, as it filters each of the dimensions of the tensor separately. The algorithm applies standard SVD to the four matrices whose columns are formed by each of the four dimensions of the tensor. In particular, if a four-dimensional tensor is considered, HOSVD is applied as follows:

$$\mathbf{V}_{ijkl} \simeq \sum_{p_1=1}^{N_v} \sum_{p_2=1}^{N_x} \sum_{p_3=1}^{N_y} \sum_{n=1}^K S_{p_1 p_2 p_3 n} \mathbf{W}_{ip_1}^{(v)} \mathbf{W}_{jp_2}^{(x)} \mathbf{W}_{lp_3}^{(y)} \mathbf{T}_{kn}, \quad (5)$$

where $S_{p_1 p_2 p_3 n}$ is the core tensor and $\mathbf{W}^{(v)}$, $\mathbf{W}^{(x)}$, $\mathbf{W}^{(y)}$ and $\mathbf{T}$ are the modes of the decomposition in each dimension. As well as in SVD, it is possible to retain a number of modes in each direction, in order to remove the noise and retain the large scales. The temporal matrix $\mathbf{T}$ is used as the reduced snapshot matrix.

In the second step, i.e., the application of the *DMD-d algorithm*, the high order Koopman assumption is applied to the reduced snapshot matrix as follows:

$$\hat{\mathbf{V}}_{d+1}^K \simeq \hat{\mathbf{R}}_1 \hat{\mathbf{V}}_1^{K-d} + \hat{\mathbf{R}}_2 \hat{\mathbf{V}}_2^{K-d+1} + \dots + \hat{\mathbf{R}}_d \hat{\mathbf{V}}_d^{K-1}. \quad (6)$$

The Koopman operators $\hat{\mathbf{R}}_1$, $\hat{\mathbf{R}}_2$, ..., $\hat{\mathbf{R}}_d$ are encompassed into a single matrix, which contains the dynamics of the system. When $d = 1$ is considered, the algorithm provides the same results as the standard DMD.

The eigenvalues provide the frequencies ω_m and the growth rates δ_m of the eigenvectors, which are used to construct the DMD modes $\mathbf{u}_m$. Finally, the amplitudes of the DMD modes a_m are obtained via least squares fitting of the expansion which was previously expressed in Equation (1). Using a second tunable tolerance ε_2 , it is possible to then select M DMD modes that are representative of the main dynamics of the system as follows:

$$a_{M+1}/a_1 \leq \varepsilon_2. \quad (7)$$

When the considered database is related to the transient state of a numerical simulation, another step is necessary to identify the leading modes driving the flow dynamics. This step consists in computing the contribution to the transient state $\mathbf{I}_m$ defined as

$$I_m = \sum_{k=1}^K |a_m e^{(\delta_m + i\omega_m)t_k}| \|u_m\|_F^2 \times \Delta t, \quad (8)$$

where $\|\cdot\|_F$ is the square of the Frobenius norm. This index I_m also takes into account the damping or growth of the modes over time, which is reflected in δ_m . A much more detailed description about this criterion can be found in [40].

Finally, the accuracy of the algorithm is measured in terms of the relative root mean square error (RRMSE) as

$$\text{RRMSE} = \sqrt{\frac{\sum_{k=1}^K \|v_k^{\text{DMD}} - v_k\|_2^2}{\sum_{k=1}^K \|v_k\|_2^2}}. \quad (9)$$

2.2. Pre-processing techniques for applying HODMD to combustion databases

When HODMD is applied to combustion databases, a pre-processing step of the data in terms of centering and scaling must be applied due to their multivariate nature [33,41]. Centering is accomplished by subtracting the temporal mean of the considered variable from each statistical observation of the dataset, so that each sample can be seen as a fluctuation from the mean. Scaling the variables, instead, is motivated by the need of having them lying in the same order of magnitude. The centered and scaled i -th statistical observation for the j -th variable is finally obtained as

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \bar{x}_{i,j}}{c_j}, \quad (10)$$

where $\bar{x}_{i,j}$ represents the temporal mean of the i -th point considering the variable's vector x_j and c_j is the associated scaling factor. Four scaling techniques were considered for the current analysis:

1. *Range scaling*: The difference between the maximum and minimum value of each variable is used as a scaling factor.
2. *Auto scaling*: the standard deviation of the j -th variable (σ_j) is used as a scaling factor.
3. *Pareto scaling*: the square root of the standard deviation of the j -th variable ($\sqrt{\sigma_j}$) is used as a scaling factor.
4. *Vast scaling*: the product of the standard deviation σ_j and the coefficient of variation $\sigma_j/\bar{x}_j$ is used as a scaling factor.

The effects of centering on DMD have been analyzed with positive results in [42]. The effects of the different scaling factors on modal decomposition techniques such as PCA and HODMD have been thoroughly analyzed for combustion applications in the literature [18,33,41]. The effects of the scaling method have also been studied in other fields of knowledge where the data also presents similar characteristics, as in molecular sciences [43] and medicine [44].

2.3. Hierarchical Higher Order Dynamic Mode Decomposition

The hierarchical higher order dynamic mode decomposition (h-HODMD) is a new algorithm which aims at iteratively identifying the main dynamics of the considered system, creating clusters of variables having the same dynamic behavior.

After applying the chosen preprocessing technique to the considered database, HODMD is applied to the scaled tensor. Once the model's order is reduced, each variable can be reconstructed back to the original space by multiplying by the scaling factor and adding the mean value, as

$$x_j^{\text{DMD}} = \tilde{x}_j^{\text{DMD}} * c_j + \bar{x}_j. \quad (11)$$

The accuracy of the reconstruction of each variable is then measured by using the relative root mean square error. The user selects a tolerance ϵ_v to retain the variables that verify the criterion.

$$\sqrt{\frac{\sum_{k=1}^K \|x_{j,k}^{\text{DMD}} - x_{j,k}\|_2^2}{\sum_{k=1}^K \|x_{j,k}\|_2^2}} < \epsilon_v. \quad (12)$$

Afterwards, HODMD is applied again to the remaining variables so that they can be grouped according to their reconstruction error. The variables of the considered system are progressively better reconstructed as HODMD is applied at each iteration, and new modes and frequencies are obtained. The reconstruction improves as there is a removal of information about the cluster, which has different dynamics from the rest of the variables. A summary of the iterative process is shown in Fig. 1: a group of variables (cluster) is retained at each iteration and is removed from the dataset for the next iteration of HODMD. In this work, the number of clusters created in each analysis (i.e., the number of iterations of the algorithm) is set to 4. The algorithm presented in this paper can be schematized as:

1. **Step 1**: Pre-process the variables as explained in Equation (10)
2. **Step 2**: Apply HODMD to the database
3. **Step 3**: Quantify the error made on the reconstruction as in Equation (4) for all the variables studied.
4. **Step 4**: Select the variables whose error is below the established tolerance ϵ_v to form the cluster and remove them from the database.
5. **Step 5**: Repeat Steps 2-4 the desired times.

The advantage of this iterative algorithm is that the clusters created are based on modes with their frequencies. Even though the frequencies ω_m should be the same for all the variables, as it is a coupled system, the amplitudes a_m related to the DMD modes may be not. Therefore, once one iteration of h-HODMD is applied, the variables better reconstructed are the ones selected for the cluster. The method can be adapted to decouple different dynamics in combustion systems. Analyzing all the variables together improve the performance of the DMD algorithm [45], and once analyzed, h-HODMD can find clusters of variables based on modes with their frequencies.

3. Numerical simulations

The CFD simulations performed to obtain the databases analyzed by h-HODMD in the following sections are representative of the oxidation of a nitrogen-diluted methane jet (65% of methane and 35% nitrogen on a molar basis) in air. The oxidizer is injected into the domain with a constant velocity of 35 cm/s, while the fuel is injected according to the following sinusoidal velocity profile:

$$v(r, t) = v_{\max} \left(1 - \frac{r^2}{R^2}\right) [1 + A \sin(2\pi f t)], \quad (13)$$

where $v_{\max}$ is the maximum velocity ($v_{\max} = 70 \text{ cm/s}$), r is the radial coordinate, R represents the internal radius of the nozzle, and A is the amplitude of the perturbation ($A = 0.25$).

Three numerical simulations were carried out changing the frequency of the perturbation f and/or the kinetic mechanism, as summarized in Table 1. The two detailed kinetic mechanisms employed for the CFD simulations were the POLIMI_C1_C3_HT_1412 [46], which takes into account 84 species and 1698 reactions, and the GRI-Mech 3.0 [47], which instead consists of 53 species and 325 reactions. The simulations carried out using the GRI-Mech 3.0 represent a particularly interesting test case for the proposed algorithm. In fact, the kinetic mechanism takes into account the chemistry associated with the formation and consumption of nitrogen oxides (NO_x), and both phenomena are well known to occur on completely different timescales with respect to other chemical species and radicals. Thus, this represents an additional degree of complexity for the formulation of an accurate reduced-order model for the dynamical systems considered in this work.

The CFD simulations were carried out by using LaminarSMOKE, an OpenFOAM-based operator-splitting solver proposed by Cuoci et al.

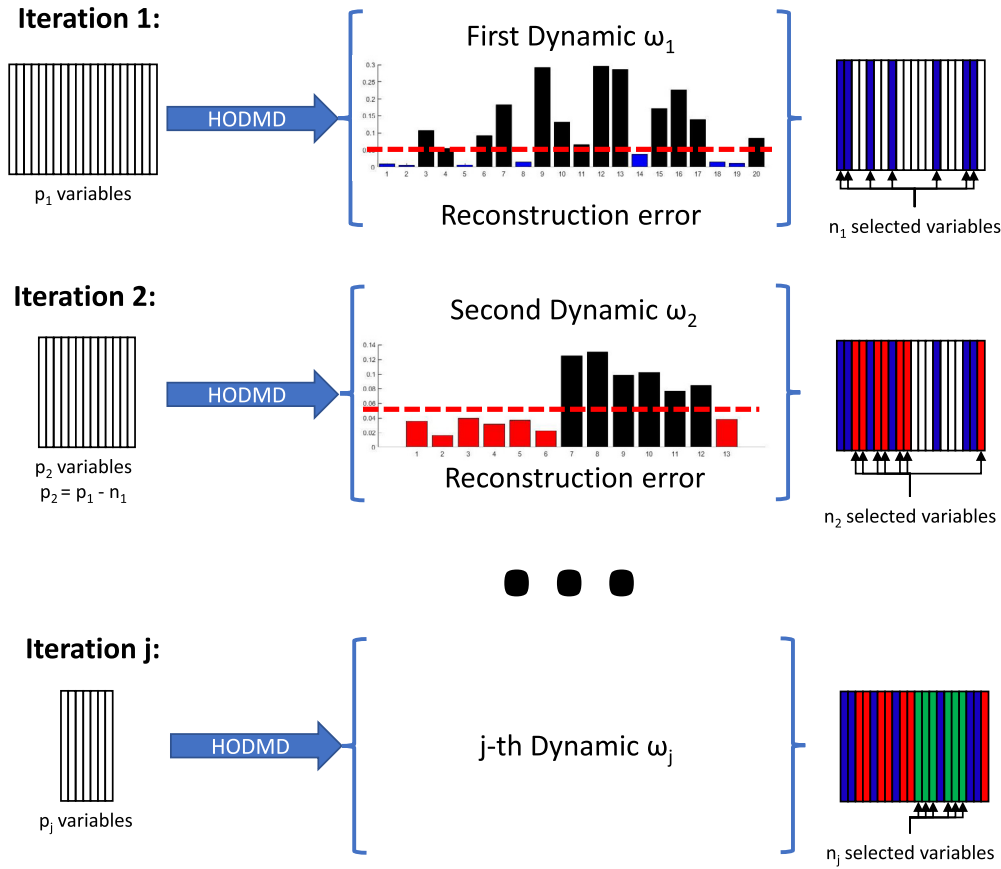


Fig. 1. Sketch for the h-HODMD algorithm: the variables of the considered dynamical system are progressively grouped together according to their reconstruction error, and clusters are obtained in an unsupervised fashion. Each of the dynamics ω_j refers to several DMD modes and the frequencies associated with them, not a single frequency.

Table 1

Frequency of the perturbation (in Hertz) and number of species and reactions being used for the CFD simulations.

	Frequency [Hz]	Number of species	Number of reactions
Simulation 1	10	84	1698
Simulation 2	10	53	325
Simulation 3	20	53	325

Table 2

Number of variables, snapshots, and completed periods for each database considered for the h-HODMD analysis.

	N_v	K	Completed periods
Database 1	83	250	1
Database 2	53	250	3
Database 3	53	250	5

[48]. Additional information regarding the numerical settings for the simulations can be found in Refs. [16,17].

For each simulation reported in Table 1, a database is extracted in the form of a fourth-order tensor having dimensions $N_v \times N_x \times N_y \times K$, where

- N_v is the number of considered variables (temperature and $n_v - 1$ chemical species, respectively);
- N_x is the number of grid points along the streamwise direction;
- N_y is the number of grid points along the spanwise direction;
- K is the number of snapshots equispaced in time, considering $\Delta t = 10^{-3}$.

For all simulations considered, it was imposed $N_x = N_y = 150$, while N_v was changing depending on the number of species originally taken into account by the kinetic mechanism, and K was also different for each database. The selected values for N_v and K for the three simulations are reported in Table 2 with the associated number of completed periods (a period is defined as: $T = 2\pi/f$). The influence of this pa-

rameter on the performance of h-HODMD will also be assessed later in Sec. 4.

4. Results

The algorithm discussed in Sec. 2 was finally applied to the data obtained from the simulations summarized in Table 1. In particular, the databases were first preprocessed by means of the centering and scaling operations reported in Sec. 2.2; afterwards, the algorithm discussed in Sec. 2.1 was applied to extract different groups of variables according to their reconstruction error. Finally, the dynamics of each group were analyzed.

To quantitatively evaluate the performance of the algorithm proposed in this work, two different metrics were used to compute the reconstruction error at each iteration. Firstly, the RRMSE (Equation (9)) after the application of HODMD with the original tensor is computed. Secondly, the average RRMSE value for the variables is calculated as:

$$RRMSE_{mean} = \frac{\sum_{j=1}^{N_v} RRMSE(x_j)}{N_v} \quad (14)$$

Table 3

Values for d and ε obtained in the HODMD calibration process for the cases considered in the analysis.

	Scaling Method							
	Auto		Range		Pareto		Vast	
	d	ε	d	ε	d	ε	d	ε
Database 1	6	4×10^{-2}	6	6×10^{-4}	6	6×10^{-4}	6	5×10^{-3}
Database 2	15	4×10^{-2}	12	6×10^{-4}	12	1×10^{-3}	12	4×10^{-3}
Database 3	12	4×10^{-2}	10	6×10^{-4}	12	2×10^{-3}	12	4×10^{-3}

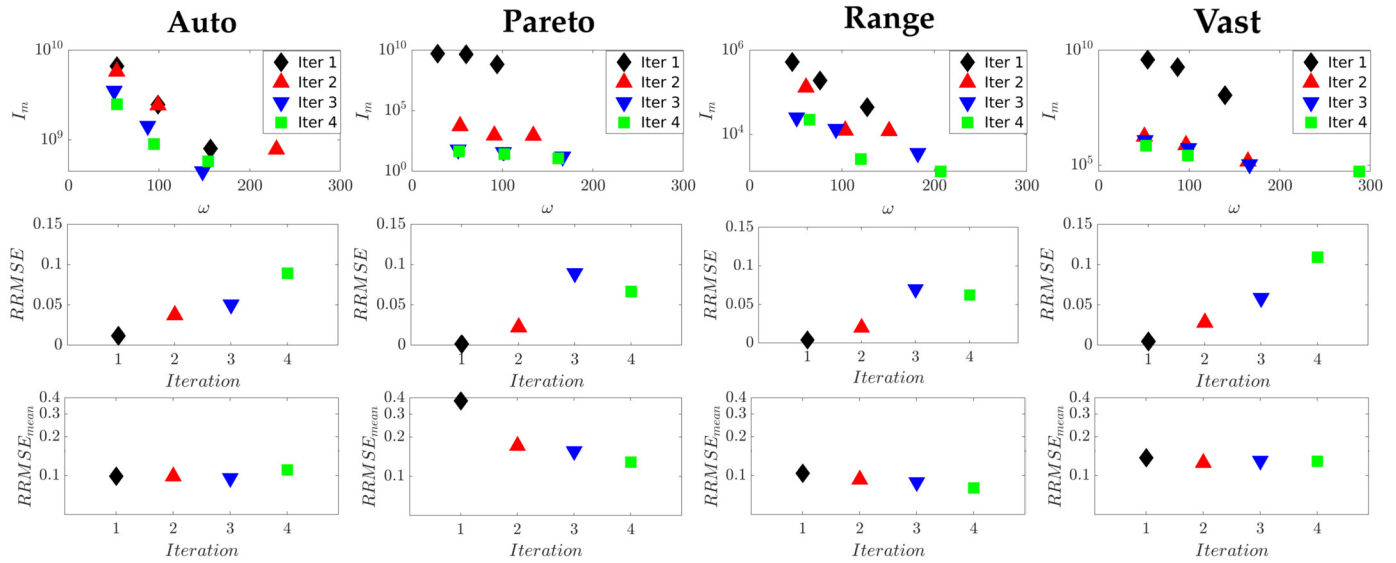


Fig. 2. From top to bottom: Three main frequencies calculated as in Equation (8), $RRMSE$ and $RRMSE_{mean}$. From left to right: auto scaling, pareto scaling, range scaling and vast scaling. The analysis corresponds to Database 1, i.e., to the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 82 species.

The algorithm was calibrated with the aim of ensuring its robustness and obtaining the lowest reconstruction error, according to the procedure presented in [33,37]. The values for d and $\varepsilon = \varepsilon_1 = \varepsilon_2$ adopted for the current work are reported in Table 3.

For the iterative process in h-HODMD, the tolerance ε_v (Equation (12)) was set equal to $\varepsilon_{v,iter} = 0.02 \times iter$, where $iter$ is the number of iterations performed by the algorithm. Thus, the tolerance becomes progressively higher with the number of iterations. The number of variables in each cluster n_{iter} is not kept constant, as this parameter is only controlled by the tolerance ε_v . Other tolerances, i.e. $\varepsilon_{v,iter} = 0.01 \times iter$ and $\varepsilon_{v,iter} = 0.05 \times iter$, have been used for the calibration. However, a high tolerance means that each iteration retains a high number of variables that can have different dynamics among them. On the other hand, a small tolerance may not include all variables with the same dynamics in a single cluster.

The first database, analyzed in Fig. 2, only considers ~ 1 cycle, so the simulation can be considered fully transient. Auto scaling appears to provide similar main dynamics at each iteration, and this is in agreement with what is expected from theory (as this scaling method gives the same importance to all the variables). Range scaling leads instead to different dynamics for each iteration, while Pareto and Vast scaling initially emphasize temperature and most stable species, respectively, to then give to the rest of variables the same importance as the number of iterations increases. Because of that, after removing the best reconstructed variables, the algorithm gives back similar dynamics between two consecutive iterations.

The reconstruction error ($RRMSE$) increases with the number of iterations, as the algorithm progressively removes the variables that are better reconstructed. The first iteration has a low $RRMSE$ for all the scaling methods. After the first iteration, the $RRMSE$ increases for

Pareto and Vast scaling, while it reaches a plateau for Range scaling. Finally, the mean $RRMSE$ ($RRMSE_{mean}$) is stable for auto scaling, and this appears to be justified, since it gives all variables the same importance. In range scaling, instead, the $RRMSE_{mean}$ slowly decreases as the number of iterations increases. When Pareto scaling is applied, $RRMSE_{mean}$ in the first iteration is much higher than in the rest of the iterations. This happens because the latter initially emphasizes temperature, thus ignoring the reconstruction of the rest of the variables. After the first iteration, the rest of the variables become more important and $RRMSE_{mean}$ decreases. The $RRMSE_{mean}$ when using vast scaling remains almost constant with the iterations. Usually, Vast scaling focuses on extremely stable variables, but this effect is mitigated as the transient behavior of the considered database for the analysis increases.

When databases 2 and 3 are analyzed (Figs. 3 and 4, respectively), similar results are obtained. For the main dynamics, Auto scaling provides similar frequencies for all the iterations, Pareto and Vast scaling have different dynamics for the first iteration and find same frequencies for the rest; while Range finds different frequencies for each iteration. Regarding the different reconstruction error metrics, the behavior of each scaling method follows the same behavior observed for the first database. The only metric that changes is the $RRMSE_{mean}$ for Vast scaling, which behaves similarly to Pareto scaling. This is mainly due to the fact that the transient behavior for the numerical simulations 2 and 3 does not dominate as more cycles are taken into account.

Summarizing the results for the three databases, Auto scaling provides the worst reconstruction error, although the clusters have similar dynamics. As this scaling method gives all the variables the same importance [41], the creation of the clusters and the calculation of the

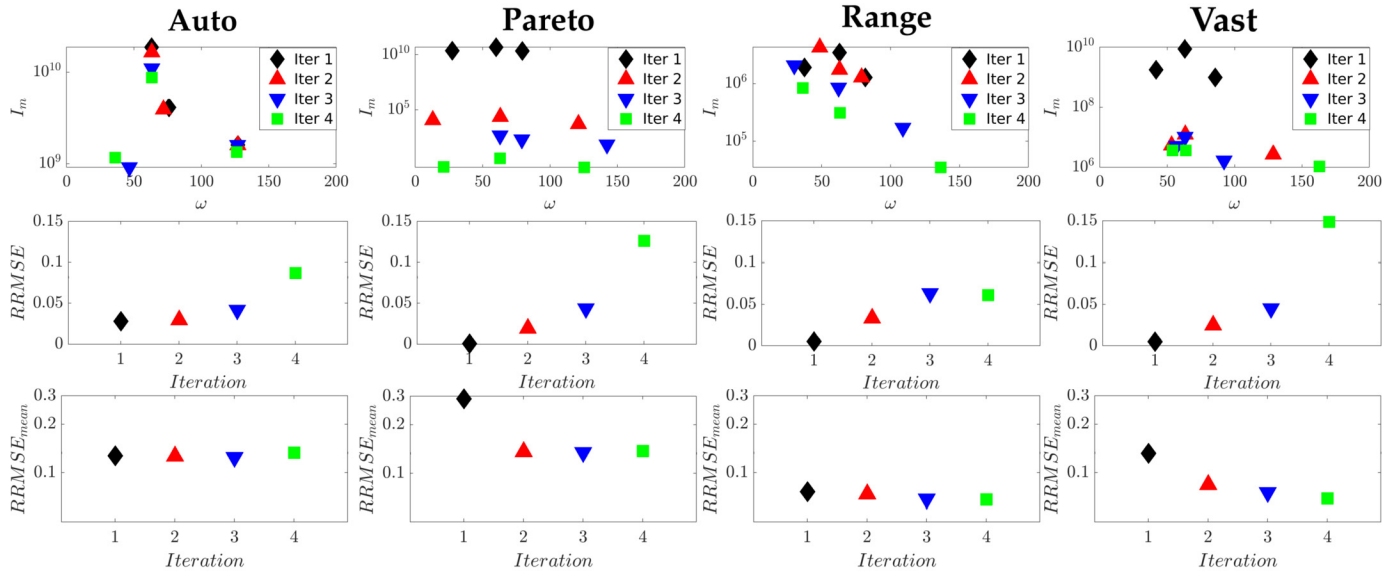


Fig. 3. From top to bottom: Three main frequencies calculated as in Equation (8), $RRMSE$ and $RRMSE_{mean}$. From left to right: auto scaling, pareto scaling, range scaling and vast scaling. The analysis corresponds to Database 2, i.e., to the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species.

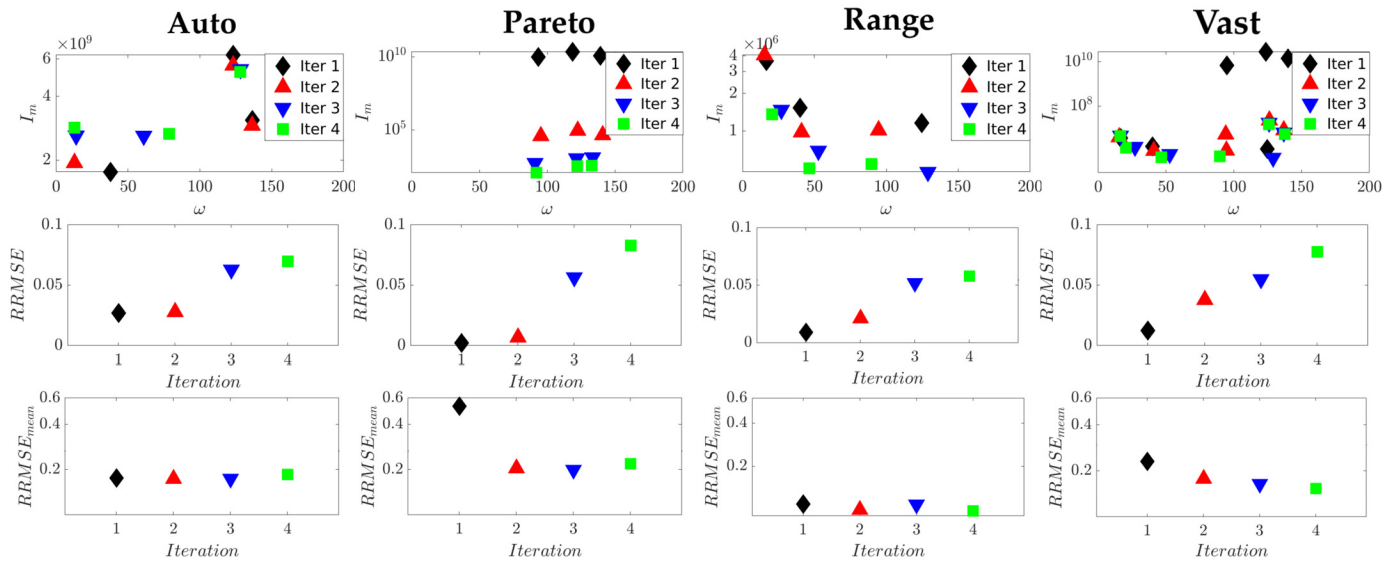


Fig. 4. From top to bottom: Three main frequencies calculated as in Equation (8), $RRMSE$ and $RRMSE_{mean}$. From left to right: auto scaling, pareto scaling, range scaling and vast scaling. The analysis corresponds to Database 3, i.e., to the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 20$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species.

frequencies are more difficult. This does not happen on the rest of scaling methods, where some of the variables emphasize some of the variables among others. Range scaling gives the lowest $RRMSE_{mean}$ on the last iteration, while each cluster of variables on this scaling method has different main dynamics. Pareto behaves similarly to Auto scaling from the second iteration onward, although it allows for a better reconstruction of the first cluster. When Vast scaling is adopted, the first cluster appears to find different dynamics with respect to the rest of iterations, and a great improvement for $RRMSE_{mean}$ is observed from the second iteration onward.

After assessing the contribution for the different scaling methods, the new iterative formulation of the algorithm was compared with the standard HODMD. To accomplish this task, the reconstruction error relative improvement (η) between the two different algorithms is analyzed, which is defined as follows:

$$\eta = \frac{RRMSE_{HODMD} - RRMSE_{h-HODMD}}{RRMSE_{HODMD}}, \quad (15)$$

Equation (15) indicates that if a given variable is better reconstructed by h-HODMD than by standard HODMD, a positive value of the improvement is obtained, whereas the opposite occurs when the aforementioned quantity is negative.

The improvement factor is shown in Fig. 5 for all databases and scaling methods. By adopting Auto scaling, the worst results are achieved as the reconstruction error gets worse for most of the species. The use of Range scaling instead improves the reconstruction error for most of the variables, and the same happens for Pareto scaling. Thus, these two scaling methods appear to be a better choice when h-HODMD is adopted. Lastly, Vast scaling appears to favor the iterative algorithm as the transient behavior decreases (e.g., as when moving from database 2 to database 3).

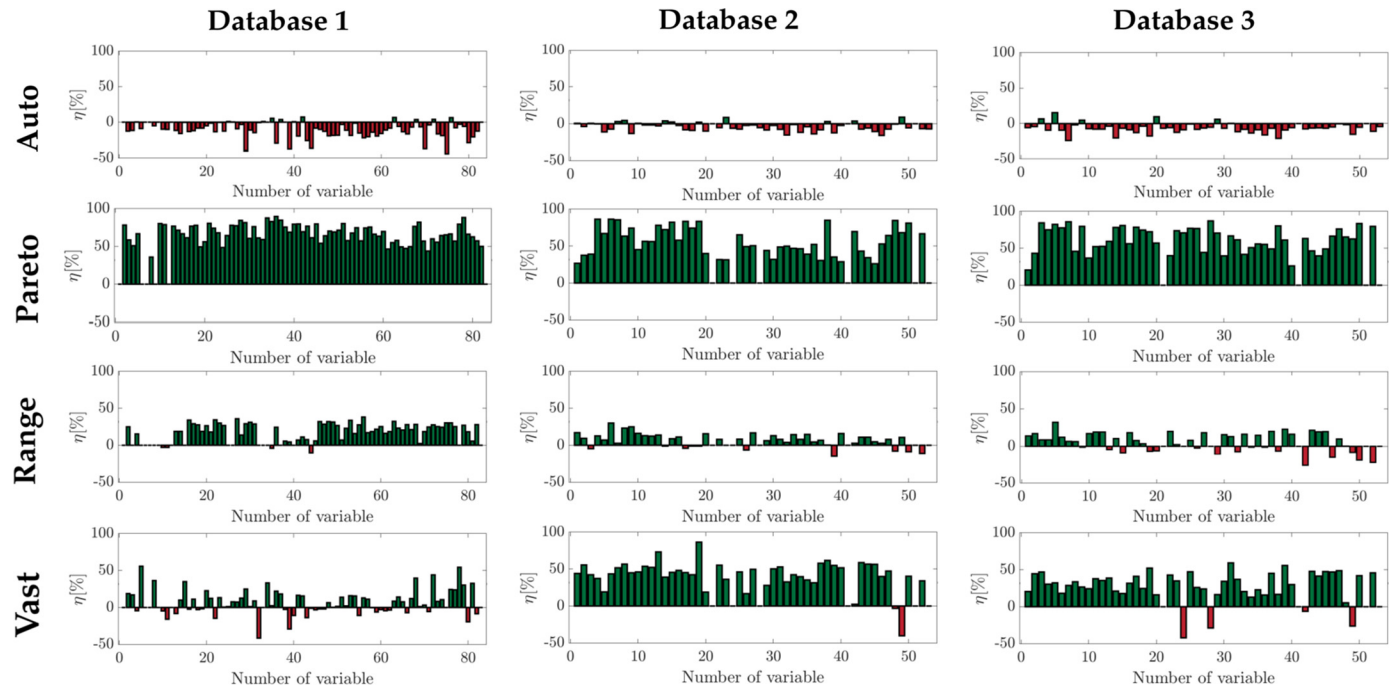


Fig. 5. Reconstruction error improvement for all the variables of the chemical system computed as reported in Equation (15) for Database 1 (simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 82 species), Database 2 (simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species), Database 3 (simulation obtained by imposing a sinusoidal perturbation having frequency $f = 20$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species), and considering different scaling criteria (from top to bottom: Auto, Range, Pareto, Vast).

Table 4

Clusters of variables obtained using h-HODMD for Database 3 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 20$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using range scaling for the preprocessing step.

Iteration 1	Iteration 2	Iteration 3	Iteration 5	Iteration 7
CH_4	CO	HO_2	CH_3	H_2O_2
CO_2	N_2O	O		
H_2O	NO	OH		
N_2	NO_2			
O_2				
T				

4.1. HODMD feature selection

As discussed in Section 2, the new iterative formulation for HODMD is capable of achieving a lower reconstruction error by grouping different variables in each iteration, i.e., applying a model reduction considering clusters of variables instead of the full input tensor.

The groups of variables that are found by the iterative algorithm for the third database considering Range scaling are listed in Table 4, while all the other cases (in terms of databases and scaling criteria) are made available to the reader in the supplemental material. In this work, only the most relevant species from a physical point of view were taken into account to perform the variables' clusters analysis; therefore, in Table 4 only temperature, fuel, oxidizer, main oxidation products, main initiation and propagation radicals, and main nitrogen oxides are reported. The tolerance ϵ_v was set equal to $\epsilon_v = tol \times iter$, where the value of tol was set equal to the RRMSE of the sixth-best reconstructed variable, so the first cluster should always contain the main variables of the reacting system (i.e., temperature, fuel, oxidizer, and the main oxidation products such as water and carbon dioxide). If an iteration

of the algorithm does not report any variable is then removed, and the algorithm runs the next iteration until there are no variables left, so in this analysis the number of iterations is not fixed.

Pareto scaling always fails to identify groups of variables that are coherent from a chemical point of view, as seen in Table 4 and in Tabs. SI to SVII of the supplemental material. In fact, O and OH are placed in different clusters, and they are grouped instead with nitrogen oxides, which have much slower dynamics. Additionally, it also requires the maximum number of iterations to identify and cluster all the selected variables. On the other hand, the application of Vast, and Range scaling yields to better results not only from a reconstruction error point of view, but also because of the possibility to have physics-based clusters consisting of species whose kinetic behavior is similar. In fact, the aforementioned scaling criteria result in two different clusters separating species having slow dynamics (such as nitrogen oxides) from fast dynamics (radicals responsible for propagation and branching reactions, such as oxygen radicals and those involved in initiation reactions, such as the methane radical). As shown in Table 4, the three main nitrogen oxides (NO, NO₂, and N₂O) are grouped when Range scaling is used, while another cluster is identified for oxy- and peroxy radicals (O, OH, and HO₂). Also, this scaling only needs seven iterations to get this result, while nine iterations are needed when Vast scaling is employed.

As can be observed in Fig. 6, range scaling allows for a significant improvement (~40%) in the reduced model accuracy for the description of radicals that are fundamental for the reacting layer dynamics and liftoff height prediction, such as OH, O, H₂O₂, and CH₃. Also, from a qualitative comparison, such as the one reported in Fig. 7 where the original (left side of each figure) and the reconstructed (right side of each figure) variables' profiles are compared, it is clear that the error observed when the proposed algorithm is adopted is negligible.

Finally, the two most relevant modes for temperature and nitrogen oxides are reported in Fig. 8. In particular, the left side of each figure

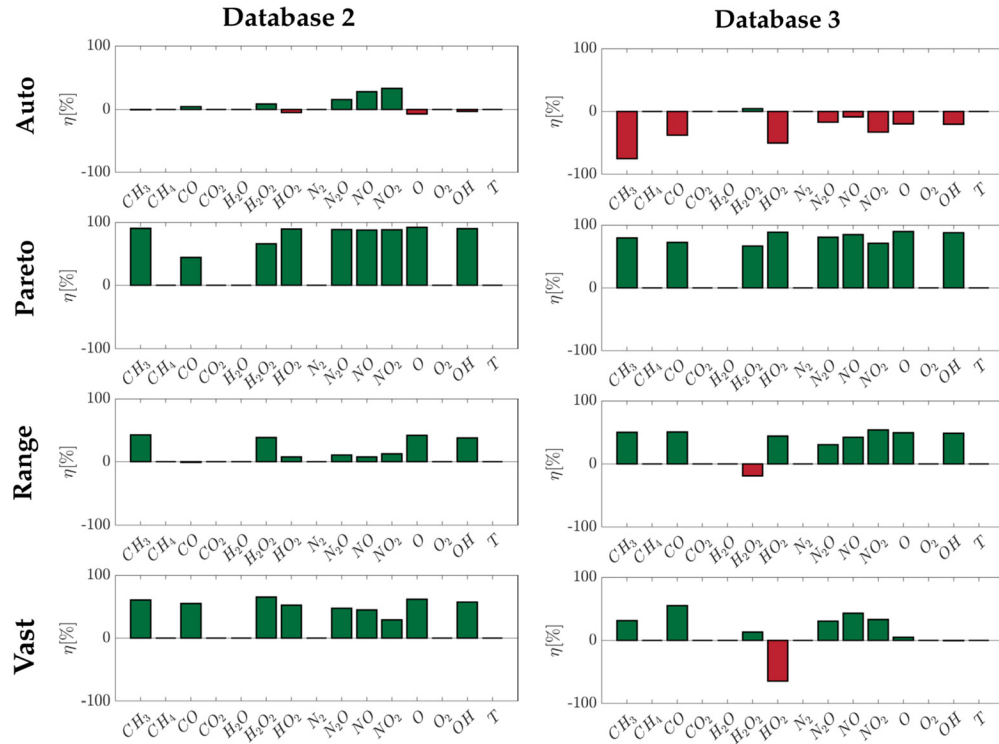


Fig. 6. Reconstruction error improvement for the main variables of the chemical system computed as reported in Equation (15) for Database 2 (simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species), and Database 3 (simulation obtained by imposing a sinusoidal perturbation having frequency $f = 20$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species), considering different scaling criteria (from top to bottom: Auto, Range, Pareto, Vast).

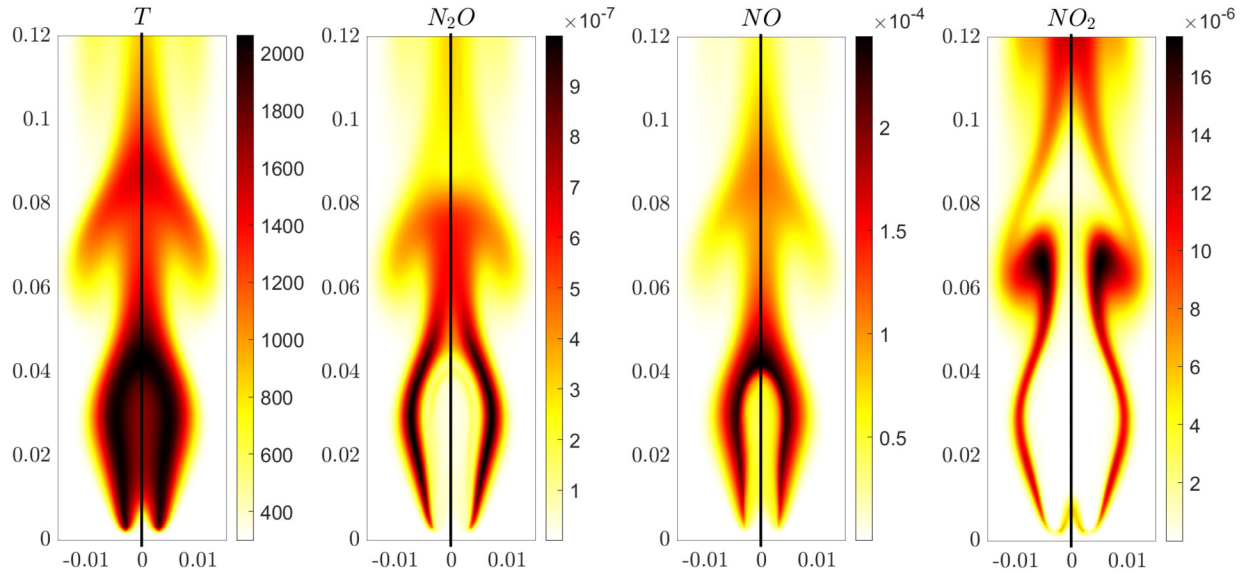


Fig. 7. Reconstruction of a representative snapshot of the flame for different variables using h-HODMD (Range scaling). The right side of each contour represents the original snapshot, while the left side of each contour represents the reconstruction with h-HODMD. Figures from left to right: contours of temperature, contours of NO, contours of NO₂, contours of N₂O.

is representative for $\omega_1 \sim 126$, which is the frequency of perturbation of the simulation ($f_1 = \omega_1/2\pi \approx 20$). This mode has, for all species, a dominant component along the streamwise direction due to the effects of the sinusoidal velocity perturbation. Also, the modes for the nitrogen oxides are non-negative in the regions where these chemical species are formed and have the maximum concentration (i.e., in the regions where the reaction terms are stronger), while negative values are observed where the transport term is dominant and no reaction is oc-

curing. On the right side of each contour in Fig. 8, the frequency with highest energy $\omega_2 = 25$ is shown. This particular frequency was selected as it is one subharmonic of the main frequency ($\omega_2 = \omega_1/5$) thus it is expected to have larger structures. For all the variables considered, it is possible to observe that the mode is focused on the large flow structures upstream and downstream, and it also has a peak in the liftoff and the flame border regions, which are due to the strong gradients (both in the streamwise and in the spanwise directions) they undergo.

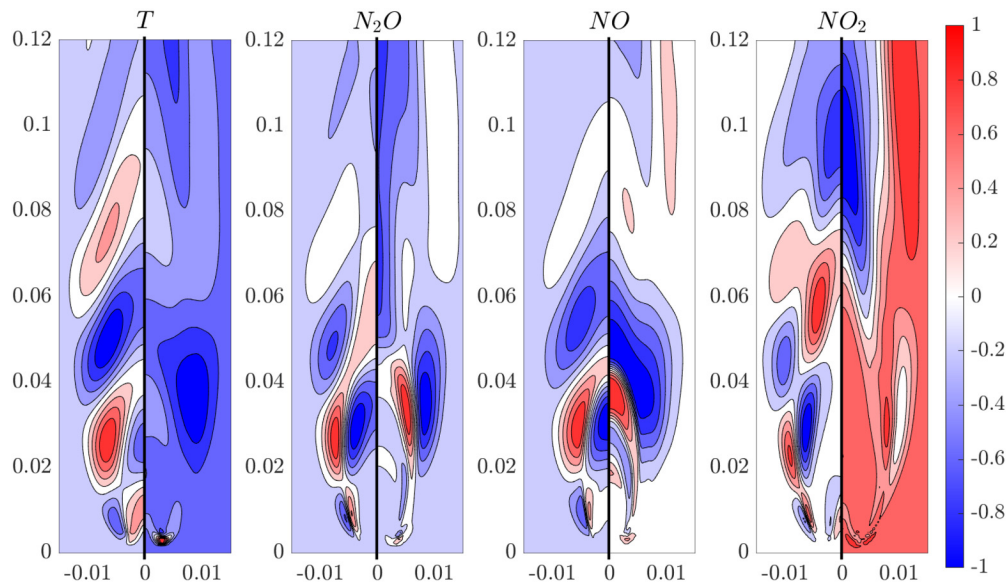


Fig. 8. Real part of the two most important DMD modes (**left:** $\omega \sim 126$, **right:** $\omega \sim 25$) for some of the most relevant species using h-HODMD with range scaling criteria. The modes have been normalized between -1 and 1 .

5. Conclusions

The article introduces an innovative iterative algorithm called hierarchical HODMD (h-HODMD) for constructing reduced order models for reacting flows. The algorithm's novelty lies in its ability to perform model reduction by grouping variables into clusters, leading to reduced error in the resulting model. Furthermore, for each cluster, it identifies a dominant dynamics that characterizes the group of variables under consideration. This technique proves to be highly effective in analyzing reacting flows, where a large number of variables are typically involved in the corresponding computational fluid dynamics (CFD) simulations.

Using modes and their frequencies, the algorithm can be adapted to decouple different dynamics in combustion systems. h-HODMD iteratively identifies and isolates the best-reconstructed variables by computing the relative root mean square error (RRMSE) for each variable and incorporating a user-defined tolerance. As a result, in addition to its application in reduced order modeling, the algorithm can also be utilized for feature selection and variable clustering tasks.

The new algorithm has been successfully applied to simulations of reactive flows using different scaling methods (Auto, Pareto, Vast and Range scaling), and the results have been compared with HODMD. The Auto scaling method aims to ensure that all clusters exhibit the same dynamics. However, it does not yield improved reconstruction compared to standard HODMD. In the case of Pareto scaling, the first cluster exhibits distinct frequencies compared to the rest. Furthermore, it significantly improves the reconstruction error compared to HODMD. Vast scaling demonstrates similar behavior to Pareto in terms of the dominant dynamics. However, it only shows notable improvements in the second and third databases. Finally, Range scaling leads to clusters with different dynamics, resulting in improvements for the majority of the variables.

This new iterative algorithm for higher order dynamic mode decomposition has shown promise in analyzing high-dimensional data, such as those obtained from reacting flow simulations conducted with detailed kinetic mechanisms. Specifically, the method proposed for feature selection and variable clustering can be valuable in the development of novel reduced order models informed by the underlying physics. This approach enables the exploration of new strategies to tackle increasingly large databases while reducing the associated computational costs.

Data availability

Data will be made available on request.

Acknowledgements

A.C. and S.L.C. acknowledge the grant PID2020-114173RB-I00 funded by MCIN/AEI/ 10.13039/501100011033. S.L.C. and A.C. acknowledge the support of Comunidad de Madrid through the call Research Grants for Young Investigators from Universidad Politécnica de Madrid. A.C. also acknowledges the support of Universidad Politécnica de Madrid, under the programme 'Programa Propio'. G.D.A. has received funding from the Fonds National de la Recherche Scientifique (FRS-FNRS) through a FRIA fellowship. A.P. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 714605.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.camwa.2024.01.003>.

References

- [1] Stewart Cant, High-performance computing in computational fluid dynamics: progress and challenges, *Philos. Trans. R. Soc. Lond. A, Math. Phys. Eng. Sci.* 360 (1795) (2002) 1211–1225.
- [2] Joan Baiges, Ramon Codina, Sergio Idelsohn, Explicit reduced-order models for the stabilized finite element approximation of the incompressible Navier–Stokes equations, *Int. J. Numer. Methods Fluids* 72 (12) (2013) 1219–1243.
- [3] Stefano Lorenzi, Antonio Cammi, Lelio Luzzi, Gianluigi Rozza, Pod-Galerkin method for finite volume approximation of Navier–Stokes and rans equations, *Comput. Methods Appl. Mech. Eng.* 311 (2016) 151–179.
- [4] Efthymios N. Karatzas, Giovanni Stabile, Leo Nouveau, Guglielmo Scovazzi, Gianluigi Rozza, A reduced basis approach for pdes on parametrized geometries based on the shifted boundary finite element method and application to a Stokes flow, *Comput. Methods Appl. Mech. Eng.* 347 (2019) 568–587.
- [5] Joseph Lorenzetti, Andrew McClellan, Charbel Farhat, Marco Pavone, Linear reduced-order model predictive control, *IEEE Trans. Autom. Control* 67 (11) (2022) 5980–5995.
- [6] Alessandro Parente, J.C. Sutherland, Bassam B. Dally, Leonardo Tognotti, P.J. Smith, Investigation of the mild combustion regime via principal component analysis, *Proc. Combust. Inst.* 33 (2) (2011) 3333–3341.

- [7] Aurélie Bellemans, Gianmarco Aversano, Axel Coussement, Alessandro Parente, Feature extraction and reduced-order modelling of nitrogen plasma models using principal component analysis, *Comput. Chem. Eng.* 115 (2018) 504–514.
- [8] Axel Coussement, Olivier Gicquel, Alessandro Parente, MG-local-PCA method for reduced order combustion modeling, *Proc. Combust. Inst.* 34 (1) (2013) 1117–1123.
- [9] Ian Jolliffe, *Principal Component Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1094–1096.
- [10] Gal Berkooz, Philip Holmes, John L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annu. Rev. Fluid Mech.* 25 (1) (1993) 539–575.
- [11] Benjamin J. Isaac, Axel Coussement, Olivier Gicquel, Philip J. Smith, Alessandro Parente, Reduced-order PCA models for chemical reacting flows, *Combust. Flame* 161 (11) (2014) 2785–2800.
- [12] Mohammad Rafi Malik, Pedro Obando Vega, Axel Coussement, Alessandro Parente, Combustion modeling using principal component analysis: a posteriori validation on Sandia flames d, e and f, *Proc. Combust. Inst.* 38 (2) (2021) 2635–2643.
- [13] Giuseppe D'Alessio, Sankaran Sundaresan, Michael E. Mueller, Automated and efficient local adaptive regression for principal component-based reduced-order modeling of turbulent reacting flows, *Proc. Combust. Inst.* (2022).
- [14] Anuj Kumar, Martin Rieth, Opeoluwa Owoyele, Jacqueline H. Chen, Tarek Echekki, Acceleration of turbulent combustion dns via principal component transport, *Combust. Flame* 255 (2023) 112903.
- [15] Aurélie Bellemans, Thierry Magin, Axel Coussement, Alessandro Parente, Reduced-order kinetic plasma models using principal component analysis: model formulation and manifold sensitivity, *Phys. Rev. Fluids* 2 (7) (2017) 073201.
- [16] Giuseppe D'Alessio, Alessandro Parente, Alessandro Stagni, Alberto Cuoci, Adaptive chemistry via pre-partitioning of composition space and mechanism reduction, *Combust. Flame* 211 (2020) 68–82.
- [17] Giuseppe D'Alessio, Alberto Cuoci, Gianmarco Aversano, Mauro Bracconi, Alessandro Stagni, Alessandro Parente, Impact of the partitioning method on multidimensional adaptive-chemistry simulations, *Energies* 13 (10) (2020) 2567.
- [18] Giuseppe D'Alessio, Antonio Attili, Alberto Cuoci, Heinz Pitsch, Alessandro Parente, Analysis of turbulent reacting jets via principal component analysis, in: *Data Analysis for Direct Numerical Simulations of Turbulent Combustion*, Springer, 2020, pp. 233–251.
- [19] Giuseppe D'Alessio, Antonio Attili, Alberto Cuoci, Heinz Pitsch, Alessandro Parente, Unsupervised data analysis of direct numerical simulation of a turbulent flame via local principal component analysis and Procrustes analysis, in: *International Workshop on Soft Computing Models in Industrial and Environmental Applications*, Springer, 2020, pp. 460–469.
- [20] S. Tipler, G. D'Alessio, Q. Van Haute, A. Parente, Francesco Contino, A. Coussement, Predicting octane numbers relying on principal component analysis and artificial neural network, *Comput. Chem. Eng.* 161 (2022) 107784.
- [21] Giuseppe D'Alessio, Alberto Cuoci, Alessandro Parente, Feature extraction and artificial neural networks for the on-the-fly classification of high-dimensional thermochemical spaces in adaptive-chemistry simulations, *Data-Cent. Eng.* 2 (2021).
- [22] Peter J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (2010) 5–28.
- [23] Souvick Chatterjee, Achintya Mukhopadhyay, 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 Thermoacoustics*, 2013, p. 13.
- [24] John M. Quinlan, T. Ben Zinn, *Transverse Combustion Instabilities: Modern Experimental Techniques and Analysis*, 2014.
- [25] Cheng Huang, William E. Anderson, Matthew E. Harvazinski, Venkateswaran Sankaran, Analysis of self-excited combustion instabilities using decomposition techniques, *AIAA J.* 54 (9) (2016) 2791–2807.
- [26] Emmanuel Motheau, Franck Nicoud, Thierry Poinsot, Mixed acoustic-entropy combustion instabilities in gas turbines, *J. Fluid Mech.* 749 (2014) 542–576.
- [27] Soledad Le Clainche, José M. Vega, Analyzing nonlinear dynamics via data-driven dynamic mode decomposition-like methods, *Complexity* (2018) 2018.
- [28] Soledad Le Clainche, José M. Vega, Higher order dynamic mode decomposition, *SIAM J. Appl. Dyn. Syst.* 16 (2) (2017) 882–925.
- [29] Adrián Corrochano, Donnatella Xavier, Philipp Schlatter, Ricardo Vinuesa, Soledad Le Clainche, Flow structures on a planar food and drug administration (fda) nozzle at low and intermediate Reynolds number, *Fluids* 6 (1) (2021) 4.
- [30] Eneko Lazpita, Álvaro Martínez-Sánchez, Adrián Corrochano, Sergio Hoyas, Soledad Le Clainche, Ricardo Vinuesa, On the generation and destruction mechanisms of arch vortices in urban fluid flows, *Phys. Fluids* 34 (5) (2022) 051702.
- [31] Soledad Le Clainche, Rubén Moreno-Ramos, Paul Taylor, José M. Vega, New robust method to study flight flutter testing, *J. Aircr.* 56 (1) (2019) 336–343.
- [32] S. Le Clainche, Daulet Izbassarov, M. Rosti, Luca Brandt, Outi Tammsola, Coherent structures in the turbulent channel flow of an elastoviscoplastic fluid, *J. Fluid Mech.* 888 (2020).
- [33] Adrián Corrochano, Giuseppe D'Alessio, Alessandro Parente, Soledad Le Clainche, Higher order dynamic mode decomposition to model reacting flows, *Int. J. Mech. Sci.* 249 (2023) 108219.
- [34] Soledad Le Clainche, Prediction of the optimal vortex in synthetic jets, *Energies* 12 (9) (2019).
- [35] Carlos Mendez, Soledad Le Clainche, Rubén Moreno, José Vega, A new automatic, very efficient method for the analysis of flight flutter testing data, *Aerosp. Sci. Technol.* 04 (2021).
- [36] Soledad Le Clainche, Zhong-Hua Han, Esteban Ferrer, An alternative method to study cross-flow instabilities based on high order dynamic mode decomposition, *Phys. Fluids* 31 (9) (2019) 094101.
- [37] Jose Manuel Vega, Soledad Le Clainche, *Higher Order Dynamic Mode Decomposition and Its Applications*, Academic Press, 2020.
- [38] Hemanth Kolla, Konduri Aditya, Jacqueline H. Chen, *Higher Order Tensors for DNS Data Analysis and Compression*, Springer International Publishing, Cham, 2020, pp. 109–134.
- [39] R. Ledyard Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika* 31 (3) (1966) 279–311.
- [40] Jiaqing Kou, Soledad Le Clainche, Weiwei Zhang, A reduced-order model for compressible flows with buffeting condition using higher order dynamic mode decomposition with a mode selection criterion, *Phys. Fluids* 30 (1) (2018) 016103.
- [41] Alessandro Parente, James C. Sutherland, Principal component analysis of turbulent combustion data: data pre-processing and manifold sensitivity, *Combust. Flame* 160 (2) (2013) 340–350.
- [42] Seth M. Hirsh, Kameron Decker Harris, J. Nathan Kutz, Bingni W. Brunton, Centering data improves the dynamic mode decomposition, *SIAM J. Appl. Dyn. Syst.* 19 (3) (2020) 1920–1955.
- [43] Jun Yang, Xinjie Zhao, Xin Lu, Xiaohui Lin, Guowang Xu, A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis, *Front. Mol. Biosci.* 2 (2015).
- [44] Robert A. van den Berg, Huub C.J. Hoefsloot, Johan A. Westerhuis, Age K. Smilde, Mariët J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, *BMC Genomics* 7 (1) (2006).
- [45] Alex Viguier, Gabriel F. Barros, Malú Grave, Alessandro Reali, Alvaro L.G.A. Coutinho, Coupled and uncoupled dynamic mode decomposition in multi-compartmental systems with applications to epidemiological and additive manufacturing problems, *Comput. Methods Appl. Mech. Eng.* 391 (2022) 114600.
- [46] Eliseo Ranzi, Alessio Frassoldati, Roberto Grana, Alberto Cuoci, Tiziano Faravelli, Andrew P. Kelley, Chung K. Law, Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels, *Prog. Energy Combust. Sci.* 38 (4) (2012) 468–501.
- [47] Gregory P. Smith, *Gri-mech 3.0*, http://www.me.berkeley.edu/gri_mech/, 1999.
- [48] Alberto Cuoci, Alessio Frassoldati, Tiziano Faravelli, Eliseo Ranzi, Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method, *Energy Fuels* 27 (12) (2013) 7730–7753.

Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 8	Iteration 9
CH_4	HO_2	CO	CH_3	H_2O_2	NO_2
CO_2		N_2O	O		
H_2O		NO	OH		
N_2					
O_2					
T					

Table 1: Clusters of variables obtained using h-HODMD for Database 2 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using auto scaling for the preprocessing step.

Iteration 1	Iteration 3	Iteration 4	Iteration 5	Iteration 6	Iteration 9
CH_4	CO	HO_2	NO	CH_3	NO_2
CO_2		O	OH		
H_2O					
N_2					
O_2					
T					
Iteration 14	Iteration 15				
H_2O_2	N_2O				

Table 2: Clusters of variables obtained using h-HODMD for Database 2 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using pareto scaling for the preprocessing step.

Iteration 1	Iteration 4	Iteration 5	Iteration 6	Iteration 14
CH_4	CO	N_2O	CH_3	H_2O_2
CO_2	HO_2	NO	O	
H_2O	NO_2		OH	
N_2				
O_2				
T				

Table 3: Clusters of variables obtained using h-HODMD for Database 2 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using range scaling for the preprocessing step.

Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 7
CH_4	CO	N_2O	CH_3	H_2O_2
CO_2	HO_2	NO	OH	
H_2O		NO_2		
N_2		O		
O_2				
T				

Table 4: Clusters of variables obtained using h-HODMD for Database 2 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 10$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using vast scaling for the preprocessing step.

Iteration 1	Iteration 3	Iteration 5	Iteration 6	Iteration 10
CH_4	CO	NO_2	CH_3	H_2O_2
CO_2	N_2O		HO_2	
H_2O	NO		O	
N_2			OH	
O_2				
T				

Table 5: Clusters of variables obtained using h-HODMD for Database 3 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 20$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using auto scaling for the preprocessing step.

Iteration 1	Iteration 2	Iteration 3	Iteration 5	Iteration 7
CH_4	CO	CH_3	NO_2	H_2O_2
CO_2	HO_2			
H_2O	N_2O			
N_2	NO			
O_2	O			
T	OH			

Table 6: Clusters of variables obtained using h-HODMD for Database 3 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 20$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using pareto scaling for the preprocessing step.

Iteration 1	Iteration 2	Iteration 3	Iteration 8	Iteration 9
CH_4	CO	N_2O	HO_2	CH_3
CO_2		NO	O	H_2O_2
H_2O		NO_2	OH	
N_2				
O_2				
T				

Table 7: Clusters of variables obtained using h-HODMD for Database 3 (i.e., for the simulation obtained by imposing a sinusoidal perturbation having frequency $f = 20$ Hz and amplitude $A = 0.25$, and carried out with a detailed kinetic mechanism accounting for 53 species) and using vast scaling for the preprocessing step.

Chapter 5

Structural sensitivity in laminar flames using data-driven methods

5.1 Structural sensitivity in non-linear flows using direct solutions

5.1.1 Publication information

Title: Structural sensitivity in non-linear flows using direct solutions

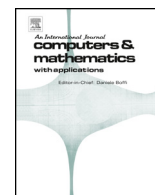
DOI: <https://doi.org/10.1016/j.camwa.2022.10.006>

Journal: Computers & Mathematics with Applications

Date: 15 December 2022

Publisher: Elsevier

Copyright: © 2022 Elsevier Ltd. All rights reserved.



Structural sensitivity in non-linear flows using direct solutions

Adrián Corrochano, Soledad Le Clainche*

School of Aerospace Engineering, Universidad Politécnica de Madrid, E-28040, Spain

ARTICLE INFO

Keywords:

Structural sensitivity
Data-driven methods
Flow patterns
Flow control
DMD
HODMD

ABSTRACT

This article introduces a novel method to calculate the structural sensitivity in non-linear flows. The method is completely data-driven and only uses direct solutions, avoiding the need to calculate the adjoint solution. Moreover, the method can be directly applied to calculate the structural sensitivity in three-dimensional flows. Depending on the nature and the complexity of the data analysed, the methodology presented to calculate this so-called *non-linear* structural sensitivity, uses dynamic mode decomposition (DMD) or higher order DMD to identify the global modes leading the flow dynamics. The performance of this novel technique has been tested in the two-dimensional wake of a circular cylinder at Reynolds number (Re) 50, 60, 85, 100, 130 and 160, the three-dimensional wake of a circular cylinder at Re = 280 and spanwise wavelength $L_z = 1.67$ and 6.99, in a Food and Drug Administration nozzle at Re = 800 and in two-concentric jets with similar velocity ratio at Re = 1500. The results have been compared with the structural sensitivity predicted by the linear theory, finding some similarities and discrepancies between both methods. Nevertheless, the highest structural sensitivity regions identified by this novel algorithm are in good agreement with the literature, where the flow instabilities have been controlled using different mechanisms of passive flow control located in the regions predicted by the non-linear structural sensitivity calculated in this article.

1. Introduction

Optimization and flow control in fluid dynamics is a research topic of high interest due to its multiple industrial applications (e.g. to improve the efficiency in industrial devices, wind turbines, aircraft, et cetera). Modelling complex flows (turbulent, noisy, multi-scale) to assess the performance of different flow control strategies is a challenging problem due to the need of large computational resources and the complexity of the underlying physical problem [1].

Structural sensitivity [2] is a strategy generally used for flow control, and also for a variety of applications, from stability theory to feedback analysis of actuator/sensor placement, among others. This concept identifies the regions of the flow most sensitive to generic structural modifications, which produce the strongest drift of the leading mode driving the flow dynamics. It is notorious that structural sensitivity can be used for any model, and do not necessarily need to be restricted to problems in fluid dynamics (e.g. see some examples of solid mechanics in Ref. [3], or see details in Ref. [4]).

Hence, computing the structural sensitivity in fluid dynamics will provide some ideas about what are the regions of the flow that should be modified to control (advancing or delaying) the main flow instabilities

driving the flow motion [5]. To calculate the structural sensitivity maps, it is necessary to combine direct and adjoint eigenfunctions, generally obtained by means of solving a direct and adjoint linear stability problem. Solving these two stability problems to model realistic flows could be costly in terms of memory and computational time, especially in three-dimensional databases. Moreover, it requires a priori knowledge of the underlying equations driving the flow, which could be considered a limitation to calculate the structural sensitivity on experimental databases.

This article introduces a novel method to calculate the structural sensitivity in fluid dynamics based on modal decompositions [6]. The advantage of the present methodology is that it is fully data-driven, it can be applied similarly in two- and three-dimensional databases concerning non-linear flows, and it only considers the direct solution of the Navier-Stokes equations. Hence, the present work can easily be extended to experimental measurements, three-dimensional and turbulent flows. Luchini et al. [7] presented an extension of the structural sensitivity to limit cycle (periodic solutions), where the adjoint modes of the Floquet transition operator were used to calculate the structural sensitivity of a periodic state in a circular cylinder. Also the work by Marquet & Lesshafft [8] follows the classical stability theory to propose

* Corresponding author.

E-mail address: soledad.leclainche@upm.es (S. Le Clainche).

an alternative definition for identifying wavemakers regions in periodic flows, which was also tested in the wake of a two-dimensional circular cylinder. However, further research should be carried out to develop novel techniques capable to identify the structural sensitivity in turbulent flows or in experimental databases. If the structural sensitivity is known using data-driven methods, either passive or active mechanisms of flow control could be applied to control the flow in experimental databases or in turbulent flows in a fast and efficient way. This article pursues this objective using some concepts that are based on adjoint equations, linear stability analysis and the definition of structural sensitivity presented in the literature [4], but introducing a completely novel approach, fully data-driven.

It is worth to mention that this novel approach introduced in the article, called as *non-linear structural sensitivity*, is expected to present some differences compared to the results obtained using the classical linear theory. Nevertheless, as in the classical approach, this method provides information related to the areas of the flow most sensitive to structural modifications (as it is presented in various examples). The main differences between these two methodologies are the following: (i) the linear theory is based on the base flow while the present method applies to non-linear solutions and is based on the mean flow, (ii) in the linear theory the structural sensitivity is based on a structural perturbation only acting on time-varying disturbances, while in the present approach for non-linear solutions, the structural perturbation acts in both the mean flow and the disturbance (as in experimental results). Although the purposes of these two methodologies are different (linear and non-linear flows), they do not contradict each other.

The article is organized as follows. Section 2 introduces the new methodology and the main results and conclusions are presented in Section 3 and Section 4.

2. Methodology

The continuity and Navier-Stokes equations are the governing equations describing the motion of viscous Newtonian flows, defined in nonlinear form for incompressible flows as:

$$\nabla \cdot \mathbf{v} = 0, \quad (1)$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla p + \frac{1}{\text{Re}} \Delta \mathbf{v}, \quad (2)$$

where $\mathbf{v}$ and p are the non-dimensional velocity vector and pressure, respectively, and Re is the Reynolds number, defined as $\text{Re} = VL/\nu$, where ν is the kinematic viscosity of the fluid, V is the characteristic flow velocity, and L is the characteristic length. Depending on the problem under study, both the velocity and the characteristic length are different. Note that the equations (1)–(2) are non-dimensionalized with the scales of time and length, L/V and L , respectively.

The solution of these equations (for simplicity with the remaining of the article, the solution is particularized for the velocity vector) can be expressed as a Fourier-like expansion of M modes defined in discrete form for the time t_k as

$$\mathbf{v}(\mathbf{x}, \mathbf{y}, \mathbf{z}, t_k) = \mathbf{v}_k \simeq \tilde{\mathbf{v}}_k = \sum_{m=1}^M a_m \mathbf{u}_m(\mathbf{x}, \mathbf{y}, \mathbf{z}) e^{(\delta_m + i\omega_m)t_k}, \quad k = 1, \dots, K, \quad (3)$$

where $\mathbf{u}_m$ are the modes and ω_m , δ_m and a_m are the mode frequencies and growth rates, and amplitudes weighting such modes, respectively. Among others, the previous expansion can be calculated by means of Fourier analysis (i.e.: fast Fourier transform) or it can also be approximated using the well known technique Dynamic Mode Decomposition (DMD) [9] or higher order DMD (HODMD) [10], a robust extension of DMD to analyse complex flows (i.e.: turbulence, multi-scale, noise). Similarly, for continuous time, considering a data set formed by the snapshot vector $\mathbf{v}$, it is also possible to approximate the simplest problem giving the dynamics in the right hand side of the previous expansion (3) by the solution of an ODE (ordinary differential equation) linear system $\frac{d\mathbf{v}}{dt} = \mathbf{A}\mathbf{v}$, where the modes $\mathbf{u}_m$ are the eigenvectors of $\mathbf{A}$ and ω_m

and δ_m are the imaginary and real components of the eigenvalues of $\mathbf{A}$. However, not all the expansions of (3) are the solution of this ODE linear system, because the number of modes provided by this expansion is limited to the spatial dimension (at most) of the data analysed (the spatial dimension is given by dimension of the span of the modes $\mathbf{u}_m$). Hence, this expansion could be generalized as the solution of a high order ODE linear system $\frac{d^d \mathbf{v}}{dt^d} = \mathbf{A}_1 \frac{d^{d-1} \mathbf{v}}{dt^{d-1}} + \mathbf{A}_2 \frac{d^{d-2} \mathbf{v}}{dt^{d-2}} + \dots + \mathbf{A}_{d-1} \frac{d \mathbf{v}}{dt} + \mathbf{A}_d \mathbf{v}$. The discretized version of these two ODE linear systems can be related to DMD (the low order system) and HODMD (the high order system) methods, see more details in Ref. [11].

Generally, the first mode (with the highest amplitude, $m = 1$) of the expansion (3) is steady, corresponding to the mean flow, that is $\mathbf{u}_1 = \bar{\mathbf{v}}$. If the flow analysed is linear, the second highest amplitude mode ($m = 2$) would correspond to the eigenmodes calculated using the linear stability theory [12], defined as $\mathbf{u}_2 e^{(\delta_2 + i\omega_2)t}$ and representing the flow fluctuations ($\mathbf{v}'$), which are leading the flow changes. Similarly, in periodic (non-linear) flow solutions, the second highest amplitude mode would also correspond to the linear stability mode (see details in Ref. [13]).

Following the Reynolds decomposition, the solution of the Navier-Stokes equation can be decomposed as a temporal averaged field plus a fluctuating field as $\mathbf{v} = \bar{\mathbf{v}} + \mathbf{v}'$ and $p = \bar{p} + p'$. Introducing this decomposition into equations (1)–(2), it yields

$$\nabla \cdot \mathbf{v}' = 0,$$

$$(4)$$

$$\frac{\partial \mathbf{v}'}{\partial t} + (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}} + (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}' + (\mathbf{v}' \cdot \nabla) \bar{\mathbf{v}} + (\mathbf{v}' \cdot \nabla) \mathbf{v}' = -\nabla(\bar{p} + p') + \frac{1}{\text{Re}} \Delta(\bar{\mathbf{v}} + \mathbf{v}'). \quad (5)$$

We recall that $\bar{\mathbf{v}}$ is independent of time. Moreover, regarding the Reynolds decomposition introduced in the continuity equation, it yields

$$\nabla \cdot (\bar{\mathbf{v}} + \mathbf{v}') = \nabla \cdot \bar{\mathbf{v}} + \nabla \cdot \mathbf{v}' = 0. \quad (6)$$

Averaging this equation, and taking into account the property of the mean of the fluctuating field defined as $\bar{\mathbf{v}'} = 0$, it yields $\nabla \cdot \bar{\mathbf{v}} = 0$ (see details in, e.g. Ref. [14]), and thus, it follows that eq. (4) also holds.

Let us re-name the convective terms in (5) as $\mathbf{M} = (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}}$, $\mathbf{L} = (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}' + (\mathbf{v}' \cdot \nabla) \bar{\mathbf{v}}$ and $\mathbf{N} = (\mathbf{v}' \cdot \nabla) \mathbf{v}'$, which correspond to a mean or averaged field ($\mathbf{M}$), a linear term ($\mathbf{L}$) and a non-linear term ($\mathbf{N}$). In classical linear stability theory, the non-linear terms are neglected to solve the linearized Navier-Stokes equations and the mean term does not appear in the equation because it balances out with the other terms related to the mean flow, since the mean flow satisfies the steady Navier-Stokes equations. Nevertheless, the present article introduces a data-driven method to approximate the structural sensitivity of the flow in non-linear solutions where the two terms $\mathbf{M}$ and $\mathbf{N}$ will recover relevance.

In Cartesian notation, the velocity vector is defined as $\mathbf{v} = (v_1, v_2, v_3)$ (corresponding to the streamwise v_x , normal v_y and spanwise v_z velocity components), connected with the spatial coordinates $\mathbf{x} = (x_1, x_2, x_3) = (x, y, z)$. Following this notation, the mean and non-linear terms can be expressed as

$$\mathbf{M} = \bar{v}_j \frac{\partial \bar{v}_i}{\partial x_j}, \quad (7)$$

$$\mathbf{N} = v'_j \frac{\partial v'_i}{\partial x_j}, \quad \text{for } i = 1, 2, 3. \quad (8)$$

Following the Reynolds decomposition, the adjoint Navier-Stokes equations are defined as

$$-\frac{\partial \mathbf{v}^*}{\partial t} + (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}} + (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}^* + (\mathbf{v}^* \cdot \nabla) \bar{\mathbf{v}} + (\mathbf{v}^* \cdot \nabla) \mathbf{v}^* = -\nabla(\bar{p} + p^*) + \frac{1}{\text{Re}} \Delta(\bar{\mathbf{v}} + \mathbf{v}^*), \quad (9)$$

where $\mathbf{v}^*$ and p^* denotes the adjoint counterpart of $\mathbf{v}'$ and p' . As in the direct equations, it is possible to distinguish the three type of convective

terms as $\mathbf{M}^* = \mathbf{M} = (\bar{\mathbf{v}} \cdot \nabla) \bar{\mathbf{v}}$ (average), $\mathbf{L}^* = (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}^* + (\mathbf{v}^* \cdot \nabla) \bar{\mathbf{v}}$ (linear) and $\mathbf{N}^* = (\mathbf{v}^* \cdot \nabla) \mathbf{v}^*$ (non-linear). Classical linear stability theory, only considers the linear term. In Cartesian notation, the non-linear term can be expressed as

$$\mathbf{N}^* = v_j^* \frac{\partial v_i^*}{\partial x_j}, \text{ for } i = 1, 2, 3. \quad (10)$$

2.1. Structural sensitivity

Structural sensitivity and the wavemaker concept in the global framework in fluid dynamics was introduced by Giannetti & Luchini [2]. It identifies the regions of the flow where it is possible to produce the strongest drift of the leading linear stability eigenmode by performing generic structural modifications of the stability problem. The wavemaker region can be determined by introducing a forcing term $\mathbf{f}$ proportional to the global velocity mode into the linearized Navier-Stokes equations, defined here as

$$\nabla \cdot \mathbf{v}' = 0, \quad (11)$$

$$\frac{\partial \mathbf{v}'}{\partial t} + (\bar{\mathbf{v}} \cdot \nabla) \mathbf{v}' + (\mathbf{v}' \cdot \nabla) \bar{\mathbf{v}} = -\nabla p' + \frac{1}{\text{Re}} \Delta \mathbf{v}' + \mathbf{f}, \quad (12)$$

where the flow fluctuations $\mathbf{v}'$ are represented by the eigenmodal solution $\mathbf{u} e^{\sigma t}$ (global velocity mode), with $\sigma = \delta + i\omega$.

For a localised feedback process in (x_0, y_0) (in two-dimensional fields), which is proportional to the local flow speed, the forcing term can be mathematically described using a matrix operator $\mathbf{C}_0$ as

$$\mathbf{f} = \mathbf{C}_0 \delta(x - x_0, y - y_0) \mathbf{u}, \quad (13)$$

where $\delta(x - x_0, y - y_0)$ is the Dirac delta function. As shown by Giannetti & Luchini [2], variations of the eigenvalues due to a small-amplitude modification of the force $\delta\sigma(x_0, y_0)$, can then be represented as

$$|\delta\sigma(x_0, y_0)| \leq \|\mathbf{C}_0\| \cdot \|\mathbf{u}(x_0, y_0)\| \cdot \|\mathbf{u}^*(x_0, y_0)\|, \quad (14)$$

being $\mathbf{u}^*$ the adjoint global stability mode, $\|\cdot\|$ the Euclidean norm of a vector and $|\cdot|$ the modulus of a complex number. Hence, the structural sensitivity calculated in linear flows is defined as

$$\mathbf{S}^{LIN} = \|\mathbf{u}\| \cdot \|\mathbf{u}^*\|, \quad (15)$$

which represents the product of the magnitude of the direct and adjoint velocities (see more details of this definition in Ref. [15]).

This article approximates the structural sensitivity in non-linear flows, where the non-linear term $\mathbf{N}$ in eq. (8) plays a fundamental role in the flow dynamics influencing the structural sensitivity of the original differential problem, in contrast to the linear theory. Moreover, this theory is extended to three-dimensional flows, which is in line with the nature of non-linear flow solutions. In non-linear dynamics, the forcing term can be defined by the non-linear term of the Navier-Stokes equations (8) (see details in Ref. [16]). Following the linear theory, this term applied to a localised feedback process in (x_0, y_0, z_0) can be mathematically described using a matrix operator $\mathbf{C}_1$ as

$$\mathbf{f} = \mathbf{C}_1 \delta(x - x_0, y - y_0, z - z_0) (\mathbf{u} \cdot \nabla) \mathbf{u}. \quad (16)$$

Then, the variations of the eigenvalues due to a small-amplitude modification of the force, can be defined as

$$|\delta\sigma(x_0, y_0, z_0)| \leq \|\mathbf{C}_1\| \cdot \|(\mathbf{u}(x_0, y_0, z_0) \cdot \nabla) \mathbf{u}(x_0, y_0, z_0)\| \cdot \|(\mathbf{u}^*(x_0, y_0, z_0) \cdot \nabla) \mathbf{u}^*(x_0, y_0, z_0)\|. \quad (17)$$

Hence, the structural sensitivity in non-linear flows can then be defined considering the previous terms and the direct and adjoint modes. It is notorious that these non-linear terms correspond to equations (8) and (10) extracted from the direct and adjoint version of the Navier-Stokes equations (5) and (9), which consider the three components of

the velocity vector. Hence, the structural sensitivity in non-linear flows is then defined as

$$\mathbf{S}^{NL} = \|\mathbf{N}\| \cdot \|\mathbf{N}^*\|. \quad (18)$$

Regarding the definition of the feedback force eqs. (16)-(17) and its physical correspondence to a flow actuation mechanism, it is known that controlling the non-linear term of the Navier-Stokes equations affects the turbulent energy production. Sharma et al. [17] presented the concept of the Navier-Stokes equations as a linear system with a non-linear feedback forcing in the context of the control theory. In linear feedback control, it is shown that the flow transition occurs for high values of the disturbance amplitudes (when these amplitudes surpass a specific threshold) [18], so to control the flow it is necessary working within the small perturbation regime. Since the linear theory is only valid for very small perturbations, this idea seems unrealistic in many applications and the control could be lost due to the presence of large magnitude external disturbances. These authors [17] presented a method working for flow disturbances of any size, which allows the application of flow control theory in the presence of non-linearities [19–21]. The method is based on the H_∞ theory, which has been successfully applied to control the performance of systems with non-linear, external and large uncertainty disturbances [22,23]. Sharma et al. [17] show that it is possible to control non-linear flows if the non-linear term of Navier-Stokes equations is characterized as positive real. To control this term, the authors show that it is possible to use a flow actuation mechanism. They applied this method to control a channel flow (with three dimensional perturbations to canonical plane Poiseuille flow) where they were able to completely suppress the streaks. The same idea is presented in this article, where the feedback force is defined using the non-linear term of Navier-Stokes equations: controlling this term, it could be possible to control the turbulent energy production.

In the structural sensitivity analysis as first introduced by Giannetti & Luchini [2], the forcing term perturbs the flow, leading to a perturbed eigenvalue problem (or linearized dynamics). As previously explained, following this theory, the extension of the structural sensitivity to limit cycle (periodic solutions) should follow the work presented by Luchini et al. [7]. These authors calculate the adjoint modes of the Floquet transition operator to calculate the structural sensitivity of a periodic state in a circular cylinder to spatially localised feedback frequency. Also the work by Marquet & Lesshafft [8] follows the classical stability theory to propose an alternative definition for identifying wavemakers regions in periodic flows. The later authors also test their novel methodology in the wake of a two-dimensional circular cylinder, showing the influence in the global frequency selection of the different flow regions. Following the previous work and the classical theory definition, the fundamentals of a structural sensitivity analysis defined in a discrete setting (as in DMD), should be based on a perturbation of the operator/matrix (i.e., defined in the previous ODE as $\mathbf{A}$, see for instance $\frac{d\mathbf{v}}{dt} = \mathbf{A}\mathbf{v}$) that propagate the snapshots. The present article introduces a novel methodology to identify the structural sensitivity in non-linear flows fully data-driven and based on DMD, which directly considers the evolution of the flow perturbation, as described in the following section. Hence, the present work does not exactly follow the classical linear theory, but it is rather based on the classical theory and introduces a novel implementation that can also be used in the analysis of experimental databases.

2.2. Practical implementation of the structural sensitivity using a data-driven method

Let us consider a group of K snapshots collected every time instant Δt , organized in matrix form as

$$\mathbf{V}_1^K = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_{K-1}, \mathbf{v}_K], \quad (19)$$

where each snapshot $\mathbf{v}_k$ represent the velocity vector at time instant t_k . For simplicity this is called as the *direct snapshot matrix*. The dimension

of each snapshot is $pJ \times 1$, where p is the number of velocity components and J is the number of grid points conforming the spatial domain analysed. This snapshot matrix is the solution of a discrete dynamical system that can be described by the state-transition equation $\mathbf{v}_{k+1} = \mathbf{A}\mathbf{v}_k$, where $\mathbf{A}$ is the linear Koopman operator, as described in DMD. Applying DMD (or HODMD in the analysis of complex flows) to the snapshot matrix, it is possible to approximate each snapshot as the expansion of DMD modes previously presented in eq. (3), for simplicity repeated here as

$$\mathbf{v}_k \simeq \tilde{\mathbf{v}}_k = \sum_{m=1}^M a_m \mathbf{u}_m e^{(\delta_m + i\omega_m)t_k}, \quad k = 1, \dots, K. \quad (20)$$

Retaining M DMD modes in this expansion, it is possible to reconstruct an approximation of the original snapshot matrix, defined as $\tilde{\mathbf{V}} \simeq \mathbf{V}_1^K$.

Let us now consider the same snapshot matrix, but organized descending in time, as

$$\mathbf{V}_K^1 = [\mathbf{v}_K, \mathbf{v}_{K-1}, \dots, \mathbf{v}_k, \mathbf{v}_{k-1}, \dots, \mathbf{v}_2, \mathbf{v}_1]. \quad (21)$$

For simplicity this is called as the *adjoint snapshot matrix*. The state-transition equation is now described as $\mathbf{v}_k = \mathbf{A}^T \mathbf{v}_{k+1}$. As explained by Luchini & Bottaro [4] (see details in the Supplemental Material of that Annual Reviews), this new dynamical system is the adjoint of the previous one. It is important to remark that an adjoint operation is not the same as the time reversal. However, the present approximation serves when advection (hyperbolic term) dominates the solution of the Navier-Stokes equations (1)-(2). Further research should be carried out for testing (and adapting if needed) the present methodology to solve problems dominated by diffusion (parabolic or elliptic problems), but this remains open for future research.

Similarly, applying DMD to this adjoint snapshot matrix, it is possible to reconstruct an approximation of the original adjoint snapshot matrix $\tilde{\mathbf{V}}^* \simeq \mathbf{V}_K^1$, where each snapshot approximates the adjoint solution of the original dataset (since the data evolve descending in time). Hence, it is possible to construct the DMD expansion associated to this second snapshot matrix as

$$\mathbf{v}^*(x, y, z, t_k) = \tilde{\mathbf{v}}_k^* \simeq \tilde{\mathbf{v}}_k^* = \sum_{m=1}^M a_m \mathbf{u}_m^*(x, y, z) e^{(-\delta_m + i\omega_m)t_k}, \quad k = 1, \dots, K, \quad (22)$$

where $\mathbf{u}_m^*$ are the adjoint DMD modes (it is notorious that these modes are not the same as the eigenmodes calculated when solving the eigenvalue problem associated to the linear adjoint equations). Each adjoint mode is related to a direct mode with similar frequency and amplitude and opposite growth rate, being orthogonal among them ($\langle \mathbf{u}_m, \mathbf{u}_m^* \rangle = 0$, with $\langle \cdot, \cdot \rangle$ as the Euclidean inner product).

The structural sensitivity in non-linear flows can then be approximated using these direct and adjoint DMD modes as explained in the following steps.

Step 1: simplifying the snapshot matrix. At this step, DMD is applied to the original snapshot matrix (19) to calculate the DMD expansion (20). Since the data analysed could be experimental or numerical, in both cases representing non-linear complex flows, it is necessary to simplify the original dataset. Hence, the DMD expansion will only retain the three highest amplitude modes: (i) representing the mean flow, this is $\mathbf{u}_1$, and (ii) representing the main flow fluctuations, these are the leading frequency mode $\mathbf{u}_2$, and its complex conjugate $\mathbf{u}_3$ (since the reconstructed solution is real). This leading frequency mode will be in charge of producing the highest changes in the mean flow. The structural sensitivity will then approximate the regions of the flow where it is possible to produce the strongest drift of such mode. Let us note that this leading mode would be the most unstable eigenmode identified if a linear stability analysis would be carried out [12]. With the simple approximation of the snapshot matrix carried out at this step, we intend to approximate the linear solution of the flow.

Step 2: direct and adjoint DMD. DMD is applied to the simplified direct and adjoint snapshot matrices, $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{V}}^*$, respectively, obtaining the *direct* and *adjoint* DMD expansions (20) and (22), respectively. To simplify the computation of the adjoint modes, the mean is subtracted to the adjoint snapshot matrix before applying the DMD method (the mean is the same in both, the direct and the adjoint matrices).

Step 4: non-linear structural sensitivity. The structural sensitivity, called as *non-linear structural sensitivity* is calculated as

$$\tilde{\Sigma}^{NL} = \omega[|\mathbf{M}| \cdot |\mathbf{M}| + 2|\mathbf{N}| \cdot |\mathbf{N}^*|], \quad (23)$$

where the DMD modes are weighted with their corresponding amplitudes, and normalized with their maximum. Note that the practical implementation of the structural sensitivity takes into account the variations of the mean flow (that also considers the non-linearities of the flow), weighted with the frequency of the leading mode, and the two non-linear terms corresponding to the leading DMD mode and its conjugate complex (see how to calculate the structural sensitivity in other types of non-linear problems in Ref. [3], where all the terms that define the non-linear equations contribute to compute the structural sensitivity).

It is important to remark the difference between this final expression for the non-linear structural sensitivity eq. (23) with the definition presented in eq. (18), where only the non-linear term is considered ($|\mathbf{N}| \cdot |\mathbf{N}^*|$). In the latter case, the definition of the structural sensitivity is based on the feedback force defined by the eigenmodal solution that represents the flow fluctuations $\mathbf{v}'$. In this practical implementation, the non-linear structural sensitivity is calculated in the instantaneous flow field $\mathbf{v}$ (composed by the mean flow and the flow perturbations as $\mathbf{v} = \bar{\mathbf{v}} + \mathbf{v}'$), as illustrated in the snapshot matrix eq. (19). Hence, as explained in Step 1 of this section, this instantaneous flow is represented by the linear combination of the three highest amplitude DMD modes: $\mathbf{u}_1$ (steady mode representing the mean flow), $\mathbf{u}_2$ (velocity fluctuations) and $\mathbf{u}_3 = \bar{\mathbf{u}}_2$, where $\bar{\mathbf{v}} = \mathbf{u}_1$ and $\mathbf{v}' = \mathbf{u}_2 + \mathbf{u}_3$. The practical implementation of the non-linear structural sensitivity should also consider the linear combination of these three terms. Since the expression for the feedback forcing to control $\mathbf{u}_2$ and $\mathbf{u}_3$ is the same, the term $|\mathbf{N}| \cdot |\mathbf{N}^*|$ is double (multiplied by 2). It is also important to take into account the variations in the mean flow $\mathbf{u}_1$, which are directly affected by changes in the flow perturbation. Based on the definition of the feedback forcing, the term associated to this mean flow variations could be defined as $|\mathbf{M}| \cdot |\mathbf{M}|$ (note that, as explained before, $|\mathbf{M}| = |\mathbf{M}^*|$). To consider the influence of the unsteady mode in the variations of the mean flow, this latter term is weighted with the mode frequency ω .

3. Structural sensitivity in non-linear flows

This section approximates the structural sensitivity in four different cases: the two- and three-dimensional wake behind a circular cylinder, the flow inside a planar Food and Drug Administration (FDA) nozzle and two axis-symmetric concentric jets with similar velocity ratio. For such aim, numerical simulations have been carried out using the numerical solver Nek5000 [24], solving the non-linear incompressible form of the Navier-Stokes equations (1)-(2). The non-linear structural sensitivity calculated in the aforementioned cases has been compared with the literature. Additionally, the numerical solver StabFem [25] has also been used to calculate the structural sensitivity in the FDA nozzle and in the two concentric jets, to simplify the reading of this article. This code solves the linearized Navier-Stokes equations (11) - (12) and calculates the structural sensitivity of the linear solution as in eq. (15).

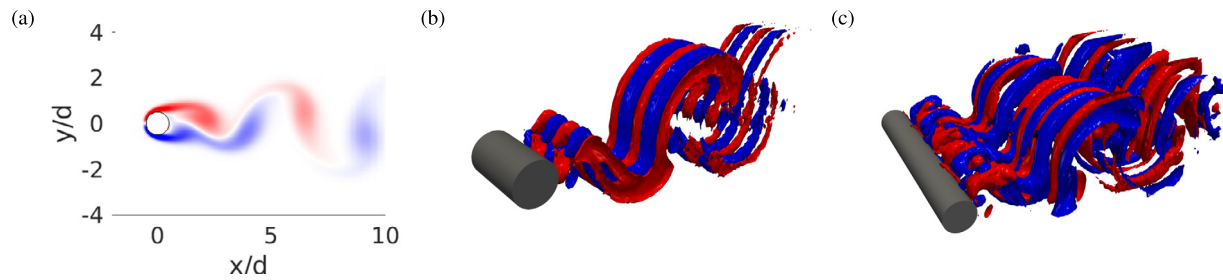


Fig. 1. Representative snapshot of the spanwise vorticity (normalized with the maximum value) for: (a) two-dimensional flow past a cylinder at $Re = 50$, (b) three-dimensional flow past a cylinder at $Re = 280$ with $L_z = 1.67$ and (c) three-dimensional flow past a cylinder at $Re = 280$ with $L_z = 6.99$.

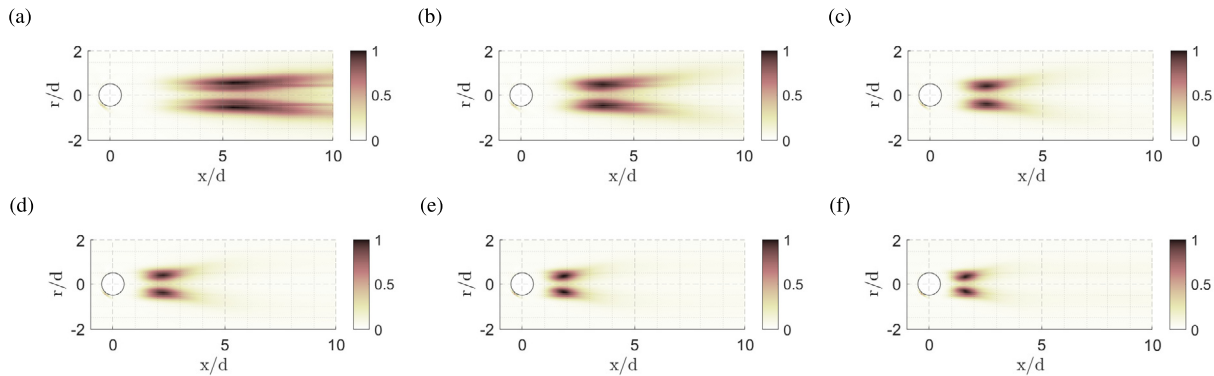


Fig. 2. Non-linear structural sensitivity in the two-dimensional wake of a circular cylinder for different Reynolds numbers. (a) $Re = 50$, leading mode frequency $St \approx 0.13$, (b) $Re = 60$, leading mode frequency $St \approx 0.14$, (c) $Re = 85$, leading mode frequency $St \approx 0.16$, (d) $Re = 100$, leading mode frequency $St \approx 0.17$, (e) $Re = 130$, leading mode frequency $St \approx 0.18$, and (f) $Re = 160$, leading mode frequency $St \approx 0.19$.

3.1. Flow past a circular cylinder

The flow past a circular cylinder is a benchmark problem in fluid dynamics. Several flow bifurcations occur in the wake of the circular cylinder as function of the Reynolds number, previously defined as $Re = Vd/\nu$, where V is the incoming free stream velocity and d is the diameter of the cylinder. At very low Reynolds number, the flow is steady. The first flow instability occurs at $Re \approx 46$: a Hopf bifurcation leads the transition from steady to unsteady two-dimensional flow, assembling a von Karman vortex street. Increasing the Reynolds number, new flow instabilities trigger the transition to three-dimensional flow for some specific values of the spanwise length $L_z = 2\pi/\beta$, where β is the spanwise wavenumber.

Numerical databases of the two- and three-dimensional cylinder wake have been analysed. For such aim, numerical simulations have been carried out using the numerical code Nek5000 to model the two-dimensional wake of a circular cylinder. The boundary conditions and dimensions of the computational domain are set as in Barkley & Henderson [26]. The boundary conditions for the inlet, upper and lower boundaries are Dirichlet for velocity ($v_x = 1$ and $v_y = 0$) and Neumann for pressure. The cylinder surface is also modelled using Dirichlet conditions for the velocity field ($v_x = 0$, $v_y = 0$) and Neumann conditions for pressure. Finally, Neumann for velocity ($\partial v_x/\partial x = 0$, $\partial v_y/\partial x = 0$) and Dirichlet for pressure ($p = 0$) boundary conditions are imposed in the outlet boundary of the computational domain. Regarding the dimensions of the computational domain, non-dimensionalized with the cylinder diameter, these are set as $L_y = \pm 15d$ in the cross-stream direction, and $L_x = 15d$ upstream the cylinder and $L_x = 50d$ downstream the cylinder. The number of elements composing the computational domain is 600, each one discretized using Gauss–Lobatto–Legendre points with a polynomial of order $\Pi = 9$. A grid independence study has been previously carried out and the solution has been validated with the literature [26].

Regarding the three-dimensional cylinder, two different databases have been studied at $Re = 280$ varying the spanwise wavelength as $L_z = 1.67$ and 6.99 . These databases are published in Ref. [27], where more details regarding the numerical set up (boundary conditions, domain dimensions and number of grid points) can be found. Fig. 1 shows a representative snapshot of the spanwise vorticity field, defined as $\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}$, revealing the main dynamics in the three-dimensional cylinder wake compared to the two-dimensional case.

Fig. 2 shows the non-linear structural sensitivity in the two-dimensional flow past a circular cylinder calculated at $Re = 50, 60, 85, 100, 130$ and 160 . The frequency of the leading mode identified in these calculations is presented in the figure caption (in good agreement with the literature [26]). The frequency f is non-dimensionalized with the Strouhal number, defined as $St = \frac{fd}{V}$, where d is the cylinder diameter and, as mentioned before, V is the characteristic flow velocity.

As expected, the most sensitive region is located in the wake of the circular cylinder. The size of this region decreases with the rise in the Reynolds number. This result is in contrast to the linear theory (see Fig. 18 in Giannetti & Luchini [2]). A distinction needs to be made between the aforementioned article and the present study. The former linearises around a base flow, i.e. the unstable solution to the steady Navier–Stokes equations. As pointed out by Zielinksa et al. [28], the recirculation bubble, and hence the region of structural sensitivity, grows as the base flow becomes more unstable. Since this paper deals with the fully developed limit cycle, i.e. vortex shedding, the recirculation bubble of the time-averaged flow shrinks as the Reynolds number increases. Hence, it is not surprising that a similar trend is observed in Fig. 2 that should not be attributed to the linear nature [2], but to the difference between linearising around the base versus the mean flow.

The present result is closer to the approximation given by Strykowski & Sreenivasan [29]. These authors perform several experiments (non-linear flow) at various Reynolds numbers, placing two small cylinders, located symmetrically with respect to the main body, at various posi-

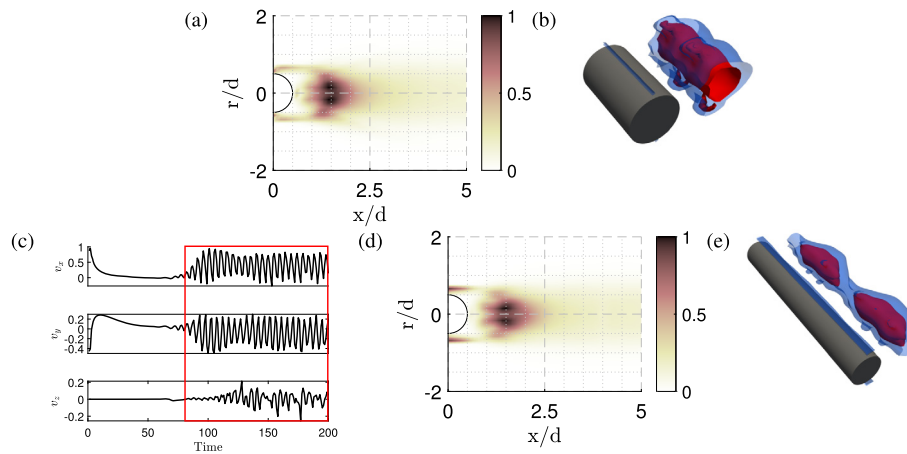


Fig. 3. Non-linear structural sensitivity in the three-dimensional wake of a circular cylinder at $Re = 280$. Short wavelength test ($L_z = 1.67$), leading mode frequency $St \approx 0.21$: (a) XY Plane extracted at $z = 0$, (b) Three-dimensional structural sensitivity. Long wavelength test ($L_z = 6.99$), leading mode frequency $St \approx 0.20$: (c) Streamwise, normal and spanwise velocities extracted at $(x, y, z) = (1, 0.5, L_z/2)$ (analysis carried out in the interval marked in red), (d) XY Plane extracted at $z = 0$, (e) Three-dimensional structural sensitivity. In (b) and (e) iso-contour of 70% and 50% of the maximum structural sensitivity value in red and blue, respectively.

tions of the cylinder wake. They identified the regions of the flow where this passive mechanism of flow control was acting to suppress the vortex shedding in the cylinder wake. These authors found that, for a similar Reynolds number, changing the diameter size of the small cylinders with respect to the main cylinder directly affects the size of the most sensitive regions to control the flow (see Fig. 5 in Ref. [29]). Also, in Fig. 20 of the mentioned paper, the authors show that the size of these high sensitivity regions decreased with the rise in the Reynolds number, in good agreement with the present results, and also in good agreement with the work presented by Marquet et al. [15] (using a methodology based on the classical linear theory, their results are probably the best in terms of matching the experimental results). It is notorious that the experimental and numerical analyses carried out using this type of passive mechanism of flow control in the flow past a circular cylinder suggest that both, the linear theory and this novel approach, provide an approximation of the most sensitive regions in the flow to control the flow instabilities, but they do not provide the exact location of such regions. The purposes of the present results (non-linear flows), the results presented by Giannetti & Luchini [2] (linear flows) and by Strykowski & Sreenivasan [29] (wind tunnel experiments) are different, and do not contradict each other. In the linear analysis, the structural sensitivity map is calculated based on a structural perturbation only acting on the time-varying disturbance, while in the experiments carried out by Strykowski & Sreenivasan [29] the structural perturbation acts in both the base flow and the disturbance. In the present results, this structural perturbation acts in both the base flow and the disturbance (as in the experimental results), but only considers the structural variations of the leading mode (as in the linear theory). Hence, this non-linear structural sensitivity could be considered as an intermediate step connecting the linear theory and the experimental analysis in non-linear flows.

The structural sensitivity is now computed in the non-linear solution of the three-dimensional flow past a circular cylinder at $Re = 280$. In the short wavelength case, it is possible to identify the mode B, while in the long wavelength case it is possible to identify both the modes A and B (see details about these type of modes in Ref. [30]). The three-dimensional structural sensitivity calculated with this novel technique is presented in Fig. 3 in the two tests performed. The frequency of the leading mode ($St \approx 0.20 - 0.21$) is in good agreement with Blackburn et al. [30]. These authors calculated this frequency based on the results by Barkley & Henderson [26], who estimated the theoretical value for this frequency as $St \approx 0.20 - 0.21$ for $Re = 280$. The experiments presented by Williamson [31] also estimate the value for this frequency as $St \approx 0.20$. See the details about the value of this frequency and a complete

description of the frequencies in the wake past a circular cylinder as function of the Reynolds number in Fig. 1 of Ref. [26]).

In both cases, the region of the highest sensitivity is in good agreement with the linear analysis presented by Giannetti et al. [32]. More specifically, in Fig. 1 (b) and Fig. 2 of the mentioned publication these authors present a two dimensional plane showing the structural sensitivity calculated for the flow past a circular cylinder at $Re = 260$ for the mode with $L_z = 2\pi/\beta = 2\pi/7.64 = 0.89$ and with $L_z = 2\pi/1.585 = 3.96$, respectively. Small differences could be found between the literature and the present test case because this article calculates the structural sensitivity in the non-linear solution of the Navier-Stokes equations (considering the mean flow), instead of performing a classical analysis based on the base flow and the linear solution of Navier-Stokes equations. Also, differences found in the Reynolds number and the wavelength between the two cases compared could be connected with the qualitative differences identified. Nevertheless, in both the present results and the literature [32], the highest intensity areas of the structural sensitivity are located in the near field of the wake of the cylinder, in $0.5 < x/d < 2$ and $-1 < r/d < 1$. Moreover, this result is in accordance with the tendency followed by the structural sensitivity previously computed in the two-dimensional cylinder wake at high Reynolds number (the highest sensitivity is located in a small region downstream and close to the body). To test the limits of this algorithm to compute the non-linear structural sensitivity, we follow two different strategies. In the short wavelength case, Fig. 3 (a) and (b), the data analysed are collected in the saturated regime of the numerical simulations (the solution is temporally converged). In contrast, in the long wavelength test, Fig. 3 (d) and (e), the algorithm is applied using information from the transient regime of the simulations, as presented Fig. 3 (c) that shows the temporal evolution in the numerical simulations of the three velocity components extracted in a representative point of the computational domain and highlights the region where the data are collected. In this second complex case, to compute the structural sensitivity, the methodology presented in Section 2.2 replaces DMD by the multi-dimensional iterative HODMD algorithm [33] (see details on how to use this method in the present test case in Ref. [13] and Ref. [27]). The same results are obtained when the analysis of this second case is carried out in the saturated regime (not shown for the sake of brevity). However, this article aims also at testing the limits of applicability of this methodology, which is mainly dependant on the type of modal decomposition method (DMD or HODMD) used for identifying the leading mode. Hence, in complex systems, by replacing DMD by HODMD it is also possible to identify the structural sensitivity of the flow using the present novel data-driven approach.

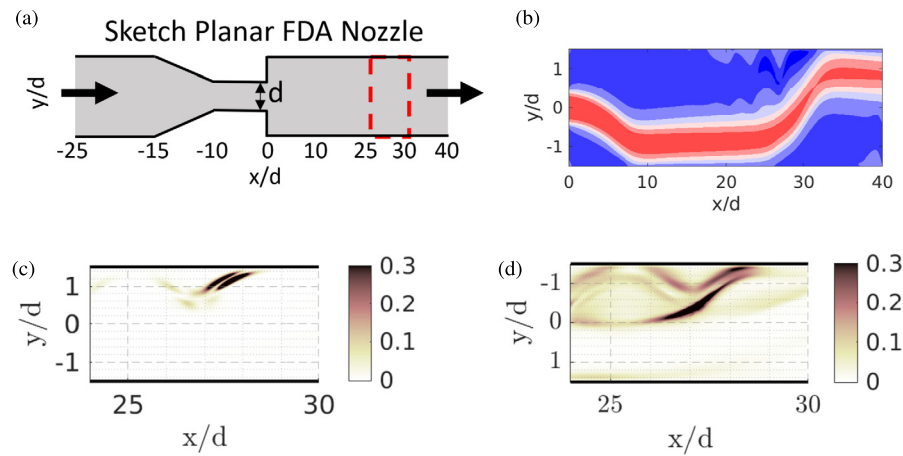


Fig. 4. (a) Sketch of the flow through a planar FDA nozzle. (b) Representative snapshot of streamwise velocity (normalized with its maximum value) representing the flow in the throat expansion ($0 \leq x/d \leq 40$) of the FDA nozzle. (c) - (d): Structural sensitivity in the FDA nozzle at $Re = 800$, with leading mode frequency $St \approx 0.02$, using (c) non-linear and (d) linear approximations. Zoomed in view in the channel downstream the throat expansion for $25 \leq x/d \leq 30$ (red rectangle in part (a)).

3.2. Flow through a planar Food and Drug Administration (FDA) nozzle

‘The critical path initiative’ program [34], created by the United States Food and Drug Administration (FDA), aims to develop robust and standardized methodologies for medical applications. Medical devices can produce some damages in the human body, generating artificially-caused disturbances and promoting the appearance of turbulent flows. Flow disturbances can provoke severe damage in blood vessels such as hemolysis, where the erythrocyte membranes are disrupted by the rise in shear forces caused by high-velocity jet streams [35,36]. The FDA is currently leading the search and the development of new medical devices minimizing health risks. In this line, a benchmark nozzle that resembles an idealized medical device has been designed to study in detail the main characteristic describing the flow inside this geometry, this is known as the FDA nozzle.

This article studies the numerical database published in Corrochano et al. [37], generated using the solver Nek5000, which reproduces the experimental model of the FDA nozzle presented in Ref. [38]. The computational domain, two-dimensional, is modelled by a tube with length and diameter $9.33d$ and $3d$, respectively, connected by a smooth contraction to a throat with length and diameter $L_t = 10d$ and $d = 1$, respectively. The origin of the expansion is set as the origin of the reference system. The Reynolds number is calculated based on the throat diameter, and is set to $Re = 800$. The wall of the geometry is modelled imposing Dirichlet boundary conditions for the velocity field ($v_x = v_y = 0$) and Neumann conditions for pressure. A Poiseuille profile is defined as inlet boundary condition, and the pressure field is set with Neumann conditions. The sum of pressure and viscous stress is imposed zero for the outlet surface, while Dirichlet conditions are imposed for the pressure. The mesh of the computational domain contains 8650 elements, each one discretized using Gauss–Lobatto–Legendre points with a polynomial of order $\Pi = 10$. More details about the numerical set-up and the validation of the numerical simulations can be found in the mentioned article (Ref. [37]). The geometry and a representative snapshot of the streamwise velocity are shown in Fig. 4 (a) and (b). As mentioned before, the numerical solver StabFem [25] has also been used to calculate the structural sensitivity (as in eq. (15)) in the FDA nozzle, to simplify the reading of this article.

Fig. 4 shows the structural sensitivity computed in the non-linear solution of a planar FDA nozzle at $Re = 800$ and compares the results with the linear structural sensitivity computed using StabFem, as presented in Fig. 4 (d). The solution of this linear stability analysis is in good agreement with the literature [37,39], where the critical Reynolds number, which differentiate the stable and unstable solutions, was accurately

computed. The non-dimensional frequency of the leading mode is for both analyses $St \approx 0.02$ (non-dimensionalized with d , the diameter of the throat in the FDA nozzle). As in the test case of the circular cylinder, some discrepancies are found between the linear and non-linear approximations. This was expected since, as explained in the previous section, the methodologies used are different and one is based on the mean flow and the other one on the base flow. Nevertheless, the regions of highest intensity for the structural sensitivity is located in both cases at $26 \leq x/d \leq 29$ and $y/d > 0$.

3.3. Flow in two axis-symmetric concentric jets with similar velocity ratio

Double concentric jets is a configuration enhancing the turbulent mixing of two jets, which is used in several industrial applications where the breakup of the jet into droplets due to flow instabilities is presented as the key technology. Combustion (i.e.: combustion chamber of rocket engines, gas turbine combustion, internal combustion engines, etc.) and noise reduction (e.g.: in turbo fan engines) are the two main applications of this geometry, although the annular jets can also be found in some other relevant applications such as ink-jet printers or spray coating.

Numerical simulations have been carried out using the numerical solver Nek5000 to study the flow in a double concentric jet configuration. The set up for the numerical simulations is the same as in Martín et al. [40], but the Reynolds number of the case studied was slightly larger than the one studied in this article. The configuration that we study is conformed by an inner rounded jet and an external annular jet. The radius of the inner jet is $d/2 = 0.5$, while for the annular jet, the internal and external radii are defined as $D_i/2 = d/2 + L$ and $D_e/2 = 3d/2 + L$, with $L = 0.1$ being the distance between the two jets. The flow is axis-symmetric at the flow conditions studied, hence the computational domain is two-dimensional and axis-symmetric boundary conditions are imposed at the bottom part of the domain. The inner jet is set as the origin of the Cartesian frame. At the entrance of the jets, free stream velocity is imposed as boundary condition for velocity (V_{in} and V_{out} for the inner and outer jets, respectively) and Neumann conditions are imposed for pressure. In the remaining surfaces at the entrance of the computational domain, the boundary conditions imposed are zero velocity (wall). In the upper and outlet surfaces, Dirichlet and Neumann boundary conditions are imposed for pressure and velocity, respectively. More details about this numerical setup can be found in Ref. [40]. The Reynolds number in the configuration studied is $Re = 1500$, defined with the diameter of the internal jet. The numerical setup and the database generated have been validated with the results presented in Canton et al. [41].

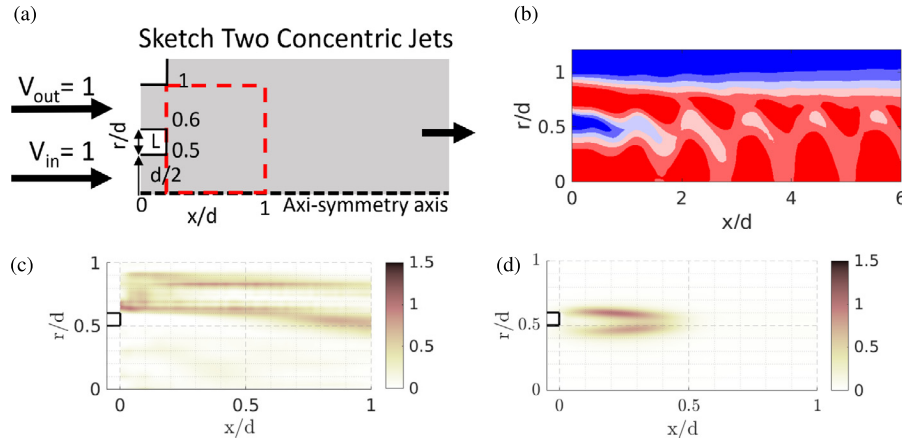


Fig. 5. (a) - (b): Sketch and a representative snapshot (streamwise velocity) of two axis-symmetric concentric jets with similar velocity ratio (the velocity field is normalized with the maximum value). (c) Non-linear and (d) linear structural sensitivity in two concentric jets at $Re = 1500$. The frequency of the leading mode presented is $St \approx 1.82$. Zoomed in view in the exit of the two jets (red rectangle in part (a)). The exit of the inner and outer jets are defined from $r/d = 0 - 0.5$ and $0.6 - 1$ at $x/d = 0$, respectively.

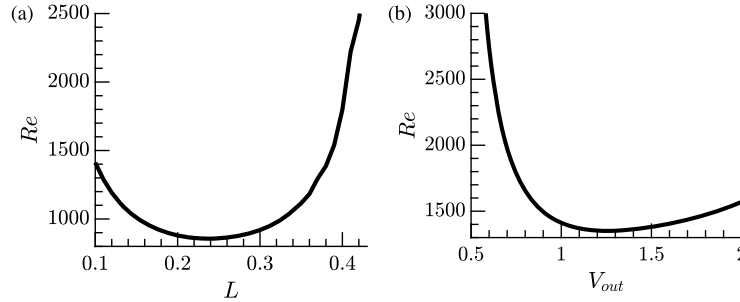


Fig. 6. Variations of the critical Reynolds number as function of (a) the distance L between the two concentric jets and (b) the velocity of the outer jet V_{out} . Results obtained from the linear stability analysis carried out using the solver StabFem [25]. Case studied in Fig. 5: $L = 0.1$, $V_{out} = 1$.

The structural sensitivity is computed in the non-linear solution of two-axis-symmetric concentric jets with similar velocity ratio at the jet nozzle ($V_{in} = V_{out} = 1$) and presented in Fig. 5. These two results have been compared with the *linear* structural sensitivity, which has been computed using StabFem, as presented in Fig. 5 (d). The solution of this linear stability analysis is in good agreement with the literature [41,40], where the critical Reynolds numbers, which differentiate the stable and unstable solutions, were accurately computed in both test cases. The non-dimensional frequency of the leading mode is $St \approx 1.82$ calculated for both the linear and non-linear test cases (the Strouhal number of the non-dimensional frequency uses d as the diameter of the inner jet in the two concentric jets). Again, some discrepancies are found between the linear and non-linear approximations. From the results obtained it is possible to see that there is a region where the structural sensitivity presents high intensity contours in both cases: for $0.6 \leq r/d \leq 0.7$ and $0 \leq x/d \leq 0.5$. This result suggests that varying the distance between jets L could be a possible mechanism of flow control. For the linear analysis, there is an additional region of high intensity at $0.4 \leq r/d \leq 0.5$ and $0 \leq x/d \leq 0.5$ that is not found in the non-linear analysis. On the contrary, for the non-linear analysis the high intensity area extends to values up to $r/d \approx 0.9$ and $x/d \approx 1$. This second result suggests that varying the velocity of the external jet, could also be a mechanism of flow control. If we pay attention to the contour maps of the non-linear structural sensitivity and the possible changes that could be performed in L to control the flow, it is possible to see that since the most sensitive regions to changes are $0.6 \leq r/d \leq 0.9$, the solution could be controlled for the following values of L : $0.1 < L \leq 0.4$, where two high intensity regions are differentiated in these results, connected to the values $0.1 < L \leq 0.2$ and $0.3 \leq L \leq 0.4$.

The results presented in Martín et al. [40] (Fig. 4) and Canton et al. [41] (Fig. 7) show that varying the velocity ratio between the two concentric jets, or varying the diameter of the outer jet, affect the critical Reynolds number, increasing or decreasing its value. This is in good agreement with the non-linear structural sensitivity computations, which suggest that the stability of the flow could be controlled by the outer jet (i.e.: varying the flow velocity or the diameter dimensions).

To test the results obtained from the non-linear structural analysis and to be sure that no assumptions are made from the interpretation presented, we have performed the linear stability analysis, where the critical Reynolds number (that differentiates stable and unstable solutions) is calculated as function of the distance between jets L and also as function of the external jet velocity. For this analysis, we have used the numerical solver StabFem [25], as described in Ref. [40] (where a similar analysis was carried out by the authors of this paper). The results obtained in this new calculations, are also in good agreement with the literature [40,41]. As presented in Fig. 6 (a), the critical Reynolds number varies with L , decreasing for $L \approx 0.1 - 0.2$ reaching its minimum value for $L \approx 0.2 - 0.3$ (flow destabilization). For values of $L \approx 0.3 - 0.4$ there is again a rise in the critical Reynolds number, showing the most stable case of the results presented for $L \approx 0.4$. This result is in good agreement with the results predicted by the non-linear structural sensitivity analysis carried out in this article. It is remarkable that in the performed non-linear analysis, two specially high intensity regions were identified for $0.1 \leq L \leq 0.2$ and $0.3 \leq L \leq 0.4$, which concur with the areas where the critical Reynolds number decrease and increase. However, further studies should be carried out in this line to better understand the nature behind this type of solutions. In Fig. 6 (b) the critical Reynolds number is calculated as function of the velocity of the outer jet. The most stable result is found for the case where

the external jet velocity is half the velocity of the internal jet. It is remarkable that for both type of flow control mechanisms applied in this article (either varying L or V_{out}) the critical Reynolds number is drastically reduced when $L = 0.4$ or when $V_{out} = 0.5$, in good agreement with the concept of the feedback force applied in the most sensitive areas of the flow. Additionally, it is worth to mention that in the case where the velocity of the external jet is modified, the most unstable case is the one with $V_{out} \in (1, 1.5)$, but increasing again this velocity value there is a rising tendency in the critical Reynolds number up to $V_{out} = 2$. This result could be connected to the extension in the streamwise domain of the highest intensity of the non-linear structural sensitivity, that reach areas up to $x/d \simeq 1$.

These two linear stability analyses carried out revealing the mechanisms of flow control in concentric jets, show once more that, although the linear and non-linear structural sensitivities identified present some differences (due to the different nature of the mentioned methodologies), the non-linear structural sensitivity locates the areas of the flow most sensitive to structural changes, shedding light on new mechanisms of flow control, with the clear advantage that the novel method presented is fully data-driven.

4. Conclusions

This article introduces a novel methodology to calculate the structural sensitivity in databases composed by non-linear flows. The method uses some ideas from the linear theory, and brings a novel implementation which is based on discrete systems and uses DMD (or HODMD in complex applications). The so-called *non-linear structural sensitivity* gives an approximation of the spatial regions of the flow more sensitive to control the global flow instabilities without the need of solving the adjoint equations (although the method does not calculate the adjoint modes as in the linear theory). The present approach could be considered as an intermediate step connecting the linear theory to non-linear solutions. The method introduced in this article, fully data-driven, could be easily extended to analyse more complex flows or experimental databases.

The performance of method is tested in four different problems: the wake behind a two-dimensional and three dimensional cylinder, the flow through a planar FDA nozzle and the flow in two axi-symmetric concentric jets. The results obtained using this novel technique present some differences compared to the structural sensitivity calculated using the linear theory, although the highest intensity structural sensitivity regions are identified in similar places. Moreover, in the case of the cylinder wake, the results obtained with the present methodology agree with the experiments carried out by Strykowski & Sreenivasan [29]. The difference between the linear theory and these experiments is reasonable, since the former linearises around a base flow (steady), while in the experiments the structural perturbation acts in both the base flow and the disturbance. In the present methodology, the structural perturbation also acts in the mean and the disturbance, but as in the linear approach, it only considers the structural variations of the mean flow, suggesting that this methodology is an intermediate step between the aforementioned experimental and linear numerical analyses.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank the support of the Comunidad de Madrid through the call Research Grants for Young Investigators from the Universidad Politécnica de Madrid and also acknowledge the grant PID2020-114173RB-I00 funded by the Spanish Ministry MCIN/AEI/10.13039/501100011033. A.C. also acknowledges the support of Universidad Politécnica de Madrid, under the programme 'Programa Propio'.

References

- [1] B.R. Noack, M. Morzynski, G. Tadmor, *Reduced-Order Modelling for Flow Control*, Springer Science and Business Media, New York, 2011, 528.
- [2] F. Giannetti, P. Luchini, Structural sensitivity of the first instability of the cylinder wake, *J. Fluid Mech.* 581 (2007) 167–197.
- [3] A. Martínez, A. Güemes, J.M. Perales, J.M. Vega, Variable thickness in plates - a solution for shm based on the topological derivative, *Sensors* 20 (2020) 2529.
- [4] P. Luchini, A. Bottaro, Adjoint equations in stability analysis, *Annu. Rev. Fluid Mech.* 46 (2014) 493–517.
- [5] E. Ferrer, O.M.E. Browne, E. Valero, Sensitivity analysis to control the far-wake unsteadiness behind turbines, *Energies* 10 (2017) 1599.
- [6] E. Ferrer, J. de Vicente, E. Valero, Low cost 3d global instability analysis and flow sensitivity based on dynamic mode decomposition and high-order numerical tools, *Int. J. Numer. Methods Fluids* 76 (2014) 169–184.
- [7] P. Luchini, F. Giannetti, J. Pralits, Structural sensitivity of the finite-amplitude vortex shedding behind a circular cylinder, in: IUTAM Symposium on Unsteady Separated Flows and Their Control, in: IUTAM Bookseries, vol. 14, 2009, pp. 151–161.
- [8] O. Marquet, L. Lesshafft, Identifying the active flow regions that drive linear and nonlinear instabilities, arXiv:1508.07620 [physics.flu-dyn].
- [9] P. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (2010) 5–28.
- [10] S. Le Clainche, J.M. Vega, Higher order dynamic mode decomposition, *SIAM J. Appl. Dyn. Syst.* 16 (2) (2017) 882–925.
- [11] S. Le Clainche, J.M. Vega, Analyzing nonlinear dynamics via data-driven dynamic mode decomposition-like methods, *Complexity* 6920783 (2018) 1–21.
- [12] F. Gómez, S. Le Clainche, P. Paredes, M. Hermanns, V. Theofilis, Four decades of studying global linear instability, *AIAA J.* 50 (12) (2012) 2731–2743.
- [13] S. Le Clainche, J.M. Pérez, J.M. Vega, Spatio-temporal flow structures in the three-dimensional wake of a circular cylinder, *Fluid Dyn. Res.* 50 (5) (2018) 051406.
- [14] S. Pope, Pdf methods for turbulent reactive flows, *Prog. Energy Combust. Sci.* 11 (1985) 119–192.
- [15] O. Marquet, D. Sipp, L. Jacquin, Sensitivity analysis and passive control of cylinder flow, *J. Fluid Mech.* 615 (2008) 221–252.
- [16] B.J. McKeon, A.S. Sharma, A critical-layer framework for turbulent pipe flow, *J. Fluid Mech.* 658 (2010) 336–382.
- [17] A.S. Sharma, B.J. McKeon, J.F. Morrison, D.J.N. Limebeer, Stabilising control laws for the incompressible Navier-Stokes equations using sector stability theory, in: 3rd AIAA Flow Control Conference AIAA-2006-3695.
- [18] M. Högberg, T.R. Bewley, D.S. Henningson, Linear feedback control and estimation of transition in plane channel flow, *J. Fluid Mech.* 481 (2003).
- [19] M.G. Safonov, E.A. Jonckheere, M. Verma, D.J.N. Limebeer, Synthesis of positive real multivariable feedback systems, *Int. J. Control* 45 (3) (1987) 817–842.
- [20] W. Sun, P. Kargonekar, D. Shim, Solution to the positive real control problem for linear time-invariant systems, *IEEE Trans. Autom. Control* 39 (10) (1994) 2034–2046.
- [21] A.S. Sharma, B.J. McKeon, J.F. Morrison, D.J.N. Limebeer, Control of incompressible flows, *Bul. Am. Phys. Soc.* 50 (9).
- [22] W.J. Green, D.J.N. Limebeer, *Linear Robust Control*, Prentice Hall, New Jersey, 1995.
- [23] J. Doyle, K. Glover, P. Kargonekar, B. Francis, State-space solutions to standard h_2 and h_∞ control problems, *IEEE Trans. Autom. Control* 34 (8) (1989) 831–847.
- [24] P.F. Fisher, J.W. Lottes, S.G. Kerkemeier, Nek5000: open source spectral element cfd solver, Available online: www.nek5000.mcs.anl.gov.
- [25] STABFEM, Gitlab repository, www.gitlab.com/stabfem/StabFem.
- [26] D. Barkley, R. Henderson, Three-dimensional Floquet stability analysis of the wake of a circular cylinder, *J. Fluid Mech.* 322 (1996) 215–241.
- [27] J.M. Vega, S. Le Clainche, Higher Order Dynamic Mode Decomposition and Its Applications, Elsevier, ISBN 9780128197431, 2020, p. 3200, Book.
- [28] B.J.A. Zielinska, S. Goujon-Durand, J. Dušek, J.E. Wesfreid, Strongly nonlinear effect in unstable wakes, *Phys. Rev. Lett.* 79 (1997) 3893.
- [29] P.J. Strykowski, K.R. Sreenivasan, On the formation and suppression of vortex 'shedding' at low Reynolds number, *J. Fluid Mech.* 218 (1990) 71–107.
- [30] H.M. Blackburn, F. Marques, J.M. Lopez, Symmetry breaking of two-dimensional time-periodic wakes, *J. Fluid Mech.* 522 (2005) 395–411.
- [31] C. Williamson, Oblique and parallel modes of vortex shedding in the wake of a circular cylinder at low Reynolds numbers, *J. Fluid Mech.* 206 (1989) 579–627.
- [32] F. Giannetti, S. Camarri, P. Luchini, Structural sensitivity of the secondary instability in the wake of a circular cylinder, *J. Fluid Mech.* 651 (2010) 319–337.
- [33] S. Le Clainche, J.M. Vega, J. Soria, Higher order dynamic mode decomposition of noisy experimental data: the flow structure of a zero-net-mass-flux jet, *Exp. Therm. Fluid Sci.* 88 (2017) 336–353.
- [34] FDA's critical path initiative [Online]; <https://www.fda.gov/science-research/science-and-research-special-topics/critical-path-initiative>. (Accessed 30 November 2020).
- [35] R.A. Malinauskas, P. Hariharan, S.W. Day, L.H. Herbertson, M. Buesen, U. Steinseifer, K.I. Aycock, B.C. Good, S. Deutsch, K.B. Manning, Fda benchmark medical device flow models for cfd validation, *ASAIO J.* 63 (2017) 150–160.
- [36] A. Tsuji, M. Tanabe, K. Onishi, T. Kitamura, T. Okinaka, M. Ito, T. Nakano, Intravascular hemolysis in aortic stenosis, *Int. Med.* 43 (2004) 935–938.

- [37] A. Corrochano, D. Xavier, P. Schlatter, R. Vinuesa, S. Le Clainche, Flow structures on a planar food and drug administration (fda) nozzle at low and intermediate Reynolds number, *Fluids* 6 (1) (2020) 4–15.
- [38] Y.T. Delorme, K. Anupindi, S.H. Frankel, Large Eddy simulation of fdas idealized medical device, *Cardiovasc. Eng. Technol.* 4 (2013) 392–407.
- [39] M. Jotkar, R. Govindarajan, Two-dimensional modal and non-modal instabilities in straight-diverging-straight channel flow, *Phys. Fluids* 31 (2019) 014102.
- [40] J.A. Martín, A. Corrochano, J. Sierra, D. Fabre, S. Le Clainche, Modeling double concentric jets using linear and non-linear approaches, in: *International Workshop on Soft Computing Models in Industrial and Environmental Applications* 1268, 2020, p. 451.
- [41] J. Canton, F. Auteri, M. Carini, Linear global stability of two incompressible coaxial jets, *J. Fluid Mech.* 824 (2017) 886–911.

5.2 Flow control with data driven approaches for laminar flames

5.2.1 Publication information

Title: Flow control with data-driven approaches for laminar flames

DOI: <https://doi.org/10.23967/eccomas.2024.085>

Journal: ECCOMAS Congress 2024 - 9th European Congress on Computational Methods in Applied Sciences and Engineering

Date: 29 October 2024

Publisher: ECCOMAS

Copyright: © 2024 CC BY-NC-SA license

FLOW CONTROL WITH DATA DRIVEN APPROACHES FOR LAMINAR FLAMES

ADRIÁN CORROCHANO^{1,*}, JUAN LUIS CANALS¹, LAURA
SAAVEDRA¹ AND SOLEDAD LE CLAINCHE^{1,2}

¹ ETSI Aeronáutica y del Espacio, Universidad Politécnica de Madrid, Madrid, Spain

* Corresponding author: adrian.corrochanoc@upm.es

Key words: Computational Fluid Dynamics, Reacting Flows, Flow control

Summary. This article introduces a novel method to identify structural sensitivity, applied for the first time, to the authors' knowledge, in reactive flows. The method, so-called, non-linear structural sensitivity, identify the regions more sensible to the flow to structural changes. The method is applied to optimize the efficiency of a laminar pulsating flame, where the NO_x emissions are also controlled. The method is based on higher order dynamic mode decomposition (HODMD), therefore it is fully data-driven and generalizable for different conditions of the flow and different geometries. The results presented show great potential for the algorithm to be used in flow control applications and the design of components for pollutant reduction.

1 INTRODUCTION

Optimization and flow control in fluid dynamics is a research topic of high interest because of its multiple industrial applications, ranging from the design of aircrafts and space vehicles to microfluidic devices [1]. An optimal design of such devices can reduce the drag or the noise and increase the efficiency of those industrial devices. Modelling complex flows is a challenging problem due to the need of large computational resources and the complexity of the underlying physical problem [2]. In terms of combustion control, in addition to these mentioned problems, the large number of species involved in combustion processes, the range of scales, and the turbulence-chemistry interaction require large computational resources [3].

Flow control encompasses active and passive flow control. Active flow control is the one which involves external energy or active mechanisms to influence the flow. On the other hand, passive flow control does not introduce any source of energy. These techniques can be achieved through the design of the industrial devices, adding objects or modifying the shape. In his work, passive flow control for reactive flows is pursued to increase the efficiency of the flame and reduce the emission of pollutants, without adding any external source of energy. In this sense, not many studies have been carried out. Karagozian [4] explored the concept of the lobed fuel injector, which enhanced the local mixing and lowering the NO_x (nitrogen oxide) emissions with the change in the shape of the injector. Other examples included the addition of atomizers in combustion chambers [5] or transverse fuel injection into supersonic airflow [6]. What these techniques have in common is the fact that there is not a guideline of the location where to apply the passive flow control.

In non-reactive flows, it is possible to calculate the *structural sensitivity* [7]. This concept identifies the regions of the flow most sensitive to changes, therefore it has applications for flow control. To calculate the structural sensitivity, it is necessary to combine direct and adjoint solutions, what in realistic problems would be expensive in terms of computational memory and time. Recently, some approaches in the field of flow control have been carried out using data-driven methods, which would mean a substantial decrease in the cost. Deep reinforcement learning (DRL) has been used for active flow control in the flow past a cylinder [8] or in turbulent channel flows [9], both for drag reduction.

The present study extends our previous work [10], where we combine the knowledge about structural sensitivity with the use of fully data driven algorithms. The method has demonstrate to provide accurate results in laminar non reactive flows. The objective of the present work is to adapt this algorithm for laminar flames to increase the efficiency in the combustion and reduce the release of pollutants.

The article is organized as follows. Section 2 introduces the flow control algorithm. The numerical setup for the CFD simulation is shown and discussed in Section 3. Finally the main results and conclusions are presented in Sections 4 and 5, respectively.

2 METHODOLOGY

The proposed algorithm for passive flow control is based on Higher Order Dynamic Mode Decomposition (HODMD) [11]. HODMD is a data-driven algorithm that needs as an input datasets in matrix form. A group of K snapshots $\mathbf{v}_k$ equidistant in time is collected. Each snapshot is composed by $J = N_v \times N_x \times N_y$ points, where N_v is the number of variables and N_x and N_y are the grid points in the streamwise and normal directions, assuming a bidimensional structured grid. Therefore, the snapshot matrix $\mathbf{V}_1^K$ of dimensions $J \times K$ can be organised as follows

$$\mathbf{V}_1^K = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_{K-1}, \mathbf{v}_K], \quad (1)$$

HODMD [11] is an extension of dynamic mode decomposition (DMD) [12] used for the analysis of complex flows, showing potential in the study of various types of flows; such as non-newtonian turbulent flows [13] or reacting flows [14] and is also capable to identify global instabilities triggering changes in the flow [15, 16]. The algorithm decomposes spatio-temporal data $\mathbf{v}_k$ into a Fourier-like expansion of M DMD modes $\mathbf{u}_m$ as

$$\mathbf{v}(x, y, z, t_k) \simeq \sum_{m=1}^M a_m \mathbf{u}_m(x, y, z) e^{(\delta_m + i\omega_m)t_k}, \quad (2)$$

for $k = 1, \dots, K$, where a_m is the amplitude, ω_m is the frequency and δ_m corresponds to the growth rate of the mode $\mathbf{u}_m$.

A more detailed description and applications of this algorithm can be found in [17], including the implementation of this algorithm in Matlab. More recently a version of the algorithm has been implemented in Python by [18].

Considering HODMD as the basis of the passive flow control algorithm, the method for calculating the so-called non-linear structural sensitivity can be summarised in four main steps:

Step 1: filtering the dynamics of the snapshot matrix. At this step, HODMD is applied to the snapshot matrix $\mathbf{V}_1^K$ to calculate the HODMD expansion (Eq. 2). The HODMD expansion will only retain the three highest amplitude modes: the mean flow (mode with $\omega = 0$ and the leading frequency mode with its complex conjugate. In this way, the dynamics of the complex flow are simplified and the noise is reduced. With the HODMD expansion, a filtered snapshot matrix $\tilde{\mathbf{V}}_1^K$ is obtained. The leading frequency mode contains information about the main flow dynamics. The structural sensitivity will approximate the parts of the domain where it is possible to produce the strongest changes in the flow, controlling the shape and even the presence of this mode.

Step 2: Direct HODMD. HODMD is applied to the filtered snapshot matrix $\tilde{\mathbf{V}}_1^K$, obtaining the *direct* HODMD expansion.

Step 3: Adjoint HODMD. To simplify the computation, the mean is subtracted to the adjoint snapshot matrix, defined as $\tilde{\mathbf{V}}_1^{K*} = \tilde{\mathbf{V}}_1^K$, before applying the HODMD method. The mean is the same in both cases, the direct and the adjoint matrices.

Step 4: non-linear structural sensitivity. The structural sensitivity, called as *non-linear structural sensitivity* is calculated as

$$\tilde{\mathbf{S}}^{NL} = \omega \|\mathbf{M}\| \cdot \|\mathbf{M}\| + 2 \|\mathbf{N}\| \cdot \|\mathbf{N}^*\|, \quad (3)$$

where $\mathbf{M} = \bar{v}_j \frac{\partial \bar{v}_i}{\partial x_j}$, being $\bar{\mathbf{v}} = (\bar{v}_i, \bar{v}_j)$ the mean flow (HODMD mode with $\omega = 0$) of the direct solution and $\mathbf{x} = (x_i, x_j)$ the spatial dimensions. The term $\mathbf{N} = v'_j \frac{\partial v'_i}{\partial x_j}$, being $\mathbf{v}' = (v'_i, v'_j)$, is the direct mode with frequency ω and lastly, $\mathbf{N}^* = v_j^* \frac{\partial v_i^*}{\partial x_j}$, where $\mathbf{v}^* = (v_i^*, v_j^*)$, is the adjoint mode with frequency ω . The HODMD modes are weighted with their corresponding amplitudes, and normalized with their maximum value.

More details about the implementation of this algorithm for passive flow control can be found in [10], as well as different applications in fluid dynamics, comparing the results with classical linear methods.

3 SIMULATION SETUP

The studied case is the oxidation of a nitrogen-diluted methane jet. CFD simulations have been carried out for this laminar flame. The oxidizer is regular air injected into the domain at constant velocity of 35 cm/s , while the fuel, composed by 65% of methane and 35% of nitrogen, is injected according to the following sinusoidal parabolic velocity profile:

$$v(r, t) = v_{max} \left(1 - \frac{r^2}{R^2} \right) [1 + A \sin(2\pi f t)], \quad (4)$$

where v_{max} is the maximum velocity ($v_{max} = 75 \text{ cm/s}$), r is the radial coordinate, R is the internal radius of the nozzle ($R = 2 \text{ mm}$), A is the amplitude of the sinusoidal perturbation ($A = 0.25$) and f is the frequency of the perturbation ($f = 10 \text{ Hz}$). The detailed kinetic mechanism used for the numerical simulation is **GRI-Mech 3.0** [19], taking into account 325 chemical reactions and 53 species, including the formation of oxides of nitrogen NO_x . The CFD

simulations were carried out using the solver *reactingFoam*, an OpenFOAM solver that accounts for transient, non-premixed combustion problems. A representative snapshot of the temperature is shown in Fig. 1. The results of the numerical simulation have been validated with previous studies [20, 21].

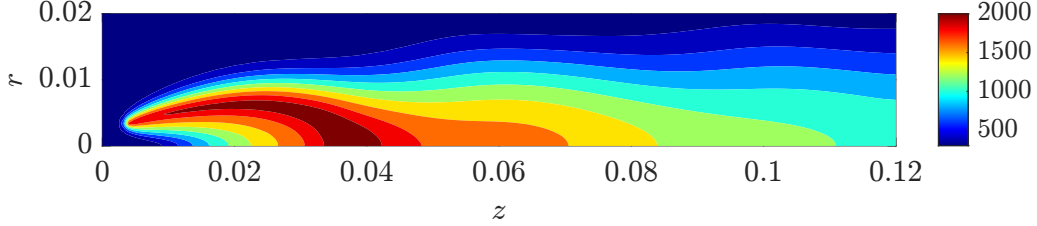


Figure 1: Instantaneous image of the temperature field of the numerical simulation analysed.

4 RESULTS

In this section, the flow control algorithm is applied to the simulation explained above. For that purpose, a database is extracted, consisting of snapshots with 19 variables (velocity field, temperature, pressure and 15 chemical species) and two spatial dimensions composed by a 150×150 grid points along the axial and radial directions. The snapshots are extracted each $\Delta t = 0.001s$ equidistant in time and a total of 1000 snapshots are extracted. Therefore the database is a fourth order tensor of dimensions $19 \times 150 \times 150 \times 1000$.

4.1 Application of HODMD and the non-linear structural sensitivity

The order of magnitude of the variables ranges from 10^5 (pressure field) to 10^{-6} (N_2O concentration), therefore the database needs to be pre-processed. The variables were centered to focus on the temporal perturbations and scaled with the range of each variable, as

$$\tilde{s}_j = \frac{s_j - \bar{s}_j}{\max(s_j) - \min(s_j)}, \quad (5)$$

where s_j is the j -th variable, $\bar{s}_j$ is its temporal mean, $\max(s_j)$ and $\min(s_j)$ are the maximum and minimum value; and $\tilde{s}_j$ is the pre-processed variable, following the advice proposed in [22, 14].

The application of HODMD to the preprocessed database can be found in Fig. 2. As it can be seen the spectrum is clearly periodic, with a frequency of $\omega \sim 63$. The frequency, as expected is equal as the one induced in the boundary condition of the fuel ($f = \omega/2\pi = 10Hz$). For the analysis of the modes, just the velocity is shown, although the analysis has been done to the 19 variables at the same time, and similar analysis can be carried out. The HODMD mode presents some coherent structures along the center line ($r = 0$), showing the movement of the reacting flow downstream.

From Eq.3, the non-linear structural sensitivity can be calculated. For the present work, both terms are going to be taken into account separately, to observe the structures provoked by the mode with 0 frequency and the mode with frequency ω . In this case, for the first term, the temporal mean of the simulation is used, as it has been removed from the database during the

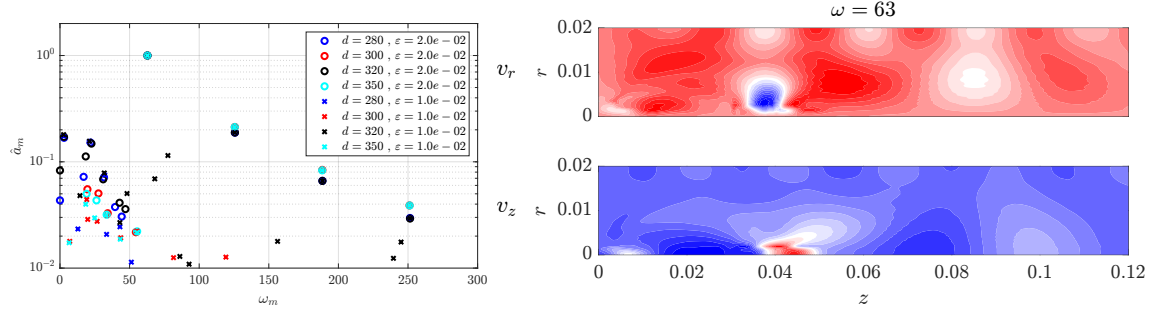


Figure 2: **Left:** Frequencies vs. Amplitudes corresponding to the HODMD modes of the studied database. **Right:** Real part of the leading mode for the velocity field, being v_r and v_z the radial and axial component, respectively. The mode has been normalized between values -1 and 1.

pre-processing. For the calculation of the sensitivity, just the velocity field is taken into account, however as the system is coupled, changes in the velocity should change the other variables.

In Fig. 3, the contours of the two terms are plotted. From the first term, the region with highest structural sensitivity appears between the two jets, while on the term related to the non-stationary mode two regions appear: one in the near field, close to the injection of fuel, and the other one, in the area where the HODMD mode has its highest amplitude. Based on these results, we study the changes produced in the flow in three different test cases, inserting a small circular cylinder (a punctual force is also valid) in the areas with the highest structural sensitivity. The position at which this cylinder is located in these three cases are the following:

1. **Case 1:** $x = 0.005$; $z = 0.004$
2. **Case 2:** $x = 0.001$; $z = 0.005$
3. **Case 3:** $x = 0.001$; $z = 0.037$

The location of the cylinders can be found in the bottom part of Fig. 3. The diameter of the cylinders is $0.0038m$, the same as the thickness of the wall separating the two jets. Other diameters can be chosen following the different characteristic length scales from the case under study (i.e., based on the temperature field, the diameter of the inner jet or the diameter of the outer jet). However this topic will remain open for future studies.

4.2 Analysis of the controlled cases

In this section, the three cases with a little cylinder are analysed. The efficiency and quality indicators are chosen attending to the different applications of the studied case. In particular, we analyse the temporal mean combustion efficiency and the concentration of NOx at the outlet. The efficiency is calculated as follows

$$\eta = \frac{1}{T} \int_0^T \left(1 - \frac{\dot{m}_{CO}Q_{CO} + \dot{m}_{H_2}Q_{H_2} + \dot{m}_{HC}Q_{HC}}{cL} \right) dt, \quad HC = CH_4, C_2H_2, C_2H_4, C_2H_6, \quad (6)$$

where T is the final time, $\dot{m}_i = Y_i \dot{m}_{out}$ and Q_i are the outlet mass flux and the lower heating value of species i , respectively, $c = Y_{CH_4} \dot{m}_{in}$ is the fuel mass flux at the inlet and L is the lower

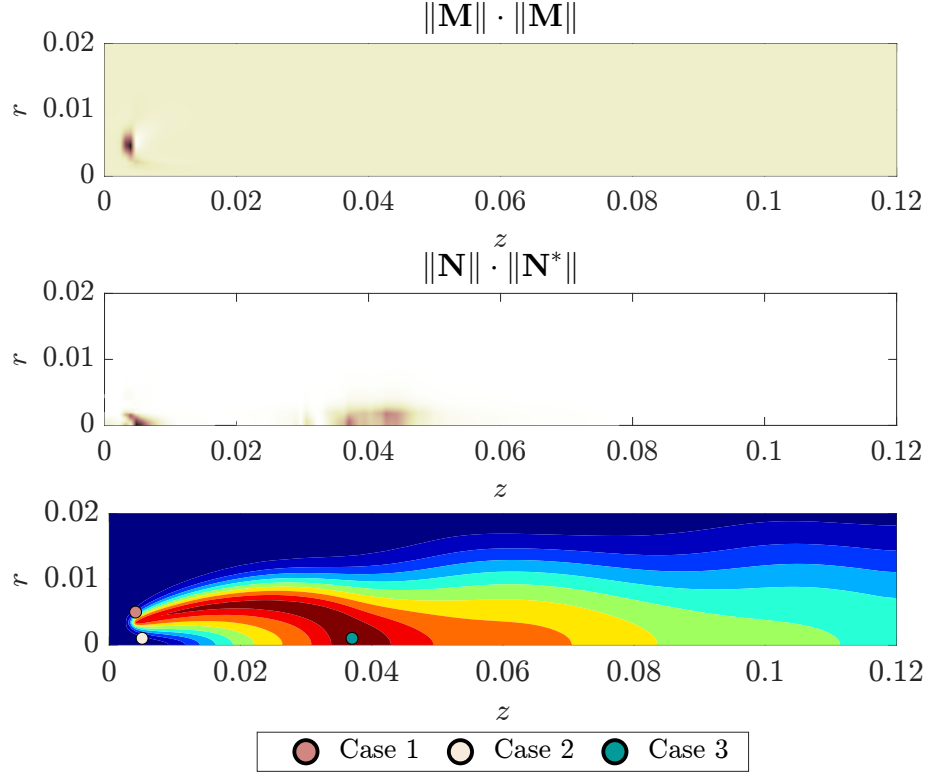


Figure 3: From top to bottom: Contour of the first term of the non-linear structural sensitivity, contour of the second term of the non-linear structural sensitivity and localization of the three cylinders over a representative snapshots of the temperature (the cylinders are bigger than in the simulations for a better visualization).

heating value of methane. This equation expresses how far is the full conversion of the fuel to the main products, which are H_2O and CO_2 . Unburnt fuel or products will decrease the efficiency.

In Tab. 1 the vales of efficiency and NO_x production are presented. Apart from the total NO_x production, also the two main components of NO_x are presented, i.e. NO and NO_2 . The first case does not provide any optimum value, however, the NO_x production is lowered, in the total sum and in each component. The third case lowers the efficiency, however it decreases the NO production at the expense of increasing the NO_2 production. Lastly, the second case is best for overall combustion efficiency, including global NO_x emissions.

For understanding the dynamics of each case and the possible explanation of this increase in the efficiency or decrease in the NO_x emissions, the three cases are analyzed with HODMD. The spectrum in the three cases is periodic with the main frequency the same as the initial case $\omega = 63$. This is trivial, as the flow is forced in the boundary condition with that frequency, therefore the frequency and the mode related are going to appear always.

In Fig. 4, the shape of the leading mode in each one of the three cases studied is presented. The first case show minor changes in the mode, mostly on the radial component of the velocity v_r and on the center line, where the initial case exhibits the greatest amplitude. The strongest

	η [%]	NO [ppm]	NO_2 [ppm]	NO_x [ppm]
Initial	93.61	51.77	12.53	64.30
Case 1	92.33	47.82 (-8%)	11.85 (-5%)	59.60 (-7%)
Case 2	96.70	45.40 (-12%)	9.51 (-24%)	54.91 (-15%)
Case 3	91.74	44.26 (-15%)	17.60 (+40%)	61.92 (-4%)

Table 1: Summary of the mean efficiency of the flame and the production of NO_x for the initial case and the three cases studied applying the flow control algorithm.

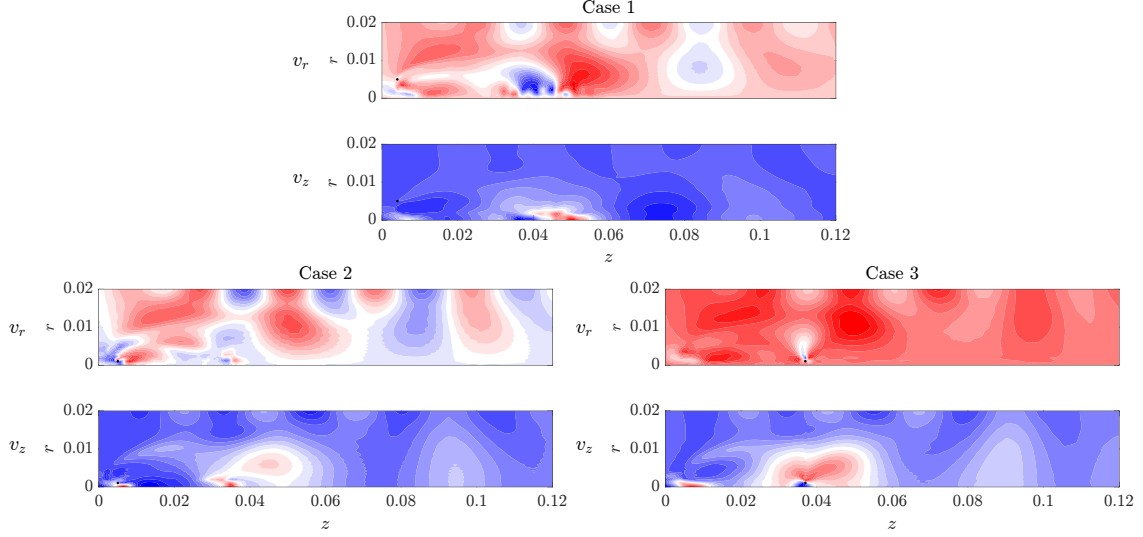


Figure 4: Real part of the leading mode for the velocity field in the three cases studied after applying the non-linear structural sensitivity algorithm, being v_r and v_z the radial and axial component, respectively. The mode has been normalized between values -1 and 1.

change in the shape of the mode appears in the second case, mostly on v_r , where the complexity of the mode increases considerably, suggesting a better mixing. Lastly in the third cases, the changes in the mode are mostly on the axial component of the velocity v_z , conducted by the appearance of a recirculation bubble downstream the cylinder inserted at $x = 0.001$; $z = 0.037$. Although this case has the lowest NO emissions, this significant change in the mode may be correlated with the highest NO_2 emissions. Further studies need to be carried out to understand what is the main mechanism triggering the reduction of emissions and rising the efficiency in the flame, but this is out of the scope of this paper.

5 CONCLUSIONS

The present work presents for the first time to the authors' knowledge a data driven algorithm for passive flow control in laminar flames. The algorithm locates the parts of the domain where the flow is more susceptible to changes. The method, as it is fully data-driven, is generalizable for different conditions of the flow and different geometries. The zones where the passive control is actuating provokes changes in the dynamics of the flow.

In particular, the algorithm has been applied to an axisymmetric, time varying, non-premixed

laminar co-flow flame. The flame has been studied with HODMD and it provides a periodic spectrum, due to the boundary condition of the fuel. The main mode locates in the center line.

The non-linear structural sensitivity provides three different parts of the domain where a little cylinder can be inserted for passive flow control. The results show a three point increase in the combustion efficiency and a reduction of 15% in the NO_x emissions, some of the most harmful pollutants. Applying HODMD to the three cases, it can be seen that the second case shows the strongest change in the mode, which appears on top of the oxidizer jet, in the main dynamics of the problem.

The results presented show great potential for the algorithm to be used in flow control applications and the design of components for pollutant reduction. However, the study should be extended to consider the influence that varying the size of the cylinder would have on the flow, as well as the inclusion of other variables in the gas mixture in the calculation of the non-linear structural sensitivity.

ACKNOWLEDGEMENTS

The authors acknowledge the grants PID2020-114173RB-I00, TED2021- 129774B-C21 and PLEC2022-009235 funded by MCIN/AEI/ 10.13039/501100011033 and by the European Union “NextGenerationEU”/PRTR, and S.L.C. acknowledges the support of Comunidad de Madrid through the call Research Grants for Young Investigators from Universidad Politécnica de Madrid. A.C. acknowledges the support of Universidad Politécnica de Madrid, under the program ‘Programa Propio’. The MODELAIR and ENCODING project has received funding from the European Union’s Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101072559 and 101072779, respectively. The results of this publication reflect only the author(s) view and do not necessarily reflect those of the European Union. The European Union can not be held responsible for them. The authors gratefully acknowledge the Universidad Politécnica de Madrid (www.upm.es) for providing computing resources on Magerit Supercomputer.

REFERENCES

- [1] Koumoutsakos, P. and Mezic I., 2006. “Control of Fluid Flow.” Lecture Notes in Control and Information Sciences. <https://doi.org/10.1007/978-3-540-36085-8>
- [2] Noack, B.R., Morzynski, M. and Tadmor G., 2011. “Reduced-Order Modelling for Flow Control.” Springer Science and Business Media, New York, 528. <https://doi.org/10.1007/978-3-7091-0758-4>
- [3] Cant, S., 2002. “High-performance computing in computational fluid dynamics: progress and challenges.” Philos. Transact. A Math. Phys. Eng. Sci., 360(1795), 1211—1225. <https://doi.org/10.1098/rsta.2002.0990>
- [4] Karagozian, A., 1999. “Control of mixing and reactive flow processes.” 30th Fluid Dynamics Conference. <https://doi.org/10.2514/6.1999-3572>
- [5] Beér, J.M. and Chigier, N.A., 1983. “Combustion Aerodynamics.” Fuel and energy science series. <https://doi.org/10.1017/S0022112072210990>

- [6] Karagozian, A.R., Wang, K.C., Le, A.-T. and Smith, O.I., 1996. "Transverse gas jet injection behind a rearward-facing step." *J. Propul. Power*, 12(6), 1129–1136. <https://doi.org/10.2514/3.24153>
- [7] Giannetti, F. and Luchini, P., 2007. "Structural sensitivity of the first instability of the cylinder wake." *J. Fluid Mech.*, 581, 167–197. <https://doi.org/10.1017/S0022112007005654>
- [8] Varela, P., Suárez, P., Alcántara-Ávila, F., Miró, A., Rabault, J., Font, B., García-Cuevas, L.M., Lehmkuhl, O. and Vinuesa, R., 2022. "Deep Reinforcement Learning for Flow Control Exploits Different Physics for Increasing Reynolds Number Regimes." *Actuators*, 11, 359. <https://doi.org/10.3390/act11120359>
- [9] Guastoni, L., Rabault, J., Schlatter, P., Azizpour, H. and Vinuesa R., 2023. "Deep reinforcement learning for turbulent drag reduction in channel flows." *Eur. Phys. J. E*, 46, 27. <https://doi.org/10.1140/epje/s10189-023-00285-8>
- [10] Corrochano, A. and Le Clainche, S., 2022. "Structural sensitivity in non-linear flows using direct solutions." *Comput. Math. Appl.*, 128, 69–78. <https://doi.org/10.1016/j.camwa.2022.10.006>
- [11] Le Clainche, S. and Vega, J.M., 2017. "Higher Order Dynamic Mode Decomposition." *SIAM J. Appl. Dyn. Syst.*, 16(2), 882–925. <https://doi.org/10.1137/15M1054924>
- [12] Schmid, P., 2010. "Dynamic mode decomposition of numerical and experimental data." *J. Fluid Mech.*, 656, 5–28. <https://doi.org/10.1137/15M1054924>
- [13] Amor, C., Corrochano, A., Foggi, G., Rosti, M.E. and Le Clainche, S., 2024. "Coherent structures in elastic turbulent planar jets." *J. Phys.: Conf. Ser.*, 2753, 012020. <https://dx.doi.org/10.1088/1742-6596/2753/1/012020>
- [14] Corrochano, A., D'Alessio, G., Parente, A. and Le Clainche, S., 2023. "Higher order dynamic mode decomposition to model reacting flows." *Int. J. Mech. Sci.*, 249, 108219. <https://doi.org/10.1016/j.ijmecsci.2023.108219>
- [15] Le Clainche, S., Han, Z. and Ferrer, E., 2019. "An alternative method to study cross-flow instabilities based on high order dynamic mode decomposition" *Phys. Fluids*, 31 (9). <https://doi.org/10.1063/1.5110697>
- [16] Le Clainche, S., Rosti, M., and Brandt, L., 2022. "A data-driven model based on modal decomposition: application to the turbulent channel flow over an anisotropic porous wall", *J. Fluid Mech.*, 939, A5. <https://doi.org/10.1017/jfm.2022.159>
- [17] Vega, J.M. and Le Clainche, S., 2020. "Higher order dynamic mode decomposition and its applications." Elsevier. <https://doi.org/10.1016/C2019-0-00038-6>
- [18] Hetherington, A., Corrochano, A., Abadía-Heredia, R., Lazpita, E., Muñoz, E., Díaz, P., Maiora, E., López-Martín, M. and Le Clainche, S., 2024. "ModelFLOWs-app: Data-driven post-processing and reduced order modelling tools." *Comput. Phys. Commun.*, 301, 109217. <https://doi.org/10.1016/j.cpc.2024.109217>

- [19] Smith, P.G., 1999. "GRI-Mech 3.0." [http://www. me. berkley. edu/gri_mech/](http://www.me.berkley.edu/gri_mech/)
- [20] D'Alessio, G., Parente, A., Stagni, A. and Cuoci, A., 2020. "Adaptive chemistry via pre-partitioning of composition space and mechanism reduction." *Combust. Flame*, 211, 68–82. <https://doi.org/10.1016/j.combustflame.2019.09.010>
- [21] D'Alessio, G., Cuoci, A., Aversano, G., Bracconi, M., Stagni, A. and Parente, A., 2020. "Impact of the partitioning method on multidimensional adaptive-chemistry simulations." *Energies*, 13(10), 2567. <https://doi.org/10.3390/en13102567>
- [22] Parente, A. and Sutherland, J.S., 2013. "Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity." *Combust. Flame*, 160(2), 340–350. <https://doi.org/10.1016/j.combustflame.2012.09.016>

Chapter 6

General Discussion

In this Chapter, a discussion explaining the main results, the descriptions of the algorithms adapted and developed for combustion databases, the main conclusions and future work is explained.

6.1 Adaptation of HODMD to model reacting flows

This section shows the application of HODMD to combustion databases. It first provides an overview of the potential of the algorithm to model combustion to develop ROMs in a conference proceedings. After that, a deep analysis of the application of HODMD to reacting flows databases is published in a research paper.

This first analysis with HODMD will provide the indispensable understanding of the leading patterns of the main dynamics. This knowledge will be essential to then develop the new algorithms in the next sections for clustering and flow control.

6.1.1 Using HODMD for the first time to model reacting flows

A first analysis with HODMD of a combustion database was carried out and the main results were presented in the *International Symposium on Unmanned Systems and the Defense Industry*, in 2021. From this talk, an extended version was published in a conference proceedings in 2023, including the main findings obtained from the research.

In that work, HODMD is applied for the first time to the authors' knowledge to a combustion database to assess the performance of the algorithm in reconstructing the main flow patterns. In this first conference proceedings, the main goal is to evaluate the suitability of HODMD to develop ROMs in combustion.

The flame examined is an axisymmetric, time varying, non-premixed laminar co-flow flame, with a sinusoidal perturbation in the velocity profile of the injection of the fuel of 10 Hz . The database comprises $\sim 0.12\text{ s}$, therefore, it has just one entire cycle of the perturbation, what makes the analysis challenging. The detailed kinetic mechanism is POLIMI_C1C3_HT_1412 (Ranzi et al., 2012), which takes into account 82 species and 1698 reactions.

Before applying HODMD, the database is preprocessed with the range of each variable, which enhances the performance for the main species, which contains the main flow dynamics. The algorithm is properly calibrated and the reconstruction of the different variables is studied quantitatively, with the RRMSE (2.20), and qualitatively, with the comparison of a representative snapshot.

The novelty of this research relies on the temporal decomposition provided by means of HODMD. Usually, machine learning techniques as PCA focus on the variability of the different chemical species, but they do not focus on the effects of the flow dynamics. However, HODMD is a data-driven method used for the identification of flow patterns and the development of ROMs based on the flow physics.

The work extends the applicability of this algorithm from fluid dynamics databases, which usually contains a maximum of 4 variables, typically the velocity field and the pressure, to combustion databases, which contain in this case 83 variables (the temperature and 82 chemical species).

The main flow dynamic of the flow have found to be periodic, with a frequency of $\omega_p = 60$. This frequency is the perturbation frequency of the inlet ($f_p = \omega_p/(2\pi) \simeq 10$). Although the main frequency is the expected before the analysis, the accuracy of the frequency is challenging due to the lack of information in the temporal dimension, just one cycle of the perturbation. In comparison with other algorithms for Fourier analysis, as Fourier Fast Transform (FFT) and Power Spectral Density (PSD, Press et al., 1988), the sampled timespan needs to be sufficiently large or to coincide with the period (Le Clainche et al., 2017).

The best reconstructed variables are the main species (air, fuel, CO_2 and water) and the temperature, with a RRMSE $\leq 1.2\%$. Moreover the nitrogen has a reconstruction error below 0.1%. On the other hand some radicals have a high reconstruction error, over 50%.

This high value of the reconstruction error can be explained looking to the representative snapshot and its reconstruction. While the main shape of the structures is recovered, HODMD faces problems when reconstructing the part of the domain with zero constant value.

Even though, as it is able to reconstruct the main structures appearing in these variables, HODMD is suitable to model combustion. This is related to one of the main goals of the Doctoral Thesis, to extend the applicability of HODMD from fluid dynamics to combustion.

6.1.2 Thorough analysis of a flame using HODMD

Given the potential of HODMD for modelling combustion shown in the previous Section, an extended study was conducted. The findings were published in 2023 in the *International Journal of Mechanical Sciences*. To the best of the authors' knowledge, this study is the first to present the application of HODMD for identifying the main dynamics presented in a combustion database.

In the mentioned work, an extensive study of the application of HODMD to a laminar flame is carried out. The laminar flame studied is the same as in Section 6.1.1, with the same kinetic mechanism.

In this case, two different scaling methods are evaluated as pre-processing techniques, the range of each variable and the standard deviation. This later scaling method, named as auto scaling, gives all the variables similar importance. The pre-processing techniques follows Eq. (2.1) and more information about the scaling methodology can be found in Sec. 2.1.

After the pre-processing, multi-dimensional HODMD (mdHODMD, Section 2.3) is applied and the algorithm is studied at each one of the two main steps of the algorithm. Firstly, HOSVD is applied to the database following Eq. (2.21), and the evaluation of the two scaling methods and the truncation in the dimension of the variables is carried out. The temporal modes are introduced in the DMD-d algorithm, following the algorithm in Sec. 2.2.2. The performance of the DMD-d algorithm for both scaling methods is studied, comparing the DMD modes obtained and the reconstruction.

The robustness of the developed reduced order model is studied by coupling HODMD with PCA with varimax rotation. PCA with varimax rotation selects progressively the number of variables to be analysed and HODMD is applied, calculating the RRMSE (2.20), for each one of these new cases.

The main goal of this work is to provide a new tool, extending its applicability from fluid dynamics to combustion, capable of modelling reacting flows accurately at a reduced computational cost. This tool, HODMD, is capable to capture the main flow physics, identifying the main frequencies and patterns driving the flow dynamics. This is important from the perspective of developing robust ROMs based on the physics of the system. This knowledge could also be utilized in future flow control applications to enhance system efficiency.

A summary of the paper can be found in Fig. 6.1. As it can be seen, the original database (a) is first preprocessed following Eq. (2.1) and the advice presented in Sec. 2.1 (b). Then, HODMD is applied to the database as in Sec. 2.3. First, HOSVD is applied to reduce the dimensionality of the system (c), following Eq. (2.21). After that, taking the temporal modes as the input for the DMD-d algorithm (Sec. 2.2.2), the DMD modes and frequencies are calculated (d). The DMD reconstruction is obtained following the DMD expansion (e), Eq. (2.2) and the original dimensions, scaled during the preprocessing, are recovered to obtain the reconstructed database (f).

The work extends the applicability of mdHODMD to combustion databases. The flame examined is an axisymmetric, time varying, non-premixed laminar co-flow flame. The oxidizer is air and it is injected with constant velocity at 35 cm/s. On the contrary, the fuel is injected with a parabolic profile with a sinusoidal perturbation, as

$$v(r, t) = v_{max} \left(1 - \frac{r^2}{R^2} \right) [1 + A \sin(2\pi f t)], \quad (6.1)$$

where v_{max} is the maximum velocity ($v_{max} = 70$ cm/s), r is the radial coordinate and R represents the internal radius of the nozzle, f and A are the frequency ($f = 10Hz$) and the amplitude ($A = 0.25$) of the perturbation, respectively, and finally, t is the time. The detailed kinetic mechanism is POLIMI_C1C3_HT_1412 (Ranzi et al., 2012), which takes into account 82 species and 1698 reactions. More details about the numerical simulation and setup can be

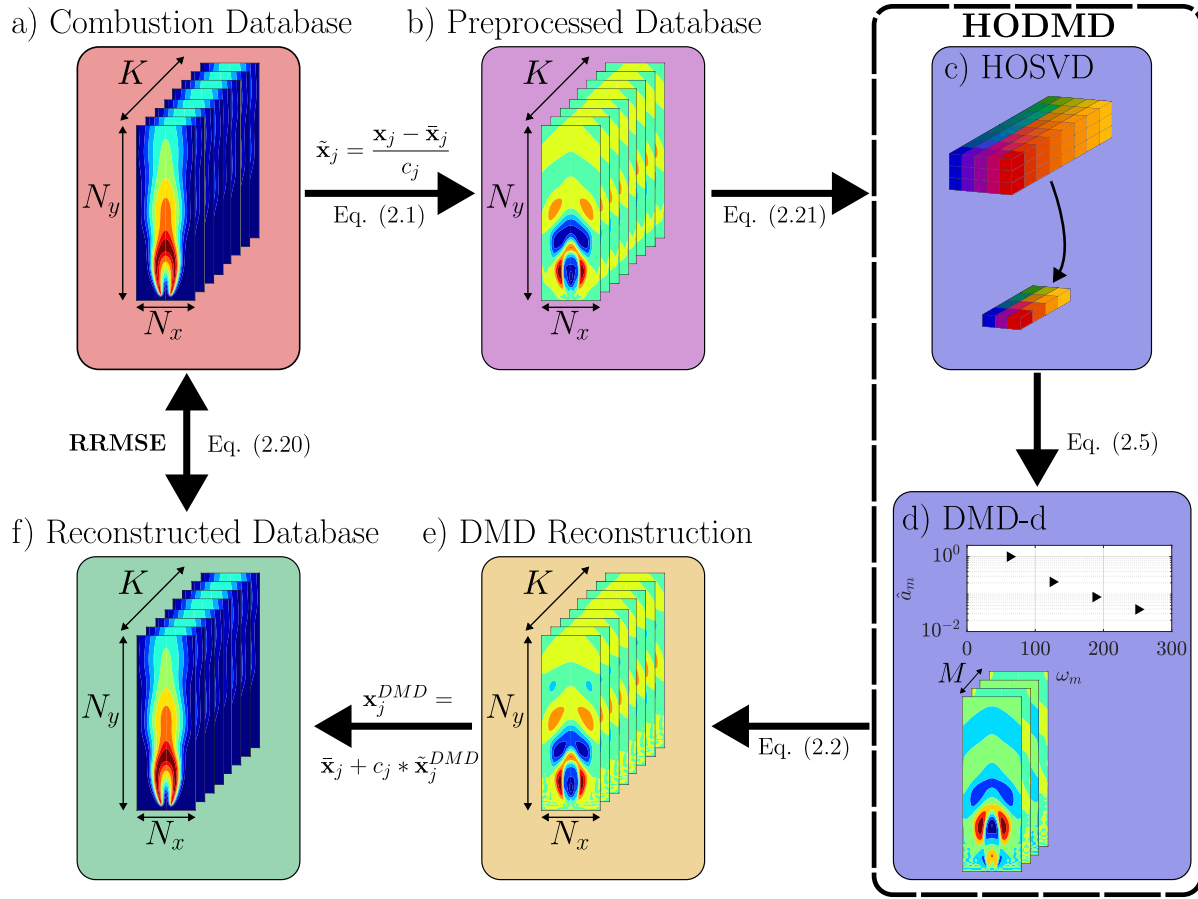


Figure 6.1: Summary of the process of the HODMD analysis of a combustion database..

found in D'Alessio et al., 2020.

From the database extracted, which comprises the temperature and the 82 chemical species, the dimension of the variables ranges several orders of magnitude, from 10^3 , in the case of the temperature, to 10^{-11} , in the case of the smallest radical computed, RALD3G. This means that there are 13 orders of magnitude difference, highlighting the importance of scaling the variables, as stated in Sec. 2.1.

In fluid dynamics, it is common to analyse only the velocity field and the pressure, therefore it is not necessary to truncate the dimension of the variables (first dimension of the tensor). However, when applying the algorithm to a combustion database, the number of variables sharply increases. In the present case, 83 variables are taken into account for the analysis. In this case, it makes sense to truncate this first dimension of the tensor when applying HOSVD (Sec. 2.3). The SVD modes obtained in this first dimension of the tensor can be related to the ones obtained in PCA.

The decay of the singular values is more pronounced when applying range scaling as the pre-processing technique than when applying auto scaling, as auto scaling give similar importance to all the variables. Range scaling, on the contrary, gives more relative importance to the main species and the temperature. This means that for a similar RRMSE, the compression

factor is higher when using range scaling. The compression factor is calculated as

$$CF = \frac{N_e}{N_{er}}. \quad (6.2)$$

This compression factor measure the ratio between the total number of elements in the tensor and the number of elements retained after applying HOSVD. Therefore, the higher is the compression factor, the greatest is the compression.

The performance of HODMD (and the DMD-d algorithm, Sec. 2.2.2) is evaluated taking a different tolerance ε_1 for each scaling method. This tunable tolerance is the one determining the number of SVD modes retained, Eq. (2.7). It is selected to have a similar compression factor.

Comparing the spectra, both scaling methods provide a clearly periodic spectrum, but with different frequency. The principal mode using range scaling is $\omega_p^{RANGE} \simeq 60$, and the one using auto scaling is $\omega_p^{AUTO} \simeq 50$. The difference between both frequencies is due to the transient regime of the simulation and the variables each scaling method enhance. When coupling HODMD with PCA for variable selection, at the end of the paper, there is also a difference when analysing the main frequency. While auto scaling always provides the same frequency independently of the number of variables selected, range scaling provides three different values of the main frequency, depending on the number of variables selected. This suggests once again that the main variables and the radicals may be driven by different dynamics, at least in the transient regime of the numerical simulation.

When comparing the DMD modes, both scaling methods provide slightly different modes. While range scaling DMD modes focus more on the wake downstream of the flow, the auto scaling DMD modes highlights the part where the oxidation reaction begins, i.e., where the oxidizer and fuel merges.

Finally, calculating the reconstruction error for each variable analysed, following Eq. (2.20), range scaling provides lower RRMSE on the main variables, as expected. These main variables are:

- The temperature: T
- The fuel components: CH_4
- The air components: N_2 and O_2
- The main oxidation products: CO_2 and H_2O

On the contrary, auto scaling performs better on the reconstruction of the radicals, as it gives a similar relative importance to all the variables. From a qualitatively perspective, both scaling methods provide good results when comparing the reconstruction of a representative snapshot.

6.2 New applications based on HODMD to model reacting flows

6.2.1 Hierarchical higher-order dynamic mode decomposition for clustering and feature selection

Once it has been proved the suitability and robustness of HODMD when applied to combustion databases, novel variants of this main algorithm have been developed in the present Doctoral Thesis. The first variant, known as *Hierarchical higher-order dynamic mode decomposition* (h-HODMD), is developed for the clustering of variables and feature extraction. Throughout this Section, the algorithm explained in Sec. 2.4 is developed and applied to different combustion databases. From this work, a research paper was published in early 2024, in the journal *Computers & Mathematics with Applications*.

In the paper published in Chapter 4, the novel algorithm explained in Sec. 2.4 is applied to a combustion database. The algorithm makes clusters of variables based on the reconstruction error when applying HODMD. Each cluster has associated its frequencies and DMD modes.

The flame examined is the same as in the previous chapter, an axisymmetric, time varying, non-premixed laminar co-flow flame with a sinusoidal perturbation in the velocity profile of the fuel’s injection. However, in this work, three different CFD simulations have been carried out, with different kinetic mechanisms and main frequency of the perturbation. The details of each simulation can be found in Tab. 6.1.

Table 6.1: Frequency of the perturbation in Hertz, kinetic mechanism and the number of variables and reactions taken into account in each CFD simulation.

	Frequency [Hz]	Kinetic mechanism	Number of species	Number of reactions
Simulation 1	10	POLIMI_C1C3_HT_1412 (Ranzi et al., 2012)	84	1698
Simulation 2	10	GRI-Mech 3.0 (Smith, 1999)	53	325
Simulation 3	20	GRI-Mech 3.0 (Smith, 1999)	53	325

Two types of analysis have been carried out to test the performance of h-HODMD. First, the algorithm has been applied in each one of the three databases to test its ability to construct a ROM and reconstruct with precision the databases. Additionally, in a reduced set of variables, where the behaviour is known, the clusters are analysed to test its capability to group the variables from the perspective of their kinetics.

This new iterative algorithm for clustering and feature extraction based on HODMD has demonstrated potential in analysing high-dimensional data, such as those generated from

reacting flow simulations that use detailed kinetic mechanisms. Specifically, the proposed method, h-HODMD, is significant when developing ROMs based on the underlying physics. This approach facilitates the exploration of novel strategies to address the challenges of increasingly large databases, while simultaneously reducing the associated computational costs in terms of CPU time and memory usage.

The algorithm is applied to the three databases obtained from the simulations in Tab. 6.1. Firstly, the databases are pre-processed by means of different centering and scaling operations, as in Sec. 2.1; afterwards, h-HODMD is applied following the algorithm explained in Sec. 2.4 and illustrated in Fig. 2.3. As a summary, once pre-processed, the algorithm applies HODMD to the database, a fourth order tensor, and reconstructs every variable following the DMD expansion, as in Eq. (2.2). Then, the reconstruction error is calculated following Eq. (2.20) and based on a tunable tolerance ε_v , the best reconstructed variables are retained, the ones which verify Eq. (2.23). These variables are separated from the rest of the database, and HODMD is applied to the rest of the variables, extracting a subset of variables each time HODMD is applied until all the variables are clustered.

In the present work, 12 different cases are evaluated, as 4 different scaling methods are evaluated in the 3 different databases. In a first study, the lowest reconstruction error is sought and the different frequencies obtained in each iteration are analysed. In this sense, as the databases are extracted in the transient state of the numerical simulation, the leading frequencies are calculated using the index I_m , calculated as

$$I_m = \sum_{k=1}^K |a_m e^{(\delta_m + i\omega_m)t_k}| \|u_m\|_F^2 \times \Delta t, \quad (6.3)$$

where $\|\bullet\|_F^2$ is the square of the Frobenius norm. The index also takes into account the growth or decay of the modes over time with the growthrate δ_m . A much more detailed description about this criterion can be found in Kou et al., 2018.

The performance of h-HODMD is compared with the standard HODMD. The main results of this first main study can be summarised as follows:

- Auto scaling ensures that all the clusters have the same dynamics. However, it is not suitable to improve the reconstruction over HODMD.
- For the rest of scaling methods, the reconstruction with h-HODMD is significantly better than the one using HODMD.
- Pareto scaling performs similar to auto scaling from the second iteration onward.
- Range scaling gives the lowest reconstruction error when applying h-HODMD and it improves the performance when the transient behaviour is high (first simulation).
- VAST scaling performs better when the transient behaviour decreases (second and third simulation).

A second study is carried out in a reduced set of variables of the second and third simulations. These two simulations are chosen because they present lower transient behaviour and it takes

into account the formation of NO_x . The 15 variables taken into account are the following:

- The temperature: T
- The fuel components: CH_4
- The air components: N_2 and O_2
- The main oxidation products: CO_2 and H_2O
- The nitrogen oxides: NO , NO_2 and N_2O
- The oxi and peroxy radicals: O , OH and HO_2
- Other minor species: CO , H_2O_2 and CH_3

In this reduced subset, h-HODMD improves the reconstruction of the dataset for most of the cases studied, as before. Moreover, an in-depth examination of the variables selected in each cluster, the three main nitrogen oxides (NO , NO_2 , and N_2O) are grouped in a separate cluster and another cluster is identified for oxy- and peroxy radicals (O , OH , and HO_2).

This new iterative algorithm, h-HODMD, allows for the construction of inexpensive ROMs that can predict flame lift-off, identify the occurrence of local extinction and blowout conditions, and facilitate control purposes. This approach facilitates the exploration of new strategies for managing increasingly large databases while minimizing computational costs.

6.2.2 Structural sensitivity in laminar flames using data-driven methods

In this section, a new algorithm to calculate the non linear structural sensitivity is proposed and the application to combustion databases is shown. The structural sensitivity identifies the regions of the spatial domain where the flow is more prone to changes. In the classical theory, the direct (2.26)–(2.27) and adjoint (2.30) linearised Navier Stokes equations are solved to obtain the global stability modes. The interaction of the direct and adjoint modes provide the linear structural sensitivity. This is an extended method for flow control, however this method is limited to laminar non-reactive Newtonian fluids.

This work introduces a novel methodology to calculate the structural sensitivity, with the advantage that the algorithm is purely data-driven, thus it is possible to apply the method to non-linear flows with numerical and experimental data, as it does not require to solve the adjoint problem. The algorithm is explained in detail in Sec. 2.5. At the inception, the algorithm was designed to work on laminar non-reacting flows, as shown in the following Section. However, due to the potential of this algorithm, different types of complex flows are analysed. Section 6.2.2 presents the application of this algorithm to non-Newtonian fluids, as some fuels (gasoline) behave, and reacting flows. The results of the flow control algorithm are shown. The highest structural sensitivity zones are modified in the domain, delaying the instability in the case of the non-Newtonian fluids (Corrochano et al., 2024a), or enhancing the combustion efficiency in reacting flows (Corrochano et al., 2024b).

Development of an algorithm to calculate the structural sensitivity in non-linear flows using direct solutions

The second variant of HODMD developed in the present Doctoral Thesis is examined in this Section. From this work, a research paper was published in late 2022, in the journal *Computers & Mathematics with Applications* and can be found in Sec. 5.1.

In this paper, the algorithm proposed in Sec. 2.5 for flow control is detailed for laminar non-reactive flows. The algorithm proposed is able to calculate the so-called non-linear structural sensitivity without the need of calculate the adjoint solution. This concept is calculated following Eq. (2.44) and following the steps detailed in Sec. 2.5.4.

The algorithm first calculates the DMD modes using HODMD and with this information generates the structural sensitivity. The algorithm has been compared with the classical linear theory, which calculates the structural sensitivity based on the solution of the direct and adjoint linearised NS equations. Results are similar, however there are some discrepancies because we apply the method directly on the non-linear NS equations. The novel algorithm is applied to four different cases: the two and three-dimensional flow past a circular cylinder, the flow through a planar Food and Drug Administration (FDA) nozzle and two axis-symmetric concentric jets with similar velocity ratio.

The method is based on HODMD, therefore the algorithm also allows to calculate the non-linear structural sensitivity in three dimensions, as in the case of the three-dimensional flow past a circular cylinder. The domain of the two concentric jets is similar to the one studied in previous sections, however in this case, the flow is non-reactive and therefore, there is no flame, just hydrodynamic instabilities. The results obtained have been compared with the literature.

STABFEM, 2020 is a Matlab interface to the FreeFem++ software allowing to use this powerful Partial Derivative Equation (PDE) solver within a practical working environment. With this solver, different linear stability analysis have been carried out during this Doctoral Thesis. A collaboration with developers of the open-source code at the IMFT, Toulouse, France, served as learning for the use of this solver to obtain the linear stability modes and the structural sensitivity from the classical theory. From this collaboration, a research paper was published in 2023 (Corrochano et al., 2023a).

For the following paper, the numerical solver STABFEM, 2020, has been used to calculate the linear structural sensitivity in the FDA nozzle and the two concentric jets. The difference between the linear theory and the proposed algorithm are explained in the paper and in the last case, the two concentric jets, a flow control strategy is proposed based on the results of the non-linear structural sensitivity. Finally, with the results obtained in 4 different cases, the method is validated. The results obtained in Sec. 6.2.2 will strengthen the robustness of the method.

Four different cases are evaluated with this novel algorithm. An outline of the geometry, a plot of the streamwise velocity in a representative snapshot of the simulation, and the Reynolds number at which the simulation has been conducted can be found in Fig. 6.2. For the case of the 3D flow past a circular cylinder, two different spanwise lengths have been analysed,

$Lz = 1.66$ and $Lz = 6.99$.

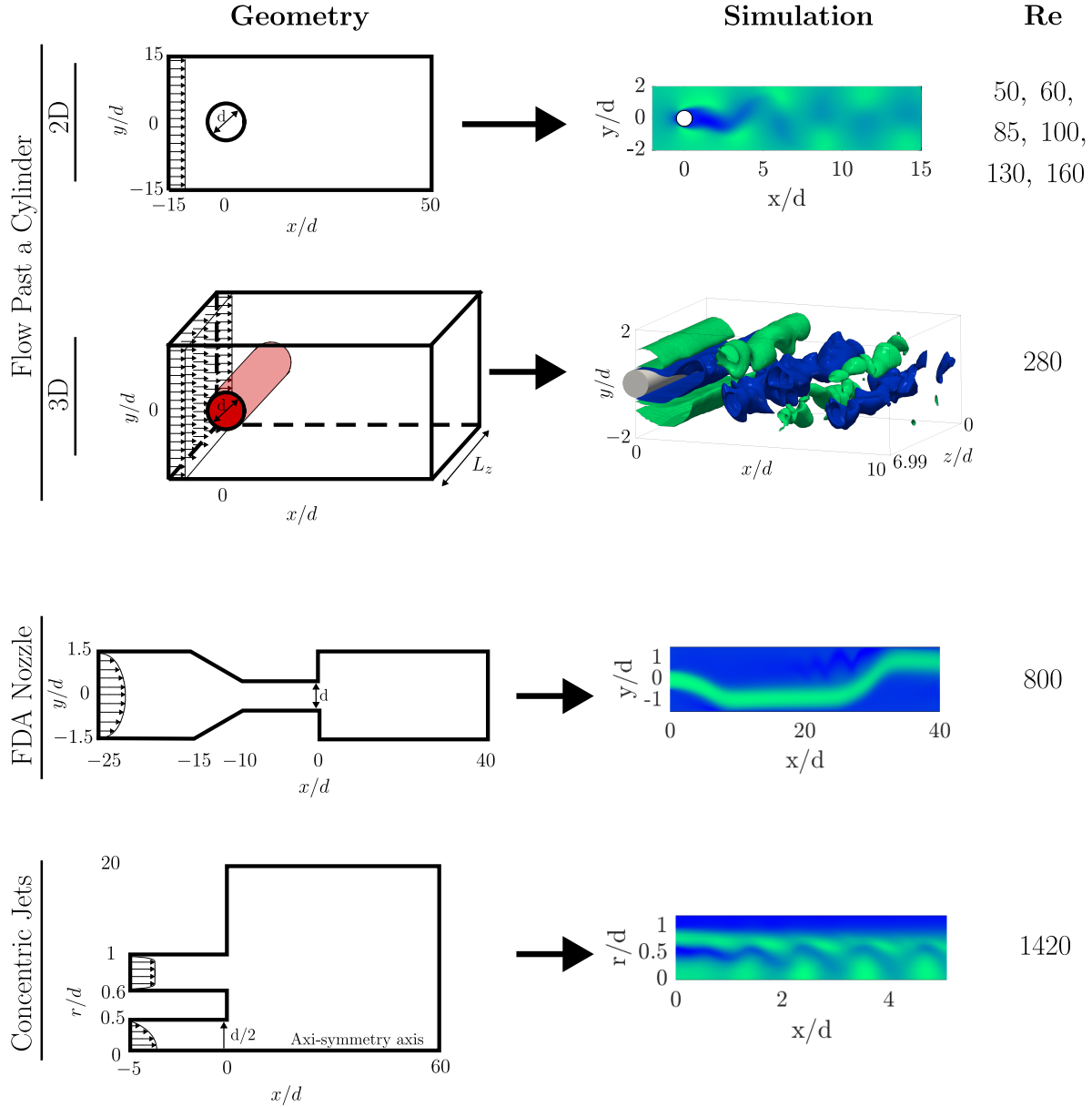


Figure 6.2: Summary of the cases analysed in Corrochano and Le Clainche, 2022. The geometry of the different cases is presented, as well as a representative snapshot and the Re conducted in the numerical simulation.

→ Two-dimensional flow past a circular cylinder

The flow past a circular cylinder is a benchmark problem in fluid dynamics. In this problem, an incoming free stream with velocity V , flows over a cylinder with diameter D , creating a wake behind the cylinder. Six different Reynolds numbers, defined as $Re = Vd/\nu$, are evaluated in this configuration ($Re = 50, 60, 85, 100, 130$ and 160). At very low Reynolds numbers, the flow remains steady. The first instability arises

at $Re \sim 46$, where a Hopf bifurcation triggers the transition from steady to unsteady two-dimensional flow, resulting in the formation of a von Kármán vortex street. The numerical simulations are carried out using Nek5000 for the present work (Fisher et al., 2008). More details of the simulations can be found in the paper below (Corrochano and Le Clainche, 2022).

Firstly, HODMD is applied to all the databases. From each database, the HODMD modes are obtained, along with their frequencies. The frequencies calculated are in good agreement with the literature (Barkley and Henderson, 1996). The structural sensitivity regions identified by the non-linear structural sensitivity are in good agreement with the literature, where this region was identified using the linear theory. Small differences can be found, since one application is solved in the non-linear solution of the Navier Stokes equations, while the other is applied to the linearized equations. Nevertheless, the high intensity regions of structural sensitivity is the same in both cases. The shape of the structural sensitivity is in contrast to the linear theory (Giannetti and Luchini, 2007). This can be explained as the present results are calculated over the non-linear solution, where the recirculation bubble decreases when the Re increases. In experimental results, the regions of high sensitivity shrink when the Re increases (Strykowski and Sreenivasan, 1990), in agreement with our results. The difference between the different methodologies (classical linear theory, the present work and the experiments) do not contradict each other; rather, this novel approach can be regarded as the intermediate step between linear theory and experimental analysis.

→ Three-dimensional flow past a circular cylinder

As the Reynolds number Re increases, additional flow instabilities emerge, initiating the transition to three-dimensional flow for specific values of the spanwise length $L_z = 2\pi/\beta$, where β represents the spanwise wavenumber. In the present study, the Re studied is 280 and two different spanwise wavelengths are examined $L_z = 1.66$ and $L_z = 6.99$. More details about the simulations carried out and the databases extracted can be found in Vega and Le Clainche, 2020, where the databases are published.

HODMD is applied again to both databases and the main frequencies and DMD modes are obtained from the analysis. The frequency of the leading mode is in agreement with the literature (Blackburn et al., 2005), as well as the shape of the region with the highest sensitivity (Giannetti et al., 2010). The differences between the linear theory and the present algorithm are small as in the linear theory, the authors calculate the structural sensitivity over the non-linear solution of the Navier Stokes equations, as they consider the mean flow.

Moreover, this result is in accordance with the trend observed in the structural sensitivity previously calculated for the two-dimensional cylinder wake at high Reynolds numbers, where the highest sensitivity is concentrated in a small region downstream and near the cylinder. The results are obtained in numerical simulations temporally converged, but also using information from the transient regime, testing the limits of applicability of this methodology.

Lastly, in the classical theory, the three-dimensional structural sensitivity is calculated

with a Fourier expansion in the spanwise direction over a 2-D plane. However, with our method, the 3-D structural sensitivity is directly calculated with the three-dimensional DMD mode.

→ **Flow through a planar FDA nozzle**

The flow through a planar FDA nozzle is enclosed in ‘The critical path initiative’ program (“FDA’s Critical Path Initiative.” [n.d.](#)), created by the United States FDA to develop standardised medical devices and methodologies. The FDA nozzle, the case under study, is a benchmark nozzle that resembles an idealized medical device with the main characteristic geometries. This nozzle has been studied with experiments (Delorme et al., [2013](#)) and numerically (Malinauskas et al., [2017](#); Corrochano et al., [2021](#)).

At low Re the flow is steady and symmetric. As seen in Corrochano et al., [2021](#), increasing the Re , the symmetry disappears, but it is still steady. Increasing the Re even more, at $Re \sim 800$, the flow becomes unsteady. At this Re , the analysis is performed. The numerical simulations are carried out using Nek5000 (Fisher et al., [2008](#)) and the databases published in Corrochano et al., [2021](#). More details of the numerical simulations can be found in this later article.

The linear structural sensitivity is calculated using STABFEM, [2020](#) and the results compared with the present studied. The solution of the linear stability analysis is in good agreement with the literature (Jotkar and Govindarajan, [2019](#); Corrochano et al., [2021](#)), as the critical Re , which differentiates the stable and unstable solutions, is accurately calculated. After that, HODMD is applied on the numerical simulation from Corrochano et al., [2021](#), obtaining the main frequency and DMD mode. The frequency calculated in both analysis (linear and present work) is the same.

The structural sensitivity calculated using both strategies is similar, with the small discrepancies observed arising from the fact that, in our algorithm, the sensitivity is solved within the non-linear solution of the Navier-Stokes equations, whereas in the linear stability analysis, it is computed using the linearised equations.

→ **Flow in two concentric jets**

Double concentric jets are configurations with multiple applications in the industry. The main application of this geometry is noise reduction, as in turbofan engines, and combustion, as in combustion chambers, among others.

In this case, the configuration and the simulation set up can be found in Martín et al., [2021](#). The Re studied in this work is slightly higher and the simulations have been validated with the work made by Canton et al., [2017](#). The linear structural sensitivity analysis has been carried out using STABFEM, [2020](#); and an exhaustive linear stability analysis has been published in Corrochano et al., [2023a](#), however, this is out of the scope of this paper.

Again, the frequency calculated using the linear theory and HODMD is the same and the structural sensitivity is similar using both approaches. With the present algorithm, the results suggest that two strategies can be a mechanism of flow control:

- Changing the distance between jets
- Changing the velocity of the external jet.

Both strategies were analysed calculating the critical Re , which differentiates the stable and unstable solutions, as function of the distance between jets and also as function of the velocity of the external jet. As expected, changing whether the distance or the velocity affects to a large extent the value of the critical Re .

Passive flow control for laminar flames with data driven approaches

The present Section combines the expertise gained in the papers published in Chapters 3 and 4 for the analysis of combustion databases with HODMD, with the knowledge of flow control using HODMD. The applicability of the method for flow control in combustion databases is explored. This work has been presented in the *9th European Congress on Computational Methods in Applied Sciences and Engineering*, in June 2024. From this talk, an extended version was published in a conference proceedings in November 2024, available in Sec. 5.2.

Due to the potential arising from the algorithm developed in Sec. 2.5 and explored in laminar non-reactive flows in the published paper in Sec. 5.1, the method is applied to different complex flows databases. Apart from the application in combustion, explained in detail in the following sections, we are working on non-Newtonian fluids. Non-Newtonian fluids are the ones whose viscosity changes depending on the applied shear stress or strain rate.

An ongoing project, we are collaborating with Professor Outi Tammisola, from the KTH, Stockholm, Sweden. In this project, we have been able to control an ElastoViscoPlastic (EVP) flow past a circular cylinder. An EVP fluid is a type of complex fluid that exhibits a combination of elastic, viscous, and plastic behaviour under different stress or deformation conditions. In preliminary results, we have identified the regions with highest structural sensitivity. Placing two small cylinders of diameter $0.1D$, being D the diameter of the cylinder, in the regions with highest sensitivity, we have been able to control the wake. The results show how an unsteady flow, becomes steady after this structural change. A summary of the main results of this ongoing project can be found in Fig. 6.3.

In the published article 5.2, the structural sensitivity is identified in reactive flows for the first time to the authors' knowledge. While a few passive flow control strategies have been applied to combustion systems, there is not a guideline of the location where to apply the passive flow control. This passive flow control strategies include a lobed fuel injector (Karagozian, 1999) and the addition of atomizers (Beér and Chigier, 1983), among others.

The novel algorithm proposed in Sec. 2.5 is applied to an axisymmetric, time varying, non-premixed laminar co-flow flame, with a sinusoidal perturbation in the velocity profile of the injection of the fuel of $10Hz$, the same configuration as the previous Chapters 3 and 4. The kinetic mechanism used in the simulations is **GRI-Mech 3.0**, which takes into account 53 chemical species and 325 reactions. This kinetic mechanisms takes also into account the formation of NOx , some of the most harmful pollutants.

As the flow is forced with a sinusoidal perturbation in the boundary conditions, the objective of this work is not the same as in the previous Sec. 6.2.2. Rather than controlling the wake

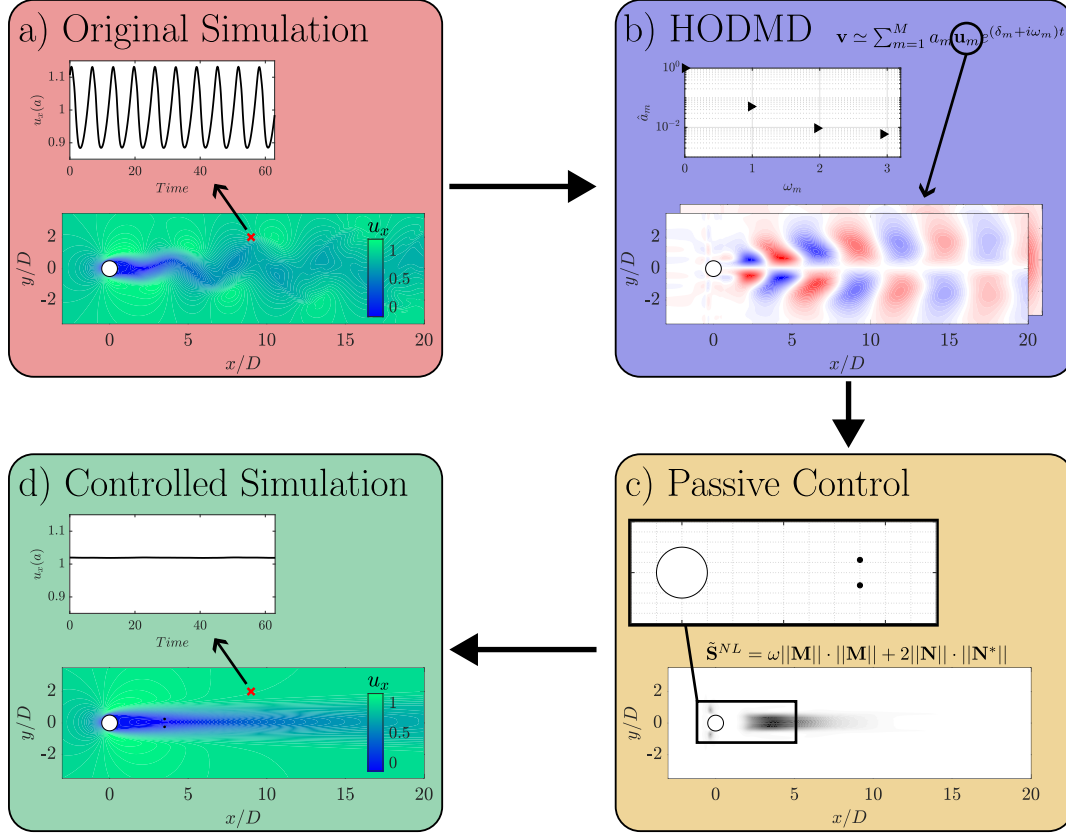


Figure 6.3: Summary of the process of passive flow control on a bidimensional EVP flow past a cylinder. a) Original unsteady simulation, represented with a representative snapshot and the plot of the streamwise velocity at a point $a = (x_a/D, y_a/D) = (9, 2)$ over time. b) HODMD analysis on the field, extracting the main frequencies and the modes associated to the dynamics of the simulation. The mode is normalised varying between -1 and 1 . c) Calculation of the non-linear structural sensitivity $\tilde{\mathbf{S}}^{NL}$ and zoom of the placement of little cylinders where $\tilde{\mathbf{S}}^{NL}$ is maximum. d) Controlled steady simulation, represented with a representative snapshot and the plot of the streamwise velocity at the same point a over time.

and attempting to make the unsteady solution steady, in this case the objective is to enhance the efficiency in the combustion and reduce the NOx emissions. The combustion efficiency is calculated as

$$\eta = \frac{1}{T} \int_0^T \left(1 - \frac{\dot{m}_{CO} Q_{CO} + \dot{m}_{H_2} Q_{H_2} + \dot{m}_{HC} Q_{HC}}{cL} \right) dt, \quad (6.4)$$

where $HC = CH_4, C_2H_2, C_2H_4, C_2H_6$ are the unburned hydrocarbons. In the previous equation, $\dot{m}_i = Y_i \dot{m}_{out}$ and Q_i are the outlet mass flux and the lower heating value of species i , respectively, and $c = Y_{CH_4} \dot{m}_{in}$ is the fuel mass flux at the inlet and L is the lower heating value of methane. This equation indicates the extent to which the fuel has been fully converted to the main products, which are H_2O and CO_2 . Every species, different from the aforementioned two, present in the outlet will decrease the efficiency.

HODMD is applied to a database extracted from the simulation following the recommendations by Corrochano et al., 2023d, and explained in Chapter 3. The database is preprocessed as in Sec. 2.1, HODMD is applied and the non-linear structural sensitivity is calculated. Three potential regions are identified and evaluated, promising results for flow control approaches, which will be explained below.

HODMD is applied to the aforementioned database. From this analysis, the obtained spectrum is periodic with a main frequency of $\omega \sim 63$. The frequency is equal to the induced in the boundary condition of the fuel ($f = \omega/2\pi = 10Hz$), as expected. The mode shown in the paper is the one related to the velocity, however, 19 different variables are analysed at the same time, and the modes related to other species can be found in Fig. 5 of Corrochano et al., 2023d (Sec. 6.1.2) and in Fig. 8 of Corrochano et al., 2024c (Chapter 4). The mode related to the velocity, the one where the structural sensitivity will be applied, present coherent structures along the centre line.

From this modes and following Eqs. (2.28), (2.29) and (2.31), the structural sensitivity can be calculated as in Eq. (2.44). For this case, the two terms are evaluated separately and three potential regions are identified. In these three regions, two different actuations could be applied. A punctual force could be applied in the highest sensitivity regions, as it is generally done in the literature (Giannetti and Luchini, 2007). Other strategy is the placement of small cylinders on the regions where the sensitivity is maximum. This approach has been explored both in experiments (Strykowski and Sreenivasan, 1990) and numerical simulations (Pralits et al., 2010). This approach is the one explored in the present work and three different simulations are carried out evaluating the potential regions placing a cylinder of diameter $0.0038m$, which is the same length as the thickness of the separating plate between jets.

In other works where the characteristic length is unique, as in Fig. 6.3, it is usual that the diameter of the small cylinder is equal to $1/10$ of the characteristic length. In this case, there are several lengths that could be used as "characteristic". For example, to name a few, the diameter of both jets, their sum, the length of the flame and the length of the structures in the HODMD modes. Therefore, a study varying the diameter of the cylinder should be carried out to analyse the influence of the diameter on the solution. However, this is out of the scope of the present study.

Table 6.3: Efficiency of the flame and the production of NOx for the initial case and the three cases studied applying the flow control algorithm. For the three cases studied, in green if the efficiency increases or the NOx emissions decrease and in red, the opposite.

	η [%]	NO [ppm]	NO_2 [ppm]	NO_x [ppm]
Initial	93.61	51.77	12.53	64.30
Case 1	92.33	47.82 (-8%)	11.85 (-5%)	59.60 (-7%)
Case 2	96.70	45.40 (-12%)	9.51 (-24%)	54.91 (-15%)
Case 3	91.74	44.26 (-15%)	17.60 (+40%)	61.92 (-4%)

The efficiency, calculated as Eq. (6.4), and the NOx emissions are summarised in Tab. 6.3.

Apart from the NOx emissions, the two main components of NOx are presented, i.e., NO and NO_2 . As seen, the best case overall is the second one, the one near the injection (Fig. 3 in the paper), with a decrease of 15% of the NOx emissions and an increase of three points in the combustion efficiency.

HODMD is applied to the three modified cases and the greatest changes in the mode appear on the second case, being a possible explanation of the highest changes in the quality indicators.

The method shows potential to be used in flow control approaches. However, there is still further improvement and the study can be extended to explore the influence of the diameter of the cylinders and a possible combination between cases 1, 2 and 3.

6.3 Conclusions

The main focus of this Doctoral Thesis has been the development of reduced order models (ROMs) for combustion systems. These ROMs include the adaptation of existing algorithms and the implementation of new ones with different purposes. Despite the fact that the algorithms explained during the thesis have been applied to combustion databases, they are fully data-driven and generalizable. Therefore, the algorithms can be used in other knowledge fields, as aerodynamics, fluid dynamics, aerobiology, molecular sciences, among others.

A diagram describing the advancements in the development of ROMs for combustion databases is shown in Fig. 6.4. From this diagram, an analysis of reacting flows can be followed, from the preprocessing of the database (Sec. 2.1), to then the analysis with HODMD (Chapter 3), up to the different purposes of the analysis, which in the present Doctoral Thesis are clustering (Chapter 4) and flow control (Chapter 5).

The database presented in Fig. 6.4 is a fourth-dimensional tensor, however we have plotted the variables in a row for visualisation purposes. The four dimensions are: variables, two spatial dimensions and the time, with K snapshots. Once centered and scaled the database does not change the dimensions of the tensor, however all the variables present the same order of magnitude. Hence, it is possible to analyse the database. HODMD is applied to the fourth-dimensional tensor, and a reduced set of M DMD modes ($M < K$) can reproduce the full database at a reduced computational cost.

The multiscale and multidimensional nature arises in combustion systems, consequently caution must be exercised before analysing reactive flows databases. In the present Doctoral Thesis, centering and scaling has been used as a preprocessing technique to uniform the order of magnitude of all the variables. Different scaling methods have been assessed in Chapters 3 and 4. As explained in Sec. 2.1, every scaling method emphasise the importance of some variables against others. This statement is verified when applying HODMD (Chapter 3) and h-HODMD (Chapter 4). The use of one scaling method over another will depend on the problem under study.

The different effect of applying HODMD to a combustion database can be read in Chapter 3. In the two papers exposed, some conclusions can be extracted from each step of the algorithm (Sec. 2.2). At the first step, the reduction of the number of the SVD modes retained in

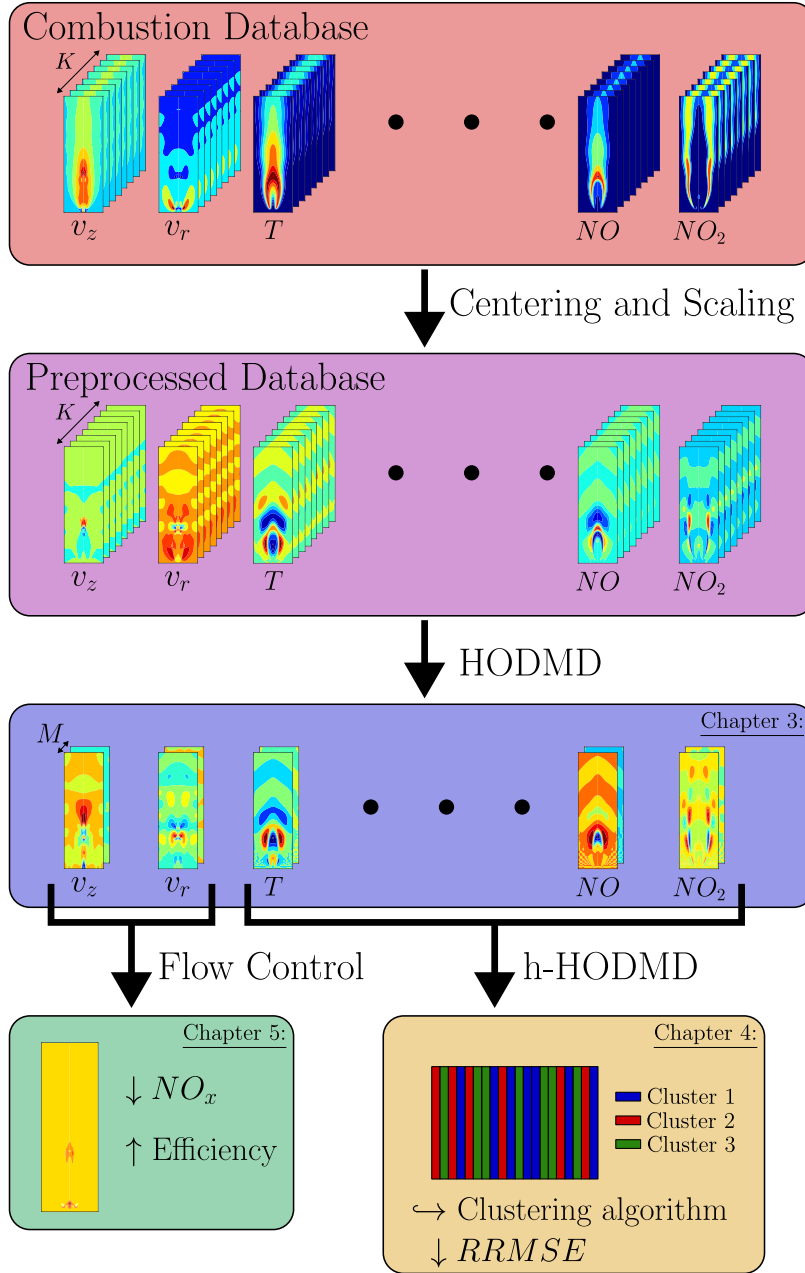


Figure 6.4: Diagram summarising the analysis of a combustion database, as proposed in the present Doctoral Thesis.

the first step with HOSVD provides an accurate approximation of the database, with low $RRMSE$, at a high compression of the database. This dimensionality reduction is drastically higher when all the different dimensions present on the tensor are reduced. The coupling of HODMD with variable selection via PCA indicates that a subset of the original variables is sufficient to capture the main dynamics of the entire system. This approach impacts feature selection and significantly reduces the computational cost associated with very large datasets.

HODMD is well-suited for constructing physics-aware ROMs based on the system's primary dynamics.

When applying HODMD to a combustion database coming from a transient regime of the simulation, different main frequencies may appear using different scaling methods, as seen in Sec. 6.1.2. These differences, due to the different variables each scaling method enhance, arise the need of an algorithm able to cluster the variables, decoupling the different dynamics in combustion systems. This algorithm, named *hierarchical-HODMD* (h-HODMD), performs HODMD iteratively, identifying and isolating the best reconstructed variables, forming a cluster and associating it with the main frequencies and DMD modes associated. The performance of h-HODMD and its ability to form clusters consistent from a kinetic perspective, are tested in Chapter 4.

A database generated during the transient state of the simulation can be used for developing ROMs that form clusters based on the main dynamics. This novel algorithm, h-HODMD (Chapter 4), outperforms HODMD in terms of reconstruction error for this type of databases and iteratively generates groups of variables which share the same main dynamics. The effect of the scaling methods on this algorithm has also been analysed. The main conclusion is that auto scaling, the one which uses the standard deviation as the scaling factor, is not a suitable scaling method for h-HODMD, due to the fact that it gives the same importance to all the variables. Therefore, every cluster has the same dynamics and the error is not improved. The method has successfully separated the variables with different dynamics from a kinetic perspective. h-HODMD offers valuable potential for developing novel, physics-informed ROMs. This method enables the exploration of innovative strategies for managing increasingly large databases while reducing the associated computational costs.

Another set of important questions is related to the increasing interest of reducing the emissions of pollutants, as it is one of the main objectives of the European Union for 2050 (European Commission and Directorate-General for Climate Action, 2019; European Commission, 2021). The improvement of the efficiency in combustion systems helps the objectives of the EU to become climate-neutral by that year. Following the growing concern, there is a need of developing new ROMs which can evaluate this problem. A new algorithm has been developed in this Doctoral Thesis based on HODMD for flow control. Being able to control the fluid flow can be one of the strategies to achieve the reduction of pollutants. This novel methodology, explained in Sec. 2.5, was developed at first for laminar non-reactive fluid flows. In this context, the objective is to reduce the drag or to achieve the flow steadiness, as seen in Sec. 6.2.2. The algorithm was tested in the laminar flame under study obtaining promising results in Sec. 6.2.2.

When analysing a fully converged database, as the one in Chapter 5, a novel methodology for flow control can be applied. From the modes obtained as in Chapter 3, the non-linear structural sensitivity can be calculated. The velocity modes are needed for the computation; nevertheless, the analysis is conducted on the entire database, as the method is based on the assumption that the dynamics are coupled. The non-linear structural sensitivity provides the parts of the flow field more prone to changes in the main dynamics; and introducing a small cylinder on the part with highest intensity, it is possible to enhance the combustion efficiency and reduce the NO_x . The presented findings in the Doctoral Thesis demonstrate significant

potential for the algorithm's application in flow control and the design of components aimed at pollutant reduction.

6.3.1 Future work

The present Doctoral Thesis present a novel methodology for analysing combustion databases. Therefore, future works could explore several directions. The most important and applicable to all the algorithms is the extension of the current models to a broader range of applications. This can include turbulent combustion or other flow configurations.

These algorithms are fully data-driven and generalizable, so future works may extend their applicability to other frameworks. HODMD has been already applied to many different scenarios before to combustion, as seen in Section 2.2. The algorithm proposed for feature selection and variable clustering, h-HODMD, can be extended to other contexts where there are multiple variables, as molecular sciences or for the study of the concentration of pollutants in cities (air quality). The algorithm for flow control has been proved to be suitable in laminar non-reactive flows, non-Newtonian fluids and laminar flames. However, this algorithm may be suitable for turbulence flows for drag reduction or noise suppression, among others.

The flow control algorithm is a preliminary work with huge potential, as the first study already reduces the NO_x emissions 15%. However, different studies may expand the present work:

- Try different configurations and flames
- Study the influence of the diameter of the cylinder. The cylinder should not disrupt the injection of the fuel or the oxidizer and many different length scales are present in the flow studied.
- When more than one region appears in the sensitivity map (as in the case studied), prove if a combination of points improves even more the combustion efficiency and reduces the emissions.

These advancements leave open a new research field centered on the control of reactive flows with the objective to improve the energy efficiency of combustion systems and the reduction of pollutants emitted to the atmosphere.

Referencias

- Abou-Taouk, A., Sadasivuni, S., Lörstad, D., Bulat, G., & Eriksson, L.-E. (2015). Cfd analysis and application of dynamic mode decomposition for resonant-mode identification and damping in an sgt-100 dle combustion system. *Proceedings of the 7th European Combustion Meeting*.
- Amor, C., Corrochano, A., Rota, G. F., Rosti, M. E., & Le Clainche, S. (2024). Coherent structures in elastic turbulent planar jets. *Journal of Physics: Conference Series*, 2753(1), 012020. <https://doi.org/10.1088/1742-6596/2753/1/012020>
- Amor, C., Pérez, J. M., Schlatter, P., Vinuesa, R., & Le Clainche, S. (2020). Modeling the Turbulent Wake Behind a Wall-Mounted Square Cylinder. *Logic Journal of the IGPL*, 30(2), 263–276. <https://doi.org/10.1093/jigpal/jzaa060>
- Aversano, G., Bellemans, A., Li, Z., Coussement, A., Gicquel, O., & Parente, A. (2019). Application of reduced-order models based on pca & kriging for the development of digital twins of reacting flow applications. *Computers & Chemical Engineering*, 121, 422–441. <https://doi.org/10.1016/j.compchemeng.2018.09.022>
- Badra, J. A., Khaled, F., Tang, M., Pei, Y., Kodavasal, J., Pal, P., Owoyele, O., Fuetterer, C., Mattia, B., & Aamir, F. (2020). Engine Combustion System Optimization Using Computational Fluid Dynamics and Machine Learning: A Methodological Approach [022306]. *Journal of Energy Resources Technology*, 143(2). <https://doi.org/10.1115/1.4047978>
- Barkley, D., & Henderson, R. (1996). Three-dimensional floquet stability analysis of the wake of a circular cylinder. *J. Fluid Mech.*, 322, 215–241.
- Beér, J., & Chigier, N. (1983). *Combustion aerodynamics*. <https://doi.org/10.1017/S0022112072210990>
- Bellemans, A., Aversano, G., Coussement, A., & Parente, A. (2018). Feature extraction and reduced-order modelling of nitrogen plasma models using principal component analysis. *Computers & chemical engineering*, 115, 504–514.
- Bellemans, A., Magin, T., Coussement, A., & Parente, A. (2017). Reduced-order kinetic plasma models using principal component analysis: Model formulation and manifold sensitivity. *Physical Review Fluids*, 2(7), 073201.
- Beltrán, V., Le Clainche, S., & Vega, J. (2022). An adaptive data-driven reduced order model based on higher order dynamic mode decomposition. *Journal of Scientific Computing*, 92. <https://doi.org/10.1007/s10915-022-01855-2>
- Bishop, C. M. (2006). Pattern recognition. *Machine learning*, 128(9).

- Blackburn, H. M., Marques, F., & Lopez, J. M. (2005). Symmetry breaking of two-dimensional time-periodic wakes. *J. Fluid Mech.*, 522, 395–411.
- Bong, C., Ali, M. H., kyun Im, S., Do, H., & Bak, M. S. (2024). Machine learning-based prediction of swirl combustor operation from flame imaging. *Engineering Applications of Artificial Intelligence*, 138, 109374. <https://doi.org/https://doi.org/10.1016/j.engappai.2024.109374>
- Bonuso, S., Gloria, P. D., Marseglia, G., Martínez, R. A. O., Mehdi, G., Shah, Z. A., Ficarella, A., & Giorgi, M. G. D. (2024). Investigation into the effect of H_2 -enriched conditions on the structure and stability of flames in a low-swirl combustor derived from aero-engine design. *Aerospace*, 11(1), 43. <https://doi.org/10.3390/aerospace11010043>
- Brouzet, D., Haghiri, A., Talei, M., Brear, M. J., Schmidt, O. T., Rigas, G., & Colonius, T. (2020). Role of coherent structures in turbulent premixed flame acoustics. *AIAA Journal*, 58(6), 2635–2642. <https://doi.org/10.2514/1.J058480>
- Cant, S. (2002). 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.
- Canton, J., Auteri, F., & Carini, M. (2017). Linear global stability of two incompressible coaxial jets. *J. Fluid Mech.*, 824, 886–911.
- Casel, M., Oberleithner, K., Zhang, F., Zirwes, T., Bockhorn, H., Trimis, D., & Kaiser, T. L. (2022). Resolvent-based modelling of coherent structures in a turbulent jet flame using a passive flame approach. *Combustion and Flame*, 236, 111695. <https://doi.org/https://doi.org/10.1016/j.combustflame.2021.111695>
- Chatterjee, S., Mukhopadhyay, A., & Sen, S. (2013). Stability study of laminar flame using proper orthogonal decomposition and dynamic mode decomposition. *33rd Int'l Summer School and Workshop on Non-Normal and Nonlinear Effects In Aero-and Thermoacoustics*, 13.
- Chen, F., Wang, B., Ni, M., Gong, Z., & Jiao, K. (2024). Online energy management strategy for ammonia-hydrogen hybrid electric vehicles harnessing deep reinforcement learning. *Energy*, 301, 131562. <https://doi.org/https://doi.org/10.1016/j.energy.2024.131562>
- Chen, J. H., Choudhary, A., de Supinski, B., DeVries, M., Hawkes, E. R., Klasky, S., Liao, W. K., Ma, K. L., Mellor-Crummey, J., Podhorszki, N., Sankaran, R., Shende, S., & Yoo, C. S. (2009). Terascale direct numerical simulations of turbulent combustion using s3d. *Computational Science & Discovery*, 2(1), 015001. <https://doi.org/10.1088/1749-4699/2/1/015001>
- Chen, J., Du, J., Liu, Y., Liu, L., Li, A., Jiang, J., & Sun, P. (2024). Numerical study of combustion flow field characteristics of industrial gas turbine under different fuel blending conditions. *Applied Thermal Engineering*, 251, 123573. <https://doi.org/https://doi.org/10.1016/j.applthermaleng.2024.123573>
- Chen, K. K., Tu, J. H., & Rowley, C. W. (2012). Variants of dynamic mode decomposition: Boundary condition, koopman, and fourier analyses. *J. Nonlin. Scien.*, 22(6), 887–915.
- Chen, X., Zhu, C., Ma, C., & Li, Q. (2024). Vortex-induced vibration of twin-box girder and its aerodynamic characteristics study based on high-order dynamic decomposition. *Journal of Wind Engineering and Industrial Aerodynamics*, 249, 105719. <https://doi.org/https://doi.org/10.1016/j.jweia.2024.105719>

- Cheng, C., Ding, J., & Zhang, Y. (2020). A koopman operator approach for machinery health monitoring and prediction with noisy and low-dimensional industrial time series. *Neurocomputing*, 406, 204–214. <https://doi.org/10.1016/j.neucom.2020.04.005>
- Choi, J., Rajasegar, R., Liu, Q., Lee, T., & Yoo, J. (2022). Modal decomposition analysis of combustion instability due to external perturbation in a mesoscale burner array. *Journal of Engineering for Gas Turbines and Power*, 144(5), 051011. <https://doi.org/10.1115/1.4053445>
- Corrochano, A., Sierra-Ausín, J., Martin, J., Fabre, D., & Le Clainche, S. (2023a). Mode selection in concentric jets: The steady–steady 1 : 2 resonant mode interaction with o(2) symmetry. *Journal of Fluid Mechanics*, 971, A30. <https://doi.org/10.1017/jfm.2023.445>
- Corrochano, A., Abadía-Heredia, R., Crialesi-Esposito, M., Brandt, L., & Le Clainche, S. (2023b). Unsteady flow feature extraction of flapping events [Conference]. *18th European Turbulence Conference*.
- Corrochano, A., Canals, J. L., Saavedra, L., & Le Clainche, S. (2024a). Flow control with data-driven approaches [Conference]. *9th European Congress on Computational Methods in Applied Sciences and Engineering*.
- Corrochano, A., Canals, J. L., Saavedra, L., & Le Clainche, S. (2024b). Flow control with data-driven approaches. *ECCOMAS Congress 2024 - 9th European Congress on Computational Methods in Applied Sciences and Engineering, Data-Enhanced Reduced Order Modeling*. <https://doi.org/10.23967/eccomas.2024.085>
- Corrochano, A., D'Alessio, G., Parente, A., & Le Clainche, S. (2023c). Higher order dynamic mode decomposition to model reacting flows. In T. H. Karakoc, N. Yilmaz, A. Dalkiran, & A. H. Ercan (Eds.), *New achievements in unmanned systems* (pp. 205–210). Springer International Publishing.
- Corrochano, A., D'Alessio, G., Parente, A., & Le Clainche, S. (2023d). Higher order dynamic mode decomposition to model reacting flows. *International Journal of Mechanical Sciences*, 249, 108219. <https://doi.org/10.1016/j.ijmecsci.2023.108219>
- Corrochano, A., D'Alessio, G., Parente, A., & Le Clainche, S. (2024c). Hierarchical higher-order dynamic mode decomposition for clustering and feature selection. *Computers & Mathematics with Applications*, 117, 25–50. <https://doi.org/10.1016/j.camwa.2024.01.003>
- Corrochano, A., Freitas, R. S. M., Parente, A., & Le Clainche, S. (2023e). A predictive physics-aware hybrid reduced order model for reacting flows.
- Corrochano, A., Freitas, R. S. M., Parente, A., & Le Clainche, S. (2023f). A predictive physics-aware machine learning model for reacting flows. In T. H. Karakoc, S. Le Clainche, X. Chen, A. Dalkiran, & A. H. Ercan (Eds.), *New technologies and developments in unmanned systems* (pp. 75–79). Springer International Publishing.
- Corrochano, A., & Le Clainche, S. (2022). Structural sensitivity in non-linear flows using direct solutions. *Computers & Mathematics with Applications*, 128, 69–78. <https://doi.org/10.1016/j.camwa.2022.10.006>
- Corrochano, A., Neves, A. F., Khanal, B., Le Clainche, S., & Lawson, N. J. (2022). Des of a slingsby firefly aircraft: Unsteady flow feature extraction using pod and hodmd. *Journal of Aerospace Engineering*, 35(5), 04022063. [https://doi.org/10.1061/\(ASCE\)AS.1943-5525.0001457](https://doi.org/10.1061/(ASCE)AS.1943-5525.0001457)

- Corrochano, A., Xavier, D., Schlatter, P., Vinuesa, R., & Le Clainche, S. (2021). Flow structures on a planar food and drug administration (fda) nozzle at low and intermediate reynolds number. *Fluids*, 6(1). <https://doi.org/10.3390/fluids6010004>
- Cuoci, A., Frassoldati, A., Faravelli, T., & Ranzi, E. (2013). Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method. *Energy & Fuels*, 27(12), 7730–7753.
- D’Alessio, G., Cuoci, A., & Parente, A. (2021). 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.
- D’Alessio, G., Parente, A., Stagni, A., & Cuoci, A. (2020). Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. *Combustion and Flame*, 211, 68–82.
- D’Alessio, G., Sundaresan, S., & Mueller, M. E. (2022). Automated and efficient local adaptive regression for principal component-based reduced-order modeling of turbulent reacting flows. *Proceedings of the Combustion Institute*.
- de Bruyn, S., Jongsma, C., Kampman, B., Görlach, B., & Thie, J.-E. (2020). *Energy-intensive industries: Challenges and opportunities in energy transition*. European Parliament.
- Delorme, Y. T., Anupindi, K., & Frankel, S. H. (2013). Large eddy simulation of fdas idealized medical device. *Cardiovasc. Eng. Technol.*, 4, 392–407.
- Díaz, P., Corrochano, A., López-Martín, M., & Le Clainche, S. (2023). Machine learning to reconstruct aeronautical databases with deep neural networks. In T. H. Karakoc, S. Le Clainche, X. Chen, A. Dalkiran, & A. H. Ercan (Eds.), *New technologies and developments in unmanned systems* (pp. 125–129). Springer International Publishing.
- Díaz-Morales, P., Corrochano, A., López-Martín, M., & Le Clainche, S. (2024). Deep learning combined with singular value decomposition to reconstruct databases in fluid dynamics. *Expert Systems with Applications*, 238, 121924. <https://doi.org/10.1016/j.eswa.2023.121924>
- Domínguez-González, A., Martínez-Ruiz, D., & Sánchez-Sanz, M. (2023). Stable circular and double-cell lean hydrogen-air premixed flames in quasi two-dimensional channels. *Proceedings of the Combustion Institute*, 39(2), 1731–1741. <https://doi.org/https://doi.org/10.1016/j.proci.2022.07.050>
- El Gammal, A., Weikmans, R., & van Ypersele, J.-P. (2022, November 25). *La cop27 que du bla-bla ?* [Conference organized by Professor Alessandro Parente, Vice-Dean for Transition, along with his student assistant, Jeanne Longlune, at the École Polytechnique de Bruxelles]. https://www.youtube.com/watch?v=QES1c6I-nTw&ab_channel=ULBTV
- European Commission. (2021). European green deal: The commission proposes transformation of eu economy and society to meet climate ambitions.
- European Commission & Directorate-General for Climate Action. (2019). *Going climate-neutral by 2050 – a strategic long-term vision for a prosperous, modern, competitive and climate-neutral eu economy*. Publications Office. <https://doi.org/doi/10.2834/02074>
- FDA’s critical path initiative. [[Online; accessed 30 November 2020]]. (n.d.).
- Fisher, P. F., Lottes, J. W., & Kerkemeier, S. G. (2008). Nek5000: Open source spectral element cfd solver. Available online: www.nek5000.mcs.anl.gov.
- Gaba, A., & Iordache, S. (2011, September). Reduction of air pollution by combustion processes. <https://doi.org/10.5772/16959>

- Gabana, P., Giménez, B., Martín Herreros, J., & Tsolakis, A. (2025). Analysis of the combustion speed in a spark ignition engine fuelled with hydrogen and gasoline blends at different air fuel ratios. *Fuel*, *381*, 133563. <https://doi.org/https://doi.org/10.1016/j.fuel.2024.133563>
- Ghani, A., Poinso, T., L.Y.M., G., & Staffelbach, G. (2015). Les of longitudinal and transverse self-excited combustion instabilities in a bluff-body stabilized turbulent premixed flame. *Combustion and Flame*, *162*, 4075–4083. <https://doi.org/10.1016/j.combustflame.2015.08.024>
- Gianneti, F., Camarri, S., & Luchini, P. (2010). Structural sensitivity of the secondary instability in the wake of a circular cylinder. *J. Fluid Mech.*, *651*, 319–337.
- Giannetti, F., & Luchini, P. (2007). Structural sensitivity of the first instability of the cylinder wake. *Journal of Fluid Mechanics*, *581*, 167–197.
- Grenga, T., & Mueller, M. (2020). Dynamic mode decomposition: A tool to extract structures hidden in massive datasets. In *Data analysis for direct numerical simulations of turbulent combustion* (pp. 157–176). Springer.
- Grenga, T., MacArt, J. F., & Mueller, M. E. (2018). Dynamic mode decomposition of a direct numerical simulation of a turbulent premixed planar jet flame: Convergence of the modes. *Combustion Theory and Modelling*, *22*(4), 795–811.
- Groun, N., Villalba-Orero, M., Lara-Pezzi, E., Valero, E., Garicano-Mena, J., & Le Clainche, S. (2022). Higher order dynamic mode decomposition: From fluid dynamics to heart disease analysis. *Computers in Biology and Medicine*, *144*, 105384. <https://doi.org/https://doi.org/10.1016/j.combiomed.2022.105384>
- Guastoni, L., Rabault, J., Schlatter, P., Azizpour, H., & Vinuesa, R. (2023). Deep reinforcement learning for turbulent drag reduction in channel flows. *European Physical Journal E*, *46*, 27. <https://doi.org/10.1140/epje/s10189-023-00285-8>
- Gunaratne, G., & Roy, S. (2021). Robust mode analysis. *Mathematics*, *9*, 1057. <https://doi.org/10.3390/math9091057>
- Gutierrez-Castillo, P., Garrido-Martin, M., Bölle, T., García-Ortiz, J. H., Aguilar-Cabello, J., & del Pino, C. (2022). Higher order dynamic mode decomposition of an experimental trailing vortex. *Physics of Fluids*, *34*(10), 107116. <https://doi.org/10.1063/5.0117611>
- Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. *Neural Networks*, *1*(1), 45–58. [https://doi.org/10.1016/S0893-6080\(05\)80065-7](https://doi.org/10.1016/S0893-6080(05)80065-7)
- Hetherington, A., Corrochano, A., Abadía-Heredia, R., Lazpita, E., Muñoz, E., Díaz, P., Maiores, E., Martín, M. L., & Clainche, S. L. (2024a). Modelflows-app: Data-driven post-processing and reduced order modelling tools. *Computer Physics Communications*, *301*, 109217. <https://doi.org/10.1016/j.cpc.2024.109217>
- Hetherington, A., Serfaty, D., Corrochano, A., Soria, J., & Le Clainche, S. (2024b). Data repairing and resolution enhancement using data-driven modal decomposition and deep learning. *Experimental Thermal and Fluid Science*, *157*, 111241. <https://doi.org/10.1016/j.expthermflusci.2024.111241>
- Huang, C., Anderson, W. E., Harvazinski, M. E., & Sankaran, V. (2016). Analysis of self-excited combustion instabilities using decomposition techniques. *AIAA Journal*, *54*(9), 2791–2807. <https://doi.org/10.2514/1.J054557>

- Huang, C., Duraisamy, K., & Merkle, C. (2022). Component-Based Reduced Order Modeling of Large-Scale Complex Systems. *Frontiers in Physics*, 10, Article 900064, 900064. <https://doi.org/10.3389/fphy.2022.900064>
- Huang, C., Duraisamy, K., & Merkle, C. (n.d.). Challenges in reduced order modeling of reacting flows. In *2018 joint propulsion conference*. <https://doi.org/10.2514/6.2018-4675>
- Huang, J., Liu, H., Wang, Q., & Cai, W. (2020). Limited-projection volumetric tomography for time-resolved turbulent combustion diagnostics via deep learning. *Aerospace Science and Technology*, 106, 106123. <https://doi.org/https://doi.org/10.1016/j.ast.2020.106123>
- International Energy Agency (IEA). (2023). Global hydrogen review 2023 [Available online: <https://www.iea.org/reports/global-hydrogen-review-2023>].
- Isaac, B. J., Coussement, A., Gicquel, O., Smith, P. J., & Parente, A. (2014). Reduced-order pca models for chemical reacting flows. *Combustion and flame*, 161(11), 2785–2800.
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd). Springer-Verlag. <https://doi.org/10.1007/b98835>
- Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. i: Artificial data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 21(2), 160–173.
- Jolliffe, I. T. (1973). Discarding variables in a principal component analysis. ii: Real data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 22(1), 21–31.
- Jones, C., & Utyuzhnikov, S. (2024). Application of noise-filtering techniques to data-driven analysis of electric power systems based on higher-order dynamic mode decomposition. *International Journal of Electrical Power & Energy Systems*, 156, 109721. <https://doi.org/https://doi.org/10.1016/j.ijepes.2023.109721>
- Jotkar, M., & Govindarajan, R. (2019). Two-dimensional modal and non-modal instabilities in straight-diverging-straight channel flow. *Phys. Fluids*, 31, 014102.
- Karagozian, A. (1999). Control of mixing and reactive flow processes. *30th Fluid Dynamics Conference*. <https://doi.org/10.2514/6.1999-3572>
- Karagozian, A., Wang, K., Le, A.-T., & Smith, O. (1996). Transverse gas jet injection behind a rearward-facing step. *Journal of Propulsion and Power*, 12(6), 1129–1136. <https://doi.org/10.2514/3.24153>
- Khalifehei, M., Yalcinkaya, Y., Padilla-Montero, I., Gungor, A. G., & Rodríguez, D. (2024). Investigation of the coupling between the dynamics of vortical structures and flame stability in bluff-body premixed combustion using extended spectral proper orthogonal decomposition. *Journal of Engineering for Gas Turbines and Power*, 147(3), 031001. <https://doi.org/10.1115/1.4066362>
- Kou, J., Le Clainche, S., & Zhang, W. (2018). 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.
- Krzanowski, W. J. (1987). 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.
- Kundashkin, A., Palkin, E., Hrebtov, M., et al. (2022). Active flow control in a model combustion chamber using additional gas injection method. *Thermophysics and Aeromechanics*, 29, 785–790. <https://doi.org/10.1134/S08698643220500171>

- Lazpita, E., Martínez-Sánchez, Á., Corrochano, A., Hoyas, S., Le Clainche, S., & Vinuesa, R. (2022). On the generation and destruction mechanisms of arch vortices in urban fluid flows. *Physics of Fluids*, 34(5), 051702. <https://doi.org/10.1063/5.0088305>
- Le Clainche, S., Izbassarov, D., Rosti, M., Brandt, L., & Tammisola, O. (2020a). Coherent structures in the turbulent channel flow of an elastoviscoplastic fluid. *Journal of Fluid Mechanics*, 888.
- Le Clainche, S., Pérez, J. M., & Vega, J. M. (2018). Spatio-temporal flow structures in the three-dimensional wake of a circular cylinder. *Fluid Dyn. Res.*, in press, 50(5), 051406.
- Le Clainche, S., & Ferrer, E. (2018). A reduced order model to predict transient flows around straight bladed vertical axis wind turbines. *Energies*, 11(3), 566.
- Le Clainche, S., Ferrer, E., Gibson, S., Cross, E., Parente, A., & Vinuesa, R. (2023). Improving aircraft performance using machine learning: A review. *Aerospace Science and Technology*, 138, 108354. <https://doi.org/10.1016/j.ast.2023.108354>
- Le Clainche, S., Moreno-Ramos, R., Taylor, P., & Vega, J. M. (2019). New robust method to study flight flutter testing. *Journal of Aircraft*, 56(1), 336–343.
- Le Clainche, S., Rosti, M., & Brandt, L. (2020b). Flow structures and shear-stress predictions in the turbulent channel flow over an anisotropic porous wall. *Journal of Physics: Conference Series*, 1522(1), 012016. <https://doi.org/10.1088/1742-6596/1522/1/012016>
- Le Clainche, S., & Vega, J. M. (2017a). Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2), 882–925.
- Le Clainche, S., & Vega, J. M. (2017b). Higher order dynamic mode decomposition to identify and extrapolate flow patterns. *Physics of Fluids*, 29(8), 084102.
- Le Clainche, S., Vega, J. M., & Soria, J. (2017). 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. <https://doi.org/10.1016/j.expthermflusci.2017.07.021>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, G., Hu, Z., Hou, F., Li, X., Wang, L., & Zhang, X. (2020). Machine learning enabled high-throughput screening of hydrocarbon molecules for the design of next generation fuels. *Fuel*, 265, 116968. <https://doi.org/10.1016/j.fuel.2019.116968>
- Liu, Y., Li, J., Tian, Z., Zhang, T., & Yan, Y. (2022). Experimental studies on suppression of combustion instability with the addition of helium. *Journal of Applied Fluid Mechanics*, 15(6), 1703–1715. <https://doi.org/10.47176/jafm.15.06.1193>
- Liu, Z., Braun, J., & Paniagua, G. (2020). Thermal power plant upgrade via a rotating detonation combustor and retrofitted turbine with optimized endwalls. *International Journal of Mechanical Sciences*, 188, 105918. <https://doi.org/10.1016/j.ijmecsci.2020.105918>
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Lorente, L., Vega de Prada, J. M., & Velázquez, Á. (2010). Compression of aerodynamic databases using high-order singular value decomposition. *Aerospace Science and Technology*, 14(3), 168–177. <https://doi.org/10.1016/j.ast.2009.12.003>
- Luchini, P., & Bottaro, A. (2014). Adjoint equations in stability analysis. *Annu. Rev. Fluid Mech.*, 46, 493–517.

- Lumley, J. L. (1967). The structure of inhomogeneous turbulent flows. In: Yaglam, A.M., Tatarsky, V.I. (eds.) *Proceedings of the International Colloquium on the Fine Scale Structure of the Atmosphere and Its Influence on Radio Wave Propagation. Doklady Akademii Nauk SSSR, Nauka, Moscow.*
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1*, 281–297.
- Maddaloni, D., Domínguez-Vázquez, A., Terragni, F., & Merino, M. (2022). Data-driven analysis of oscillations in hall thruster simulations. *Plasma Sources Science and Technology, 31*(4), 045026. <https://doi.org/10.1088/1361-6595/ac6444>
- Maged, A., & Nour, M. (2025). Prediction of combustion pressure with deep learning using flame images. *Fuel, 380*, 133203. <https://doi.org/https://doi.org/10.1016/j.fuel.2024.133203>
- Malik, M. R., Vega, P. O., Coussement, A., & Parente, A. (2021). Combustion modeling using principal component analysis: A posteriori validation on sandia flames d, e and f. *Proceedings of the Combustion Institute, 38*(2), 2635–2643.
- Malinauskas, R. A., Hariharan, P., Day, S. W., Herbertson, L. H., Buesen, M., Steinseifer, U., Aycock, K. I., Good, B. C., Deutsch, S., & Manning, K. B. e. a. (2017). Fda benchmark medical device flow models for cfd validation. *ASAIO J., 63*, 150–160.
- Mariappan, S., Nath, K., & Em Karniadakis, G. (2024). Learning thermoacoustic interactions in combustors using a physics-informed neural network. *Engineering Applications of Artificial Intelligence, 138*, 109388. <https://doi.org/https://doi.org/10.1016/j.engappai.2024.109388>
- Marquet, O., Sipp, D., & Jacquin, L. (2008). Sensitivity analysis and passive control of cylinder flow. *Journal of Fluid Mechanics, 615*, 221–252.
- Martín, J. A., Corrochano, A., Sierra, J., Fabre, D., & Le Clainche, S. (2021). Modeling double concentric jets using linear and non-linear approaches. In Á. Herrero, C. Cambra, D. Urda, J. Sedano, H. Quintián, & E. Corchado (Eds.), *15th international conference on soft computing models in industrial and environmental applications (soco 2020)* (pp. 451–459). Springer International Publishing.
- Martínez, A., Güemes, A., Perales, J. M., & Vega, J. M. (2020). Variable thickness in plates—a solution for shm based on the topological derivative. *Sensors, 20*(9), 2529. <https://doi.org/10.3390/s20092529>
- Martínez-Sánchez, Á., Lazpita, E., Corrochano, A., Le Clainche, S., Hoyas, S., & Vinuesa, R. (2023). Data-driven assessment of arch vortices in simplified urban flows. *International Journal of Heat and Fluid Flow, 100*, 109101. <https://doi.org/10.1016/j.ijheatfluidflow.2022.109101>
- McKeon, B. J., & Sharma, A. S. (2010). A critical-layer framework for turbulent pipe flow. *Journal of Fluid Mechanics, 658*, 336–382.
- McQuarrie, S. A., Huang, C., & Willcox, K. E. (2021). Data-driven reduced-order models via regularised operator inference for a single-injector combustion process. *Journal of the Royal Society of New Zealand, 51*(2), 194–211. <https://doi.org/10.1080/03036758.2020.1863237>
- Mekhaël, N., Pasquato, M., Carenini, G., Braga, V. F., Trevisan, P., Bono, G., & Hezaveh, Y. (2024). Variable star light curves in koopman space.

- Mirgolbabaei, H., & Echehki, T. (2013). A novel principal component analysis-based acceleration scheme for les-odt: An a priori study. *Combustion and Flame*, 160(5), 898–908. <https://doi.org/https://doi.org/10.1016/j.combustflame.2013.01.007>
- MODELAIR. (2023–2026). Groundbreaking tools and models to reduce air pollution in urban areas.
- Muñoz, E., & Le Clainche, S. (2022). On the topology patterns and symmetry breaking in two planar synthetic jets. *Physics of Fluids*, 34(2), 024103. <https://doi.org/10.1063/5.0080834>
- Özden, A., Procacci, A., Malpica Galassi, R., Contino, F., & Parente, A. (2024). Incremental sampling methods for multi-fidelity surrogate modeling: Application on a furnace operating in mild combustion conditions. *Applied Thermal Engineering*, 246, 122902. <https://doi.org/https://doi.org/10.1016/j.applthermaleng.2024.122902>
- Palkin, E. V., Hrebtov, M. Y., Slastnaya, D. A., Mullyadzhhanov, R. I., Vervisch, L., Sharaborin, D. K., Lobasov, A. S., & Dulin, V. M. (2022). Influence of a central jet on isothermal and reacting swirling flow in a model combustion chamber. *Energies*, 15(5). <https://doi.org/10.3390/en15051615>
- Palomba, I., Richiedi, D., & Trevisani, A. (2016). Mode selection for reduced order modeling of mechanical systems excited at resonance. *International Journal of Mechanical Sciences*, 114, 268–276. <https://doi.org/https://doi.org/10.1016/j.ijmecsci.2016.05.026>
- Palomeque-Santiago, R., Domínguez-González, A., Martínez-Ruiz, D., Rubio-Rubio, M., Fernández Tarrazo, E., & Sánchez-Sanz, M. (2024). Unveiling the bi-stable character of stealthy hydrogen–air flames. *Physics of Fluids*, 36(8), 087137. <https://doi.org/10.1063/5.0222108>
- Parente, A., & Sutherland, J. C. (2013). Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combustion and flame*, 160(2), 340–350.
- Parente, A., Sutherland, J. C., Tognotti, L., & Smith, P. J. (2009). Identification of low-dimensional manifolds in turbulent flames. *Proceedings of the Combustion Institute*, 32(1), 1579–1586.
- Parente, A., Sutherland, J., Dally, B. B., Tognotti, L., & Smith, P. (2011). Investigation of the mild combustion regime via principal component analysis. *Proceedings of the Combustion Institute*, 33(2), 3333–3341.
- Parente, A., & Swaminathan, N. (2024). Data-driven models and digital twins for sustainable combustion technologies. *iScience*, 27(4), 109349. <https://doi.org/https://doi.org/10.1016/j.isci.2024.109349>
- Perales-Díaz, J., Domínguez-Vázquez, A., Fajardo, P., & Ahedo, E. (2023). Simulations of driven breathing modes of a magnetically shielded hall thruster. *Plasma Sources Science and Technology*, 32(7), 075011. <https://doi.org/10.1088/1361-6595/ace651>
- Pralits, J. O., Brandt, L., & Gianetti, F. (2010). Instability and sensitivity of the flow around a rotating circular cylinder. *Journal of Fluid Mechanics*, 650, 513–536. <https://doi.org/10.1017/S0022112009993764>
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1988). *Numerical recipes in c: The art of scientific computing*. Cambridge University Press.
- Procacci, A., Amaduzzi, R., Coussement, A., & Parente, A. (2023). Adaptive digital twins of combustion systems using sparse sensing strategies. *Proceedings of the Combustion*

- Institute*, 39(4), 4257–4266. <https://doi.org/https://doi.org/10.1016/j.proci.2022.07.029>
- Qiu, P., Lu, C., Zhang, L., Xing, C., Cao, Z., Liu, L., Peng, J., & Yu, X. (2024). Extraction of turbulent flame structures and dynamic modes with corrected oh-plif images for a hydrogen micromix burner. *Sustainable Energy Fuels*, 8, 3865–3876. <https://doi.org/10.1039/D4SE00023D>
- Quan Reyes, D., Roekaerts, D., & van Oijen, J. (2024). Direct numerical simulation of igniting non-premixed hydrogen combustion for the argon power cycle. *Proceedings of the Combustion Institute*, 40(1), 105553. <https://doi.org/https://doi.org/10.1016/j.proci.2024.105553>
- Ranzi, E., Frassoldati, A., Grana, R., Cuoci, A., Faravelli, T., Kelley, A. P., & Law, C. K. (2012). Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels. *Progress in Energy and Combustion Science*, 38(4), 468–501.
- Richecoeur, F., Hakim, L., Renaud, A., & Zimmer, L. (2012, December). DMD algorithms for experimental data processing in combustion. In *Proceeding of the 2012 Summer Program* (pp. 459–468). Center for Turbulence Research, Stanford University.
- Rieth, M., Gruber, A., Williams, F. A., & Chen, J. H. (2022). Enhanced burning rates in hydrogen-enriched turbulent premixed flames by diffusion of molecular and atomic hydrogen [A dedication to Professor Kenneth Noel Corbett Bray]. *Combustion and Flame*, 239, 111740. <https://doi.org/https://doi.org/10.1016/j.combustflame.2021.111740>
- Rodríguez-Pastor, M., Koumides, P., Pérez-Sánchez, E., García-Tíscar, J., Broatch, A., & Mira, D. (2024). Numerical analysis of laminar velocity-forced premixed slit flames using modal decomposition techniques. *Combustion and Flame*, 269, 113661. <https://doi.org/https://doi.org/10.1016/j.combustflame.2024.113661>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by backpropagating errors* (Vol. 323). Nature.
- Salvi, B., & Subramanian, K. (2015). Sustainable development of road transportation sector using hydrogen energy system. *Renewable and Sustainable Energy Reviews*, 51, 1132–1155. <https://doi.org/10.1016/j.rser.2015.07.059>
- Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656, 5–28.
- Schmidt, O. T., & Henningson, D. S. (2018). Spectral proper orthogonal decomposition of fluid flows. *Journal of Fluid Mechanics*, 842, 131–159. <https://doi.org/10.1017/jfm.2018.209>
- Sharma, A., De, A., & Kumar, S. (2024). Investigation of injector-coupled combustion dynamics in a methane–oxygen combustor using large eddy simulation and dynamic mode decomposition. *Physics of Fluids*, 36(6), 065116. <https://doi.org/10.1063/5.0212604>
- Sharma, A. J., Johnson, R. F., Kessler, D. A., & Moses, A. (2020). Deep learning for scalable chemical kinetics. In *Aiaa scitech 2020 forum*. <https://doi.org/10.2514/6.2020-0181>
- Shoji, T., Iwasaki, Y., Kato, K., Yoshida, S., Tachibana, S., & Yokomori, T. (2020). Effects of flame structures on direct combustion noise produced by lean-premixed H_2 /air low-swirl jet flames. In *Aiaa scitech 2020 forum*. <https://doi.org/10.2514/6.2020-0918>
- Sirovich, L. (1987). Turbulence and the dynamics of coherent structures, parts i-iii. *Quarterly of Applied Mathematics*, 45(3), 561–590.

- Smith, G. P. (1999). Gri-mech 3.0. http://www.me.berkley.edu/gri_mech/.
- Son, H., & Lee, M. (2024). A pinn approach for identifying governing parameters of noisy thermoacoustic systems. *Journal of Fluid Mechanics*, 984, A21. <https://doi.org/10.1017/jfm.2024.219>
- STABFEM. (2020). Gitlab repository: www.gitlab.com/stabfem/StabFem.
- Staffell, I., Scamman, D., Abad, A., Balcombe, P., Dodds, P., Ekins, P., Shah, N., & Ward, K. (2019). The role of hydrogen and fuel cells in the global energy system. *Energy & Environmental Science*, 12, 463–491. <https://doi.org/10.1039/C8EE01157E>
- Strykowski, P. J., & Sreenivasan, K. R. (1990). On the formation and suppression of vortex 'shedding' at low reynolds number. *J. Fluid Mech.*, 218, 71–107.
- Suzuki, S., Iwata, K., Kai, R., & Kurose, R. (2024). A dns study of detonation in h2/o2 mixture with variable-intensity turbulences. *Proceedings of the Combustion Institute*, 40(1), 105337. <https://doi.org/https://doi.org/10.1016/j.proci.2024.105337>
- Swischuk, R., Kramer, B., Huang, C., & Willcox, K. (2020). Learning physics-based reduced-order models for a single-injector combustion process. *AIAA Journal*, 58(6), 2658–2672. <https://doi.org/10.2514/1.J058943>
- Tagliaferro, S., Corrochano, A., Marchetti, P., Marcon, A., & Le Clainche, S. (2024). A new method based on physical patterns to impute aerobiological datasets. *PLOS ONE*, 19(11), 1–14. <https://doi.org/10.1371/journal.pone.0314005>
- Tipler, S., D'Alessio, G., Van Haute, Q., Parente, A., Contino, F., & Coussement, A. (2022). Predicting octane numbers relying on principal component analysis and artificial neural network. *Computers & Chemical Engineering*, 161, 107784.
- Torrez, S. M., Driscoll, J. F., Ihme, M., & Fotia, M. L. (2011). Reduced-order modeling of turbulent reacting flows with application to ramjets and scramjets. *Journal of Propulsion and Power*, 27(2), 371–382. <https://doi.org/10.2514/1.50272>
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3), 279–311.
- Turns, S. R., & Haworth, D. C. (2021). *An introduction to combustion: Concepts and applications* (5th). McGraw-Hill.
- United Nations Environment Programme. (1985). Vienna convention for the protection of the ozone layer [United Nations, Vienna, Austria].
- United Nations Environment Programme. (1987). Montreal protocol on substances that deplete the ozone layer [United Nations, Montreal, Canada].
- van den Berg, R., Hoefsloot, H., Westerhuis, J., Smilde, A., & van der Werf, M. (2006). Centering, scaling and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*, 7, 142. <https://doi.org/10.1186/1471-2164-7-142>
- Varela, P., Suárez, P., Alcántara-Ávila, F., Miró, A., Rabault, J., Font, B., García-Cuevas, L. M., Lehmkuhl, O., & Vinuesa, R. (2022). Deep reinforcement learning for flow control exploits different physics for increasing reynolds number regimes. *Actuators*, 11(12). <https://doi.org/10.3390/act11120359>
- Vega, J., & Le Clainche, S. (2020). Higher order dynamic mode decomposition and its applications. *BOOK: Elsevier*.
- Wong, P., Tam, L., Li, K., & Vong, C. (2010). Engine idle-speed system modelling and control optimization using artificial intelligence [Cited by: 58]. *Proceedings of the Institution*

- of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 224(1), 55–72. <https://doi.org/10.1243/09544070JAUTO1196>
- World Meteorological Organization (WMO). (2022). *Scientific assessment of ozone depletion: 2022* (tech. rep. No. GAW Report No. 278). World Meteorological Organization. Geneva.
- Yahou, T., Dawson, J. R., Poinso, T., Selle, L., & Schuller, T. (2024). Flame front dynamics during ignition of lean premixed h₂/air and ch₄/air flames. *Combustion and Flame*, 269, 113708. <https://doi.org/https://doi.org/10.1016/j.combustflame.2024.113708>
- Yang, J., Zhao, X., Lu, X., Lin, X., & Xu, G. (2015). A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis. *Frontiers in Molecular Biosciences*, 2, 4. <https://doi.org/10.3389/fmolb.2015.00004>
- Zdybal, K., D'Alessio, G., Aversano, G., Malik, M. R., Coussement, A., Sutherland, J. C., & Parente, A. (2023). Advancing reacting flow simulations with data-driven models. In M. A. Mendez, A. Ianiro, B. R. Noack, & S. L. Brunton (Eds.), *Data-driven fluid mechanics: Combining first principles and machine learning* (pp. 304–329). Cambridge University Press. <https://doi.org/10.1017/9781108896214.022>
- Zdybał, K., D'Alessio, G., Attili, A., Coussement, A., Sutherland, J. C., & Parente, A. (2023a). Local manifold learning and its link to domain-based physics knowledge. *Applications in Energy and Combustion Science*, 14, 100131. <https://doi.org/https://doi.org/10.1016/j.jaecs.2023.100131>
- Zdybał, K., Parente, A., & Sutherland, J. C. (2023b). Improving reduced-order models through nonlinear decoding of projection-dependent outputs. *Patterns*, 4(11), 100859. <https://doi.org/https://doi.org/10.1016/j.patter.2023.100859>
- Zhang, S., Zhang, C., & Wang, B. (2024). Crk-pinn: A physics-informed neural network for solving combustion reaction kinetics ordinary differential equations. *Combustion and Flame*, 269, 113647. <https://doi.org/https://doi.org/10.1016/j.combustflame.2024.113647>
- Zhou, L., Song, Y., Ji, W., & Wei, H. (2022). Machine learning for combustion. *Energy and AI*, 7, 100128. <https://doi.org/https://doi.org/10.1016/j.egyai.2021.100128>