



HAL
open science

Reduced Order Model of Diffusion Flames Based on Multi-scale Data from Detailed CFD: The Impact of preprocessing

Nicole Lopes Junqueira, Louise da Costa Ramos, Luís Fernando Figueira da Silva

► **To cite this version:**

Nicole Lopes Junqueira, Louise da Costa Ramos, Luís Fernando Figueira da Silva. Reduced Order Model of Diffusion Flames Based on Multi-scale Data from Detailed CFD: The Impact of preprocessing. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 2024, 46 (4), pp.215. 10.1007/s40430-024-04749-6 . hal-03738492v2

HAL Id: hal-03738492

<https://hal.science/hal-03738492v2>

Submitted on 2 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reduced Order Model of Diffusion Flames Based on Multi-scale Data from Detailed CFD: The Impact of preprocessing

Nicole Lopes Junqueira^{1,2,3*}, Louise da Costa Ramos^{4†}
and Luís Fernando Figueira da Silva^{1,2†}

^{1*}Department of Mechanical Engineering, Pontificia Universidade Católica do Rio de Janeiro, 22451-900, Rio de Janeiro, Brazil.

²Institut Pprime UPR 3346 CNRS, ENSMA and University of Poitiers, 86961, Futuroscope Chasseneuil, France.

³Ansys Fr, 69100, Villeurbanne, France.

⁴GIPSA-lab, Université Grenoble Alpes, Grenoble INP, Grenoble, France.

*Corresponding author(s). E-mail(s):

nicole.lopes-monteiro-de-barros-junqueira@ensma.fr;

Contributing authors: louise.da-costa-ramos@grenoble-inp.fr;

luis-fernando.figueira-da-silva@ensma.fr;

†These authors contributed equally to this work.

Abstract

Machine learning techniques, such as reduced order models (ROM), have demonstrated low cost when creating models of complex systems while aiming at the same accuracy as high fidelity models, such as Computational Fluid Dynamics (CFD). Here, ROM are created using data from CFD simulations of non-premixed laminar flames with detailed chemistry and transport. The data obtained for variable fuel velocity are reduced using singular value decomposition (SVD) and then a genetic aggregation response surface algorithm is applied to predict the properties fields for an arbitrary velocity. This work analyzes the effect of different data preprocessing approaches on the ROM, i.e., (1) the properties treated as a uncoupled or as a coupled system, (2) normalization of different properties, and (3) the logarithm of the chemical species.

For all constructed ROM, the energy content of the reduction process and the reconstructed fields of the flame properties evidence the slow convergence of SVD modes for the uncoupled ROM, while a faster one is seen when the logarithm preprocessing is applied. Also, the learning is shown to be achieved with a smaller number of modes for two of the coupled ROM and the ROM using the logarithm. The reconstruction of the mass fraction fields is characterized by regions of negative values, underscoring that the baseline ROM methodology does not preserve the properties monotonicity, positivity and boundedness. The proposed logarithm preprocessing enables to overcome such problems and to accurately reproduce the original data.

Keywords: Machine learning, computational fluid dynamics, non-premixed flames, methane/air combustion.

1 Ethical statement

- **Funding:** Nicole Lopes Junqueira and Luís Fernando Figueira da Silva received funding from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Part of this work was developed while Louise da C. Ramos received funding from the European Union's Horizon 2020 research and innovation programme under grant Agreement nr 766264.
- **Conflict of Interest:** the authors have no competing interests to declare that are relevant to the content of this article.
- **Ethical approval:** not applicable.
- **Informed consent:** not applicable.
- **Author contribution:** all authors contributed to the study conception and design. The first draft of the manuscript was written by Nicole Lopes Junqueira, Louise de Costa Ramos and Luis Fernando Figueira da Silva and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.
- **Academic partnership:** Ansys/France (TwinBuilder Team) and Pontifícia Universidade Católica do Rio de Janeiro (Brazil).

2 Introduction

The combustion process is a multi-scale phenomenon, which means that different physico-chemical processes occur at different time scales, spanning several orders of magnitude. In non-premixed flames, which are of interest in this work, the transport of fuel and oxidant towards the reaction zone is controlled by diffusion. The multi-step chemical reaction present in such flames occurs in the vicinity of the stoichiometric surface, from which the combustion products and the heat released are transported towards the fresh unburned gases. Due to such process being present in the flame, these time scales manifest

as length scales also. Even after decades of a steady increase in the available computational resources and algorithm improvement which enable the study of combustion, the direct numerical simulation of hydrocarbon-air flames in scenarios of industrial interest remains challenging. Indeed, the accurate numerical modeling of such flame requires that such time and length scales are fully captured, which often entails a large computational cost. For this reason the present work endeavors to contribute to the reduced-order model development of a non-premixed laminar flame configuration.

The studied flame configuration is that of a classical Gülder burner flame [1–5]. This burner has already been used to compare the effects of nitrogen dilution and flame temperature on soot formation in ethylene diffusion flames. More recently, studies on the Gülder burner have been performed focusing on the soot formation characteristics in non-premixed laminar flames of a mixture of n-heptane/butanol isomer and air, and also for soot characterization when ammonia is diluted to n-heptane fuel. Despite the availability of experimental data for comparison purposes, the aim of the present work is not on comparing computational and experimental results. Nevertheless, the focus here is on improving the learning process in a specific class of machine learning algorithms, known as reduced-order models (ROM).

Machine learning (ML) is an evolving branch of computational algorithms. In particular, this approach is based on fundamental mathematics, linear algebra, optimization, and regression, and its performance improves as it is exposed to more and different type of data [6]. Ideally, ML generalizes the given data, learning its patterns and correlating it with the outcomes that are intended to be predicted. Once the model is validated, it can be applied to new database values [7]. Examples of machine learning applications include identifying objects in images, selecting relevant search results, and machine translation [8].

Machine learning algorithms are classified according to its structure, which can be supervised or unsupervised, depending on the information available to the model [7]. Supervised learning is the most common form of ML, where the goal is to make predictions of a target by having expert knowledge learning, providing corrective information to the algorithm [8]. In unsupervised approach, the learning occurs without training data being labeled, where the goal is to find a structure in the data [6, 7].

Regarding the learning available data, machine learning approaches can be divided into online and offline models. Online models learn the characteristics of a system while the data is being collected, e.g., artificial neural networks applied to image recognition [8]. Concerning fluid dynamics applications, the estimation of the eddy viscosity has been developed using online identification, which enabled an adaption of a reduced order model to changes of the flow configuration [9]. Offline models, however, learn the behavior based on previously collected data. For instance, the ROM of the Navier-Stokes equations of a flow passing a cylinder has already been done using an offline data collection [10].

The applications of these algorithms embrace different fields of knowledge, such as engineering, biomedical, and finance [11–13]. For instance, the ML has already been applied to the analysis of genome sequencing data sets or applications in agricultural supply chains in different phases [14, 15]. However, such algorithms are not yet widely accepted in the operation of engineering systems, since they are often considered as black-box models, i.e., no prior knowledge about the underlying physics of the problem or its restrictions is considered for the learning [8, 16]. To overcome such limitation, it has been proposed to combine ML with first principles models of engineering systems [16, 17]. An example is the physics-informed machine learning used to predict the critical heat flux with superior performance over standalone approaches [18].

Concerning the combustion process, machine learning techniques applications have been used for over two decades [19]. For instance, a reduced order model based on CFD simulations results for oxy-coal combustion enabled the estimation of the average outlet temperature of the burnt gases for a given fuel and oxidant mass flow rates, and also to determine the inlet mass flow rate required to obtain the desired temperature [20]. In a different application, a non-intrusive reduced order model has been applied for an unstable flow using a proper orthogonal decomposition combined with a feed-forward neural network approach [21]. Recently, a non-intrusive methodology, using the proper orthogonal decomposition and an interpolation method, has been applied to construct a digital twin using CFD simulations and real-time measurements of an industrial furnace [22].

The reduced order model is a ML technique which applies a reduction method to sort the data. For instance, singular value decomposition (SVD) or proper order decomposition (POD) are usually applied as reduction methods. Subsequently, an interpolation/integration method is applied to learn the behavior of the reduced system. As an example, the application of the POD approach for a ROM of the Navier-Stokes equation has been used with different methods of integration/interpolation, comparing the impact of each method on the ROM [10]. Another recent application of ROM, based on SVD, has been the coupling with a genetic aggregation response surface (GARS) to predict the behavior of a laminar premixed inverted conical flame [23].

The particular application of reduced order model to combustion problems covered different topics. For instance, ROM has been used to the description of properties of a laminar premixed flame based on steady CFD data [23]. Furthermore, the analysis of the infrared radiation emitted by a reacting, supersonic, turbulent jet has been developed using a ROM based on numerical simulation [24]. Also, an approach that blends data-driven learning with theoretical foundations has been proposed for a single injector combustor of a rocket engine, willing to predict its combustion related properties, such as pressure and temperature [25]. A similar study on physics-based data-driven methods has been proposed to learn the behavior of a single-injector combustor ROM using high-fidelity simulations [26].

Even though turbulent flames typifies most practical combustion industrial processes, the multi-scale physics of laminar non-premixed flames enables the study of combustion from the perspective of the computational cost of high-fidelity CFD simulations. The associated multiple time and length scales present in such flames increase the complexity of such models and its learning process. Furthermore, the application of reduced order model methodology is not as explored for laminar non-premixed flames, as it is for turbulent flames [19, 23].

Regarding the data preprocessing, it is vastly known that it could either ease or complicate the ML data based model predictions [8, 16]. In particular, the data normalization is a practice widely used to transform different data values to a common scale, without distorting the differences in the value ranges [7, 11]. Furthermore, to ensure that algorithms learn with prior knowledge about the underlying physics of a problem or its restrictions, it has been proposed to combine ML with first principles models of engineering system, e.g., the physics informed machine learning used to predict the critical heat flux, with superior performance over standalone approaches [16–18]. Here, the focus is on using data preprocessing to solve the monotonicity problem found in ROM applied to laminar flames using known combustion principles, such as the exponential decay of chemical species with distance [27–30].

The main objective of this work is to develop reduced order models of a methane/air laminar non-premixed flame stabilized on a Glder burner, using computational fluid dynamic modeling results to construct ROM. The corresponding specific objectives are: To analyze the influence of preprocessing the learning data, which are the CFD modelled combustion properties, on the ROM results. More specifically, the impact of creating a ROM for each uncoupled flame property or treating the properties as a coupled system; The effect of normalizing the data set, including the species data since some minority species mass fraction are of the order of magnitude of 10^{-6} ; The influence of applying a logarithm transformation of species mass fraction to construct the ROM; To develop an analysis of the reconstructed modes energy content of each methodology applied to the learning data; To compare the CFD and ROM results.

3 Methodology

In this section, the steps to model a non-premixed diffusion flame are presented. First, the studied diffusion flame and burner are given, based on experimental knowledge, followed by the experimental and geometrical setup definition. Then, the two used models are outlined, i.e., the computational fluid dynamics and the reduced order models. Lastly, the methods applied for preprocessing the learning data used for the ROM are presented.

3.1 Studied Non Premixed Flame Configuration

In non-premixed flames the transport of fuel and oxidant happens due to diffusion towards the reaction region, where chemical reactions occur. Hydrocarbon/air diffusion flames are known to have a high production of soot when compared to lean premixed flames, and have been extensively used on the study of soot formation [27, 31]. In this work, the study and modeling of non-premixed diffusion flames at the well-known Gülder burner is developed.

The Gülder burner has a simple geometry and is widely used on the study of stable, axisymmetric, non-premixed laminar flames [1–3, 32]. This burner presents an axial symmetric flow entering the system through two inlets [2], as shown in Fig. 1a. The non-premixed fuel enters through the central tube, which inner diameter is 11 mm, whereas the air enters through the annular region with a radius of 50 mm, being responsible for the flame stabilization. Figure 1b presents an example of an ethylene/air laminar diffusion flame stabilized on a Gülder burner [2].

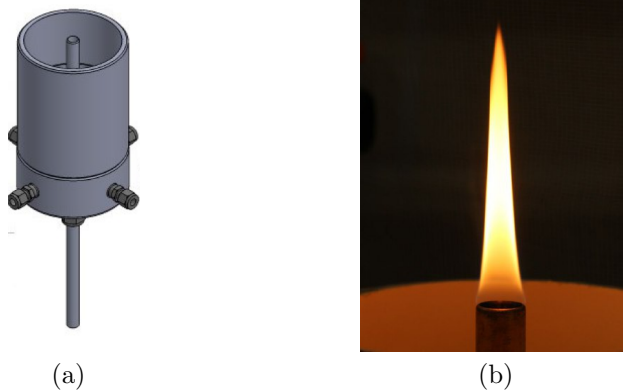


Fig. 1 Gülder burner (a) schematic representation and (b) stabilized ethylene/air flame [2].

Based on the burner characteristics [28], one can define the non-premixed flame experimental setup and geometry and boundary conditions, as presented on the following section, which are necessary to the development of the combustion modeling of interest.

To model the Gülder burner flame, it is first necessary to define the system geometry and mesh, according to the problem of interest. Since the flame is supposed to be axisymmetric, only a slice of the actual physical domain is modeled, such that a 2D model represents the burner and computational domain, as shown at Fig. 2.

The geometry dimensions of the fuel and air inlets, and the burner wall thickness are the same as for the experiments, i.e., radius of 5.5 mm for the fuel inlets, and a tube wall thickness of 1 mm. The two dimensional geometry given at Fig. 2 represents the domain of interest and its boundary conditions, using

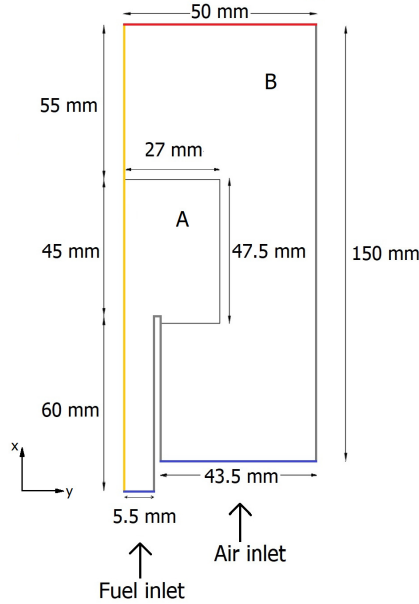


Fig. 2 Glder burner computational domain and its dimensions. (A): internal part, (B): external part.

different colors to represent each boundary. The symmetry axis, represented by the yellow line, has a length of 160 mm. The four gray lines represent the walls, whereas the outlet and the two inlets are the red and the blue lines, respectively. Two types of walls are considered in this domain; (1) the burner wall, which has an adiabatic, non catalytic, no-slip condition, and (2) the one representing the outer boundary, which has a slip and constant temperature condition. The walls between the burner inlets are denominated as burner walls, and the free boundary is the external right boundary, with a length of 150 mm.

The burner is designed to ensure that the flame is laminar, and the fuel flow at the outlet of the fuel tube is fully developed. Therefore, the tube length is long enough to guarantee the laminar fuel flow [33]. The length of the fuel and air feeding tubes are 60 mm and 50 mm, respectively. In this experiment, air and fuel enter the domain through different inlets, as shown in Fig. 2, such that the mixing of reactants only occurs at the reaction zone. Regarding the inflow conditions, at the computational model inlets, the species molar fraction, temperature, pressure and the inlet velocity are given. The air is a mixture composed by 21% O_2 and 79% N_2 , and the fuel is methane. Air and fuel have prescribed temperature and pressure of 300 K and 1 atm, respectively.

Based on the described system geometry and boundary conditions, the following sections of this work present the overview physical and computational models applied in this work.

3.2 Physical Mathematical Model Summary

The studied methane/air diffusion flames are modeled with transport equations of species, mass, momentum and energy in an axisymmetric reference frame. The species mass transport equation is solved accounting for the multi-species diffusion coefficients, the Fick's law of diffusion, the Soret thermodiffusion effects, but the barodiffusion effects are neglected since the flame is isobaric. The solved energy transport equation accounts for the non unity Lewis number effects, but neglecting thermal radiation since the modeled flame here does not present soot formation. Indeed, the participating medium emission relies on CO₂ and H₂O at the burnt gases only, thus the flame is considered transparent regarding the absorption process. Since such mathematical formulation has been extensively studied in the context of combustion of laminar flames, no further discussion is going to be developed this work, and one can refer to [27, 34, 35] for detailed formulations. The momentum transport equation considers the buoyancy effect, and Newtonian fluid behavior is assumed.

To model the chemical kinetic process a skeletal model called DRM19 is applied [36], which is a reduced GRI-Mech 1.2 kinetic model. The DRM19 skeletal kinetic model is composed of 19 (plus N₂) species and 84 reactions [36]. Such model has already been used to model laminar flames in different configurations and one can refer to [23, 37, 38] for such applications.

3.3 Computational Fluid Dynamics Model Overview

In this work, the model of the diffusion flame is developed using *Ansys* Fluent 2020 R2, which applies a finite volume technique, using the following solution methods: coupled for the pressure-velocity coupling, Presto for the spatial discretization of pressure, and the second order upwind to calculate the spatial discretization of energy, species and momentum [34]. Since those methods are quite standard, no further discussion is developed here.

The stiff chemistry solver is applied to account for the reaction in the species transport equation, consisting of a fractional step algorithm for pressure-based unsteady simulations [39]. Moreover, the In Situ Adaptive Tabulation technique (ISAT table) is used to solve the chemistry in each cell for a constant pressure. Here, the convection and diffusion terms are treated as in a non-reacting simulation.

The detailed chemical kinetics mechanisms contain several intermediate species in addition to the principal species, and these intermediate species evolve at different reaction rates, leading to a large variation of time scales for species formation and consumption [27]. The accurate simulation of such system, thus, requires small time steps, at mili-second order, increasing the necessary computational time. Nevertheless, the in-situ adaptive tabulation integration (ISAT) tool [40] enables the reduction of the computational burden.

Indeed, the ISAT has been widely used as an integration method in combustion simulations [41, 42]. It has been employed to integrate the stiff chemistry

to reduce the burden of a direct integration of the chemistry [40, 41, 43–45]. However, the performance of ISAT decreases in flames with large time scales, since more work is required on the integrator of the ordinary differential equation [41]. In this work, the ISAT table is used as integration method, with an error tolerance of 10^{-4} .

Moreover, to properly describe the combustion process and all of its scales, a refined mesh is necessary on the domain, but using such refined mesh may lead to an infeasible simulations cost. To overcome this problem, as shown in Fig. 2, the domain is divided in two parts, one internal, representing the combustion region (A), and one external, representing the fresh and burnt gases (B). Since the part (A) contains the reactive zone, it has a refined mesh. Such internal part has a rectangular mesh, with an initial size of $100\ \mu\text{m}$ and $1\ \text{mm}$ in the radial and axial directions, respectively. The external part (B) has a uniform mesh with a size of $1\ \text{mm}$. Even though the two mesh blocks are used to decrease the simulation burden, it is not enough to properly compute the combustion scales, especially the minor species. Therefore, a mesh adaptation tool is applied to refine and coarsen the mesh when needed.

The mesh adaption tool is then applied to reduce the computational burden linked to the detailed combustion models, thus accurately solving the flow characteristics. Such adaptive mesh refinement is usually performed to reduce the numerical error with reduced numerical cost. It enables the refinement and/or coarsening of the mesh during the numerical simulation [46, 47], adding or deleting mesh cells/nodes based on a pre-determined property range. For instance, one can set the adaption based on the temperature gradient, and then set the minimum and maximal ranges for such property [48].

To overcome the burden linked to modeling such complex combustion systems, a gradient mesh adaptation based on the temperature is applied in this work, where the refinement and coarsening thresholds are $10\ \text{K/m}$ and $300\ \text{K/m}$, respectively. A trial and error procedure, based on the temperature sensitivity test in mesh adaptation [38] has been performed to determine the best threshold values for the case of the laminar diffusion flame, with a fuel inlet velocity of $17.5\ \text{cm/s}$. For the sake of brevity, these tests are not included in this work, one may consult [37] for a detailed explanation of the mesh adaption impact on the modeling of a lean premixed flame.

The adaption based on temperature gradient is well known to ensure an appropriate adaptation criterion for combustion systems [36]. In this work, an adaptation based on a single property, temperature, is effected every 25 iterations. To exemplify the impact of such adaption procedure on the current work, the initial mesh has 11,667 nodes, however, when the simulation for the fuel inlet velocity is $4.38\ \text{cm/s}$ converges, it has 81,501 more nodes than the original mesh.

It is worth to mention that the ignition of such reactive simulation is performed by imposing a high-temperature region over a fully developed fuel/air isothermal field. More precisely, departing from a converged isothermal simulation, a patch of $3\ \text{mm}$ in the radial and $5\ \text{mm}$ in the axial direction is set

at the stoichiometric line ($Y_{CH_4,st} = 0.055$), 1 mm above the fuel feed tube outlet. This patch is initialized with a temperature of 1,800 K, which is of the order of the methane adiabatic flame temperature. Once this reactive simulation converges, the properties fields obtained are used as a ignition trigger initialization for subsequent simulations. Such initialization method decreases the simulation time necessary for the flame to stabilize, since the ignition phase is taken into account each time.

Now that the model has been set, it is used to generate data representative of a Glder burner flame for different configurations, which are presented later in this work. However, in order to obtain high-fidelity CFD simulations, there is a high computational cost, especially in relation to memory and simulation time, which often makes a parametric study unfeasible. Therefore, a machine learning technique, i.e., a reduced order model has been applied to overcome this shortcoming [10, 19]. More specifically, CFD data generated through the previously presented model is used to train a reduced order model, and the next section presents the methodology that is applied to develop the ROM in this work.

3.4 Reduced Order Models

Reduced order models (ROM) are known to enable a simple representation of complex systems, but keeping the main characteristics of those systems [7, 19, 49]. In this work the software Static ROM, of Twin Builder from *Ansys*, is used to create reduced models for combustion. Such software builds the model based on four main steps, as shown in Fig. 3. Its construction steps are the following; (1) choosing the learning data set, (2) decomposing the data in simplified form, (3) retaining the main characteristics of the learning set and (4) applying a machine learning interpolation.

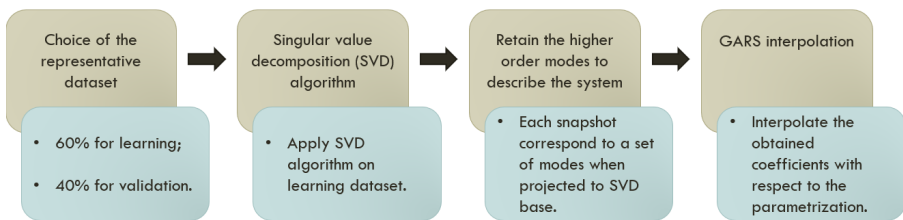


Fig. 3 Procedure scheme to construct a reduced order model with *Ansys* Static ROM.

The first step consists on obtaining representative data of the system of interest, among which, in this case, 60% is used to learn the model, and the remaining 40% to its validation. One may note that the learning set must include the extremes of the parametric space, in order to avoid an extrapolation when the obtained ROM is used. In the second step, a singular value decomposition (SVD) is applied to decompose the system and obtain its modes. The SVD decomposition organizes the modes matrix in decreasing order, such that

the higher mode values are on the top of the matrix, and those retain the most important characteristics of the system of interest [50]. Nevertheless, based on the SVD results, one can reduce the order of the model simply by choosing the first modes of the system.

It is worth stressing that the data matrix that is decomposed to create the ROM in the second step of Fig. 3 contains one learning CFD case in each column, and each row is the information that corresponds to a domain cell. Note that the each data provided to create the ROM has the same number of rows, which means that even though CFD simulations use adaptive meshing, the data provided to the ROM is extracted from the initial uniform mesh of only the part A from the computational domain.

The third step constitutes the filtering of the system through the SVD, where only the most important characteristics are retained, i.e., the higher order modes. This procedure generates a reduced base which is then used to describe the entire system [50].

The last step of the modeling is, then, the interpolation with Genetic Aggregation Response Surface (GARS), which applies a selection process to determine what type of response surface (RS) better describes the system [51, 52]. Such method is defined as a weighted average of a set of response surfaces, and has a cross-validation process, where the risk of an algorithm spreading the data set is estimated, and thus present an improved reliability when compared to other classical response surface methodologies [51, 53].

Moreover, the genetic aggregation algorithms (GA) are based on the principle of natural selection, improving the population from generation to generation, and making the algorithm more effective [6, 53]. Such algorithms can be based on different processes, such as crossover, mutation or replication [51]. More specifically, the Genetic Aggregation Response Surface (GARS) is based on different response surface methods as population, which, in this work, corresponds to the integration methods of polynomial regression, Kriging, support vector regression and moving least squares [54–57]. The GARS algorithm, thus, chooses the combination of response surface (RS) that best describes the system, which, here, are the flame properties fields.

To summarize, a reduced model is created, based on the chosen modes and RS, describing the behavior of the system. Furthermore, the model validation is performed through the available validation data. It is worth to note that the accuracy of the ROM is influenced by several factors. For instance, the data available (learning and testing), the number of modes used on the reduction of the learning set, and the error of interpolation method [25]. Varying such parameters might strongly change the ROM and its accuracy.

One may note that, in classical combustion problems, it is important to account for some physical restrictions. More specifically, with respect to the combustion scalars, monotonicity, positivity, and boundedness which are critical properties [34, 35]. Therefore, a ROM methodology for describing a non premixed flame must guarantee these three properties. Note also, that other problems, such as bankruptcy predictions, or medical diagnosis that

use machine learning algorithms, are similarly concerned by monotonicity restrictions [58].

3.5 Preprocessing Methods

On Machine Learning studies and applications, it is well known that the data provided for learning an specific system has a remarkable impact on the obtained model. As a consequence, knowing the main characteristics of the data is important to achieve a representative ROM. Thereby, data processing tools are often used to provide insights on the data, thus, facilitating the learning process.

In this work, a ROM of a laminar diffusion flame based on several flame properties obtained through CFD is created, since difficulties linked to the learning such different flame scales have been seen on previous studies. For instance, an issue related to the monotonicity of the predicted combustion properties has been observed [28]. To overcome such monotonicity problem [28], different methodologies to preprocess the available data are proposed. Firstly, a simple methodology regarding the properties being treated as an uncoupled or as a coupled system in the construction of the ROM is proposed.

For the uncoupled method, each property has its respective ROM. Thus, for a property A, the available data influences only itself and its reduced order model, and the same goes for properties B, C, etc. Nevertheless, when the properties are considered as a coupled system, the data of all properties construct a single reduced order model. Therefore, the data of property A influences the data of other properties in the coupled ROM.

Moreover, willing to facilitate learning process, other preprocessing methodologies are applied to the learning data. More specifically, the normalization of the properties when the system is being treated as coupled, and the logarithm of the chemical species for the coupled and uncoupled data.

3.5.1 Properties normalization

More precisely, in the studied case, the data of the combustion properties present scales that differs in several orders of magnitude. For instance, the maximum temperature reaches about 2,100 K, whereas the minor chemical species, such as the CH₂ radical, have a maximum mass fraction of $3 \cdot 10^{-5}$. Since the singular values are sorted in hierarchical order on the SVD, the properties with smaller scales might be ignored when the model reduction is applied. Thus, when such a scale disparity is present, one should take normalization into account when the coupled method is envisaged to model the system.

Moreover, normalization of the learning properties has often been used on machine learning methodologies. In this work, a normalization method is applied to harmonize the variance of the learning data, reaching a common scale, without biasing the differences in the value ranges. Since the studied

flame is multi-scale, the normalization of temperature, velocity components, and species are:

$$T^* = \frac{T(k) - T_0}{T_{ad} - T_0}, \quad u^* = \frac{u(k)}{u_{max}}, \quad u_{max} = u_0 \frac{T_{ad}}{T_0}, \quad Y_i^* = \frac{Y_i(k)}{Y_{i,max}}, \quad (1)$$

where T^* , u^* and Y_i^* are the normalized temperature, velocity component and species i mass fraction. T_0 is the temperature of fresh gas, 300 K; T_{ad} is the adiabatic flame temperature of stoichiometric methane/air mixtures of 2,236 K; u_0 is the air inlet velocity, 60 cm/s, and $Y_{i,max}$ is the maximum concentration of the species i found when the CFD simulation converges.

One may note that, to estimate the temperature and velocity bounds in the reactive case, the reference temperature in the chemical equilibrium (T_{ad}) and the inlet velocity are well known, and computed through mathematical straightforward methods. However, each chemical species maximum concentration is *a priori* unknown, thus, the estimation of a reference value for such properties is not straightforward.

Furthermore, now that the normalization method of the coupled combustion system has been given, this work goes beyond the common preprocessing methodologies. In the following section, a preprocessing method for ROM models focused on combustion system is proposed.

3.5.2 Logarithm preprocessing

Here, the application of the logarithm on the computed species mass fractions reactions is proposed. Such method is chosen based on its classical known link to combustion systems. Indeed, away from the flame reaction zone the spatial decrease of the mass fractions due to diffusion and convection process is exponential [27]. Therefore, the use of the logarithm in ML applications related to chemical kinetics has a theoretical basis [29, 30], easing the learning process linked to the combustion chemical properties.

Studies on the impact of the logarithm on combustion chemical properties have already been developed. For instance, in order to reduce the cost linked to solving the chemical source term of reaction, a neural network architecture was proposed to approximate chemical kinetics in an efficient and scalable way [30]. This architecture is composed of 12 neural sub-networks trained on temperature, pressure and 10 chemical species, where the sum of the mass fractions of the chemical species has to be equal to one in order not to violate the mass conservation. Willing to strengthen an accurate prediction of the different scales of the chemical species, especially the minor and radical species, a logarithmic normalization of the input data and a loss function using mean squared logarithmic error for the chemical species concentrations is proposed. This architecture enabled the prediction of temperature, pressure and species concentration with a high degree of accuracy compared to the numerical results

for the hydrogen combustion mechanism, with an average species mass fraction error of the order of 10^{-5} [30].

Another example of the use of logarithm, as preprocessing method, is the artificial neural network (ANN) constructed for the tabulation of chemical reaction terms for premixed flames and ignition flames [29]. The use of a log-scale normalization for minor and radical species has been adopted, since the distribution of such species is strongly skewed towards 0. Thus, a standard normalization would underestimate the contributions of these minor species concentrations to the reaction kinetics and predictions. The results show that the ANN with log-scale normalization is more accurate, in particular for species with small mass fractions.

In this work, the logarithm preprocessing method is then applied to all the chemical species from the chemical mechanism applied. However, since there are species regions of the domain where the mass fraction of some species are zero, the logarithm method might present some instabilities. Thus, to overcome the problem linked to the regions where chemical species are absent, a truncation parameter (ϵ_c) is imposed on the mass fractions data. Then, the \log_{10} is applied to the data set before the ROM construction, i.e.,

$$\mathcal{Y}_i = \log_{10} \left(\max(Y_i, 10^{-\epsilon_c}) \right). \quad (2)$$

This operation is performed for all the chemical species, since the logarithm has an know impact on the chemical kinetic properties of the combustion. The ROM is then computed over the transformed chemical species \mathcal{Y}_i only, and thus no pressure, velocity or temperature are included on the model. After the model is learned, it can be applied to validation, where the logarithm results obtained with the ROM are recovered to its initial form ($Y_i = 10^{\mathcal{Y}_i}$).

Furthermore, all of the different preprocessed ROM results are interpolated on the uniform mesh through Fluent, then those are analyzed and compared with each other. In this work, a study of the influence of the truncation parameter ϵ_c on the final result is also performed, aiming to determine a value, which should best suit all species reconstruction. Such impact is developed by analyzing the influence of the singular value on the result and the reconstructed mass fractions. The three truncation parameter values are chosen as $\epsilon_c = 8, 10$ and 12 , which are all smaller than the maximum mass fractions found for the minor species.

4 Results and Discussion

In this work, the fields of several combustion properties modeled with CFD are used to create a reduced order model representative data set of the methane/air laminar non-premixed flame. Such data is obtained from 20 reactive CFD simulations, where only one parameter changes from one simulation to the other. Here, the parameter chosen is the fuel inlet velocity, which varies between 1.75 and 4.38 cm/s with an equally spaced interval.

From each CFD case, i.e. for each inlet velocity, a total of 23 property fields are extracted to build the data set (temperature; velocity components and the species mass fraction, except for the activated CH_2^*). Depending on the methodology applied in learning data preprocessing step, such properties are treated either separately or as a coupled system. Since 60% of the CFD available data is used as learning data, 12 of the 20 numerical simulations are chosen to construct the reduced order model. Note that all ROM built in this work have the same learning cases, corresponding to 60% of the data set. For the purpose of comparison, all the ROMs have the same number of modes, five, as well as the same learning cases. Then, for the sake of brevity, from the total of 8 validation cases, only the case of fuel inlet velocity of 3.1 cm/s is discussed here.

It should be stressed that each CFD computation required around 15 days to reach convergence, using 18 cores of a 24 CPU in a computer with 32 GB of memory using Windows 10. The processor used is the AMD Ryzen 9 3900X 12-core, running at 3.79 GHz. The software used is *Ansys* Fluent, version 2020 R2, for flow analysis and flame calculation. Moreover, static ROM from *Ansys* Twin Builder, version 2021 R1, is applied to create the ROM and it takes a few minutes to train the model and seconds to apply it.

4.1 SVD Energy Analysis

To study the influence of varying the number of the ROM modes on the precision of the different models, Figs. 4 - 6 present the singular values, λ_n , from the SVD for the decoupled, coupled and the logarithm species ROM approach, respectively. The vertical axis displays the singular values normalized by the sum on a \log_{10} scale, and the horizontal axis, the number of modes (n) in a linear scale. The logarithm scale has been chosen to underscore the order of magnitude variations of the singular values computed.

In a general overview, all properties show similar behavior; i.e., the first SVD modes have higher singular values and, as the number of modes increases, the singular value decreases. Such feature comes by construction from the SVD decomposition, where the singular values are organized in decreasing order, such that that the information given by the first modes retains the main characteristics of the considered system.

The SVD energy analysis for the uncoupled reduced order models is presented at Fig. 4. For the sake of brevity, only five different flame properties are shown in such figure: OH, CH_2 and CO_2 mass fractions, temperature, and radial velocity. The temperature, radial velocity and CO_2 mass fraction have a steeper property decrease when compared to OH and CH_2 mass fractions. Such difference becomes even more evident when considering a horizontal line at $\lambda_n^s = 10^0$ and analyzing the mode number for which the singular value first lies below this line. Concerning the temperature, radial velocity and CO_2 five modes are required for the SVD to go below this line. Nevertheless, for OH and CH_2 this is only observed at the ninth and twelfth modes, respectively. This

underscores that learning these minor species behavior is harder than learning the velocity components or temperature.

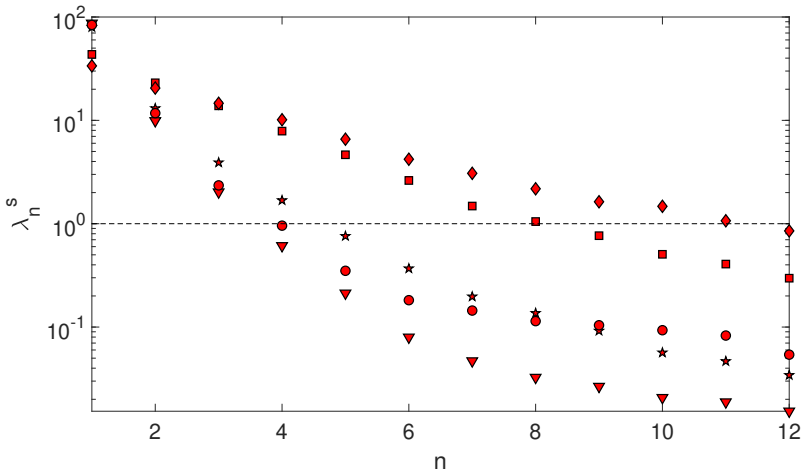


Fig. 4 Singular values (%) normalized by the sum (λ_n^s), as a function of the number of modes (n) for five properties of the uncoupled ROM. ●: v_y ; ▼: T ; ■: Y_{OH} ; ◆: Y_{CH_2} ; ★: Y_{CO_2} .

In order to evaluate the effect of the different coupling methodologies, the comparison of the energy related to the reduced order model, with the properties treated as uncoupled and coupled, is given in Fig. 5. More specifically, the results corresponding to three methodologies of the coupled properties is presented; one without normalization, the other two cases with normalization being applied, using Eq. (1): one having only the temperature and velocity normalized, but not the species, and another with all properties normalized.

In Fig. 5, the energy required to reconstruct the ROM using the coupled properties without a normalization is similar to the energy necessary to the uncoupled temperature ROM. This is explained by considering that the singular values are controlled by the highest absolute property value, which in this case is the temperature, of the order of 10^3 . Despite the normalization of the temperature and velocity, the energy content is still similar to that required to reconstruct the temperature, which might be explained by the remarkable difference between the orders of magnitude of the different species, being significantly smaller than one. Nevertheless, when considering the coupled case with all properties normalized, the decrease of the energy present in each mode is more similar to the uncoupled OH ROM, i.e., it is slower than the other coupled models. To better illustrate such phenomena, an horizontal line is drawn at $\lambda_n^s = 1$. Indeed, for both coupled ROM, without normalization and with normalization of temperature and velocity, the energy is smaller than 1 at the fourth mode. On the other hand, for the coupled ROM with all properties

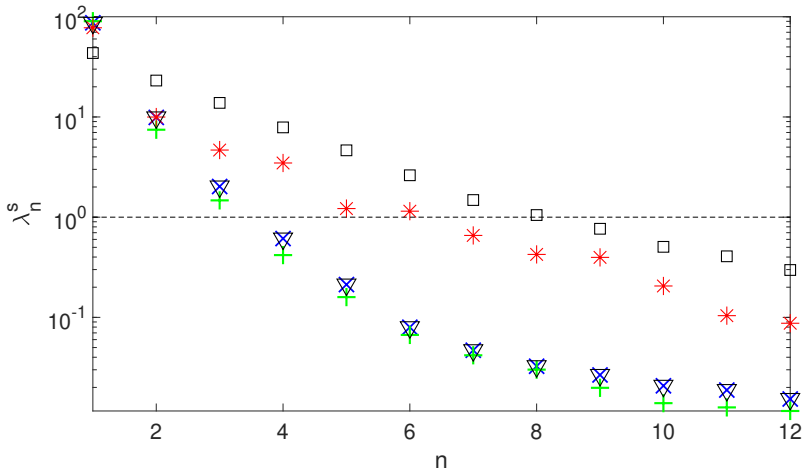


Fig. 5 Singular values (%) normalized by the sum (λ_n^s), as a function of the number of modes (n). \times : coupled without normalization; $+$: coupled with normalization of temperature and velocity; $*$: coupled with all properties normalized; ∇ : uncoupled temperature ROM; \square : uncoupled Y_{OH} ROM.

normalized, this occurs only in the seventh mode. Therefore, the ROM with coupled normalized properties has a learning process limited by the combustion scalars, i.e., the intermediate minor species, that are “harder” to learn. This suggests that the absence of normalization could be influencing the minor species learning, which would lead to incorrect mass fraction fields predictions, as will be explored in section 4.2.

Aiming to overcome the monotonicity problem seen in [28], the logarithm is applied to the learning data as given by Eq. (2). To study the influence of this method, the effect of truncation parameter is analyzed, for three values of ϵ_c , where ϵ_c is the negative exponent of the truncation parameter. The values of ϵ_c are arbitrarily chosen as 8, 10 and 12.

Accordingly, three reduced order models, using the logarithm technique, are constructed for each ϵ_c . Figure 6 presents the corresponding singular values normalized by the sum, as a function of the number of modes, for two species; one minor (CH_2) and one major species (CO_2).

A qualitatively similar behavior is noticed for those ROM shown in Fig. 6. Indeed, the singular values of the species for the uncoupled ROM are larger than those found in the ROM where the logarithm is applied, for almost every mode. Also, for the CH_2 , the singular values of the uncoupled ROM are at least an order of magnitude larger than the one corresponding to the logarithm ROM.

Regarding the minor species CH_2 , given in Fig. 6a, one may note a difference between the singular values starting from the second mode. Also, when $\epsilon_c = 12$, the decrease on the values is smaller when compared to $\epsilon_c = 10$ and 8, suggesting that this threshold process could be removing learning data from

this minor species field, which will be further characterized in section 4.2. On the contrary, for the major species, Fig. 6b, the singular values corresponding to the three ϵ_c chosen are remarkably similar.

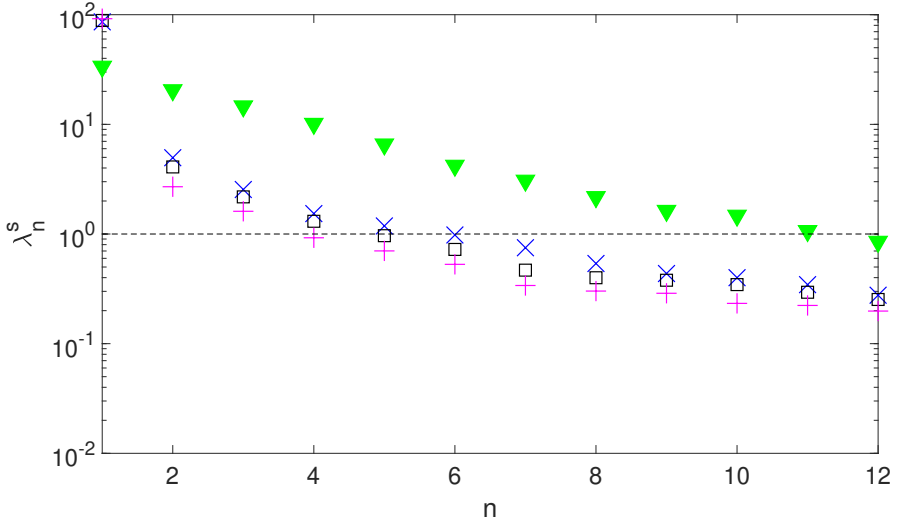
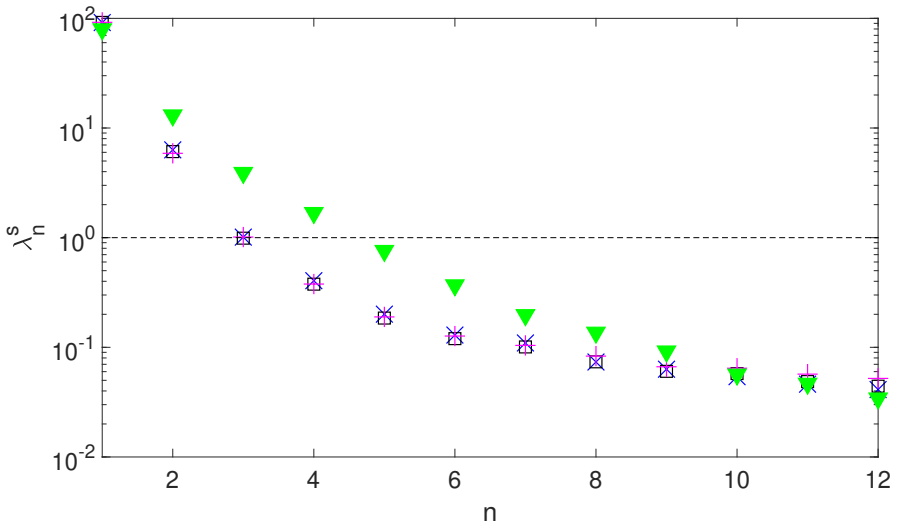
(a) Y_{CH_2} (b) Y_{CO_2}

Fig. 6 Singular values (%) normalized by the sum as a function of the number of modes, for the species mass fraction uncoupled ROM and the ROM using the logarithm approach, for three different truncation parameters. ϵ_c is the negative exponent of the truncation parameter chosen as the minimum value of the mass fraction range. ∇ : uncoupled ROM; +: $\epsilon_c = 8$; \square : $\epsilon_c = 10$; \times : $\epsilon_c = 12$.

Comparing the results between the minor and major species it is noted that learning the behavior of the major species requires a smaller number of modes than for the minor species, and consequently this facilitates learning the behavior of CO₂. This behavior can be exemplified by drawing a horizontal line at $\lambda_n^s = 10^0$, as done in the previous energy analysis. In the case of CH₂, the first ROM below this line is $\epsilon_c = 8$ at the fourth mode, followed by the ROM of ϵ_c equal to 10 and 12, in the respective fifth and sixth modes. For CO₂, the logarithm ROM are below the line in the third mode. Furthermore, the energy decay in each mode is steeper for the major species than for the minor species. The same behavior is found when comparing the species in uncoupled ROM. Again, this underscore that minor species are harder to learn than major species.

4.2 Properties Reconstruction Field Comparison

In the previous section, five reduced order models have been presented, with different preprocessing methodologies. Nevertheless, for the sake of brevity, only some of the reconstruction results will be presented here, i.e., the most remarkable ones. Indeed, the results of the coupled ROM without normalization and the one with temperature and velocity normalized present no significant differences when compared to the uncoupled ROM. In particular, the three models still present negative mass fraction regions [28]. Moreover, for the coupled ROM with all properties normalized, a smaller region of negative mass fraction and a degradation in the profiles of the chemical species reconstruction is seen, e.g., a spatial bifurcation present in some species such as OH. Such phenomena has been observed, in the description of chemical species, however it could be reduced by increasing the number of learning cases [28].

Figure 7 shows the mass field of three chemical species (OH, CH₂ and CO) obtained from CFD, for the uncoupled and logarithmic ROM reconstructions. Each column of the figure represents one of the three methods used to obtain the field, whereas each row represents a given species. Those species have been chosen to illustrate the impact of each model on the intermediate species, with different length scales. Moreover, for these species the maximum concentrations are around 10^{-3} for OH, 10^{-5} for CH₂, and 10^{-2} for CO.

The computational domain portion used to visualize the results corresponds to the air and fuel inlets velocity. The fuel inlet is limited by the burner wall, which has a thickness of 1 mm (represented by the white rectangle in the lower part of Fig. 7). Such 1 mm white rectangle provides a reference scale for the dimensions of all figures displaying property fields, facilitating the analysis. In the following, one may note that the reference to the air side or fuel side refer to the reactant inlets. More specifically, the air side is located in the lower left part of the displayed domain, whereas the fuel side lies at the lower right one part.

Regarding the CFD species profiles, at the first column of Fig. 7, the OH mass fractions, Fig. 7a, presents a maximum concentration of the species near the anchoring region, at the air-side burner wall, exhibiting a local maximum at

the symmetry axis. Also, the OH concentration along the stream wise direction decreases due to the progressive dilution by combustion products. The CH₂ mass fraction field, depicted at Fig. 7d, is characterized by a thin (0.8 mm at 1 mm after the fuel feeding tube outlet) and long profile, having its maximum near the anchoring region, at the air side. One may note that this species has the smallest time and length scales, being quickly produced and consumed, within the reactive region [23]. Accordingly, CH₂ is an effective indicator of the flame front position, since it has the smallest scale in comparison to all the combustion properties [23]. Lastly, the mass fraction profile of carbon monoxide is given in Fig. 7g. Unlike the other species produced from methane oxidation, CO shows a maximum concentration along the inner region of the flame, and not at the flame front. Indeed, the concentration of this species first increases along the axis of symmetry and, as it is oxidized, the concentration decreases.

Following, the uncoupled ROM results given at the second column of Fig. 7, present a negative mass fraction region, represented by the white color areas on the OH and CH₂ fields. Such negative mass fraction has no physical meaning, since the mass fraction are bounded between 0 and 1, thus these negative values being an artifact of ROM prediction. This negative mass fraction in the uncoupled ROM reconstruction is present on the predicted result of 15 out of 20 combustion species, with various orders of magnitude, which is not shown here for the sake of brevity. One may note that the result for OH, Fig. 7b, shows a smaller region of negative mass fraction than the CH₂ field (Fig. 7e), meaning that there is a higher difficulty linked to the prediction of the minor chemical species, such as, CH₂.

Concerning the carbon monoxide mass fraction field of the uncoupled ROM, Fig. 7h, one may note that the reconstructed field is slightly longer than the CFD one (Fig. 7g). Also, the ROM result predicts a region where the maximum mass fraction is 0.053, which is almost 4% higher than the CFD one, of 0.051. The fact that the ROM predicts a result that is outside in the CFD learning data range, underscores that the ROM methodology does not preserve neither the monotonicity nor the boundedness of the properties. Nevertheless, the monotonicity and the boundedness of the properties are indispensable for combustion studies, and must be preserved to apply the ROM in real industrial combustion scenarios. Despite these shortcomings, the species fields obtained by the uncoupled ROM are similar to those obtained by CFD.

Since the ROM results show the presence of a negative mass fraction, mass conservation might not be guaranteed on such models. Therefore, to confirm that the ROM methodology does not violate the mass conservation, the analysis of the residual of the sum of mass for each species i , i.e., $r_{Y_i} = 1 - \sum Y_i$, is conducted. The results show that the conservation of mass is not violated despite the negative mass fraction presented for some chemical species. Indeed, the maximum standard deviation is 10^{-7} , and it is randomly distributed.

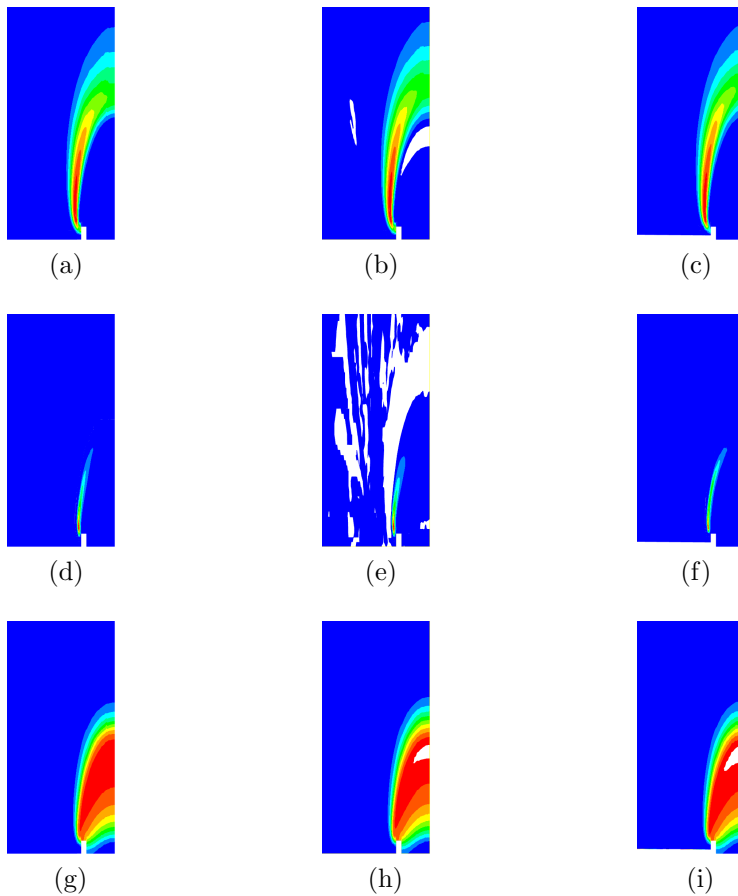


Fig. 7 Selected flame properties fields, in a $20 \times 43 \text{ mm}^2$ area of interest. The fuel inlet is located at the bottom right side, and at the left side is the air inlet. The sub-figures represent the OH (a, b, c), CH_2 (d, e, f) and CO (g, h, i) species mass fraction obtained with CFD (first column), uncoupled ROM (second column) and the one using the logarithm of the chemical species data as a learning data with $\epsilon_c = 10$ (third column). All sub-figures correspond to the validation case with prescribed fuel inlet velocity of 3.1 cm/s . The color map limits are $Y_{OH} \in [0, 3.9 \cdot 10^{-3}]$; $Y_{\text{CH}_2} \in [0, 2.8 \cdot 10^{-5}]$; $Y_{CO} \in [0, 5.1 \cdot 10^{-2}]$. The color map goes from blue (minimum) to red (maximum). The white areas correspond to nonphysical negative values.

The ROM results using the logarithm exhibits, third column of Fig. 7, significant differences in comparison to the other methods, used and are nearly indistinguishable from the CFD learning data. Indeed, none of chemical species predictions have a negative mass fraction region, corresponding to the physically expected behavior. In fact, the minimum mass fractions of all predicted properties are positive and greater than zero, on the order of magnitude of 10^{-11} . Also, some of the predicted fields using the logarithm preprocessing methodology are similar to the others ROM. It is worth to note that the CO mass fraction field, Fig. 7i, still presents a region of high mass fraction, with

a $Y_{CO} > 0.051$, similar to the predicted with the uncoupled Y_{CO} ROM. Concerning the continuity criteria, the maximum standard deviation increases to 10^{-5} , but the error (r_{Y_i}) still remains randomly spatially distributed.

To further characterize the effect of the proposed logarithm preprocessing on the ROM, Fig. 8 shows the mass fractions of OH along the burner symmetry axis. The origin of this axis lies at the fuel tube entrance, and the tube exit is at 60 mm. More specifically, Fig. 8a gives the mass fraction of OH along this axis for the CFD results, uncoupled ROM and logarithm ROM. One may note that the curve corresponding to the OH uncoupled ROM exhibits a non-monotonic behavior, between 70 and 80 mm, underscoring the previously discussed phenomena that is seen in Fig. 7. The OH mass fraction increases and then decreases to a negative nonphysical value, returning to a positive value near 80 mm. Regarding the ROM using the logarithm preprocessing approach, Fig. 8a shows that a distinction on the behavior between the models with different ϵ_c values and the CFD result is not observed until 80 mm. The CFD result has a maximum OH mass fraction slightly upstream, when compared to the ROM methods. Moreover, comparing the curves it is noticed that for CFD and $\epsilon_c = 8$ exhibit the highest values, and the peak of the curve of the uncoupled ROM the lowest. Nevertheless, the curves of ROM using the logarithm approach for $\epsilon_c = 10$ and 12 are quite similar to each other.

Regarding now the logarithm representation of the OH mass fraction, given at Fig. 8b, one can further observe the influence of the chosen truncation value, ϵ_c . As expected, a truncating $\epsilon_c = 8$ loses the most information, i.e., nearly all the information between the 60 and 65 mm. However, for $\epsilon_c = 12$, a close resemblance to the CFD behavior is seen. For $\epsilon_c = 10$, a smaller amount of information is lost in comparison to the $\epsilon_c = 8$ behavior in the same interval. Using this particular graphical representation, downstream 65 mm, no significant difference between the CFD results and the different ROM is seen. Furthermore, the logarithm representation enables to observe that, regarding the uncoupled ROM, an Y_{OH} increase occurs upstream, in comparison to the other results, and resemblance is seen downstream 80 mm only. Moreover, between 70 and 75 mm, an oscillation corresponding to the non-monotonic behavior is seen, as previously observed in Fig. 8a. It is worth to stress, though, that in the logarithm of OH mass fraction of the uncoupled ROM, the segment without the dashed line, between 75 - 80 mm, corresponds to the same location where the OH mass fraction is negative in Fig. 8a.

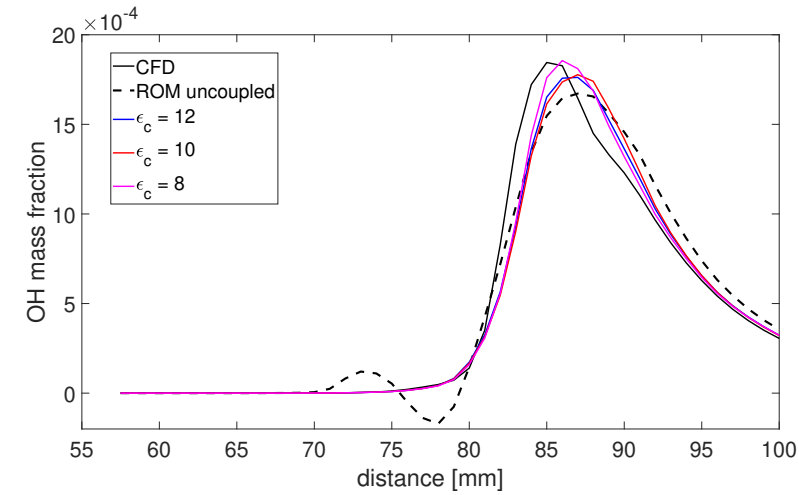
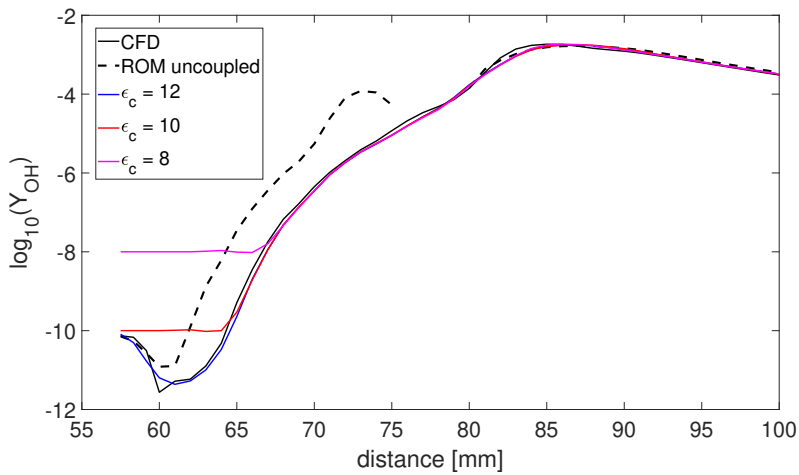
(a) Y_{OH} .(b) $\log_{10} Y_{OH}$.

Fig. 8 Mass fractions of OH and its logarithm along the symmetry axis for the CFD, the uncoupled and the logarithm ROM for different truncation parameters. The y axis is the species mass fraction and the x axis is the distance in the flow direction [mm], where 60 mm is the outlet of the fuel feeding tube.

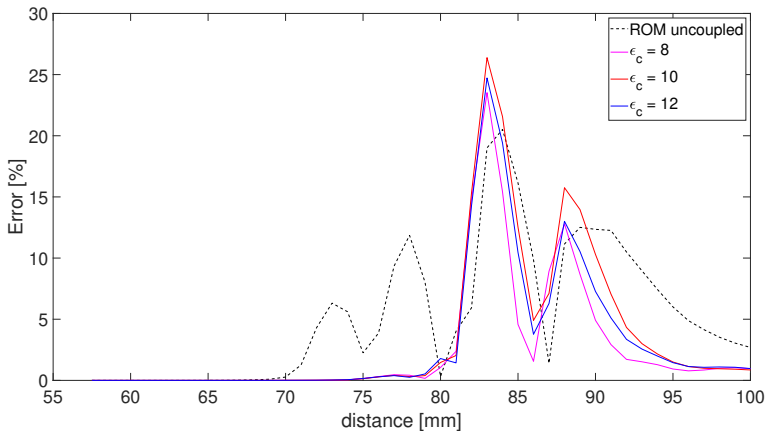


Fig. 9 Absolute relative error (%) of the different ROM predicted OH mass fraction (Y_{OH}) in relation to the CFD expected OH mass fraction, along the symmetry axis for different ROM.

Figure 9 shows the absolute relative error of the Y_{OH} ROM in comparison to the CFD results, along the symmetry axis shown in Fig. 8a. One may note that the uncoupled ROM (dotted curve) has a distributed error along the symmetry, in between 70 and 100 mm, whereas the other logarithm ROM present an error concentrate between 80 and 90 mm. Such difference between the ROM is linked to the non monotonicity previously viewed on Fig. 8, leaving evident that the uncoupled ROM presents a higher more spread, thus worse, error when compared to the other models. When comparing the error, Fig. 9, to the OH mass fraction, Fig. 8a, its seen that the error local maximums for the logarithm ROM correspond to the region where a delay between the ROM predicted mass fraction and the CFD one. More precisely, the ROM predicted Y_{OH} mass fraction (Fig. 8a) are displaced to the right when compared to the CFD one, thus corresponding to the elevated error at local regions. Moreover, it is evident that, despite either ROM models present a maximum error around 23%, the logarithm ROM corresponding to $\epsilon_c = 8$ presents the best prediction. Therefore, the application of the logarithmic a preprocessing method for combustion ROM increases the precision of such models when compared to the decoupled ones.

Nevertheless, proposed logarithm transformation thus increase the precision of the ROM of diffusion flames. The success of this strategy may be attributed to the exponential tails of the chemical species spatial distribution, which contains relevant information to be learned, and are thus better captured. Furthermore, this preprocessing step effectively spatially widens the fields of minor species, which also eases the learning. Finally, this preprocessing also should provide a more generalized re-scaling of species mass fractions and other properties that span over several orders of magnitude.

5 Conclusion

This work shows that there are remarkable impact of preprocessing the learning data when developing reduced order models of steady-state two-dimensional laminar methane air diffusion flames. To create the reduced model, high-fidelity CFD data have been used, which feature an adaptive mesh strategy based on the temperature gradient, and a skeletal chemical kinetics mechanism. In particular, the reduced order model was built from the fields of 23 computed flame properties that present different time and space scalars from a combustion system, and only the fuel inlet velocity varies.

The learning data set is composed of 12 simulated cases, and the validation data set is 8, which is a relatively small amount of data when compared to typical machine learning applications. The learning data set is then reduced using the singular value decomposition, leading to a reduced representation of the system. Subsequently, the machine learning algorithm GARS is applied to this reduced base to create a ROM model of the system and then predict the flow properties fields for arbitrary values of the inlet fuel velocity. For each of the modelings, the computational cost has a huge role. For instance, each CFD simulation required an average of 15 days to reach convergence, whereas the ROM requires a few seconds to be generated.

One may note that, here, the ROM model is directly dependent on the data coming from CFD simulation. Thus, there is also a cost linked to the generation of this data. Nevertheless, once the ROM model is created, its computational cost is considerably small compared to CFD. Accordingly, the use of such models enables to study of complex systems with precision and efficiency, as was shown in this work.

Here, a study on the impact of the data on the resultant ROM has been developed for different scenarios. Such scenarios consider the preprocessing type applied to the data. Specifically, a total of five preprocessing approaches have been studied, leading to the main conclusions listed below:

- For either the uncoupled or the logarithmic ROM, the learning of the major chemical species is easier than learning than minor species, i.e., they require less SVD modes to create the ROM when targeting the same SVD energy. Such difficulty is attributed to the low scale of the minor species when compared to the other flame properties.
- The coupling strategy applied for the ROM of the flame properties is not enough to account for all of the time and space combustion scales. Regardless of the coupling, the coupled ROM exhibits nonphysical non-monotonic behavior regarding the minor species fields and, in some cases, even negative mass fractions.
- The logarithm preprocessing is, on the scope of this study, the best solution to add prior knowledge of the combustion system to the ROM, and such methodology presented remarkable results when compared to the other methods. The logarithm ROM has been shown be the only method to ease the learning process of the ROM and guarantee the combustion properties

restrictions, i.e., the monotonicity, positivity and boundedness of the flame properties.

Moreover, the results of this work underscore the benefits of using the logarithm as a preprocessing of the combustion learning data. Indeed, both monotonicity and boundedness properties have been preserved, and also the learning phase has been achieved with a smaller number of modes when compared to the unprocessed data, thus easier than other methods. To conclude, the proposed logarithm transformation, thus, is sufficiently general to be considered as a learning strategy improvement in any combustion problem.

Acknowledgment

For the purpose of Open Access, a CC-BY public copyright licence has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission.

References

- [1] Gülder, Ö.L., Snelling, D.R.: Influence of nitrogen dilution and flame temperature on soot formation in diffusion flames. *Combustion and Flame* **92**(1), 115–124 (1993). [https://doi.org/10.1016/0010-2180\(93\)90202-E](https://doi.org/10.1016/0010-2180(93)90202-E)
- [2] Jerez, A., Cruz Villanueva, J.J., Figueira da Silva, L.F., Demarco, R., Fuentes, A.: Measurements and modeling of PAH soot precursors in coflow ethylene/air laminar diffusion flames. *Fuel* **236**, 452–460 (2019). <https://doi.org/10.1016/j.fuel.2018.09.047>
- [3] Escudero, F., Fuentes, A., Consalvi, J.-L., Liu, F., Demarco, R.: Unified behavior of soot production and radiative heat transfer in ethylene, propane and butane axisymmetric laminar diffusion flames at different oxygen indices. *Fuel* **183**, 668–679 (2016). <https://doi.org/10.1016/j.fuel.2016.06.126>
- [4] Liu, Y., Cheng, X., Qin, L., Wang, X., Yao, J., Wu, H.: Experimental investigation on soot formation characteristics of n-heptane/butanol isomers blends in laminar diffusion flames. *Energy* **211**, 118714 (2020). <https://doi.org/10.1016/j.energy.2020.118714>
- [5] Cheng, X., Li, Y., Xu, Y., Liu, Y., Wang, B.: Study of effects of ammonia addition on soot formation characteristics in n-heptane co-flow laminar diffusion flames. *Combustion and Flame*, 111683 (2021). <https://doi.org/10.1016/j.combustflame.2021.111683>
- [6] Brunton, S.L., Kutz, J.N.: *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press,

- Cambridge (2019). <https://doi.org/10.1017/9781108380690>
- [7] Brunton, S.L., Noack, B.R., Koumoutsakos, P.: Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics* **52**(1), 477–508 (2020). <https://doi.org/10.1146/annurev-fluid-010719-060214>
- [8] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**, 436–444 (2015). <https://doi.org/10.1038/nature14539>
- [9] Pyta, L., Abel, D.: Online model adaption of reduced order models for fluid flows. *IFAC-PapersOnLine* **50**(1), 11138–11143 (2017). <https://doi.org/10.1016/j.ifacol.2017.08.1006>
- [10] Xiao, X., Fang, F., Buchan, A.G., Pain, C.C., Navon, I.M., Muggerridge, A.: Non-intrusive reduced order modelling of the Navier-Stokes equations. *Computer Methods in Applied Mechanics and Engineering* **293**, 522–541 (2015). <https://doi.org/10.1016/j.cma.2015.05.015>
- [11] Angra, S., Ahuja, S.: Machine learning and its applications: A review. *Proceedings of the 2017 International Conference On Big Data Analytics and Computational Intelligence, ICBDACI 2017*, 57–60 (2017). <https://doi.org/10.1109/ICBDACI.2017.8070809>
- [12] Sun, H., Burton, H.V., Huang, H.: Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering* **33**, 101816 (2021). <https://doi.org/10.1016/J.JOBE.2020.101816>
- [13] Gogas, P., Papadimitriou, T.: Machine learning in economics and finance. *Computational Economics* 2021 57:1 **57**, 1–4 (2021). <https://doi.org/10.1007/S10614-021-10094-W>
- [14] Libbrecht, M.W., Noble, W.S.: Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16**, 321–332 (2015). <https://doi.org/10.1038/nrg3920>
- [15] Sharma, R., Kamble, S.S., Gunasekaran, A., Kumar, V., Kumar, A.: A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Computers & Operations Research* **119**, 104926 (2020). <https://doi.org/10.1016/J.COR.2020.104926>
- [16] Bikmukhametov, T., Jäschke, J.: Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Computers & Chemical Engineering* **138**, 106834 (2020). <https://doi.org/10.1016/J.COMPCHEMENG.2020.106834>

- [17] Rai, R., Sahu, C.K.: Driven by data or derived through physics? a review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus. *IEEE Access* **8**, 71050–71073 (2020). <https://doi.org/10.1109/ACCESS.2020.2987324>
- [18] Zhao, X., Shirvan, K., Salko, R.K., Guo, F.: On the prediction of critical heat flux using a physics-informed machine learning-aided framework. *Applied Thermal Engineering* **164**, 114540 (2020). <https://doi.org/10.1016/J.APPLTHERMALENG.2019.114540>
- [19] Kalogirou, S.A.: Artificial intelligence for the modeling and control of combustion processes: a review. *Progress in Energy and Combustion Science* **29**(6), 515–566 (2003). [https://doi.org/10.1016/S0360-1285\(03\)00058-3](https://doi.org/10.1016/S0360-1285(03)00058-3)
- [20] Chakravarthy, S.R., Rowan, S.L., Celik, I.B., Gutierrez, A.D., Escobar Vargas, J.: A reduced order model for the design of oxy-coal combustion systems. *Journal of Combustion* **2015**(943568), 1–9 (2015). <https://doi.org/10.1155/2015/943568>
- [21] Wang, Q., Hesthaven, J.S., Ray, D.: Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem. *Journal of Computational Physics* **384**, 289–307 (2019). <https://doi.org/10.1016/j.jcp.2019.01.031>
- [22] Aversano, G., Ferrarotti, M., Parente, A.: Digital twin of a combustion furnace operating in flameless conditions: reduced-order model development from CFD simulations. *Proceedings of the Combustion Institute* **38**(4), 5373–5381 (2021). <https://doi.org/10.1016/j.proci.2020.06.045>
- [23] Da Costa Ramos, L., Di Meglio, F., Figueira da Silva, L.F., Morgenthaler, V.: Reduced order model of laminar premixed inverted conical flames. In: *AIAA SciTech Forum*, Orlando, USA (2020). <https://doi.org/10.2514/6.2020-0416>
- [24] Alomar, A., Nicole, A., Sipp, D., Rialland, V., Vuillot, F.: Reduced-order model of a reacting, turbulent supersonic jet based on proper orthogonal decomposition. *Theoretical and Computational Fluid Dynamics* **34**, 49–77 (2020). <https://doi.org/10.1007/s00162-019-00513-y>
- [25] McQuarrie, S.A., Huang, C., Willcox, K.E.: 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 (2021). <https://doi.org/10.1080/03036758.2020.1863237>
- [26] Swischuk, R., Kramer, B., Huang, C., Willcox, K.: Learning physics-based reduced-order models for a single-injector combustion process. *AIAA Journal* **58**(6), 2658–2672 (2020). <https://doi.org/10.2514/1.J058943>

- [27] Law, C.K.: *Combustion Physics*, pp. 1–722. Cambridge University Press, Cambridge (2006). <https://doi.org/10.1017/CBO9780511754517>
- [28] Junqueira, N.L., Figueira da Silva, L., da Costa Ramos, L., de Paula, I.B.: The influence of the learning data on the reduced order model of laminar non-premixed flames. In: *Proceedings of the 26th ABCM International Congress of Mechanical Engineering*, online (2021). <https://doi.org/10.26678/abcm.cobem2021.cob2021-0110>. hal-03357849
- [29] Chi, C., Janiga, G., Thévenin, D.: On-the-fly artificial neural network for chemical kinetics in direct numerical simulations of premixed combustion. *Combustion and Flame* **226**, 467–477 (2021). <https://doi.org/10.1016/j.combustflame.2020.12.038>
- [30] Sharma, A.J., Johnson, R.F., Kessler, D.A., Moses, A.: Deep learning for scalable chemical kinetics. In: *AIAA Scitech Forum*, Orlando, Florida (2020). <https://doi.org/10.2514/6.2020-0181>
- [31] Zhao, F., Yang, W., Yu, W.: A progress review of practical soot modelling development in diesel engine combustion. *Journal of Traffic and Transportation Engineering (English Edition)* **7**(3), 269–281 (2020). <https://doi.org/10.1016/j.jtte.2020.04.002>. Special Issue: Clean Alternative Fuels for Transport Vehicles
- [32] Liu, F., Hua, Y., Wu, H., Lee, C.-f., He, X.: An experimental study on soot distribution characteristics of ethanol-gasoline blends in laminar diffusion flames. *Journal of the Energy Institute* **91**(6), 997–1008 (2018). <https://doi.org/10.1016/j.joei.2017.07.008>
- [33] Incropera, F.P., Dewitt, D.P., Bergman, T.L., Lavine, A.S.: *Fundamentals of Heat and Mass Transfer*, 6th edn., pp. 1–997. John Wiley & Son, ??? (2007)
- [34] Poinso, T., Veynante, D.: *Theoretical and Numerical Combustion*, 2nd edn., pp. 1–522. Edwards, ??? (2005). <https://doi.org/10.1201/9781482234213>
- [35] Turns, S.R.: *An Introduction to Combustion: Concepts and Applications*, 2nd edn., pp. 1–676. McGraw-Hill, ??? (2006)
- [36] Kazakov, A., Frenklach, M.: *Reduced Reaction Sets based on GRI-Mech 1.2*. The Combustion Laboratory at the University of California, Berkeley. <http://combustion.berkeley.edu/drm/>. Accessed in Nov 2020. (1984)
- [37] da Costa Ramos, L.: *Numerical study of an unstable premixed laminar flame and numerical Luenberger observers*. Thesis, Université Paris

- sciences et lettres (September 2021). <https://pastel.archives-ouvertes.fr/tel-03417236>
- [38] da Costa Ramos, L., da Silva, L.F.F., Meglio, F.D., Morgenthaler, V.: Modelling of pulsating inverted conical flames: a numerical instability analysis. *Combustion Theory and Modelling* **26**(2), 260–288 (2022). <https://doi.org/10.1080/13647830.2021.2011961>
- [39] Young, T.R., Boris, J.P.: A numerical technique for solving stiff ordinary differential equations associated with the chemical kinetics of reactive-flow problems. *The Journal of Physical Chemistry* **81**, 2424–2427 (1977)
- [40] ANSYS: Ansys Fluent 12.0 Theory Guide. ANSYS. <https://www.afs.enea.it/project/neptunius/docs/fluent/html/th/node1.htm>. Accessed in 10 Oct 2021. (2009)
- [41] García, A.M., Rendon, M.A., Amell, A.A.: Combustion model evaluation in a CFD simulation of a radiant-tube burner. *Fuel* **276**(25), 118013 (2020). <https://doi.org/10.1016/j.fuel.2020.118013>
- [42] Hiremath, V., Ren, Z., Pope, S.B.: Combined dimension reduction and tabulation strategy using ISAT–RCCE–GALI for the efficient implementation of combustion chemistry. *Combustion and Flame* **158**(11), 2113–2127 (2011). <https://doi.org/10.1016/j.combustflame.2011.04.010>
- [43] Cunha Jr, A., Figueira da Silva, L.F.: Assessment of a transient homogeneous reactor through in situ adaptive tabulation. *J Braz. Soc. Mech. Sci. Eng.* **36**, 377–391 (2014). <https://doi.org/10.1007/s40430-013-0080-4>
- [44] Cunha Jr, A., Figueira da Silva, L.F.: Crflowlib — chemically reacting flow library. *Software Impacts* **11**, 100206 (2021). <https://doi.org/10.1016/j.simpa.2021.100206>
- [45] Celis, C., Figueira da Silva, L.F.: Computational assessment of methane-air reduced chemical kinetic mechanisms for soot production studies. *J Braz. Soc. Mech. Sci. Eng.* **36**, 2225–2244 (2016). <https://doi.org/10.1007/s40430-016-0494-x>
- [46] Marrocu, M., Ambrosi, D.: Mesh adaptation strategies for shallow water flow. *International Journal for Numerical Methods in Fluids* **31**, 497–512 (1999). [https://doi.org/10.1002/\(SICI\)1097-0363\(19990930\)31:2](https://doi.org/10.1002/(SICI)1097-0363(19990930)31:2)
- [47] Kallinderis, Y., Vijayan, P.: Adaptive refinement-coarsening scheme for three-dimensional unstructured meshes. *AIAA Journal* **31**(8), 1440–1447 (1993). <https://doi.org/10.2514/3.11793>
- [48] Walter, M.A.T., Abdu, A.A.Q., Figueira da Silva, L.F., Azevedo, J.L.F.:

- Evaluation of adaptive mesh refinement and coarsening for the computation of compressible flows on unstructured meshes. *Int. J. Numer. Meth. Fluids* **49**, 999–1014 (2005). <https://doi.org/10.1002/fld.1037>
- [49] Luboz, V., Bailet, M., Grivot, C.B., Rochette, M., Diot, B., Bucki, M., Payan, Y.: Personalized modeling for real-time pressure ulcer prevention in sitting posture. *Journal of Tissue Viability* **31**, 54–58 (2018). <https://doi.org/10.1016/j.jtv.2017.06.002>
- [50] Trefethen, D. L. N.; Bau: *Numerical Linear Algebra*, 1st edn. Society for Industrial and Applied Mathematics, USA (1997)
- [51] Viana, F.A.C., Haftka, R.T., Steffen, V.: Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. *Structural and Multidisciplinary Optimization* **39**, 439–457 (2009). <https://doi.org/10.1007/s00158-008-0338-0>
- [52] Ben Salem, M., Roustant, O., Gamboa, F., Tomaso, L.: Universal prediction distribution for surrogate models. *SIAM/ASA Journal on Uncertainty Quantification* **5**, 1086–1109 (2017). <https://doi.org/10.1137/15M1053529>
- [53] Wang, S., Jian, G., Xiao, J., Wen, J., Zhang, Z.: Optimization investigation on configuration parameters of spiral-wound heat exchanger using genetic aggregation response surface and multi-objective genetic algorithm. *Applied Thermal Engineering* **119**, 603–609 (2017). <https://doi.org/10.1016/j.applthermaleng.2017.03.100>
- [54] Ostertagová, E.: Modelling using polynomial regression. *Procedia Engineering* **48**, 500–506 (2012). <https://doi.org/10.1016/j.proeng.2012.09.545>
- [55] Aversano, G., D’Alessio, G., Coussement, A., Contino, F., Parente, A.: Combination of polynomial chaos and kriging for reduced-order model of reacting flow applications. *Results in Engineering* **10**, 100223 (2021). <https://doi.org/10.1016/j.rineng.2021.100223>
- [56] Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. *Statistics and Computing* **14**, 199–222 (2004). <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- [57] Lancaster, P., Salkauskas, K.: Surfaces generated by moving least squares methods. *Mathematics of Computation* **37**(155), 141–158 (1981). <https://doi.org/10.1090/S0025-5718-1981-0616367-1>
- [58] Cano, J.-R., Gutiérrez, P.A., Krawczyk, B., Woźniak, M., García, S.:

Monotonic classification: An overview on algorithms, performance measures and data sets. *Neurocomputing* **341**, 168–182 (2019). <https://doi.org/10.1016/j.neucom.2019.02.024>