Model identification in reactor-based combustion closures using sparse symbolic regression

Model identification in reactor-based combustion closures using sparse symbolic regression

Rodolfo S. M. Freitas[a,b,*], Arthur Péquin[a,b], Riccardo Malpica Galassi[a,b], Antonio Attili[c], Alessandro Parente[a,b,*]

[a] *Université Libre de Bruxelles, École Polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Brussels, Belgium*
[b] *Université Libre de Bruxelles and Vrije Universiteit Brussel, Brussels Institute for Thermal-Fluid Systems and Clean Energy (BRITE), Brussels, Belgium*
[c] *University of Edinburgh, School of Engineering, Institute of Multiscale Thermofluids, The King's Buildings, Mayfield Road, Edinburgh, EH9 3FD, United Kingdom*

Abstract

In Large Eddy Simulations (LES) of combustion, the accuracy of predictions might be heavily affected by deficiencies in traditional/simplified closure models, especially when employed to simulate non-conventional fuels and combustion regimes. The increasing availability of data from experiments and higher-fidelity numerical simulations offers attractive opportunities for improving combustion models with data-driven techniques. In this work, we focus on sub-grid turbulence-chemistry interactions with the Partially Stirred Reactor (PaSR) model and its associated cell reacting fraction sub-model. We combine machine learning and sparsity-promoting techniques to improve the predictive capabilities of PaSR by discovering new functional forms of the cell reacting fraction sub-model from data. The obtained models are parsimonious models that balance accuracy with model complexity to avoid over-fitting. We employ the proposed model identification approach on data from a Direct Numerical Simulation (DNS) of a three-dimensional non-premixed n-heptane/air jet flame. As a result, we single out the most plausible model form of the cell reacting fraction, expressed as a function of

[*] *Corresponding authors at*: Aero-Thermo-Mechanics Laboratory, École Polytechnique de Bruxelles, Université Libre de Bruxelles, Brussels, Belgium.

Email addresses: `rodolfo.da.silva.machado.de.freit@ulb.be` (Rodolfo S. M. Freitas), `Alessandro.Parente@ulb.be` (Alessandro Parente)

the local Damköhler number. Then, the capability of the model to generalize properly to new, previously unseen data is tested. The results demonstrate the ability of the machine learning approaches to infer robust corrections for turbulence-chemistry reactor-based combustion models.

Keywords: Machine learning, Sparse regression, Turbulent combustion modeling, Direct Numerical Simulation (DNS), Partially Stirred Reactor (PaSR)

---

1. Introduction

Computer models of practical combustion systems must combine the capability to predict the relevant flow features with the ability to solve the chemistry involved in combustion processes, as well as any other important physical phenomena [1]. The Reynolds-averaged Navier-Stokes (RANS) approach is still the workhorse for industrial flow modeling, given the large domain sizes and the relatively low computational effort, which enables outer-loop applications such as optimization. On the other hand, Large Eddy Simulation (LES) is gaining momentum thanks to the increasing availability of computational resources and the improved turbulence representation that allows the study of unsteady phenomena such as local extinction and re-ignition [2, 3]. Predictive inaccuracies are present in both RANS and LES due to deficiencies in the closure models. This is the case of subgrid-scale (SGS) turbulence-chemistry interactions (TCI) closures for LES, wherein the filtered reaction rates cannot be directly computed from the filtered thermophysical quantities, due to non-linearity [4, 5].

Following the vast development of LES in the last decade, reactor-based models [6, 7] have attracted interest as TCI closures since they can deal with finite-rate chemistry at an affordable computational expense [8, 9, 10]. In such models, the computational cell is partitioned into two distinct regions: a reacting zone called *fine structures*, where all the chemical reactions take place, and a non-reacting zone defined as *surroundings*, solely driven by turbulent mixing. Under this assumption, the chemical source terms are calculated from a mass balance between the two regions. Such model formulation needs to be closed by determining the fraction of the cell occupied by the fine structures. Chomiak [7] proposed the Partially Stirred Reactor (PaSR) model, in which the cell reacting fraction is computed as a function of the local mixing and chemical characteristic timescales. The PaSR

model has been tested for various types of flames [11, 12, 13, 14], sprays [15], and non-conventional combustion regimes [16, 17, 18, 19]. Alternative formulations of the PaSR model have been investigated in RANS and LES simulations [20, 21, 22], and the effect of the mixing and chemical timescales estimation on the reacting fraction has been assessed in [18].

The construction of turbulent combustion closures might lead to model deficiencies. This can be related to incorrect parameter estimation, impacted by strong fluctuations of the thermo-physical properties and indirect measurements, or they can result from employing simplified physics or chemistry (model errors) [23, 24, 25]. One method for the development and validation of combustion models involves an *a priori* assessment, where the SGS modeled quantities of interest are compared with the corresponding filtered DNS [26]. Minamoto and Swaminathan [27] performed an *a priori* analysis on a three-dimensional dataset to develop and validate a representative canonical reactor to model the filtered reaction rates in Moderate and Intense Low-oxygen Dilution (MILD) combustion. Recently, Péquin et al. [12] performed an *a priori* assessment of the PaSR combustion model in LES and pointed out potential limitations of the model associated with the definition of the cell reacting fraction.

An approach for developing and improving the predictive ability of SGS closures in turbulent reacting flows is via data-driven approaches [28]. The use of data from experiments and simulations to improve the understanding and modeling capabilities of reacting flows has become a new challenge and research opportunity [29, 30, 31]. Several works have focused on the construction of predictive data-driven machine learning (ML) models able to return accurate predictions at a low cost [32, 33, 34, 35]. Furthermore, data-driven machine learning has been demonstrated to be a reliable tool to build constitutive relations of material properties [36, 37]. Also, ML models can be used to characterize the error inherent to the use of low-fidelity physics-based models. More specifically, neural network architectures can be constructed to characterize the discrepancy term due to the use of limited physics and be embedded in the computational model [38], allowing improved predictions and uncertainty quantification of low-fidelity models [23, 39].

Further, the advances of ML models for fluid dynamics were reviewed by Brunton et al. [40]. They highlight the capabilities of physics-aware machine learning to improve the simulation of fluids. Nakazawa et al. [41] explored the possibility and limitation of a machine learning-based approach for turbulent combustion modeling by applying physics-guided deep neural networks

(DNNs) for the prediction of the mean reaction rates for chemical species. The capability of DNNs to predict joint filtered density functions of mixture fraction and progress variable from DNS of non-conventional combustion regime were shown in an *a priori* study by Chen et al. [42], in which the results from the ML tool outperformed algebraic models at larger filter widths. Jigjid et al. [43] proposed a machine learning aided model derived from a PaSR approach to close the filtered reaction rate. A combination of neural networks, with a local combustion mode prediction, was used to predict the fraction of the reactive structure for both premixed and non-premixed conditions. Convolutional neural networks (CNNs) were also applied for modeling closures for turbulence combustion [44, 45] since convolutional architectures enable learning underlying physics from spatial data. Going further, physics-guided generative adversarial networks (GANs) were shown to perform typically better when compared with traditional CNNs for sub-filter modeling in turbulent reactive flows [46]. Despite the deep learning tools have been widely used in turbulence modeling, these methods might suffer from a lack of interpretability, which limits the understanding of the underlying physics of the resulting models and also hampers physical insights to derive new modeling ideas. In the current context, interpretability means the ability to correctly and efficiently understand the ML model and present the underlying basis behind predictions and decision-making in a way that is physically explainable.

Recently, machine learning and sparse-promoting techniques have been combined to discover parsimonious models that balance accuracy with model complexity [47]. In this context, Schmelzer et al. [48] introduced a novel sparse symbolic regression approach to discover algebraic stress models for the closures of RANS. Also, a general framework that combines symbolic regression with graph neural networks (GNNs) capable of providing plausible analytical expressions was introduced [49], being able to be used as a practical tool to discover analytical equations for fluid properties [50, 51]. More recently, Chung et al. [52] employed a random forest regressor with sparse symbolic regression to enable the discovery of analytical models for SGS terms of a turbulent transcritical flame.

The present work aims at improving the prediction capabilities of the reactor-based TCI models by leveraging sparse symbolic regression. More specifically, the present study has the following objectives: (i) to discover novel model forms of the cell reacting fraction using sparse-promoting techniques; (ii) to utilize the interpretable machine learning models to get physi-

cal insights and develop novel turbulence-chemistry interactions approaches; (iii) to compare different cell reacting fraction formulations and quantify their impact on the reactor-based combustion model predictions.

The remainder of this paper is organized as follows. The reactor-based combustion models and data-driven models employed in the present work are described in Sec. 2. Details regarding the DNS configuration are presented in Sec. 3. In Section 4, results of an *a priori* assessment are presented. The concluding section includes a summary of our main findings and future remarks.

2. TCI closures and data-driven model

2.1. The Partially Stirred Reactor (PaSR) model

In the Partially Stirred Reactor model, each computational cell is partitioned into two locally uniform regions: a non-reacting zone, solely driven by turbulent mixing and referred to as the *surroundings*, and a reacting part called the *fine structures* [7, 12]. The mean cell value is then defined as a weighted sum from the two regions,

$$\bar{Y}_i = \gamma Y_i^* + (1-\gamma) Y_i^0, \tag{1}$$

where γ (or κ in other works [7, 12]) represents the cell volume fraction available for chemical processes, Y_i^* and Y_i^0 are the i-th species mass fractions in the fine structures and in the surroundings, respectively [7]. Combustion takes place in the fine structures where the fuel and oxidizer are assumed perfectly mixed. The total time associated with this sequential process is estimated as the sum of the mixing timescale τ_m and the chemical timescale τ_c. The fraction of the cell occupied by the fine structures γ is modeled as the timescale ratio between the chemical reaction time and the total combustion time [7],

$$\gamma = \frac{\tau_c}{\tau_c + \tau_m}. \tag{2}$$

One can define the local Damköhler number $D_a = \tau_m/\tau_c$ as the ratio between the mixing and the chemical timescales and rewrite Eq. (2),

$$\gamma = \frac{1}{1 + D_a}. \tag{3}$$

From Eq. (1), the filtered chemical reaction rates for the i-th species can be defined as: $\bar{\dot{\omega}}_i = \gamma\dot{\omega}_i^* + (1-\gamma)\dot{\omega}_i^0$, where $\dot{\omega}_i^*$ and $\dot{\omega}_i^0$ are the chemical source terms from the fine structures and the surroundings, respectively. Assuming that the contribution of the *surroundings* to the chemical processes is negligible [53], i.e., $\dot{\omega}_i^0 \approx 0$, the filtered source terms, to be employed in the filtered species transport equations, are expressed as,

$$\bar{\dot{\omega}}_i = \gamma\dot{\omega}_i^* = \gamma\frac{\bar{\rho}(Y_i^* - \tilde{Y}_i)}{\tau^*}, \tag{4}$$

where $\bar{\rho}$ is the Reynolds-averaged density, τ^* represents the residence time in the fine structures. In the original model [15], the residence time in the PaSR model was estimated using the mixing timescale only, but other works [20, 18, 12, 22] demonstrated improved model prediction capabilities when τ^* equals the minimum between the mixing and the chemical timescales. Therefore, the present work follows the latter definition of the reactor residence time, defined as: $\tau^* = \min(\tau_m, \tau_c)$.

The estimate of Y_i^* results from the time integration of the fine structures that are treated as ideal reactors. Both Perfectly Stirred Reactor (PSR) and Plug-Flow Reactor (PFR) formulations have been investigated for this task [15, 10]. The PFR allows for alleviated computational costs without loss of accuracy to simulate the fine structures and is often preferred over the PSR [54, 55]. In the present work, we use a PFR evolving from Y_i^0, which can be approximated by the filtered mass fraction $\tilde{Y}_i$ [15], over τ^*. This leads to the following set of mass fraction equations,

$$\frac{dY_i^*}{dt} = \frac{\dot{\omega}_i}{\bar{\rho}} \tag{5}$$

where the term $\dot{\omega}_i$ is the instantaneous reaction rate of the i-th species. The integration of Eq. (5) over the residence time τ^* results in the estimation of the final state[1] of the sub-grid quantity Y_i^*, used in Eq. (4). As far as the mixing timescale is concerned, the *scalar mixing timescale*, which has demonstrated its effectiveness among other formulations [56, 18], is computed as

[1]Note that, between each simulation time-step, the temporal evolution of the fine structures towards local chemical equilibrium is modelled through the time integration of ideal reactors at a rate given by chemical reactions following Arrhenius laws. It is to be noted that the model part that deals with the evolution of the chemical state remains unaltered in this study, which instead focuses on the definition of the reacting portion of the cell.

the ratio of the mixture fraction variance to the mixture fraction dissipation rate $\tau_m = \widetilde{Z''^2}/\tilde{\chi}$. For the chemical timescale, Chomiak [7] proposed an estimation based on the formation rates of the species acting as fuel and oxidizer,

$$\frac{1}{\tau_c^{\text{Ch.}}} = \max\left(\frac{-\dot{\omega}_F}{Y_F}, \frac{-\dot{\omega}_O}{Y_O}\right)/\rho, \tag{6}$$

where $\dot{\omega}$ is the global conversion rate, Y is the species mass fraction, and the subscripts F and O stand for fuel and oxidizer respectively. Chomiak's formulation extracts the scales associated with the limiting reactant, i.e., the chemical species which is consumed first.

2.2. Quasi-Laminar Finite Rate model (QLFR)

Acting as a simplified PaSR model, the Quasi-Laminar Finite Rate model [12] enforces the assumption that the fine structures occupy the total volume of each computation cell, i.e., $\gamma = 1$. This corresponds to a mixing timescale which is much faster than the chemical timescale, and to a cell that can be entirely treated as an ideal reactor, see Eq. (5), of residence time $\tau^* = \tau_m$ in the present case. The QLFR filtered source term is then rewritten from Eq. (4) as follows:

$$\bar{\dot{\omega}}_i = \frac{\bar{\rho}(Y_i^* - \tilde{Y}_i)}{\tau^*}. \tag{7}$$

The QLFR approach is to be used in restricted conditions [12], for relatively fine LES grids before losing accuracy at coarser grids, and generally is prone to overestimating chemical reaction rates. Employing the QLFR model allows for quantifying the importance of modeling the cell reacting fraction in reactor-based subgrid-scale models. In addition, both the QLFR and PaSR models allow for contrasting the potential improvements obtained from the data-driven combustion models.

2.3. Sparse regression for TCI model identification

Sparse symbolic regression, which will be indicated as SpaR hereafter, exploits the fact that in most physics-based models, only a few nonlinear terms are needed to characterize the system, resulting in sparse models in a high-dimensional nonlinear feature space [47]. The main advantage of the sparse identification lies in the fact that the model form is parsimonious, i.e., the

model intrinsically balances complexity with accuracy, thus avoiding data overfitting [57]. In particular, the complexity is measured by the number of terms in the model to accurately represent the data. Moreover, since it is assumed that the underlying structure of the models has only a few important terms, so avoiding the discovery of overfitted models [58]. Finally, this method involves the choice of predetermined symbolic candidate functions, among which the most plausible in describing the data will be identified, resulting in interpretable models.

To identify a model $\mathcal{M} : \mathbb{R} \to \mathbb{R}$; $x \longmapsto \mathcal{M}(x)$ given a dataset of n observations $\{x_i, y_i\}_{i=1}^{n}$, we consider that a model that relates input-output data generically can be expressed as a linear combination of p regressor variables $\mathbf{X} \in \mathbb{R}^{n \times p}$

$$\mathbf{Y} = \mathcal{M}(\mathbf{X}) + \epsilon = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{8}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients, $\mathbf{X}$ is the matrix of the column vectors $\mathbf{X}_j \in \mathbb{R}^n$, $j = 1, \dots p$, and $\epsilon \in \mathbb{R}^n$ is the discrepancy between the model predicted quantities and the observations $\mathbf{Y} \in \mathbb{R}^n$.

From a sub-grid models perspective, the closure models are built upon the combination of nonlinear functions of the independent quantities x. That allows replacing the regressors $\mathbf{X}_j$ with candidate non-linear functions $\boldsymbol{\Theta}(\mathbf{X}_j)$ of the original variables, where each column of $\boldsymbol{\Theta}(\mathbf{X})$ represents a candidate function of the model:

$$\mathcal{M} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\beta}. \tag{9}$$

Here, there is no restriction in the choice of the candidate non-linear functions since only a few of the non-linear functions will be deemed important to characterize the model form. For instance, the non-linear functions may consist of constant, d-order polynomial, and predefined functions:

$$\boldsymbol{\Theta}(\mathbf{X}) = \begin{bmatrix} | & | & | & & | & | & | & \\ \mathbf{1} & \mathbf{X} & \mathbf{X}^2 & \cdots & \mathbf{X}^d & \text{erf}(\mathbf{X}) & \log_b(\mathbf{X}) & \cdots \\ | & | & | & & | & | & | & \end{bmatrix}. \tag{10}$$

However, the choice of the candidate functions can be problem dependent and has the potential to embed prior knowledge. Indeed, $\boldsymbol{\Theta}(\mathbf{X})$ can be constructed using physics-based analyses, aiming at providing a mathematical ansatz with the underlying physics of the system before identification.

Since we expect the model to be sparse in the high-dimensional nonlinear feature space, we will regress the sparse vector of model coefficients $\boldsymbol{\beta}$ by setting up a sparse regression problem, such as the least absolute shrinkage and selection operator (*lasso*) method [59], which is an l_1-regularized regression that encourages sparsity. Therefore, the aim of the *lasso* regression is to solve

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathcal{M} - \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1, \tag{11}$$

where λ is a regularization parameter that prescribes the amount of shrinkage (or sparsity constraint) in the model coefficients, which will result in a small number of non-zero predictors in the final regressed model. Hence, the *lasso* approach helps in reducing the model complexity and the multi-colinearity.

In this work, the model to be regressed is the cell reacting fraction submodel γ, as a non-linear function of the local filtered thermo-physical quantities. Figure 1 shows an overview of the SpaR approach for model discovery of the cell reacting fraction γ for the PaSR model. The main goal of SpaR is to combine functions from a predefined library of candidates to provide an ensemble of identified models for the cell volume fraction, i.e., $\mathcal{M} = \gamma$. The main steps consist in:

1. building a library of candidate functions $\boldsymbol{\Theta}(\mathbf{X})$, for which the independent variables are the filtered thermo-physical quantities, $\mathbf{X} = \bar{\boldsymbol{\Phi}}$;
2. identifying the model using lasso regression, where the coefficients β are regressed by minimizing the discrepancy between the filtered DNS source terms $\bar{\dot{\omega}}^{DNS}$ and the PaSR modeled source terms $\dot{\omega}^{PaSR}(\boldsymbol{\Theta}(\boldsymbol{\Phi})\beta)$;
3. learning the model coefficients $\hat{\boldsymbol{\beta}}$.

Furthermore, the model identification approach is not expensive in terms of computational cost, allowing model identification at a moderate price even for high-dimensional problems. Such an approach has the advantage to be able to use efficient optimization algorithms developed for parameter estimation in machine learning, which rapidly converge to the sparse solution with a small number of iterations [48].

2.3.1. Choosing plausible candidate functions for γ

The sparse symbolic regression methodology requires the specification of the candidate functions whose - preferably sparse - linear combination identifies the parsimonious model. Hence, constructing a library of functions is

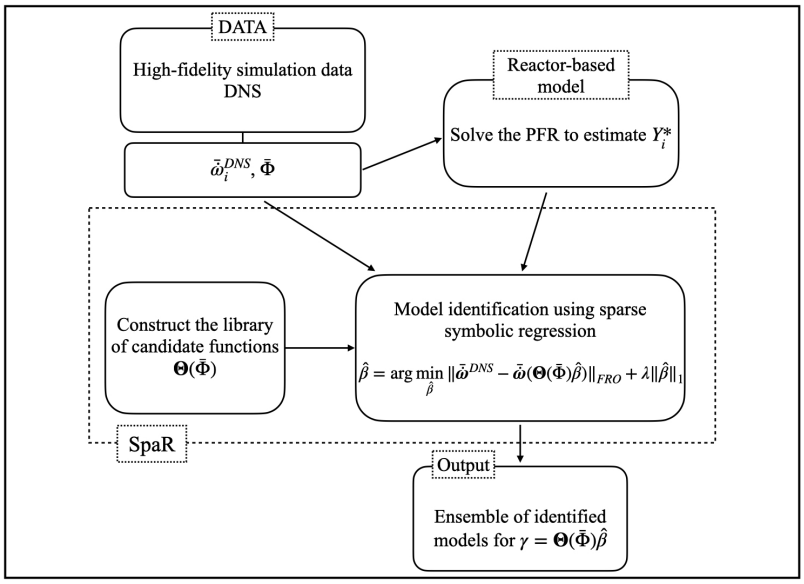


Figure 1: Overview of the sparse symbolic regression (SpaR) approach for TCI model identification.

an essential element of the entire framework, and must accommodate relevant candidates that facilitate a physical explanation of the data. However, in many complex scenarios, embedding prior knowledge may be challenging, and the library of candidate functions can be overparameterized. This can be mitigated by using key features that explain the data from which a model is identified through a linear combination of the candidates. The extraction of the key features describing the data can be achieved by correlation or sensitivity analysis to reduce the search space dimensionality [60].

The approach pursued in the present work relies on the PaSR model and we aim to find models for γ in Eq. (4) given the local Damköhler number $Da = \tau_m/\tau_c$ as an independent variable, following the genesis of Eq. (3). Indeed, we are inspired by the work proposed by Magnussen [61]. Specifically, it is assumed that the fine structures are vortex-sheet or vortex tubes of random extension folded in the flow and that chemical reactions take place where the species are mixed at the molecular scale in the fine structure regions. In order to improve the treatment of the chemical reactions within these regions, Magnussen proposed modifications to the cell reacting fraction by introducing different positive exponents, i.e., γ^n, based on the assumptions that the exchange between the fine structures and the surroundings is

a surface or volumetric process. In the present work, we propose a set of candidate functions by introducing different exponents to the cell reacting fraction of the PaSR model. The resulting library is therefore

$$\gamma(Da) := \begin{cases} \dfrac{1}{(1+Da)^n} \\ \dfrac{1}{1+Da^n} \qquad n = \dfrac{1}{5}, \dfrac{1}{4}, \dfrac{1}{3}, \dfrac{1}{2}, \dfrac{2}{3}, 1, \dfrac{3}{2}, 2, 3, 4, 5. \\ \dfrac{1}{Da} \end{cases} \tag{12}$$

In case of infinitely fast chemistry, i.e., $\tau_c << \tau_m$, the model form $1/Da$ is a fair approximation of $1/(1+Da)$. This hypothesis justifies the presence of the $1/Da$ function within the candidates at high Damköhler numbers. It is to be noted that such functions will greatly differ from the original formulations at small Damköhler numbers by exhibiting a hyperbolic behavior, not ensuring the natural upper bound of unity. In the approach being pursued here, n is assumed to be a positive rational exponent.

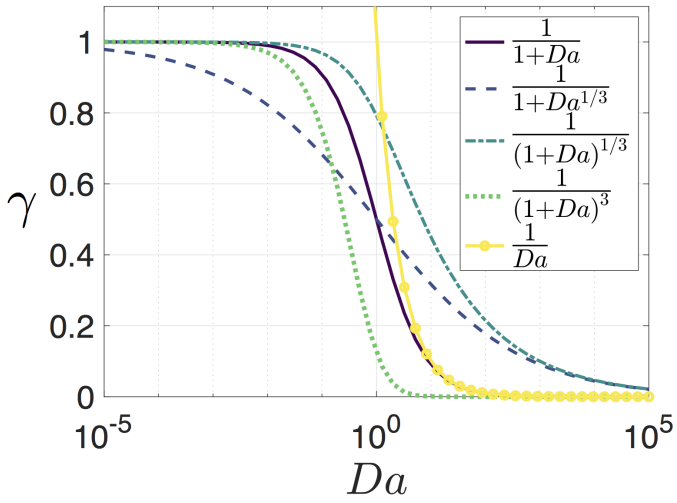


Figure 2: Graphical representation of sparse candidate functions for the PaSR cell reacting fraction as a function of the Damköhler number.

A selection of functions is presented in Fig. 2. While the first four functions preserve the original bounds $\gamma \in [0, 1]$, the last form, i.e., $1/x$-like

function, diverges at small Damköhler numbers. In addition, the formulation $1/(1+Da^n)$ yields 0.5 at the condition $\tau_c = \tau_{mix}$, as the original γ formulation, and so for any n. For the specific case of the functional forms $1/(1+Da)^n$, the condition $\tau_c = \tau_{mix}$ may yield to γ values below or above the reference value 0.5 depending on n. As shown in Fig. 2, an above-unity exponent yields crossing values below 0.5. Conversely, higher than 0.5 values are obtained when considering $n < 1$. In addition, we add as candidates the nonlinear functions $1-\sigma(Da)$ where σ is the sigmoid function, and the complementary error function ($erfc = 1 - erf(Da)$) [62]. This partially reduces the bias that we induced by tailoring the functional search space using prior knowledge. Furthermore, this allows us to show the robustness of the proposed approach to correctly identify the functional form of the cell reacting fraction with an overparameterized library of candidate functions.

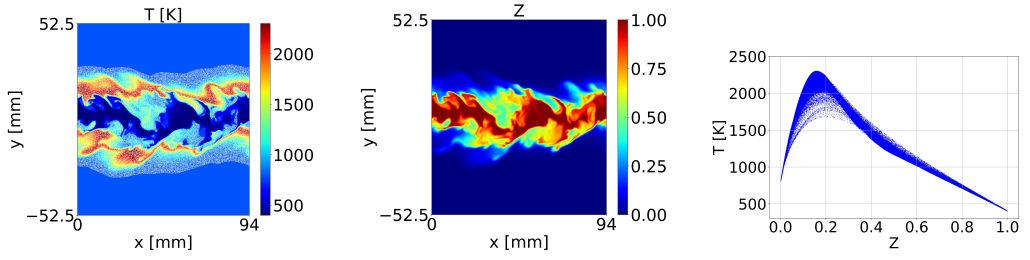
3. Direct Numerical Simulation data

The generality of machine learning algorithms is a key feature for the development of reliable combustion models [63]. In this regard, miscellaneous reacting flow conditions should be considered to strengthen the prediction capabilities of new data-driven models. We consider three-dimensional Direct Numerical Simulation datasets of a non-premixed n-heptane/air jet flame [64, 65, 66] for model training and validation, along with a turbulent methane-air-jet flame [67, 68] for extrapolation and transfer learning purposes. Details about the datasets are reported hereafter.

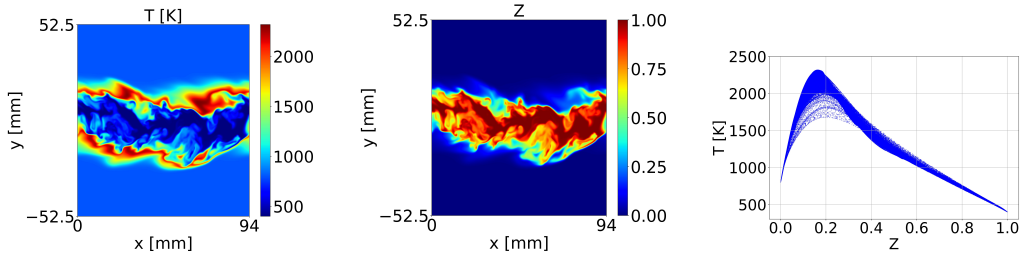
3.1. 3D DNS of a non-premixed n-heptane/air jet flame

Attili et al. [64, 65, 66] performed 3D DNS of a turbulent non-premixed sooting flame. The fuel stream consists of n-heptane diluted with 85% (by volume) nitrogen at a temperature of 400 K. The oxidizer stream consists of air at 800 K. The pressure was set at the atmospheric value, and the stoichiometric mixture fraction is Z_{st} was evaluated at 0.147. The jet centerline initial velocity was imposed at 8.74 m/s, while the reverse coflow velocity was -8.74 m/s yielding a jet Reynolds number of Re =15000. Combustion was modeled using a reduced mechanism for the oxidation of n-heptane comprising $N_S = 47$ species and $N_R = 290$ reactions. The domain dimensions in the periodic directions are $L_x = 94$ mm (6.3 times the initial jet width $H = 15$ mm) and $L_z = 47$ mm ($L_z/H = 3.1$). The crosswise coordinate spans the range $[-L_y/2, L_y/2]$, where $L_y = 105$ mm ($L_z/H = 7$). The domain was

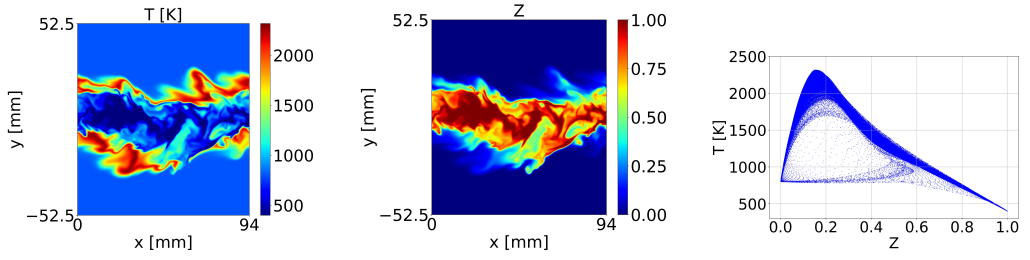
discretized with $1024 \times 1024 \times 512 \approx 500$ million points. The mesh was homogeneous in all directions, with grid resolution $\delta = 91$ μm. The minimum Kolmogorov scale η was 110 μm so that $\delta/\eta = 0.82$ in this case. Moreover, a minimum of 10 grid points were included in the OH layer, i.e., $\delta_{OH} = 10\delta$. More details on the DNS can be found in Ref. [64]. In this study, the dataset represents three 2D planes extracted from the 3D domain at 15 ms. More specifically, the 2D planes are obtained from the 3D flame at locations $L_z/2$, $L_z/4$, and $3L_z/4$, and hereafter called Planes A, B, and C, respectively. The DNS temperature field T, mixture fraction Z, and scatter plot of the Z-T for the three planes are presented in Fig. 3.



(a) Plane A



(b) Plane B

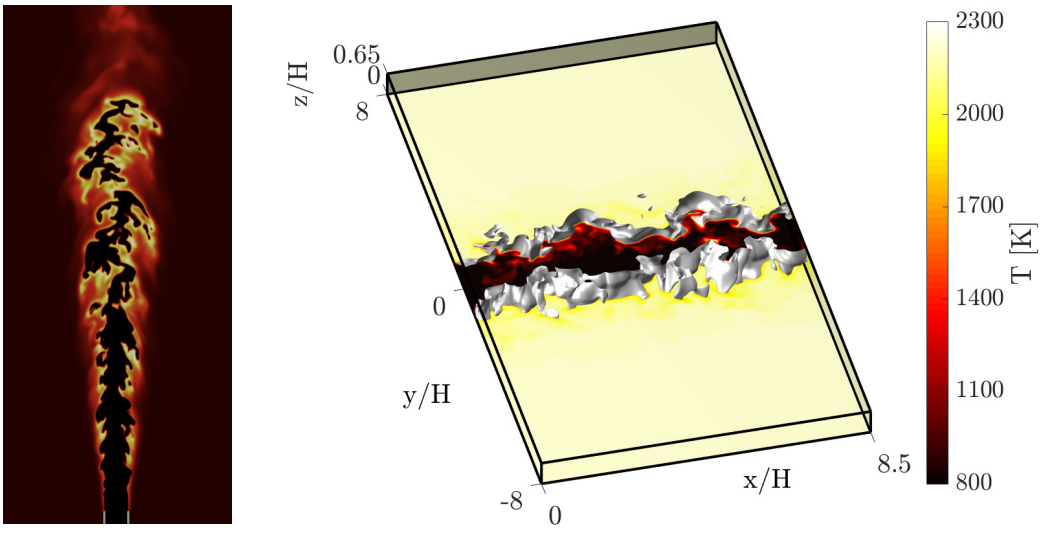


(c) Plane C

Figure 3: DNS temperature field, mixture fraction, and scatter plot of the Z-T of the non-premixed turbulent sooting flame [64]. The training data points are marked with white points at Plane A.

3.1.1. 3D turbulent premixed methane-air flame DNS data

3D DNS of premixed turbulent jet flames with varying Reynolds numbers were performed by Luca et al. [67]. In particular, the case with moderate Reynolds number $Re = 5600$, referred to as the R2 slot jet flame in Luca et al. [69, 67, 68] was selected in the present work, see Fig. 4a.



(a) (b)

Figure 4: (a) 2D visualisation of the full flame and (b) 3D subset temperature field of the Direct Numerical Simulation of a three-dimensional premixed turbulent methane-air flame [67, 68]. The progress variable $C_{O_2} = 0.73$ is marked as a white isosurface on the 3D plot.

The configuration is a turbulent lean methane-air slot jet flame of equivalence ratio $\Phi = 0.7$ surrounded by a coflow of burnt gases [67, 68]. The bulk velocity of the jet is $U = 100m.s^{-1}$ yielding to a Reynolds number $Re = UH/\nu = 5600$ where $H = 1.2mm$ is the slot width and ν is the kinematic viscosity of the reactants. The whole 3D domain [67], of size 24H $\times$ 16H $\times$ 4.3H in the streamwise x, the crosswise y, and the spanwise z direction, is discretized with $1440 \times 962 \times 256 \approx 354$ million points. The domain is periodic in z, open boundary conditions are prescribed at the outlet in x and no-slip conditions are imposed at the boundaries in y. The mesh is uniform in all three directions with a resolution $\delta = 20\mu m$. Only a 20 million point subset was considered in the present work of size 8.5H $\times$ 16H $\times$

0.65H, i.e., $512 \times 962 \times 40$. The 3D subset is displayed in Fig. 4b along with the isosurface of an O_2 based progress variable $C_{O_2} = 0.73$ (flame surface of peak methane consumption in a one-dimensional unstretched laminar flame) [67, 68]. The flame lies in the thin reaction zone regime with a characteristic thermal thickness $\delta^{th} = 110\mu m$. The simulation features finite rate chemistry, described by a skeletal methane mechanism with 16 species and 72 reactions [69]. Additional information about the DNS dataset can be found in Refs. [67, 68].

3.2. Extraction of LES filtered quantities

In the present study, the LES quantities of interest are obtained from the DNS by applying Gaussian filtering, denoted by the operator $\overline{(.)}$, and Favre averaging operations [70, 18, 12]. Any thermo-chemical state variable is then written as $\tilde{\phi} = \overline{\rho\phi}/\overline{\rho}$. The function associated to the Gaussian filter $G(r)$ is defined as

$$G(r) = \left(\frac{6}{\pi\bar{\Delta}^2}\right)^{1/2} \exp\left(-\frac{6r^2}{\bar{\Delta}^2}\right), \tag{13}$$

in accordance with Pope [71] and Leonard [72], where $\bar{\Delta} = \Delta\delta_{\mathrm{DNS}}$ is the LES filter width and δ_{DNS} the DNS grid resolution. Different values of the normalised filter width Δ are investigated. For the non-premixed case, the filter widths $\Delta = 2$ and 5 are considered to match fine and moderate LES grids, following a similar criteria adopted by [12].

The premixed data, instead, is used to test the extrapolation capabilities of the model. More specifically, the purpose of the extrapolation test is to investigate the ability of the identified model to generalize to a complex flame configuration that is different from the training data. To achieve this goal, we follow the approach proposed by Nista et al. [73] that identifies the key flow characteristics that should be preserved for a GAN model to perform well on a very different flame configuration. In particular, they show that a generalization criterion is to conserve the ratio between the filter size and the Kolmogorov scale $\bar{\Delta}/\eta$. Given the Kolmogorov scale of the premixed jet flame, we apply a LES filter of size $\Delta = 3$ to match the $\bar{\Delta}/\eta$ ratio of the non-premixed case calculated at $\Delta = 2$. Table 1 sums up the similar and different flow variables between the employed LES filter widths for the DNS datasets.

Table 1: DNS characteristic scales and LES filter widths.

DNS dataset	Re	δ_{DNS} [μm]	η [μm]	Δ
3D non-premixed [64]	15000	91	110	2 - 5
3D premixed [67]	5600	20	23	3

4. Results and Discussions

In this section, we present a sub-grid machine learning modeling strategy, based on the sparse symbolic regression approach, to discover new functional forms of the cell reacting fraction and improve the predictivity of the reactor-based TCI models. We assume that, in each computational cell, a single value of γ closes the set of Eqs. (4) for all reacting species, as it is commonly done in literature for ensuring mass conservation[2]. First, we *train* the model in Eq.(9), where $\boldsymbol{\Theta}(\mathbf{X})$ are plausible candidate functions for γ given by Eq. (12), using observations from a 2-dimensional plane (Plane A) extracted from the 3D dataset. We embed prior knowledge in the model by introducing plausible candidate functions that satisfy underlying physical constraints. Afterward, we *test* the discovered model on previously unseen data, *i.e.*, other planes (Planes B and C) from the 3D dataset. Finally, we explore the ability of the discovered model to generalize to a flame configuration that is substantially different from the training one.

4.1. Training strategy

For the learning process, 100000 data points of the chemical source terms are randomly selected from the filtered DNS field. The non-reacting region is excluded by the random selection, to avoid dataset imbalances toward non-informing observations. More specifically, the learning data is conditioned to $T > 800$ K. Figure 3 (a) shows the DNS temperature field with the randomly selected points.

To avoid overfitting the resulting models, we use k-fold cross-validation [57], where the dataset is randomly partitioned into k equally-sized subsets. The

[2]The use of species-dependent values γ_i of the cell reacting fraction, for groups of species or individual species, can introduce mass imbalance unless a constraint is imposed as $\sum_i \gamma_i \dot{\omega}_i^* = 0$ [12].

model is trained using $k-1$ folds as the training data, and the remaining left-out fold is used as the validation set. Here, we split the initial dataset into $k = 5$ folds, which is equivalent to an 80% - 20% partitioning for training and validation datasets, respectively, in each optimization run. Finally, we employ Planes B and C as testing datasets for the performance evaluation of the models. The lasso regression objective function is not differentiable, however, a wide variety of techniques from convex analysis and optimization theory have been developed to compute the solution of lasso regression. In the present work, we employ the gradient descend optimizer L-BFGS-B [74] to minimize the cost function:

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{N_c \times N_s} \sum_{c}^{N_c} \sum_{i}^{N_s} \Delta h_{i,c}^{f} \times \left(\bar{\dot{\omega}}_{i,c}^{DNS} - \bar{\dot{\omega}}_{i,c}^{\mathcal{M}}(\boldsymbol{\beta})\right)^2 + \lambda \|\boldsymbol{\beta}\|_1, \tag{14}$$

where the residual sum of squares between the chemical source terms from filtered DNS $\bar{\dot{\omega}}_i^{DNS}$, and the chemical source terms modeled by the reactor-based model $\bar{\dot{\omega}}_i^{\mathcal{M}}$ is weighted with the enthalpy of formation Δh_i^f of the ith chemical species for the data points c. This provides physics-guided information for the learning process, giving priority to ameliorating the predictions of the species which have major contributions to the heat release rate $\dot{Q}$.

The L-BFGS-B algorithm uses curvature information and is computationally efficient to estimate the inverse Hessian matrix to drive the search space. The inverse Hessian matrix is updated in each iteration based on the change in gradients. Wolfe line search is used to ensure that the curvature condition is satisfied and the algorithm update is stable [75]. The optimization algorithm is implemented via the SciPy package for Python [76].

The training procedure comprises a least-squares minimization problem successively solved for $\boldsymbol{\beta}$, in which after each iteration, the vector of parameters is updated, and least-squares is reapplied until a tolerance is attained or a limiting number of iterations of the gradient descent solver is achieved. The tolerance for the optimization is $\epsilon \leq 1 \times 10^{-6}$ and the maximum number of iterations is set to 1000. The regularization parameter λ shapes the final mathematical ansatz of the model. For larger values of λ, most of the model coefficients are reduced to zeros, while for smaller λ values the amount of non-zero coefficients increases. The specification of the regularization parameter is critical since the choice controls the balance between model accuracy and model complexity. Thus, we specify a set of λ values,

for which the minimization of Eq. (14) is solved:

$$\lambda = [0.05, 0.06, 0.07, 0.08, 0.09, 1.0, 1.5, 2.0]^T, \tag{15}$$

which allows us to cover a substantial range of values of the regularization parameter. A different model is identified, with different coefficients $\boldsymbol{\beta}_{\lambda_i}$, for each value λ_i. The computational cost associated with the model identification and calibration is of the order of minutes.

4.2. Training results

Figure 5 shows the ensemble of identified models at the filter width $\Delta = 5$, for increasing values of the regularization parameter λ. Each row represents one candidate model term. The red colour intensity is proportional to the value of the fitted model coefficient of that candidate function, averaged over the 5-folds, corresponding to one λ value (x-axis). If a candidate function is not selected, i.e., it was fitted by a zero coefficient, the corresponding field is coloured in gray. The final model structure depends on the regularization parameter λ. The smaller the regularization parameter, the higher the number of non-zero coefficients, the more complex is the overall model, which comprises more terms, as expected. However, when favouring the sparsity with larger λ values, the learning process reveals a stable pattern, in which the model given by the single term $1/(1+Da)^{1/3}$ is mostly supported by the data, across different regularization parameters and k-fold randomness. A similar trend is found for the smaller filter width $\Delta = 2$.

The performances of the models $\mathcal{M}$ in the 5-folds cross-validation are measured with the Root Mean Squared Error (RMSE) with respect to the filtered DNS, computed on the left-out test data according to

$$\varepsilon(\bar{\dot{Q}}^{\mathcal{M}}) = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\bar{\dot{Q}}_i^{\mathrm{DNS}} - \bar{\dot{Q}}_i^{\mathcal{M}}\right)^2}, \tag{16}$$

where N_t is the number of test points in the left-out fold. The RMSE values for the two considered filter widths are displayed in Fig. 6 normalized with the RMSE of the baseline PaSR model. The RMSE computed with the 5 test data folds are shown for all the regularization parameter values λ. A value of 1 indicates that the sparse regression model exhibits comparable predictive performances to the baseline PaSR. We note that the regressed models $\mathcal{M}$ improve the accuracy of the baseline model, as all $\varepsilon(\bar{\dot{Q}}^{\mathcal{M}})/\varepsilon(\bar{\dot{Q}}^{\mathrm{PaSR}})$ ratios are

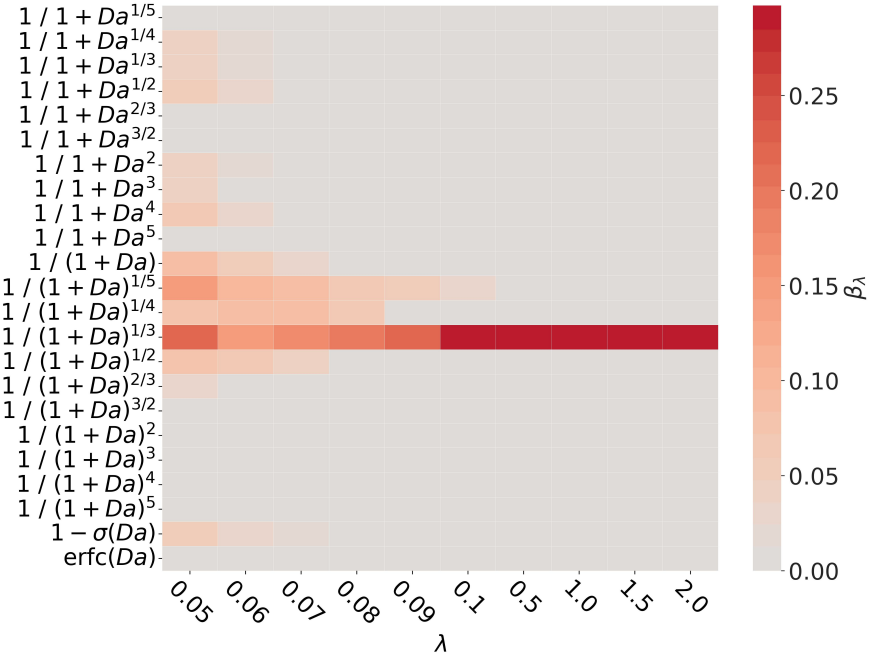


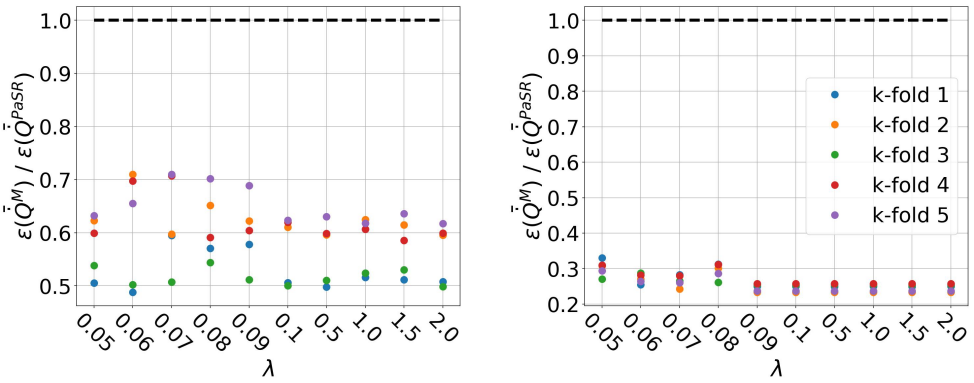
Figure 5: Fitted model coefficients of the selected model forms, for increasing values of the regularization parameter λ, averaged over the 5-folds, at filter width $\Delta = 5$.

below unity, for all the values of λ and in all the 5-folds. While a moderate improvement is observed at filter width $\Delta = 2$, a significant amelioration is obtained at filter width $\Delta = 5$, where the error on $\dot{Q}$ is $\sim 25\%$ of the original formulation. In addition, smaller errors are exhibited for higher values of λ with a lower scattering level across the folds, hence supporting the quest for sparsity in the regressed models.

Table 2: Best predictive models for the fraction of cell occupied by fine structures for 3D DNS data at different filter widths. The normalized RMSE is obtained by the testing set.

Δ	λ	C_Δ	$\varepsilon(\bar{\dot{Q}}^{SpaR})/\varepsilon(\bar{\dot{Q}}^{\mathrm{PaSR}})$
2	1.0	0.5687	0.4942
5	1.0	0.2967	0.1995

Given the cross-validation assessment, we select model forms based on the lowest normalized RMSE per filter width, and the lowest number of



(a) $\Delta = 2$ (b) $\Delta = 5$

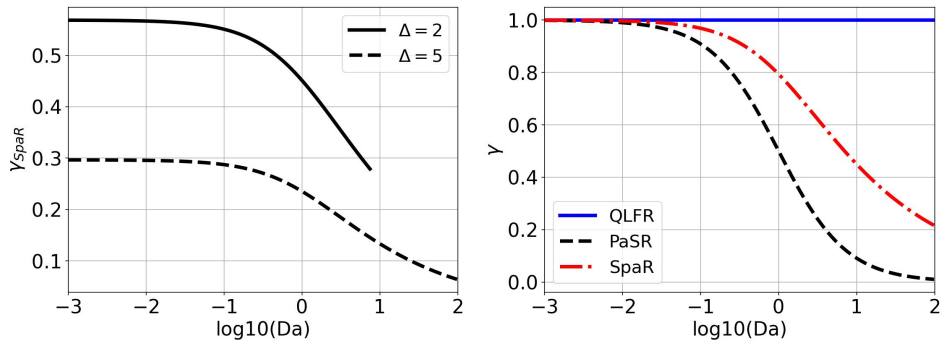
Figure 6: The root mean squared error in the heat release rate normalised by the root mean squared error of the baseline PaSR model.

candidate functions employed, thus leveraging the inherent advantage that a simple model form has in reducing overfitting issues [77]. So, the SpaR identified model form is given by

$$\gamma_{SpaR} = C_\Delta \times \frac{1}{(1 + Da)^{1/3}}, \tag{17}$$

where the model coefficient C_Δ is the respective component of the vector of model coefficients $\boldsymbol{\beta}_\lambda$ multiplying the best candidate function. Table 2 shows the value of the fitted coefficient for the two differently filtered DNS datasets, *i.e.*, $\Delta = 2$ and $\Delta = 5$. Here, the model coefficient C_Δ may be interpreted as a numerical correction of the filtered thermo-physical quantities obtained at fine and moderate LES grids. We observe that, as the filter width increases, the model coefficient decreases. That might be possibly explained by the fact that the cell reacting fraction γ represents the sub-grid fraction of the fine structures in the computational cell, and when larger filters are applied, a smaller fraction of the fine structures are expected in the cell volume, as shown in Fig. 7(a). Also, Figure 8 shows the Damköhler number at different filter widths. It shows how the filter operator affects the estimation of the thermo-physical quantities: an overestimation of the Damköhler number is observed at larger filter widths compared to the fine filter width. This directly influences the model formulation through the model coefficient C_Δ. Thus, we conclude that enforcing the assumption $\gamma = 1$ at a very small Damköhler number, as done in the PaSR model, can be not true when larger filter widths

are applied.



(a) (b)

Figure 7: Cell reacting fractions γ with the Damköhler number. (a) SpaR cell reacting fractions at different filter widths Δ and (b) Different model forms of the cell reacting fraction.

Moreover, we show in Fig 7(b) a comparison of the cell reacting fractions as a function of the Damköhler number (Da) for different model forms: the QLFR (in blue), the original PaSR (black) and the presently learned SpaR, net of the coefficient C_Δ (red). The QLFR model enforces $\gamma = 1$ for the whole range of the Da number. Péquin et al. [12] used the concept of similarity function and showed that at fine filter widths the QLFR assumption is validated if the functions similar to the cell reacting fraction provide values close to unity. However, that assumption might not be valid when larger filters are applied, since larger proportions of the mixing layer thickness are unresolved. In the PaSR formulation, the cell reacting fraction has an abrupt decay in the range $Da \in [10^{-1},\ 10^{1}]$. In the case of a perfectly mixed cell ($\tau_m << \tau_c$), the reacting fraction occupies the entire cell volume, $\gamma = 1$, and if the chemical timescale is much faster than the mixing timescale ($\tau_c << \tau_m$) the reacting fraction is infinitely thin, $\gamma = 0$. Finally, the SpaR model (with $C_\Delta = 1$) has a similar behavior to the PaSR formulation for the perfectly mixed scenario but has a smoother decay from $Da \approx 10^{-1}$ until it reaches the infinitely fast chemistry assumption.

Lastly, we discuss regression diagnostics to assess if the model adequately represents the structure of the data. Our regression model is defined as $\mathcal{M} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is a statistical model for the discrepancy with the data. As commonly done in linear regression, we assume that discrepancies are additive, normal, independently distributed with zero mean and constant

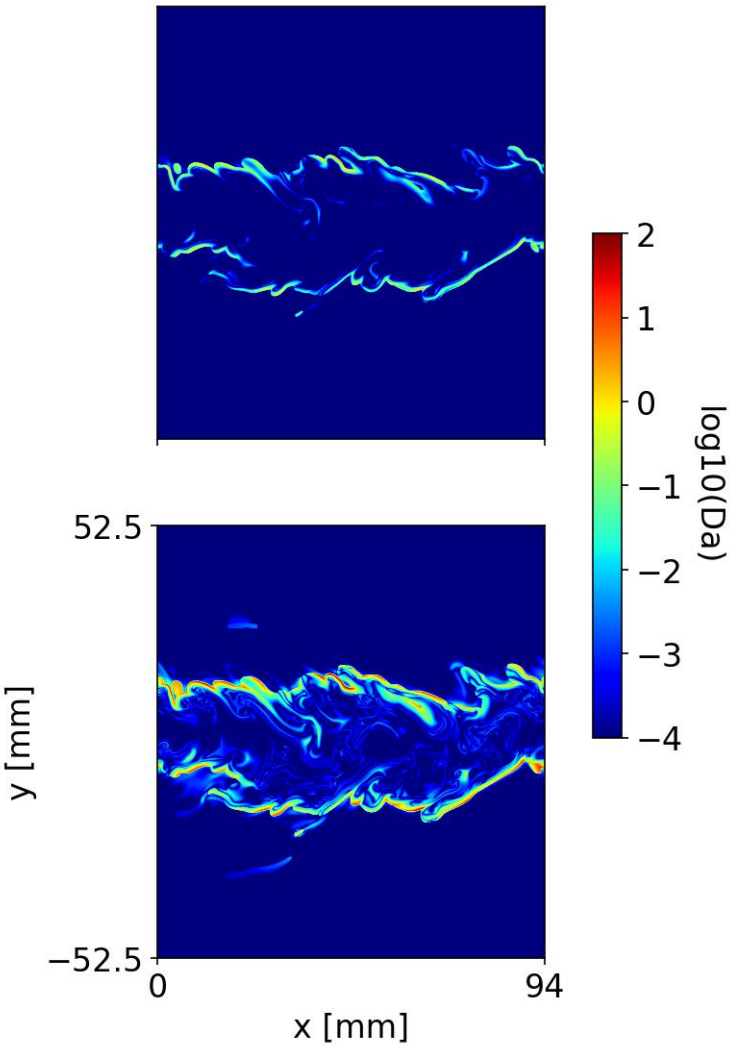
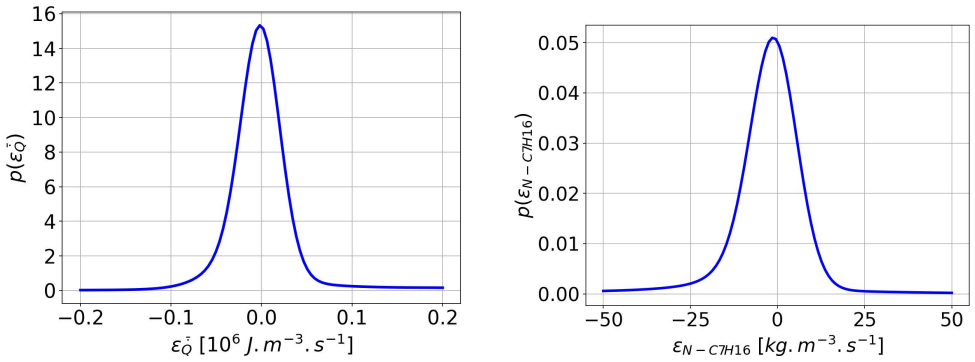


Figure 8: Damköhler number at different filter widths Δ. From top to bottom are presented the fine and moderate filter widths at Plane A.

variance [78], i.e., *i.i.d.*. However, this implies that such discrepancies are only due to uncorrelated noise/fluctuations in the data, i.e., not to structural deficiencies of the model. Figure 9 shows the probability density function of the discrepancies of the heat release rate and the n-heptane source term, computed as $\epsilon_c = \bar{\dot{\omega}}_c^{DNS} - \bar{\dot{\omega}}_c^{SpaR}$ in all grid points c. The slight asymmetries in the shown PDFs are symptoms of violations of the *i.i.d.* hypothesis of

the data fluctuations around the model. This may introduce biases on the estimation of the regression coefficients. A more advanced learning strategy would encompass the modeling of the model discrepancy, which should not be intended as a statistical quantity, but rather as a systematic augmenting term, as done in [79]. This will be pursued in a follow-up work. In this work, we adhere to the classic linear regression framework, but we interpret non-*i.i.d.* residuals as indications of model structural deficiency.



(a) (b)

Figure 9: Probability distribution of the residuals of the (a) heat release rate ($\dot{Q}$) and (b) N-C7H16 net production rate, at filter width $\Delta = 5$.

4.3. Testing the model on unseen data

In this section, we test the performance of the SpaR model on unseen data. The test dataset corresponds to planes B and C extracted from the 3D domain, which are used to obtain an unbiased evaluation of the model fit. Although such planes are extracted from the same domain, the flame structure differs locally from plane A, used in the training process.

The performance of the SpaR model is assessed with the mean absolute error (MAE) between the chemical source terms computed from filtered DNS and the quantities modeled by the reactor-based models, for the i-th chemical species,

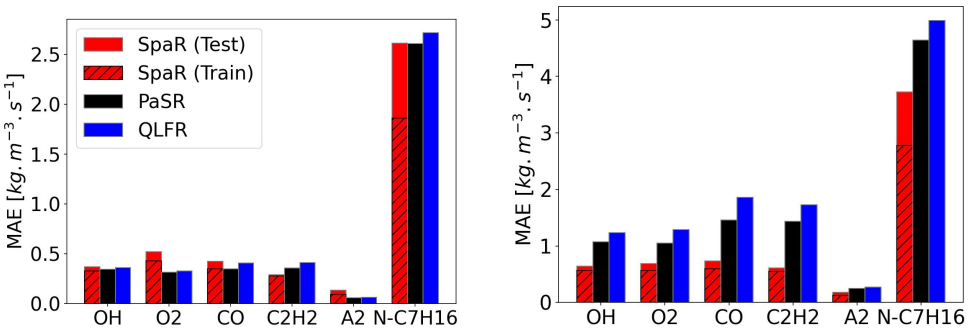
$$\mathrm{MAE}_i^{\mathrm{train/test}} = \frac{1}{N_{\mathrm{train/test}}} \sum_{c=1}^{N_c} |\bar{\dot{\omega}}_{i,c}^{DNS} - \bar{\dot{\omega}}_{i,c}^{M}| \tag{18}$$

evaluated point-wise on all grid points N_{train} of the training dataset and N_{test} of the test dataset. The comparison between error metrics evaluated on the

train and test datasets is useful to assess the quality of fitting: if the training error is lower than the test error, the model has overfitted the data and it is not able to generalize on unseen data. Figure 10 shows in red the test error $\mathrm{MAE}_i^{\mathrm{test}}$ for select species, namely OH, O_2, CO, C_2H_2, A2, and N-C_7H_16, at both filter widths. The training error $\mathrm{MAE}_i^{\mathrm{train}}$ is overlapped with a striped background. The test error is slightly higher than the training error: this is evidence of the absence of overfitting. Then, we compare the SpaR error metrics against the same metrics obtained with the QLFR and PaSR models (in blue and black, respectively). The errors show that all the models have structural deficiencies that prevent a perfect agreement with the filtered chemical source terms extracted from DNS, especially for n-heptane. We observe that at fine filter width ($\Delta = 2$) no substantial improvements are promoted by the SpaR formulation over the QLFR and PaSR ones. This corroborates the results of Péquin et al. [12], where at fine filter width the cell fraction occupied by the fine structures tends to 1, and the assumption of $\gamma = 1$, as enforced in the QLFR model, returns satisfying predictions of the filtered thermo-physical quantities. In such cases, improvements should be sought through alternative model forms, which relax the assumption of linearity in the term representing the mass exchange between the fine structures and the surroundings in Eq. (4), which is beyond the scope of the present work.

Nonetheless, Figure 10(b) shows that the SpaR model is indeed able to reduce the discrepancies with the filtered DNS chemical source terms for the larger filter width $\Delta = 5$. This result demonstrates that new model forms of the cell reacting fraction are a viable route to improve reactor-based models. Such evidence is further confirmed by comparing the DNS conditional mean profiles with reactor-based modeled profiles, evaluated on the test dataset, as shown in Fig. 11. Conditional mean profiles provide a higher-level approach to compare DNS and LES [80], than comparing simply the mean profiles of the chemical source terms.

The mean of the n-heptane (N-C_7H_{16}) source term conditioned to the mixture fraction Z at filter width $\Delta = 5$ is shown in Fig. 11(a). The profile computed with the SpaR model agrees well with the DNS profile at lean and stoichiometric conditions. However, for very rich conditions ($Z > 0.4$), the SpaR model underestimates the fuel consumption. This might be explained by the fact that the SpaR model assumes a global numerical correction of the cell reacting fraction, given by the model coefficient C_Δ. A possible solution would be to enforce a local numerical correction in which C_Δ is built as a function of the local quantities. Deep learning models can provide



(a) $\Delta = 2$ (b) $\Delta = 5$

Figure 10: Mean absolute errors (MAE) between the filtered DNS chemical source terms with those modeled by the reactor-based models for the Planes at different filter widths Δ. The species MAE have been rescaled for visual representation: A2 ($\times 10^2$).

highly accurate and flexible approaches to ameliorate the correction term using physics-guided information. The construction of such a term is beyond the scope of this paper but represents a proper direction to further improve the predictivity of the reactor-based models. Moreover, the QLFR and PaSR models present a considerable overprediction of the consumption of fuel, especially in the rich region with mixture fractions ranging from the stoichiometric mixture fraction ($Z_{st} = 0.147$) to $Z = 0.5$, and an underestimation of the consumption beyond $Z = 0.5$.

Figures 11(b) and (c) show conditional means of acetylene (C_2H_2) and naphthalene (A2), respectively. These species play a major role in the formation and evolution of soot [64]. Thus, the statistics of such species are important to understand soot formation, growth, and consumption. The C_2H_2 conditional mean profile predicted by the SpaR model recovers the DNS data fairly well, while in the case of A2 the SpaR model underestimates the A2 production at very rich conditions, mainly in the mixture fractions ranging from $Z = 0.3$ to $Z = 0.5$. Note that, the baseline reactor-based models fail to capture the conditional means, overestimating both the consumption and the production of A2, specifically from $Z_{st} = 0.147$ to $Z = 0.5$. Attili et al. [64] showed that naphthalene is significantly affected by the small-scale variations of the turbulent field compared to other species controlled by oxidation chemistry. This might explain the worse predictions of reactor-based models which, due to the naphthalene sensitivity to turbulent mixing, suffer the stronger intermittencies of the field [66].

The conditional means of the heat release rate $\dot{Q}$ for the DNS and the reactor-based models at filter width $\Delta = 5$ are shown in Fig. 11(d). We observe that the SpaR model recovers the DNS profile satisfactorily well. Instead, similarly to the major species presented in Fig. 11(a), (b) and (c), the QLFR and PaSR return an overestimation of the heat release rate profiles from the stoichiometric mixture fraction to the rich side.

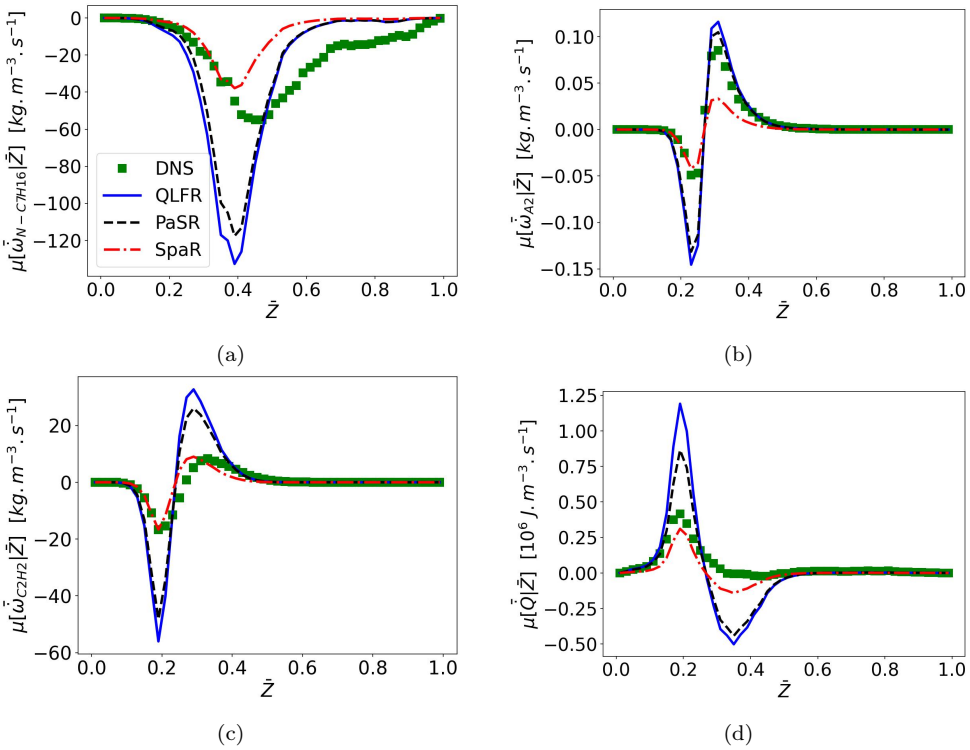


Figure 11: Comparisons of conditional mean profiles provided by DNS and computed from reactor-based models for (a) N-C7H16, (b) A2, (c) C2H2 and (d) heat release rate ($\dot{Q}$) at filter width $\Delta = 5$.

4.4. Extrapolating the model to a different flame configuration

We assess the performance of the identified SpaR model in an extrapolation test on a substantially different flame configuration, to verify the methodology's robustness and the generalization capabilities of the model. Therefore, in addition to the turbulent non-premixed sooting flame, we consider DNS data from a turbulent premixed jet flame. As discussed in Section 3.2, the approach proposed by Nista et al. [73] is followed to preserve

the key flow characteristics across DNS datasets. A filter of size $\Delta = 3$ on the premixed DNS data approximates well the $\bar{\Delta}/\eta$ ratio calculated on the non-premixed case at $\Delta = 2$. Therefore, the value of the model coefficient $C_\Delta = 0.5687$ is selected for the extrapolation test.

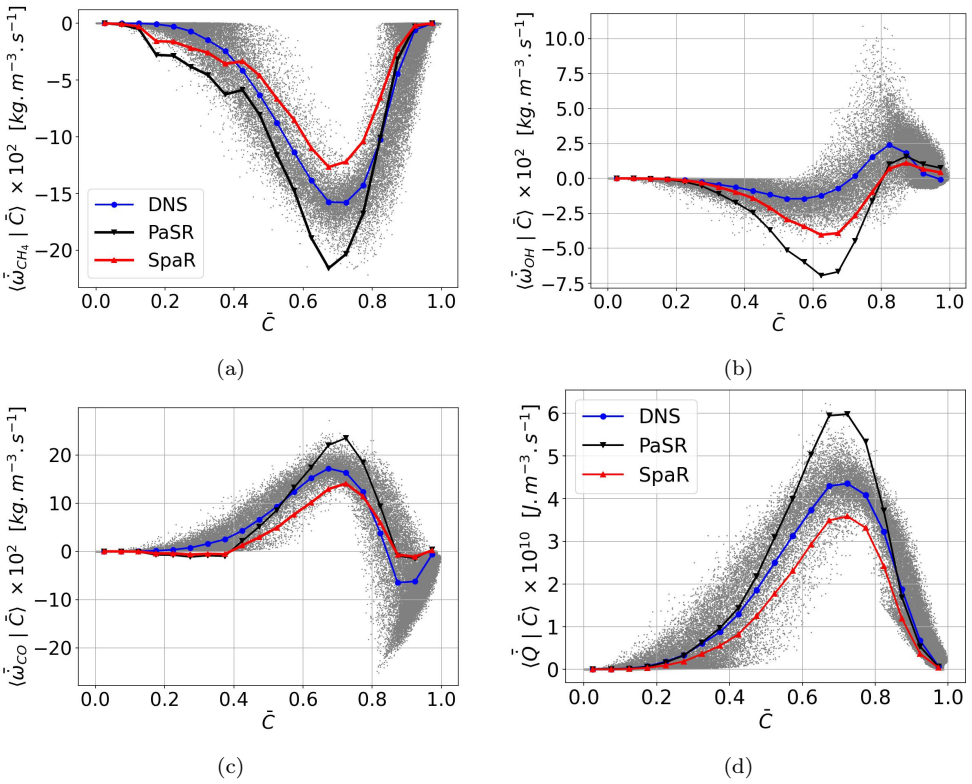


Figure 12: Comparisons of conditional mean profiles provided by DNS and computed from reactor-based models for (a) CH4, (b) OH, (c) CO and (d) heat release rate ($\dot{Q}$) at filter width $\Delta = 3$.

The means for species source terms and heat release rate $\dot{Q}$ conditioned to the progress variable C for the DNS and the reactor-based models at filter width $\Delta = 3$ are shown in Fig. 12. Note that, the SpaR model recovers the DNS data fairly well, while the baseline reactor-based model (PaSR) overestimates both the consumption of the fuel and the production of the species. This is further confirmed when the performance of the SpaR model is assessed through the MAE metric for selected species, namely H_2, H, O, O_2, OH, H_2O, CH_4, CO, and CO_2. Figure 13 compares the SpaR error metrics against the same metrics obtained with the PaSR model. The errors

show that both models have deficiencies that prevent a perfect agreement with the filtered chemical source terms extracted from DNS. However, the identified SpaR model is able to reduce the discrepancies with the filtered DNS chemical source terms.

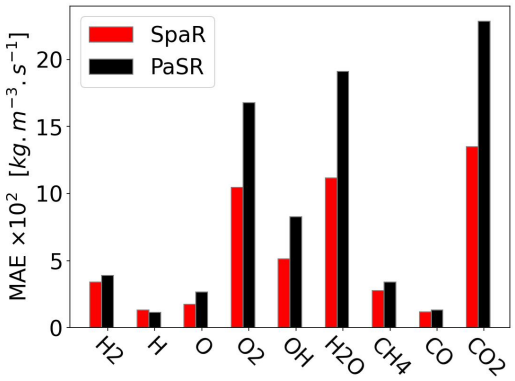


Figure 13: Mean absolute errors (MAE) between the filtered DNS chemical source terms with those modeled by the reactor-based models. The species MAE have been rescaled for visual representation: H2 and H ($\times 10^2$), and O, OH and CH4 ($\times 10^1$).

Going further, we observe that the heat release rate predicted by the SpaR model recovers the DNS fairly well but slightly underestimates it near the peak methane consumption ($C_{O_2} = 0.73$), as shown in Fig. 12(d). Instead, the PaSR model returns an overestimation of the heat release rate. To quantitatively evaluate both reactor-based models in the heat release rate prediction accuracy, the quadratic error [81] is computed between the predicted heat release rates and the filtered DNS. The identified SpaR model deviates from the DNS by 7.7%, while the baseline PaSR model deviates by 10.5%. Hence, we conclude that the SpaR model gives overall better predictions than the PaSR model in considerably different test conditions, confirming the generalization capabilities of the identified model.

5. Conclusions and Future Remarks

In this work, we contribute to the emerging area of physics-guided machine learning models in combustion. More specifically, a data-driven, reactor-based turbulence-chemistry interaction modeling framework is proposed. The study is based on sparse-promoting techniques to discover novel model forms for the cell reacting fraction employed in the Partially Stirred Reactor closure. We employed this framework to improve the predictivity of reactor-based models in the context of LES. We used data from a three-dimensional DNS dataset of a non-premixed n-heptane/air jet flame. We built a library of plausible candidate functions and we performed a sparse symbolic regression (SpaR) in order to retain the physical interpretability of the cell reacting fraction model. The results showed that data mostly supported the model form $\gamma_{\mathrm{SpaR}} = 1/(1 + Da)^{1/3}$, independently of the filter width employed for generating the synthetic LES data from the DNS. However, a dependence on the filter width was observed in the fitted model coefficient, which may be interpreted as a correction term for the filter size. The data-driven model returned more accurate conditional means profiles compared with the baseline PaSR model on unseen data, at larger filter widths.

Future work will be dedicated to ameliorating the correction term using physics-guided machine learning models, providing a local correction to the cell reacting fraction. Also, further research efforts will be focused on building alternative model forms, which relax the assumption of linearity in the term representing the mass exchange between the fine structures and surroundings, and on the a-posteriori assessment of such modeling approach on CFD calculations.

Acknowledgements

A. Péquin acknowledges the financial support of the Fonds National de la Recherche Scientifique (FRS-FNRS). Also, this work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 714605 and Marie Skłodowska-Curie grant agreement No 801505.

References

[1] T. Poinsot, D. Veynante, Theoretical and Numerical Combustion, 2nd Edition, R.T. Edwards, Inc, 2005.

[2] N. Ren, D. Zeng, K. V. Meredith, Y. Wang, S. B. Dorofeev, Modeling of flame extinction/re-ignition in oxygen-reduced environments, Proceedings of the Combustion Institute 37 (3) (2019) 3951–3958. doi:https://doi.org/10.1016/j.proci.2018.06.076.
URL https://www.sciencedirect.com/science/article/pii/S1540748918302591

[3] Z. X. Chen, I. Langella, R. S. Barlow, N. Swaminathan, Prediction of local extinctions in piloted jet flames with inhomogeneous inlets using unstrained flamelets, Combustion and Flame 212 (2020) 415–432. doi:https://doi.org/10.1016/j.combustflame.2019.11.007.
URL https://www.sciencedirect.com/science/article/pii/S0010218019305127

[4] D. Veynante, L. Vervisch, Turbulent combustion modeling, Prog. Energy Combust. Sci. 28 (2002) 193–266.

[5] B. Fiorina, R. Vicquelin, P. Auzillon, N. Darabiha, O. Gicquel, et al., A filtered tabulated chemistry model for les of premixed combustion, Combust. Flame 157 (3) (2010) 465–475.

[6] B. MAGNUSSEN, On the structure of turbulence and a generalized eddy dissipation concept for chemical reaction in turbulent flow. arXiv:https://arc.aiaa.org/doi/pdf/10.2514/6.1981-42, doi:10.2514/6.1981-42.
URL https://arc.aiaa.org/doi/abs/10.2514/6.1981-42

[7] J. Chomiak, Combustion a study in theory, fact and application (1 1990).
URL https://www.osti.gov/biblio/5894595

[8] A. De, A. Dongre, Assessment of turbulence-chemistry interaction models in mild combustion regime, Flow, Turbulence and Combustion 94 (2) (2015) 439–478. doi:10.1007/s10494-014-9587-8.
URL https://doi.org/10.1007/s10494-014-9587-8

[9] H. Lu, W. Chen, C. Zou, H. Yao, Large-eddy simulation of sandia flame f using structural subgrid-scale models and partially-stirred-reactor approach, Physics of Fluids 31 (4) (2019) 045109. arXiv:https://doi.org/10.1063/1.5087078, doi:10.1063/1.5087078.
URL https://doi.org/10.1063/1.5087078

[10] Z. Li, A. Cuoci, A. Parente, Large eddy simulation of mild combustion using finite rate chemistry: Effect of combustion sub-grid closure,

Proceedings of the Combustion Institute 37 (4) (2019) 4519–4529. doi:https://doi.org/10.1016/j.proci.2018.09.033.
URL https://www.sciencedirect.com/science/article/pii/S1540748918306382

[11] V. Sabelnikov, C. Fureby, Extended les-pasr model for simulation of turbulent combustion, EUCASS Proceedings Series 4 (2013) 539–568. doi:10.1051/eucass/201304539.

[12] A. Péquin, S. Iavarone, R. Malpica Galassi, A. Parente, The partially stirred reactor model for combustion closure in large eddy simulations: Physical principles, sub-models for the cell reacting fraction, and open challenges, Physics of Fluids 34 (5) (2022) 055122. arXiv:https://doi.org/10.1063/5.0090970, doi:10.1063/5.0090970.
URL https://doi.org/10.1063/5.0090970

[13] R. Amaduzzi, A. Bertolino, A. özden, R. Malpica Galassi, A. Parente, Impact of scalar mixing uncertainty on the predictions of reactor-based closures: Application to a lifted methane/air jet flame, Proceedings of the Combustion Institute (2022). doi:10.1016/j.proci.2022.06.028.

[14] H. Liu, Z. Yin, W. Xie, B. Zhang, J. Le, H. Liu, Numerical and analytical assessment of finite rate chemistry models for les of turbulent premixed flames, Flow, Turbulence and Combustion (2022) 1–24.

[15] P. Nordin, Complex chemistry modeling of diesel spray combustion (2001) 1–55.

[16] M. Ferrarotti, M. Fürst, E. Cresci, W. de Paepe, A. Parente, Key modeling aspects in the simulation of a quasi-industrial 20 kw moderate or intense low-oxygen dilution combustion chamber, Energy & Fuels 32 (10) (2018) 10228–10241. arXiv:https://doi.org/10.1021/acs.energyfuels.8b01064, doi:10.1021/acs.energyfuels.8b01064.
URL https://doi.org/10.1021/acs.energyfuels.8b01064

[17] Z. Li, A. Parente, A review of the numerical investigations of jet-in-hot-coflow burner with reactor-based models, Frontiers in Mechanical Engineering 6 (2020) 512501.

[18] S. Iavarone, A. Péquin, Z. X. Chen, N. A. K. Doan, N. Swaminathan, A. Parente, An a priori assessment of the partially

stirred reactor (pasr) model for mild combustion, Proceedings of the Combustion Institute 38 (4) (2021) 5403–5414. doi:https://doi.org/10.1016/j.proci.2020.06.234.
URL https://www.sciencedirect.com/science/article/pii/S1540748920303266

[19] Z. Li, S. Tomasch, Z. X. Chen, A. Parente, I. S. Ertesvåg, N. Swaminathan, Study of mild combustion using les and advanced analysis tools, Proceedings of the Combustion Institute 38 (4) (2021) 5423–5432. doi:https://doi.org/10.1016/j.proci.2020.06.298.
URL https://www.sciencedirect.com/science/article/pii/S1540748920303904

[20] S. Iavarone, M. Cafiero, M. Ferrarotti, F. Contino, A. Parente, A multiscale combustion model formulation for nox predictions in hydrogen enriched jet flames, International Journal of Hydrogen Energy 44 (41) (2019) 23436–23457. doi:https://doi.org/10.1016/j.ijhydene.2019.07.019.
URL https://www.sciencedirect.com/science/article/pii/S036031991932539X

[21] M. Ferrarotti, Z. Li, A. Parente, On the role of mixing models in the simulation of mild combustion using finite-rate chemistry combustion models, Proceedings of the Combustion Institute 37 (4) (2019) 4531–4538. doi:https://doi.org/10.1016/j.proci.2018.07.043.
URL https://www.sciencedirect.com/science/article/pii/S1540748918304619

[22] E. Quadarella, A. Péquin, A. Stagni, A. Parente, T. Faravelli, H. G. Im, A generalized partially stirred reactor model for turbulent closure, Proceedings of the Combustion Institute (2022). doi:https://doi.org/10.1016/j.proci.2022.08.061.
URL https://www.sciencedirect.com/science/article/pii/S1540748922003583

[23] R. S. M. Freitas, F. A. Rochinha, D. Mira, X. Jiang, Parametric and model uncertainties induced by reduced order chemical mechanisms for biogas combustion, Chemical Engineering Science Journal (2020).

[24] R. E. Morrison, T. A. Oliver, R. D. Moser, Representing model inadequacy: A stochastic operator approach, SIAM/ASA Journal on Uncertainty Quantification 6 (2018) 457–496.

[25] M. R. E., Embedded discrepancy operators in reduced models of interacting species, arXiv preprint arXiv:1910.08191 (2019).

[26] P. Trisjono, H. Pitsch, Systematic analysis strategies for the development of combustion models from dns: A review, Flow, Turbulence and Combustion 95 (09 2015). doi:10.1007/s10494-015-9645-x.

[27] Y. Minamoto, N. Swaminathan, Subgrid scale modelling for mild combustion, Proceedings of the Combustion Institute 35 (3) (2015) 3529–3536. doi:https://doi.org/10.1016/j.proci.2014.07.025.
URL https://www.sciencedirect.com/science/article/pii/S1540748914003356

[28] L. Zhou, Y. Song, W. Ji, H. Wei, Machine learning for combustion, Energy and AI 7 (2022) 100128. doi:https://doi.org/10.1016/j.egyai.2021.100128.
URL https://www.sciencedirect.com/science/article/pii/S2666546821000756

[29] M. Ihme, W. T. Chung, A. A. Mishra, Combustion machine learning: Principles, progress and prospects, Progress in Energy and Combustion Science 91 (2022) 101010. doi:https://doi.org/10.1016/j.pecs.2022.101010.
URL https://www.sciencedirect.com/science/article/pii/S0360128522000193

[30] M. Aliramezani, C. R. Koch, M. Shahbakhti, Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions, Progress in Energy and Combustion Science 88 (2022) 100967. doi:https://doi.org/10.1016/j.pecs.2021.100967.
URL https://www.sciencedirect.com/science/article/pii/S0360128521000654

[31] A. Oishi, G. Yagawa, Computational mechanics enhanced by deep learning, Computer Methods in Applied Mechanics and Engineering 327 (2017) 327–351, advances in Computational Mechanics and Scientific Computation—the Cutting Edge. doi:https://doi.org/10.1016/j.cma.2017.08.040.
URL https://www.sciencedirect.com/science/article/pii/S0045782517306199

[32] Y. Yang, P. Perdikaris, Conditional deep surrogate models for stochastic, high-dimensional, and multi-fidelity systems, Computational Mechanics 64 (2019) 417–434. doi:https://doi.org/10.1007/s00466-019-01718-y.

[33] Y. Zhu, N. Zabaras, Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification, Journal of

Computational Physics 366 (2018) 415–447.
URL https://doi.org/10.1016/j.jcp.2018.04.018

[34] G. Aversano, M. Ferrarotti, A. Parente, Digital twin of a combustion furnace operating in flameless conditions: reduced-order model development from cfd simulations, Proceedings of the Combustion Institute 38 (4) (2021) 5373–5381. doi:https://doi.org/10.1016/j.proci.2020.06.045.
URL https://www.sciencedirect.com/science/article/pii/S1540748920300742

[35] R. S. M. Freitas, C. H. S. Barbosa, G. M. Guerra, A. L. G. A. Coutinho, F. A. Rochinha, An encoder-decoder deep surrogate for reverse time migration in seismic imaging under uncertainty, Comput Geosci 25 (2021) 1229–1250. doi:https://doi.org/10.1007/s10596-021-10052-3.

[36] D. Field, Y. Ammouche, J.-M. Peña, A. Jérusalem, Machine learning based multiscale calibration of mesoscopic constitutive models for composite materials: application to brain white matter, Comput Mech 67 (2021) 1629–1643. doi:https://doi.org/10.1007/s00466-021-02009-1.

[37] G. Chen, Recurrent neural networks (rnns) learn the constitutive law of viscoelasticity, Comput Mech 67 (2021) 1009–1019. doi:https://doi.org/10.1007/s00466-021-01981-y.

[38] J. R. Holland, J. D. Baeder, K. Duraisamy, Towards Integrated Field Inversion and Machine Learning With Embedded Neural Networks for RANS Modeling. arXiv:https://arc.aiaa.org/doi/pdf/10.2514/6.2019-1884, doi:10.2514/6.2019-1884.
URL https://arc.aiaa.org/doi/abs/10.2514/6.2019-1884

[39] R. Tripathy, I. Bilionis, Deep uq: Learning deep neural network surrogate models for high dimensional uncertainty quantification, Journal of Computational Physics 375 (2018) 565–588.

[40] S. L. Brunton, B. R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, Annual Review of Fluid Mechanics 52 (1) (2020) 477–508. arXiv:https://doi.org/10.1146/annurev-fluid-010719-060214, doi:10.1146/annurev-fluid-010719-060214.
URL https://doi.org/10.1146/annurev-fluid-010719-060214

[41] R. Nakazawa, Y. Minamoto, N. Inoue, M. Tanahashi, Species reaction rate modelling based on physics-guided machine learning, Combustion and Flame 235 (2022) 111696. doi:https://doi.org/10.1016/j.combustflame.2021.111696.
URL https://www.sciencedirect.com/science/article/pii/S0010218021004399

[42] Z. X. Chen, S. Iavarone, G. Ghiasi, V. Kannan, G. D'Alessio, A. Parente, N. Swaminathan, Application of machine learning for filtered density function closure in mild combustion, Combustion and Flame 225 (2021) 160–179. doi:https://doi.org/10.1016/j.combustflame.2020.10.043.
URL https://www.sciencedirect.com/science/article/pii/S0010218020304636

[43] K. Jigjid, Y. Minamoto, N. A. K. Doan, M. Tanahashi, Sgs reaction rate modelling for mild combustion based on machine-learning combustion mode classification: Development and a priori study, Proceedings of the Combustion Institute (2022). doi:https://doi.org/10.1016/j.proci.2022.07.020.
URL https://www.sciencedirect.com/science/article/pii/S1540748922000499

[44] C. J. Lapeyre, A. Misdariis, N. Cazard, D. Veynante, T. Poinsot, Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates, Combustion and Flame 203 (2019) 255–264. doi:https://doi.org/10.1016/j.combustflame.2019.02.019.
URL https://www.sciencedirect.com/science/article/pii/S0010218019300835

[45] Z. M. Nikolaou, C. Chrysostomou, L. Vervisch, S. Cant, Progress variable variance and filtered rate modelling using convolutional neural networks and flamelet methods, Flow, Turbulence and Combustion (103) (2019) 485–501. doi:10.1007/s10494-019-00028-w.
URL https://hal.archives-ouvertes.fr/hal-02420299

[46] M. Bode, M. Gauding, Z. Lian, D. Denker, M. Davidovic, K. Kleinheinz, J. Jitsev, H. Pitsch, Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows, Proceedings of the Combustion Institute 38 (2) (2021) 2617–2625. doi:https://doi.org/10.1016/j.proci.2020.06.022.
URL https://www.sciencedirect.com/science/article/pii/S1540748920300481

[47] S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems,

Proceedings of the National Academy of Sciences 113 (15) (2016) 3932–3937. arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1517384113, doi:10.1073/pnas.1517384113.
URL https://www.pnas.org/doi/abs/10.1073/pnas.1517384113

[48] M. Schmelzer, R. Dwight, P. Cinnella, Discovery of algebraic reynolds-stress models using sparse symbolic regression, Flow Turbulence and Combustion 104 (2020) 579–603. doi:10.1007/s10494-019-00089-x.

[49] M. Cranmer, A. Sanchez Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, S. Ho, Discovering symbolic models from deep learning with inductive biases, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, Vol. 33, Curran Associates, Inc., 2020, pp. 17429–17442.
URL https://proceedings.neurips.cc/paper/2020/file/c9f2f917078bd2db12f23c3b41

[50] K. Papastamatiou, F. Sofos, T. E. Karakasidis, Machine learning symbolic equations for diffusion with physics-based descriptions, AIP Advances 12 (2) (2022) 025004. arXiv:https://doi.org/10.1063/5.0082147, doi:10.1063/5.0082147.
URL https://doi.org/10.1063/5.0082147

[51] F. Sofos, A. Charakopoulos, K. Papastamatiou, T. E. Karakasidis, A combined clustering/symbolic regression framework for fluid property prediction, Physics of Fluids 34 (6) (2022) 062004. arXiv:https://doi.org/10.1063/5.0096669, doi:10.1063/5.0096669.
URL https://doi.org/10.1063/5.0096669

[52] W. T. Chung, A. A. Mishra, M. Ihme, Interpretable data-driven methods for subgrid-scale closure in les for transcritical lox/gch4 combustion, Combustion and Flame 239 (2022) 111758, a dedication to Professor Kenneth Noel Corbett Bray. doi:https://doi.org/10.1016/j.combustflame.2021.111758.
URL https://www.sciencedirect.com/science/article/pii/S0010218021005010

[53] V. Sabelnikov, C. Fureby, Extended les-pasr model for simulation of turbulent combustion, Progress in Propulsion Physics 4 (2013) 539–568.

[54] M. Bösenhofer, E.-M. Wartha, C. Jordan, M. Harasek, The eddy dissipation concept—analysis of different fine structure treatments for classical combustion, Energies 11 (7) (2018) 1902.

[55] Z. Li, A. Cuoci, A. Sadiki, A. Parente, Comprehensive numerical study of the adelaide jet in hot-coflow burner by means of rans and detailed chemistry, Energy 139 (2017) 555–570. doi:https://doi.org/10.1016/j.energy.2017.07.132.
URL https://www.sciencedirect.com/science/article/pii/S0360544217313142

[56] V. Raman, H. Pitsch, A consistent LES/filtered-density function formulation for the simulation of turbulent flames with detailed chemistry, Proc. Combust. Inst. 31 (2) (2007) 1711–1719.

[57] S. L. Brunton, J. N. Kutz, Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control, Cambridge University Press, 2019. doi:10.1017/9781108380690.

[58] B. Everitt, The Cambridge dictionary of statistics, Cambridge University Press, Cambridge, UK; New York, 2002.

[59] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288.
URL http://www.jstor.org/stable/2346178

[60] I. Ben Hassan Saïdi, M. Schmelzer, P. Cinnella, F. Grasso, Cfd-driven symbolic identification of algebraic reynolds-stress models, Journal of Computational Physics 457 (2022) 111037. doi:https://doi.org/10.1016/j.jcp.2022.111037.
URL https://www.sciencedirect.com/science/article/pii/S0021999122000997

[61] B. F. Magnussen, The eddy dissipation concept a bridge between science and technology, in: ECCOMAS Thematic Conference on Computational Combustion, 2005.

[62] L. C. Andrews, Special Functions of Mathematics for Engineers, Oxford University Press, 1997. doi:https://doi.org/10.1117/3.270709.

[63] V. Xing, C. Lapeyre, T. Jaravel, T. Poinsot, Generalization capability of convolutional neural networks for progress variable variance and reaction rate subgrid-scale modeling, Energies 14 (16) (2021). doi:10.3390/en14165096.
URL https://www.mdpi.com/1996-1073/14/16/5096

[64] A. Attili, F. Bisetti, M. E. Mueller, H. Pitsch, Formation, growth, and transport of soot in a three-dimensional turbulent non-premixed jet flame, Combustion and flame 161 (7) (2014) 1849–1865.

[65] A. Attili, F. Bisetti, M. E. Mueller, H. Pitsch, Damköhler number effects on soot formation and growth in turbulent nonpremixed flames, Proceedings of the Combustion Institute 35 (2) (2015) 1215–1223.

[66] A. Attili, F. Bisetti, M. E. Mueller, H. Pitsch, Effects of non-unity lewis number of gas-phase species in turbulent nonpremixed sooting flames, Combustion and Flame 166 (2016) 192–202.

[67] S. Luca, A. Attili, E. Lo Schiavo, F. Creta, F. Bisetti, On the statistics of flame stretch in turbulent premixed jet flames in the thin reaction zone regime at varying Reynolds number, Proceedings of the Combustion Institute 37 (2) (2019) 2451–2459. doi:10.1016/j.proci.2018.06.194.
URL https://linkinghub.elsevier.com/retrieve/pii/S1540748918303808

[68] A. Attili, S. Luca, D. Denker, F. Bisetti, H. Pitsch, Turbulent flame speed and reaction layer thickening in premixed jet flames at constant Karlovitz and increasing Reynolds numbers, Proceedings of the Combustion Institute 38 (2) (2021) 2939–2947. doi:10.1016/j.proci.2020.06.210.
URL https://linkinghub.elsevier.com/retrieve/pii/S1540748920303023

[69] S. Luca, A. N. Al-Khateeb, A. Attili, F. Bisetti, Comprehensive Validation of Skeletal Mechanism for Turbulent Premixed Methane–Air Flame Simulations, Journal of Propulsion and Power 34 (1) (2018) 153–160, publisher: American Institute of Aeronautics and Astronautics. doi:10.2514/1.B36528.
URL https://arc.aiaa.org/doi/10.2514/1.B36528

[70] J. C. Sutherland, Evaluation of mixing and reaction models for large-eddy simulation of nonpremixed combustion using direct numerical simulation, Ph.D. thesis, Department of Chemical and Fuels Engineering, University of Utah Salt Lake (2004).

[71] S. B. Pope, S. B. Pope, Turbulent flows, Cambridge university press, 2000.

[72] A. Leonard, Energy cascade in large-eddy simulations of turbulent fluid flows, in: Advances in geophysics, Vol. 18, Elsevier, 1975, pp. 237–248.

[73] L. Nista, C. Schumann, T. Grenga, A. Attili, H. Pitsch, Investigation of the generalization capability of a generative adversarial network for large eddy simulation of turbulent premixed reacting flows, Proceedings of the Combustion Institute (2022). doi:https://doi.org/10.1016/j.proci.2022.07.244.
URL https://www.sciencedirect.com/science/article/pii/S1540748922002851

[74] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific Computing 16 (5) (1995) 1190–1208. arXiv:https://doi.org/10.1137/0916069, doi:10.1137/0916069.
URL https://doi.org/10.1137/0916069

[75] J. Nocedal, S. J. Wright, Numerical Optimization, Springer New York, NY, 2006. doi:https://doi.org/10.1007/978-0-387-40065-5.

[76] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods 17 (2020) 261–272. doi:10.1038/s41592-019-0686-2.

[77] S. H. Rudy, S. L. Brunton, J. L. Proctor, J. N. Kutz, Data-driven discovery of partial differential equations, Science Advances 3 (4) (2017) e1602614. arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.1602614, doi:10.1126/sciadv.1602614.
URL https://www.science.org/doi/abs/10.1126/sciadv.1602614

[78] N. Altman, M. Krzywinski, Regression diagnostics, Nature Methods 13 (2016) 385–386.

[79] K. Sargsyan, X. Huan, H. N. Najm, Embedded model error representation for bayesian model calibration, International Journal for Uncertainty Quantification 9 (4) (2019) 365–394.

[80] J. F. Driscoll, J. H. Chen, A. W. Skiba, C. D. Carter, E. R. Hawkes, H. Wang, Premixed flames subjected to extreme turbulence: Some questions and recent answers, Progress in Energy and Combustion Science 76 (2020) 100802. doi:https://doi.org/10.1016/j.pecs.2019.100802. URL https://www.sciencedirect.com/science/article/pii/S036012851930036X

[81] C. Wall, B. Boersma, P. Moin, An evaluation of the assumed beta probability density function subgrid-scale model for large eddy simulation of nonpremixed, turbulent combustion with heat release, Physics of Fluids 12 (10) (2000) 2522–2529.