Proceedings of the Combustion Institute 40 (2024) 105311

Contents lists available at ScienceDirect

Proceedings of the Combustion Institute

journal homepage: www.elsevier.com/locate/proci

Augmenting filtered flame front displacement models for LES using machine learning with *a posteriori* simulations

Jen Zen Ho [a], Mohsen Talei [a,*], Davy Brouzet [b], Wai Tong Chung [b], Pushan Sharma [b], Matthias Ihme [b,c]

[a] Department of Mechanical Engineering, The University of Melbourne, Australia
[b] Department of Mechanical Engineering, Stanford University, CA, USA
[c] Department of Photon Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

ARTICLE INFO

Keywords:
Large eddy simulation
Machine learning
Combustion modelling
Flame surface density

ABSTRACT

The Flame Surface Density (FSD) model is an affordable combustion model that has been widely used in simulating turbulent premixed flames. In Large Eddy Simulations (LES) with FSD, the combined effect of reaction and diffusion is governed by the Filtered Flame Front Displacement (FFFD) term. While the existing modelling approaches for this term are computationally cost-effective, their predictions still show inconsistencies in certain cases. This study aims to address these inconsistencies by generating Machine Learning (ML) models for the FFFD and FSD terms using the DNS data of a turbulent premixed jet flame. With this approach, the relevance of certain input parameters as well as certain modelling assumptions used for the FFFD term are assessed. Overall, it is found that the resolved curvature term is the most important input parameter to consider and that the resolved progress variable should also be considered in the models. It is shown that the ML models perform significantly better than legacy, algebraic formulations using *a priori* testing. To further assess the performance of ML, one of the ML models is employed in a *a posteriori* LES and compared against simulations with the algebraic model. The ML simulation is stable and yields encouraging improvements on key physical parameters regarding the flame length and the FFFD distribution.

Novelty and Significance Statement: This research is of importance because it answers fundamental and practical questions related to the use of combustion modelling approaches, specifically the Flame Surface Density (FSD) and the Filtered Flame Front Displacement (FFFD) models, by means of Machine Learning (ML) algorithms. From a fundamental aspect, we show that two features which are typically not considered as inputs in combustion models, i.e., the progress variable and the resolved curvature, are key to consider for improved predictions of the model, more so than features which are typically used in FSD modelling, i.e., u'_Δ, Δ, and $|\nabla\tilde{c}|$. From a practical standpoint, we demonstrate a framework to use the developed ML combustion model *a posteriori* in a LES, without any stability issues. Overall, these findings are key to guide further traditional and ML improvement efforts on combustion models.

1. Introduction

Simulations are a useful tool to guide the design process of turbulent combustors, but due to the large range of scales present in the flame and flow field, high-fidelity simulations of practical combustors are still very expensive. Large Eddy Simulation (LES) separates small and large scales and reduces the mesh resolution required and is therefore a powerful technique to reduce cost [1]. LES uses models to approximate the effect of the small scales on the large scale, and its accuracy relies on the fidelity of these models.

An important model for the subgrid scale terms is the combustion model used for the filtered reaction and diffusion terms. A popular model for these terms in premixed flames is the Flame Surface Density (FSD) model [1,2]. Several studies [3–5] have reported significant inconsistencies in the model's predictions against Direct Numerical Simulation (DNS) data.

FSD models introduce a progress variable, c, to track the state of the fluid from unburned gas ($c = 0$) to burnt gas ($c = 1$). This allows all chemical species to be tracked by one transported variable which significantly reduces memory and computational cost. For LES, the filtered transport equation for c is [6]

$$\frac{D(\overline{\rho}\widetilde{c})}{Dt} + \nabla\cdot(\overline{\rho}\widetilde{\boldsymbol{u}}\widetilde{c} - \overline{\rho}\widetilde{\boldsymbol{u}}\widetilde{c}) = \nabla\cdot\overline{\rho D_c\nabla c} + \overline{\dot{\omega}}_c, \tag{1}$$

* Corresponding author.
E-mail address: mohsen.talei@unimelb.edu.au (M. Talei).

https://doi.org/10.1016/j.proci.2024.105311
Received 4 December 2023; Accepted 29 May 2024
Available online 24 June 2024

where $\tilde{\cdot}$ denotes density-weighted (Favre) filtering, $\bar{\cdot}$ denotes spatial filtering, ρ is the density, $\boldsymbol{u}$ is the velocity vector, $\dot{\omega}_c$ is the production rate, and D_c is the diffusion coefficient of c.

In FSD modelling, the right-hand side of Eq. (1) is written as the Filtered Flame Front Displacement (FFFD) term [2]:

$$\nabla \cdot \overline{\rho D_c \nabla c} + \overline{\dot{\omega}}_c = \overline{\rho s_d |\nabla c|}, \tag{2}$$

where s_d is the flame displacement speed. The FFFD term is then modelled as [1]

$$\overline{\rho s_d |\nabla c|} = \overline{(\rho s_d)}_s \Sigma, \tag{3}$$

where $\overline{(\rho s_d)}_s$ is the surface-averaged density weighted displacement speed and $\Sigma = \overline{|\nabla c|}$ is the FSD. These two terms must be modelled.

A common model for the displacement speed term is $\overline{(\rho s_d)}_s = \rho_u s_L$, where s_L is the unstretched laminar flame speed. Other models that have been proposed either rely on assumptions about curvature and employ test filters [3], or statistically represent the displacement speed terms [7].

For modelling Σ, it is possible to either solve for a transport equation [8] or model Σ using an algebraic function [4]. Focusing on algebraic models since they are generally computationally less expensive than solving another transport equation, the FSD is typically modelled using a wrinkling factor, $\Xi = \overline{|\nabla c|}/|\nabla \bar{c}|$.

There have been studies that have pointed out deficiencies with the current models of Ξ and $\overline{(\rho s_d)}_s$ [4,5]. Chakraborty and Klein [4] tested different Ξ models with a turbulent premixed flame using DNS data. They found that models that do not take into account the tangential diffusion component of the displacement speed tend to overpredict FSD. Panek et al. [5] also investigated the performance of Ξ and $\overline{(\rho s_d)}_s$ models for predicting sound generation. They found that several well-established models, including Charlette's [9], did not properly close FSD for regions near a critical point in the progress variable field. These critical points feature flame front interactions, hence they have $|\nabla c| \approx 0$ while having high Σ. These cause $\overline{|\nabla c|}/|\nabla \bar{c}| \to \infty$, making these regions hard to model and no existing FSD model has been shown to correctly represent the regions.

The inability of FSD models to capture flame behaviour near critical points and the complexity of closure models leads to the first open question that this work will attempt to address: What are the most important quantities captured in LES? Identifying these input parameters is an important step to understanding what flame physics should be incorporated to improve current models or develop new ones. To do this, Machine Learning (ML) algorithms can be helpful.

Machine learning has shown promise in combustion modelling [10]. For example, extensive *a priori* studies have been performed to demonstrate that deep learning models can provide accurate closures for turbulent combustion models [11,12]. However, many ML methods tend to be uninterpretable, thereby offering little insight towards the discovery of physical properties and earning their reputation as *black-box models*. However, not all ML models deserve this reputation. For instance, random forest models [13] can provide significant information on the importance of individual input components, and have previously been employed towards discovering closure models in turbulent combustion DNS [14].

Another weakness of ML-generated models is shared with many models developed in *a priori* studies — the stability of the model is not guaranteed in *a posteriori* simulations, and the model may cause the simulation to become unstable. For example, Wu et al. [15] found that an ML model developed to predict the Reynolds stresses using *a priori* data became unstable due to accumulation of small errors when used in Reynolds Averaged Navier Stokes simulations and treatment was required. This leads to the second open question this work will attempt to address: is it possible to run an LES with combustion models derived from ML models while minimising stability issues?

The goal of this work is therefore twofold: (1) identify the most important model input parameters that contribute towards improved predictions of FFFD and FSD using ML models and (2) investigate whether such ML-derived closure models are stable when used in *posteriori* LES. To this end, ML models are generated and their performance tested against the Charlette model [9], which was found by Panek et al. [5] to have the least error in FSD predictions compared to other models. Physical insights with the importance of the features of the ML models are generated. Finally, LES are performed to assess the robustness of the ML models in *a posteriori* applications.

2. Methods

2.1. A priori study

This study considers a turbulent jet flame configuration employing a finite rate chemistry approach (FRC) from the BLASTNet database [16]. The configuration is that of a DNS of a stoichiometric methane-fuelled jet flame at atmospheric pressure preheated to 700 K [17]. The jet Reynolds number based on the inlet diameter is 5,300 with a Mach number of 0.36. The unstretched laminar flame speed, $s_L = 1.85$ m/s, and the thermal flame thickness of an unstretched laminar flame, $\delta = (T_b - T_u)/\max(dT/dx) = 292$ μm, which results in a Karlovitz number of 15.43, and the flame lies in the thin reaction zone. Further details can be found in Brouzet et al. [17].

Gaussian filtering was applied to the DNS dataset to obtain $\overline{|\nabla c|}$, subgrid turbulent velocity, $u'_\Delta = \sqrt{(\widetilde{u_i^2} - \tilde{u}_i^2)/3}$, where u is the velocity, $|\nabla \bar{c}|$, $\bar{c}$, and the resolved density-weighted curvature term of displacement speed, $(\rho s_{d,c})_r$, defined as:

$$(\rho s_{d,c})_r = -\overline{\rho D_c} \nabla \cdot \left(-\frac{\nabla \bar{c}}{|\nabla \bar{c}|} \right) = -\overline{\rho D_c} \nabla \cdot \boldsymbol{n}_r, \tag{4}$$

where the resolved curvature is defined as $\nabla \cdot \boldsymbol{n}_r = \nabla \cdot (-\nabla \bar{c}/|\nabla \bar{c}|)$, and $\boldsymbol{n}_r$ is the resolved normal vector pointed towards the unburned gas. The progress variable is defined using the sensible enthalpy, $c = (h_s - h_{s,u})/\Delta h_s$, where Δh_s is the difference between the burnt and unburned gas h_s. The main reason for choosing u'_Δ and $|\nabla \bar{c}|$ as features in the ML modelling is because they are common features in existing algebraic models. Filter sizes, $\Delta = 0.5\delta$, 1δ, and 2δ were used, T is the temperature, T_b is the burnt gas temperature, and T_u is the unburned gas temperature. $|\nabla \tilde{c}|$, $\tilde{c}$, and the resolved curvature part of $(\rho s_{d,c})_r$ are directly calculable from LES, while u'_Δ is usually extracted from turbulence subgrid scale models. Note that $\bar{c}$ and $\tilde{c}$ can be related to each other using models such as the one described by Chakraborty and Cant [3], but it was found that for this dataset $\bar{c} \approx \tilde{c}$.

2.2. Additional features

The important parameters that could be used as features are determined in this section. Using theoretical arguments, a framework is built to examine the importance of these features in modelling FFFD.

Firstly, the displacement speed for an isosurface defined by sensible enthalpy is [6]:

$$s_d = \frac{1}{\rho |\nabla h_s|} \left[\dot{\omega}_c \Delta h_s + \frac{Dp}{Dt} + \nabla \cdot (\lambda \nabla T) - \nabla \cdot \left(\rho \sum_{k=1}^{N} h_{s,k} Y_k \boldsymbol{V}_k \right) + \boldsymbol{\tau} : \nabla \boldsymbol{u} \right], \tag{5}$$

where λ is the thermal conductivity, p is the pressure, $h_{s,k}$ is the sensible enthalpy for species k for a gas with a total number of species N, Y_k is the mass fraction of species k, $\boldsymbol{V}_k$ is the mass diffusion velocity of species k, and τ is the viscous stress tensor.

Substituting Eq. (5) into $\overline{\rho s_d |\nabla c|}$ yields the features that the FFFD is dependent on. The supplementary material contains the full derivation; the most important steps are stated here. Assuming that c_p is constant

Table 1
Models investigated. The RMSE values are non-dimensionalised by δ for $\overline{|\nabla c|}$, and $1/\dot{\omega}_{c,\max}$ for $\overline{\rho s_d|\nabla c|}$.

Model	Algorithm	Feature	Model output	Target variable	RMSE model	RMSE Charlette
1	RF	u'_Δ, Δ	$\Xi_{\mathrm{ML}}\|\nabla\bar{c}\|$	$\overline{\|\nabla c\|}$	0.377	0.236
2	RF	$u'_\Delta, \Delta, \bar{c}, \|\nabla\bar{c}\|, (\rho s_{d,c})_r$	$(\overline{\|\nabla c\|})_{\mathrm{ML}}$	$\overline{\|\nabla c\|}$	0.067	0.236
2a	RF	$u'_\Delta, \Delta, \bar{c}, \|\nabla\bar{c}\|, (\rho s_{d,c})_r$	$\rho_u s_L(\overline{\|\nabla c\|})_{\mathrm{ML}}$	$\overline{\rho s_d\|\nabla c\|}$	0.264	0.368
3	FGM	$\Delta, \bar{c}, (\rho s_{d,c})_r$	$(\overline{\rho s_d\|\nabla c\|})_{\mathrm{FGM}}$	$\overline{\rho s_d\|\nabla c\|}$	0.338	0.368
4	RF	$u'_\Delta, \Delta, \bar{c}, \|\nabla\bar{c}\|, (\rho s_{d,c})_r$	$(\overline{\rho s_d\|\nabla c\|})_{\mathrm{ML}}$	$\overline{\rho s_d\|\nabla c\|}$	0.129	0.368

to decompose the diffusion into normal and tangential diffusion, the FFFD can be written as:

$$\overline{\rho s_d|\nabla c|} = \bar{\dot{\omega}}_c + \overline{\frac{\partial}{\partial n}\left(\rho D_c \frac{\partial c}{\partial n}\right)} - \overline{|\nabla c|\rho D_c \nabla\cdot\frac{\nabla c}{|\nabla c|}}, \tag{6}$$

where n is the direction normal to the flame front. Assuming that a 1D FPF can be used to approximate the reaction and normal diffusion terms, it is possible to show that FFFD must be a function of $\bar{c}$, Δ, and $\overline{\rho s_{d,c}|\nabla c|}$, where $s_{d,c}$ is the tangential diffusion term. Considering $\overline{\rho s_{d,c}|\nabla c|}$ for high and low curvature flamelets separately, and assuming that the resolved-scale curvature contains information about the subgrid scale curvature finally yields $\overline{\rho s_d|\nabla c|} = f(\bar{c}, |\nabla\bar{c}|, u'_\Delta, (\rho s_{d,c})_r, \Delta)$. The assumption that the resolved curvature contains information regarding subgrid scale curvature will be revisited in Section 3.2. We note that the features $\bar{c}$ and $(\rho s_{d,c})_r$ are not usually considered in algebraic FSD models, and we will examine the importance of including these quantities.

2.3. Machine learning

The total number of flow-through times of the turbulent jet is $\tau_{ft} = 1.5$. The data consisted of a total of 550 cross-sections through the mid-plane of the flame. The ML models were trained on the first 0.8 τ_{ft}, and then $0.5\tau_{ft}$ were discarded to ensure the turbulence in the first $0.8\tau_{ft}$ does not directly affect the remaining 0.2 τ_{ft}, which was the testing dataset.

The flame region corresponding to $0.29 < c < 0.68$ was used in the analysis as this range has > 50% of the maximum heat release rate in a 1D unstrained FPF. This range was chosen to represent the flame and remove the unburned and burned regions because these regions comprise the majority of the domain and hence would significantly skew the training dataset. A random sample of 100,000 points was used to train the ML models.

The Random Forest (RF) [13] algorithm was used to generate the ML models using ten estimators with a maximum depth of 100. These parameters were chosen such that they err towards generating low bias models to ensure that the variance between different parts of the flame is captured, which may allow the ML models to capture statistically low probability occurrences such as flame-flame annihilation. They were built using Scikit-learn [18] version 0.3.21 in Python 3.10.6.

A summary of the models investigated is shown in Table 1. The values used to non-dimensionalise the Root Mean Square Error (RMSE) are δ and the maximum production rate of c in a 1D FPF, $\dot{\omega}_{c,\max}$. Model 1 is the RF model trained on Ξ data to target $\overline{|\nabla c|} = \Xi_{\mathrm{ML}}|\nabla\bar{c}|$, where subscript ML denotes the variable that the ML models are trained on. The RMSE is calculated based on the deviation of the output of the models to the target variable. Model 2 is the RF model trained on $\overline{|\nabla c|}$ directly. Model 2a is the output of Model 2 multiplied by $\rho_u s_L$. The Flamelet Generated Manifold (FGM)-like model, Model 3, will be discussed in Section 3.2. Model 4 is trained on filtered $\rho s_d|\nabla c|$ values.

2.4. LES of the turbulent jet

A posteriori simulations of the present configuration are performed by employing an unstructured compressible finite-volume solver [19], with integrated ML capabilities provided by the `OpenCV` library [20]. A central scheme, which is 4th-order accurate on uniform meshes, is used along with a 2nd-order ENO scheme that is activated by a gradient-threshold sensor. A strong stability preserving 3rd-order Runge–Kutta (SSP-RK3) scheme is used for time advancement. The turbulent stresses are accounted for via the Vreman [21] sub-grid scale (SGS) model and both turbulent Prandtl and Schmidt numbers are set to 0.7. In this work, we compare LES that employ two FFFD models: (i) one based on the Charlette model, and (ii) an ML model. In ML models, out-of-distribution predictions can pose a significant challenge to numerical stability [10]. To ameliorate this issue, the ML model is only activated when the input conditions are within the range of the progress variable $\bar{c}$ found in training data. Outside this range, the Charlette model is employed. We note that this dynamic combustion submodel assignment approach has been validated in previous work [19,22].

3. Results and discussion

3.1. Model results for FSD

Table 1 shows the RMSE for Model 1, which uses only u'_Δ and Δ as features to model $\overline{|\nabla c|} = \Xi|\nabla\bar{c}|$. The RMSE is actually larger than the RMSE using Charlette's model. This shows that Charlette's model is accurate at predicting the wrinkling factor using information from u'_Δ and Δ, and this translates to a more accurate prediction of FSD.

Fig. 1 shows a scatter plot from the test dataset of the FSD plotted against $|\nabla\bar{c}|$, coloured by $\tilde{c}$. It has been shown before that algebraic models are not able to capture flame interaction events, which are events with high $\overline{|\nabla c|}$, low $|\nabla\bar{c}|$, and low $\bar{c}$ values [5], shown in Region A. Both the Charlette model and Model 1 show an inability to capture the shape of the distribution of points, especially Region A. However, Model 2, which is trained on $u'_\Delta, \Delta, \bar{c}, |\nabla\bar{c}|$, and $(\rho s_{d,c})_r$, is able to capture Region B, and it has a lower RMSE of 0.067 compared to the Charlette model's 0.236 and Model 1's 0.377. This comparison shows that information from the additional parameters is useful to model the flame interaction events.

3.2. Model results for FFFD

Next, the ability of Model 3, which is capable of detecting flame interaction events, is tested for its ability to predict FFFD by multiplying its results by $\rho_u s_L$ (Model 2a). Fig. 2 shows the distribution of the FFFD for Models 2a, 3, and 4.

It is evident that both the Charlette model and Model 2a do not recover the DNS scatter distribution for FFFD. Specifically, the high FFFD with low $|\nabla\bar{c}|$ (Regions C) and the negative FFFD (Region F) are not captured. The tendency for low $\tilde{c}$ to appear at high FFFD values for a fixed $|\nabla\bar{c}|$ is captured by Model 2a but not by the Charlette model. The main reason these FSD models do not capture the distribution correctly is because of the direct multiplication of the model results with $|\nabla\bar{c}|$ and the fact that $(\rho s_d)_s \approx \rho_u s_L$ is positive, which forces the shape of the distribution to follow an approximately straight line.

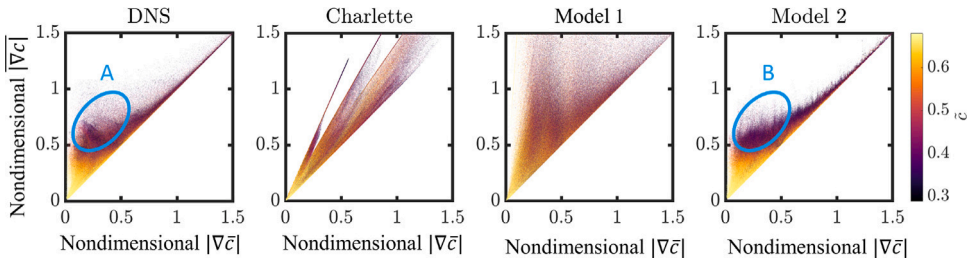

Fig. 1. Scatter plot of FSD versus $|\nabla \bar{c}|$, both non-dimensionalised by δ, coloured by $\tilde{c}$, for the filtered DNS results and the FSD predicted by the Charlette model, Model 2, and Model 3 for all filter sizes $\Delta/\delta = 0.5, 1.0, 2.0$.

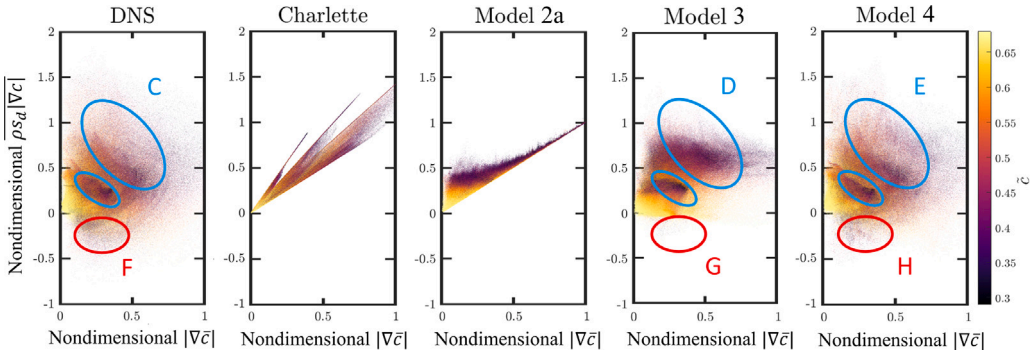

Fig. 2. Scatter plot of FFFD nondimensionalised by $1/\dot{\omega}_{c,\max}$ versus $|\nabla \bar{c}|$ nondimensionalised by $\rho_u s_L/\dot{\omega}_{c,\max}$, coloured by $\tilde{c}$, for the filtered DNS results and the FFFD predicted by Charlette model, Model 2a, Model 3, and Model 4 for all filter sizes $\Delta/\delta = 0.5, 1.0, 2.0$.

Instead of assuming that FFFD has a linear relationship with $|\nabla \bar{c}|$, the reaction and normal diffusion terms can be approximated by a 1D FPF, which is a common assumption used in FGM models [23]. It is interesting to compare this approach against that of models generated by ML. In FGM, usually only the reaction rate term is tabulated while the diffusion term is modelled. In this work, because the diffusion term has been split into normal and tangential terms, it is possible to use a 1D FPF to tabulate the results for the normal diffusion term as well.

Model 3 is the FGM-like model given as:

$$\overline{\rho s_d |\nabla c|} = \left(\bar{\dot{\omega}}_c + \overline{\frac{\partial}{\partial n}\left(\rho D_c \frac{\partial c}{\partial n}\right)} \right)_{\text{tab}} k_\Delta + (\rho s_{d,c})_r, \tag{7}$$

where the subscript tab denotes that the terms are tabulated from a 1D FPF solution, and $k_\Delta = 1$ if $\Delta/\delta \leq 1$ and $k_\Delta = \delta/\Delta$ if $\Delta/\delta > 1$. The k_Δ term is multiplied to the tabulation model to account for the effect of the filter size. Specifically, as the filter size increases, the volume increases by a magnitude of Δ larger compared to the surface area for a planar flamelet, which means that more unburned and burnt gases are taken into account by the filter. These unburned and burnt regions have FFFD = 0, hence they reduce the FFFD by a proportion approximately equal to the difference between volume and surface area increase. The tangential diffusion term is not modelled using a 1D FPF, so only the resolved curvature is added directly.

Fig. 2 shows the scatter plot of FFFD. The reason there are multiple regions with high FFFD (circled in blue) is because the results from multiple Δ are plotted on top of each other. This is more clearly seen in the supplementary material, which shows the results for each separate Δ. The FFFD for Model 3 and its RMSE is 0.338, which is lower than Charlette's model but higher than Model 4. Without the use of ML algorithms, it is able to capture the approximate spread of points and is also able to capture Region C. In addition, the k_Δ term is able to model the effect of a more compact distribution of FFFD as Δ increases for $\Delta/\delta > 1$, but is unable to capture the effect of filtering for filter sizes smaller than δ. This model is also not able to capture regions with low $\tilde{c}$ and low FFFD (Region F). The performance of the ML models is tested in Section 3.3.

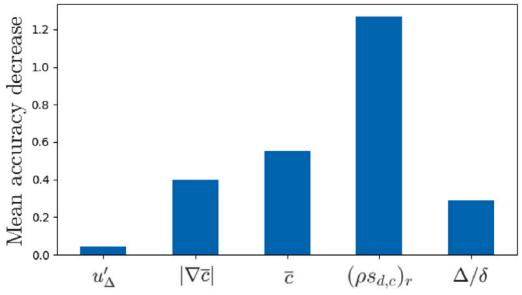

Fig. 3. Feature importances using permutation on Model 4.

In contrast to the performance of Model 3, Model 4 is capable of recovering the distribution of both Regions C and Region F in the DNS FFFD (see Fig. 2) and has a low RMSE of 0.129 compared to 0.368 and 0.264 for the Charlette model multiplied by $\rho_u s_L$ and Model 3, respectively.

To identify the most important features, Fig. 3 shows the importance of each feature in Model 4 using the decrease of the accuracy when a single feature value is randomly permuted [13]. It shows that the $(\rho s_{d,c})_r$ term is most important, followed by $\bar{c}$, $|\nabla \bar{c}|$, Δ, and u'_Δ, respectively. This shows that information from the resolved curvature can be used to model curvature effects on FFFD, which strengthens the arguments of models that use resolved scale curvature to close the $\overline{(\rho s_d)_s}$ term [3,7]. This quantity depends on gradients and other subgrid quantities, which may be inaccurate in LES due to numerical and modelling errors hence the effect of these errors should be investigated further in future work.

A weakness of the feature importance approach is that it is possible for statistically related features to have their importances split between them, which makes the calculated importance lower than its true value. To address this, Fig. 4 shows the hierarchical clustering dendrogram of Model 4 which was constructed by calculating Ward's method [24]

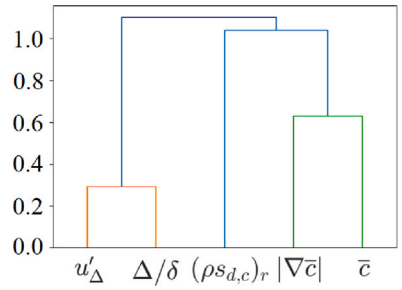

Fig. 4. Hierarchical clustering dendrogram of Model 4.

from the Spearman correlation coefficient, which measures how monotonically correlated two variables are with each other. The lower the value, the smaller the distance is between two variables hence the more correlated they are. It shows that u'_Δ and Δ are strongly correlated with each other, hence together these variables have similar importance to $|\nabla\bar{c}|$. This is likely because a larger filter size will have larger and stronger turbulent eddies within the filter region. This is a consequence of the geometry of the turbulent jet since the integral length scale is $\approx$ 1.9 times larger than δ, and the filter sizes considered are 0.5–2 δ, hence increasing filter sizes means it filters more energetic eddies.

Overall, the resolved curvature term is the most important and is also uncorrelated to the other features. This suggests that the addition of curvature information will significantly improve FFFD models. Furthermore, $\bar{c}$ has been shown to be of approximately equal importance to the FSD, which suggests that the dependence of FFFD on the location within the flame is also important to model.

To assess the sensitivity of the results to different ML algorithms, a Multi-Layer Perceptron Regressor (MLPR) model [25] was trained with 10,000 maximum iterations using the same features and output as RF model 4. Results for the FFFD distribution and the feature importance are reported in the supplementary material. Overall, the FFFD predictions, RMSE, and feature importance are similar to the RF model, which suggests that neural networks may be suitable for combustion modelling.

3.3. LES of the turbulent jet

To assess the stability and effects of the newly devised ML approaches in a practical sense, an *a posteriori* study was conducted by running an LES of the turbulent premixed flame using a hybrid MLPR-Charlette approach for combustion modelling. The MLPR algorithm (discussed in Section 3.2) generates a lightweight Artificial Neural Network (ANN) which is fast and more practical in terms of deployment in production runs.

Two different mesh resolutions are used, both with the Charlette model and the ML model. A DNS-like mesh with a resolution of $\Delta/\delta = 0.125$ gives an indication of the ML model performance when SGS terms are negligible, while a mesh with a characteristic grid size $\Delta/\delta = 0.25$ is informative in an LES context. Note that a coarser grid, as the ones found typically in LES for industrial applications, was impractical in the present configuration because it led to too few points at the inflow which results in strong dampening of turbulent fluctuations.

Fig. 5 displays instantaneous solutions of the LES, showing the progress variable at the centerplane obtained with the Charlette and ML model after 80,000 iterations, corresponding to 1 ms of physical time. In Fig. 5(b-c), we observe that the LES performed with the ML model does not demonstrate visible signs of physical inconsistency and remains numerically stable, even after $\approx 2\tau_{ft}$. This result demonstrates that the present *a posteriori* hybrid ML combustion modelling approach can avoid issues tied to physically inconsistent out-of-distribution predictions that often result in numerical blow-up [26].

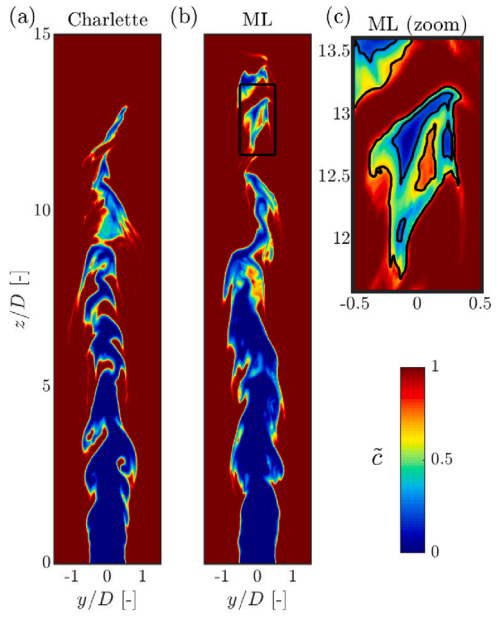

Fig. 5. Instantaneous solutions of the progress variable field for the simulations performed with (a) the Charlette model and (b-c) the hybrid MLPR model using the DNS mesh. The isolines in (c) show the region in which the MLPR model is used, i.e., in the range $0.29 < \tilde{c} < 0.68$.

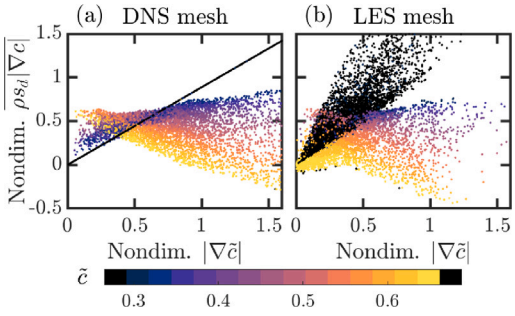

Fig. 6. Scatter plot of the FFFD term (non-dimensionalised by $1/\dot{\omega}_{c,\max}$) versus $|\nabla\tilde{c}|$ (non-dimensionalised by $\rho_u s_L/\dot{\omega}_{c,\max}$) and coloured by $\tilde{c}$, for the *a-posteriori* LES results. The black points represent regions where the Charlette model is used, i.e., for $\tilde{c} < 0.29$ and $\tilde{c} > 0.68$.

Fig. 6 shows the FFFD term as a function of $|\nabla\tilde{c}|$ for ML model simulations. The regions where the ML model is used are represented by the coloured points while the Charlette model is used for the black points. Different behaviour is clearly observed between the two approaches. With both meshes, the ML approach leads to a lower FFFD compared to the algebraic model for large $|\nabla\tilde{c}|$ values, with a significant portion featuring negative FFFD. However, for small $|\nabla\tilde{c}|$, the ML model predicts a higher FFFD, consistent with the *a priori* results.

In light of those results, the flame length is investigated. Fig. 5 suggests that the ML model leads to a slightly longer flame, with fresh reactants further downstream. To confirm this observation, Fig. 7 shows the average (left) and variance (right) of the FFFD term, integrated in the radial direction and shown as a function of the streamwise location. The coarse mesh results show that the ML and algebraic models yield similar statistics, indicating that the ML approach performs reasonably well in a *a posteriori* LES context.

The results obtained with the DNS mesh show that the FSD and ML results are both quite different from the simulation employing FRC, which can be explained by the inherent differences in the combustion modelling. With the ML approach, however, the amplitude of the source

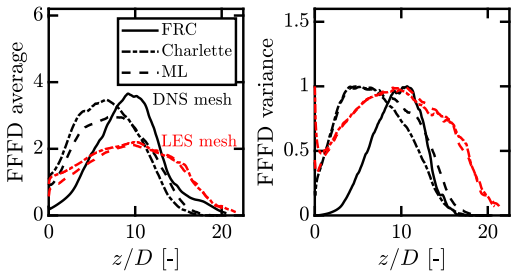

Fig. 7. Radially integrated average (left) and normalised variance (right) of the FFFD term.

term is smaller in the first half of the domain. Consequently, more pockets of fresh gases are present downstream, leading to a flame brush that extends by 10%, and which is marginally closer to the FRC results. The normalised variance is similar for $x/D < 8$ between the ML and Charlette cases, suggesting that the differences arise at the flame tip, where most of the flame interaction events are located [27]. As mentioned in Section 1, legacy algebraic models were found to typically over-predict the FSD term, and perform poorly for flame interaction events. In contrast, the MLPR model results in a lower FFFD average, and significantly different flame tip behaviour are encouraging in that regard, highlighting that core modelling deficiencies might be addressed by data-driven methods.

While these results are encouraging, they also highlight the challenges to couple the ML model in LES. For instance, the flame lengthening as the LES mesh increases is a sign of a non-optimal SGS model, and the hybrid ML-Charlette framework could be improved to a full-ML approach. Analysis of the flame dynamics and flame/turbulence interaction could shed more light on the necessary improvements to yield an ML combustion model that fully addresses the limitations of the legacy approaches.

4. Conclusions

This study investigated the origins of inconsistencies obtained with the wrinkling factor, FSD, and FFFD approaches by training ML models on DNS data of a turbulent premixed jet flame. First, theoretical analysis showed that the filtered progress variable $\overline{c}$, the norm of its gradient $|\nabla\overline{c}|$, and the resolved curvature term of the FFFD $(\rho s_{d,c})_r$, should all be relevant features to predict the FFFD term. While the well-known Charlette model [9] was able to predict the wrinkling factor more accurately compared to a random forest approach, it performed worse for the FSD term. Testing on the different modelling terms revealed two main results: (1) the approximation $(\rho s_d)_s = \rho_u s_L$ yielded a significant accuracy decrease for the FFFD term, and (2) ML models using $\overline{c}$ and $(\rho s_{d,c})_r$ as input parameters, in addition to the usual features in existing algebraic models, captured the FFFD distribution and reduced the overall FFFD modelling error. Furthermore, it was shown that $(\rho s_{d,c})_r$ is the most important feature in capturing FFFD, followed by $\overline{c}$, $|\nabla\overline{c}|$, u'_Δ, and Δ. Of note, an FGM-like model using these features had a better prediction than the Charlette model, but performed worse than the ML approaches, which is likely explained by the absence of subgrid curvature modelling. Finally, one of the ML models was employed *a posteriori* in LES and was compared against simulations using the Charlette model. The ML approach was stable and was able to address some of the deficiencies of the algebraic approach, notably regarding the FFFD distribution and the flame length.

Novelty and Significance Statement

This research is of importance because it answers fundamental and practical questions related to the use of combustion modelling approaches, specifically the Flame Surface Density (FSD) and the Filtered Flame Front Displacement (FFFD) models, by means of Machine Learning (ML) algorithms. From a fundamental aspect, we show that two features which are typically not considered as inputs in combustion models, i.e., the progress variable and the resolved curvature, are key to consider for improved predictions of the model, more so than features which are typically used in FSD modelling, i.e., u'_Δ, Δ, and $|\nabla c|$. From a practical standpoint, we demonstrate a framework to use the developed ML combustion model a posteriori in a LES, without any stability issues. Overall, these findings are key to guide further traditional and ML improvement efforts on combustion models.

CRediT authorship contribution statement

Jen Zen Ho: Designed research, Developed methodology, Performed analysis, Wrote paper. **Mohsen Talei:** Designed research, Revised paper. **Davy Brouzet:** Designed research, Performed simulations, Performed analysis, Revised paper. **Wai Tong Chung:** Designed research, Performed simulations, Developed methodology, Revised paper. **Pushan Sharma:** Developed methodology. **Matthias Ihme:** Designed research, Revised paper.

Declaration of competing interest

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

Acknowledgements

This research was supported by the Australian Government National Collaborative Research Infrastructure Strategy, with computational resources provided by the National Computational Infrastructure and Pawsey Supercomputing Centre through the National Computational Merit Allocation Scheme. Computational resources from the National Energy Research Scientific Computing Center are also acknowledged. MT acknowledges the Australian Research Council through the DECRA Fellowship (DE180100416). This research is also supported by the U.S. Department of Energy, National Nuclear Security Administration, USA (DE-NA0003968) and the AeroAcoustics Research Consortium, USA.

Appendix A. Supplementary data

Provided in 'SuppMat.pdf'.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.proci.2024.105311.

References

[1] M. Boger, D. Veynante, H. Boughanem, A. Trouvé, Direct numerical simulation analysis of flame surface density concept for large eddy simulation of turbulent premixed combustion, Symp. Combust. Proc. 27 (1998) 917–925.
[2] A. Trouve, T. Poinsot, The evolution equation for the flame surface density in turbulent premixed combustion, J. Fluid Mech. 278 (1994) 1–31.
[3] N. Chakraborty, R.S. Cant, A priori analysis of the curvature and propagation terms of the flame surface density transport equation for large eddy simulation, Phys. Fluids 19 (2007).
[4] N. Chakraborty, M. Klein, A priori direct numerical simulation assessment of algebraic flame surface density models for turbulent premixed flames in the context of large eddy simulation, Phys. Fluids 20 (2008).
[5] P. Panek, D. Brouzet, M. Talei, R.L. Gordon, A priori assessment of flame surface density modelling for large-eddy simulation of sound generation by turbulent premixed flames, Combust. Flame 112143 (2022).
[6] T. Poinsot, D. Veynante, Theoretical and Numerical Combustion, second ed., 2005, p. 21.
[7] E. Suillaud, K. Truffin, O. Colin, D. Veynante, Direct numerical simulations of high Karlovitz number premixed flames for the analysis and modeling of the displacement speed, Combust. Flame 236 (2022) 111770.
[8] C. Fureby, A fractal flame-wrinkling large eddy simulation model for premixed turbulent combustion, Proc. Combust. Symp. 30 (2005) 593–601.

[9] F. Charlette, C. Meneveau, D. Veynante, A power-law flame wrinkling model for LES of premixed turbulent combustion part I: Non-dynamic formulation and initial tests, Combust. Flame 131 (2002) 159–180.
[10] M. Ihme, W.T. Chung, A.A. Mishra, Combustion machine learning: Principles, progress and prospects: Combustion machine learning, Prog. Energy Combust. Sci. 91 (2022).
[11] C.J. Lapeyre, A. Misdariis, N. Cazard, D. Veynante, T. Poinsot, Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates, Combust. Flame 203 (2019) 255–264.
[12] A. Seltz, P. Domingo, L. Vervisch, Z.M. Nikolaou, Direct mapping from LES resolved scales to filtered-flame generated manifolds using convolutional neural networks, Combust. Flame 210 (2019) 71–82.
[13] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[14] W.T. Chung, A.A. Mishra, M. Ihme, Interpretable data-driven methods for subgrid-scale closure in LES for transcritical LOX/GCH4 combustion, Combust. Flame 239 (2022).
[15] J.L. Wu, H. Xiao, E. Paterson, Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework, Phys. Rev. Fluids 7 (2018).
[16] W.T. Chung, K.S. Jung, J.H. Chen, M. Ihme, BLASTNet: A call for community-involved big data in combustion machine learning, Appl. Energy Combust. Sci. 12 (2022).
[17] D. Brouzet, M. Talei, M.J. Brear, B. Cuenot, The impact of chemical modelling on turbulent premixed flame acoustics, J. Fluid Mech. 915 (2021) 1–33.
[18] F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, D. Édouard, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[19] W.T. Chung, A.A. Mishra, N. Perakis, M. Ihme, Data-assisted combustion simulations with dynamic submodel assignment using random forests, Combust. Flame 227 (2021) 172–185.
[20] G. Bradski, The OpenCV library, Dr. Dobb's J. Softw. Tools (2000).
[21] A.W. Vreman, An eddy-viscosity subgrid-scale model for turbulent shear flow: Algebraic theory and applications, Phys. Fluids 16 (10) (2004) 3670–3681.
[22] D. Mohaddes, D. Brouzet, M. Ihme, Cost-constrained adaptive simulations of transient spray combustion in a gas turbine combustor, Combust. Flame 249 (2023) 112530.
[23] J.A. van Oijen, L.P. de Goey, Modelling of premixed laminar flames using flamelet-generated manifolds, Combust. Sci. Technol. 161 (2000) 113–137.
[24] J.H. Ward, Hierarchical grouping to optimize an objective function, J. Amer. Statist. Assoc. 58 (1963) 236–244.
[25] F. Murtagh, Multilayer perceptrons for classification and regression, Neurocomputing 2 (1991) 183–197.
[26] Y. Guan, A. Chattopadhyay, A. Subel, P. Hassanzadeh, Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning, J. Comput. Phys. 458 (2022) 111090.
[27] D. Brouzet, A. Haghiri, M. Talei, M.J. Brear, Annihilation events topology and their generated sound in turbulent premixed flames, Combust. Flame 204 (2019) 268–277.