Accepted for publication in Combustion and Flame on 21/October/2020

Application of machine learning for filtered density function closure in MILD combustion

Zhi X. Chen[a,b,*], Salvatore Iavarone[a], Golnoush Ghiasi[a], Veeraraghavan Kannan[a], Giuseppe D'Alessio[c,d,e], Alessandro Parente[c,d], Nedunchezhian Swaminathan[a]

[a]*Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK*
[b]*Robinson College, University of Cambridge, Grange Road, Cambridge CB3 9AN, UK*
[c]*Université Libre de Bruxelles, Avenue F.D Roosevelt 50, 1050 Brussels, Belgium*
[d]*Combustion and Robust Optimization Group (BURN), Université Libre de Bruxelles and Vrije Universiteit Brussel, 1050 Brussels, Belgium*
[e]*CRECK Modeling Lab, Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy*

Abstract

A machine learning algorithm, the deep neural network (DNN)[1], is trained using a comprehensive direct numerical simulation (DNS) dataset to predict joint filtered density functions (FDFs) of mixture fraction and reaction progress variable in Moderate or Intense Low-oxygen Dilution (MILD) combustion. The important features of the DNS cases include mixture fraction variations, turbulent mixing lengths, exhaust gas recirculation (EGR) dilution levels, etc., posing a great challenge for data-driven modelling. The DNN architecture is built and optimised with extreme care to achieve high robustness and accuracy, resorting to dimensionality reduction techniques such as principal component analysis (PCA) to identify and remove the outliers in the training data. To better interpret the predictive ability

*Corresponding author.
Email address: zc252@cam.ac.uk (Zhi X. Chen)

[1]The DNN code and a sample training dataset are available for public access at https://github.com/zhixchen/ml-mild-combustion.

of the DNN, two analytical joint FDF models respectively using two independent β and *copula* distributions, are also employed for a detailed comparison with the DNS data.

The FDFs in MILD combustion behave differently compared to those in conventional flames because the reaction zones are more distributed. They generally exhibit non-regular (neither Gaussian nor bi-modal) distributions and strong cross correlations, which cannot be captured adequately by the analytical models. However, the DNN is well suited for this physico-chemically complex problem and its predictions are in excellent agreement with the DNS data for a broad range of mixture conditions and filter sizes. Furthermore, *a priori* assessment is conducted for filtered reaction rate closure. It is found that the DNN model significantly outperforms the analytical models for all cases showing very good predictions for the filtered reaction rate for a range of filter sizes. The DNN prediction improves as the filter size becomes larger than the characteristic reaction zone thickness while the analytical models works relatively better for smaller filter sizes. This is a clear advantage for the DNN to be used in practical LES applications.
Keywords: Machine learning; Deep neural networks; Filtered density function; MILD combustion; Subgrid scale modelling

---

1. Introduction

The increasingly stringent regulations on pollutant emission drive novel technologies towards cleaner combustion regimes. Moderate or Intense Low-oxygen Dilution (MILD) combustion has been widely recognised as a promising candidate to lead the upcoming *green* revolution of combustion devices [1, 2]. Due to the small temperature rise during combustion (typically few hundreds of Kelvin),

the formation of major combustion pollutants such as NO_x is largely reduced. In addition, MILD combustion is known to have smaller temperature gradients between the reactants and products, resulting in more homogeneous and distributed reaction zones [3], which can be beneficial for practical combustion systems [4].

Because of these promising features, MILD combustion research has become increasingly popular in recent years, including early experimental studies [5–7], and also numerical investigations using perfectly stirred reactor (PSR) [8], laminar counterflow [9, 10] and freely propagating premixed flames [11]. While these pioneering studies mainly focused on the overall burning characteristics, laminar reaction zone behaviours and pollutant formation, recent efforts have shifted towards understanding MILD combustion in turbulent flows and seeking viable modelling approaches (see [4] for a review). Direct numerical simulation (DNS) is a powerful tool for fundamental studies and model development. It has been applied for auto-igniting mixing layer [12] and EGR-type [13] MILD configurations. Minamoto *et al.* [14, 15] showed that the MILD reaction zones are broadly distributed with strong interactions, which differs significantly from the classical sheet-like flamelet structures of conventional combustion. Moreover, they also found that autoigniting and propagating zones coexist in MILD combustion and their intensities depend on local behaviours of scalars and their gradients. These observations were further confirmed in recent DNS studies [16–18] involving mixture fraction variations to elucidate the effects of scalar mixing on the reaction zone behaviours. These subtleties and complexities suggest that the turbulence–chemistry interaction models developed for conventional combustion may be inadequate for MILD combustion.

Among the numerous laboratory-scale MILD burners, the Jet in Hot Co-flow

(JHC) configurations [19, 20] have received major attentions from the modelling community. The Adelaide JHC burner [20], for instance, has been simulated with converging good results by many research groups using a full range of state-of-the-art approaches (not detailed here, see [4] and Refs. therein) with both Reynolds-averaged Navier-Stokes (RANS) and large eddy simulation (LES) methodologies. However, other types of MILD burners including those with cyclonic [21] and reversed [22] EGR flow motions, which are more industry-relevant, have mostly been modelled using the computationally more affordable tabulated chemistry approaches [23–26] with quite limited success. The predictive discrepancies can be attributed to two main modelling aspects. Most tabulation approaches involve probability density functions (PDFs) for RANS, or filtered density functions (FDFs) for LES, and the presumed shapes, e.g., β-PDF, may not represent the correct scalar statistical behaviour over the homogeneous and distributed MILD reaction zones. Furthermore, when the burnt mixture is compositionally inhomogeneous, i.e., with mixture fraction variations, interactions between scalar fluctuations arise over a range of length scales, leading to significant statistical correlation [27]. This interaction is expected to be stronger under MILD conditions because of the broadened reaction zones. Therefore, the widely used statistical independence assumption could result in additional modelling errors. For LES, these statistical behaviours come into play through the FDF [28] involving some randomness and filter-size dependency [29], and hence pose further challenges to the modelling.

The present work aims to address these issues using the DNS datasets generated in [16–18] and to shed light on the FDF behaviours and subgrid modelling for LES of MILD combustion, which have not been contemplated in previous works.

First, a major focus is given to the FDF, which is essential for many modelling approaches. Given the complex and unconventional features of scalar statistical behaviours in MILD combustion, machine learning (ML) techniques are well suited for the prediction of FDF. A recent study by de Frahan *et al.* [30] compared different ML algorithms for FDF prediction in a conventional swirl methane−air premixed flame, and a favourable performance was shown for the feed-forward deep neural network (DNN). Another attractive option is convolutional neural network (CNN), well known for its excellent ability for image recognition and recently it was successfully applied for prediction of subgrid flame wrinkling [31]. Since FDF represents more of the statistical features than topology of the reaction zones, DNN is chosen for this study and we focus on feature selection/extraction, architecture and hyperparameter optimisation for the MILD combustion DNS datasets to achieve the best predictive performance for the FDF. The DNN predictions are compared against those obtained using classical β and *copula* (correlated multivariant [32]) functions. Moreover, *a priori* assessment of the filtered reaction rate closure using these different FDFs is performed to assess their respective accuracies.

The remainder of this paper is organised as follows. Section 2 describes the DNS datasets and the various FDF modelling approaches used with a particular focus on the DNN learning procedures. The FDF results are then compared and discussed in Section 3, followed by the *a priori* assessment of the filtered reaction rate closure in Section 4. Finally, the conclusions are summarised in Section 5.

2. Data analysis and modelling methodology

2.1. DNS datasets

DNS of MILD combustion with varying mixture fraction and internal EGR was performed and reported in [16], and the datasets generated were used to study the role of radicals [18] and the relative roles of autoignition and flame propagating phenomena [17] under MILD conditions. In this work, we use these datasets to explore the FDF modelling aspects of this combustion regime to facilitate future simulations of laboratory and industry scale burners. Detailed descriptions of the DNS methodology and initial data preprocessing can be found in [16], and only the essential information is provided here.

All three DNS cases, differing in mixing lengthscale and dilution level, are considered in this study, and their initial conditions are summarised in Table 1. The initial turbulence is the same for the three cases having an integral length scale of $\Lambda_0 \approx 4.12$ mm and a global velocity fluctuation of $u' \approx 16.66$ m/s. The corresponding turbulence and Taylor-microscale Reynolds numbers are $\mathrm{Re}_t \approx 96$ and $\mathrm{Re}_\lambda \approx 35$, respectively. The characteristic length for the mixture fraction field, l_Z, is about $1.67\Lambda_0$ for case AZ1, larger than $1.27\Lambda_0$ for AZ2. This means that the fuel/oxidiser mixing is faster and more homogeneous in case AZ2. The third case BZ1 has the same mixing length scale as AZ1 but with a higher dilution level of 2% O_2 (by volume) compared to that of 3.5% for cases AZ1 and AZ2. Thus, this leads to a smaller stoichiometric mixture fraction value and an overall lower reactivity for case BZ1 compared to the other two cases. The reaction progress variable fields based on normalised fuel mass fraction used for the three cases have similar statistics as listed in Table 1.

The DNS was performed inside a cubic numerical domain of size $L_x \times L_y \times L_z =$

Table 1: Summary of initial conditions for the DNS cases considered.

Case	l_Z/Λ_0	$X^{ox}_{O2}\%$	Z_{st}	$\langle Z\rangle$	$\sigma_Z/\langle Z\rangle$	$\langle c\rangle$	$\sigma_c/\langle c\rangle$
AZ1	1.67	3.5	0.01	0.008	1.05	0.56	0.46
AZ2	1.27	3.5	0.01	0.008	1.31	0.56	0.5
BZ1	1.67	2	0.0058	0.0046	1.23	0.56	0.46

$10\times10\times10$ mm. 512 uniformly distributed grid points were used for each physical direction and this resolution with $\delta x \approx 20$ μm is sufficient to resolve all turbulent flow and chemical lengthscales of interest. A modified mechanism of Smooke and Giovangigli [33] with OH^* chemistry [34] was used for the methane/air combustion considered. Detailed description and validation are in [16]. The high-order compressible DNS code SENGA2 was used. The inflow and non-reflective outflow boundary conditions were used for the x-direction and periodic conditions were employed for the transverse y- and z- directions. For each DNS case, simulation was first run for one flow-through-time, $\tau_f = L_x/U_{\text{in}}$, where $U_{\text{in}} = 20$ m/s is the inflow bulk mean velocity, to allow the transients to pass out of the domain. The subsequent about 60 snapshots spanning over $0.5\tau_f$ were then taken for statistics and these snapshots are used for the data training and subsequent analyses in this study.

2.2. FDF and presumed models

In transported filtered density function models, the FDFs of all compositional space scalars can be solved using various methods (e.g. Lagrangian particles [35], Eulerian stochastic fields [36] and multi-environment [37]), and the reaction source terms appear in closed form. However, such sophisticated modelling is quite computationally demanding and hence many combustion models, especially for those

widely used by industry, rely on a presumed statistical description of the subgrid scalar fluctuations to close the reaction source terms in the LES transport equations [38, 39]. Few key variables are often chosen to represent the leading-order physical processes that influence the reaction rates, such as fuel/oxidiser mixing, reaction progress, flame stretch/straining, non-adiabaticity, etc. The subgrid reaction rate can be obtained using an appropriate moment of the joint FDF of these variables. A major advantage of this approach is that only the statistical moments (usually the mean and variance) of the key variables need to be transported in the LES, whereas other quantities of interest can be retrieved from a tabulated manifold through post-processing. Therefore, the computational cost is significantly reduced, typically by orders of magnitude.

It is important to note that in the LES context, there is a conceptual difference between FDF and subgrid PDF [38]. The FDF for a given point at a given time can be extracted by applying a fine-grain filtering function to a single ensemble (or realisation) of the fully resolved DNS or experimental data [40]. It has all statistical properties of a PDF but also involves some randomness due to the unsteady nature of single realisations [28]. To remove this randomness, one needs to collect the samples in a given physical subgrid space at a given time over many ensembles having the same resolved fields [29]. This would require excessively long runtimes for either experiments or simulations with enormous data storage requirements, which may not be practical. In this work, similar to [27], filtering is applied to single DNS snapshots and thus what we obtain are FDFs with certain levels of randomness. However, this randomness is removed to a good extent if the training data for machine learning are selected over many DNS realisations at a statistically stationary state, which is discussed further in Section 3.2.

Two commonly used conditioning variables, Bilger mixture fraction, Z, and a temperature-based reaction progress variable, c_T, are used for this study, and the latter is defined as

$$c_T = \frac{T - T_u}{T_b(Z) - T_u}, \tag{1}$$

where T_u is 1500 K for all three cases listed in Table 1 as in the DNS initial conditions. The burnt temperature T_b is determined using the MILD Flame Element (MIFE) method [14], which depends on the local mixture fraction. Favre-filtered fields can be obtained by performing a low-pass *box* filtering procedure, e.g, for mixture fraction and a given filter width Δ:

$$\widetilde{Z}(\boldsymbol{x}, t) = \frac{1}{\overline{\rho}(\boldsymbol{x}, t)} \int_{\boldsymbol{x}-\frac{\Delta}{2}}^{\boldsymbol{x}+\frac{\Delta}{2}} \rho(\boldsymbol{x}', t)\, Z(\boldsymbol{x}', t) d\boldsymbol{x}', \tag{2}$$

where $\overline{\cdot}$ and $\widetilde{\cdot}$ denote the Reynolds and Favre filtering respectively, and ρ is the mixture density. The coordinate vectors, $\boldsymbol{x}$ and $\boldsymbol{x}'$, correspond to the reference frames of the entire DNS domain and the filter sub-space, respectively. The subgrid variance is obtained using

$$\widetilde{\sigma_Z^2}(\boldsymbol{x}, t) = \frac{1}{\overline{\rho}(\boldsymbol{x}, t)} \int_{\boldsymbol{x}-\frac{\Delta}{2}}^{\boldsymbol{x}+\frac{\Delta}{2}} \rho(\boldsymbol{x}', t) \left[Z(\boldsymbol{x}', t) - \widetilde{Z}(\boldsymbol{x}, t) \right]^2 d\boldsymbol{x}'. \tag{3}$$

The $\widetilde{c_T}$ and $\widetilde{\sigma_{c_T}^2}$ fields can be calculated in a similar manner.

The Z-c_T joint FDF can be obtained using

$$\widetilde{P}(\xi, \eta; \boldsymbol{x}, t) = \frac{1}{\overline{\rho}(\boldsymbol{x}, t)} \int_{\boldsymbol{x}-\frac{\Delta}{2}}^{\boldsymbol{x}+\frac{\Delta}{2}} \rho(\boldsymbol{x}', t)\, \delta\left[\xi - Z(\boldsymbol{x}', t)\right] \times \delta\left[\eta - c_T(\boldsymbol{x}', t)\right] d\boldsymbol{x}', \tag{4}$$

where ξ and η are the sample-space variables and $\delta[\cdot]$ is the Dirac delta function. The discrete FDF can be obtained for a given point in a given DNS snapshot by

binning the Z and c_T samples in the corresponding filtering sub-space. This procedure is illustrated in Fig. 1 for an arbitrarily chosen filtering point at the snapshot of $t = 1.5\tau_f$ for case AZ1. The Z and c_T fields are shown for the bottom and side planes of the DNS domain respectively. Uniformly spaced 31 bins between 0 and 1 are used for c_T, whereas 35 non-uniform bins are used for Z with the majority clustered between 10^{-3} and 10^{-2} (see Fig. 1b) to accommodate most points in all three DNS cases listed in Table 1. Note that for notational concision Z and c_T are used directly instead of the sample-space variables to denote the FDFs, e.g., $\widetilde{P}(Z, c_T)$, $\widetilde{P}(Z)$, $\widetilde{P}(c_T)$, hereafter in this work. It can be seen in Fig. 1 for the FDF shape that, (i) the distribution in the progress variable space is quite broad and irregular compared to the bi-modal behaviour in conventional flames; and (ii) there is an evident cross correlation in subgrid Z and c_T fluctuations. Both features are difficult to capture using commonly employed analytical distribution models, which are briefly described next.

The Favre FDF of mixture fraction with a presumed *β-distribution* is calculated as

$$\widetilde{P}_\beta(\xi; \widetilde{Z}, \widetilde{\sigma_Z^2}) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, \xi^{a-1}\,(1-\xi)^{b-1} \tag{5}$$

with $a = \widetilde{Z}\,(1/\widetilde{g_Z} - 1)$ and $b = \left(1 - \widetilde{Z}\right)(1/\widetilde{g_Z} - 1)$, where Γ is the *gamma function* and $\widetilde{g_Z} = \widetilde{\sigma_Z^2}\ / \left(\widetilde{Z}(1-\widetilde{Z})\right)$ is the segregation factor. With the progress variable FDF, $\widetilde{P}_\beta(\eta; \widetilde{c}, \widetilde{\sigma_{c_T}^2})$, obtained using a similar procedure, the joint FDF is modelled as

$$\widetilde{P}\,(Z, c_T) = \widetilde{P}_\beta\left(\xi; \widetilde{Z}, \widetilde{\sigma_Z^2}\right) \widetilde{P}_\beta\left(\eta; \widetilde{c}_T, \widetilde{\sigma_{c_T}^2}\right), \tag{6}$$

using the common assumption of statistical independence between the subgrid fluctuations of Z and c_T. Although this assumption has been broadly used for LES

of conventional combustion showing promising outcomes [38, 41], its validity is questionable for MILD conditions since the distributed reaction zone behaviour is likely to result in stronger subgrid interactions of scalar fluctuations [15].

To consider the cross correlation, Darbyshire and Swaminathan [32] proposed a correlated joint PDF model using the Plackett *copula* method [42] for stratified premixed flames, and it was also applied for lifted flames [43, 44] showing improved prediction of the lift-off height as compared to the double-β approach in Eq. (6) in the RANS context. As for LES, one would expect the cross correlation effect to be less influential at the subgrid scales as shown in a recent DNS analysis [27] on a lifted hydrogen flame. In this study, it is of interest to explore the correlation effects in MILD combustion and hence the *copula* method is adopted to construct the correlated joint FDF. The subgrid covariance, $\widetilde{\sigma}_{Zc_T}$, computed as

$$\widetilde{\sigma_{Zc_T}}(\boldsymbol{x},t) = \frac{1}{\overline{\rho}(\boldsymbol{x},t)} \int_{\boldsymbol{x}-\frac{\Delta}{2}}^{\boldsymbol{x}+\frac{\Delta}{2}} \rho(\boldsymbol{x}',t) \left[Z(\boldsymbol{x}',t) - \widetilde{Z}(\boldsymbol{x},t) \right] \times \left[c_T(\boldsymbol{x}',t) - \widetilde{c_T}(\boldsymbol{x},t) \right] \, d\boldsymbol{x}' \tag{7}$$

is used in the *copula* method to couple the univariate marginal distributions, $\widetilde{P}_\beta(Z)$ and $\widetilde{P}_\beta(c_T)$. For non-zero values of $\widetilde{\sigma}_{Zc_T}$, the correlated joint FDF is calculated as

$$\widetilde{P}(Z,c_T) = \frac{\theta \, \widetilde{P}_\beta(Z) \widetilde{P}_\beta(c_T) \, (\mathscr{A} - 2\mathscr{B})}{\left(\mathscr{A}^2 - 4\theta\mathscr{B}\right)^{3/2}} \tag{8}$$

with

$$\mathscr{A} = 1 + (\theta - 1) \left[\widetilde{\mathscr{C}_\beta}(Z) + \widetilde{\mathscr{C}_\beta}(c_T) \right] \quad \text{and} \tag{9}$$

$$\mathscr{B} = (\theta - 1) \widetilde{\mathscr{C}_\beta}(Z) \widetilde{\mathscr{C}_\beta}(c_T) \, , \tag{10}$$

where $\widetilde{\mathscr{C}_\beta}$ is the β cumulative distribution function (CDF) and θ is the odds ratio calculated using a Monte Carlo approach. More details can be found in [32] and [43].

Alternative presumed PDF models have also been proposed, e.g. the statistically most likely distribution [45], laminar flamelet PDF [46], top-hat distribution [47], etc. Although these models were validated for specific configurations, they are less commonly used and often require on-the-fly calculation. For the purpose of this work, which is to demonstrate the potentials of using ML for FDF against the classical presumed models, only the β and *copula* models are considered for comparison.

2.3. Machine learning procedures for FDF

A feed-forward DNN algorithm is constructed and optimised for the FDF model training and validation in this study. For any ML activity, the pre-learning data selection, arrangement and manipulation play a vital role in the final performance of the trained model. Additional care is also required for physics-based ML as intended here, because the standard over-fitting and outliers treatment may be constrained by relevant physical processes [48]. Therefore, these details are described and discussed in this subsection in four progressive steps as follows.

2.3.1. Step (i): data extraction from DNS

First, eight 3D (512×512×512) DNS field matrices are prepared to extract the training samples for the DNN. The unfiltered ρ, Z and c_T fields are used to obtain the Favre FDFs for the target matrix $\mathbf{Y}$. The filtered $\widetilde{Z}$, $\widetilde{c_T}$, $\widetilde{\sigma_Z^2}$, $\widetilde{\sigma_{c_T}^2}$ and $\widetilde{\sigma_{Zc_T}}$ fields are used to construct the input matrix $\mathbf{X}$. This procedure is shown schematically in Fig. 2 for a typical snapshot of case AZ1. The filtered fields are presented in 2D

with the thin DNS grid-lines for visual clarity. The looping for "LES filter cubes" (red boxes) is performed in 3D as illustrated using the indices i, j and k for the x, y and z directions, respectively. The filter size used in this figure is $\Delta = 80\delta x$, which corresponds to a normalised filter size of $\Delta^+ = 1$, defined as

$$\Delta^+ = \frac{\Delta}{\delta_{th}^{st}}, \tag{11}$$

where $\delta_{th}^{st} = 1.6$ mm is the reference thermal thickness obtained from the MILD Flame Element (MIFE) laminar calculation [14] using the stoichiometric mixture fraction. It should be noted here that overlapping is applied between neighbouring cubes (see Fig. 2b) to increase the number of samples per DNS snapshot, which is equal to n_{cube}^3. The overlap length, i.e., the distance between two neighbouring cube centres (red points), is adjusted for different filter sizes to have a similar number of samples. When the filter size increases, the number of samples per DNS snapshot decreases at a cubic rate and hence more snapshots are required to keep a similar number of total training samples. Each snapshot is about 10 GB and only one snapshot is loaded at a time while the training data are collected accumulatively. Different Δ^+ relevant to common LES practice are investigated to elucidate the effect of filter size on the model performances for the three DNS cases. These details are summarised in Table 2. It can be seen that the number of training sets is kept similar across all cases to ensure comparable learning qualities.

2.3.2. Step (ii): manipulation of the training datasets

As required by the DNN, the two multi-dimensional matrices extracted from Step (i), $\mathbf{X}$ and $\mathbf{Y}$, are transformed (flattened) to two-dimensional input and target matrices with the same number of rows, i.e., n_{cube}^3. Each input and output row con-

Table 2: Summary of filter sizes applied for the DNS cases.

Case	Δ^+	δ^{st}_{th}[mm]	$\Delta/\delta x$	Training sets	DNS snapshots
AZ1 & AZ2	0.5	1.6	40	27648	2
AZ1 & AZ2	1	1.6	80	23958	18
AZ1 & AZ2	1.5	1.6	120	25515	35
BZ1	0.5	3	74	24192	14
BZ1	1	3	148	28672	56

tains 5 and 1085 components (columns), which are the statistical moments and the discrete FDF values, respectively (see Fig. 2). Appropriate centring and scaling are generally beneficial for ML, and it was found that the training predictions improve significantly when subtracting the mean and then dividing by the standard deviation of the n^3_{cube} rows for each column of $\mathbf{X}$. Other scaling approaches were also tested with no further improvement observed.

Applying a similar preprocessing for $\mathbf{Y}$, however, gave much worse predictions compared to using the FDFs directly as the output of the DNN. One way to explain this is that the statistical moments can be seen as independent inputs and their relative values within each row are not influential, whereas the FDF must be considered as a whole within each target row and the relative significance across the components bins must be kept the same for the training. On the other hand, the FDF is mathematically unbounded and its values can vary across several orders of magnitudes, as shown in Fig. 1, which is not ideal for the DNN. To address this issue, the values of the discretised FDF in each row are transformed into probability ones (by multiplying every FDF value by the area of the corresponding bin of the Z-cT filtering sub-space, shown in Fig. 2). As such, every number in Y

varies between 0 and 1, and they also sum up to unity across each target row. Another advantage of using probabilities is that the *softmax* activation function (see Eq. (19) for details) can be employed for the output layer of the DNN [30] to ensure the summation of outputs to unity.

2.3.3. Step (iii): identifying and removing outliers

Subsequent to the above scaling procedures, the outliers in the training datasets need to be identified and removed to further improve the training quality. The usual procedure for outlier detection in multivariate data analysis is to measure the distance of the i-th observation from the sample mean using the so-called Mahalanobis distance:

$$D_M = \left(\mathbf{X} - \overline{\mathbf{X}}\right)^T \mathbf{S}^{-1} \left(\mathbf{X} - \overline{\mathbf{X}}\right), \tag{12}$$

where $\overline{\mathbf{X}}$ is a matrix containing the average values, $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$, of the dataset variables, and $\mathbf{S}$ is the covariance matrix defined later. The observations associated with large values of D_M are classified as outliers and then discarded. A robust methodology based on the Principal Component Analysis (PCA) can be leveraged for outlier identification and removal given the known relationship between the Principal Components (PCs) and the Mahalanobis distance. Before showing this relationship, a brief description on PCA is given as follows.

PCA is a statistical method used to reduce a large number of interdependent variables to a smaller number of uncorrelated variables, while retaining as much as possible of the original data variance [49]. Given a dataset $\mathbf{X}$ of n rows (observations) and m columns (variables), which is originally centred and scaled (by the m means and standard deviations of the observations, respectively), it is possible to compute the $m \times m$ covariance matrix $\mathbf{S} = \frac{1}{(n-1)}\mathbf{X}^T\mathbf{X}$, and then to obtain its eigen-

value decomposition: $\mathbf{S} = \mathbf{A}\mathbf{L}\mathbf{A}^T$. The eigenvectors of the covariance matrix, i.e., the columns of the orthogonal matrix $\mathbf{A}$, are called Principal Components (PCs), while the eigenvalues located on the main diagonal of the matrix $\mathbf{L}$ represent the share of the dataset variance they account for. The matrix $\mathbf{X}$ can be expressed as a function of the Principal Components through the scores matrix $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{X}\mathbf{A}. \tag{13}$$

The original variables are thus transformed into a new set of uncorrelated variables. The dimensionality reduction is achieved by considering only a subset containing $q \leq m$ PCs, which contains the most energetic eigenvectors, i.e., the columns of $\mathbf{A}$ associated with the largest eigenvalues of $\mathbf{L}$. By considering the submatrix $\mathbf{A}_q$, the original dataset $\mathbf{X}$ can be correctly compressed to the chosen reduced dimensionality by finding the matrix of the scores $\mathbf{Z_q}$, through $\mathbf{Z_q} = \mathbf{X}\mathbf{A_q}$, similarly to Eq. (13). The original data matrix $\mathbf{X}$ can be reconstructed from the lower dimensional space $\mathbf{Z_q}$:

$$\mathbf{X} \approx \mathbf{X}_q = \mathbf{Z}_q\mathbf{A}_q^T. \tag{14}$$

It has been shown [49] that the sum of the squares of the PC scores, normalised by the corresponding eigenvalue, equals the Mahalanobis distance for the i-th observation:

$$\sum_{k=1}^{m} \frac{z_{ik}^2}{l_k} = D_{M,i}, \tag{15}$$

where m is the number of PCs, z_{ik} is the k-th PC score for the i-th observation, and l_k is the associated eigenvalue. In this work, the outliers are identified by examining the last few PCs, which represent linear combinations of the original variables

with minimal variance. Therefore, the last PCs are sensitive to the observations that are inconsistent with the covariance structure of the data, even if these observations are not outliers with respect to the original individual variables [49]. Thus, the summation of Eq. (15) was calculated on the last PCs of the input matrix $\mathbf{X}$ of the DNN. Given the already low dimensionality of $\mathbf{X}$ (5 variables), only the last two PCs were considered. An observation $X(i,:)$ is classified as a *leverage* outlier if

$$\sum_{k=m-q+1}^{m} \frac{z_{ik}^2}{l_k} > \psi_1, \tag{16}$$

where q is the number of last PCs considered (equal to 2 in this case), and ψ_1 is chosen as the 98th quantile of the empirical distribution of $\sum_{k=m-q+1}^{m}(z_{ik}^2/l_k)$.

Subsequently, an additional outlier removal method was applied. The method focuses on identifying the so-called *orthogonal* outliers [50] by examining the reconstruction error ϵ between the input matrix $\mathbf{X}$ and the matrix $\mathbf{X}_q$ (computed according to Eq. (14) with q=2). The i-th element of the vector ϵ is given by

$$\epsilon_i = \sqrt{\sum_{j=1}^{n}(x_{ij} - x_{q,ij})^2}, \tag{17}$$

where n is the number of variables. The i-th observation $\mathbf{x}_i$ is classified as a *orthogonal* outlier if its corresponding $\epsilon_i > \psi_2$, being ψ_2 the 98th quantile of the empirical distribution of ϵ. Once leverage and orthogonal outliers are removed from the dataset, the DNN training is then performed on the remaining observations as discussed next.

2.3.4. Step (iv): DNN architecture and training

A feed-forward, fully connected DNN was constructed for the prediction of 31×35 joint probability values. The network consists of two hidden layers and

an output layer as schematically shown in Fig. 3. The hidden layers have 256 and 512 fully connected neurons respectively, and a leaky rectified linear unit (LeakyReLU) activation function:

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ \alpha x_i & \text{otherwise} \end{cases} \tag{18}$$

where x_i is the weighted sum of the neuron input, y_i is the neuron output, and $\alpha = 0.01$ is the slope. Each hidden layer is followed by a batch normalization layer [51]. A softmax activation function is used for the output layer, composed of 1085 neurons, to predict the probability values:

$$y_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}, \tag{19}$$

where x_i and y_i correspond to the i-th neuron of the output layer. The softmax function ensures that $\sum_{i=1}^{n} y_i = 1$ and $y_i \in [0, 1]$ $\forall i = 1, ..., n$, with $n = 1085$ in our case. The loss function for the network is the binary cross entropy between the target t and the output y:

$$\mathcal{L}(y, t) = \frac{1}{n} \sum_{i=1}^{n} (t_i \log(y_i) + (1 - t_i) \log(1 - y_i)), \tag{20}$$

which represents a proper metric for measuring differences between probability distributions. The Adam optimizer [52] with an initial learning rate of 10^{-4} is used as gradient-descent algorithm. The training occurs for a maximum of 1000 epochs, i.e., training cycles through the entire training data. An early stopping method is used to avoid overfitting and a split of 80/20% between training and validation samples is applied. Validation samples are held out from the training of the neural network. The training terminates when the loss function, calculated at

the end of each epoch for the validation samples, does not decrease after a certain number of epochs (10 in this work). For each epoch, the training data are fully shuffled and divided into batches with 256 training samples per batch.

In summary, the DNN used in the work has the following architecture:

- linear layer with 5 input features and bias, LeakyReLU activation function with α=0.01, and 256 output features;
- batch normalization with 256 input and output features, and momentum equals to 0.9;
- linear layer with 256 input features and bias, LeakyReLU activation function with α=0.01, and 512 output features;
- batch normalization with 512 input and output features and momentum equals to 0.9;
- linear layer with 512 input features and bias, Softmax activation function, and 1085 output features.

The loss function is the binary crossentropy. Additional DNN hyperparameters are the learning rate (=10^{-4}), the number of epochs (=1000), the batch size (=256), and the patience (=10) of the early stopping method.

The above learning algorithm is applied for the different DNS cases with a range of filter sizes listed in Table 2, followed by a prediction step on the entire training datasets. The time required to generate the DNN model depends on the particular case because more than 95% is spent on Step (i), where multiple large DNS snapshots are read and preprocessed. The most costly case in this work is therefore the last case in Table 2 (BZ1 with $\Delta^+ = 1$), which took about 4 hours

using a 3.0 GHz single core with GPU acceleration for the learning. Once the the DNN is generated, the computational cost for the prediction step is negligible (as it will be in *a posteriori* LES) since it only involves simple algebraic operations and function evaluations using already calibrated node weights and biases. The predicted joint FDFs for the different cases listed in Table 2 are compared with the DNS and analytically modelled results in the following section.

3. FDF behaviour and model prediction

Owing to the distinct reaction zone and scalar gradient behaviours in MILD combustion [14, 16], the corresponding FDF characteristics are little known and expected to be significantly different from those in conventional combustion. These attributes have not been reported in previous studies. Therefore, in this section it is of interest to examine the basic features of FDFs obtained from the three DNS cases listed in Table 1, and then their conditional means are compared with those obtained by the DNN and analytical models described earlier in Section 2. Furthermore, the importance of subgrid covariance is studied by training the DNN with and without it in the input matrix. The results are compared with the DNS data to assess model performance under different conditions.

3.1. Basic instantaneous features

The x-y midplane instantaneous contours of Z, c_T and $\dot{\omega}_{c_T}$ at $t = 1.5\tau_f$ with typical reaction zone FDFs are shown in Figs. 4, 5 and 6 for cases AZ1, AZ2 and BZ1, respectively. The mean flow moves in the x-direction. First of all, it is worth recalling the different characteristics of the three DNS cases summarised in Table 1. AZ1 is the base case with a maximum oxygen mole fraction ($X_{\mathrm{O2}}^{\mathrm{ox}}$) of 3.5% and a mixing length (l_Z) larger than reaction zone length scale (l_c). Case

AZ2 differs from AZ1 only in the turbulent mixing characteristics with $l_Z \approx l_c$. One can easily observe that the Z field of AZ2 in Fig. 5b is more wrinkled as compared to that of AZ1 in Fig. 4b. As a result, the reaction zones shown in Fig. 5c are more distributed in the computational domain for AZ2 compared to those in Figs. 4c for AZ1, which is a common feature of MILD combustion as the Damköhler number tends toward unity [15, 53]. Another difference caused by the mixing length is that the mixture fraction fluctuations are statistically stronger in AZ2 having a standard deviation about 30% higher than that in AZ1 (see Table 1). Case BZ1 has the same mixing characteristics as AZ1 but is more diluted with $X_{\mathrm{O2}}^{\mathrm{ox}} = 2\%$. As one would expect, the peak reaction rate shown in Fig. 6c drops significantly due to the dilution. The stronger turbulence effect also gives rise to rather distributed reaction zones similar to AZ2 despite a larger mixing length.

The FDFs shown in Figs. 4, 5 and 6 are extracted for an arbitrarily chosen point (marked by black dot) in the given x-y plane for each case, such that this point is located in a reaction zone, and also the mixing and reaction states, i.e., local (Z/Z_{st}) and c values, are similar among the three cases. Three filter sizes with $\Delta^+ = 0.5$, 1 and 1.5 are applied for cases AZ1 and AZ2. These filter sizes are illustrated using the to-the-scale boxes with corresponding colours. Only $\Delta^+ = 0.5$ and 1 are explored for BZ1 which has a reference thermal thickness nearly twice that for AZ1 and AZ2 (see Table 1), and a larger Δ^+ leads to insufficient number of samples for the subsequent DNN training. Nevertheless, larger filter sizes ($\Delta >$ 3 mm) are not of interest as they would not resolve sufficient amount of turbulent kinetic energy in *a posteriori* LES. The marginal FDFs, $P(Z) = \int P(Z, c_T)\, dc_T$ and $P(c_T) = \int P(Z, c_T)\, dZ$, are also shown in the small inset figures. The following observations are made by a careful and close examination of the FDFs in Figs. 4,

5 and 6:

- case AZ1 with conditions relatively *less* MILD among the three cases, exhibits FDF behaviours that share more similarities with those appearing in conventional flames. For example, the distribution in c_T-space shown in Figs. 4d, 4e and 4f is more or less bi-modal. This is not at all surprising if one sees the $flamelet\text{-}like$ structures (as commonly found in conventional combustion) inside the filter boxes in Fig. 4c, where the probabilities of falling into either unburnt ($c_T = 0$) or burnt ($c_T = 1$) mixtures are much higher. The corresponding local scalar gradient of c_T in Fig. 4a is also quite strong, clearly marking the wrinkled local flame fronts. Such structures are not as evident in cases AZ2 and BZ1 due to the more distributed reaction zones.

- The marginal FDF for mixture fraction has a single pronounced peak and behaves quite consistently across all three cases. This is expected for the passive scalar. For the progress variable, however, its statistical distributions in cases AZ2 and BZ1 become very different in their own ways compared to the bi-modal shapes observed for AZ1 in Fig. 4. Specifically, for AZ2 one can identify a peak on one side of the marginal PDF of c_T, accompanied with a plateau extending towards the other side. As Δ^+ becomes larger the height of this plateau increases and its value can be quite close to the peak as can be seen in Fig. 5f. Differently in Fig. 6 for case BZ1, $P(c_T)$ is rather broadly distributed between 0 and 1 without dominant (or with multiple) peaks. Both of these are difficult to capture using classical presumed models.

- There is a negative cross correlation between Z and c_T in all three cases, which is in line with previous numerical [27] and experimental [54] studies on jet flames. This cross correlation is stronger in the diluted case BZ1, which is not surprising since the distribution of c_T spreads more evenly over a broader range resulting in a stronger interaction with variations of Z as compared to cases AZ1 and AZ2. Modelling the cross correlation in the FDF is also quite challenging.

Next, it is of interest to examine the DNN and analytical model abilities to capture the above complex FDF behaviours, which are presented in Figs. 7, 8 and 9 for cases AZ1, AZ2 and BZ1, respectively. The corresponding temporal and spatial points are the same as those marked in Figs. 4, 5 and 6. Before making the comparisons, it is stressed here again that one must be cautious with the FDFs directly extracted from the instantaneous DNS snapshots. As already discussed in Section 2.2, these FDFs are random variables (containing subgrid statistical information) but not statistics [35, 38], whereas the often used analytical models, e.g., β and *copula* PDFs, presume statistical distributions. Although strictly they should not be compared directly against one another, this comparison using single snapshot data is made here to provide a clear view of the typical individual model performance for the specific filter points shown in the previous figures. Detailed quantitative comparison of the statistics is presented later in Section 3.2.

It can be seen in these figures that the DNN model significantly outperforms both analytical models and its prediction agrees very well with the DNS data for all cases with different filter sizes and initial conditions. This is particularly evident for the marginal distributions of c_T, where both the β and *copula* models fail to predict with satisfactory accuracy, although the latter seems to capture the corre-

lation qualitatively well (the correct slope is observed). Relatively, the prediction by DNN is even better at larger Δ^+ showing excellent agreement with the DNS. As a general observation, the marginal FDF of progress variable tends to always have different non-regular shapes from one case to another. The DNN captures these shapes quite well, as compared to those given by the β and *copula* models always having *Gaussian-like* distributions. This difference has important implications for the reaction rate modelling, because the pronounced peak in the Gaussian FDF leads to larger reaction rate gradients through the double-integration procedure, as one shall see later in Section 4. For the mixture fraction, however, all models give good results but only the DNN is able to capture the asymmetry of the FDF as can be seen more clearly in Figs. 9b and 9d for case BZ1. These results indicate promising capabilities of the DNN to predict the complex subgrid scalar fluctuations in MILD combustion, and this is further confirmed by comparing the FDF statistics next.

3.2. Statistics: conditional FDF & Jensen-Shannon divergence

Following the experimental studies of Tong *et al.* [29, 55, 56], the instantaneous FDFs obtained from the DNS are conditioned on the resolved scalars, $\widetilde{Z}$ and $\widetilde{c_T}$, and then ensemble-averaged. It would be insightful to condition the FDFs also on the variances, ideally all five statistical moments, but there is not sufficient number of samples available to perform a statistically meaningful averaging. Thousands of DNS snapshots would be required, which is impractical for the state-of-the-art computational infrastructures. Thus, only two conditional variables are considered in this work. The resolved mixture fraction and progress variable are chosen here so that the selected samples are located in the reaction zone ($\widetilde{c_T} \approx 0.5$) and with a relevant lean equivalence ratio.

Figures 10 and 11 show the conditional FDFs, $\left\langle \widetilde{P}(Z, c_T) \,\middle|\, \widetilde{Z}, \widetilde{c_T} \right\rangle$, for cases AZ1 and BZ1, respectively. The results for AZ1 and AZ2 are similar and hence the latter is not shown here (can be found in Supplementary Material). The corresponding values for the conditioning variables can be found in the figure captions. These values are chosen such that the conditional FDFs shown here consistently reflect the statistics of the previous instantaneous FDFs in Figs. 7 and 9. Comparing with those instantaneous FDFs, the random fluctuations are smoothed out by the conditional averaging as one would expect, which makes it more valid to be compared with the β and *copula* distributions. Similarly, the conditional DNN results also appear to be relatively more statistically converged. The key observations from Figs. 10 and 11 are summarised as follows.

- The DNN is able to remarkably reproduce the conditional FDFs precisely with a maximum error smaller than 0.1% across all cases considered. This includes all the joint FDFs and marginal ones for both Z and c_T. It also captures the significant changes in the FDF shape with the varying filter size, especially for the progress variable.

- For case AZ1 (and also AZ2), both the β and *copula* models over-predict the peak when $\Delta^+ \leqslant 1$ for both Z and c_T distributions. However, for $\Delta^+ = 1.5$, the overall prediction is quite good for $\widetilde{P}(Z)$ and the peak of $\widetilde{P}(c_T)$ is also similar to the DNS value, but the shape show significant deviations.

- For case BZ1, the mixture fraction distribution is generally predicted well by all models for different Δ^+ values. However, the both analytical models fail to capture the *bimodal-plateau* shape of $\widetilde{P}(c_T)$, which is typical of MILD combustion but rarely seen in conventional flames.

In order to demonstrate the model performances in terms of statistics over a large number of test points under various conditions, a commonly used metric is employed to quantify the deviation of the model predictions from the DNS data. The Jensen-Shannon divergence [57], which measures the similarity of two probability distributions, $Q_1(n)$ and $Q_2(n)$, is computed using

$$\begin{aligned} \mathrm{JSD}(Q_1 \| Q_2) = & \frac{1}{2} \sum_{n=1}^{N} \left\{ Q_1(n) \ln \left[\frac{Q_1(n)}{Q_2(n)} \right] \right. \\ & + \left. Q_2(n) \ln \left[\frac{Q_2(n)}{Q_1(n)} \right] \right\} \end{aligned} \tag{21}$$

for a given filter point (see number of samples in Table 2). Q_1 and Q_2 are taken to be the DNS and modelled FDFs, respectively. The JSD is mathematically bounded between 0 and ln(2), with 0 indicating $Q_1 = Q_2$ and hence the smaller the JSD value the more similar the two distributions. The PDFs of JSD for different DNS cases and filter sizes are presented in Fig. 12. As can be seen, the JSD values given by the DNN are clustered near zero and overall much lower than those for the β and *copula* models. Moving from small to large filter sizes, all three models give improved predictions as the PDFs shift towards smaller JSD values and the improvement is particularly significant for the DNN having more than 95% lower than 0.05 as seen in Figs. 12c, 12f and 12h. The DNN model seems to be insensitive to the scalar considered, whereas the β and *copula* models perform better for Z, which is consistent with the observations in previous figures.

3.3. Importance of covariance

It is of interest to understand how the covariance influences the DNN performance. To this end, the DNN training was repeated using the same DNS cases, snapshots and filter sizes with the only change made by removing the covari-

ance from the input matrix, which reduces to $\mathbf{X}^4_{i,j,k} = \{\widetilde{Z}, \widetilde{c_T}, \widetilde{\sigma^2_Z}, \widetilde{\sigma^2_{c_T}}\}_{i,j,k}$. The JSD of the four-variable DNN predictions are calculated in a similar manner as before, and compared with the five-variable DNN results in Figs. 13 and 14 for the progress variable and mixture fraction, respectively. Overall, the model behaviour and prediction trends are similar between the two scalars and the results generally improve as the filter size increases. Without considering the covariance, the DNN predictions for cases AZ1 and AZ2 become considerably poorer, i.e., the PDF shifts towards larger JSD values, but little difference is observed for the more diluted case BZ1.

While all LES sample cubes from the DNS domain were used to reconstruct these PDFs, additional analysis was also carried out to investigate the influence of covariance at different axial locations (x-direction). This was motivated by Ref. [16], which showed that different combustion modes occur as the mean flow convects the reacting mixtures from the inlet to outlet: autoigniting spots first form in the lean mixture upstream, and subsequently these hot spots trigger propagating fronts moving towards richer mixtures. The JSD results at different x-locations (see Supplementary Material), however, were found to be quite similar to the global behaviours seen in Figs. 13 and 14, suggesting that the influence of covariance on the FDF prediction is independent of the physical locations, thermochemical state or burning modes. Therefore, this demonstrates the promising general applicability of the DNN method for various conditions which may occur in different regions of the computational domain.

3.4. DNN prediction using "unknown" inputs

The results presented so far are still *self-predictions*, meaning that the predictions are performed on the training datasets, which is essential for model assess-

ment. However, a key ability determining the genuine usefulness of ML-based approaches is that when a model is trained and validated using a given set of data, it should be able to take unknown inputs from other datasets (often having similar features) which have not been explicitly included in the learning, and accurately predict the corresponding outcomes. To demonstrate this ability for the present DNN approach, the sampling method for the training data selection is revised such that the DNN model learns from the snapshots spanning between $t = \tau_f$ and $1.2\tau_f$, and subsequently the predictions are made for snapshots between $t = 1.4\tau_f$ and $1.5\tau_f$. There are substantial levels of temporal variations in the turbulent mixing characteristics and composition (including EGR dilution) for the inflow mixtures and hence the MILD combustion behaviours within each snapshot, as time evolves from τ_f to $1.5\tau_f$ (see [16] for details). Therefore, such a DNN setup poses considerable challenges for the model to handle scenarios that have not been explicitly *seen* during its training process.

The PDFs of JSD for the self- and unknown-predictions of the FDF are compared in Fig. 15. The filter size of $\Delta^+ = 1$ is used for all cases. As shown consistently across three cases, the DNN provides a similar level of predictive accuracy when unknown inputs are fed into the model with a very small drop from the self-predictions. The JSDs are generally ($> 80\%$) smaller than 0.05. This is to be compared with the results given by the β and *copula* FDF models in Fig. 12 (middle column), both having significant amount of JSD values larger than 0.1 indicating significant deviations from the DNS data.

Since the above "unknown" inputs come from the DNS case on which the DNN was trained, these inputs may still have similar physical patterns and features despite the large variations in turbulence and composition. Therefore, a fur-

ther step is considered by training and validating the DNN model using the data collected from cases AZ1 and BZ1, and subsequently the testing is performed on case AZ2. Note that the three DNS cases were simulated independently with very different initial fields (see Table 1), leading to distinct turbulent mixing and combustion behaviours as demonstrated earlier in Figs. 4 to 6. Figure 16 compares the JSD results obtained from this new test with the self-predictions for $\Delta^+ = 0.5$. Although a considerable drop is seen at the peak of the PDF near JSD = 0.02, the overall performance is still quite good with the majority of the JSD smaller than 0.1 for both the mixture fraction and progress variable FDFs. This suggests that the present DNN approach is robust and able to predict the FDFs accurately with the capability to handle the unknown variations in the input data, which is promising for broader applications in MILD combustion modelling.

4. *A priori* assessment of reaction rate modelling

4.1. Filtered reaction rate closure

Once the joint FDF, $\widetilde{P}(Z, c_T)$, has been obtained through the models described earlier, now one is able to close the reaction rate source term in the transport equation for filtered progress variable using

$$\overline{\dot{\omega}}_{c_T}(\boldsymbol{x}, t) = \overline{\rho}(\boldsymbol{x}, t) \int_0^1 \int_0^1 \langle \dot{\omega}_{c_T} \rangle \widetilde{P}(Z, c_T; \boldsymbol{x}, t) dZ \, dc_T, \tag{22}$$

where $\langle \dot{\omega}_{c_T} \rangle = \left\langle \dot{\omega}_{c_T}(\boldsymbol{x}, t)/\rho(\boldsymbol{x}, t) \middle| Z, c_T \right\rangle$ is the doubly conditional mean reaction rate obtained from the DNS data. According to Eq. (1), the instantaneous reaction rate of c_T is calculated as, $\dot{\omega}_{c_T} = \dot{q}/[c_p(T_b - T_u)]$, with $\dot{q}$ and c_p being the volumetric heat release rate and mixture specific heat capacity respectively.

The conditional averages are performed using samples collected over the entire DNS computational domain including all snapshots available (≈ 60) to achieve good statistical convergence. It was verified (not shown) that the doubly conditional mean rates only have negligible variations in time and space within the DNS datasets. This supports the fundamental assumption for many turbulent combustion models that the conditional means have relatively small temporal and spatial gradients, if appropriate conditioning variables are chosen. These models include the two main categories, namely flamelets [58–61] and conditional moment-based approaches (with transport equation closure − CMC [62] or source-term estimation − CSE [63]).

Figure 17 presents the conditional mean reaction rates of progress variable for cases AZ1 and BZ1. Note that case AZ2 was found to have very similar thermochemical characteristics as AZ1 and thus is not considered further in this section. A *heart-shaped* distribution pattern is seen for the conditional mean reaction rates for both cases. Also, high reaction rates are located mostly in lean regions although mixtures spanning over the entire flammable range are present in the DNS domain. The reaction rates seen here are very different to those occurring in conventional combustion where they are usually clustered near the stoichiometric mixture fraction and an intermediate value of progress variable, typically forming a Gaussian bell shape. This can be explained as follows. When the inflow mixtures enter the DNS domain, reactions immediately occur as autoignition (AI) spots due to the high temperature and presence of radials [16]. AI happens in the very lean mixtures which have much smaller ignition delay times as marked using stars in Fig. 17. These spots then diffuse into slightly richer mixtures promoting to form lean flame fronts. This process involves a temperature drop and we

can see the reaction rates in regions with small c_T. Once the fronts are formed, then they propagate into even richer mixtures resulting in a large amount of heat release. This behaviour is observed for both AZ1 and BZ1 cases, with the latter completing combustion earlier (no rich flames seen) because of the higher dilution level, i.e., lower fuel concentrations. Such observations for the conditional reaction rates are consistent with the findings in [17] using Lagrangian particles to track the autoigniting spots and propagating flame fronts, further confirming the multi-regime nature of MILD combustion. Next, the conditional reaction rates, along with the FDFs obtained in Section 3, are used to model the filtered reaction rate via Eq. (22), and the results are then compared with the explicitly filtered reaction rate to assess the model performance.

Figure 18 shows the typical instantaneous mid-plane contours of the unfiltered, filtered and modelled reaction rates for case AZ1 and $\Delta^+ = 0.5$ (see results for other Δ^+ in Supplementary Material). In this figure, the filtering effect can be seen clearly by comparing the $\dot{\omega}_{c_T}$ and $\overline{\dot{\omega}}_{c_T}$ contours. It is also important to note that $\overline{\dot{\omega}}_{c_T}^{\text{m-DNS}}$ is the best possible result one would obtain using two-scalar (Z and c_T) based modelling approaches (e.g., flamelets, CMC, CSE), since both the conditional mean reaction rate and FDF in Eq. (22) are directly obtained from the DNS data. It is shown in Fig. 18 that the $\overline{\dot{\omega}}_{c_T}^{\text{m-DNS}}$ field is in good agreement with $\overline{\dot{\omega}}_{c_T}$ although the former seems slightly more diffusive, which may be due to the limited samples and number of bins used for the conditional averaging and FDF construction. A more quantitative comparison is shown in Fig. 19 for the joint PDF of $\overline{\dot{\omega}}_{c_T}$ and $\overline{\dot{\omega}}_{c_T}^{\text{m-DNS}}$ for different filter sizes. It can be observed clearly that the joint PDFs fall within the vicinity of the diagonal suggesting an excellent agreement between $\overline{\dot{\omega}}_{c_T}$ and $\overline{\dot{\omega}}_{c_T}^{\text{m-DNS}}$ for a range of Δ^+ values. The observations

in Figs. 18 and 19 for case AZ1 are also representative for the other cases AZ2 and BZ1 (included in Supplementary Material). Thus, these results demonstrate that Eq. (22) is an adequate and robust model for the filtered reaction rate closure in MILD combustion under different turbulence and thermochemical conditions. The key question now is, however, whether this model performs accurately with the modelled FDFs, although the qualitative contours in Fig. 18 (lower row) seem quite promising.

The scatter plots of $\overline{\dot{\omega}}_{c_T}^{\mathrm{m-DNS}}$ and the reaction rates computed using β, *copula* and DNN FDF models are presented in Fig. 20. The qualitative behaviours and trends are similar among the three cases (see Supplementary Material) and hence only AZ1 is shown here. It can be seen that although all models give reasonable predictions near the diagonal line, the DNN outperforms the analytical models for all filter sizes. The DNN predictions generally exhibit good symmetry about the diagonal, indicating a bias towards neither under- nor over-prediction. As Δ^+ increases, the DNN prediction improves considerably. The performance of the analytical models, by contrast, shows more complex behaviour with respect to the filter size change. While the scatters for both the β and *copula* models are asymmetric, there seems to be a trend in the off-diagonal samples moving from under-predictions at small Δ^+ to over-predictions at larger Δ^+.

The model performance can be examined quantitatively via the normalised prediction error, which is defined as

$$\varepsilon^{\mathrm{m}} = \frac{\overline{\dot{\omega}}_{c_T}^{\mathrm{m}} - \overline{\dot{\omega}}_{c_T}^{\mathrm{m-DNS}}}{\overline{\dot{\omega}}_{c_T}^{\mathrm{m-DNS}}}\,, \tag{23}$$

for every filter point. Figure 21 compares the PDFs of ε^{m} for the DNN and analytical models, with the samples collected from all three cases AZ1, AZ2 and BZ1

for each filter size. The results shown here generally support the observations and arguments made earlier for Fig. 20. The DNN gives accurate, robust and unbiased predictions with the ε^{m} values located symmetrically near zero. For the analytical models, the asymmetry in the reaction rate prediction observed earlier is further confirmed, as well as the shift from under- to over-predictions when Δ^+ increases. The overall accuracy also decreases slightly at larger Δ^+ for both the β and *copula* models.

5. Concluding remarks

In this work, the DNN machine learning algorithm is applied to train a predictive model for the joint FDF of mixture fraction and progress variable in the context of MILD combustion. A state-of-the-art DNS dataset considering detailed-chemistry methane/air combustion with EGR dilution is used for the DNN training and validation. Three DNS cases, differing in the inflow turbulent mixing characteristics and dilution levels, are investigated, firstly to provide physical insight into the FDF behaviours with respect to those found in conventional combustion, and then secondly to assess the accuracy and robustness of the DNN model under various MILD conditions of practical interest. In addition, a range of filter sizes, of the order of the thermal thickness for a reference planar laminar flame, are studied for each DNS case, composing a comprehensive group of test cases for model development and validation. For comparison, two commonly used analytical FDF models, presuming double independent β and correlated *copula* distributions respectively, are used to show the relative performance of the DNN model. Moreover, the mean reaction rate doubly conditioned on the mixture fraction and progress variable is computed using the DNS data for different MILD

conditions, showing complex behaviours that significantly deviate from those observed in conventional combustion. This conditional reaction rate is then used along with the presumed and DNN-predicted FDFs to compute the filtered reaction rate source term for *a priori* assessment. The main findings of the above investigations in this work are summarised as follows.

- The FDFs in MILD combustion are found to be non-regular and significantly different from the classical shapes, e.g., Gaussian and bi-modal commonly found in conventional non-premixed and premixed jet flames. The differences are particularly evident for the DNS cases where the characteristic mixing and reaction lengths are comparable or dilution level is high, both promoting more distributed reaction zone behaviours.

- The DNN is generally able to capture the complex FDF behaviours and their variations with excellent accuracy across various turbulence and MILD thermochemical conditions for a range of filter sizes. This can be achieved with very little effort in parametric tuning, typically fixing network architecture and all hyper-parameters using one case and then applying to others as a *black-box*. It is also shown that the DNN can handle unknown inputs quite well, suggesting a good model robustness. By contrast, the β and *copula* models in most cases cannot capture the MILD FDF behaviours especially in the progress variable space, although the latter could provide an accurate cross-correlation slope in the joint space.

- The doubly conditional reaction rate of the progress variable exhibit complex multi-regime behaviours, with autoignition and flame propagation dominating in different regions in the joint conditional space. The former always

appears in leanest mixtures, which have much shorter ignition delay times, whereas the latter tends to be initiated by the autoignited spots and then propagates from lean to richer mixtures. Also, autoignition-driven heat release occurs mostly in the upstream (close to the inlet), and the flame fronts can be found all over the entire DNS domain.

- Despite its physical-space dependent behaviour above, the conditional reaction rate is found to vary only slightly ($< 5\%$ maximum) in space and time, at least for the DNS cases considered. As a result, the filtered reaction rate closure based on the joint FDF performs consistently well for all cases. The DNN predictions show much better agreement with the DNS reaction rates than those given by the analytical models for a range of filter sizes.

Although the above observations clearly demonstrate the attractive potential of using Eq. (22) and DNN-based FDF model for the reaction rate modelling in MILD combustion, there is still a remaining challenge which requires further investigation. Even though the FDF may be modelled well using the DNN, estimating the mean reaction rate conditioned on the mixture fraction and progress variable prior to *a posteriori* LES is challenging. In conventional combustion, the initial thermochemical state can be determined by mixture fraction alone (at least for two-stream problems). However, as a unique feature of MILD combustion, the mixing between unburnt reactants and recirculated burnt products (often cooled and thus non-adiabatic) can change the reactivity of local mixtures, and it is difficult to describe in a pre-tabulation step. All canonical flame/reactor calculations require well defined initial thermochemical state and/or boundary conditions, which are not straightforward to specify for MILD combustion, particularly for cases involving mixture fraction variations. An additional parametric variable

marking the local dilution level may be necessary, which will be a central topic for follow-up works.

Acknowledgements

ZXC and NS acknowledge the support of Mitsubishi Heavy Industries, Takasago, Japan. SI acknowledges the financial support of the Fondation Wiener Anspach. VK acknowledges the financial support of the Cambridge Trusts. GDA has received funding from the Fonds National de la Recherche Scientifique (FRS-FNRS) through a FRIA fellowship. This work has received funding from the Fondation Wiener Anspach, through the collaborative project Alchemy - Machine Learning for Complex Multiphysics Problems. This work has also received funding from the European Research Council, Starting Grant No. 714605. The authors gratefully acknowledge Dr. N.A.K. Doan for helping with the DNS data.

References

[1] J.A. Wünning, J.G. Wünning, Flameless oxidation to reduce thermal no-formation, Prog. Energy Combust. Sci. 23 (1997) 81–94. doi:10.1016/S0360-1285(97)00006-3.

[2] A. Cavaliere, M. de Joannon, Mild Combustion, Prog. Energy Combust. Sci. 30 (4) (2004) 329–366. doi:10.1016/j.pecs.2004.02.003.

[3] N. Swaminathan, Physical Insights on MILD Combustion From DNS, Front. Mech. Eng. 5 (2019) 59. doi:10.3389/fmech.2019.00059.

[4] André A.V. Perpignan, A. Gangoli Rao, D.J.E.M. Roekaerts, Flameless combustion and its potential towards gas turbines, Prog. Energy Combust. Sci. 69 (2018) 28–62. doi:10.1016/j.pecs.2018.06.002.

[5] İ. B. Özdemir, N. Peters, Characteristics of the reaction zone in a combustor operating at mild combustion, Experiments in Fluids 30 (6) (2001) 683–695. doi:10.1007/s003480000248.
URL https://doi.org/10.1007/s003480000248

[6] M. de Joannon, A. Cavaliere, T. Faravelli, E. Ranzi, P. Sabia, A. Tregrossi, Analysis of process parameters for steady operations in methane mild combustion technology, Proc. Combust. Inst. 30 (2) (2005) 2605–2612. doi:10.1016/j.proci.2004.08.190.
URL http://linkinghub.elsevier.com/retrieve/pii/S0082078404002425

[7] T. Plessing, N. Peters, J.G. Wünning, Laseroptical investigation of highly preheated combustion with strong exhaust gas recirculation, Proc. Combust. Inst. 27 (1998) 3197–3204.

[8] M. de Joannon, A. Saponaro, A. Cavaliere, Zero-dimensional analysis of diluted oxidation of methane in rich conditions, Proc. Combust. Inst. 28 (2) (2000) 1639–1646. doi:10.1016/S0082-0784(00)80562-7.

[9] M. de Joannon, A. Matarazzo, P. Sabia, A. Cavaliere, Mild Combustion in Homogeneous Charge Diffusion Ignition (HCDI) regime, Proc. Combust. Inst. 31 (2) (2007) 3409–3416. doi:10.1016/j.proci.2006.07.039.

URL http://linkinghub.elsevier.com/retrieve/pii/S1540748906000460

[10] M. de Joannon, G. Sorrentino, A. Cavaliere, MILD combustion in diffusion-controlled regimes of Hot Diluted Fuel, Combust. Flame 159 (5) (2012) 1832–1839. doi:10.1016/j.combustflame.2012.01.013.
URL http://dx.doi.org/10.1016/j.combustflame.2012.01.013

[11] J. Sidey, E. Mastorakos, R.L. Gordon, Simulations of autoignition and laminar premixed flames in methane/air mixtures diluted with hot products, Combust. Sci. Technol. 186 (4-5) (2014) 453–465. arXiv:https://doi.org/10.1080/00102202.2014.883217, doi:10.1080/00102202.2014.883217.
URL https://doi.org/10.1080/00102202.2014.883217

[12] J.A. Van Oijen, Direct numerical simulation of autoigniting mixing layers in MILD combustion, Proc. Combust. Inst. 34 (1) (2013) 1163–1171. doi:10.1016/j.proci.2012.05.070.
URL http://dx.doi.org/10.1016/j.proci.2012.05.070

[13] Y. Minamoto, T.D. Dunstan, N. Swaminathan, R.S. Cant, DNS of EGR-type turbulent flame in MILD condition, Proceedings of the Combustion Institute 34 (2) (2013) 3231–3238. doi:https://doi.org/10.1016/j.proci.2012.06.041.
URL http://www.sciencedirect.com/science/article/pii/S1540748912001496

[14] Y. Minamoto, N. Swaminathan, Scalar gradient behaviour in

MILD combustion, Combust. Flame 161 (4) (2014) 1063–1075. doi:10.1016/j.combustflame.2013.10.005.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84894254713{&}partnerID=tZOtx3y1

[15] Y. Minamoto, N. Swaminathan, R.S. Cant, T. Leung, Reaction Zones and Their Structure in MILD Combustion, Combust. Sci. Technol. 186 (8) (2014) 1075–1096. doi:10.1080/00102202.2014.902814.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84903649697{&}partnerID=tZOtx3y1

[16] N.A.K. Doan, N. Swaminathan, Y. Minamoto, DNS of MILD combustion with mixture fraction variations, Combust. Flame 189 (2018) 173–189.

[17] N.A.K. Doan, N. Swaminathan, Autoignition and flame propagation in non-premixed MILD combustion, Combust. Flame 201 (2019) 234–243. doi:10.1016/j.combustflame.2018.12.025.

[18] N.A.K. Doan, N. Swaminathan, Role of radicals on MILD combustion inception, Proc. Combust. Inst. 37 (4) (2019) 4539–4546. doi:10.1016/j.proci.2018.07.038.
URL https://doi.org/10.1016/j.proci.2018.07.038

[19] E. Oldenhof, M.J. Tummers, E.H. van Veen, D.J.E.M. Roekaerts, Ignition kernel formation and lift-off behaviour of jet-in-hot-coflow flames, Combust. Flame 157 (6) (2010) 1167 – 1178. doi:https://doi.org/10.1016/j.combustflame.2010.01.002.

URL http://www.sciencedirect.com/science/article/pii/S0010218010000131

[20] B.B. Dally, A.N. Karpetis, R.S. Barlow, Structure of turbulent non-premixed jet flames in a diluted hot coflow, Proc. Combust. Inst. 29 (x) (2002) 1147–1154. doi:10.1016/S1540-7489(02)80145-6.

[21] G. Sorrentino, P. Sabia, M. de Joannon, A. Cavaliere, R. Ragucci, The effect of diluent on the sustainability of mild combustion in a cyclonic burner, Flow Turbul. Combust. 96 (2) (2016) 449–468. doi:10.1007/s10494-015-9668-3.
URL https://doi.org/10.1007/s10494-015-9668-3

[22] A.S. Veíssimo, A.M. Rocha, M. Costa, Operational, combustion, and emission characteristics of a small-scale combustor, Energy Fuels 25 (6) (2011) 2469–2480.

[23] J. Lamouroux, M. Ihme, B. Fiorina, O. Gicquel, Tabulated chemistry approach for diluted combustion regimes with internal recirculation and heat losses, Combust. Flame 161 (8) (2014) 2120–2136.

[24] G. Sorrentino, G. Ceriello, M. de Joannon, P. Sabia, R. Ragucci, J. van Oijen, A. Cavaliere, L.P.H. de Goey, Numerical investigation of moderate or intense low-oxygen dilution combustion in a cyclonic burner using a flamelet-generated manifold approach, Energy Fuels 32 (10) (2018) 10242–10255. doi:10.1021/acs.energyfuels.8b01099.
URL https://doi.org/10.1021/acs.energyfuels.8b01099

[25] Z.X. Chen, N.A.K. Doan, X.J. Lv, N. Swaminathan, G. Ceriello, G. Sorrentino, A. Cavaliere, A numerical study of a cyclonic combustor under MILD conditions using non-adiabatic tabulated chemistry, Energy Fuels 32 (2018) 10256–10265. doi:10.1021/acs.energyfuels.8b01103.

[26] G. Ceriello, G. Sorrentino, A. Cavaliere, P. Sabia, M. de Joannon, R. Ragucci, The role of dilution level and canonical configuration in the modeling of MILD combustion systems with internal recirculation, Fuel 264 (2020) 116840. doi:10.1016/j.fuel.2019.116840.

[27] Z.X. Chen, N.A.K. Doan, S. Ruan, I. Langella, N. Swaminathan, *A priori* investigation of subgrid correlation of mixture fraction and progress variable in partially premixed flames, Combust. Theory Model. 22 (2018) 862–882.

[28] F. Gao, E.E. O'Brien, A large-eddy simulation scheme for turbulent reacting flows, Phys. Fluids A 5 (1993) 1282–1284.

[29] C. Tong, Measurements of conserved scalar filtered density function in a turbulent jet, Phys. Fluids 13 (2001) 2923.

[30] M.T.H. de Frahan, S. Yellapantula, R. King, M.S. Day, R.W. Grout, Deep learning for presumed probability density function models, Combust. Flame 208 (2019) 436–450.

[31] Corentin J. Lapeyre, Antony Misdariis, Nicolas Cazard, Denis Veynante, Thierry Poinsot, Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates, Combust. Flame 203 (2019) 255–264. doi:10.1016/J.COMBUSTFLAME.2019.02.019.

URL https://www.sciencedirect.com/science/article/pii/S0010218019300835?dgcid=rss{_}sd{_}all

[32] O.R. Darbyshire, N. Swaminathan, A presumed joint pdf model for turbulent combustion with varying equivalence ratio, Combust. Sci. Technol. 184 (12) (2012) 2036–2067.

[33] M.D. Smooke, V. Giovangigli, Formulation of the premixed and non-premixed test problems, in: M. D. Smooke (Ed.), Reduced Kinetic Mechanisms and Asymptotic Approximations for Methane-air Flames, Vol. 384 of Lecture Notes in Physics, Springer-Verlag, Berlin, 1991.

[34] T. Kathrotia, U. Riedel, A. Seipel, K. Moshammer, A. Brockhinke, Experimental and numerical study of chemiluminescent species in low-pressure flames, Appl. Phys. B Lasers Opt. 107 (3) (2012) 571–584. doi:10.1007/s00340-012-5002-0.
URL https://doi.org/10.1007/s00340-012-5002-0

[35] S.B. Pope, Pdf methods for turbulent reactive flows, Prog. Energy Combust. Sci. 11 (1985) 119–192.

[36] W.P. Jones, M. Kakhi, Pdf modeling of finite-rate chemistry effects in turbulent nonpremixed jet flames, Combust. Flame 115 (1–2) (1998) 210–229.

[37] R.O. Fox, Computational Models for Turbulent Reacting Flows, Cambridge University Press, Cambridge, UK, 2003.

[38] H. Pitsch, Large-Eddy Simulation of turbulent combustion, Annu. Rev. Fluid Mech. 38 (2006) 453–482.

[39] S.B. Pope, Small scales, many species and the manifold challenges of turbulent combustion, Proc. Combust. Inst. 34 (2013) 1–31.

[40] S.B. Pope, Computations of turbulent combustion: progress and challenges, Proc. Combust. Inst. 23 (1990) 591–612.

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

[42] R.L. Plackett, A class of bivariate distributions, J. Am. Stat. Assoc. 310 (1965) 516–522.

[43] S. Ruan, N. Swaminathan, O.R. Darbyshire, Modelling of turbulent lifted jet flames using flamelets: *a priori* assessment and *a posteriori* validation, Combust. Theory Model. 18 (2) (2014) 295–329.

[44] Z. Chen, S. Ruan, N. Swaminathan, Simulation of turbulent lifted methane jet flames: effects of air-dilution and transient flame propagation, Combust. Flame 162 (2015) 703–716.

[45] M. Ihme, H. Pitsch, Prediction of extinction and reignition in nonpremixed turbulent flames using a flamelet/progress variable model: 1. *a priori* study and presumed pdf closure, Combust. Flame 155 (1-2) (2008) 70–89.

[46] K.N.C. Bray, Laminar flamelets in turbulent combustion modeling, Combust. Sci. Technol. 188 (2016) 1372–1375.

[47] J. Floyd, A.M. Kempf, A. Kronenburg, R.H. Ram, A simple model for the filtered density function for passive scalar combustion LES, Combust. Theory Model. 13 (4) (2009) 559–588. doi:10.1080/13647830802632200.

[48] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. 52 (1) (2020) 477–508. arXiv:1905.11075, doi:10.1146/annurev-fluid-010719-060214.

[49] I.T. Joliffe, Principal Component Analysis, 2nd Edition, Springer-Verlag, New York, USA, 2002.

[50] T. Verdonck M. Hubert, P. Rousseeuw, Robust PCA for skewed data and its outlier map, Comput. Stat. Data Anal. 53 (2009) 2264–2274.

[51] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift (2015) arXiv:1502.03167.

[52] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization (2014) arXiv: 1412.6980.

[53] Y. Minamoto, N. Swaminathan, R.S. Cant, T. Leung, Morphological and statistical features of reaction zones in MILD and premixed combustion, Combust. Flame 161 (11) (2014) 2801–2814. doi:10.1016/j.combustflame.2014.04.018.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84901181588{&}partnerID=tZOtx3y1

[54] R.S. Barlow, G. Magnotti, H.C. Cutcher, A.R. Masri, On defining progress variable for Raman/Rayleigh experiments in partially-premixed methane flames, Combust. Flame 179 (2017) 117–129. doi:http://dx.doi.org/10.1016/j.combustflame.2017.01.027.

[55] D. Wang, C. Tong, R.S. Barlow, A.N. Karpetis, Experimental study

of scalar filtered mass density function in turbulent partially premixed flames, Proc. Combust. Inst. 31 (1) (2007) 1533–1541. doi:10.1016/j.proci.2006.07.099.
URL http://www.sciencedirect.com/science/article/pii/S1540748906001088

[56] J. Cai, D. Wang, C. Tong, R.S. Barlow, A.N. Karpetis, Investigation of subgrid-scale mixing of mixture fraction and temperature in turbulent partially premixed flames, Proc. Combust. Inst. 32 (1) (2009) 1517–1525. doi:10.1016/j.proci.2008.05.026.
URL http://linkinghub.elsevier.com/retrieve/pii/S1540748908001612

[57] D.M. Endres, J.E. Schindelin, A new metric for probability distributions, IEEE Trans. Inform. Theory 49 (7) (2003) 1858–1860. doi:10.1109/TIT.2003.813506.

[58] D. Bradley, P.H. Gaskell, A.K.C. Lau, A mixedness-reactedness flamelet model for turbulent diffusion flames, Proc. Combust. Inst. 23 (1) (1990) 685–692.

[59] B. Fiorina, R. Baron, O. Gicquel, D. Thevenin, S. Carpentier, N. Darabiha, Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM, Combust. Theory Model. 7 (2003) 449–470.

[60] C.D. Pierce, P. Moin, Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion, J. Fluid Mech. 504 (2004) 73–97.

[61] J.A. van Oijen, A Donini, R.J.M. Bastiaans, J.H.M. ten Thije Boonkkamp, L.P.H. de Goey, State-of-the-art in premixed combustion modeling using flamelet generated manifolds, Prog. Energy Combust. Sci. 57 (2016) 30–74.

[62] A.Y. Klimenko, R.W. Bilger, Conditional moment closure for turbulent combustion, Prog. Energy Combust. Sci. 25 (6) (1999) 595–687.

[63] H. Steiner, W.K. Bushe, Large eddy simulation of a turbulent reacting jet with conditional source-term estimation, Phys. Fluids 13 (3) (2001) 754–769. arXiv:https://doi.org/10.1063/1.1343482, doi:10.1063/1.1343482.
URL https://doi.org/10.1063/1.1343482

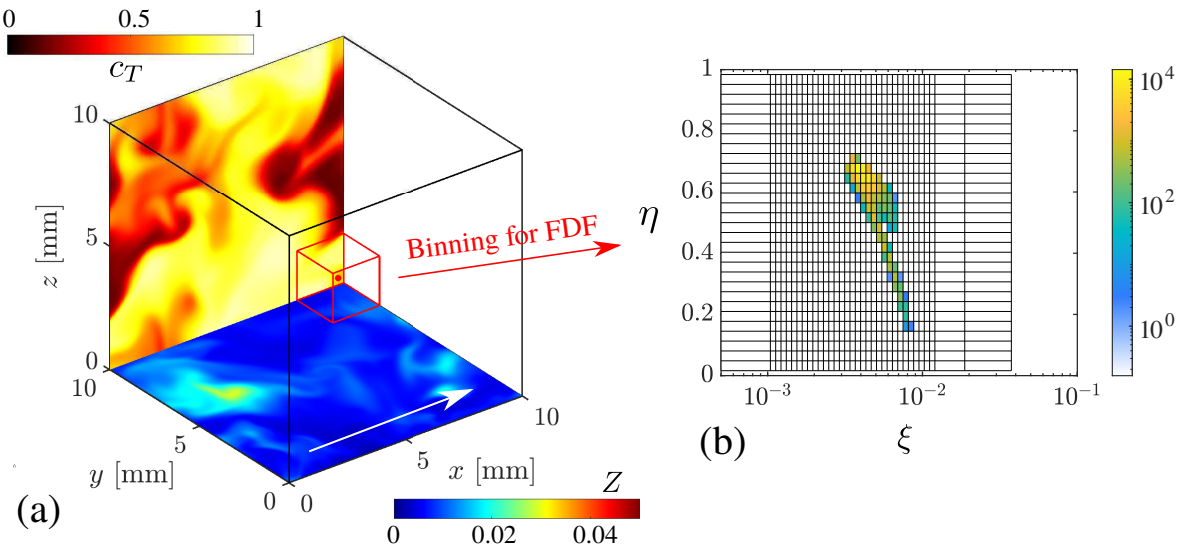

Fig. 1: Illustration of the joint FDF binning procedure using a typical DNS snapshot for case AZ1. White arrow in (a) indicates the mean flow direction. Colourbar values in (b) are normalised counts (i.e., densities).

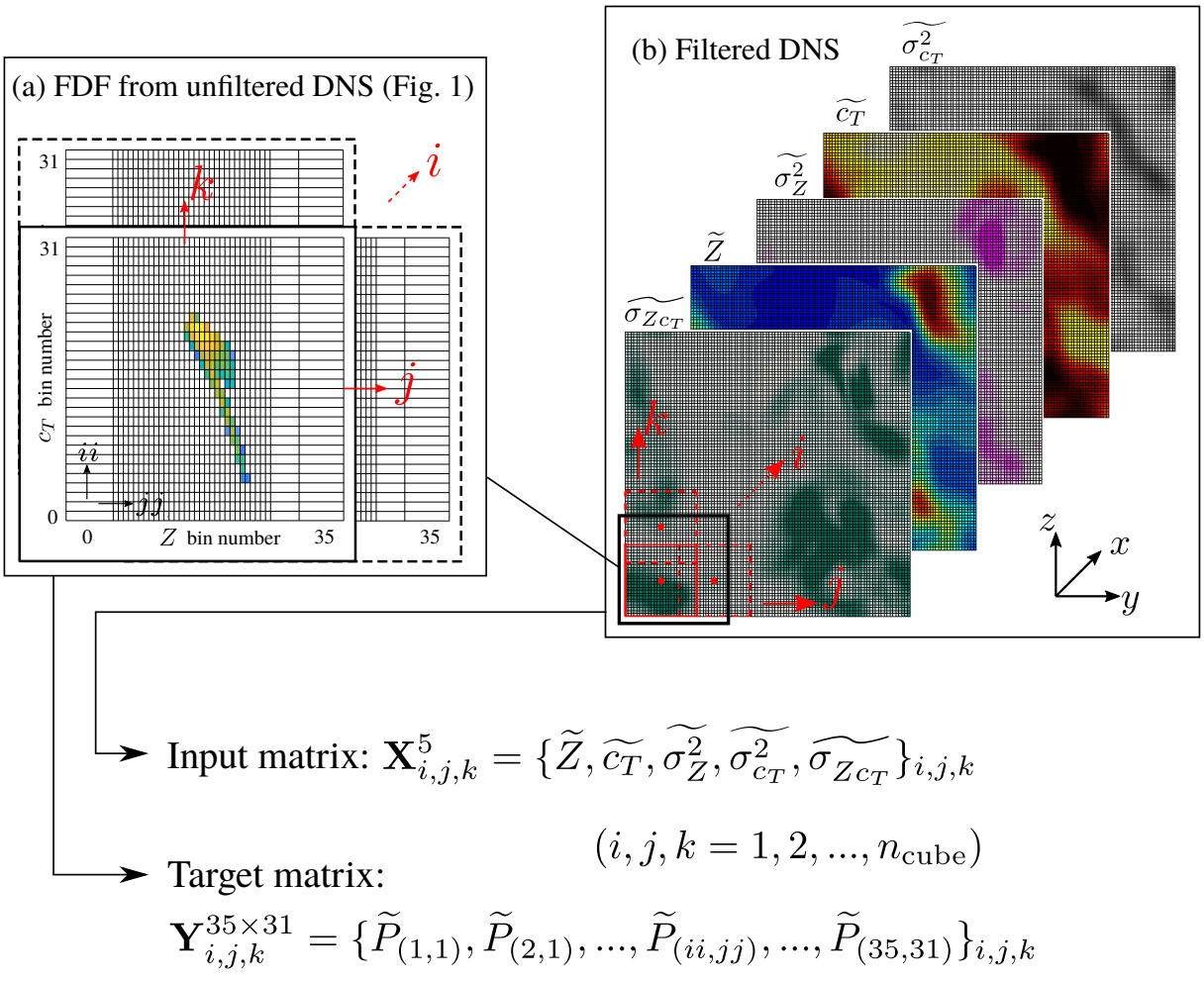

Fig. 2: Schematic demonstration of the ML training data extraction process from the DNS (same snapshot as in Fig. 1).

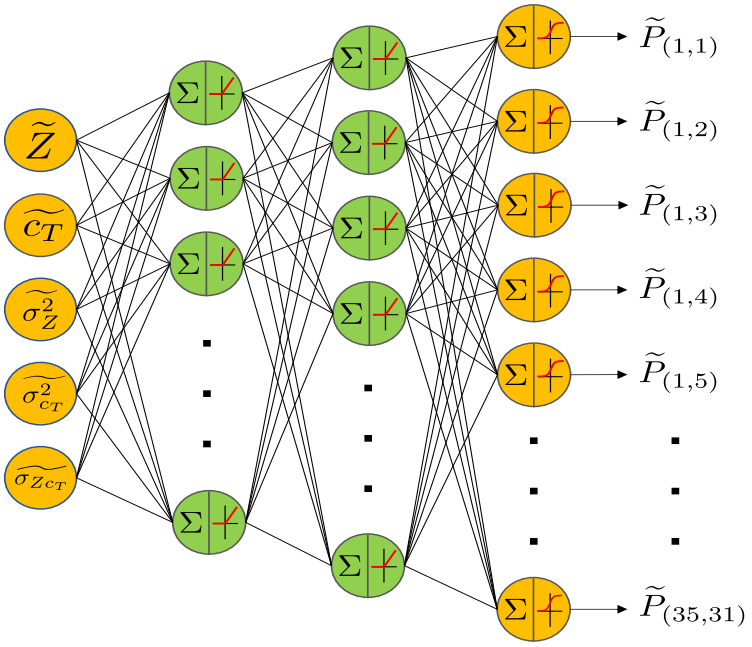

Fig. 3: Schematic of the DNN architecture with input and output matrices (yellow symbols) and hidden layers (green symbols).

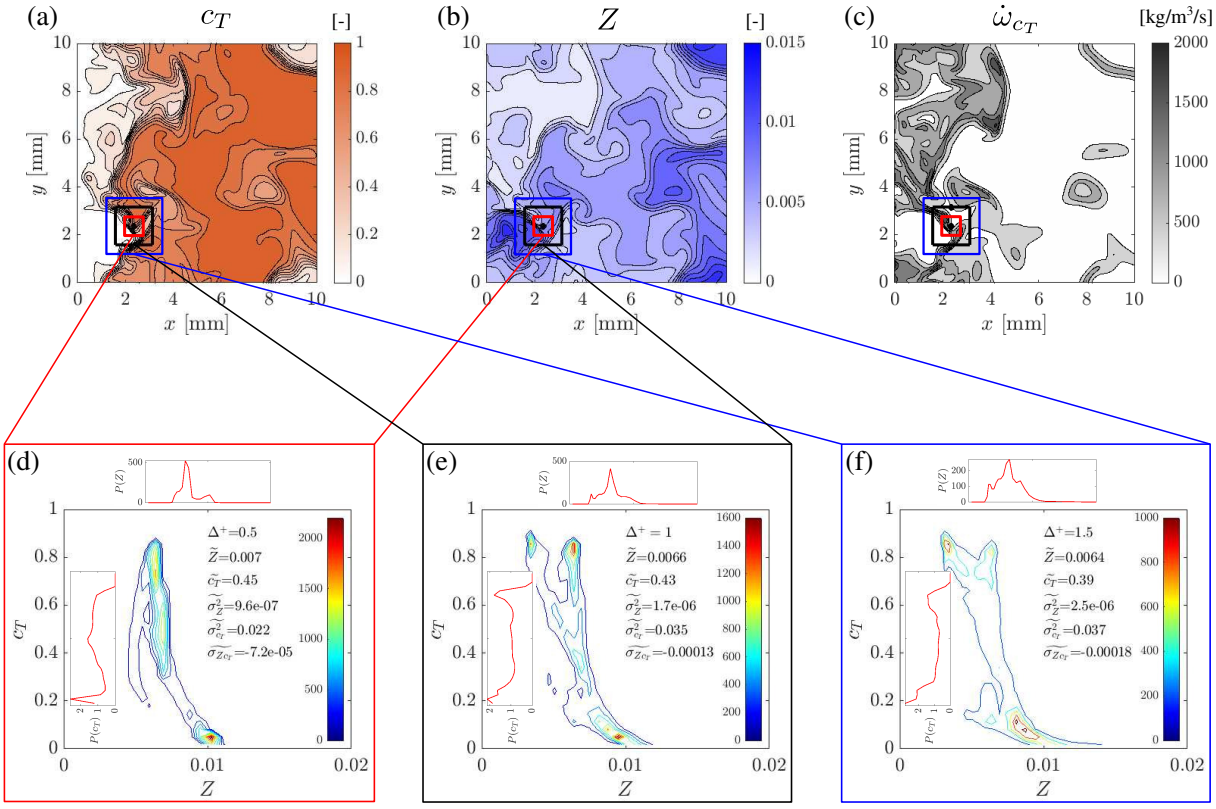

Fig. 4: Case AZ1: typical instantaneous x-y plane contours of unfiltered fields for (a) c_T, (b) Z and (c) $\dot{\omega}_{c_T}$; and FDF behaviours with three different filter sizes, (d) $\Delta^+ = 0.5$, (e) 1 and (f) 1.5, at an arbitrarily chosen point.

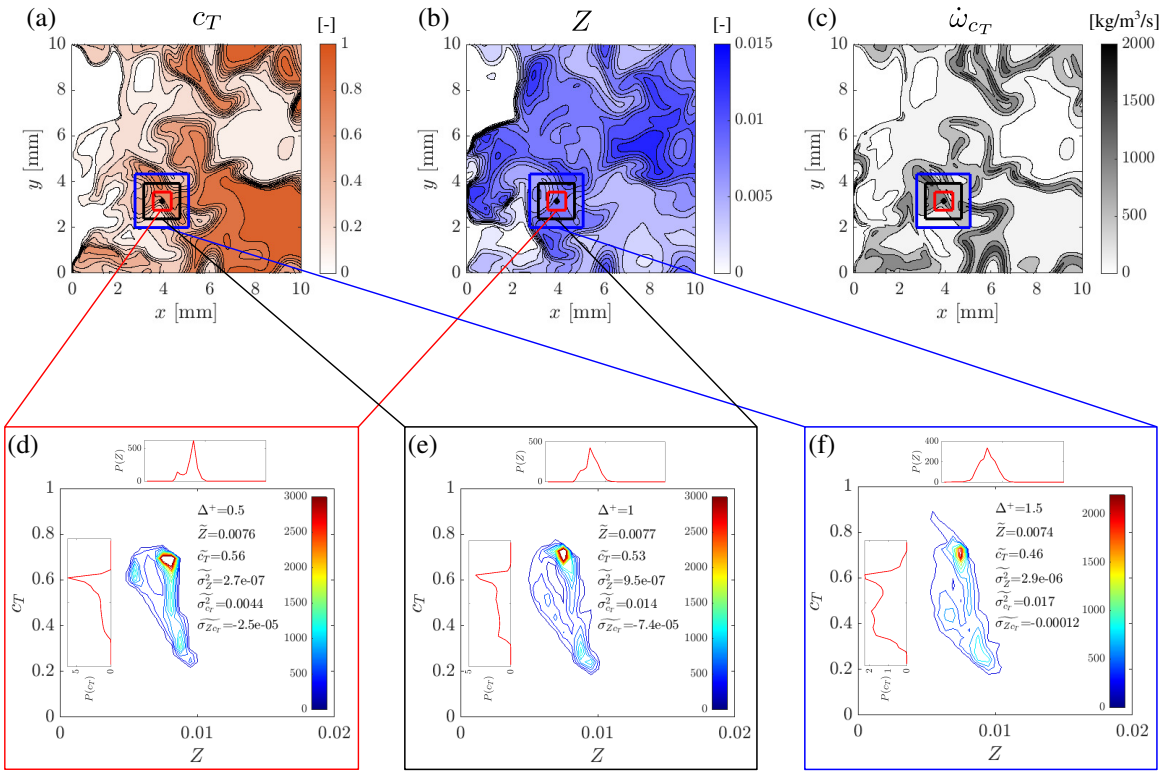

Fig. 5: Case AZ2: typical instantaneous x-y plane contours of unfiltered fields for (a) c_T, (b) Z and (c) $\dot{\omega}_{c_T}$; and FDFs with three different filter sizes, (d) $\Delta^+ = 0.5$, (e) 1 and (f) 1.5, at an arbitrarily chosen point.

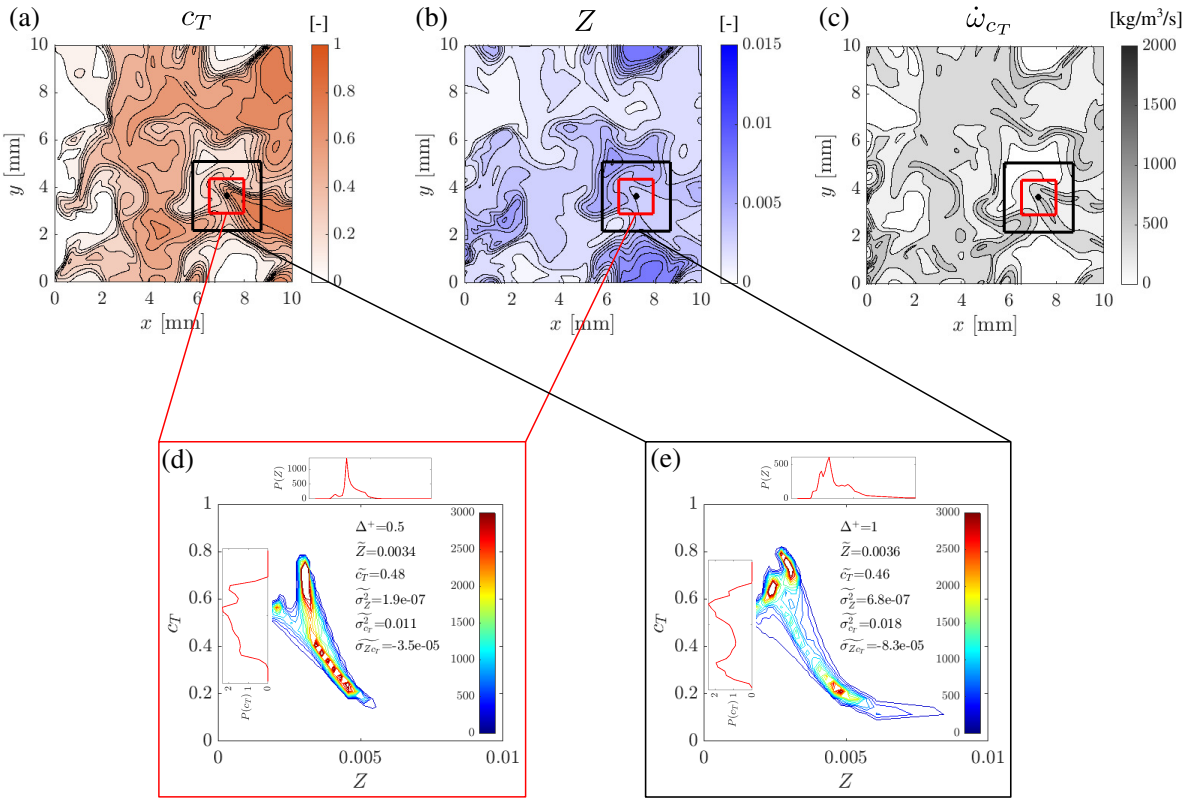

Fig. 6: Case BZ1: typical instantaneous x-y plane contours of unfiltered fields for (a) c_T, (b) Z and (c) $\dot{\omega}_{c_T}$; and FDF behaviours with two different filter sizes, (d) $\Delta^+ = 0.5$ and (e) 1, at an arbitrarily chosen point.

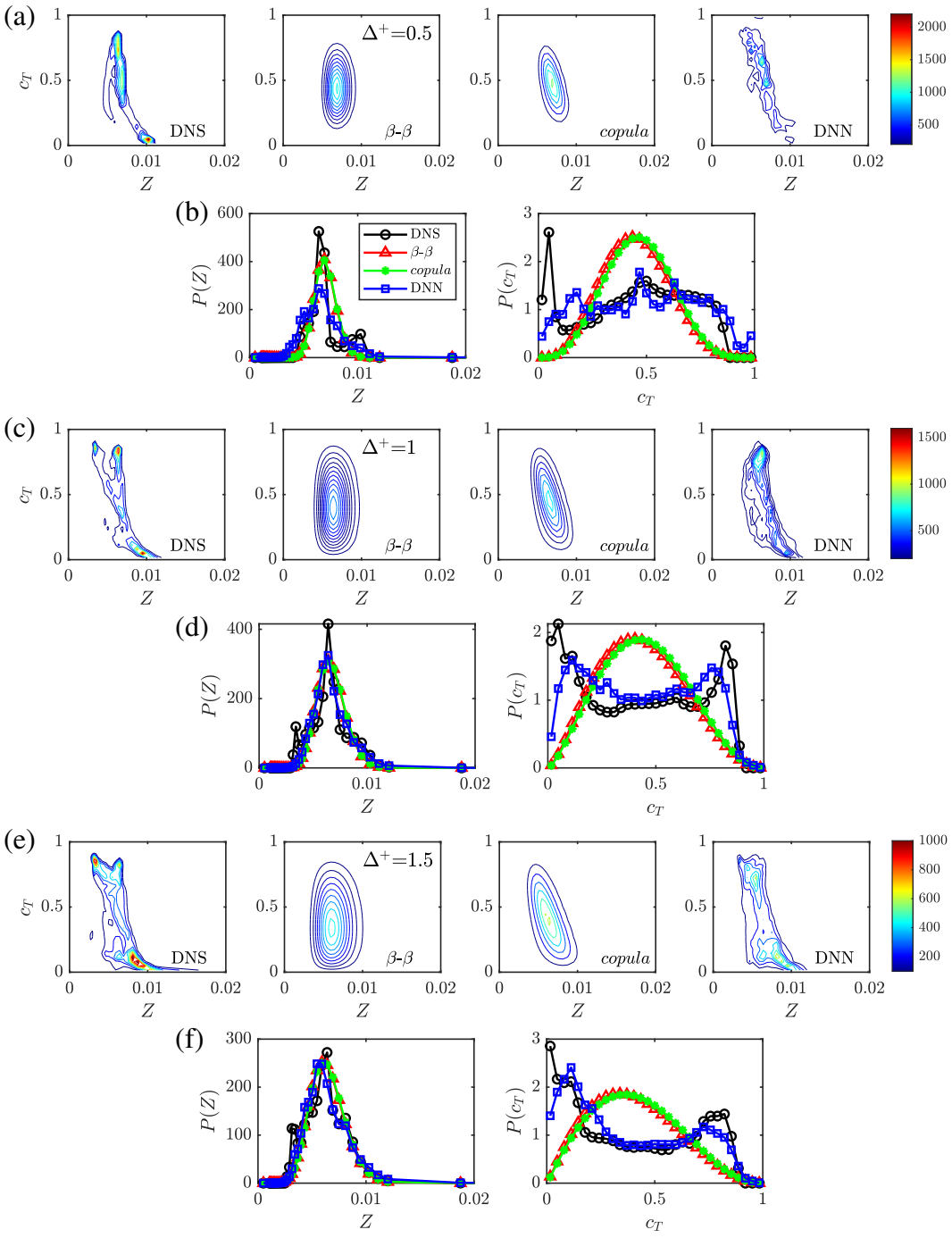

Fig. 7: Case AZ1: comparison of joint and marginal FDFs between DNS and model predictions for filter sizes of (a)-(b) $\Delta^+ = 0.5$, (c)-(d) $\Delta^+ = 1$ and (e)-(f) 1.5 for the filter point shown Fig. 4.

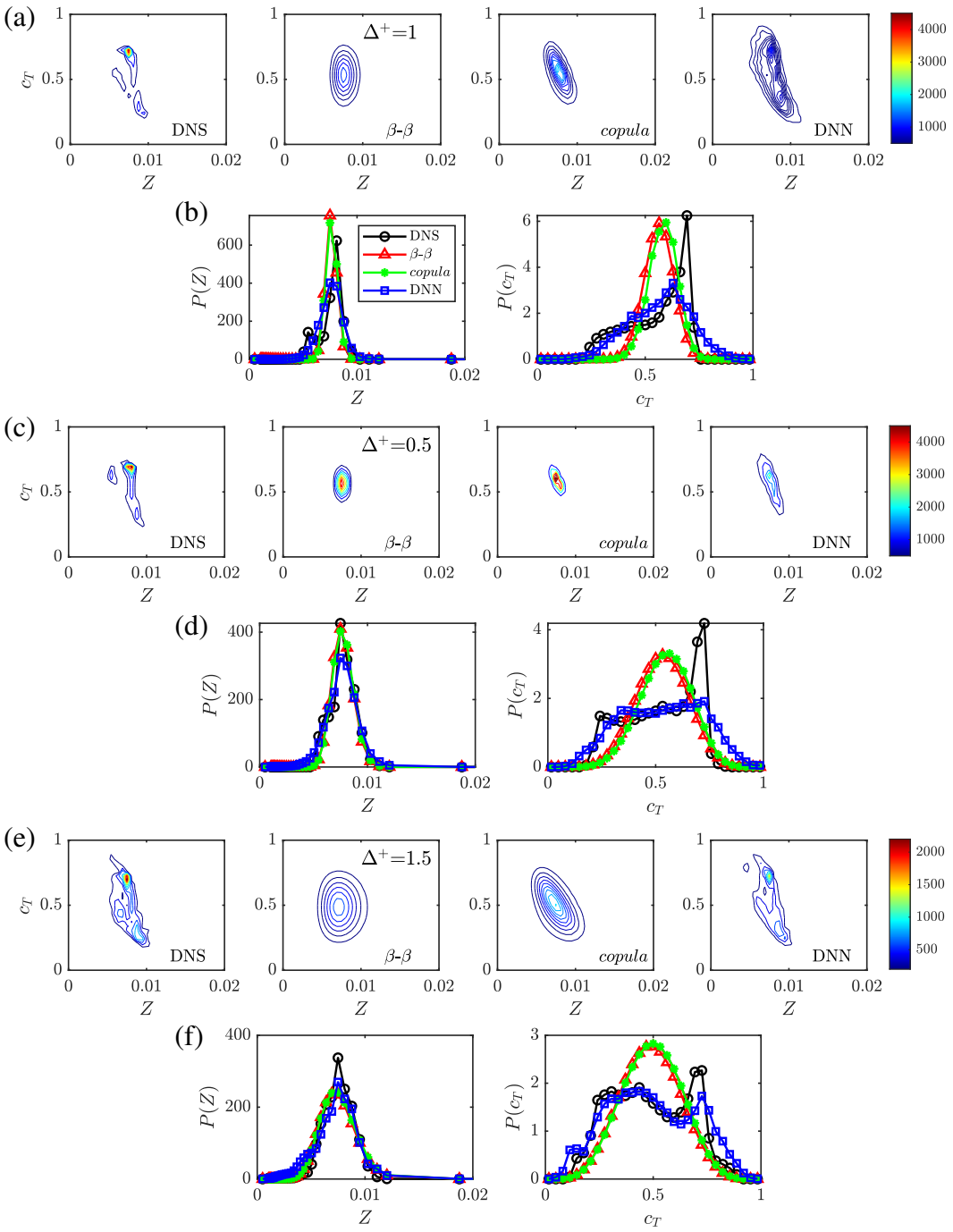

Fig. 8: Case AZ2: comparison of joint and marginal FDFs between DNS and model predictions for filter sizes of (a)-(b) $\Delta^+ = 0.5$, (c(-(d) $\Delta^+ = 1$ and (e)-(f) 1.5 for the filter point shown Fig. 5.

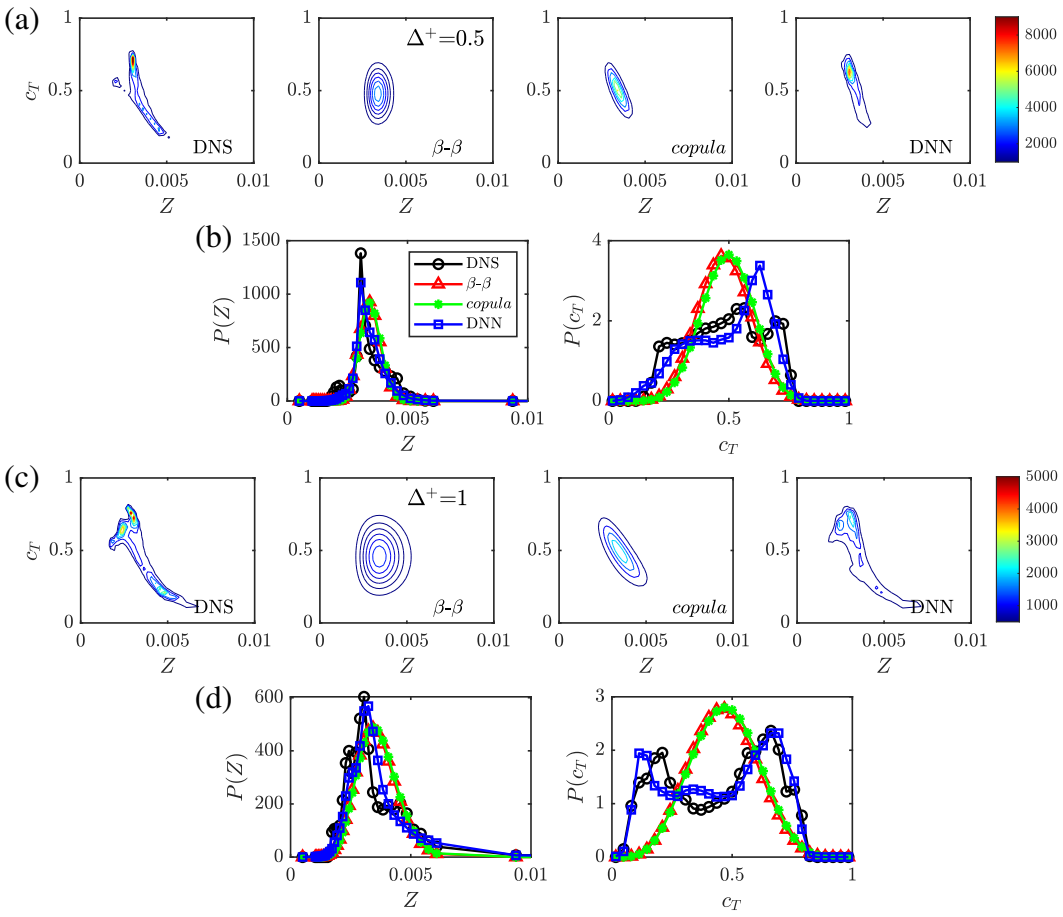

Fig. 9: Case BZ1: comparison of joint and marginal FDFs between DNS and model predictions for filter sizes of (a)-(b) $\Delta^+ = 0.5$ and (c)-(d) $\Delta^+ = 1$ for the filter point shown Fig. 6.

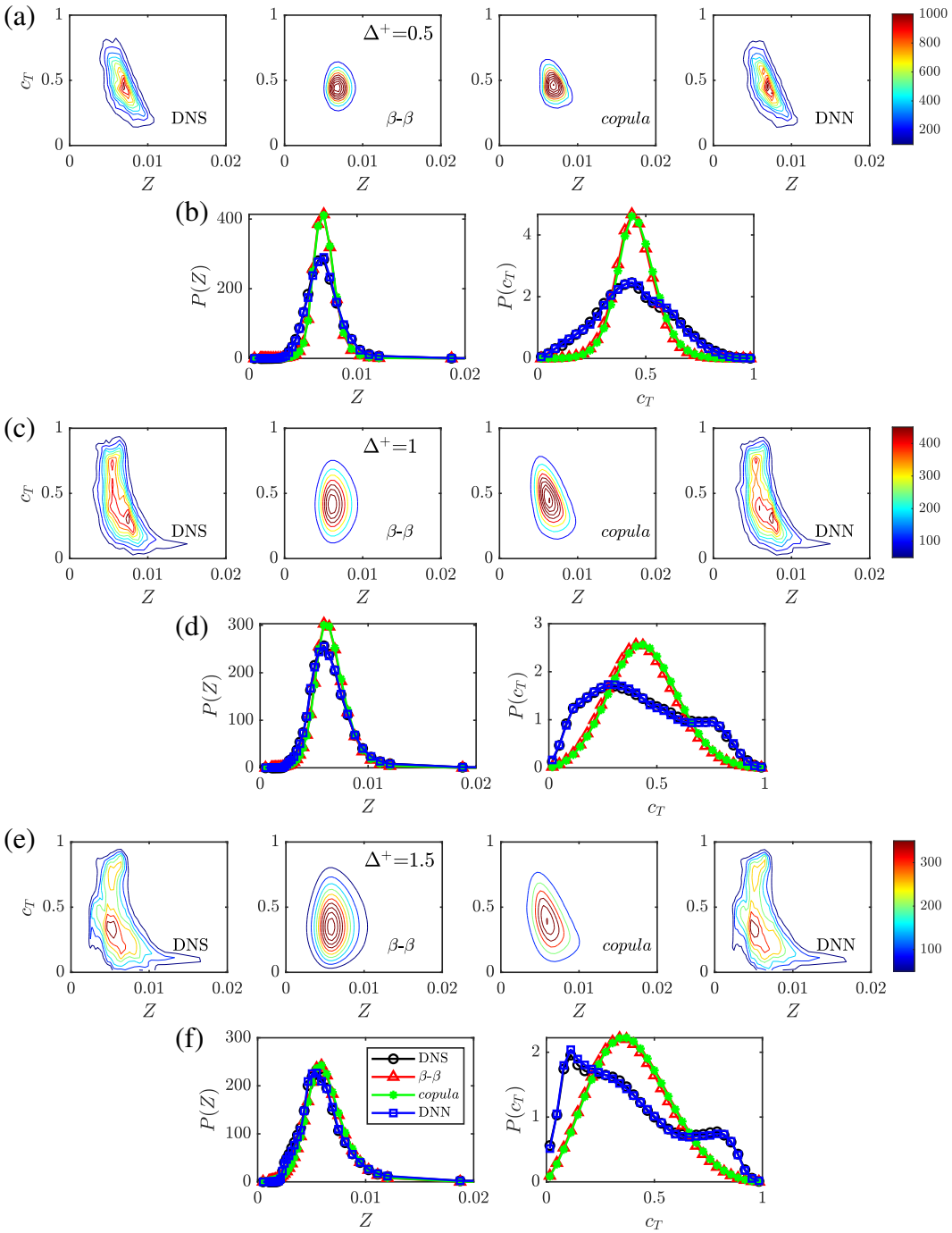

Fig. 10: Case AZ1: conditional joint and marginal FDFs for (a)-(b) $\Delta^+ = 0.5$, $\widetilde{Z} = 0.007$, $\widetilde{c_T} = 0.45$; (c)-(d) $\Delta^+ = 1$, $\widetilde{Z} = 0.0066$, $\widetilde{c_T} = 0.43$; and (e)-(f) $\Delta^+ = 1.5$, $\widetilde{Z} = 0.0064$, $\widetilde{c_T} = 0.39$.

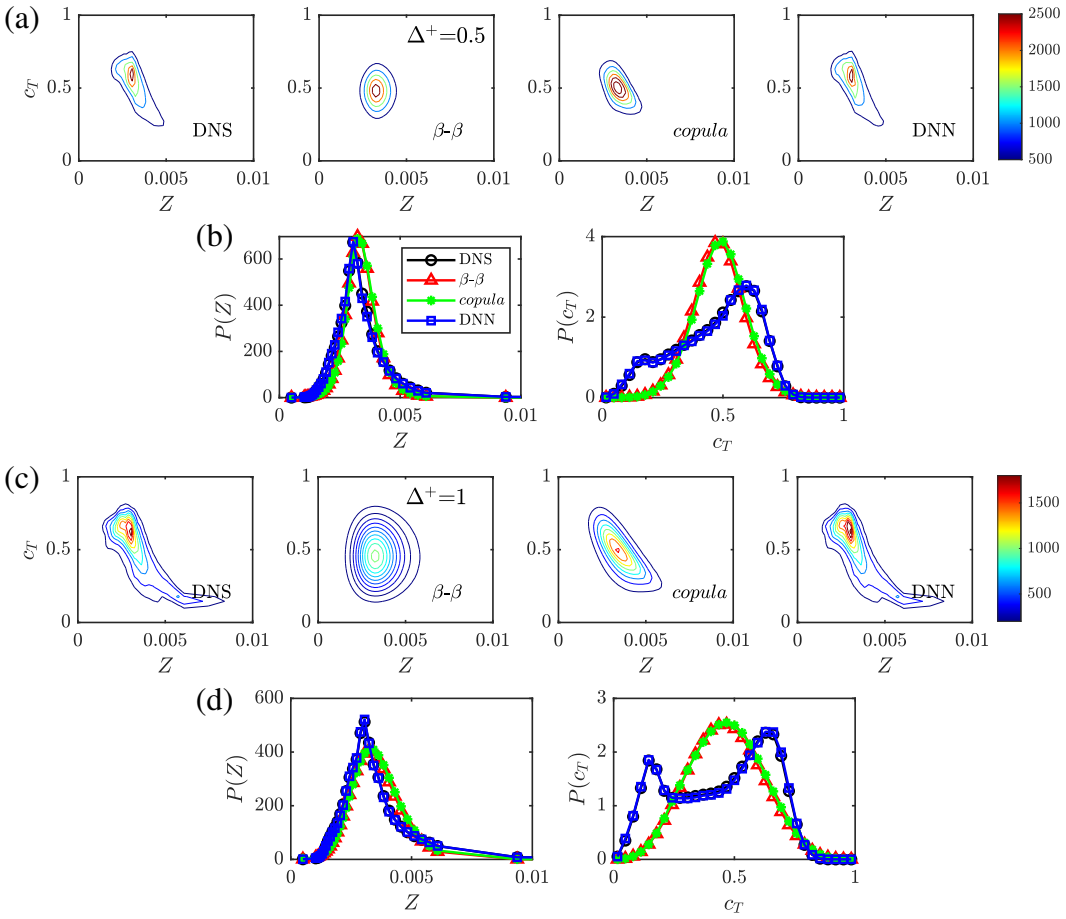

Fig. 11: Case BZ1: conditional joint and marginal FDFs for (a)-(b) $\Delta^+ = 0.5$, $\widetilde{Z} = 0.00034$, $\widetilde{c_T} = 0.48$; and (c)-(d) $\Delta^+ = 1$, $\widetilde{Z} = 0.0036$, $\widetilde{c_T} = 0.46$.

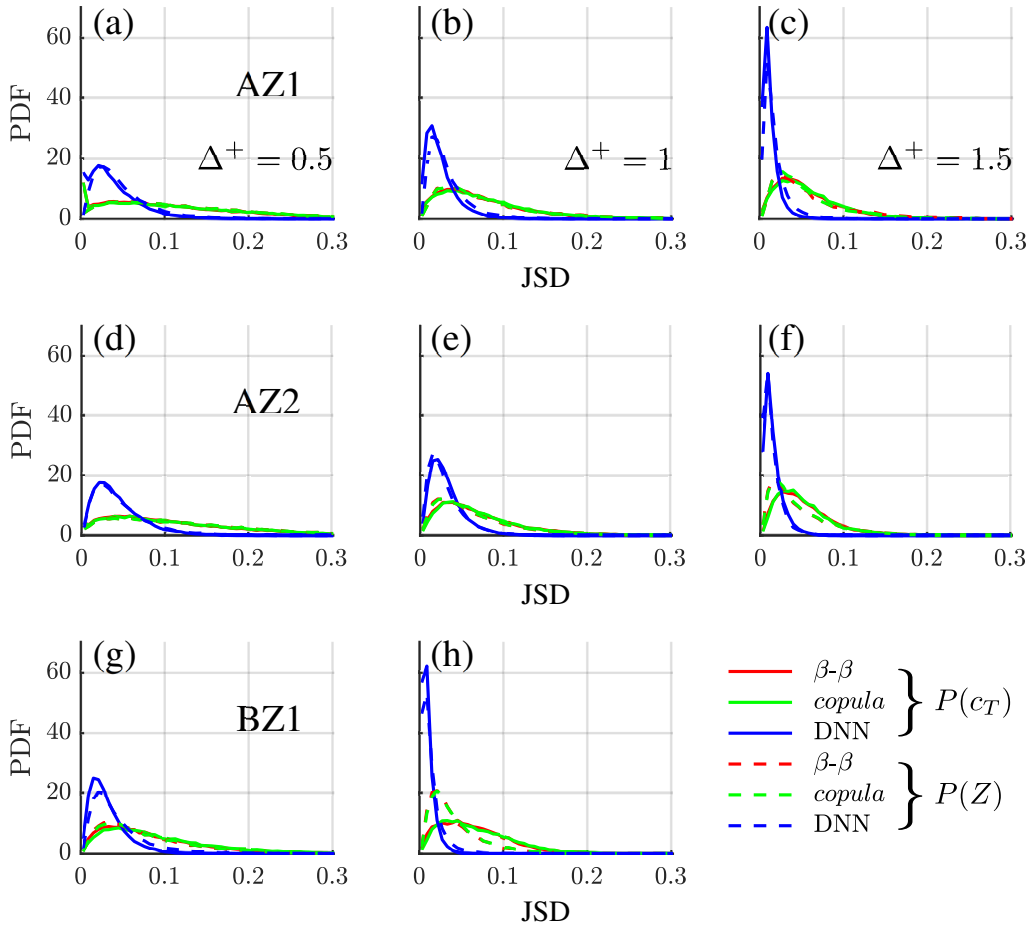

Fig. 12: Probability density function of Jansen-Shannon divergence for DNS and modelled FDFs. Solid and dished lines correspond to progress variable and mixture fraction, respectively.

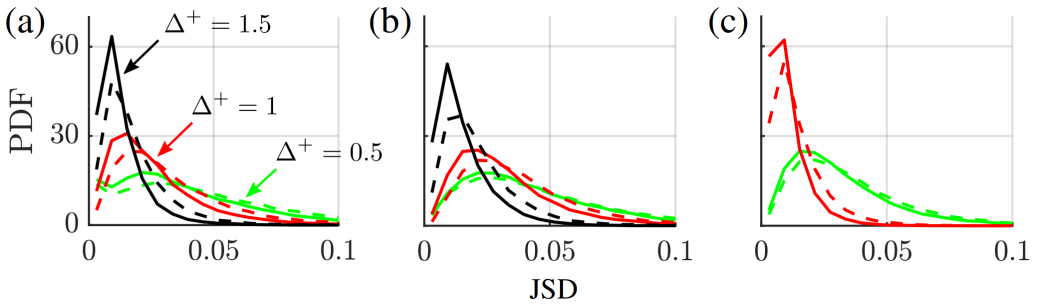

Fig. 13: Probability density function of Jansen-Shannon divergence for the DNN predicted progress variable FDFs for cases (a) AZ1, (b) AZ2 and (c) BZ1. Solid and dished lines correspond to learning with and without the covariance.

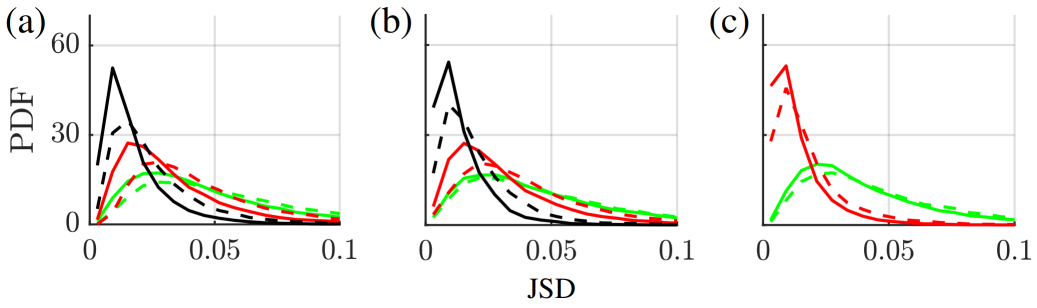

Fig. 14: Probability density function of Jansen-Shannon divergence for the DNN predicted mixture fraction FDFs for cases (a) AZ1, (b) AZ2 and (c) BZ1. Solid and dished lines correspond to learning with and without the covariance. See colour legend in Fig. 13.

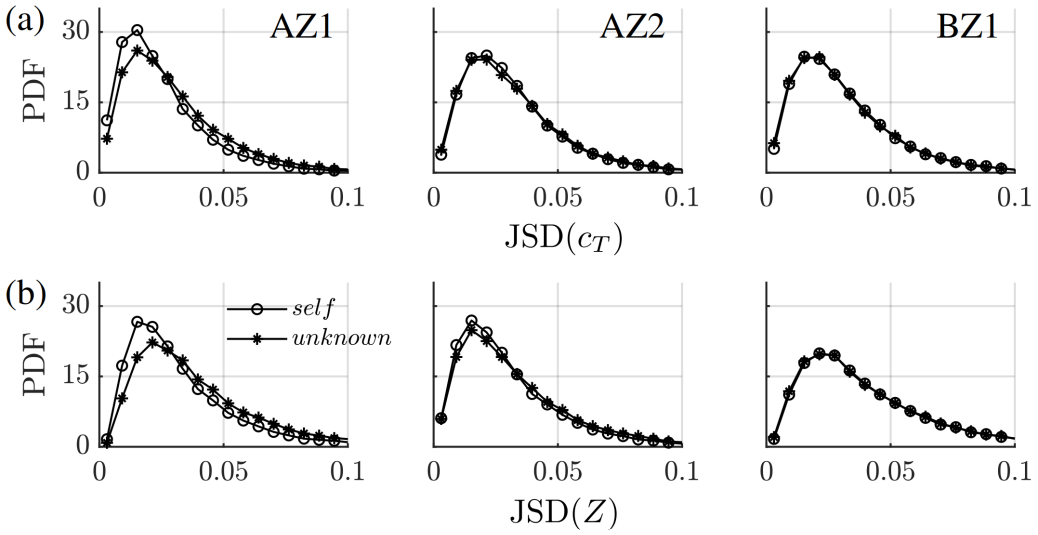

Fig. 15: Comparison of Jansen-Shannon divergence for DNN self- and unknown-predictions of FDF for (a) progress variable and (b) mixture fraction . The filter size for all cases is $\Delta^+ = 1$.

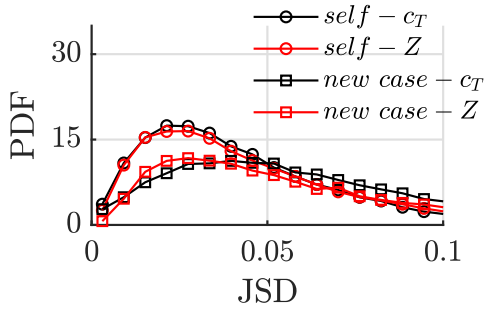

Fig. 16: Comparison of Jansen-Shannon divergence for case AZ2 FDF predictions using self and new case inputs. The filter size is $\Delta^+ = 0.5$.

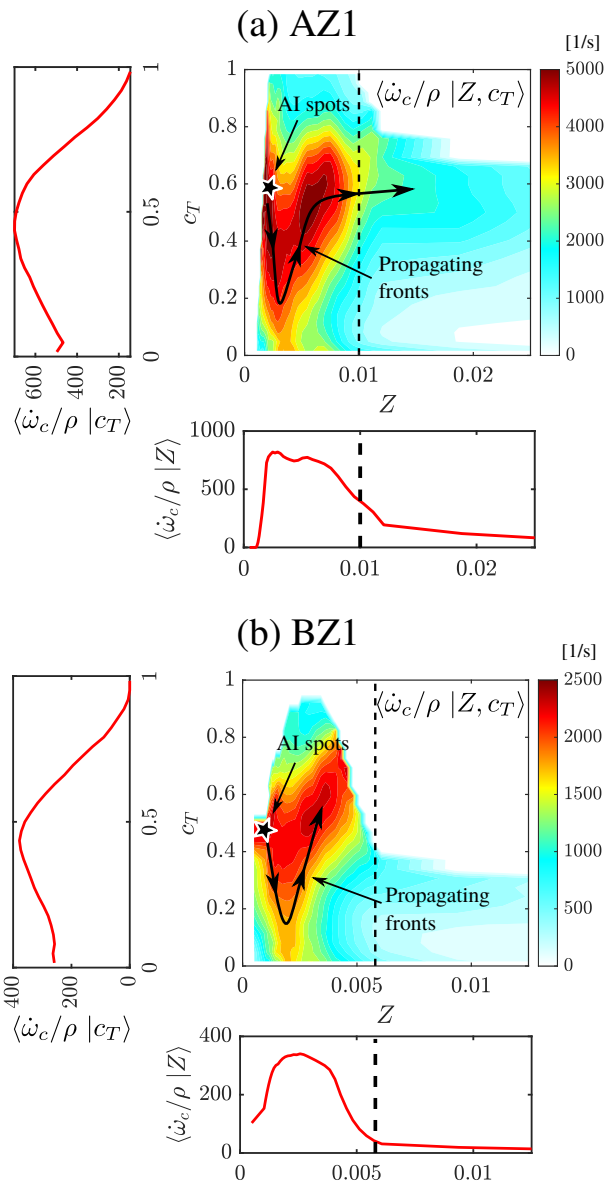

Fig. 17: Conditional mean reaction rates of progress variable. Dashed lines indicate stoichiometric mixture fraction.

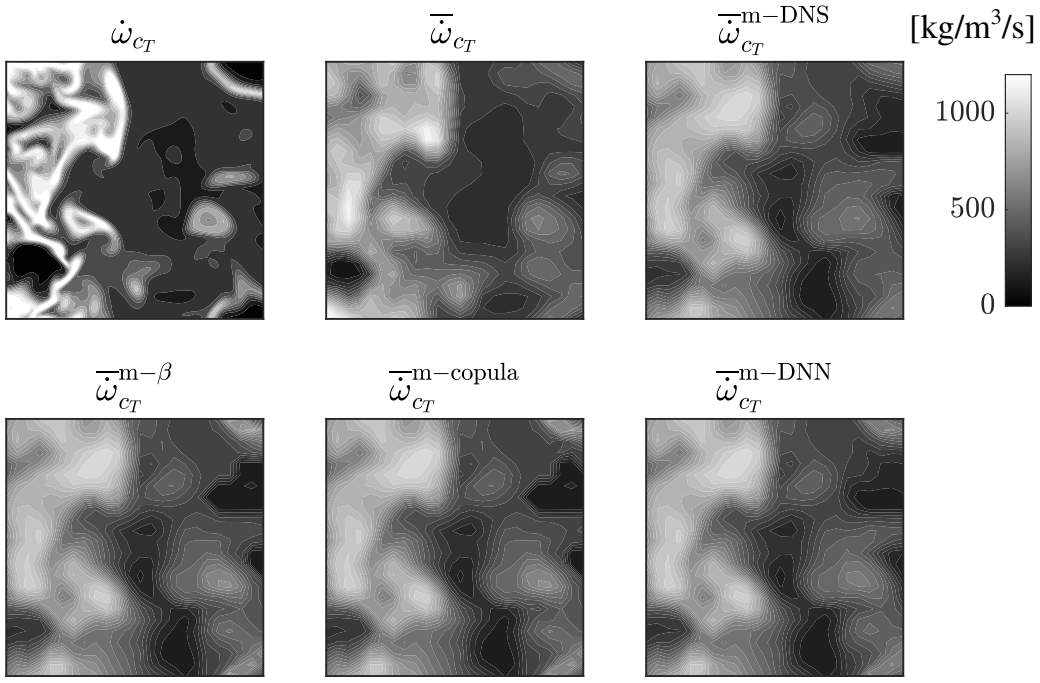

Fig. 18: Typical reaction rate contours for case AZ1 with $\Delta^+ = 0.5$. Mean flow moves from left to right.

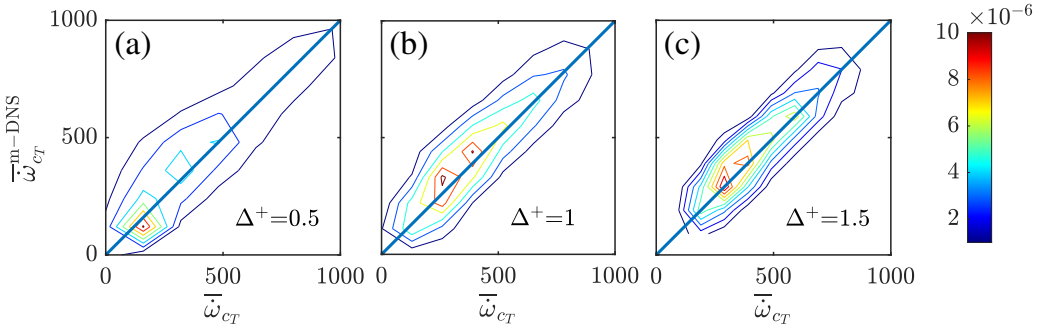

Fig. 19: Joint PDF of $\overline{\dot{\omega}}_{c_T}$ and $\overline{\dot{\omega}}^{\mathrm{m-DNS}}_{c_T}$ (in units of kg/m^3/s) for case AZ1 with different filter sizes.

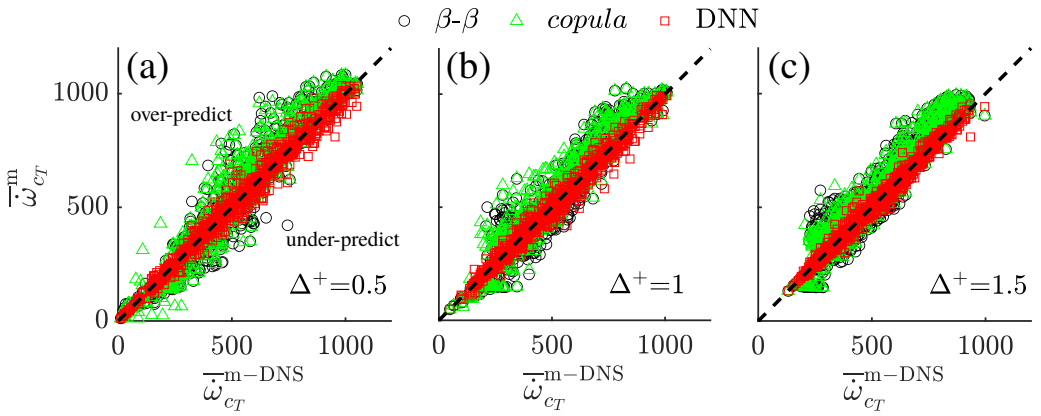

Fig. 20: Scatter plot of $\overline{\dot{\omega}}^{\mathrm{m-DNS}}_{c_T}$ and $\overline{\dot{\omega}}^{\mathrm{m}}_{c_T}$ (in units of kg/m^3/s) modelled using different FDF models (denoted using different markers) for case AZ1 with different filter sizes.

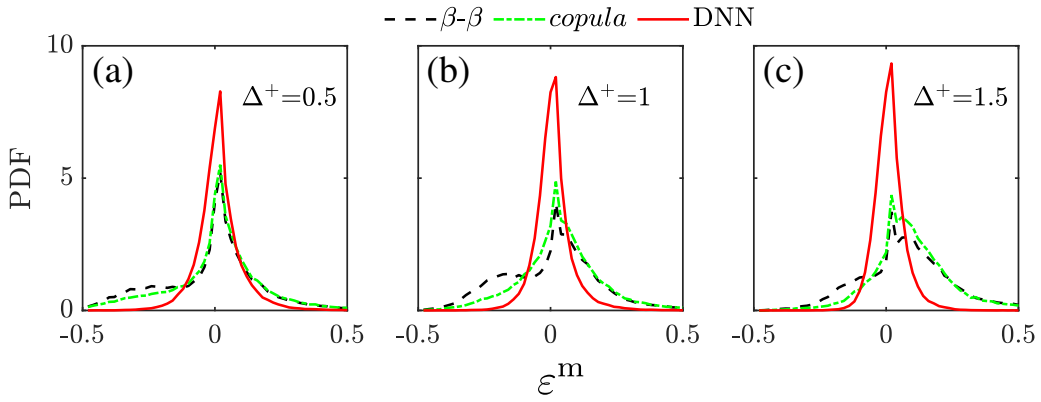

Fig. 21: PDFs of normalised prediction error for the filtered reaction rate using different FDF models. The samples for each filter size are collected from all three cases for each model.