

# JPResUnet: A Joint Probability Density Function Translation Model in Partially Premixed Flames

Hanying Yang<sup>a</sup>, James C. Massey<sup>a,b</sup>, Nedunchezhian Swaminathan<sup>a</sup>

<sup>a</sup>*Department of Engineering, University of Cambridge, Trumpington Street,  
Cambridge CB2 1PZ, United Kingdom*

<sup>b</sup>*Robinson College, University of Cambridge, Grange Road,  
Cambridge CB3 9AN, United Kingdom*

---

## Abstract

Machine learning (ML) models are often constrained by their limitations in extrapolation, which restricts their applicability in engineering contexts. Conversely, while exhibiting broad generality, many established scientific models seem to lack the necessary accuracy. This study addresses these challenges by introducing JPResUnet (Joint PDF Residual U-net), a novel model that integrates the strengths of both ML and traditional scientific approaches to predict sub-grid joint probability density functions (PDFs) in partially premixed flames. JPResUnet employs a residual U-Net architecture to translate classic  $\beta$ -PDFs to sub-grid PDFs. The model is trained using direct numerical simulation (DNS) data from methane–air moderate or intense low-oxygen dilution (MILD) combustion and is initially tested through *a priori* assessments on out-of-sample data. Comparative analyses against an artificial neural network (ANN) and the  $\beta$ -PDF approach demonstrate that JPResUnet consistently outperforms these methods in capturing complex sub-grid features with greater accuracy and robustness for both box and Gaussian kernels of varying widths. Further evaluations on a different case

reveal the model’s generalisability, where the ANN is unable to produce satisfactory prediction. Subsequent *a posteriori* assessment involves two versions of JPResUnet with different output PDF resolutions, which are deployed for large eddy simulation (LES) of a multi-regime burner through the look-up table (LUT) approach. The higher resolution model yields improvements in temperature estimates compared to the conventional LUT method. This highlights the potential of the JPResUnet model for robust and accurate LES of reacting flows with ML.

*Keywords:* Joint probability density function, Partially premixed flame, Residual U-net, MILD combustion, Multi-regime combustion

---

## 1. Introduction

Large eddy simulation (LES) has emerged as an important tool for the simulation of large-scale combustion systems, particularly in scenarios where fully resolving flow fields is computationally prohibitive. By employing a spatial filter, LES effectively resolves physical processes associated with scales larger than the filter width  $\Delta$ , while smaller scales, collectively referred to as the sub-grid scale (SGS) range, are modelled. Over the past half-century, turbulent convection and diffusion within the SGS range have been extensively studied, leading to the development of accurate models for both non-reacting [1] and reacting flows [2]. However, modelling sub-grid reactions presents significant challenges due to the highly nonlinear nature of chemical processes and their complex interactions with turbulence [2]. Although presumed probability density functions (PDFs) have been widely utilised to capture sub-grid fluctuations in thermo-chemical quantities for premixed [3, 4], non-premixed [5–7], and partially premixed [8–11] flames, this approach can encounter limitations under specific conditions, such as combustion with multiple fuel streams of varying composition, multi-regime combustion, and multi-phase flows. In the context of partially premixed flames, additional difficulties arise due to the strong correlation between the progress variable and the mixture fraction, whereas the presumed joint PDF approach typically assumes statistical independence between these two scalars at the sub-grid level [12, 13].

The use of machine learning (ML) as an alternative to traditional reaction rate modelling has been explored for over 30 years, with early studies focusing on the development of artificial neural networks (ANNs) to model

chemical kinetics [14–22]. Ihme *et al.* [23, 24] later proposed a systematic approach to optimise these techniques. The advent of deep learning has led to the exploration of more complex ANN structures for sub-grid PDF modelling. De Frahan *et al.* [25] validated the predictive capabilities of ANNs for the marginal PDF of the progress variable in swirling methane–air premixed flames. Similarly, Yao *et al.* [26] demonstrated that ANNs could predict the marginal PDF of the mixture fraction in turbulent spray flames more accurately than conventional presumed PDFs. Chen *et al.* [27] further extended this approach to predict the joint PDF of the progress variable and mixture fraction in a partially premixed flame, specifically in moderate or intense low-oxygen dilution (MILD) combustion. Despite these advancements, most studies are self-testing, meaning that the training and testing data share similar thermo-chemical and -physical conditions. To assess the generalisability of ANNs, Yang *et al.* [28] tested a model trained on MILD combustion data across five different premixed flames. The ANN performed well when applied to flames with the same fuel as in the MILD cases. However, its predictive accuracy diminished when there were deviations from the training conditions, such as increased filter size or a shift to hydrogen–air flames, underscoring the challenges for developing ML models with broad generalisability.

The rapid advancements in deep learning have led to the emergence of novel ML architectures with exceptional predictive capabilities, particularly in domains such as computer vision and natural language processing [29, 30]. These advances have also been leveraged in many fields, including biology and material science [31, 32]. In the context of combustion, Xing *et al.* [33] employed a deep convolutional neural network (CNN) with a U-net archi-

tecture, commonly used in image processing [34, 35], to predict the progress variable SGS variance. Trained on a planar flame in homogeneous isotropic turbulence and tested on a complex slot jet flame, the model demonstrated good agreement with direct numerical simulation (DNS), thus verifying its generalisability. Nista *et al.* [36, 37] utilised a generative adversarial network (GAN), originally designed for computer vision [38], to recover fully resolved flow fields from LES input. This GAN outperformed other models across various thermo-chemical and -physical test conditions.

Inspired by the image translation techniques in computer vision, this study aims to develop and test a U-net-based model, referred to as JPResUnet (Joint PDF Residual U-net), to infer joint sub-grid PDFs based on presumed distributions while maintaining generalisability. The model is trained on data from a single MILD combustion case and tested on unseen samples from different cases with varying filter sizes and types. Additionally, *a posteriori* assessment through LES for a multi-regime burner (MRB) is conducted to evaluate the model's practical performance.

The structure of the paper is as follows. Section 2 provides a theoretical overview of turbulent combustion modelling in LES and the concept of PDF translation. Section 3 details the methodology for training the JPResUnet model and describes the training dataset. Section 4 compares the joint PDFs predicted by the JPResUnet and corresponding values from DNS, alongside a discussion of the filtered reaction rate modelled via these PDFs. Results from *a posteriori* assessment are discussed in Section 5, and conclusions are summarised in the final section.

## 2. Theoretical background

### 2.1. Governing equations

In terms of the LES implemented by present work, the mass, momentum and total enthalpy are transported as

$$\frac{\partial \bar{\rho}}{\partial t} + \nabla \cdot (\bar{\rho} \tilde{\mathbf{U}}) = 0, \quad (1)$$

$$\bar{\rho} \frac{D \tilde{\mathbf{U}}}{Dt} = -\nabla \bar{p} + \nabla \cdot \left[ \bar{\tau} - \left( \bar{\rho} \tilde{\mathbf{U}} \tilde{\mathbf{U}} - \bar{\rho} \tilde{\mathbf{U}} \tilde{\mathbf{U}} \right) \right], \quad (2)$$

$$\bar{\rho} \frac{D \tilde{h}}{Dt} = \nabla \cdot \left[ \bar{\rho} \alpha \nabla \tilde{h} - \left( \bar{\rho} \tilde{\mathbf{U}} \tilde{h} - \bar{\rho} \tilde{\mathbf{U}} \tilde{h} \right) \right]. \quad (3)$$

The symbols  $\rho$ ,  $\mathbf{U}$  and  $h$  represent density, velocity vector and thermochemical enthalpy (sum of sensible and chemical enthalpies) respectively. Filtering and Favre-filtering operations are denoted by  $\bar{\cdot}$  and  $\tilde{\cdot}$ , respectively. The operator  $D/Dt = \partial/\partial t + \tilde{\mathbf{U}} \cdot \nabla$  is the material derivative. The shear stress  $\tau$  in Eq. (2) is calculated as  $\bar{\tau} = 2\mu[\tilde{\mathbf{S}} - 1/3(\nabla \cdot \tilde{\mathbf{U}})\mathbf{I}]$ , where  $\mu$  is the molecular dynamic viscosity,  $\tilde{\mathbf{S}}$  is the strain rate tensor defined as  $0.5[\nabla \tilde{\mathbf{U}} + (\nabla \tilde{\mathbf{U}})^T]$ , and  $\mathbf{I}$  is the identity matrix. The term within the parentheses in Eq. (2) represents the residual stress tensor  $\bar{\tau}^R$ . It is unclosed and modelled as  $\bar{\tau}^R = 2(\bar{\rho} \tilde{k}_{sgs} \mathbf{I})/3 - 2\bar{\rho} \nu_T [\tilde{\mathbf{S}} - (\nabla \cdot \tilde{\mathbf{U}})\mathbf{I}/3]$ , where  $\tilde{k}_{sgs}$  is the turbulent kinetic energy at the sub-grid scale and  $\nu_T$  is the turbulent viscosity. The latter is modelled using the  $\sigma$ -model as  $\nu_T = (C_\sigma \Delta)^2 \mathcal{D}_\sigma$  [39], where the model constant  $C_\sigma = 1.5$  is decided similar to [40]. The operator  $\mathcal{D}_\sigma$  denotes the differential operator related to the singular values of the Jacobian matrix of the velocity. The molecular thermal diffusivity  $\alpha$  in Eq. (3) is calculated as  $\bar{\rho} \alpha = \bar{\rho} \nu / \text{Pr}$ , where  $\nu$  is the molecular kinematic viscosity, and  $\text{Pr} = 0.7$  is the Prandtl number. The sub-grid scalar flux in Eq. (3) is

modelled as  $\bar{\rho}\widetilde{\mathbf{U}}\widetilde{h} - \bar{\rho}\widetilde{\mathbf{U}}\widetilde{h} = -\bar{\rho}(\nu_T/\text{Pr}_T)\nabla\widetilde{h}$ , where  $\text{Pr}_T = 0.7$  is the turbulent Prandtl number following the previous study on the MRB considered in the *a posteriori* assessment [40].

## 2.2. Combustion closure

A joint PDF-based tabulation method [12, 41–43] is used for combustion modelling. During the simulation, unclosed thermo-chemical scalars are retrieved from a four-dimensional look-up table (LUT) parameterised by the filtered progress variable  $\widetilde{c}$ , the mixture fraction  $\widetilde{Z}$  and their respective sub-grid variance  $\widetilde{\sigma}_{c,\text{sgs}}^2$  and  $\widetilde{\sigma}_{Z,\text{sgs}}^2$ . The progress variable  $c$  is defined as  $(Y_{\text{CO}} + Y_{\text{CO}_2})/(Y_{\text{CO}} + Y_{\text{CO}_2})^{\text{eq}}$ , where the superscript ‘eq’ denotes the equilibrium state for a given mixture fraction following previous studies that used methane–air mixtures [12, 44]. The mixture fraction  $Z$  is defined according to Bilger [45]. These quantities are transported using

$$\bar{\rho}\frac{D\widetilde{c}}{Dt} \approx \nabla \cdot \left( \bar{\rho}\overline{\mathcal{D}} + \bar{\rho}\frac{\nu_T}{\text{Sc}_T} \right) \nabla\widetilde{c} + \bar{\omega}^*, \quad (4)$$

$$\bar{\rho}\frac{D\widetilde{Z}}{Dt} \approx \nabla \cdot \left( \bar{\rho}\overline{\mathcal{D}} + \bar{\rho}\frac{\nu_T}{\text{Sc}_T} \right) \nabla\widetilde{Z}, \quad (5)$$

$$\begin{aligned} \bar{\rho}\frac{D\widetilde{\sigma}_{c,\text{sgs}}^2}{Dt} &\approx \nabla \cdot \left( \bar{\rho}\overline{\mathcal{D}} + \bar{\rho}\frac{\nu_T}{\text{Sc}_T} \right) \nabla\widetilde{\sigma}_{c,\text{sgs}}^2 - 2\bar{\rho}\widetilde{\chi}_{c,\text{sgs}} \\ &+ 2\bar{\rho}\frac{\nu_T}{\text{Sc}_T}(\nabla\widetilde{c} \cdot \nabla\widetilde{c}) + 2(\bar{c}\bar{\omega}^* - \widetilde{c}\bar{\omega}^*), \end{aligned} \quad (6)$$

$$\begin{aligned} \bar{\rho}\frac{D\widetilde{\sigma}_{Z,\text{sgs}}^2}{Dt} &\approx \nabla \cdot \left( \bar{\rho}\overline{\mathcal{D}} + \bar{\rho}\frac{\nu_T}{\text{Sc}_T} \right) \nabla\widetilde{\sigma}_{Z,\text{sgs}}^2 - 2\bar{\rho}\widetilde{\chi}_{Z,\text{sgs}} \\ &+ 2\bar{\rho}\frac{\nu_T}{\text{Sc}_T}(\nabla\widetilde{Z} \cdot \nabla\widetilde{Z}), \end{aligned} \quad (7)$$

where the turbulent Schmidt number  $\text{Sc}_T$  is assigned as a constant value of 0.4 following the previous study on the MRB considered in the *a posteriori*

assessment [40]. The term  $\overline{\rho\mathcal{D}}$  is calculated as  $\bar{\rho}\nu/\text{Sc}$  with the molecular Schmidt number  $\text{Sc} = 0.7$  [46]. The sub-grid scalar dissipation rate (SDR) for the progress variable and mixture fraction are denoted respectively as  $\tilde{\chi}_{c,\text{sgs}}$  and  $\tilde{\chi}_{Z,\text{sgs}}$ .  $\tilde{\chi}_{c,\text{sgs}}$  is modelled using a model proposed by Dunstan et al. [47], which has been validated by many studies [12, 41, 42, 46].  $\tilde{\chi}_{Z,\text{sgs}}$  is modelled using a linear relaxation model as  $\tilde{\chi}_{Z,\text{sgs}} = C_Z(\nu_T/\Delta^2)\sigma_{Z,\text{sgs}}^2$  with a constant  $C_Z = 2$  [48–50].

The source term  $\overline{\dot{\omega}^*}$  in Eq. (4) and (6) originates from three components [10, 51]: premixed combustion  $\overline{\dot{\omega}}_p$ , non-premixed combustion  $\overline{\dot{\omega}}_{np}$ , and their interaction through the cross dissipation rate  $\overline{\dot{\omega}}_{cdr}$ . This cross dissipation term is neglected following previous studies [51, 52]. The premixed combustion term is modelled as

$$\overline{\dot{\omega}}_p = \overline{\rho} \int_0^1 \int_0^1 \frac{\dot{\omega}(\eta, \xi)}{\rho(\eta, \xi)} \widetilde{P}(\eta, \xi) d\eta d\xi, \quad (8)$$

where  $\eta$  and  $\xi$  are the sample space variables for  $c$  and  $Z$  respectively. The flamelet reaction rate  $\dot{\omega}$  and density  $\rho$  are obtained by solving one-dimensional unstrained planar laminar premixed flames over the flammability of the methane–air flame using the GRI-Mech 3.0 chemical mechanism [53] in Cantera [54]. The non-premixed combustion mode is modelled by using the marginal PDF of the mixture fraction as

$$\overline{\dot{\omega}}_{np} = \overline{\rho c} \widetilde{\chi}_Z \int_0^1 \frac{1}{\psi^{\text{eq}}(\xi)} \frac{d^2\psi^{\text{eq}}(\xi)}{d\xi^2} \widetilde{P}(\xi) d\xi, \quad (9)$$

where  $\widetilde{\chi}_Z = \widetilde{D}(\nabla \widetilde{Z} \cdot \nabla \widetilde{Z}) + \widetilde{\chi}_{Z,\text{sgs}}$ , as seen in the work [12] and  $\psi = Y_{\text{CO}} + Y_{\text{CO}_2}$ . Another source term in Eq. (6) is assumed as  $\overline{c\dot{\omega}^*} \approx \overline{c\dot{\omega}_p}$  [52], and it is modelled similar to Eq. (8).

### 2.3. Presumed joint PDF

Typically, the joint PDF in Eq. (8) is calculated as the product of two marginal PDFs of progress variable and mixture fraction [40, 43], based on the assumption of the statistical independence between these two scalars at sub-grid level. Hence,

$$\tilde{P}(\eta, \xi) = \tilde{P}_\beta(\eta; \tilde{c}, \tilde{\sigma}_{c,\text{sgs}}^2) \times \tilde{P}_\beta(\xi; \tilde{Z}, \tilde{\sigma}_{Z,\text{sgs}}^2), \quad (10)$$

where the marginal PDF is presumed with the  $\beta$ -PDF distribution. This presumed distribution is calculated as

$$\tilde{P}(\eta; \tilde{c}, \tilde{\sigma}_{c,\text{sgs}}^2) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \eta^{a-1} (1-\eta)^{b-1}, \quad (11)$$

where  $a = \tilde{c}(1/\tilde{g}_c - 1)$ ,  $b = (1-\tilde{c})(1/\tilde{g}_c - 1)$  and  $\Gamma$  is the *gamma* function.  $\tilde{g}_c$  is the segregation factor, calculated as  $\tilde{g}_c = \tilde{\sigma}_{c,\text{sgs}}^2 / (\tilde{c}(1-\tilde{c}))$ . The marginal PDF for the mixture fraction is calculated in a similar way.

The statistical independence in Eq. (10) may not hold in partially premixed flames due to the evident correlation between the progress variable and mixture fraction at the SGS level [27, 55] when the grid width is large. Alternative models, such as the *copula* [56], which accounts for the cross-correlation, have been suggested [55] but require additional transport equations for sub-grid covariance between  $c$  and  $Z$ , leading to increased complexity and need for further modelling. Consequently, the present work utilises the presumed joint PDF formulation (Eq. (10)) and considers incorporating *copula* models in future studies.

### 2.4. PDF-to-PDF translation model

To improve the accuracy of PDF predictions, this study develops the JPResUnet model, a PDF-to-PDF translation framework inspired by su-

pervised image-to-image translation techniques. This model translates from a source probability space  $\{\Omega, \mathcal{F}, \mathcal{P}\}_A$  to the target probability space  $\{\Omega, \mathcal{F}, \mathcal{P}\}_B$ , where  $\Omega$  is the sample space,  $\mathcal{F}$  denotes the  $\sigma$ -algebra on  $\Omega$ , and  $\mathcal{P}$  is the probability measure. Regarding the partially premixed flame, the sample space  $\Omega$  is 2-dimensional, constituted by the progress variable  $c \in [0, 1]$  and mixture fraction  $Z \in [0, 1]$ . The model is designed for self-translation, implying that the source and target probability spaces originate from the same flame field, with identical  $\sigma$ -algebras, i.e.,  $\mathcal{F}_A = \mathcal{F}_B = \mathcal{F}$ . Therefore, the translation is constrained to the probability measure  $\mathcal{P}$ , which is described through the PDF  $P$  for the continuous random variables. This process is displayed as,

$$P_{AB}(\eta, \xi) = \mathcal{M}_\theta^{A \rightarrow B}(P_A(\eta, \xi)), \text{ for } (\eta, \xi) \in \mathcal{F}, \quad (12)$$

where the subscript  $\theta$  denotes the parameter of the model  $\mathcal{M}$ , and the outcome of the model,  $P_{AB}$ , is associated with the probability space  $B$ .

The model is optimised by minimising the loss function  $\mathcal{L}$  as

$$\begin{aligned} \min_\theta \mathcal{L} &= \min_\theta \mathbb{E}_{\eta, \xi} [\| P_{AB}(\eta, \xi) - P_B(\eta, \xi) \|_2] \\ &= \min_\theta \mathbb{E}_{\eta, \xi} [\| \mathcal{M}_\theta(P_A(\eta, \xi)) - P_B(\eta, \xi) \|_2], \end{aligned} \quad (13)$$

which is calculated by using the L2 norm, to ensure that the translated PDF  $P_{AB}$  is indistinguishable from the target PDF  $P_B$ . The trained model is expected to be applied to cases, where the distance between  $P_A$  and  $P_B$  in PDF space is not much larger than the training case. Given that the input  $P_A$  is based on the presumed joint PDF which is generally consistent with the target sub-grid distribution  $P_B$ , the scope of application is supposed to cover most combustion cases, thus the model's generalisability is enhanced.

For this study, the joint PDF is calculated using the  $\beta$ -PDF distributions in Eq. (10).

In summary, JPResUnet is trained using pairs of joint PDFs derived from the  $\beta$ -distribution and DNS as input and target, respectively, following the optimisation framework given in Eq. (13). A detailed discussion of data extraction and model structure follows in the subsequent sections.

### 3. Data preprocessing and numerical setup

#### 3.1. DNS cases

The JPResUnet is trained by using the DNS dataset of MILD combustion with varying mixture fractions and internal recirculation of exhaust gases (EGR). This combustion features a partially premixed combustion mode, indicated by the flame index observed across a wide field of the computational domain [57, 58]. In addition, a broad reaction zone is noted in MILD combustion [57, 58], which implies significant subgrid-scale fluctuations in the thermo-chemical and -physical properties of the reacting mixture. These fluctuations are advantageous for training purposes, as they contribute to the development of a robust and generalised model.

The study focuses on two MILD combustion cases with different levels of dilution, labelled ‘AZ1’ and ‘BZ1’. The initial thermo-chemical and -physical conditions for these cases are detailed in Table 1. For both cases, the initial root-mean-square (RMS) value of the velocity fluctuation  $u'_{\text{rms}}$  is around 16.66 m/s and the integral length scale of the turbulence  $\Lambda_0$  is around 1.42 mm. Based on these quantities, the turbulent Reynolds number  $\text{Re}_T$  and the Taylor micro-scale Reynolds number  $\text{Re}_\lambda$  are 96 and 34.73,

respectively. The ratio between the integral length scales of the progress variable and mixture fraction fields is  $l_c/l_Z \approx 0.77$ . The averaged progress variable is  $\langle c \rangle = 0.56$ , and the RMS value of the initial fluctuation in the progress variable field is  $\sigma_c/\langle c \rangle = 0.46$ . Compared with the case AZ1, the dilution is enhanced for the case BZ1, where the oxygen level is reduced to 2% by volume, inducing the difference in the mixture fraction field. The stoichiometric mixture fraction  $Z_{\text{st}}$  for the case AZ1 and BZ1 are 0.01 and 0.0058, respectively. With a similar equivalence ratio, the averaged mixture fractions are  $\langle Z \rangle = 0.008$  for the case AZ1 and  $\langle Z \rangle = 0.0046$  for BZ1. The RMS values of the initial fluctuation of the mixture fraction  $\sigma_Z/\langle Z \rangle$  for AZ1 and BZ1 are 1.05 and 1.23, respectively.

The combustion behaviour varies significantly with different dilution levels. In AZ1, both thin and thick heat release zones, with a large variation in typical thickness, are observed, where the thickened zone is induced by the interaction of reaction zones. In contrast, BZ1 exhibits a much thicker heat release zone nearly over the whole computational domain, indicating more frequent interactions of reaction zones [57]. To capture a wide range of combustion phenomena, training data is collected from the AZ1 case, while the BZ1 case is used for testing, as outlined in Table 2.

Table 1: Initial conditions of the MILD combustion.

Case	$\Lambda_0/l_Z$	$l_c/l_Z$	$X_{\text{O}_2}^{\max}$	$Z_{\text{st}}$	$\langle Z \rangle$	$\sigma_Z/\langle Z \rangle$	$\langle c \rangle$	$\sigma_c/\langle c \rangle$
AZ1	0.60	0.77	0.035	0.01	0.008	1.05	0.56	0.46
BZ1	0.60	0.77	0.020	0.0058	0.0046	1.23	0.56	0.46

A cubic computational domain of dimensions  $L_x \times L_y \times L_z = 10 \times 10 \times 10$  mm<sup>3</sup> was used, with 512 grid points in each spatial direction [57]. The resulting grid size is  $\delta x \approx 20 \mu\text{m}$ , providing approximately 30 grid points within the smallest chemical thickness of methane–air combustion. Combustion chemistry is modelled using a modified chemical mechanism, MS-58, which is based on the Smooke and Giovangigli scheme [59] and has been enhanced to include OH\* chemistry [60]. The MILD combustion was simulated using a DNS code, SENGA [61], with a timestep of  $\delta t = 1$  ns. After the first flow-through time,  $\tau_f = L_x/U_{\text{in}}$ , where  $U_{\text{in}} = 20$  m/s representing the inflow bulk mean velocity, the initial transient exited the computational domain, and the simulation continued for another half of  $\tau_f$  for data collection (approximately 60 snapshots). Further details regarding the numerical schemes, chemical mechanism, boundary conditions, and initial conditions can be found in [57].

### 3.2. PDF extraction

The PDFs are extracted from the sub-filter space within the DNS fields. For this study, the input PDFs are computed using a  $\beta$ -PDF in Eq. (11), for given Favre filtered quantities,  $\tilde{c}$ ,  $\tilde{\sigma}_{c,\text{sgs}}^2$ ,  $\tilde{Z}$  and  $\tilde{\sigma}_{Z,\text{sgs}}^2$ . The progress variable is defined using temperature,  $c = (T - T_u)/(T_b(Z) - T_u)$ , where  $T_u = 1500$  K is the initial temperature for the unburnt mixture. The burnt mixture temperature,  $T_b$ , is calculated by using the local mixture fraction [58]. Since combustion is adiabatic, this progress variable is equal to the one defined using species (with unity Lewis number) mass fractions, described previously and used for *a posteriori* assessment. Hereinafter, the temperature-based progress variable is represented by  $c_T$  to avoid ambiguity.

The progress variable-related quantities are calculated as

$$\tilde{c}_T(\mathbf{x}, t) = \frac{1}{\bar{\rho}(\mathbf{x}, t)} \int_{\mathbf{x}-\frac{\Delta}{2}}^{\mathbf{x}+\frac{\Delta}{2}} \rho(\mathbf{x}', t) c_T(\mathbf{x}', t) \mathcal{G}(\mathbf{x}') d\mathbf{x}', \quad (14)$$

$$\begin{aligned} \tilde{\sigma}_{c_T, \text{sgs}}^2(\mathbf{x}, t) &= \\ &\frac{1}{\bar{\rho}(\mathbf{x}, t)} \int_{\mathbf{x}-\frac{\Delta}{2}}^{\mathbf{x}+\frac{\Delta}{2}} \rho(\mathbf{x}', t) [c_T(\mathbf{x}', t) - \tilde{c}_T(\mathbf{x}, t)]^2 \mathcal{G}(\mathbf{x}') d\mathbf{x}', \end{aligned} \quad (15)$$

where the prime  $\mathbf{x}'$  represents local position inside the filter of size  $\Delta$ , and  $\mathcal{G}$  is the filter kernel. The Favre filtered mixture fraction and its variance are calculated using similar procedures on the DNS data. Two types of filter kernels are used: a spatial box filter and a Gaussian filter, and they are expressed as

$$\mathcal{G}_{\text{box}}(\mathbf{x}') = \begin{cases} \frac{1}{\Delta} & \text{if } |\mathbf{x}' - \mathbf{x}| \leq \frac{\Delta}{2} \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

$$\mathcal{G}_{\text{Gaussian}}(\mathbf{x}') = \left( \frac{6}{\pi \Delta^2} \right)^{\frac{1}{2}} \exp \left( -\frac{6\mathbf{x}'^2}{\Delta^2} \right). \quad (17)$$

As noted in Table 2, the normalised box filter kernel width of  $\Delta^+ = \Delta/\delta_{\text{th}}^{\text{st}} = 1$  is employed for training. The term  $\delta_{\text{th}}^{\text{st}}$  is the reference thermal thickness of stoichiometric laminar flame with the size of 1.6 mm ( $80\delta x$ ) and 3 mm ( $148\delta x$ ) for AZ1 and BZ1, respectively. Different filter sizes and Gaussian kernels are used during the testing phase.

The target joint distribution, denoted as  $\tilde{P}(\eta, \varphi; \mathbf{x}, t)$ , is calculated by using the kernel density estimation (KDE) as

$$\tilde{P}(\eta, \varphi; \mathbf{x}, t) = \frac{1}{nh} \sum_{i=1}^n K \left[ \frac{(\eta, \varphi) - \frac{1}{\bar{\rho}(\mathbf{x}, t)} (\rho c_T, \rho \ln \frac{Z}{Z_{\text{st}}})_i}{h} \right], \quad (18)$$

where  $K$  and  $h$  denote the kernel function and bandwidth, respectively. In this study, a bandwidth of 0.1 is used, and the Epanechnikov kernel [62] is chosen for its computational efficiency, as it requires fewer samples  $n$  only. This kernel is defined as

$$K(s) = \begin{cases} \frac{3}{4}(1 - s^2) & \text{if } |s| \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

The sample space variables  $\eta$  and  $\varphi$  in Eq. (18) are for the density-weighted progress variable and scaled mixture fraction, respectively. The scaled mixture fraction is expressed as  $\widehat{Z} = \ln(\widetilde{Z}/Z_{\text{st}})$  [28], which generates a relatively uniform distribution for flames with different flammability limits and stoichiometric values of the mixture fraction. The random variable space is discretised into a grid with dimensions  $N_\eta \times N_\varphi = 80 \times 100$ . The progress variable dimension is linearly discretised, while the scaled mixture fraction dimension is split into two segments,  $[-1.5, 0.3]$  and  $[0.3, 1.8]$ , with 65 and 33 points allocated respectively following earlier study [28]. The prediction is transferred back to the  $\eta$ - $\xi$  space as  $\widetilde{P}(\eta, \xi; \mathbf{x}, t) = \widetilde{P}(\eta, \varphi; \mathbf{x}, t)/\xi$  [28]. For notational simplicity and clarity,  $c_T$  and  $Z$  are used directly instead of the sample-space variables for the PDFs, e.g.,  $\widetilde{P}(c_T, Z)$ ,  $\widetilde{P}(c_T)$ ,  $\widetilde{P}(Z)$ , hereafter in this work.

It is noted that the joint distribution  $\widetilde{P}(c_T, Z)$  obtained above is the filter density function (FDF) [49]. The FDF is constructed from sub-filter space samples at a specific spatial location in a single DNS snapshot. Since DNS realisations are inherently unsteady, the FDF includes random variations [63], differing from the expected sub-grid PDF  $\widetilde{P}(c_T, Z)$ . To minimise this randomness, more samples from the sub-filter space would be required

across multiple realisations with identical resolved fields [64], which is computationally expensive. Alternatively, this randomness can be significantly reduced if the training dataset for machine learning incorporates FDF samples collected over many realisations during a statistically stationary state [27]. For this study, 25 DNS snapshots of the AZ1 MILD combustion case are selected. In each snapshot, the sub-filter space is systematically marched with a fixed spatial step to extract pairs of input-target PDFs. This process yields a training dataset of 33,275 samples, with 20% reserved for validation to prevent overfitting.

### 3.3. Machine learning algorithm

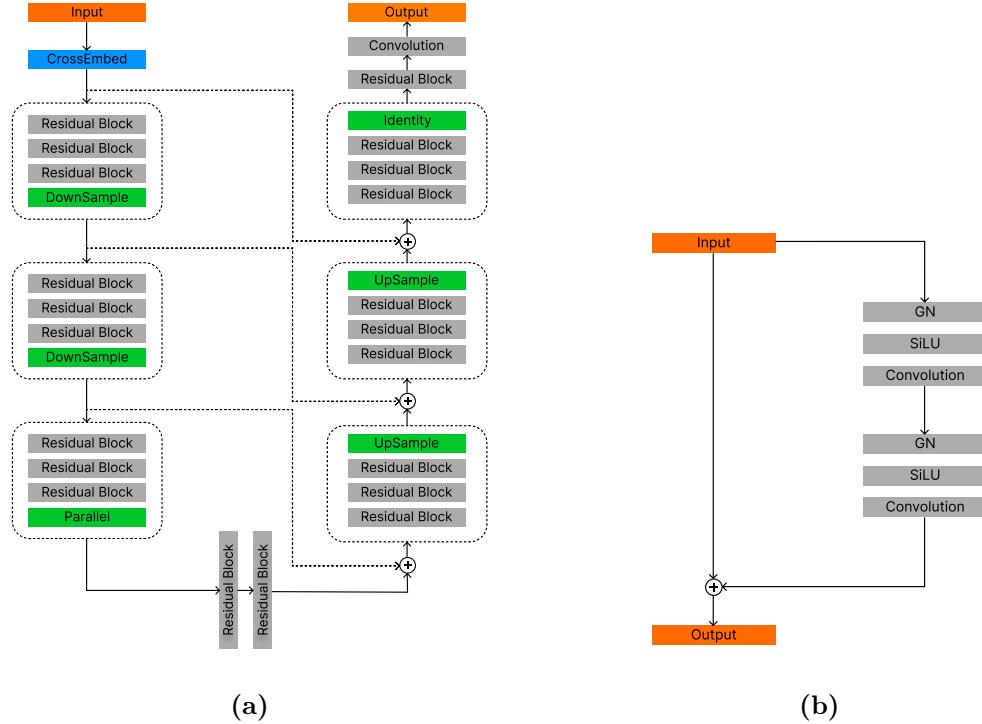


Fig. 1: Structure of (a) JPResUnet and (b) residual block.

The JPResUnet model features an encoder-decoder structure with skip connections, and each level within this structure utilises residual blocks, as shown in Fig. 1a. Its architecture can be summarised as follows:

- A cross-embedding layer and a residual block with an additional convolutional layer are deployed at the encoder-decoder’s inlet and outlet, respectively.
- The encoder and decoder are organised into three levels, each comprising three consecutive residual blocks followed by a size-changing unit.
- Levels are connected via skip connections that integrate the input at each encoder level into the corresponding decoder level.
- Two residual blocks are positioned at the bottleneck of the model.

The cross-embedding layer (highlighted in blue in Fig. 1a) adjusts the input to match the number of channels required by the encoder. This is achieved through three groups of convolutional kernels with sizes of  $3 \times 3$ ,  $7 \times 7$ , and  $15 \times 15$ . The output from each kernel is padded to maintain the same size, and these outputs are concatenated along the channel dimension to reach the desired number of channels  $N_{\text{ch}}$ . The kernel allocation is divided into  $N_{\text{ch}}/2$ ,  $N_{\text{ch}}/4$  and  $N_{\text{ch}}/4$  for the respective groups.

The residual blocks address the degradation issues often encountered in deep models by adding the input upstream to the residual function’s output [65]. This function includes a combination of group normalisation [66], a Sigmoid-Weighted Linear Unit (SiLU) [67], and a  $3 \times 3$  convolution, which is repeated once, as illustrated in Fig. 1b. The SiLU activation function, which outperforms the commonly used rectified linear unit (ReLU) in deep models

by preventing inactive neurons during training, is defined as

$$\text{SiLU}(x) = x * S(x), \quad (20)$$

where  $S(x) = 1/(1 + e^x)$  is the sigmoid function.

The size-changing unit, highlighted in green in Fig. 1a, adjusts the input size to the required value. For an input with the shape  $(N_{\text{ch}}^i, N_{c_T}, N_Z)$ , the down-sampling unit changes it to  $(N_{\text{ch}}^{i+1}, N_{c_T}/2, N_Z/2)$  by using a  $4 \times 4$  convolutional layer, where the superscript  $i$  represents the model's level. The up-sampling unit, combining interpolation with a  $3 \times 3$  convolution, restores the output to  $(N_{\text{ch}}^{i+1}, 2N_{c_T}, 2N_Z)$ . A parallel unit uses two groups of convolutions with kernel sizes of  $3 \times 3$  and  $1 \times 1$  respectively, both producing outputs of shape  $(N_{\text{ch}}^{i+1}, N_{c_T}, N_Z)$ , which are then summed.

To preserve spatial information and facilitate gradient flow, skip connections are employed between the corresponding levels of the encoder and decoder. They scale the input at the  $i^{\text{th}}$  level of the encoder by  $2^{-1/2}$  and concatenate it with the input at the corresponding decoder level along the channel dimension, resulting in a new input shape of  $(N_{\text{ch}}^{i,\text{ec}} + N_{\text{ch}}^{i,\text{dc}}, N_{c_T}, N_Z)$ , where ‘ec’ and ‘dc’ denote the encoder and decoder respectively.

The JPResUnet’s input PDF is a single-channel image sized at  $80 \times 100$ . The cross-embedding layer expands this into a shape of  $(32, 80, 100)$ . The encoder reduces the image resolution to  $(128, 20, 25)$ , and the decoder, which mirrors the encoder’s structure, restores the resolution to  $(32, 80, 100)$ . The final residual block and convolution ensure the output PDF has the same size as the input. The detailed input-output information for every component of the JPResUnet is listed in Appendix A. It is noted that many iterations of the model’s architecture have been tested, including the number of channels

of the image generated by the cross-embedding layer, the number of levels in the encoder and decoder, and the number of residual blocks at each level. The current layout of the model shows the best performance and is adopted by the present study.

The JPResUnet is implemented using PyTorch [68]. Optimisation in Eq. (13) is handled by the AdamW algorithm [69], with a weight decay of 0.01, an initial learning rate of  $10^{-4}$ , and a linear decay to  $10^{-7}$  in 200 epochs. The batch size is set to 32. These hyperparameters were fine-tuned using the grid-search method. Training was halted at around 121 epochs as no further drop in validation loss was observed, taking about 6 hours on an NVIDIA GeForce RTX 4090 GPU.

#### 4. Performance of JPResUnet on different testing cases

The generality and effectiveness of the JPResUnet model were evaluated through a comparative study across different cases with varying levels of extrapolation, as outlined in Table 2. To demonstrate the performance of JPResUnet, it was compared with an artificial neural network (ANN) consisting of three fully connected layers, with 256 and 512 neurons in the two hidden layers, respectively. This ANN was based on a previous work [27] that successfully predicted the sub-grid PDF for MILD combustion. The ANN was modified for extrapolation using a methodology described in [28]. The input features for the ANN included the Favre-filtered quantities  $\tilde{c}$ ,  $\tilde{g}_c$ ,  $\hat{Z}$  and  $\tilde{g}_Z$ , while the output was the joint PDF  $\tilde{P}(c_T, \hat{Z})$ . Both the JPResUnet and ANN models were trained on the same dataset as described in Section 3.2 of the study.

Table 2: List of training and testing information. The filter kernels "B" and "G" denote the box and Gaussian kernels, respectively

Data case	Filter kernels ( $\Delta^+$ )	Size of dataset
<i>Training</i>		
AZ1	B(1)	33275
<i>In-sample validation</i>		
AZ1	B(1)	3993
<i>Out-of-sample prediction: same combustion case</i>		
AZ1	B(1) B(1.5) B(2)	3993 3000 2187
AZ1	G(1) G(1.5) G(2)	3993 3000 2187
<i>Out-of-sample prediction: different combustion case</i>		
BZ1	B(1) B(1.5)	3000 1536
BZ1	G(1) G(1.5)	3000 1536

#### 4.1. Validation: *in-sample prediction*

The extrapolation capability of JPResUnet was tested on data that is slightly different from the training dataset. Within the temporal domain spanned by snapshots of the AZ1 flame used for training, three additional realisations of the flame were selected for in-sample predictions.

The JPResUnet model's ability to predict the joint PDF for a sub-grid space within the reaction region of AZ1 was compared with that of the ANN and the analytical model  $\beta$ -PDF. The results, presented in Fig. 2 of the study, include the marginal PDFs  $\tilde{P}(c_T) = \int \tilde{P}(c_T, Z) dZ$  and  $\tilde{P}(Z) = \int \tilde{P}(c_T, Z) dc_T$ . JPResUnet captures the bi-modal distribution of the progress

variable and a negative correlation between  $c_T$  and  $Z$ , characteristic of MILD combustion under the AZ1 conditions [27], with contours closely matching DNS data. The ANN's prediction shows a similar agreement with the DNS, which is expected as many studies have verified such excellent in-sample prediction [27, 28]. In contrast, the analytic model,  $\beta$ -PDF, fails to capture the DNS data's shape and peak.

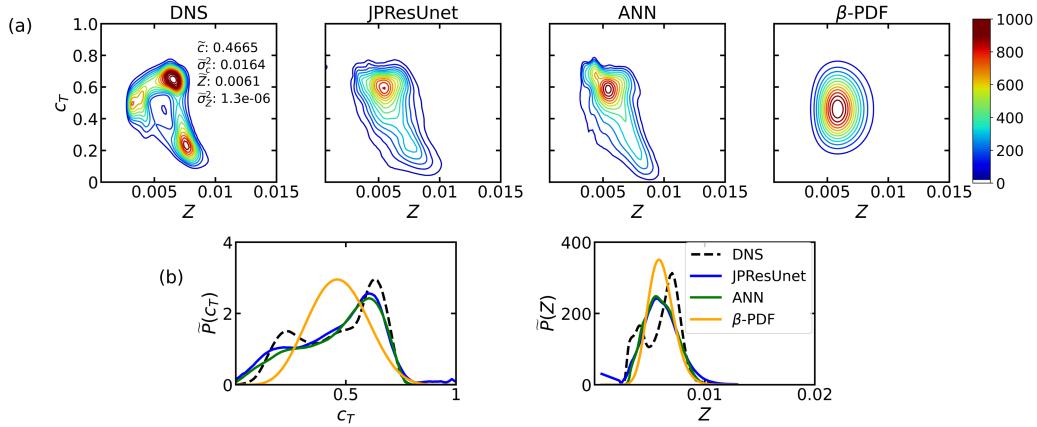


Fig. 2: Comparative analysis of the joint and marginal PDFs between DNS and model in-sample predictions.

The overall accuracy of the JPResUnet model across the entire testing dataset (as listed in Table 2) was assessed by comparing the predicted PDFs to those obtained from DNS using the Jensen-Shannon divergence (JSD) [70]. The JSD measures the similarity between two distributions  $P_1$  and  $P_2$  and is calculated as

$$\text{JSD}(P_1||P_2) = \frac{1}{2} \sum_{n=1}^N \left( P_1(n) \ln \frac{P_1(n)}{P_2(n)} + P_2(n) \ln \frac{P_2(n)}{P_1(n)} \right), \quad (21)$$

where  $N$  denotes the total discretised points in the random variable space. The JSD value is bounded between 0 and  $\ln(2)$ , with smaller values indicating

higher similarity between the two distributions. The PDFs from the model and DNS were taken as  $P_1$  and  $P_2$ , respectively, and the JSD values for the marginal PDFs of the progress variable and the mixture fraction were calculated and plotted in Fig. 3.

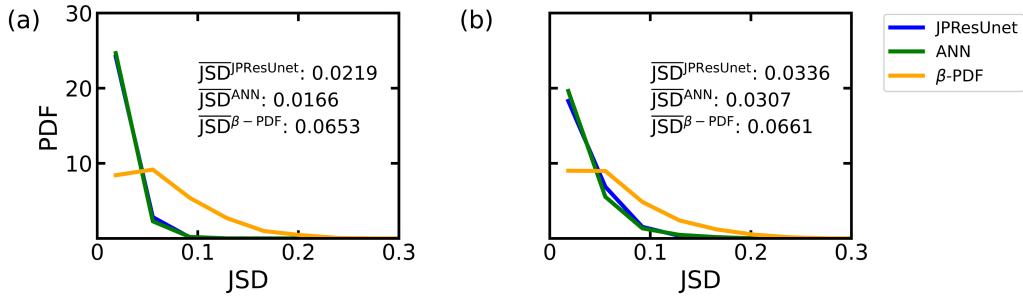


Fig. 3: PDF of JSD for marginal PDFs of (a) progress variable  $c_T$  and (b) mixture fraction  $Z$ , predicted on in-sample dataset.

Both the JPResUnet and ANN models exhibited high accuracy compared to the  $\beta$ -PDF model, as their JSD plots for  $\tilde{P}(c_T)$  and  $\tilde{P}(Z)$  are clustered closer to zero. The mean JSD values for these models were approximately one-third and one-half of the corresponding values for the  $\beta$ -PDF model, respectively. This finding reaffirms the superiority of the ANN model for in-sample predictions. The JPResUnet's similar performance to the ANN confirms its accuracy, though further improvements in in-sample prediction are beyond the scope of this study.

#### 4.2. Testing with different filter widths and kernels

Due to constraints in computational and experimental resources, the availability of high-fidelity data for turbulent flames is currently limited,

which makes it challenging to comprehensively cover practical scenarios, especially complex geometries. This limitation necessitates development of models that can be applied to scenarios significantly different from the training dataset. To validate this capability, the JPResUnet model was evaluated using data sampled from realisations of AZ1 at temporal steps beyond the range used for training. Additionally, since most LES utilise an implicit filter, where the filter type and size are unknown, the model was tested on data derived using various filter kernels and sizes to ensure its practical applicability.

The performance of the models was evaluated using data extracted with box and Gaussian filter kernels of widths  $\Delta^+ = 1, 1.5$ , and  $2$ . Larger filter widths were not considered due to the size of DNS domain, as  $\Delta^+ = 3$  approaches half the side length of the domain, resulting in an insufficient number of samples extracted from selected snapshots. Moreover, large filters fail to encompass a sufficient range of the turbulent kinetic energy, rendering the validated models impractical for *a posteriori* simulations and testing, as discussed in [27]. The predicted joint and marginal PDFs for a local sub-grid space with the box and Gaussian kernels at the three filter sizes are presented in Fig. 4 and 5, respectively.

The predicted, both joint and marginal, PDFs are observed to be insensitive to the choice of the filter kernel, specifically for  $\Delta^+ = 1$ . At this width, there is minimal variation in filtered quantities, as shown in Fig. 4a and 5a. Consequently, the models, including the ANN and  $\beta$ -PDF approach, generate consistent predictions despite the subtle input variations. The JPResUnet's predictions also remain constant since the inputs from the  $\beta$ -distribution un-

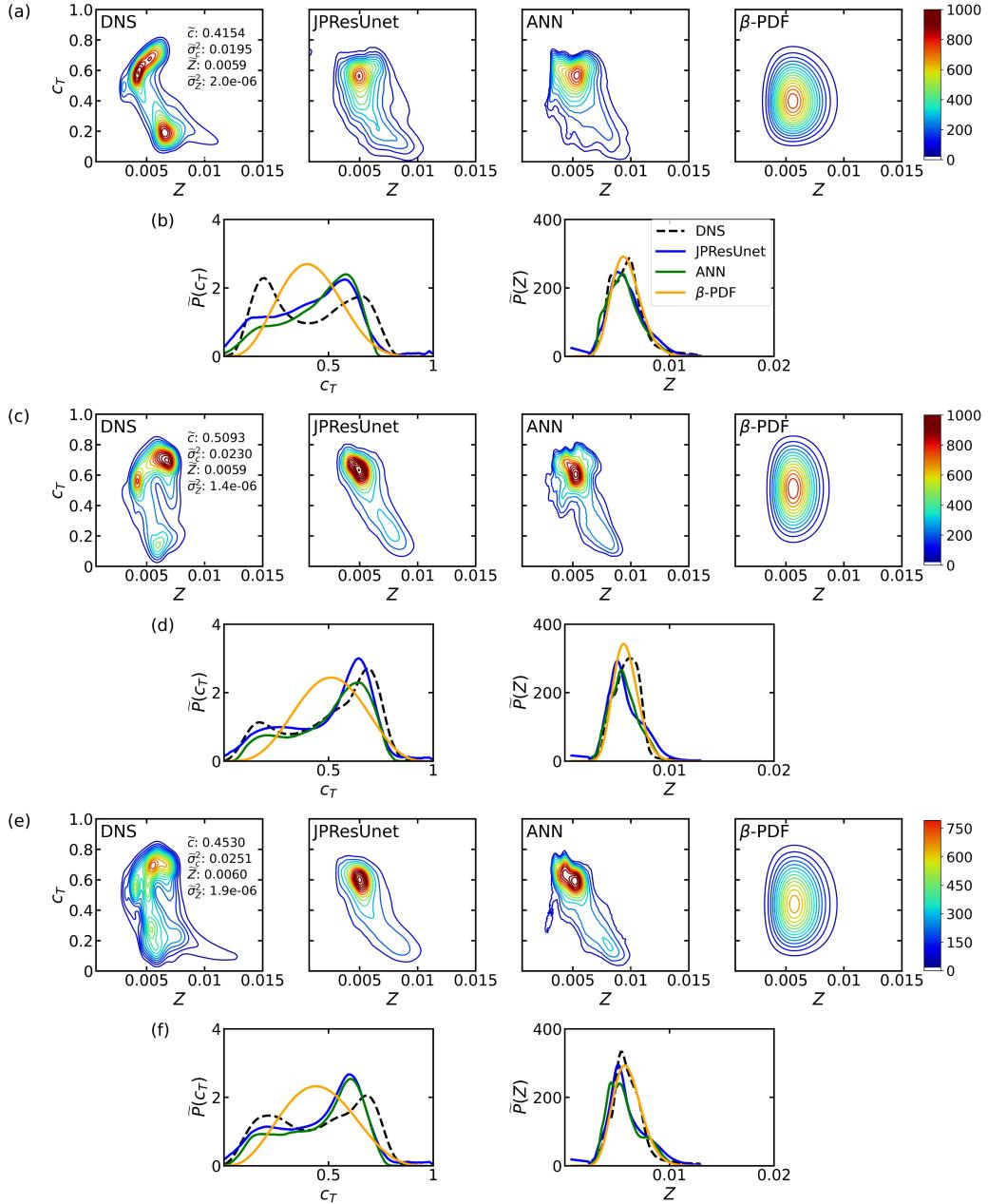


Fig. 4: Comparative analysis of the joint and marginal PDFs between DNS and model out-of-sample predictions for the case AZ1, utilising a box filter with widths of (a)-(b)  $\Delta^+ = 1$ , (c)-(d)  $\Delta^+ = 1.5$  and (e)-(f)  $\Delta^+ = 2$ .

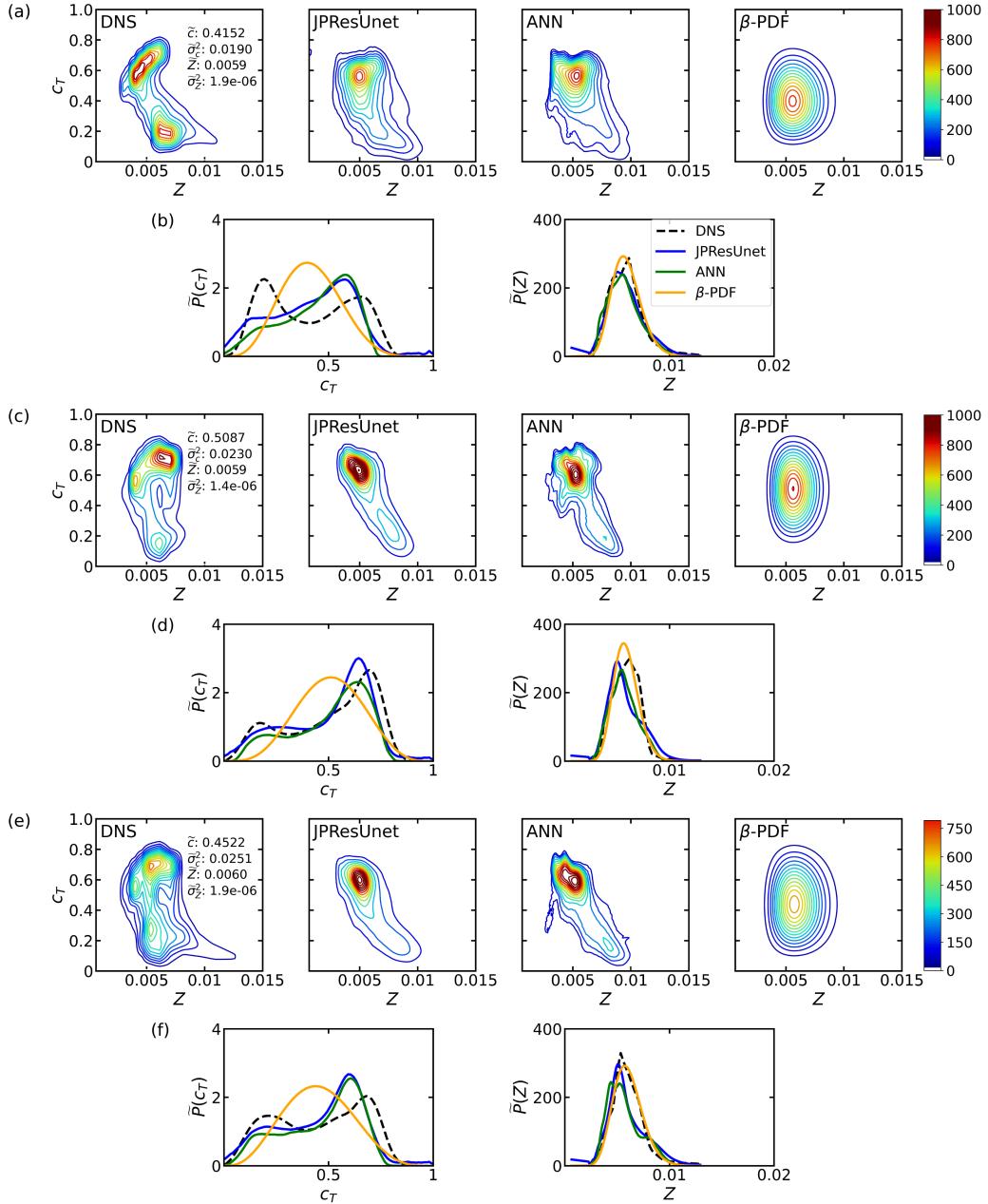


Fig. 5: Comparative analysis of the joint and marginal PDFs between DNS and model out-of-sample predictions for the case AZ1, utilising a Gaussian filter with widths of (a)-(b)  $\Delta^+ = 1$ , (c)-(d)  $\Delta^+ = 1.5$  and (e)-(f)  $\Delta^+ = 2$ .

der both box and Gaussian filters are the same. Therefore, the discussion on local sub-grid PDF predictions focuses on the results obtained using the box filter only (Fig. 4).

The JPResUnet demonstrates excellent performance, accurately reproducing the shape and peak of joint PDF contours across all filter widths, with close agreement with DNS results. This accuracy is particularly evident in  $\tilde{P}(c_T)$  at larger filter widths (as shown in Fig. 4d and 4f). In contrast, the  $\beta$ -PDF approach fails to provide satisfactory predictions except for the marginal distribution of the mixture fraction. This analytical model generates statistical distributions, as the Gaussian-like shape observed in the  $\tilde{P}(c_T)$  plots, while missing the instantaneous features. This deficiency compromises the accuracy of the PDF-based approach in modelling instantaneous reaction rates, as will be further demonstrated later in this section.

The ANN model performs comparably to JPResUnet but shows noticeable under-predictions in marginal PDFs as filter width increases (as seen in Fig. 4f). Additionally, the joint PDF contours predicted by ANN exhibit wrinkled shapes, attributed to the unsteadiness of the instantaneous sub-grid PDFs (FDFs) used during training. JPResUnet avoids these issues by leveraging the statistical structure of  $\beta$ -PDF while preserving sub-grid complexity, resulting in smooth and consistent predictions across filter sizes. This robustness underscores JPResUnet’s high generalisation capability.

Figure 6 presents the PDF of the JSD values for model predictions at varying filter widths and kernels (solid and dashed lines correspond to the box and Gaussian filters, respectively). The mean JSD values for each model’s predictions with the box (and also Gaussian) filter are listed. JPResUnet

demonstrates consistent high accuracy across all filter widths and types, with over 80% of its JSD values concentrated below 0.05 and decreasing mean JSD value as filter width increases. This reduction indicates the benefit of translating the statistical  $\beta$ -distribution, which shows a decline as well. While ANN performs well at  $\Delta^+ = 1$ , its accuracy decreases with filter width, leading to the highest mean JSD values at  $\Delta^+ = 2$ , reflecting the negative influence of randomness in the training data.

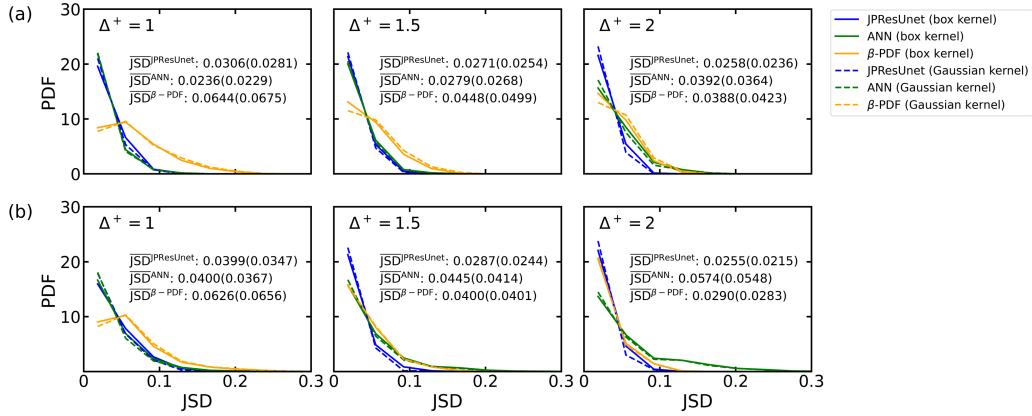


Fig. 6: PDF of JSD for marginal PDF of (a) progress variable  $c_T$  and (b) mixture fraction  $Z$  for  $\Delta^+ = 1, 1.5$  and  $2$ , predicted on out-of-sample dataset of the case AZ1. Solid and dashed lines correspond to the box and Gaussian filters, respectively. The mean JSD values for models' prediction based on the box filter (Gaussian filter) are also listed.

Using the predicted joint PDF, the filtered reaction rate source term  $\bar{\omega}_{c_T}$  is calculated similarly to Eq. (8), where the flamelet reaction rate and density in the integrand are replaced with the doubly conditional averaging counterparts  $\langle \dot{\omega}_{c_T} / \rho | c_T, Z \rangle$  over the DNS data (60 snapshots in total), as verified in [27]. The reaction rate of  $c_T$  is calculated as  $\dot{\omega}_{c_T} = \dot{q}/[c_p(T_b - T_u)]$ , where  $\dot{q}$  and  $c_p$  are the volumetric heat release rate and mixture specific heat

capacity respectively. The comparison between the filtered reaction rates calculated by using the joint PDF from models  $\bar{\omega}_{c_T}^m$  and from DNS  $\bar{\omega}_{c_T}^{m\text{-DNS}}$  with different filter widths is illustrated in the scatter plot shown in Fig. 7. Due to the similarity of the joint PDFs predicted using the box and Gaussian filters, the modelled filtered reaction rates are very similar, so only the results for the box filter are shown here (the results for the Gaussian filter are in the supplementary material).

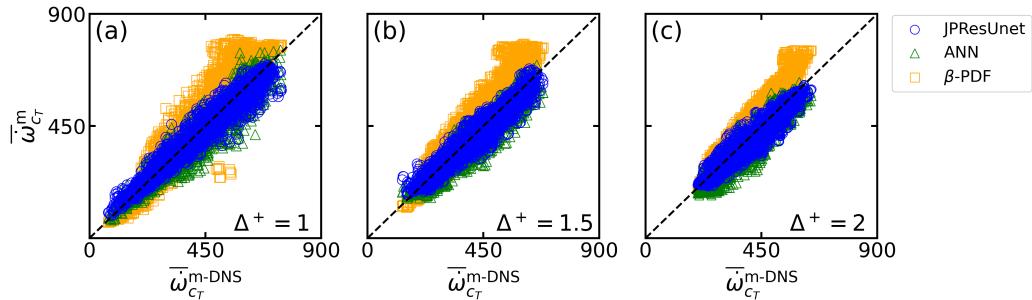


Fig. 7: Scatter plot of the progress variable source term modelled by joint PDFs from different models and from the DNS for the case AZ1, using the box filter with different filter widths.

Among all the models, JPResUnet demonstrates the best performance across all filter sizes, accurately modelling the filtered reaction rate closely clustered around the diagonal without noticeable over- or under-prediction. Similarly, ANN provides comparable predictions with substantial overlap with JPResUnet. However, as filter size increases, ANN's predictions become more scattered, reflecting a decline in predictive performance discussed previously. The analytical  $\beta$ -PDF approach produces reasonable estimates but consistently over-predicts the reaction rate  $\bar{\omega}_{c_T}^m$  across all filter widths, aligning with prior observations [27].

The accuracy of filtered reaction rate modelling is assessed using the root-mean-square of a normalised error (RMSE), defined as

$$\text{RMSE}^m = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{\bar{\omega}_{c_T,i}^m - \bar{\omega}_{c_T,i}^{\text{m-DNS}}}{\bar{\omega}_{c_T,i}^{\text{m-DNS}}} \right)^2}, \quad (22)$$

where the index  $i$  denotes each sample in the test dataset with the size  $N$ . The RMSE values for each model with the box (and Gaussian) filter at all filter widths are listed in Table 3. The RMSE values for JPResUnet are the lowest for all filter widths and remain below 0.1 for both box and Gaussian kernels. By contrast, the RMSE value for the ANN increases significantly with the filter width, reaching around 1.5 times the value of JPResUnet's predictions. Although the  $\beta$ -PDF approach shows improved performance at higher filter widths, its RMSE remains the highest, due to the over-prediction observed in Fig. 7c.

Table 3: Root-mean-square error (RMSE) for filtered reaction rate by using different models for AZ1 with the box (Gaussian) filter at different filter widths.

Model	$\Delta^+ = 1$	$\Delta^+ = 1.5$	$\Delta^+ = 2$
JPResUnet	0.0868 (0.0849)	0.0904 (0.0864)	0.0982 (0.0950)
ANN	0.101 (0.0999)	0.112 (0.107)	0.135 (0.130)
$\beta$ -PDF	0.165 (0.169)	0.152 (0.157)	0.147 (0.150)

### 4.3. Extrapolating to a higher dilution level

The extrapolation capability of the JPResUnet model has been rigorously assessed and shown to be robust across different filter widths, demonstrating improved accuracy compared to both the state-of-the-art ML model and the

widely used analytical model. To further evaluate this capability, JPResUnet was tested on a distinct MILD combustion case, BZ1, which exhibits different thermo-chemical and -physical characteristics from the training case AZ1 (detailed in Section 3.1). The model evaluations were conducted using data extracted with two filter widths,  $\Delta^+ = 1$  and  $1.5$ .

The local sub-grid distributions within the reaction zone obtained using the box kernel are depicted in Fig. 8. The results for the Gaussian kernel are similar, which are included in Appendix B. Unlike the AZ1 case, the DNS joint PDFs for BZ1 display a less pronounced bimodal distribution and a more dispersed shape, indicating a more distributed reaction zone. JPResUnet accurately captures these characteristics across all filter widths, with its predictions aligning closely with DNS. In contrast, the ANN model predicts the joint PDF contour with a discontinuity at  $\Delta^+ = 1$ , as evidenced by fluctuations in the marginal PDF for the mixture fraction (Fig. 8b), and significantly underpredicts the marginal PDF for the progress variable. At the larger filter width, ANN fails to produce a reasonable PDF, with the output resembling a highly concentrated delta-distribution, as illustrated in Figs. 8c and 8d. These results underscore the excellent generality of JPResUnet. By leveraging the PDF-translation mechanism introduced in Section 2.4, it is well-suited for cases with a reasonable statistical distribution of the sub-grid space. Conversely, ANN’s performance is limited by its training dataset, resulting in poor predictions when the test case deviates significantly from the training data.

The overall accuracy of the models over the test dataset is compared using JSD of the PDF predictions in Fig. 9. Similar to the previous observation

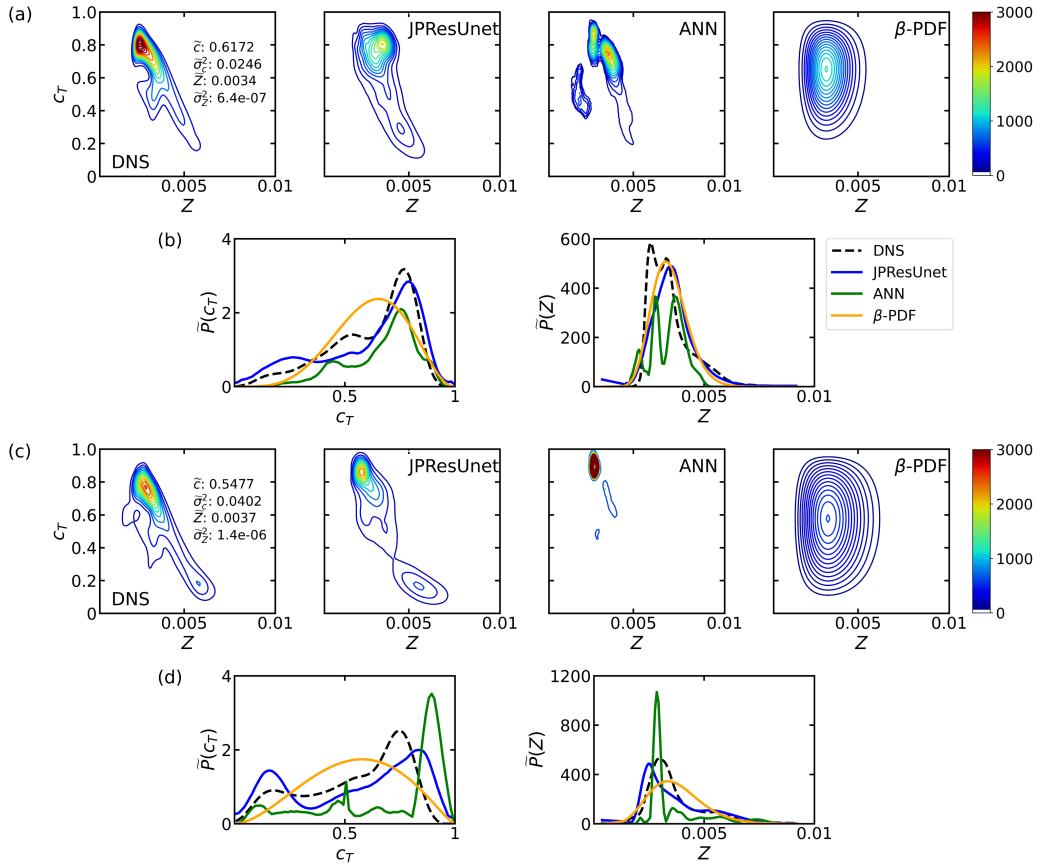


Fig. 8: Comparative analysis of the joint and marginal PDFs between DNS and model predictions for the case BZ1, utilising a box filter with widths of (a)-(b)  $\Delta^+ = 1$  and (c)-(d)  $\Delta^+ = 1.5$ .

(Fig. 6), JPResUnet achieves the highest predictive accuracy, with over 80% JSD values below 0.05, leading to the lowest mean value (less than 0.03). In contrast, the ANN model exhibits a significant number of JSD values exceeding 0.1, particularly at  $\Delta^+ = 1.5$ , where its mean JSD values are double those of JPResUnet, reflecting a low-fidelity prediction.

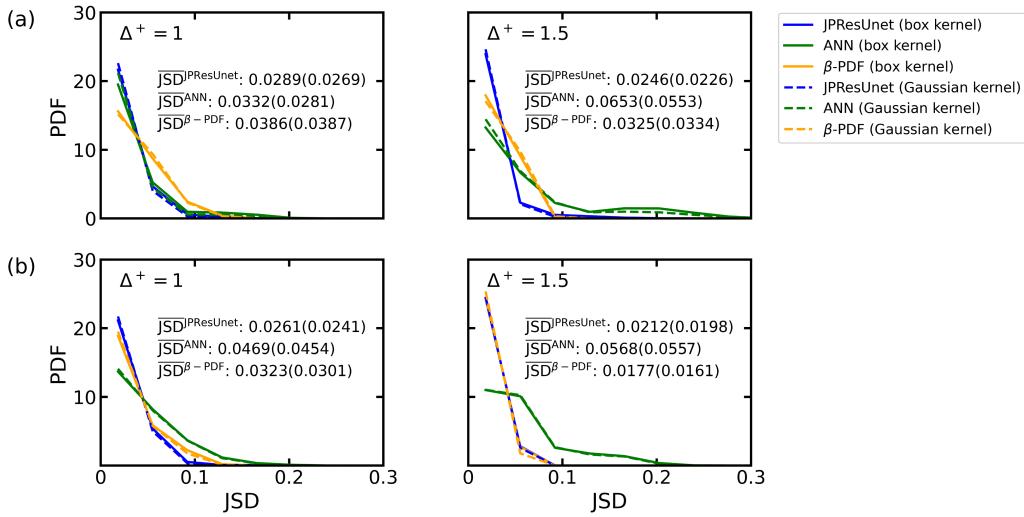


Fig. 9: PDF of JSD for marginal PDF of (a) progress variable  $c_T$  and (b) mixture fraction  $Z$  for  $\Delta^+ = 1$  and 1.5, predicted on the case BZ1. Solid and dashed lines correspond to the box and Gaussian filters, respectively. The mean JSD values for models' prediction based on the box filter (Gaussian filter) are also listed.

The filtered reaction rate modelled by different approaches with the box filter is compared to DNS results in Fig. 10, with results for the Gaussian filter provided in Appendix B. Among all models, JPResUnet delivers the most accurate predictions, which are tightly clustered around the diagonal, despite slight underprediction at  $\Delta^+ = 1.5$ . In contrast, the ANN model exhibits significant under-predictions due to inaccurate PDF predictions, while

the  $\beta$ -PDF approach shows over-prediction for the region of intense reaction. The RMSE values presented in Table 4 confirm the very good performance of JPResUnet, remaining the lowest RMSE (around 0.1). By comparison, the ANN model’s inferior performance is quantitatively evidenced by high RMSE values, exceeding 0.2 at  $\Delta^+ = 1.5$ .

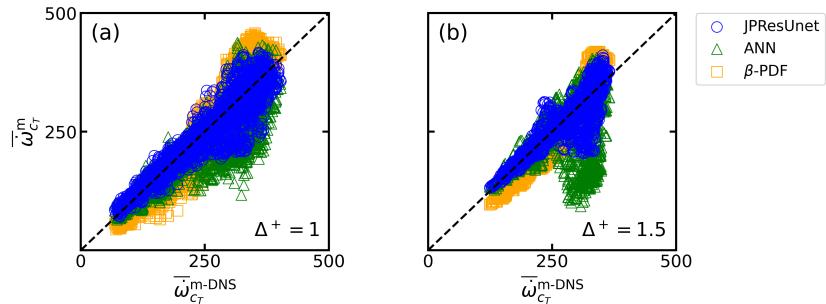


Fig. 10: Scatter plot of the progress variable source term modelled by joint PDFs from different models and from the DNS for the case BZ1, using the box filter with different filter widths.

Table 4: Root-mean-square error (RMSE) for filtered reaction rate by using different models for BZ1 with the box (Gaussian) filter at different filter widths.

Model	$\Delta^+ = 1$	$\Delta^+ = 1.5$
JPResUnet	0.103 (0.102)	0.0997 (0.0989)
ANN	0.156 (0.151)	0.244 (0.238)
$\beta$ -PDF	0.175 (0.179)	0.131 (0.134)

## 5. *a posteriori* assessment

The preceding section illustrates the performance of the JPResUnet model in predicting sub-grid PDFs for various MILD combustion scenarios

within an *a priori* assessment framework. This performance is benchmarked against the ANN and  $\beta$ -distribution approach on out-of-sample filter widths and kernels. Notably, JPResUnet maintains consistently high predictive accuracy across these conditions. Consequently, the model is evaluated in an *a posteriori* LES to substantiate its robustness and potential in practice.

### 5.1. LES setup

The experimental setup of the multi-regime burner (MRB) is shown schematically in Fig. 11. This involves three streams with different flow rates and equivalence ratios, resulting in inhomogeneous mixing of reactants downstream of the central jet. The bluff body positioned between slots 1 and 2 generates a recirculation zone composed of burnt products, stabilising the inner and outer flames. A high-velocity, fuel-rich premixed methane–air mixture is issued through the central jet, while pure air is supplied through slot 1. A range of equivalence ratios of the central premixed jet and bulk-mean air velocity for slot 1 is considered in the experimental study of [71]. One of those cases, MRB26b, is considered for the *a posteriori* testing of the ML models presented in the previous section. This specific case is considered because it showed clear multi-regime combustion including local mixtures having mixture fraction values beyond the flammability limits [40]. The equivalence ratio of the central methane–air jet is 2.6, with a velocity of 105 m/s. The bulk-mean air velocity is 15 m/s for slot 1. The lean premixed methane–air mixture with an equivalence ratio of 0.8 flows through slot 2 at a velocity of 20 m/s. The flame is shielded from the external disturbances by using a low-velocity air co-flow around the burner. The temperatures of the mixtures introduced through the jet, slot 1, slot 2, and the co-flow are 309, 333,

307, and 298 K respectively. Further details on the burner configuration and measurement techniques are discussed by Butz et al. [71, 72].

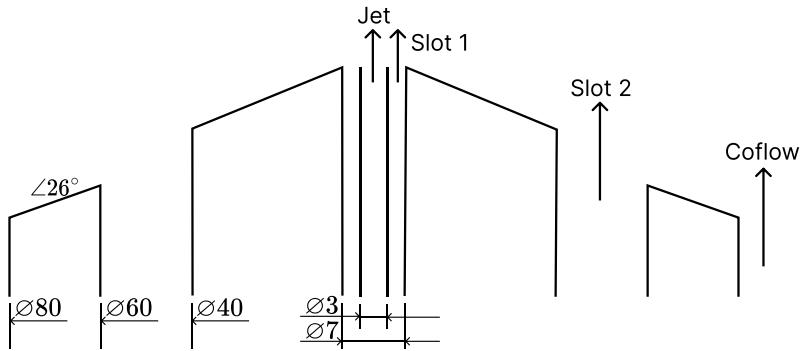


Fig. 11: Schematic of the multi-regime burner (MRB) with diameters in mm.

The numerical setup follows the previous study [40]. The computational domain is cylindrical, with a radius of 600 mm and a length of 600 mm from the exit of the central jet nozzle, which corresponds to approximately five times the flame length. This domain is discretised using about 3.5M hexahedral numerical cells. The finest resolution is applied within the central jet region, which has the highest velocity. The mesh size ranges from 0.1 to 0.3 mm in the inner flame region. The outer flame region has cell sizes ranging from 0.4 to 1 mm. The velocity boundary conditions at the inlets utilise an inflow turbulence generator based on the synthetic eddy method [73]. A three-dimensional steady RANS with the Reynolds stress equation model is conducted to obtain the mean profiles of velocity, the Reynolds stress tensor, and the streamwise integral length scale required for the LEMOS inflow generator [74]. Wave transmissive boundary conditions are imposed on the outflow boundaries to prevent acoustic wave reflection. All burner geometry

walls are set to be no-slip and adiabatic.

In the *a priori* assessment, JPResUnet has demonstrated its effectiveness in capturing the instantaneous features of sub-grid distributions, indicating the potential for real-time inference during the simulation. However, the primary challenge associated with on-the-fly deployment is the computational cost. The numerous nonlinear operations inherent in ML models significantly increase the computational burden compared to widely used tabulation methods, which rely on linear interpolation within a relatively low-dimensional look-up table (LUT). The computational costs increase with model complexity. Consequently, a tabulation approach was employed as a practical compromise with tables generated using ML model, JPResUnet, tested in the previous section. The filtered reaction rate source terms  $\bar{\omega}_p$  and  $\bar{c\omega^*}$  modelled by using JPResUnet were stored in a new LUT, which were retrieved during the simulation. Future research will explore more promising alternatives, such as integrating JPResUnet in LES using GPU acceleration for on-the-fly inference.

The simulations are performed using OpenFOAM v7 with a modified PIMPLE algorithm (rhoPimpleFoam solver). Second-order central difference schemes are used for velocity, and an implicit Euler scheme is employed for time marching, with a small variable time step on the order of  $\mathcal{O}(10^{-7}\text{s})$  to ensure the CFL number remains below 0.4 across the entire domain. Time-averaged statistics are obtained over a period of 25 flow-through times, which is necessary to achieve convergence due to the presence of low-velocity, large-scale structures within the recirculation zone. The flow-through time is based on the size of the recirculation zone, the distance upstream to slot 2, and the

velocity of slot 2; one flow-through time is approximately 2 ms, and five flow-through times are required for the flames to stabilise after ignition. The simulation is run on ARCHER2, a UK high-performance computing facility, using 1920 cores for 62 h of wall clock time.

### 5.2. Results

The results of *a posteriori* assessment of the JPResUnet are compared with the experiment as well as the prior LES study [40] employing a  $\beta$ -PDF-based LUT (referred to as Look-up $_{\beta}$ ). It is worth noting that the case Look-up $_{\beta}$  uses the PDF with the resolution of  $400 \times 500$  in the  $c \times Z$  space for the reaction rate modelling in Eq. (8), significantly finer than the resolution of  $80 \times 100$  for JPResUnet. To investigate the effect of the resolution, a higher-resolution version with  $384 \times 384$  is also used ( $400 \times 500$  is not used because of excessive demand for computer memory). Hereinafter, results from this higher-resolution JPResUnet are denoted with the subscript ‘hr’, while those from the lower-resolution JPResUnet tested in Section 4 are denoted with the subscript ‘lr’.

Figure 12 presents the radial profiles of time-averaged and root-mean-squared (RMS) axial velocity, mixture fraction, and temperature at various streamwise locations, with all results averaged in the azimuthal direction. The velocity field is not affected by PDF models, with mean and RMS axial velocities (Fig. 12a) showing indistinguishable profiles across all approaches and aligning well with experimental data. The mixture fraction profiles (Fig. 12b) exhibit similar trends, with all approaches producing comparable results that closely match experimental data in the inner reaction region and over- and under-predict the mean and RMS values, respectively, for the outer shear

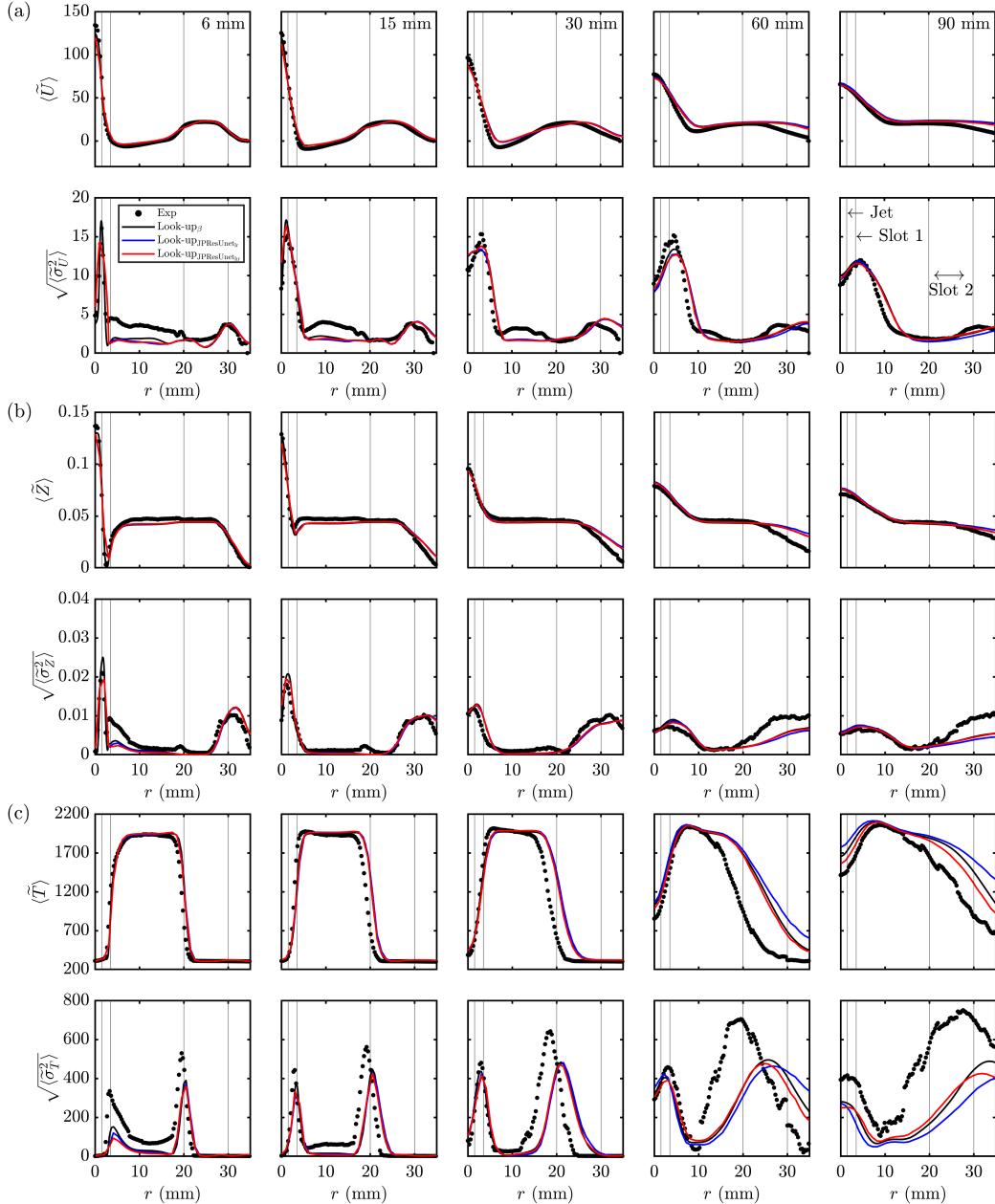


Fig. 12: Radial profiles at different streamwise locations of the (a) axial velocity, (b) mixture fraction, and (c) temperature for the simulations (lines) and measurements (symbols).

layer of slot 2. Notably, the sensitivity to PDF resolution of JPResUnet is detected, with the lower-resolution JPResUnet (represented by blue lines) showing slightly more deviations than other simulations.

Regarding the temperature field (Fig. 12c), all simulations exhibit similar radial profiles that align well with experimental data in the near-field region (below 30 mm). Further downstream (60 to 90 mm), simulations show a good agreement with the measurement around the jet region with a slight over-prediction of the mean value. For the outer flame ( $r \geq 15$  mm), the mean temperature is over-predicted, indicating insufficient air entrainment [40], and the peak of the RMS value is shifted towards a higher radial position. The high-resolution JPResUnet (represented by red lines) improves agreement with experimental data, particularly at 90 mm, where it reduces both the mean temperature over-prediction and RMS peak shift. Conversely, the low-resolution model amplifies these deviations, underscoring the critical role of PDF resolution in achieving accurate predictions.

Overall, with proper resolution in the PDF space, JPResUnet demonstrates very good accuracy in *a posteriori* assessments, especially in regions where the  $\beta$ -PDF model struggles. It is stressed here again that the current deployment of the JPResUnet is a compromise due to the limited computational resources. As a model trained by an instantaneous dataset, on-the-fly inference should be conducted in future to further validate and strengthen the above findings.

## 6. Conclusion

This study developed a novel PDF-translation model, JPResUnet, inspired by image-to-image translation techniques and utilising a residual U-net architecture. The model predicts the sub-grid joint probability density function (PDF) of the progress variable and mixture fraction in partially premixed flames, leveraging the analytical  $\beta$ -PDF as input. This approach ensures consistent translation within the PDF space during training and testing, enhancing the model's generalisability across diverse applications. Training was conducted using direct numerical simulation (DNS) data from the methane–air Moderate or Intense Low-oxygen Dilution (MILD) combustion, with a box filter at unit-normalised width.

The performance of JPResUnet was first validated through an *a priori* assessment against a well-studied Artificial Neural Network (ANN) and the  $\beta$ -PDF approach. On in-sample data, JPResUnet achieved predictive accuracy comparable to the ANN while outperforming the  $\beta$ -distribution, as demonstrated by the contours of local sub-grid distributions and Jensen-Shannon Divergence (JSD) values.

For out-of-sample data with varying filter kernels and widths, JPResUnet provided consistently accurate predictions of local sub-grid PDFs and produced smoother contours compared to ANN, reducing sensitivity to unsteadiness in the training data. It also demonstrated good performance in modelling the filtered reaction rate source term, achieving the lowest root-mean-square error (RMSE) across all filter widths, while the ANN performance declined with increasing filter widths.

JPResUnet's generalisability was further assessed in the case of MILD

combustion with a higher dilution level than the training case at two filter widths. JPResUnet’s predictions closely matched the DNS results, while the ANN’s predictions deteriorated significantly to entire failure at  $\Delta^+ = 1.5$ . The JSD values for JPResUnet remained the lowest, while those for ANN increased to approximately twice that of JPResUnet at the large filter width. JPResUnet also aligned closely with DNS in modelling the filtered reaction rate source term, with the lowest RMSEs, in contrast to ANN’s substantial under-prediction, particularly at  $\Delta^+ = 1.5$ .

An *a posteriori* assessment of JPResUnet was conducted through a large eddy simulation (LES) of the multi-regime burner (MRB). To address the computational cost of on-the-fly inference, JPResUnet was implemented via a look-up table (LUT), and compared against experimental data and a conventional  $\beta$ -PDF-based LUT. An additional JPResUnet with a higher PDF resolution ( $384 \times 384$  in  $c \times Z$ ) was tested to assess the influence of resolution. While the velocity and mixture fraction fields showed negligible differences, the high-resolution JPResUnet reduced deviations in the temperature field, particularly in the outer reaction region at downstream locations. In contrast, increased deviations were observed for low-resolution JPResUnet, implying the performance is sensitive to the PDF resolution.

In conclusion, JPResUnet demonstrates robust performance across various combustion scenarios, surpassing traditional methods in accurately capturing complex features and exhibiting better generality compared to the ANN. Its ability to reduce deviations in both *a priori* and *a posteriori* assessments underscores the potential for LES applications. Future research will focus on optimising computational efficiency to enhance the model’s ap-

plicability in practice.

### Acknowledgements

J.C.M. and N.S. acknowledge the financial support from Mitsubishi Heavy Industries, Ltd., Takasago, Japan. This work used the ARCHER2 UK National Supercomputing Service ( <https://www.archer2.ac.uk> ). The authors are grateful to ARCHER2 for the financial and computational allocation as a part of an ARCHER2 Pioneer Project (e808).

### References

- [1] S.B. Pope, Turbulent flows, Cambridge University Press, Cambridge, UK, 2000.
- [2] T. Poinsot, D. Veynante, Theoretical and numerical combustion, 3rd ed., R.T. Edwards, Inc., Philadelphia, USA, 2005.
- [3] A.W. Vreman, J.A. van Oijen, L.P.H. de Goey, R.J.M. Bastiaans, Subgrid scale modeling in large-eddy simulation of turbulent combustion using premixed flamelet chemistry, *Flow Turbul. Combust.* 82 (2009) 511–535.
- [4] F.E. Hernández-Pérez, F.T.C. Yuen, C.P.T. Groth, Ö.L. Gülder, LES of a laboratory-scale turbulent premixed bunsen flame using FSD, PCM-FPI and thickened flame models, *Proc. Combust. Inst.* 33 (2011) 1365–1371.
- [5] A.W. Cook, J.J. Riley, A subgrid model for equilibrium chemistry in turbulent flows, *Phys. Fluids* 6 (1994) 2868–2870.

- [6] M. Ihme, H. Pitsch, Prediction of extinction and reignition in non-premixed turbulent flames using a flamelet/progress variable model: 1. A priori study and presumed PDF closure, *Combust. Flame* 155 (2008) 90–107.
- [7] M. Ihme, H. Pitsch, Prediction of extinction and reignition in non-premixed turbulent flames using a flamelet/progress variable model: 2. Application in LES of Sandia flames D and E, *Combust. Flame* 155 (2008) 70–89.
- [8] D. Bradley, P.H. Gaskell, A.K.C. Lau, A mixedness-reactedness flamelet model for turbulent diffusion flames, *Symp. (Int.) Combust.* 23 (1991) 685–692.
- [9] A. Donini, R.J.M. Bastiaans, J.A. van Oijen, L.P.H. de Goey, A 5-d implementation of FGM for the large eddy simulation of a stratified swirled flame with heat loss in a gas turbine combustor, *Flow Turbul. Combust.* 98 (2017) 887–922.
- [10] P. Domingo, L. Vervisch, K. Bray, Partially premixed flamelets in LES of nonpremixed turbulent combustion, *Combust. Theor. Model.* 6 (2002) 529–551.
- [11] J.-B. Michel, O. Colin, C. Angelberger, D. Veynante, Using the tabulated diffusion flamelet model ADF-PCM to simulate a lifted methane-air jet flame, *Combust. Flame* 156 (2009) 1318–1331.
- [12] Z.X. Chen, S. Ruan, N. Swaminathan, Large eddy simulation of flame

- edge evolution in a spark-ignited methane–air jet, Proc. Combust. Inst. 36 (2017) 1645–1652.
- [13] Y.C. See, M. Ihme, Large eddy simulation of a partially-premixed gas turbine model combustor, Proc. Combust. Inst. 35 (2015) 1225–1234.
- [14] F.C. Christo, A.R. Masri, E.M. Nebot, T. Turanyi, Utilising artificial neural network and repro-modelling in turbulent combustion, Proceedings of ICNN’95—International Conference on Neural Networks, volume 2, IEEE, 1995, pp. 911–916.
- [15] F.C. Christo, A.R. Masri, E.M. Nebot, Artificial neural network implementation of chemistry with pdf simulation of H<sub>2</sub>/CO<sub>2</sub> flames, Combust. Flame 106 (1996) 406–427.
- [16] F.C. Christo, A.R. Masri, E.M. Nebot, S.B. Pope, An integrated PDF/neural network approach for simulating turbulent reacting systems, Symposium (International) on Combustion 26 (1996) 43–48.
- [17] J.A. Blasco, N. Fueyo, C. Dopazo, J. Ballester, Modelling the temporal evolution of a reduced combustion chemical system with an artificial neural network, Combust. Flame 113 (1998) 38–52.
- [18] J.A. Blasco, N. Fueyo, J.C. Larroya, C. Dopazo, Y.-J. Chen, A single-step time-integrator of a methane–air chemical system using artificial neural networks, Comput. Chem. Eng. 23 (1999) 1127–1133.
- [19] B.A. Sen, S. Menon, Turbulent premixed flame modeling using artificial neural networks based chemical kinetics, Proc. Combust. Inst. 32 (2009) 1605–1611.

- [20] B.A. Sen, S. Menon, Linear eddy mixing based tabulation and artificial neural networks for large eddy simulations of turbulent flames, *Combust. Flame* 157 (2010) 62–74.
- [21] B.A. Sen, E.R. Hawkes, S. Menon, Large eddy simulation of extinction and reignition with artificial neural networks based chemical kinetics, *Combust. Flame* 157 (2010) 566–578.
- [22] A.K. Chatzopoulos, S. Rigopoulos, A chemistry tabulation approach via rate-controlled constrained equilibrium (RCCE) and artificial neural networks (ANNs), with application to turbulent non-premixed  $\text{CH}_4/\text{H}_2/\text{N}_2$  flames, *Proc. Combust. Inst.* 34 (2013) 1465–1473.
- [23] M. Ihme, A.L. Marsden, H. Pitsch, Generation of optimal artificial neural networks using a pattern search algorithm: Application to approximation of chemical systems, *Neural Comput.* 20 (2008) 573–601.
- [24] M. Ihme, C. Schmitt, H. Pitsch, Optimal artificial neural networks and tabulation methods for chemistry representation in LES of a bluff-body swirl-stabilized flame, *Proc. Combust. Inst.* 32 (2009) 1527–1535.
- [25] 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.
- [26] S. Yao, B. Wang, A. Kronenburg, O.T. Stein, Modeling of sub-grid conditional mixing statistics in turbulent sprays using machine learning methods, *Phys. Fluids* 32 (2020) 115124.

- [27] Z.X. Chen, S. Iavarone, G. Ghiasi, V. Kannan, G. D'Alessio, A. Parente, N. Swaminathan, Application of machine learning for filtered density function closure in MILD combustion, *Combust. Flame* 225 (2021) 160–179.
- [28] H. Yang, T. Kobayashi, S. Iavarone, J.C. Massey, Z.X. Chen, Y. Minamoto, N. Swaminathan, Towards a generalised artificial neural network for sub-grid filtered density function closure in turbulent combustion, *Appl. Energy Combust. Sci.* 14 (2023) 100142.
- [29] T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D.M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, *Adv. Neur. In* 33 (2020) 1877–1901.
- [30] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth  $16 \times 16$  words: Transformers for image recognition at scale, arXiv preprint arXiv: 2010.11929 (2020).
- [31] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pachol-

- ska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature* 596 (2021) 583–589.
- [32] A. Merchant, S. Batzner, S.S. Schoenholz, M. Aykol, G. Cheon, E.D. Cubuk, Scaling deep learning for materials discovery, *Nature* 624 (2023) 80–85.
- [33] V. Xing, C. Lapeyre, T. Jaravel, T. Poinsot, Generalization capability of convolutional neural networks for progress variable variance and reaction rate subgrid-scale modeling, *Energies* 14 (2021) 5096.
- [34] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [35] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested U-net architecture for medical image segmentation, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 3–11.
- [36] L. Nista, C.D.K. Schumann, T. Grenga, A. Attili, H. Pitsch, Investigation of the generalization capability of a generative adversarial network for large eddy simulation of turbulent premixed reacting flows, *Proc. Combust. Inst.* 39 (2023) 5279–5288.
- [37] L. Nista, H. Pitsch, C.D.K. Schumann, M. Bode, T. Grenga, J.F. MacArt, A. Attili, Influence of adversarial training on super-resolution turbulence reconstruction, *Phys. Rev. Fluids* 9 (2024) 064601.

- [38] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017, pp. 4681–4690.
- [39] F. Nicoud, H.B. Toda, O. Cabrit, S. Bose, J. Lee, Using singular values to build a subgrid-scale model for large eddy simulations, *Phys. Fluids* 23 (2011) 085106.
- [40] J.C. Massey, Z. Li, Z.X. Chen, Y. Tanaka, N. Swaminathan, Large eddy simulation of multi-regime combustion with a two-progress variable approach for carbon monoxide, *Proc. Combust. Inst.* 39 (2023) 2117–2127.
- [41] I. Langella, N. Swaminathan, Unstrained and strained flamelets for LES of premixed combustion, *Combust. Theor. Model.* 20 (2016) 410–440.
- [42] J.C. Massey, Z.X. Chen, N. Swaminathan, Modelling heat loss effects in the large eddy simulation of a lean swirl-stabilised flame, *Flow Turbul. Combust.* 106 (2021) 1355–1378.
- [43] J.C. Massey, Z.X. Chen, M. Stöhr, W. Meier, N. Swaminathan, On the blow-off correlation for swirl-stabilised flames with a precessing vortex core, *Combust. Flame* 239 (2022) 111741.
- [44] B. Fiorina, R. Baron, O. Gicquel, D. Thevenin, S. Carpentier, N. Darabiha, Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM, *Combust. Theor. Model.* 7 (2003) 449–470.

- [45] R.W. Bilger, The structure of turbulent nonpremixed flames, *Symp. (Int.) Combust.* 22 (1989) 475–488.
- [46] Z.X. Chen, I. Langella, R.S. Barlow, N. Swaminathan, Prediction of local extinctions in piloted jet flames with inhomogeneous inlets using unstrained flamelets, *Combust. Flame* 212 (2020) 415–432.
- [47] T.D. Dunstan, Y. Minamoto, N. Chakraborty, N. Swaminathan, Scalar dissipation rate modelling for large eddy simulation of turbulent premixed flames, *Proc. Combust. Inst.* 34 (2013) 1193–1201.
- [48] C.D. Pierce, P. Moin, A dynamic model for subgrid-scale variance and dissipation rate of a conserved scalar, *Phys. Fluids* 10 (1998) 3041–3044.
- [49] H. Pitsch, Large-eddy simulation of turbulent combustion, *Annu. Rev. Fluid Mech.* (2006).
- [50] A. Mura, V. Robin, M. Champion, Modeling of scalar dissipation in partially premixed turbulent flames, *Combust. Flame* 149 (2007) 217–224.
- [51] K. Bray, P. Domingo, L. Vervisch, Role of the progress variable in models for partially premixed turbulent combustion, *Combust. Flame* 141 (2005) 431–437.
- [52] S. Ruan, N. Swaminathan, O. Darbyshire, Modelling of turbulent lifted jet flames using flamelets: *a priori* assessment and *a posteriori* validation, *Combust. Theor. Model.* 18 (2014) 295–329.

- [53] G.P. Smith, D.M. Golden, M. Frenklach, N. Moriarty, B. Eiteneer, M. Goldenberg, C.T. Bowman, R.K. Hanson, S. Song, W.C. Gardiner, Jr., V.V. Lissianski, Z. Qin, GRI-Mech 3.0, 2000. [http://www.me.berkeley.edu/gri\\_mech/](http://www.me.berkeley.edu/gri_mech/).
- [54] D.G. Goodwin, H.K. Moffat, I. Schoegl, R.L. Speth, B.W. Weber, Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes, 2017. <https://www.cantera.org>.
- [55] 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.
- [56] R.L. Plackett, A class of bivariate distributions, J. Am. Stat. Assoc. 60 (1965) 516–522.
- [57] N.A.K. Doan, N. Swaminathan, Y. Minamoto, DNS of MILD combustion with mixture fraction variations, Combust. Flame 189 (2018) 173–189.
- [58] Y. Minamoto, N. Swaminathan, Scalar gradient behaviour in MILD combustion, Combust. Flame 161 (2014) 1063–1075.
- [59] M.D. Smoke, V. Giovangigli, Formulation of the premixed and non-premixed test problems, in: M.D. Smoke (Eds.), Reduced Kinetic Mechanisms and Asymptotic Approximations for Methane–Air Flames, Springer, Berlin, Germany, 1991, pp. 1–28.

- [60] T. Kathrotia, U. Riedel, A. Seipel, K. Moshammer, A. Brockhinke, Experimental and numerical study of chemiluminescent species in low-pressure flames, *Appl. Phys. B* 107 (2012) 571–584.
- [61] R.S. Cant, SENG2 user guide, Technical Report CUED/A-THERMO/TR67, Cambridge University Engineering Department, 2012.
- [62] V.A. Epanechnikov, Non-parametric estimation of a multivariate probability density, *Theory Probab. Its Appl.* 14 (1969) 153–158.
- [63] F. Gao, A large-eddy simulation scheme for turbulent reacting flows, *Phys. Fluids A-Fluid* 5 (1993) 1282–1284.
- [64] C. Tong, Measurements of conserved scalar filtered density function in a turbulent jet, *Phys. Fluids* 13 (2001) 2923–2937.
- [65] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 770–778.
- [66] Y. Wu, K. He, Group normalization, *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [67] S. Elfwing, E. Uchibe, K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, *Neural Networks* 107 (2018) 3–11.
- [68] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imper-

ative style, high-performance deep learning library, Advances in Neural Information Processing Systems, volume 32, 2019, pp. 8024–8035.

- [69] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 (2019).
- [70] D.M. Endres, J.E. Schindelin, A new metric for probability distributions, IEEE Trans. Inf. Theory 49 (2003) 1858–1860.
- [71] D. Butz, S. Hartl, S. Popp, S. Walther, R.S. Barlow, C. Hasse, A. Dreizler, D. Geyer, Local flame structure analysis in turbulent CH<sub>4</sub>/air flames with multi-regime characteristics, Combust. Flame 210 (2019) 426–438.
- [72] D. Butz, A. Breicher, R.S. Barlow, D. Geyer, A. Dreizler, Turbulent multi-regime methane-air flames analysed by Raman/Rayleigh spectroscopy and conditional velocity field measurements, Combust. Flame 243 (2022) 111941.
- [73] N. Jarrin, S. Benhamadouche, D. Laurence, R. Prosser, A synthetic-eddy-method for generating inflow conditions for large-eddy simulations, Int. J. Heat Fluid Flow 27 (2006) 585–593.
- [74] N. Kornev, E. Hassel, Method of random spots for generation of synthetic inhomogeneous turbulent fields with prescribed autocorrelation functions, Commun. Numer. Meth. Eng 23 (2007) 35–43.

## Appendix A. Detailed JPResUnet structure

Table A.5: Network structure of JPResUnet.

Unit level	Level components	Filter (size/number)	Stride	Padding	Output size (channels × height × width)
Input				$1 \times 80 \times 100$	
CrossEmbed		$3 \times 3/16$	1	1	
		$7 \times 7/8$	1	3	$32 \times 80 \times 100$
		$15 \times 15/8$	1	7	
Encoder	Level 1	Residual Block	$3 \times 3/32$	1	1
		Residual Block	$3 \times 3/32$	1	1
		Residual Block	$3 \times 3/32$	1	1
		DownSample	$4 \times 4/32$	2	1
	Level 2	Residual Block	$3 \times 3/32$	1	1
		Residual Block	$3 \times 3/32$	1	1
		Residual Block	$3 \times 3/32$	1	1
		DownSample	$4 \times 4/64$	2	1
	Level 3	Residual Block	$3 \times 3/64$	1	1
		Residual Block	$3 \times 3/64$	1	1
		Residual Block	$3 \times 3/64$	1	1
		Parallel	$3 \times 3/128$	1	1
			$1 \times 1/128$	1	0
Bridge	Level 4	Residual Block	$3 \times 3/128$	1	1
	Level 5	Residual Block	$3 \times 3/128$	1	1
Decoder	Level 6	Residual Block	$3 \times 3/128$	1	1
		Residual Block	$3 \times 3/128$	1	1
		Residual Block	$3 \times 3/128$	1	1
		UpSample	$3 \times 3/64$	1	1
	Level 7	Residual Block	$3 \times 3/64$	1	1
		Residual Block	$3 \times 3/64$	1	1
		Residual Block	$3 \times 3/64$	1	1
		UpSample	$3 \times 3/32$	1	1
	Level 8	Residual Block	$3 \times 3/32$	1	1
		Residual Block	$3 \times 3/32$	1	1
		Residual Block	$3 \times 3/32$	1	1
		Identity			$32 \times 80 \times 100$
Output				$1 \times 80 \times 100$	

## Appendix B. Results for the Gaussian filter

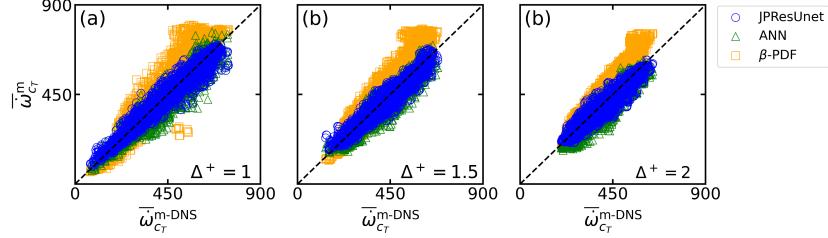


Fig. B.13: Scatter plot of the progress variable source term modelled by joint PDFs from different models and from the DNS for the case AZ1, using the Gaussian filter with different filter widths.

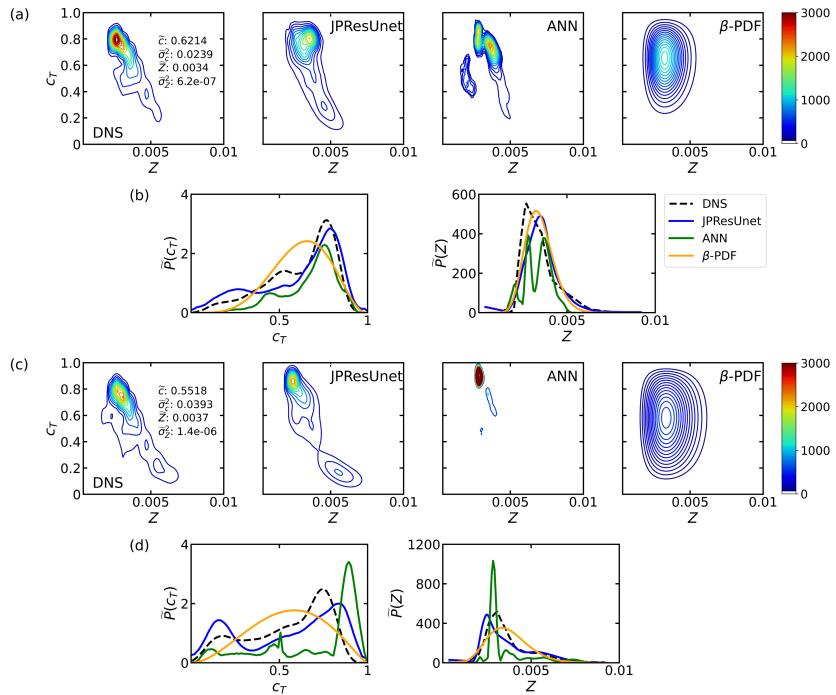


Fig. B.14: Comparative analysis of the joint and marginal PDFs between DNS and model predictions for the case BZ1, utilising a Gaussian filter with widths of (a)-(b)  $\Delta^+ = 1$ , (c)-(d)  $\Delta^+ = 1.5$ .

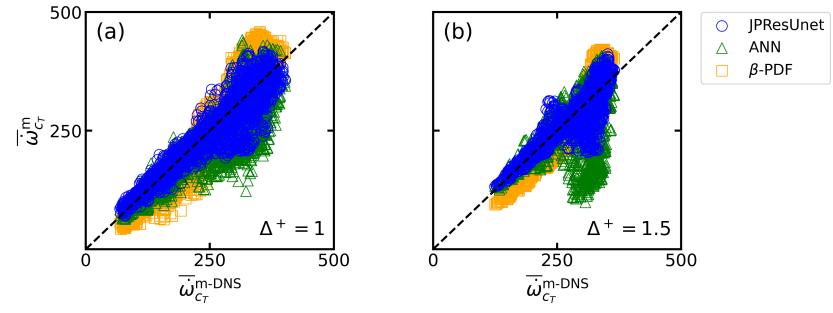


Fig. B.15: Scatter plot of the progress variable source term modelled by joint PDFs from different models and from the DNS for the case BZ1, using the Gaussian filter with different filter widths.