



Investigation of deep learning methods for efficient high-fidelity simulations in turbulent combustion

Kevin M. Gitushi^a, Rishikesh Ranade^b, Tarek Echekki^{a,*}

^a Department of Mechanical and Aerospace Engineering, North Carolina State University, Campus Box 7910, Raleigh 27695, NC, USA

^b Ansys, Inc., Canonsburg 15317, PA, USA

ARTICLE INFO

Article history:

Received 18 May 2021

Revised 11 October 2021

Accepted 12 October 2021

Available online 31 October 2021

Keywords:

DeepONet

Kernel density estimation

Principal component analysis

Independent component analysis

ABSTRACT

Turbulent combustion modeling often faces a trade-off between the so-called flamelet-like models and PDF-like models. Flamelet-like models, are characterized by a choice of a limited set of prescribed moments, which are transported to represent the manifold of the composition space and its statistics. PDF-like approaches are designed to directly evaluate the closure terms associated with the nonlinear chemical source terms in the energy and species equations. They generate data on the fly, which can be used to accelerate the simulation of PDF-like based models. Establishing key ingredients for implementing acceleration schemes for PDF-like methods by constructing flamelet-like models on the fly can potentially result in computational saving while maintaining the ability to resolve closure terms. These ingredients are investigated in this study. They include a data-based dimensional reduction of the composition space to a low-dimensional manifold using principal component analysis (PCA). The principal components (PCs) serve as moments, which characterize the manifold; and conditional means of the thermo-chemical scalars are evaluated in terms of these PCs. A second ingredient involves adapting a novel deep learning framework, DeepONet, to construct joint PCs' PDFs as alternative methods to presumed shapes common in flamelet-like approaches. We also investigate whether the rotation of the PCs into independent components (ICs) can improve their statistical independence. The combination of these ingredients is investigated using experimental data based on the Sydney turbulent nonpremixed flames with inhomogeneous inlets. The combination of constructed PDFs and conditional mean models are able to adequately reproduce unconditional statistics of thermo-chemical scalars, and establish acceptable statistical independence between the PCs, which simplify further the modeling of the joint PCs' PDFs.

© 2021 The Combustion Institute. Published by Elsevier Inc. All rights reserved.

1. Introduction

State-of-the-art turbulent combustion models can be classified under two common classes defined by Pope [1] as flamelet-like [2,3] and PDF-like models [4]. Flamelet-like models such as the flamelet [2] or the conditional moment closure (CMC) [5,6] models, rely on the assumption that the combustion composition space lies within a prescribed low-dimensional manifold that is characterized by a reduced set of moments, such as the mixture fraction and the progress variable. These parameters and their moments are transported. The shape of their probability density functions (PDFs) is often assumed along with their statistical independence.

PDF-like approaches are designed to generate statistics to directly compute the key closure terms in turbulent combustion, which pertain primarily to the chemical source closure terms in

the species and energy equations. Such approaches include transported PDF models [4] and coupled low-dimensional stochastic models with large-eddy simulations (LES). These coupled models combine the linear-eddy model (LEM) with LES, the LEMLES approach, [7–10] or the one-dimensional turbulence (ODT) model with LES, the LES-ODT approach [11–16].

These approaches invariably generate data on the fly, which can be used to determine unconditional means for the reactive scalars and their chemical source terms. This data also can be used to construct a low-dimensional description of the accessed composition space in the simulation and key statistics needed to transport the moments, which characterize the low-dimensional manifolds of this space.

Given the stated trade-off between flamelet-like and PDF-like models, can we construct, on the fly, flamelet-like models using statistics generated from PDF-like simulations? And can we exploit this capability to accelerate PDF-like model-based simulations with a hybrid scheme that accommodates a PDF-like solution and its generated flamelet-like solutions? Addressing these questions must

* Corresponding author.

E-mail address: techekk@ncsu.edu (T. Echekki).

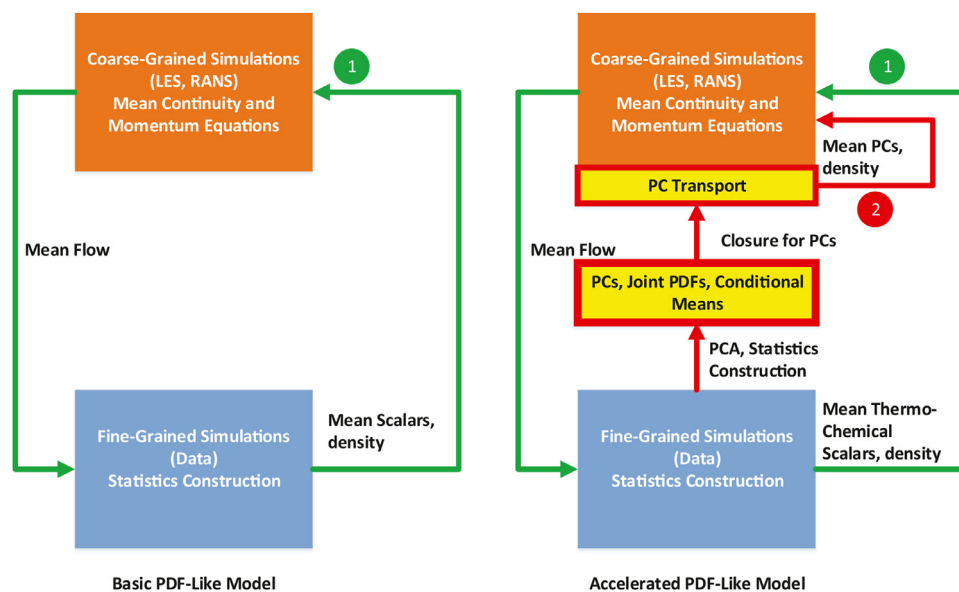


Fig. 1. Schematic of the procedure to accelerate PDF-like simulations.

resolve the challenges that can be faced while implementing such hierarchical schemes.

The first set of challenges is related to the PDF-like solution. Beyond the choice of the PDF-like model, the quality of the generated statistics, including whether the data is sufficiently adequate for constructing statistics and whether it is representative of all scenarios encountered in the simulation, remains a key challenge. Metrics to assess this quality as well as methods to accommodate potentially sparse data may be critical towards addressing this challenge.

The second set of challenges is related to the on-the-fly construction of the flamelet-like model. Recently, Lacey et al. [17] proposed an *in situ* adaptive manifolds (ISAM) approach by enabling on-the-fly calculation of the low-dimensional manifolds of the composition space and the corresponding moments PDFs based on the flamelet approach. By relying on a flamelet assumption or, in general, a prescribed low-dimensional manifold for the composition space, there is an inherent risk that such assumptions may not apply for a given problem. Therefore, a systematic approach for constructing the low-dimensional manifold in which the moments that characterize the manifold are not known *a priori* is desirable. Much of the scope of the present study is related to addressing the on-the-fly construction of the flamelet-like model.

The third set of challenges is related to the coupling of the PDF-like and flamelet-like solutions. Therefore, they are related to how this coupling is implemented, when a PDF-like solution is implemented and when it is replaced by a flamelet-like solution in either zones in the computational domain or at a time interval of the solution.

The study by Wu et al. [18] provides an illustration of how a hierarchy of models can be implemented within the same simulation. Their approach relies on a Pareto-efficient framework to assign submodels subject to metrics of accuracy. Within the framework of combining PDF-like and flamelet-like simulations, such metrics can set criteria for when the low-dimensional manifold must be updated to accommodate additional conditions not accessed by the original data. More recently, Chung et al. [19] proposed and demonstrated a data-assisted dynamic submodel assignment approach using trained random forest classifiers. These classifiers identify the assignment of a particular submodel from a selection of submodels based on the estimated error of quantities of

interest, which correspond to thermophysical quantities. The *in situ* adaptive tabulation (ISAT) approach [20] for accelerating chemistry provides another example of coupling a hierarchy of simulation approaches within the context of combustion.

Figure 1 illustrates how a PDF-like model-based simulation can be augmented/accelerated with a flamelet-like solution. On the left sub-figure is the basic set up for a PDF-like simulation. It involves the coupling of a coarse-grained solution for continuity and momentum based on LES or Reynolds-averaged Navier-Stokes (RANS) with fine-grained, low-dimensional stochastic solutions for reactive scalars. In transported PDF models, the chemical state and associated statistics are tracked through notional particles from which both joint scalar or scalar-velocity probability density functions can be constructed. In LEMLES and LES-ODT approaches, the low-dimensional 1D solutions embedded within or across the LES cells also carry distinct chemical states along their 1D profiles. Therefore, the fine-grained simulations can generate statistics on the fly.

In a coupled PDF-like and flamelet-like model-based framework, which is shown on the right sub-figure in Fig. 1, the flamelet-like model and associated low-dimensional manifold is constructed and updated on the fly using statistics from the PDF-like model-based simulation from the current and previous time steps of the solution. We propose principal component analysis (PCA) [21] as a viable approach for constructing this manifold. Accordingly, the thermo-chemical scalars (species and energy) are replaced by a much smaller subset represented by the principal components (PCs) from PCA. Adopting a low-dimensional manifold also means that unconditional statistics of thermo-chemical scalars can be constructed based on the retained moments of the low-dimensional manifold, which correspond to the PCs. To evolve the solution in a flamelet-like model-based simulation, the unconditional moments of the PCs are transported along with the continuity and momentum equations as shown in Fig. 1. The PC-transport and momentum equations include closure terms, such as the unconditional means for the density and the chemical source terms for the PCs, which are modeled in terms of the unconditional means of the PCs [22,23]. Similar closure for the mean molecular diffusion terms for PCs and momentum can also be constructed in terms of the PCs' unconditional means. The solution of PC-transport equations [22–26] is an alternative approach to the transport of the traditional moments used in flamelet-like models,

such as the mixture fraction and the progress variable and their moments.

As the coarse-grained solution evolves, the range of accessed states can be updated by fine-grained solutions and the corresponding PCs can be adaptively updated, as well. It is important to emphasize that fine-grained simulations within PDF-like models can be used to generate both means for thermo-chemical scalars conditioned upon the manifold variables as well as statistical distributions (PDFs) of the manifold variables. Within this strategy, the solution of the PC-transport equations provide a mechanism for evaluating unconditional thermo-chemical scalars' statistics (e.g. unconditional means of temperature or species). In flamelet-like models, it is common to use a presumed PDF shape. However, the availability of statistics from PDF-like model-based solutions results in additional, and potentially more effective, methods to construct these PDFs.

In the present study, we investigate the key ingredients to construct a flamelet-like model from PDF-like simulation data and extract important statistics for its closure. These ingredients include the construction of low-dimensional manifolds, conditional means and statistical distributions associated with the manifold parameters. The use of PCA to construct a low-dimensional manifold for the composition space provides a systematic method to accommodate complex problems, which may not be characterized by a single combustion mode or regime or well-defined reference streams or measures of the completion of reaction. We also rely on machine learning (ML) approaches as efficient tools to construct conditional means for thermo-chemical scalars based on PCs using deep neural networks and joint distributions for these PCs using the kernel density estimation (KDE) [27] approach and the DeepONet architecture [28].

In this study we also examine strategies to establish statistical independence of the PCs using independent component analysis (ICA) and the generation of independent components (ICs). Establishing statistical independence can help further simplify the description of the joint PCs/ICs PDFs and enable the use of more sparse data for their construction. We use experimental data based on the Sydney flames [29–31]. These flames provide a rich data set that exhibits different dominant combustion modes and non-equilibrium effects.

In Section 2, we outline the general implementation of PCA to construct low-dimensional manifolds from PDF-like model-based simulations. In Section 3 the various machine learning methods are discussed. Next, results are presented and discussed in Section 4.

2. PCA and ICA-based low-dimensional manifolds for the composition space

The identification of a data-based low-dimensional manifold for the composition space is one of the key ingredients towards the acceleration of PDF-like model simulations. The resulting moments for this manifold replace the standard prescribed moments in traditional flamelet-like models. In this study, we propose the use of PCA to identify the composition space low-dimensional manifold.

PCA has been used in combustion for chemistry reduction (see for example Vajda et al. [32]) as well as for the parameterization of the composition space [24,26,33–38]. Sutherland and Parente [24] derived governing equations for PCs starting from the transport equations for the thermo-chemical scalars (species and temperature). This work has motivated a number of subsequent studies that explored PC-transport within the context of DNS [25], one-dimensional turbulence (ODT) [38], LES [26] and RANS [22,23]. In the work of Ranade et al. [22,23], a flamelet-like model based on PC transport is constructed using experimental data. In this model, the unconditional means of the thermo-chemical scalars and the

PCs' source terms are evaluated by constructing models for their means conditioned on the retained PCs and the PCs' joint PDFs.

When adopting PCA, the corresponding PCs' transport equations in a RANS or LES formulation are very similar in form to those of passive/reactive scalars' moments common in combustion modeling [22,23]. The most important advantage, of course, is in the reduced description of the composition space and the lower dimension of the statistical distributions for these PCs. These transport equations require closure for the unconditional means of the PCs source terms. These terms can be evaluated directly from the PDF-like model-based simulation statistics, including means for the PCs chemical source terms conditioned on the PCs as well as the joint PCs PDFs or filtered density functions (FDFs) for LES.

PCA results from a linear transformation of the vector of thermo-chemical scalars (e.g. temperature and species mass fractions) to yield uncorrelated scalars, the PCs:

$$\phi = \mathbf{A}^T \psi, \quad (1)$$

where ψ is the vector of N thermo-chemical scalars normalized to yield comparable magnitudes using maximum and minimum values of these scalars: $\psi = (T, Y_1, Y_2, \dots, Y_{N-1})$. The PCs vector ϕ includes N_{pc} principal components, $\phi = (\phi_1, \phi_2, \dots, \phi_{N_{pc}})$, which account for a threshold percentage of the data variance. For example, the data from the Sandia flames [39] and the Sydney flames [29–31] require 2 to 3 PCs to capture the entire range of flow conditions and reaction scenarios (mixing, extinction and reignition) in these flames. The matrix $\mathbf{A}$ includes the leading N_{pc} eigenvectors of the normalized thermo-chemical scalars covariance matrix.

Unconditional means for the k th thermo-chemical scalar's source term can be expressed using the joint thermo-chemical scalars PDFs as follows:

$$\overline{\omega_k(\psi)} = \int_{\psi} \omega_k(\psi) p(\psi) d\psi. \quad (2)$$

In this expression, $p(\psi)$ is the joint thermo-chemical scalars' PDF, which can be highly dimensional, and the instantaneous thermo-chemical scalars source is $\omega_k(\psi)$.

Mapping the composition space into a low-dimensional manifold based on the PCs yields a similar form for the thermo-chemical scalars unconditional means:

$$\overline{\omega_k(\phi)} = \int_{\phi} \langle \omega_k | \phi \rangle p(\phi) d\phi. \quad (3)$$

In this expression, $\langle \omega_k | \phi \rangle$ is the mean of the thermo-chemical scalars conditioned on the PCs. Conditional means are constructed from data; they can be viewed as generalizations of flamelet libraries where the parameterization is based on the PCs instead of prescribed parameters (e.g. mixture fraction, progress variable). The joint PCs' PDF, $p(\phi)$, has a significantly reduced dimension compared to the original joint thermo-chemical scalars' PDF, $p(\psi)$.

A similar expression to Eq. (3) for thermo-chemical scalars' unconditional means can be written:

$$\widetilde{\psi_k(\phi)} = \int_{\phi} \langle \psi_k | \phi \rangle \tilde{p}(\phi) d\phi, \quad (4)$$

where, the superscript “~” refers to a density-weighted or Favre averaging. Again, $\langle \psi_k | \phi \rangle$ represents the k th thermo-chemical scalar's mean conditioned on the PCs. $\tilde{p}(\phi)$ is the density-weighted PCs' PDF, which may be expressed as:

$$\tilde{p}(\phi) = \frac{\langle \rho(\phi) \rangle}{\bar{\rho}} p(\phi). \quad (5)$$

The next challenge is the representation of the thermo-chemical scalars conditional means and the joint PCs' PDF. Although we can carry out subsequent analysis with an arbitrary number of retained PCs, we will discuss primarily here the case where $N_{pc} = 2$. This choice is made to simplify the analysis and also because the data

used for validating the proposed approach can adequately be represented by 2 PCs. In this scenario, the joint PCs' PDF can be expressed as: $\tilde{p}(\phi) = \tilde{p}(\phi_1, \phi_2)$.

The resulting PCs are uncorrelated; yet, PCA does not guarantee their statistical independence. A related approach, independent component analysis (ICA) [40,41] determines a basis for the data in which the new set, the independent components (ICs), are as statistically independent as possible. ICA yields a "rotation" of the PC vector:

$$\zeta = \mathbf{B}^T \phi. \quad (6)$$

In this expression, ζ represents the vector of N_{pc} independent components and $\mathbf{B}$ is the rotation or mixing matrix. This rotation is obtained through optimization algorithms that yields variables $\zeta_1, \zeta_2, \dots, \zeta_{N_{pc}}$ that are as statistically independent as possible [41,42]. Transport equations for the ICs also can be derived similarly to those of the PCs.

Representing the data in IC space results in a simplified expression for the ICs joint PDFs in terms of the product of the individual ICs' marginal PDFs. The value of determining ICs in combustion resides in this convenience where the assumption of statistical independence does not need to be invoked to model joint PDFs. Regardless, both PCA [21] and ICA [41] seek to statistically project the original data into a new space through a linear transformation of the original variables.

The principal limitation of ICA is that it cannot be used for model reduction. Instead, it has to be coupled with PCA as an initial step [43,44]. Within this procedure, PCA generates a reduced set of parameters to represent the multi-dimensional composition space; subsequently, the PCs are "rotated" to generate a more statistically independent set of parameters in terms of ICs.

3. Machine learning tools for the construction of statistics

The construction of flamelet-like model requires, in addition to the choice of the transported moments, the development of approaches to construct conditional means and the PCs' joint PDFs. The statistics must be constructed on the fly and can be evaluated incrementally as additional data from the PDF-solution becomes available. If this data corresponds to states and statistics that have not been accessed previously, considerations for the cost of 'training' a new flamelet-like model become important. Machine learning approaches rely on complex neural networks whose parameters (e.g. the weights associated with neuron connections) can be determined through an optimization process that maps required inputs to desired outputs. If the data is introduced incrementally, the new network parameters may not significantly deviate from the previous ones, and the cost of training can be significantly reduced. This feature is perhaps one of the key motivations for adopting machine learning to construct *in situ* statistics for flamelet-like models. Additional features are discussed for the specific networks adopted for these statistics.

3.1. Computation of the conditional means

One of the principal ingredients for the construction of statistics on the fly from PDF-like model-based simulations is the evaluation of the means of thermo-chemical scalars conditioned on the PCs. Eq. (1) provides a direct expression of the PCs in terms of the thermo-chemical scalars. This relation can be inverted to convert from the full PCs' vector to the original thermo-chemical scalars. However, to recover the original thermo-chemical scalars from a subset of the PCs, a nonlinear mapping relation is needed. A nonlinear mapping is equally needed to express the conditional means of the PCs' chemical source terms. We propose to carry

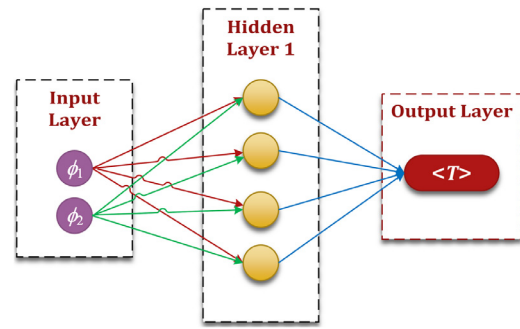


Fig. 2. Illustration of an ANN network for determining temperature conditional means.

out this mapping using an ANN with a required input corresponding to the retained N_{pc} PCs and the output corresponding to the thermo-chemical scalars or their derived quantities (e.g. the thermo-chemical scalars' source terms).

The data required for training the network is based on the raw data in which the PCs are calculated along with the computed or measured thermo-chemical scalars to yield conditional means for these scalars, $\langle \psi_k | \phi \rangle$. The choice of using ANN as a regression tool to construct conditional means is motivated by a few attributes of ANN. First, the cost of implementing ANN training can be significantly reduced if the training data is incrementally added, such that the ANN optimization can be completed within a few epochs. Second, ANN tabulation with proper clustering can be a computationally efficient alternative to, and requires less storage than, interpolation methods for relatively high-dimensional tables [45,46]. For low-dimensional tabulation, such as the case in the present study, the advantage of ANN-based regression can be associated with the ability to 1) generate smooth functions even when the data is sparse across the composition space, and 2) incrementally add data by training primarily on the new data and a subset of the old data.

A representative ANN architecture is shown in Fig. 2 for the calculation of the temperature conditional means starting with 2 PCs. Similar networks can be adopted for the other thermo-chemical scalars. The input layer contains neurons associated with the input variables, ϕ_1 and ϕ_2 , hidden layers (only one hidden layer with 4 neurons is shown in the figure), and an output layer, containing the temperature conditional mean. Although not shown on the figure, additional bias neurons are added to the input and all hidden layers. Within the context of the present study, the network is trained on the available measurements in the Sydney flame based on all the available downstream distance and radial positions. The training determines the strength of the connections between the previous and subsequent layers from input to output as indicated by the arrows in the figure.

To illustrate the relation between the input and the output layers, we use the network illustrated in Fig. 2. The output $\langle T | \phi \rangle$ is expressed in terms of the hidden layer:

$$\langle T | \phi \rangle = f\left(\sum_{i=1}^4 w_i^{(1)} a_i^{(1)} + b^{(1)}\right), \quad (7)$$

where the superscript (1) corresponds the first hidden layer with weights $w_i^{(1)}$ and values $a_i^{(1)}$ at the i th neuron in the hidden layer. $b^{(1)}$ is the bias value at the hidden layer and f is the activation function. We use the sigmoid function for activation in the ANN networks for the conditional means. The values of the i th neuron, $a_i^{(1)}$, in the hidden layer can be related to the input values as follows:

$$a_i^{(1)} = f(w_{i1}^{(0)} \phi_1 + w_{i2}^{(0)} \phi_2 + b^{(0)}). \quad (8)$$

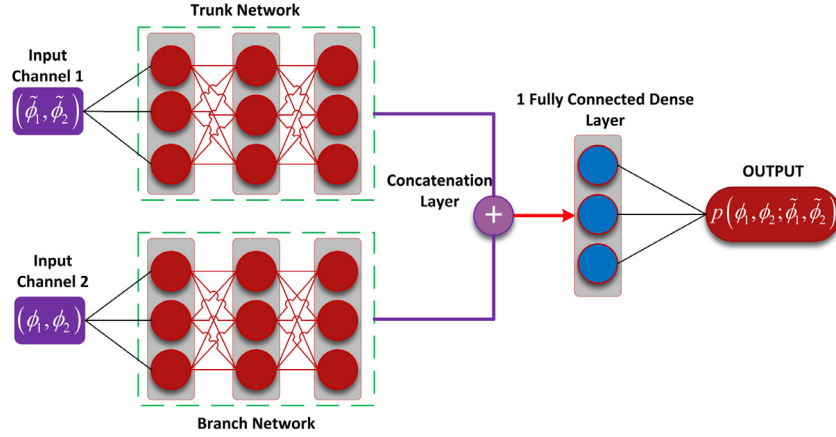


Fig. 3. Illustration of a DeepONet for determining a PC's generalized PDF.

Here $w_{1i}^{(0)}$ and $w_{2i}^{(0)}$ correspond to the weights of the connections between the input layer and the i th neuron in the first hidden layer associated with inputs ϕ_1 and ϕ_2 , respectively. The network is trained to determine the weights of all connections from input to output layers and the bias values.

3.2. Approximations of the PCs joint PDF

The use of fine-grained simulations to construct directly PDFs of thermo-chemical scalars via tabulation have provided a viable alternative to the presumed shape PDF assumption [47–50]. However, tabulation is not immune to the curse of dimensionality and may present challenges for memory and storage in computations.

Machine learning methods offer an alternative to the approximation based on presumed shapes, the solution of the transport equations or the tabulation of these PDFs. Although joint PDFs can be readily available from the fine-grained simulations' data, the abundance of this data or the lack of it can limit the ability to accurately reconstruct these distributions.

In their recent work, de Frahan et al. [51] investigated three machine learning techniques to reconstruct PDFs for LES parameterized with the mixture fraction and the progress variable. These techniques include the use of random forests, deep learning neural networks and conditional variational autoencoders. The different techniques rely on a general strategy of single input channels that yield an output, which represents the PDF. The implementations of de Frahan et al. [51] of the 3 machine learning techniques showed that the predictions of the joint PDFs yield significant improvements on β -function based presumed PDFs using DNS data.

Kernel density estimation (KDE) [27] provides a strategy for establishing arbitrary shapes for a PDF given discrete data at a given spatial position. In a KDE, a PDF can be expressed as a weighted sum of discrete kernel functions, K (e.g. Gaussians) centered at discrete values of the parameters, say the PCs:

$$p(\phi; \tilde{\phi}) = \frac{1}{n h} \sum_{i=1}^n K\left(\frac{\phi - \hat{\phi}_i}{h}\right). \quad (9)$$

In this expression, ϕ represents the vector of retained PCs, K is the kernel function, h is the so-called bandwidth, which controls the smoothing of the approximation, and $\hat{\phi}_i$ is the i th sample of ϕ values out of a total of n samples. Such a distribution can be evaluated at a specific spatial point, which can be parameterized in terms of an unconditional mean for the PCs, $\tilde{\phi}$. Therefore, different shapes can be evaluated at different spatial positions. However, to determine a “generalized” PDF that can be adopted for a wide

range of unconditional means for the PCs or spatial positions, additional steps are needed.

3.3. A generalized machine learning reconstruction of the PCs' joint PDF

Although ML-based methods to reconstruct parameterized PDFs have been proposed in the past [51], there is still need to construct PDF shapes that can accommodate different PDFs without resorting to presumed shapes. The most recent work by Lu et al. [28] offers a potentially more powerful strategy to reconstruct joint or marginal PDFs from combustion data. The approach is based on the so-called DeepONet network, which is designed to learn non-linear functions and operators from data. Within the context of the present work, DeepONet takes different functions, PDF shapes, from different positions in the flame, which are labeled with the corresponding PCs' unconditional means; then, the network determines other functions at prescribed PCs unconditional means. Therefore, instead of training different networks for different PDFs at prescribed PCs unconditional means, the DeepONet is used to incorporate these means into a single network that accommodates these PDFs. This attribute of learning operators can reduce network complexity and require less training data to achieve the same predictions [28].

The DeepONet architecture that corresponds to the present implementation is illustrated in Fig. 3. The network consists of 2 separate channels: the branch channel and the trunk channel. The input to the branch channel corresponds to the PC vectors sampled in PC space, identified in the figure with the pair (ϕ_1, ϕ_2) . The input to the trunk channel corresponds to the spatial unconditional mean of the PCs, identified in the figure with the pair (ϕ_1, ϕ_2) . For the present implementation of the DeepONet, the branch and trunk networks consist of 2 dense hidden layers with 64 neurons each. The outputs of these layers are concatenated and passed through an additional dense layer of 64 neurons followed by an output layer of 1 neuron. This output corresponds to the PCs' PDF.

KDE is used to fit PDF shapes at individual measurement positions. These positions are labeled with the retained PCs unconditional means, $\tilde{\phi}_1$ and $\tilde{\phi}_2$. For each position, 200 PC vectors $(\tilde{\phi}_1, \tilde{\phi}_2)$ are sampled, which span the entire range of the PC values. The set of PC vectors and associated PCs' unconditional means and output PDF values constitute the training data for the DeepONet to predict PDFs. As pointed out in our previous studies [22,23], the use of the unconditional variances of the PCs does not alter, in any discernable way, the PDF predictions. We believe that the main reason for this resides in the uncorrelated nature of the PCs. Pairs of PCs un-

conditional means in the flame are reasonably unique and uncorrelated, like their instantaneous counterparts; and, thus, they tend to generate unique distributions that can be described in terms of these unconditional means.

As outlined earlier, PCs define a low-dimensional manifold of the composition space. However, PCA generates a set of uncorrelated PCs, which may or may not be statistically independent. The joint PCs' PDFs are inherently multi-dimensional. Therefore, their reconstruction requires an adequate amount of data to span the distributions among different values of the PCs. It is desirable to adopt an even lower dimensional representation for these PDFs, such as in terms of marginal PDFs of the individual PCs. For this, 2 different strategies are considered:

- The simplest form is to assume that the PCs are statistically independent at least for a range of conditions where the shape and magnitude of the PDFs matters. Within this assumption, the joint PCs' PDFs can be expressed for a 2-PC parameterization as:

$$p(\phi_1, \phi_2) = f(\phi_1) g(\phi_2), \quad (10)$$

where f and g are the marginal PDFs of ϕ_1 and ϕ_2 , respectively. The idea of exploring the assumption of statistical independence of the PCs is not an unreasonable one. Uncorrelatedness for PCs represents a weaker manifestation and an essential criteria of statistical independence. Although not pursued in the present study, we have also investigated the use of copulas [52] as a way to "bind" the PCs marginal PDFs and account for their potential statistical independence. However, we have found that copulas do not provide any significant correction to the relation provided in Eq. (10).

- The second strategy is to construct independent components from the PCs using ICA [40,41].

As stated earlier, PCA is a desirable step prior to ICA to enable a reduced parameterization of the combustion composition space [42]. By using PCA in a pre-processing step ahead of ICA, ICA still yields a lower-dimensional manifold description of this space.

4. Results

In this section, we present results of the *a priori* validation of the outlined data processing based on machine learning techniques discussed in Section 3.

4.1. Experimental conditions

The dataset used in the following analysis is based on multi-scalar measurements in the Sydney piloted jet flames by Meares et al. [29–31]. The burner is configured to have 3 concentric tubes in which the fuel flows through the most inner tube, surrounded by co-flow air with a bulk velocity 15 m/s and outer tube containing hot gases from a pilot flame. The 3 tube diameters are 4, 7.5 and 18 mm, respectively. The most prominent feature of the burner configuration is the ability to control the recess distance of the inner-most tube relative to the co-flow air. A fully recessed fuel tube allows for maximum premixing of the fuel and air; while, zero recess results in a purely pilot-stabilized nonpremixed flame. The fuel is methane (CP grade) with 99% CH₄ at a temperature of 294 K. The pilot stream flame uses a 5-gas mixture of C₂H₂, H₂, CO₂ and N₂ with a bulk inlet velocity of 3.7 m/s.

This flame is characterized by the presence of different dominant combustion modes evolving from primarily premixed mode near the jet inlet to a primarily non-premixed mode further downstream with a transition in between. Like the Sandia flames [39], the flame is also characterized by the presence non-equilibrium

effects, including extinction and reignition. These attributes associated with evolving burning modes and burning stabilities result in important variations in the joint and marginal PDFs of thermo-chemical scalars as well as the PCs. Given all these features, the composition space for the Sydney flames cannot be parameterized in terms of the conventional mixture fraction/progress variable parameters alone across the various downstream conditions.

In the present study, the composition space associated with the different multi-shot, multiscalar measurements is parameterized with 2 PCs (i.e. $N_{pc} = 2$) out of 7 PCs. This number is deemed sufficient for the reconstruction of most statistics. However, it is likely that an additional PC is needed to fully capture statistics near the inlet. Adding more PCs can improve the accuracy of the flamelet-like model. This comes at the cost of a higher dimensionality for the manifold, the conditional means and the joint PDFs. The trade-off between a higher dimensionality and model fidelity is an important consideration during the flamelet-like model construction.

Finally, PCA is carried out using a subset of the thermo-chemical scalars' vector. This subset corresponds to the measured quantities, temperature and the following reported species mass fractions for: CH₄, O₂, CO₂, H₂O, CO, H₂. These scalars adequately represent the evolution of mixing and chemistry, including events of extinction and reignition. In the past, we have found it to be more convenient to carry out PCA using a set of representative thermo-chemical scalars and reconstruct the remaining ones through their correlations with the representative scalars [25,34–38].

In the present study, we consider one flame condition corresponding to recess distance $L_r = 75$ mm and a bulk jet velocity of 57 m/s (the so-called FJ200-5GP-Lr75-57 flame) [29–31]. In the Sydney flames, the vector of PCs is related to the normalized measured thermo-chemical scalars as follows:

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \phi_5 \\ \phi_6 \\ \phi_7 \end{pmatrix} = \begin{pmatrix} -0.59 & 0.57 & 0.12 & -0.35 & -0.30 & -0.01 & -0.32 \\ 0.24 & 0.46 & -0.83 & 0.10 & 0.11 & -0.01 & -0.17 \\ 0.31 & -0.13 & 0.22 & -0.01 & 0.20 & -0.04 & -0.90 \\ 0.66 & 0.48 & 0.37 & 0.03 & -0.42 & 0.00 & 0.12 \\ 0.12 & 0.31 & 0.20 & -0.48 & 0.76 & -0.04 & 0.22 \\ 0.23 & -0.36 & -0.27 & -0.80 & -0.33 & -0.02 & 0.01 \\ 0.02 & 0.01 & 0.01 & -0.04 & 0.03 & 0.10 & -0.03 \end{pmatrix} \begin{pmatrix} T^* \\ Y_{O_2}^* \\ Y_{CH_4}^* \\ Y_{H_2O}^* \\ Y_{CO}^* \\ Y_{H_2}^* \\ Y_{CO}^* \end{pmatrix} \quad (11)$$

The superscript "*" indicates a normalization of the thermo-chemical scalars to yield values from -1 to 1. The weights in the conversion matrix also measure the contribution of the thermo-chemical scalars to the PCs. The first 2 PCs contribute 94% of the data variance. Both temperature and the O₂ mass fractions constitute the main contributions to the first PC; while the second PC's main contribution comes from the reactants, the O₂ and the CH₄ mass fractions. The first and second PCs correlate reasonably well with a progress variable and a mixture fraction at downstream distances beyond 10 diameters of the fuel jet where nonpremixed burning is dominant. However, they deviate from these parameters near the jet inlet where premixed burning is dominant. The intermediates H₂ and CO exhibit the most influence on the seventh and third PCs, respectively.

4.2. Determination of the conditional means

As stated earlier, the conditional means for all species and temperature are recovered using an inversion of the relation in Eq. (1). They are implemented using a neural network where the 2 instantaneous PCs are prescribed as inputs and the various thermo-chemical scalars are prescribed as output. The data used to construct the conditional means is drawn from all downstream and radial positions in the flame.

Table 1 summarizes information about the architecture of the various networks used to train temperature and species O₂, H₂O,

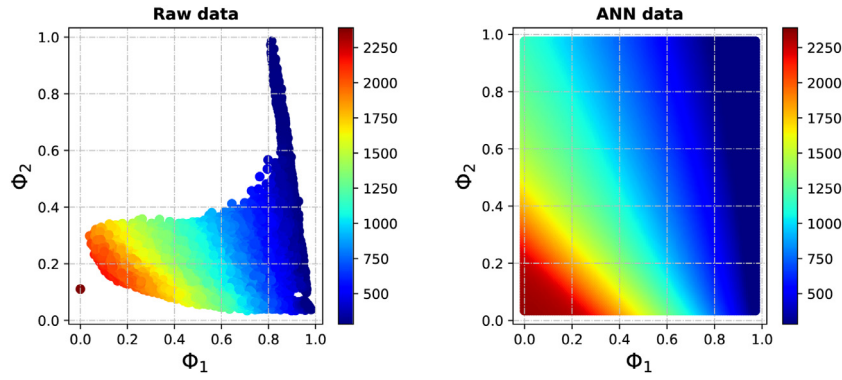


Fig. 4. Conditional means of temperature based on raw data (Left) and ANN regression (right).

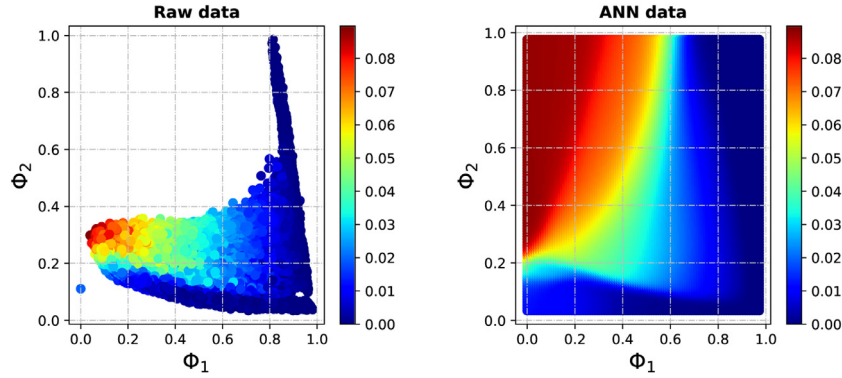


Fig. 5. Conditional means of the CO mass fraction based on raw data (Left) and ANN regression (right).

Table 1

ANN architectures for different output variables.

Output	Layer 1	Layer 2	Layer 3	Layer 4	Weights
T	40	20	10	N/A	1090
H ₂	128	128	64	N/A	24,896
O ₂	40	20	10	N/A	1090
H ₂ O	40	20	10	N/A	1090
CH ₄	40	20	10	N/A	1090
CO	128	128	64	10	25,482
CO ₂	40	20	10	N/A	1090
Density	40	20	10	N/A	1090

CH₄, CO₂ mass fractions conditional means. The training for the different quantities has different requirements depending on the complexity of the correlations between PCs and species and temperature. For example, if a PC has a strong correlation with a thermo-chemical scalar, then, the architecture tends to be simpler. The number of hidden layers ranges from 3 to 4 with the number of neurons per hidden layer ranging from 10 to 128, resulting in a total number of trainable weights ranging from approximately 1000 to more than 25,000. The average training time on Google Colab ranges from approximately 20 sec to approximately 2 minutes (for H₂ and CO). The sigmoid activation function is used for these networks. The number of epochs needed for convergence ranges from 100 for the simpler architectures to 500 for CO and H₂.

Figures 4 and 5 show the conditional means for temperature and the CO mass fractions, respectively, based on the raw data and the ANN regression. The evaluation of the means conditioned on the PCs based on the raw data is implemented by binning the PC

values across their entire range over different bins and calculating averages in each bin. The number of bins adopted for each PC is 20. The extent of the composition space is indicated in the raw data. Although, ANN is able to provide regression outside these boundaries. While such a regression represents an extrapolation for ANN, it is not explicitly used in determining the conditional means since the corresponding PDFs are zero.

The results in Figs. 4 and 5 clearly show that the statistics based on the raw data and those based on the ANN regression exhibit both qualitative and quantitative agreements. However, the ANN-based regression can result in smoother profiles. Hence, it can accommodate more sparse data. Moreover, the resulting regression requires less memory as only the trained parameters of the network need to be retained instead of the storage of a multi-dimensional data. From a computational perspective, lower memory requirements may enable the use of accelerated hardware, such as GPUs.

4.3. Statistical independence of PCs and ICs

In this section, we investigate the extent of statistical independence between PCs and ICs using different correlations and independence coefficients: 1) the Pearson correlation, K_{Pearson} , 2) the Kendall's τ , K_{Kendall} , and 3) the Spearman's ρ , K_{Spearman} . The Pearson correlation for the PCs can be evaluated using the ratio of the covariance of the PCs or ICs and the products of the individual PCs standard deviations:

$$K_{\text{Pearson}} = \frac{\text{cov}(\phi_1, \phi_2)}{\sigma_{\phi_1} \sigma_{\phi_2}}. \quad (12)$$

A similar expression may be written for the ICs. Based on its form, the Pearson's correlation is a measure of the linear relationship

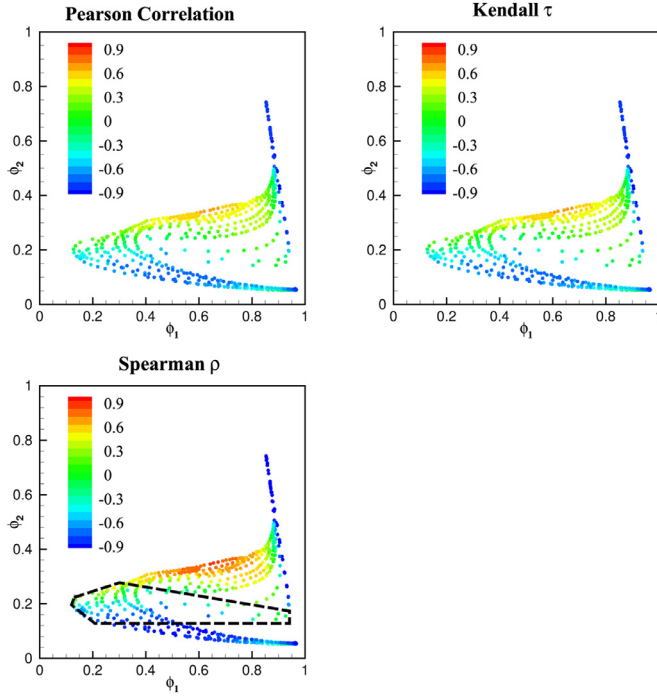


Fig. 6. Measures of correlations of the data parameters at different PCs unconditional means.

between the two variables ϕ_1 and ϕ_2 . Both the Kendall's τ and the Spearman's ρ provide standard measures of statistical independence. The Spearman's ρ looks at the monotonic relationship between the two variables ϕ_1 and ϕ_2 . To evaluate the correlation, the pairs of variables are ranked (e.g. from smallest to highest) and are assigned their rank (e.g. 1 for the lowest and 2 for the next lowest), i_{ϕ_1} and i_{ϕ_2} , respectively. The Spearman's ρ is then calculated using the Pearson's correlation on the ranks (and not on the values of the PCs):

$$K_{\text{Spearman}} = \frac{\text{cov}(i_{\phi_1}, i_{\phi_2})}{\sigma_{i_{\phi_1}} \cdot \sigma_{i_{\phi_2}}} \quad (13)$$

The Kendall's τ measures the difference of the probabilities of concordant pairs of PCs and discordant pairs of PCs. The criteria for concordant and discordant pairs is based on the comparison of two data pairs identified by their indices i and j : $(\phi_1, \phi_2)_i$ and $(\phi_1, \phi_2)_j$ where $1 \leq j \leq n$. Here, n is the total number of data points. Two conditions apply to make the pair concordant: either $(\phi_1)_i > (\phi_1)_j$ and $(\phi_2)_i > (\phi_2)_j$ or $(\phi_1)_i < (\phi_1)_j$ and $(\phi_2)_i < (\phi_2)_j$. Otherwise, the pair is discordant. The Kendall's τ is then expressed as:

$$K_{\text{Kendall}} = \frac{n_{\text{concordant}} - n_{\text{discordant}}}{\frac{1}{2}n(n-1)} \quad (14)$$

where $n_{\text{concordant}}$ and $n_{\text{discordant}}$ are the numbers of concordant and discordant pairs in the data, respectively. All correlation coefficients range from -1 to 1. The value of 1 corresponds to a maximum positive correlation while the value of -1 correspond to a maximum negative correlation. A correlation of 0 or a low value of the correlation is an indicator of statistical independence for the Spearman's ρ and Kendall's τ .

Figure 6 shows scatter plots of the Pearson's correlation, the Kendall's τ and the Spearman's ρ at different measurement positions in Sydney flame FJ200-5GP-Lr75-57. The different positions are parameterized by the unconditional Favre means of the first two PCs, $\tilde{\phi}_1$ and $\tilde{\phi}_2$. The different subfigures are shown with the

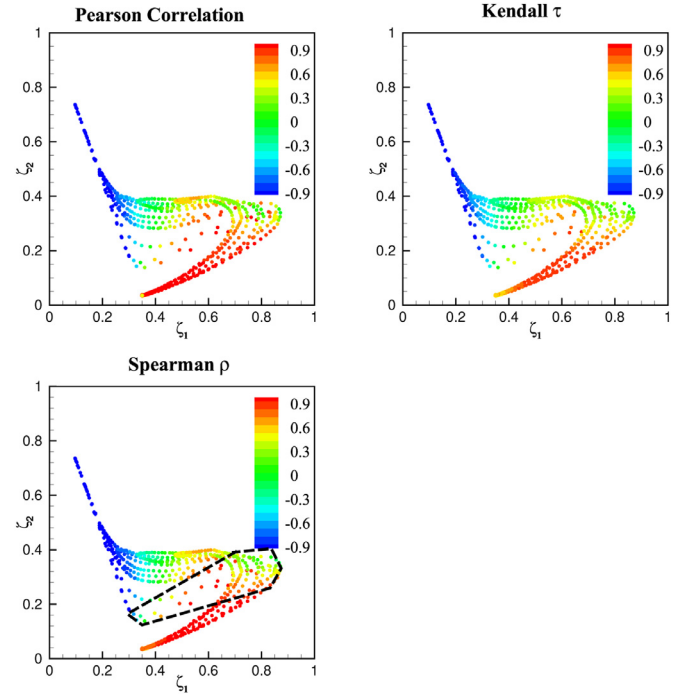


Fig. 7. Measures of correlations of the data parameters at different ICs unconditional means.

same range for the correlation values. The box drawn in dashed lines on the Spearman's ρ subfigure indicates a range of mean mixture fractions between 0.05 and 0.1, which represents conditions around the flame position.

The figure shows that all correlation parameters exhibit qualitatively similar trends indicating essentially where these correlations exhibit the most statistical independence and where the PCs are positively (red) or negatively (blue) correlated. The conditions where the correlations are near zero correspond to regions of flame crossings where the unconditional means for the mixture fraction are near the stoichiometric value of 0.055. There are, however, some important quantitative differences, which may be attributed to the definitions of the different correlations. The figure also shows that neither the PCs nor the ICs exhibit statistical independence everywhere in the flame. Such a condition is not an absolute requirement since the value of statistical independence is most important when there is significantly variability (i.e. high values of the PDFs) at a given spatial position.

Finally, it is important to note that although PCA is established to generate uncorrelated PCs, this process is implemented on the entire set of data. This process does not guarantee that most measurement conditions will also yield lower correlation values, including the Pearson correlation. The same applies to the ICA results discussed below.

Figure 7 shows the same coefficients shown in Fig. 6, but now considering the statistics of the ICs. Also shown on the Spearman's ρ subfigure is the region of mean mixture fractions between 0.05 and 0.1, indicating conditions near the flame zone. Given the rotation associated with ICA, the shape of the scatter is different from that of the PCs. However, all figures exhibit similar quantitative trends as seen in the PCs' plots. The ICs are most statistically independent around the flame zone. More importantly, there are now more measurements points that exhibit low correlation values compared to the PCs. The implementation of ICA has slightly improved the statistical independence of the parameters of the low-dimensional manifold.

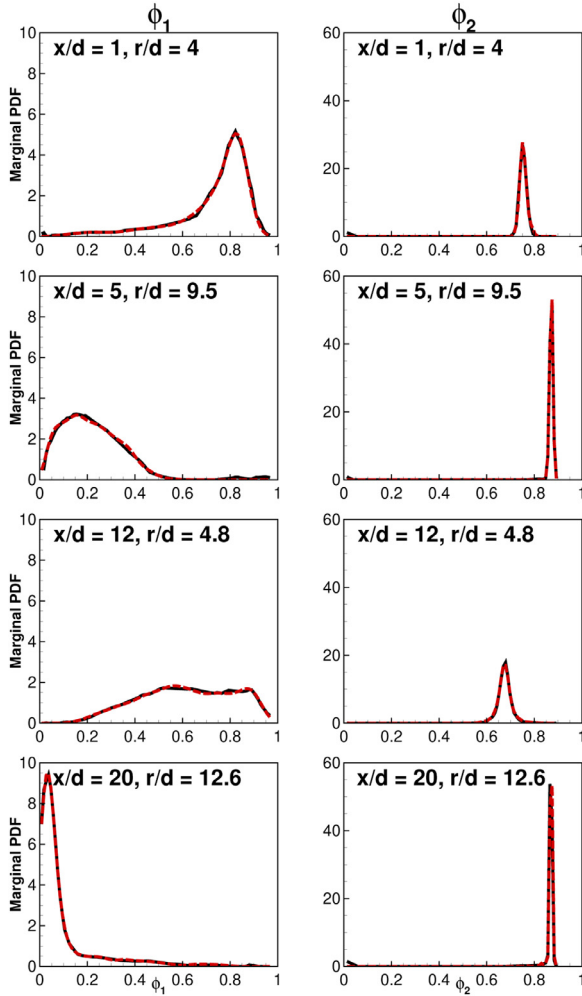


Fig. 8. Comparisons of marginal PDFs of ϕ_1 and ϕ_2 between individual KDE profiles at different spatial positions and the generalized DeepONet network parameterized with the unconditional means of ϕ_1 and ϕ_2 . Solid black: DeepONet network, red dashed: KDE. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.4. Parameterized marginal PDFs

In this section, we present results of the DeepONet-based evaluation of general PDFs. The network is trained in TensorFlow [53] using the Sydney flame experimental data. All the hidden layers use a $\tanh$ activation function, except for the output layer, which is linear. The training data set contains data related to marginal PDFs corresponding to only 50% of the spatial locations and these are randomly selected. No data augmentation is implemented. Additionally, a sampling strategy is used to maintain the balance of low and high probabilities in the training set. The total number of parameters (weights and bias values) is approximately 25,000. A mean squared error (MSE) loss is used to update the network weights and the training set is split such that 10% is reserved for validation and the remaining data is split almost evenly between training and testing. The network weights are tuned for about 1000 epochs and the training is stopped when the mean squared error loss on the validation set goes below 5×10^{-7} . However, to avoid over-fitting during training, early stopping is implemented, such that training stops when the validation loss does not improve for 10 consecutive epochs. We have also used the default Adam optimizer. The entire training procedure requires about 10 minutes using a NVIDIA Tesla V-100 GPU.

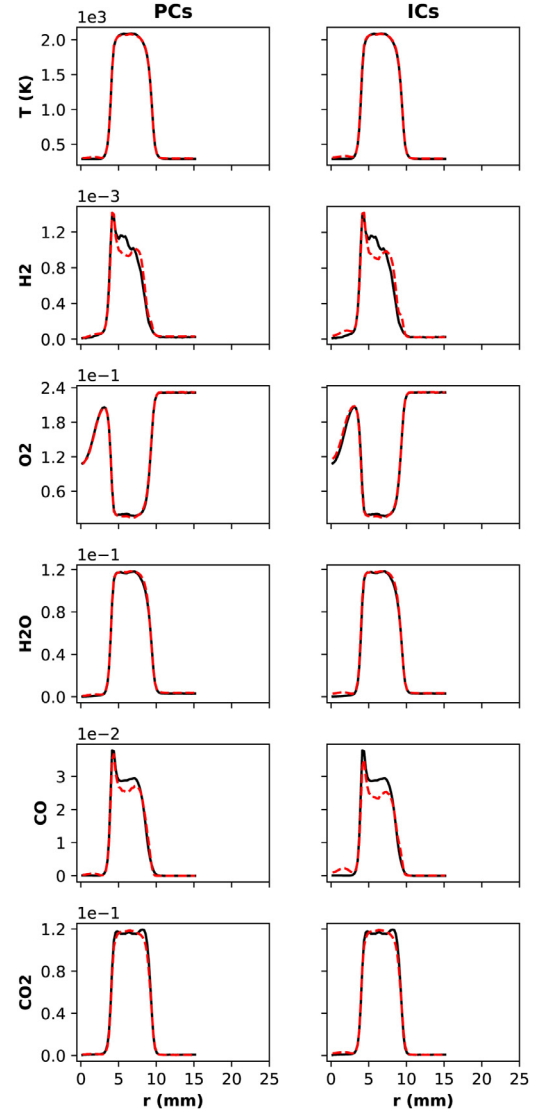


Fig. 9. Favre-averaged radial profiles based on PCs statistical independence of temperature, H_2 , O_2 , H_2O , CO , and CO_2 mass fractions at $x/d = 1$. Solid black line: experimental data; Red dashed line: PCA-ML reconstruction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the present study, the analysis is carried out on the marginal PDFs of the 2 PCs, which define the low-dimensional manifold of the flame data. This is consistent with our attempt to implement an approach that can be applied to as much sparse data as possible.

Figure 8 shows comparisons of different marginal PDFs for ϕ_1 (left column) and ϕ_2 (right column) at 4 different downstream x/d and radial positions r/d within the flame. Here x is the downstream distance from the burner inlet, r is the radial distance from the burner axis and d is the fuel jet diameter. The DeepONet predictions are compared with the local approximations of the marginal PDFs using KDE. The figure shows that the evolution of the mixture from a primarily premixed combustion regime (at $x/d = 1$) to a transitional combustion mode (at $x/d = 5$ and 10) and a non-premixed combustion mode further downstream is characterized by different marginal PDF shapes. These shapes are adequately captured by the more generalized model that adopts as inputs their parameterization in terms of the unconditional means for the 2 PCs. The two sets of profiles are indistinguishable at the different flame positions.

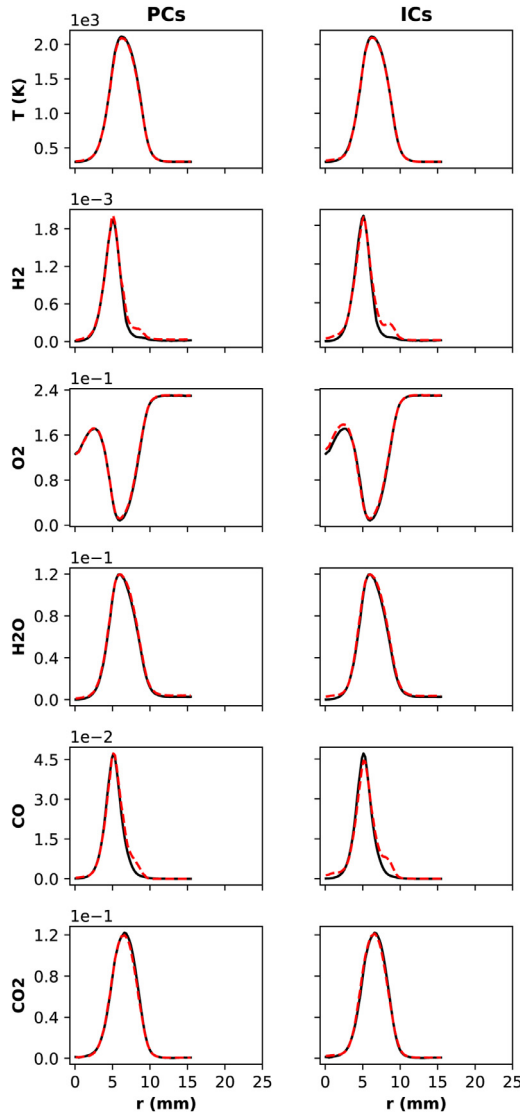


Fig. 10. Favre-averaged radial profiles based on PCs statistical independence of temperature, H_2 , O_2 , H_2O , CO , and CO_2 mass fractions at $x/d = 5$. Solid black line: experimental data; Red dashed line: PCA-ML reconstruction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

At this stage, it is noteworthy to state that KDE along with interpolation functions between KDEs at different spatial positions could have been adopted as an alternative to the present approach employing DeepONet. However, the implementation of DeepONet presents several advantages as compared to a KDE-interpolation approach. First, the DeepONet is memory efficient. It encapsulates a wide range of marginal PDF shapes with a network containing only 25,000 trainable weights. This approach requires a much lower memory than keeping the parameters of the KDEs associated with different spatial positions. Second, with the current implementation, the evaluation of DeepONet on the trained data is approximately 3 to 5 times faster than the corresponding KDE-based evaluations. Further speed up may be achieved with neural network parameter optimization using various tools available as open source such as AutoML [54]; however, such optimizations have not been investigated as part of this work.

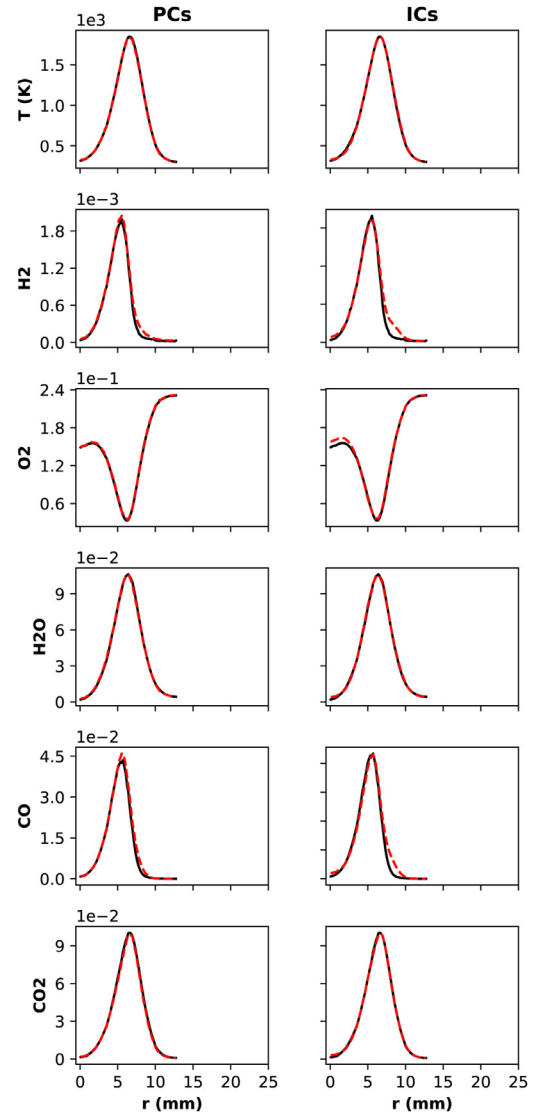


Fig. 11. Favre-averaged radial profiles based on ICs statistical independence of temperature, H_2 , O_2 , H_2O , CO , and CO_2 mass fractions at $x/d = 10$. Solid black line: experimental data; Red dashed line: PCA-ML reconstruction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.5. The reconstruction of unconditional means of thermo-chemical scalars

Next, we present comparisons of Favre-averaged radial profiles of selected thermo-chemical scalars at different downstream distances. These profiles are generated from direct calculations of the unconditional means from the data as well as the profiles reconstructed using Eq. (4) based on the conditional means, $\langle \psi_k | \phi \rangle$, and the joint PCs or ICs PDFs. Again, we consider comparisons for the Sydney flame FJ200-5GP-Lr75-57. Given the similarity between the generalized PDF and the local reconstructed PDFs using KDE, no additional comparisons are carried out using these 2 approaches. Instead, we report comparisons of radial profiles of unconditional means of the measured quantities based on their direct measurements from the data as well as by using Eq. (4) and the models for the conditional means and the reconstructed PDFs using KDE.

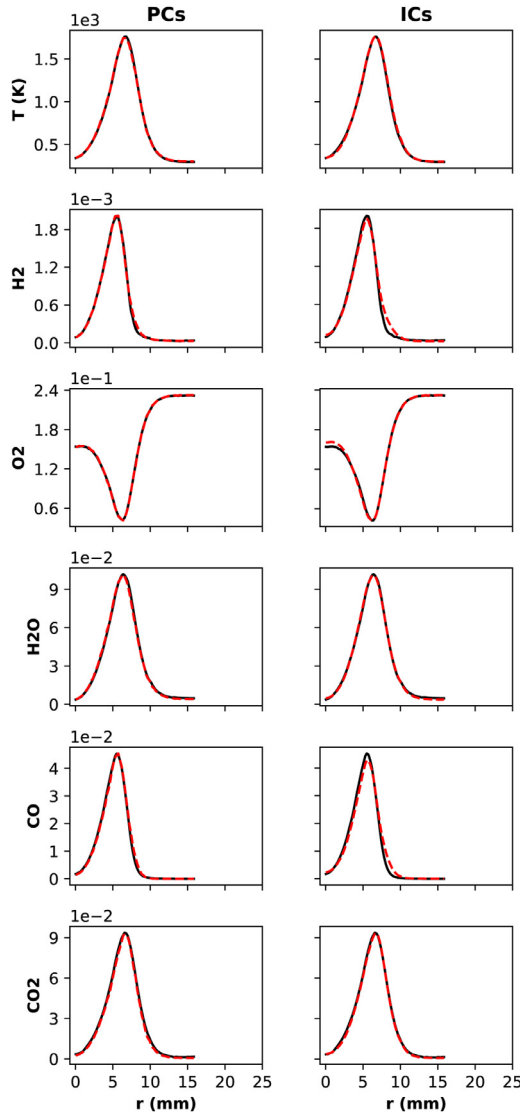


Fig. 12. Favre-averaged radial profiles based on PCs statistical independence of temperature, H_2 , O_2 , H_2O , CO , and CO_2 mass fractions at $x/d = 12$. Solid black line: experimental data; Red dashed line: PCA-ML reconstruction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figures 9–14 compare the radial profiles of Favre-averaged temperature, H_2 , O_2 , H_2O , CO , and CO_2 mass fractions at $x/d = 1, 5, 12, 15, 20$ and 30 , respectively, based on the available data and using the assumption of statistical independence between ϕ_1 and ϕ_2 .

The figures show an excellent reconstruction of the unconditional means of temperature, O_2 , H_2O and CO_2 at all shown downstream distances. The maximum percentage error for temperature, O_2 , H_2O , CH_4 and CO_2 mass fractions ranges from approximately 1% to 4, 3, 6 and 13 and 7%, respectively. The figures also show some discrepancies in the predictions of CO and H_2 unconditional means at $x/d = 1$ and 5 . The maximum percent error for H_2 and CO mass fractions ranges from 5% and 6% to 15% and 10%, respectively. The predictions for these 2 quantities is significantly improved further downstream. Much of the difference, we believe, can be attributed to the choice of the number of retained PCs. During PCA, we retained a number of PCs that represents 94% of the data variance. Including more PCs, such as the third PC for CO and

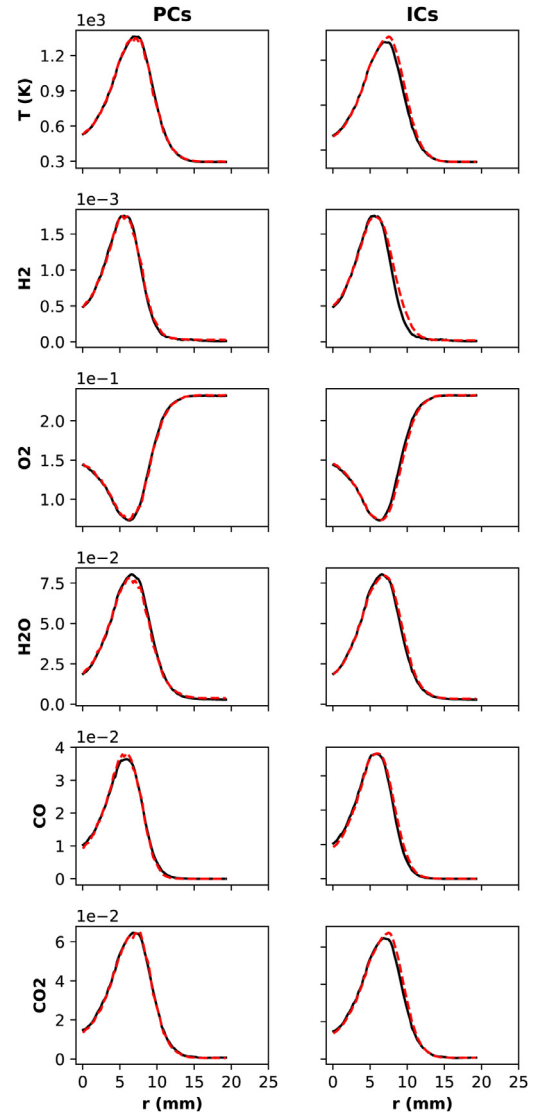


Fig. 13. Favre-averaged radial profiles based on ICs statistical independence of temperature, H_2 , O_2 , H_2O , CO , and CO_2 mass fractions at $x/d = 20$. Solid black line: experimental data; Red dashed line: PCA-ML reconstruction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the seventh PC for H_2 , can improve these predictions with an increased cost for the conditional means and the joint and marginal PDFs.

The figures show that there is no real advantage to the use of ICs as parameters for the composition space despite the fact that we have established a broader range of conditions for statistical independence with ICA. However, this enhanced statistical independence did not alter the evaluation of the convolution relation (4). In fact, while ICA appears to perform better for some of the radial profiles, it generally has a poorer performance than PCA-based statistics for H_2 and CO . This may appear to be counter-intuitive. However, the discrepancy between PCA and ICA-based statistics can be related to the optimization algorithm for determining the ICs. This algorithm yields a single rotation matrix to convert from PCs to ICs and is based on all the available data. This rotation may not be optimum for all positions considered.

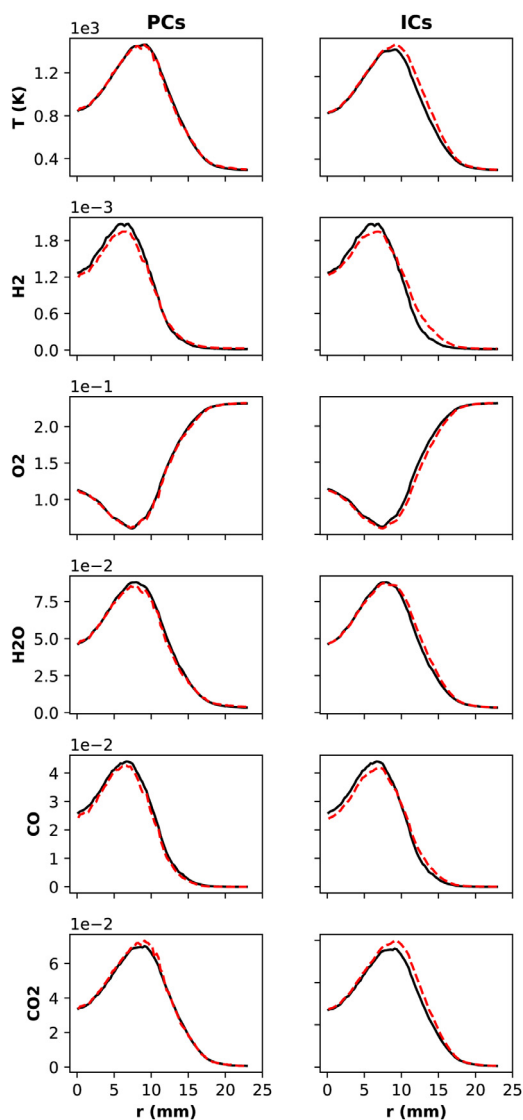


Fig. 14. Favre-averaged radial profiles based on PCs statistical independence of temperature, H_2 , O_2 , H_2O , CO , and CO_2 mass fractions at $x/d = 30$. Solid black line: experimental data; Red dashed line: PCA-ML reconstruction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions and implications to turbulent combustion modeling

In this paper, we investigate the key ingredients needed to construct flamelet-like models from PDF-like model-based simulations. A hybrid approach based on the implementation of flamelet-like model simulations can be used to accelerate simulations based on PDF-like models.

One of the key ingredients is the determination of low-dimensional manifolds from the computational data using PCA, which generates a significant reduction in the dimensionality of the composition space. We also investigate a rotation of the PCs using ICA to examine the assumption of statistical independence of the low-dimensional manifold parameters. We show that both sets of PCs and ICs parameters are reasonably independent, at least within the reaction region where statistics matter the most. Invariably, the assumption of statistical independence, while not applicable everywhere in the flame, is most important where the PDF magnitude is sufficiently high to impact statistics. Regardless, this

independence enables a simplified expression for the joint PCs/ICs PDFs in terms of their marginal PDFs.

The marginal PDFs are constructed at each spatial position from the Sydney flame data [29,30] using the KDE approach [27]; then, these regressions for the PCs' marginal PDFs are combined to be parameterized in terms of the unconditional means of the PCs using the novel DeepONet architecture. The *a priori* results show that the KDE-DeepONet approach for the PCs' marginal PDFs is able to reconstruct these distributions under different conditions and shapes. The use of machine learning also overcomes the curse of dimensionality by retaining only the networks' trainable parameters. Therefore, this potential saving in memory enables further software and hardware acceleration within adaptive schemes for PDF-like model simulations.

The above results establish a clear motivation to explore the different elements of this work within a framework that can be combined with PDF-like models' simulations to obtain efficient computations while maintaining the key advantages of these models.

Declaration of Competing Interest

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

Acknowledgments

This work is supported by the [National Science Foundation](#) under grant no. 1941430. The authors would like to thank Prof. As-saad Masri for sharing the data on the Sydney flames experiments.

References

- [1] S. Pope, Small scales, many species and the manifold challenges of turbulent combustion, *Proc. Combust. Inst.* 34 (2013) 1–31.
- [2] N. Peters, Laminar flamelet concepts in turbulent combustion, *Symp. (Int.) Combust.* 21 (1988) 1231–1250.
- [3] N. Peters, *Turbulent Combustion*, Cambridge Monographs on Mechanics, first ed., Cambridge University Press, Cambridge UK, 2000.
- [4] S. Pope, Pdf methods for turbulent reacting flows, *Prog. Energy Combust. Sci.* 11 (1985) 119–192.
- [5] A. Klimenko, Multicomponent diffusion of various mixtures in turbulent flow, *Fluid Dyn.* 25 (1990) 327–331.
- [6] R. Bilger, Conditional moment closure for turbulent reacting flow, *Phys. Fluids A* 5 (1993) 436–444.
- [7] V. Chakravarthy, S. Menon, Subgrid modeling of turbulent premixed flames in the flamelet regime, *Flow Turbul. Combust.* 65 (2000) 133–161.
- [8] V. Sankaran, S. Menon, Subgrid combustion modeling of 3-d premixed flames in the thin-reaction-zone regime, *Proc. Combust. Inst.* 30 (2005) 575–582.
- [9] H. El-Arag, S. Menon, Large eddy simulation of bluff-body stabilized swirling non-premixed flames, *Proc. Combust. Inst.* 31 (2007) 1747–1754.
- [10] S. Undapalli, S. Srinivasan, S. Menon, LES of premixed and non-premixed combustion in a stagnation point reverse flow combustor, *Proc. Combust. Inst.* 32 (2009) 1537–1544.
- [11] S. Cao, T. Echehki, A low-dimensional stochastic closure model for combustion large-Eddy simulation, *J. Turbul.* 9 (2008) 1–35.
- [12] J. Park, T. Echehki, LES-ODT study of turbulent premixed interacting flames, *Combust. Flame* 159 (2012) 609–620.
- [13] S. Ben Rejeb, T. Echehki, Thermal radiation modeling using the LES-ODT framework for turbulent combustion flows, *Int. J. Heat Mass Transf.* 104 (2017) 1300–1316.
- [14] Y. Fu, T. Echehki, Upscaling and downscaling approaches in LES-ODT for turbulent combustion flows, *Int. J. Multiscale Sci. Eng.* 16 (2018) 45–76.
- [15] A.F. HOFFIE, T. Echehki, A coupled LES-ODT model for spatially-developing turbulent reacting shear layers, *Int. J. Heat Mass Transf.* 127 (2018) 458–473.
- [16] J. Miles, T. Echehki, A one-dimensional turbulence-based closure model for combustion LES, *Combust. Sci. Technol.* 192 (2020) 78–111.
- [17] C. Lacey, A. Novoselov, M. Mueller, In-situ adaptive methods: enabling computationally efficient simulations of complex turbulent reacting flows, *Proc. Combust. Inst.* 38 (2021) 2673–2680.
- [18] H. Wu, Y.C. See, Q. Wang, M. Ihme, A pareto-efficient combustion framework with submodel assignment for predicting complex flame configurations, *Combust. Flame* 162 (2015) 4208–4230.
- [19] W.T. Chung, A.A. Mishra, N. Perakis, M. Ihme, Data-assisted combustion simulations with dynamic submodel assignment using random forests, *Combust. Flame* 227 (2021) 172–185.

- [20] S. Pope, Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation, *Combust. Sci. Tech.* 1 (1997) 41–63.
- [21] I. Jolliffe, *Principal Component Analysis*, second ed., Springer-Verlag, New York, 2002.
- [22] R. Ranade, T. Echekki, A framework for data-based turbulent combustion closure: A priori validation, *Combust. Flame* 206 (2019) 490–505.
- [23] R. Ranade, T. Echekki, A framework for data-based turbulent combustion closure: A posteriori validation, *Combust. Flame* 210 (2019) 279–291.
- [24] J. Sutherland, A. Parente, Combustion modeling using principal component analysis, *Proc. Combust. Inst.* 32 (2009) 1563–1570.
- [25] O. Owoyele, T. Echekki, Toward computationally efficient combustion DNS with complex fuels via principal component transport, *Combust. Theo. Model.* 21 (2017) 770–798.
- [26] R. Malik, P. Vega, A. Coussement, A. Parente, Combustion modeling using Principal Component Analysis: a posteriori validation on Sandia flames D, E and F, *Proc. Combust. Inst.* 38 (2021) 2635–2643.
- [27] A. Bowman, A. Azzalini, *Applied Smoothing Techniques for Data Analysis: the Kernel Approach With S-Plus Illustrations*, Oxford University Press, Oxford, 1997.
- [28] L. Lu, P. Jin, Z. Zhang, G. Karniadakis, Learning nonlinear operators with Deep ONet based on the universal approximation theorem of operators, *Nature Mach. Intel.* 3 (2021) 218–229.
- [29] S. Meares, A. Masri, A modified piloted burner for stabilizing turbulent flames of inhomogeneous mixtures, *Combust. Flame* 161 (2014) 484–495.
- [30] S. Meares, V. Prasad, G. Magnotti, R. Barlow, A. Masri, Stabilization of piloted turbulent flames with inhomogeneous inlets, *Proc. Combust. Inst.* 35 (2015) 1477–1484.
- [31] R. Barlow, S. Meares, G. Magnotti, H. Cutcher, A. Masri, Local extinction and near-field structure in piloted turbulent CH_4 < /UPPER – CASE > /air jet flames with inhomogeneous inlets, *Combust. Flame* 162 (2015) 3516–3540.
- [32] S. Vajda, P. Valko, T. Turányi, Principal component analysis of kinetic models, *Int. J. Chem. Kinet.* 17 (2006) 55–81.
- [33] S. Danby, T. Echekki, Proper orthogonal decomposition analysis of autoignition simulation data of nonhomogeneous hydrogen-air mixtures, *Combust. Flame* 144 (2006) 126–138.
- [34] H. Mirgolbabaie, T. Echekki, A novel principal component analysis-based acceleration scheme for LES-ODT: an a priori study, *Combust. Flame* 160 (2013) 898–908.
- [35] H. Mirgolbabaie, T. Echekki, N. Smaoui, A nonlinear principal component analysis approach for turbulent combustion composition space, *Int. J. Hydrogen Energy* 39 (2014) 4622–4633.
- [36] H. Mirgolbabaie, T. Echekki, Nonlinear reduction of combustion composition space with kernel principal component analysis, *Combust. Flame* 161 (2014) 118–126.
- [37] H. Mirgolbabaie, T. Echekki, The reconstruction of thermo-chemical scalars in combustion from a reduced set of their principal components, *Combust. Flame* 162 (2015) 1650–1652.
- [38] T. Echekki, H. Mirgolbabaie, Principal component transport in turbulent combustion: a posteriori analysis, *Combust. Flame* 162 (2015) 1919–1933.
- [39] R. Barlow, J. Frank, Effects of turbulence on species mass fractions in methane/air jet flames, *Proc. Combust. Inst.* 27 (1998) 1087–1095.
- [40] P. Comon, Independent component analysis, a new concept, *Signal Process.* 36 (1994) 287–314.
- [41] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Netw.* 13 (2000) 411–430.
- [42] A.H.J. Karhunen, E. Oja, *Independent Component Analysis*, first ed., John Wiley & Sons, New York, 2001.
- [43] M. Bartlett, J. Movellan, T. Sejnowski, Face recognition by independent component analysis, *IEEE Trans. Neural Netw.* 13 (2002) 1450–1464.
- [44] B. Draper, K. Baek, M. Bartlett, J. Beveridge, Recognizing faces with PCA and ICA, *Comput. Vis. Image Understand.* 91 (2003) 115–137.
- [45] O. Owoyele, P. Kundu, M.M. Ameen, T. Echekki, S. Som, Application of deep artificial neural networks to multi-dimensional flamelet libraries and spray flames, *Int. J. Engine Res.* 21 (1, SI) (2020) 151–168.
- [46] R. Ranade, G. Li, S. Li, T. Echekki, An efficient machine-learning approach for PDF tabulation in turbulent combustion closure, *Combust. Sci. Technol.* 193 (7) (2021) 1258–1277.
- [47] G. Goldin, S. Menon, A scalar PDF construction model for turbulent non-premixed combustion, *Combust. Sci. Technol.* 125 (1997) 47–72.
- [48] G. Goldin, S. Menon, A comparison of scalar PDF turbulent combustion models, *Combust. Flame* 113 (1998) 442–453.
- [49] G. Goldin, A priori investigation of the constructed PDF model, *Proc. Combust. Inst.* 30 (2005) 785–792.
- [50] V. Sankaran, T.G. Drozda, J.C. Oefelein, A tabulated closure for turbulent non-premixed combustion based on the linear eddy model, *Proc. Combust. Inst.* 32 (2009) 1571–1578.
- [51] M. de Frahan, S. Yellapantula, R. King, M. Day, R. Grout, Deep learning for presumed probability density function models, *Combust. Flame* 208 (2019) 436–450.
- [52] R.B. Nelson, *An Introduction to Copulas*, second ed., Springer-Verlag, New York, 2006.
- [53] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, *Tensorflow: A system for large-scale machine learning*, 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), USENIX (2016), pp. 256–283.
- [54] *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff, J. Vanschoren (Eds.), Springer, 2018. In press, available at <http://automl.org/book>.