



An a priori analysis on principal component analysis based conditional source-term estimation model for Sandia jet flames

N. Sekularac^a, W.K. Bushe^b, X.H. Fang^{c,d,*}

^a CERFACS, 42 Avenue Gaspard Coriolis, 31057, Toulouse, France

^b Department of Mechanical Engineering, University of British Columbia, V6T 2G9, Vancouver, Canada

^c Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, UK

^d Department of Mechanical & Manufacturing Engineering, Schulich School of Engineering, University of Calgary, T2L 1Y6, Calgary, Canada

ARTICLE INFO

Keywords:

Numerical simulation
Turbulent combustion
Principal component analysis
Conditional moment method

ABSTRACT

Data from all spatial locations and two turbulent flames in the Sandia/TUD database are used to explore the feasibility of adopting principal components (PC) as conditional variables in the conditional source-term estimation (CSE) model. Principal component analysis (PCA) is applied to both Flame C and F to generate the new set of controlling variables, PC-scores. Two PCA scaling methods have been adopted, namely Pareto and Auto-scaling (AS). Regardless of the scaling option selected and the flame investigated, it was found that a single principal component score (PC1-score) correlated with temperature accounts for the largest amount of variance. As such, the conditional space fluctuations and normalized RMS of both flames' reactive scalars around PC1-Pareto and PC1-AS are examined and compared against the ones obtained with the mixture fraction, Z . The results indicate that both PC1-scores are not able to accurately quantify the thermo-chemical state-space of Flame C compared to mixture fraction, in particular for the fuel and all the intermediate species. Interestingly, using a single principal component score for Flame F significantly improved the conditional fluctuations, suggesting that a single PC-score well correlated with temperature can more effectively reduce the spatial gradient and Reynolds numbers effects than mixture fraction. The results are further depicted by comparing the trends obtained for Flame F with two-condition conditional averages around Z and four different progress variable definitions, c . While doubly conditioning enabled to detach the conditional averages of all the scalars from the physical domain, the results obtained with PC1-Pareto and PC1-AS were found to not deviate by much (excluding the mass fractions of CH_4). This leads to believe that a conditional moment closure-based model such as CSE, coupled with PCA, can perhaps recover with satisfactory levels of approximation the thermo-chemical state-space of Flame F and separate the conditional manifold from the real domain.

1. Introduction

The most engineering-relevant cases involving complex reacting flows remain out of reach for Direct Numerical Simulations (DNS) due to the computational time required to resolve the vast range of scales present in both turbulence and chemistry. As such, reduced-order models are needed to mitigate the costs of fully resolving turbulent reactive systems. In practice, turbulence is generally modeled using either Reynolds-Averaged Navier–Stokes (RANS) or Large Eddy Simulation (LES). To lower the computational costs for solving transport equations for all chemical species, one can parameterize the thermo-chemical state-space using a reduced set of scalars (found to be closely

related to the chemical state) to construct a reduced-dimension composition space. As such, only a portion of selected variables need to be transported in the Computational Fluid Dynamics (CFD) solver and used as inputs for the low-dimensional manifold look-up process.

In recent years, data-driven techniques have gained considerable attention in combustion applications for building low-dimensional manifolds while preserving an adequate representation of the thermo-chemical state [1]. Among many others, principal component analysis (PCA) can be used to describe a system using a reduced number of optimal scalars identified in the directions of maximal data variance, namely principal components (PCs). Projecting the state-space on those PCs gives the PC-scores, and transporting only a subset of those scores in a numerical framework was shown to improve computational costs.

* Corresponding author at: Department of Mechanical & Manufacturing Engineering, Schulich School of Engineering, University of Calgary, T2L 1Y6, Calgary, Canada.

E-mail address: xiaohang.fang@ucalgary.ca (X.H. Fang).

¹ Also commonly referred to as “turbulence-chemistry interaction effects”.

Sutherland & Parente [2] and Parente et al. [3] derived the transport equations for the PC-scores, from which originated the so-called PC-score combustion model. The model was further extended by adding, in conjunction with PCA, a non-linear regression technique to avoid increasing the number of components that need to be retained to counter the non-linearity of the source terms that results in errors [4]. Malik et al. [5] consolidated the modeling capabilities of the PC-score approach to more complex fuels involving methane and propane using the same perfectly stirred reactor (PSR) formalism as in [4]. The PC-score has been examined under various configurations and conditions [6–9] with satisfactory results compared to the benchmarks used. It is only with the recent work of Malik et al. [10] that the potentiality of adopting the PC-score approach in the framework of turbulent combustion modeling with a fully three-dimensional LES was assessed. The model results were compared with the experimental measurements of the Sandia/TUD jet flames [11]. The obtained trends indicated that the model could predict the entire set of flames sufficiently well, particularly flames E and F, which are known to exhibit complex structures. However, applying such an approach in the context of numerical modeling still requires further research, particularly when incorporating filter-chemistry interaction effects.¹

Biglari & Sutherland [12] showed that the functional dependency between the thermo-chemical variables and the PC-scores is independent of the filter width, suggesting that no explicit closure model is required for the source quantities. While this conclusion was deduced from two datasets that exhibit relatively small turbulence intensity, further research might be needed to assess the importance of an adequate subgrid closure at higher levels of turbulence fluctuation. Recently, Malik et al. [13] used a presumed joint probability density function (PDF) closure for the source terms in the PC-scores transport equations. The model was coupled with a subgrid closure to improve the Cabra et al. flame's [14] inner structure, mixing zone, lift-off height, and mean flow field predictions. However, the model exhibited an extended ignition region with temperatures in the post-ignition zone that deviated ~400 K from measurements. This could be attributed: (i) to the assumption of statistical independence applied between the two PC-scores which are orthogonally uncorrelated but might still exhibit strong statistical dependence, and (ii) to the beta-PDF that was used for both PC-scores, which might not adequately describe the statistical distribution of both PCs. This raises an essential aspect of the model, where presuming the shape of the PCs' joint PDF is not trivial considering that principal components are often difficult to associate with control variables such as the mixture fraction or progress variables. This makes the physical interpretations not always straightforward, depending on the studied case. Facing that, one could avoid this concern by incorporating variance terms in the look-up table where the dimension of the latter can, however, be particularly restrictive (i.e., four dimensions assuming that both PC-scores' mean and variance are accounted for) and where additional interpolation might be involved having a significant impact on the memory usage and CPU time.

Among different reduced-order combustion modeling approaches, conditional moment closure-based models (CMC) assume that fluctuations in scalar quantities can be associated with the fluctuations around values of conditioning variables [15]. Assuming small fluctuations around conditional averages, filter-chemical source terms can be closed using a Taylor series expansion with a first-order truncation. Originating from CMC, conditional source-term estimation (CSE) avoids solving conditionally-averaged transport equations by inversion of the integral functions, significantly reducing the computational cost while maintaining the accuracy [16,17]. Within the CSE framework, inverting all the species to retrieve the conditionally-averaged quantities would be computationally prohibitive. Several techniques based on chemistry tabulation have been incorporated in CSE and tested over the past years to lower the number of required inversions, of which the most used techniques referred to the trajectory generated low-dimensional manifolds (TGLDM) [18–22]. However, studies [22,23] have suggested

that the TGLDM approach might not be suitable for heavy fuels such as ethanol or diesel and gasoline surrogates. As such, other methods derived from flamelets have also been tested with CSE, such as the well-known flamelet generated manifold (FGM), which enabled to improve the tendency [24]. Filter-chemistry interaction effects are incorporated in the form of a presumed PDF. The PDF of the conditioning variable selected predicts the complete statistical information about the reactive scalars, making it possible to predict any filtered moment of interest. While CSE with a single conditioning variable has been found suitable for non-premixed and premixed flames [23–27], for partially-premixed or stratified combustion operating under high turbulence intensity (i.e., $I = \sqrt{u'^2}/\bar{U} \sim 10\%$, where $\sqrt{u'^2}$ is the root mean square value of the fluctuating velocity component and \bar{U} the local mean velocity), a unique conditioning variable is not sufficient [28]. Consequently, doubly CSE (DCSE) with often the mixture fraction and the progress variable as conditioning variables have been developed to simulate partially-premixed flames, lifted flames, and spray flames [29,30]. The CSE and DCSE concepts have been examined in *a-priori* analysis in DNS for high-pressure conditions [31]. The results highlighted that DCSE and double conditioning are likely needed for cases closer to real practical combustion applications. However, the inversion process used to calculate the conditional scalar values coupled with the current presumed shapes of the joint PDF are found to be the two most significant sources of error associated with DCSE [32]. The assumption of using a beta-PDF for the reaction progress variable was shown to be a particularly poor choice, as is the assumption that the conditioning variables are statistically independent.

The incentive of the current study arises from both the pros and cons of the above-mentioned two approaches for reduced order modeling. And the main hypothesis is whether we can use other sets of conditioning variables in CSE to separate the conditional manifold from the real domain and whether PC-scores would be the appropriate choice for that. If the first few PC-scores reproduce most of the variation while containing the major dynamics of a system, one can suppose that the number of conditioning scalars needed in CSE with PC-based variables can be reduced while preserving a sufficiently accurate representation of the thermo-chemical state. As such, a single PC-score (that has the highest correlation with the reactive quantities) may be employed as a control variable in CSE and avoid the selection of a set of conditioning variables, increasing the manifold's dimensionality. It is expected that this approach would result in smaller discrepancies for the unconditional averages compared to previous implementations of CSE using either the mixture fraction or the progress variable as single conditioning variables, in particular for highly turbulent combustion cases. The complementarity of the two models can be further extended considering that CSE can be used as a subgrid closure model in PCA, and as such, account for the effect of small turbulent scales on combustion while avoiding incorporating mean and variance terms in the look-up process, and potentially increasing the dimensionality of the problem.

In this study, the conditional space fluctuations of Sandia/TUD jet flames C and F obtained with single and doubly conditional averages are examined and compared with the new set of variables obtained with the PCA model, where all one-point, one-time measurements are included. First, the PCA and CSE approaches are presented, with particular attention brought to the possible CSE-PCA routine structure. The methodology employed for assessing the feasibility of adopting a single PC-score in CSE is then introduced, followed by the properties and working conditions of the burner. Finally, the obtained results are presented and discussed, with physical insights provided based on the observations.

2. Theory

2.1. Principal component analysis

The mathematical approach to compute the principal components of a given dataset X ($n \times Q$) reduces to an eigenvalue decomposition problem, where rows n represent individual measurements of Q variables. Suppose X has been appropriately standardized (i.e., centered and scaled), PCA projects all Q variables onto a rotated basis obtained from the eigenvalue decomposition of the covariance matrix S ($Q \times Q$) as

$$S = \frac{1}{n-1} X^T X = A \Lambda A^T \quad (1)$$

where A is the ($Q \times Q$) matrix whose columns are the eigenvectors of S , and Λ is a ($Q \times Q$) diagonal matrix containing the eigenvalues of S . Following the details of the PCA reduction provided in [2,33], PC-scores Ψ are obtained as

$$\Psi = XA \quad (2)$$

where Ψ is an ($n \times Q$) matrix. Each column of A describes the weight between the Q variables of X and the corresponding principal component. The dimensionality reduction is undertaken by truncating A , such that only the first q PCs that account for the maximum variance are retained, with $q < Q$. The original dataset X is retrieved as

$$X \approx X_q = \Psi_q A_q^T \quad (3)$$

where X_q is the approximation of X based on the first q eigenvectors of A , and Ψ_q is the ($n \times q$) matrix of the principal component scores. A detailed mathematical formulation of PCA can be found in the literature [34]. The notations Ψ_1 and Ψ_2 will be used hereafter to designate the PC1-scores and PC2-scores, respectively.

2.2. Conditional source-term estimation

In CSE, the concept of conditional averages is used to determine the unconditional Favre-averaged quantities. The conditional averages are calculated at a specific sampling space value η in the conditioning variable domain. Following the same first-order hypothesis as CMC for closing chemical source terms, and therefore neglecting the spatial conditional fluctuations, the mean conditional chemical source-term $\langle \dot{\omega}_k | \eta \rangle$ is obtained as

$$\langle \dot{\omega}_k | \eta \rangle \approx \dot{\omega}_k(\langle Y_k | \eta \rangle, \langle T | \eta \rangle, \langle \rho | \eta \rangle) \quad (4)$$

where $\langle T | \eta \rangle$ is the conditional temperature, $\langle Y_k | \eta \rangle$ and $\langle \rho | \eta \rangle$ are the conditional mass fractions of species k and the conditional density, respectively. The conditionally-averaged scalars are obtained by inverting the following Fredholm integral equation of the first kind

$$\tilde{Y}_k(x_j, t) = \int_{\eta_1}^{\eta_2} \langle Y_k | \eta \rangle \tilde{P}(\eta, x_j, t) d\eta \quad (5)$$

where x_j is the spatial coordinate, t the simulation time, and η_1 and η_2 being the lower and upper bounds of the sampling space integral, respectively. The unconditional mean chemical source terms are obtained by integrating the conditional value over the averaged PDF of the conditioning variable η , such as

$$\bar{\dot{\omega}}_k(x_j, t) = \int_{\eta_1}^{\eta_2} \langle \dot{\omega}_k | \eta \rangle \bar{P}(\eta, x_j, t) d\eta \quad (6)$$

The inversion process used to calculate the conditional averages will not be elaborated on here, as it is believed to be out of the scope of this study. Further information can be found in the recent work of Mahdipour & Salehi [27] which proposes a new alternative approach to calculate the conditional averages and bypass the challenges of the conventional inversion method.

Considering the conclusions drawn from previous PCA studies, where two principal components seem to account for more than 90% of

the data variance [3,13,35], a low-dimensional manifold parametrized by Ψ_1 and Ψ_2 can be constructed containing all the thermo-chemical quantities. As such, only two PC-scores can be transported into the CFD solver and used to look-up in the low-dimensional manifold. Based on the work conducted in [2,3], the PC-score transport equations derived from the general species transport equation read

$$\frac{\partial \tilde{\rho} \tilde{\Psi}_i}{\partial t} + \frac{\partial}{\partial x_j} (\tilde{\rho} \tilde{u}_j \tilde{\Psi}_i) = \frac{\partial}{\partial x_j} \left(\left(D_{\Psi_i} + \frac{\mu_t}{Sc_i} \right) \frac{\partial \tilde{\Psi}_i}{\partial x_j} \right) + \bar{\dot{\omega}}_{\Psi_i} \quad \text{with } i = 1, 2 \quad (7)$$

where $D_{\Psi_i} = A_{i,q}^T D_X A_{i,q}$ with D_X the diagonal matrix of diffusion coefficients for species [36,37], μ_t and Sc_i denote respectively the turbulent viscosity and the turbulent Schmidt number, and $\dot{\omega}_{\Psi_i} = \dot{\omega}_X A_{i,q}^T$ with $\dot{\omega}_X$ corresponding to the source terms of the reactive scalars of X . Assuming that a single principal component score Ψ_1 accounts for the largest amount of data variance, it can be retained and used as a conditioning variable. The resulting unconditional PC1-score source terms can be calculated as

$$\bar{\dot{\omega}}_{\Psi_1}(x_j, t) = \int_{\eta_1}^{\eta_2} \langle \dot{\omega}_{\Psi_1} | \eta \rangle \bar{P}(\eta, x_j, t) d\eta \quad (8)$$

where η is, in this case, the sampling space variable of Ψ_1 . The Ψ_2 source terms can be retrieved using the same formalism as in Eq. (8). Fig. 1 illustrates a possible routine structure coupling CSE and PCA. The CFD code provides the mean and variance of the PC1-score to the PDF sub-model and the unconditional averages of the PC2-score. Assuming that the PC1-score is a reactive scalar, it is possible to retrieve the PC1-score variance balance equation analogous to the progress variable variance derived in [38], such that

$$\widetilde{\Psi_1^2} = \bar{\Psi_1^2} - \bar{\Psi}_1^2 \quad (9)$$

where $\bar{\Psi_1^2}$ is the square of the mean PC1-score, whereas $\widetilde{\Psi_1^2}$ is the mean PC1-score squared. The PDF constructs the kernel of the inverse problem for the inversion process in the CSE module for obtaining the conditionally-averaged Ψ_2 . The latter is used with $\tilde{\Psi}_1$ to look-up the conditionally-averaged source terms and mass fractions from the PCA manifold generated beforehand. The manifold can be obtained using the recently developed open-source Python package PCAfold 2.0 [39,40] enabling to generate, analyze and improve low-dimensional data manifolds. This tool uses a novel cost function that characterizes the quality of the manifold topology in conjunction with an artificial neural network (ANN) model to accurately project the data in PC-space. The scalars are then integrated using the PDF module to calculate the unconditional quantities. It should be noted that the unconditional mass fractions are not required in the proposed CFD solver configuration. Interestingly, previous studies [10,36,37] showed that one of the first PCs was often found to be highly correlated with the mixture fraction Z for non-premixed flames. This trend has been thoroughly validated for Sandia/TUD jet flames, suggesting that Z could perhaps replace a PC-score.

3. Methodology

3.1. Experimental setup

Experimental measurements of piloted CH_4/Air jet flames from the well-known Sandia/TUD burner are used in this study. Multiscalar data of four turbulent non-premixed flames (referred to as flames C, D, E, and F) are examined with: (i) increasing velocity in the main jet and pilot, and (ii) increasing probability of localized extinction, depicted in Table 1.

The burner, illustrated in Fig. 2, consists of an axisymmetric jet centered in an annulus pilot in which a heat source from premixed flames is used to stabilize the main jet. The main jet composition is set to 25% methane and 75% dry air (by volume), enabling to reduce the flame length and producing a more robust flame than pure

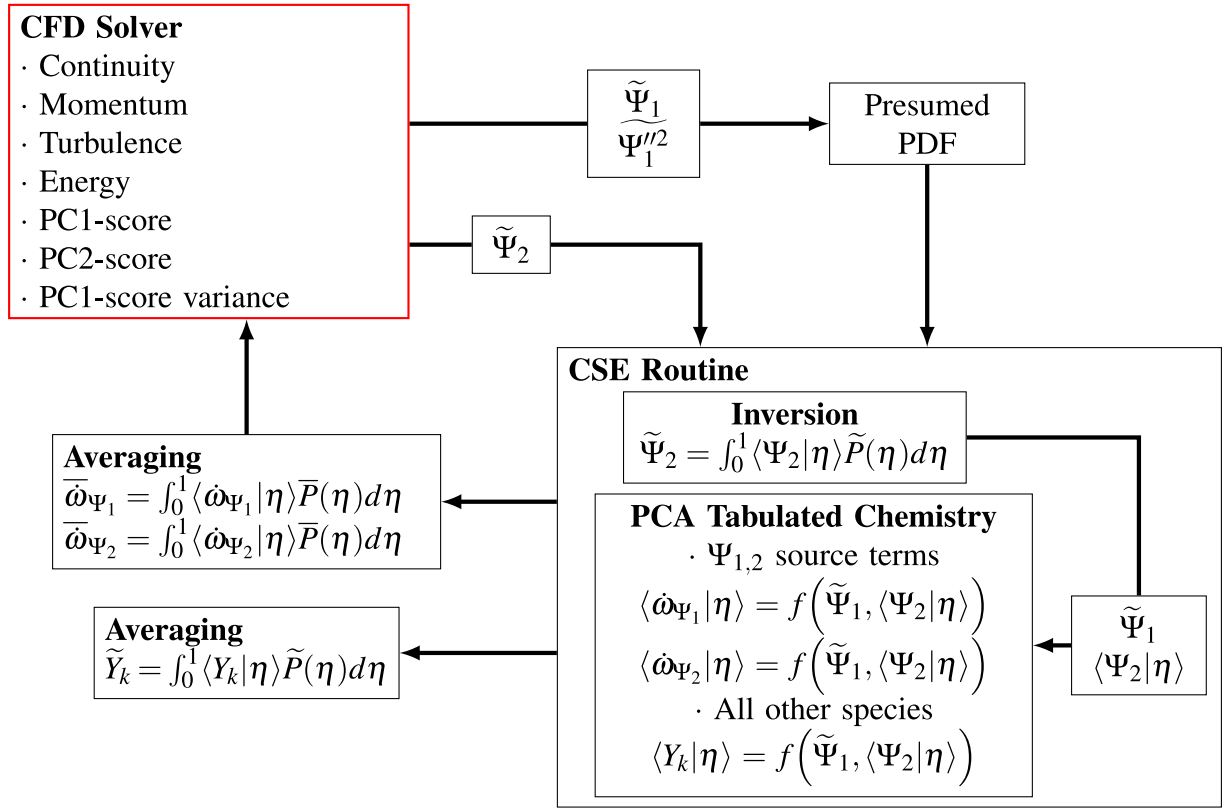


Fig. 1. Possible CSE-PCA routine structure. The labels Ψ_1 and Ψ_2 denote the PC1-scores and PC2-scores, respectively. The variable η is the sampling space variable of the PC1-score. The limits on the integrals assume that Ψ_1 is defined in the interval $[0, 1]$.

Table 1

Main flow parameters of Sandia flames C-F with $U_{j,b}$: the bulk velocity for the fuel jet, $U_{p,b}$: the bulk velocity for the pilot, T_p : the pilot temperature. In all cases, $\phi_p = 0.77$.

Flame	Re_j	$U_{j,b}$ [m/s]	$U_{p,b}$ [m/s]	T_p [K]	Local Extinction
C	~13400	29.7	6.8	~1920	None
D	~22400	49.6	11.4	~1880	Little
E	~33600	74.4	17.1	~1880	Moderate
F	~44800	99.2	22.8	~1860	Intense

methane or nitrogen-diluted methane [41–43]. As such, the flames can operate at reasonably high Reynolds numbers with little or no local extinction. The pilot is a lean mixture of C_2H_2 , H_2 , CO_2 , and air, with the same nominal enthalpy and equilibrium composition as CH_4 /Air at 0.77 equivalence ratio. The piloted jet burner is centered in an unconfined co-flowing stream of air with a velocity fixed at 0.9 m/s for all cases.

The scalars measurements include temperature, mixture fraction, and the mass fractions of O_2 , N_2 , H_2 , H_2O , CH_4 , CO_2 , OH , NO , CO , at different radial and axial profiles. Samples were taken between 10 and 19 different radial locations for all six axial positions via Rayleigh and Raman scattering to capture the temperature and species, respectively. Laser-induced fluorescence (LIF) was used to determine instantaneous OH and NO . Carbon monoxide was measured by Raman scattering and two-photon LIF (TPLIF). The CO-TPLIF data are preferred over CO-Raman measurements as the latter are more strongly affected by hydrocarbon fluorescence interference, causing errors in the conditional means in the region of high interference on the fuel-rich side of the reaction zone. Further information on the measurement techniques, the experimental setup, and the burner's characteristics can be found in [11,44]. The simplicity of the flow (without the soot formation and thermal radiation complications) and the existence of a fully turbulent region of the flame where the chemical kinetics effects are significant

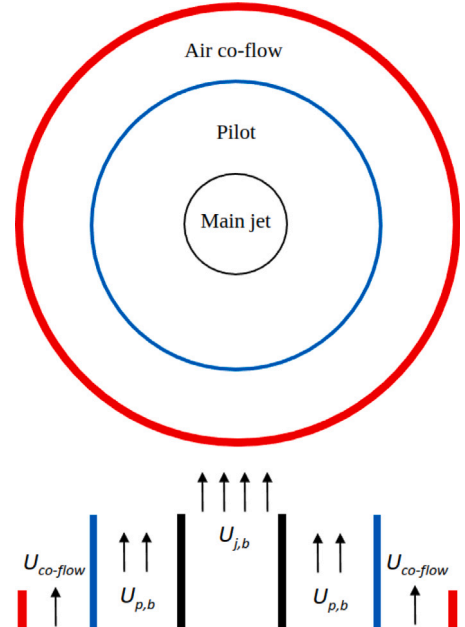


Fig. 2. Plan view schematic of the exit geometry in the Sandia/TUD burner, showing a plan view and a cross-section through the burner axis. In the cross-section view, $U_{j,b}$ and $U_{p,b}$ denote the bulk velocity for the fuel jet and the bulk velocity for the pilot, respectively. In all cases, $U_{co-flow} = 0.9$ m/s.

makes this burner an ideal test case for evaluating the possibility of using a principal component as a conditioning variable in CSE.

3.2. Data-processing

While flames C, D, and arguably E burn as non-premixed flames with a single reaction zone near the stoichiometric mixture fraction, Flame F exhibits partially-premixed characteristics with the evidence of premixed reaction flame fronts. As a result of the increased turbulence intensity, Flame F has a significant probability of local extinction in regions where chemical and turbulent mixing time scales are of the same orders of magnitude. Further downstream the burner, re-ignition pockets appear within the flow where turbulent mixing rates are less intense. This makes this burner, and particularly Flame F, ideal for studying filter-chemistry interaction effects. As such, only the datasets of Flame C and Flame F will be used hereafter to investigate the effects of spatial coordinates and Reynolds numbers on the conditionally-averaged reactive scalars.

The data was first “cleaned” to remove mass fractions displaying negative values associated with experimental uncertainty. Instantaneous single-point measurements of mixture fraction following Bilger's definition are used, as

$$Z = \frac{2 \frac{Y_C - Y_{C,2}}{M_C} + \frac{1}{2} \frac{Y_H - Y_{H,2}}{M_H}}{2 \frac{Y_{C,1} - Y_{C,2}}{M_C} + \frac{1}{2} \frac{Y_{H,1} - Y_{H,2}}{M_H}} \quad (10)$$

where only the elemental mass fractions of hydrogen and carbon are included. This is because the main jet and co-flow boundary conditions for the elemental oxygen mass fraction are relatively close, and shot noise in the measurements of elemental oxygen mass fraction causing additional noise in the mixture fraction [11]. Additionally, it was shown that averaged experimental results using the Bilger definition and the present modification do not differ significantly due to differential diffusion effects that are relatively small in these flames [45]. For the considered sets of flames, the stoichiometric mixture fraction lies at $Z_{\text{stoich}} = 0.351$.

Four of the most common combustion progress variable c used by the community are calculated for every data point in the datasets, defined as

$$c_k = \frac{\phi_k - \phi_{k,\min}}{\phi_{k,\max} - \phi_{k,\min}} \quad (11)$$

where ϕ_1 denotes the temperature, while ϕ_2 , ϕ_3 and ϕ_4 are the mass fractions of CO_2 , $\text{CO} + \text{CO}_2$ and $\text{CO} + \text{H}_2 + \text{H}_2\text{O} + \text{CO}_2$, respectively. The local maximum and minimum are determined using a function that returns the upper and lower peak envelopes of the scalar k selected to define c . The progress variables will be used later on for the two-condition conditional averages analysis.

A second condition based on mixture fraction and the four progress variables (i.e., $Z, c_k < 0$ and $Z, c_k > 1$) is applied to take potential outliers from the analysis as these points are considered to be unphysical. After executing all previous steps, each database consists of 15,862 measurements for Flame C and 22,009 for Flame F.

Principal component analysis requires high-fidelity datasets to generate the PC-basis and accurately describe the thermo-chemical state-space. The experimental measurements of both datasets fed to PCA have been “cleaned out” following the steps presented previously. It should be noted that the mixture fraction and the progress variables have not been included in the two databases before being passed to PCA. The PCA analysis is carried out using two scaling factors, namely Pareto and Auto-scaling (AS), assuming that the datasets have been previously centered. Pareto scaling, which adopts the square root of the standard deviation, was recognized as having a distinct advantage for major species and source terms reconstruction while needing fewer components [4]. Auto-scaling option, which uses the standard deviation as the scaling factor, was found to need often more components to achieve the same reconstruction accuracy obtained with Pareto [5]. Previously, Parente & Sutherland [46] found that Auto-scaling is more adapted when an exploratory analysis on the chemical manifold should

be performed. In contrast, Pareto appears more suitable for capturing the systems' principal features and the main species' behavior. Parente & Sutherland also [46] showed that the square root of the standard deviation enhances the temperature scalar in carrying most of the data variance, thus, forcing the first principal component to align with temperature. For this reason, the temperature was excluded from the two flames' datasets passed on to PCA. PCA's real utility comes by finding correlations among the variables defining the state-space. A new coordinate system is identified in the directions of maximal data variance, allowing less important dimensions to be eliminated while maintaining the primary structure of the original data. In that sense, one can suppose applying PCA to a given reactive flow where no prior knowledge about the physical and chemical phenomena is known, and generate the adequate control variables needed to accurately quantify the thermo-chemical state-space within a manifold. In order to determine the amount of information captured by each principal component and thus replace the Q elements of X by $q < Q$ principal components, the fraction of total variance accounted by each PC is calculated as

$$t_{q_i} = \frac{\sum_{k=1}^{q_i} l_k}{\sum_{k=1}^Q l_k} \quad (12)$$

where i and l_k denote a single PC and the variance located on the diagonal of the covariance matrix S , respectively. Since outputs resulting from the two scaling methods have different numerical ranges, their PC-scores have been scaled to the interval $[-1, 1]$.

To examine the conditional space of Sandia/TUD flames with PC-scores obtained from the above-mentioned methodology, the approach proposed in [47,48] is followed and applied to both datasets. The conditional fluctuations of species mass fractions and temperature around one-condition (PC1-score Pareto, PC1-score AS, and Z) conditional averages are first calculated at each axial location, such that

$$\begin{aligned} f'_{i,k} &= f_i - \langle f | \eta = \Psi_1 \rangle(x) \\ f'_{i,k} &= f_i - \langle f | \eta = Z \rangle(x) \end{aligned} \quad (13)$$

where i denotes a single-point measurement, $f'_{i,k}$ is the fluctuation of either mass fraction or temperature around one-condition, f_i is the point measurement of that reactive scalar and $\langle f | \eta \rangle(x)$ is the conditional average of that reactive scalar evaluated by averaging all of the measurements of the chosen dataset at all radial locations together at each downstream distance. The conditional averages are obtained via a discrete process involving binning, dividing each PC1-score and mixture fraction dimension into 50 bins. It should be noted that conditional averages are investigated along the axial direction as conditional fluctuations are larger in the axial positions and are spatially independent in radial locations [19,48]. That was shown to be particularly true for jet flames.

If the conditional average is a good representation of the local thermo-chemical state, then the mean of the conditional fluctuations is expected to be zero. However, the RMS of those fluctuations is clearly not, as shown by Bushe [48] using the Sandia/TUD database and the Sydney swirl burner [49,50]. Therefore, the square root of the average of the square of conditional fluctuations can be computed, as

$$\text{RMS}_{i,k} = \sqrt{\langle f_{i,k}^2 \rangle} \quad (14)$$

The proposed RMS is normalized by the maximum value of the considered reactive scalar. This last step is justified in two ways: (i) to compare the different scalars to one another which the relative magnitude should be comparable, and (ii) the data has been filtered to eliminate outliers, suggesting that the maximum measured value is unlikely to be the consequence of a major measurement error.

4. Results & discussion

4.1. Principal component analysis

Fig. 3 presents the variance accounted by both flames' principal components with all radial and axial locations grouped together. The

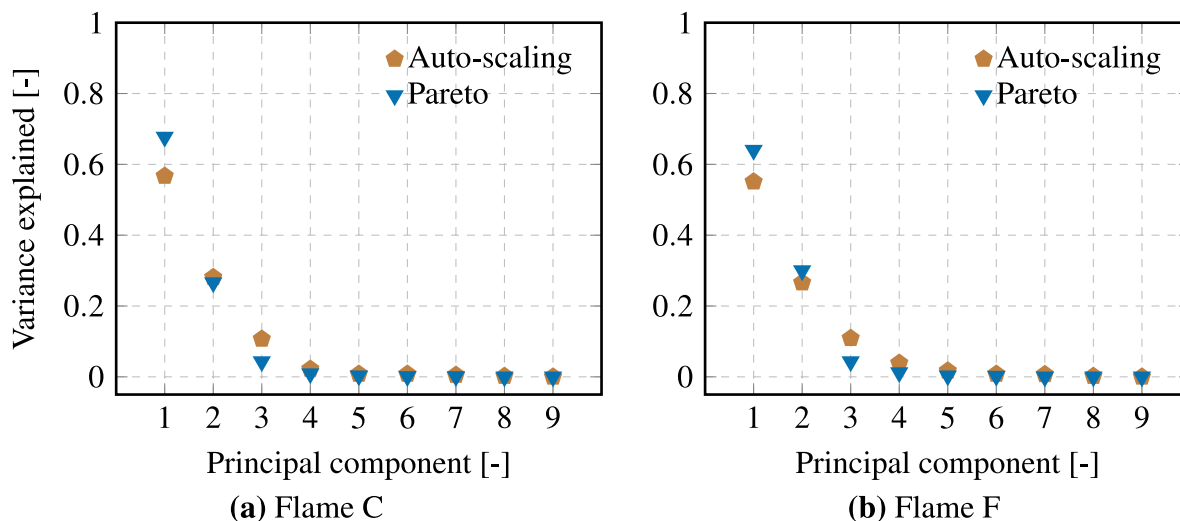


Fig. 3. Comparison of variance explained with Pareto (triangles) and Auto-scaling (pentagons) for each principal component of (a) Flame C and (b) Flame F.

two scaling methods output similar results for both flames. Regardless of the scaling option adopted, two principal components seem to account for the largest amount of variance present in both datasets, with ~ 0.6 for PC1 and ~ 0.3 for PC2. Slightly more variance captured by PC1 using Pareto is observed compared to AS. The total variance explained by both PCs is in good agreement with the ~ 0.9 threshold proposed by Parente et al. [3]. It was shown that all main species and temperature can be recovered with satisfactory levels of approximation. Consequently, the physical interpretation of all other principal components is omitted in this work, as it is believed to be out of the scope of this study. To clarify, figures depicting PCA results do not include the temperature scalar within the analysis.

The weights of the original variables characterizing the two flames (i.e., matrix A) are presented in Fig. 4 for both scaling methods.

Regardless of the scaling option selected and the flame studied, it is interesting to note that PCA is able to automatically distinguish reactants from products with PC1, where PC1-Pareto and PC1-AS are here negatively correlated with the mass fractions of CH_4 and O_2 , depicted in Fig. 4(a)–(b). It should be noted that the results of PC1-AS for Flame C have been flipped to make the comparison with PC1-Pareto clearer/easier. Considering the criterion proposed by Ranade & Echekki [51], only coefficients with magnitudes ≥ 0.4 are kept to help identify the more prominent contributors to both PCs. Regardless of the dataset, it appears that the mass fractions of O_2 , H_2O and CO_2 have the most important contributions to PC1-Pareto, with coefficients equal to approximately 0.60, 0.50 and 0.45, respectively. This trend is also apparent for PC1-AS, where the same three species have dominant weights with ~ 0.4 for both flames. Interestingly, regardless of the flame and the PC investigated, Auto-scaling appears to have non-negligible weights on intermediate species, as opposed to Pareto, which clearly emphasizes main species. This observation agrees with the study undertaken by Parente et al. [46], which showed that the variance accounted for minor species by Auto-scaling is up to $\sim 20\%$ higher than that explained by the other scaling methods investigated in their work. For the second principal components, PC2-Pareto and PC2-AS exhibit the same positive/negative contributions, with a shift occurring from Flame C to Flame F, illustrated in Figs. 4(c)–(d). Adopting the Pareto scaling, the mass fractions of CH_4 and N_2 have the largest weights on PC2 for both flames, with coefficients equal to ~ 0.65 . While the most important contributions to PC2-AS are the mass fractions of H_2 (~ 0.55), CO (~ 0.48) and N_2 (~ 0.48) for Flame C, for Flame F, nitrogen, methane, and H_2 coefficients have significant weights with values of approximately 0.57, 0.53 and 0.42, respectively.

Table 2

Summary of the conditioning variables investigated in this study for one-condition conditional averages.

Label	Scalar	Mark
Pareto	PC1-score Ψ_1	×
AS	PC1-score Ψ_1	*
Z	Mixture fraction	▼

In view of the presented results, all three species, namely O_2 , H_2O and CO_2 , are known to behave linearly with temperature, suggesting (irrespective of the scaling option) that PC1 is perhaps correlated/aligned with temperature. This trend is illustrated in Fig. 5(a)–(b), where results of both flames' datasets considered herein promote a PC1 behaving monotonically with temperature. The dominant contribution to PC2-Pareto is associated with the fuel for both flames, which is highly correlated with Z. This tendency is shown in Figs. 5(c)–(d). For Flame F, the correlation between PC2-Pareto and Z is more emphasized. Although N_2 , H_2 and CO are also known to be aligned with the mixture fraction, the results obtained for Flame C with AS promote a non-linear PC2 in the Z-space. Therefore, it could perhaps represent a measure of the degree of chemical reactivity as PC2-AS is mainly dominated by intermediate species. For Flame F, a similar behavior is visualized in Fig. 5(d). The correlation with the mixture fraction is more accentuated due to the non-negligible weights of methane. Overall, the presented results are in good agreement with the analysis previously conducted in [51,52].

As the first PCs account for the largest amount of variance, only the PC1-scores obtained from both scaling methods will be used as variables to evaluate the first-moment conditional fluctuations.

4.2. One conditioning variable

After identifying the principal components' structures, one can calculate the conditional averages of the Q variables describing the flames' thermo-chemical states and consequently determine the conditional fluctuation associated with each experimental data point. Here, conditional fluctuations are calculated at each axial position combining data from all radial locations for each given axial position. The single conditional averages using the first principal component scores of Pareto and AS are investigated and compared against the one obtained using the mixture fraction. For clarity purposes, the labels and markers depicted in Table 2 will be used throughout this study.

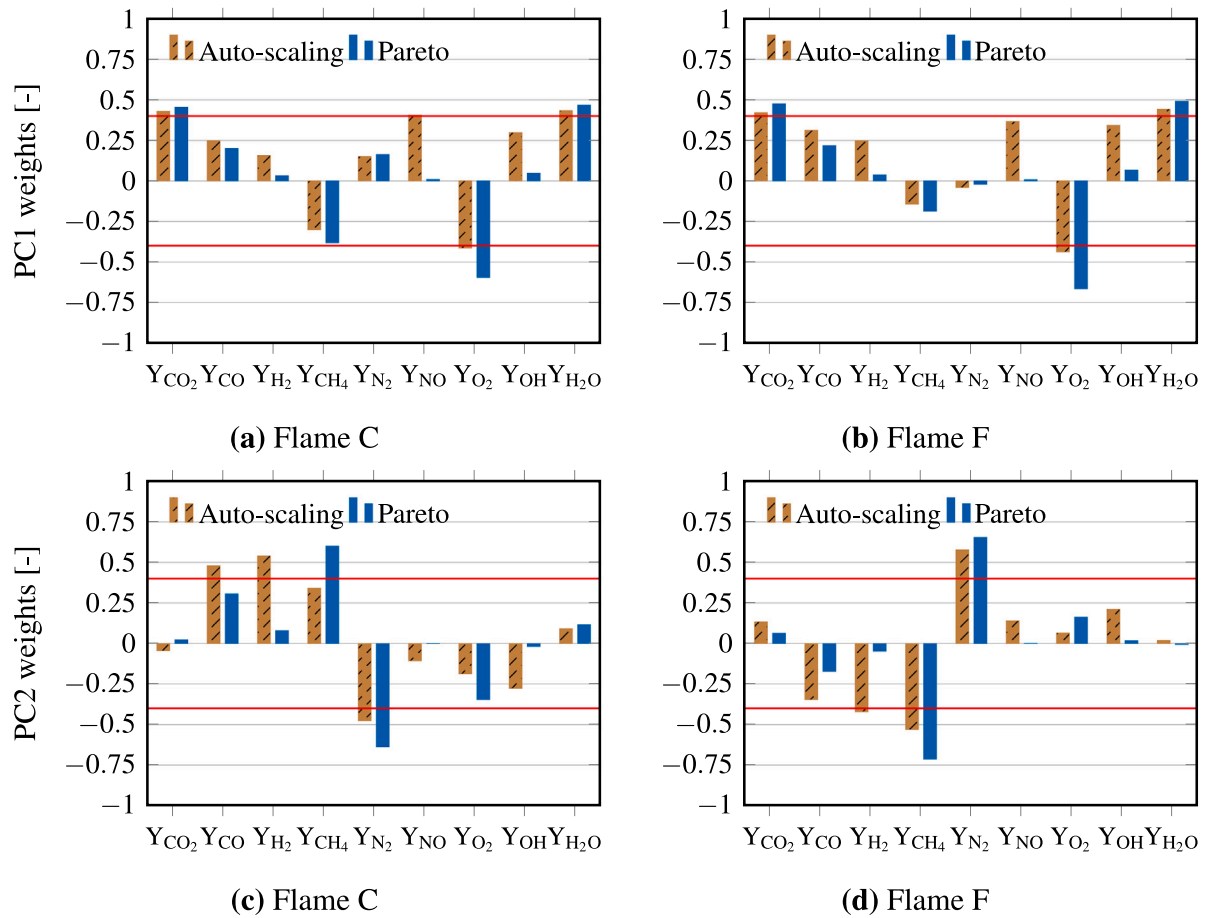


Fig. 4. Comparison of weights obtained with Pareto and Auto-scaling (with strips) for the leading principal component and scalars of (a) Flame C and (b) Flame F, and the second principal component and scalars of (c) Flame C and (d) Flame F. The solid red lines indicate the 40% threshold used in this study.

Fig. 6 illustrates the conditional fluctuations of temperature and various species mass fractions calculated from the Flame C dataset. Although Flame C does not exhibit severe probabilities of localized extinction, the fluctuations induced by the turbulence effects can be clearly identified at axial positions equal to 15, 30 and 45. This trend is visually highlighted by conditional fluctuation points spread far from zero. The conditional fluctuations of the mass fractions of CH_4 , CO , OH and H_2 around PC1-Pareto and PC1-AS exhibit a clear functional dependence on the physical domain compared to Z . This can be attributed to the fact that all four previous species' mass fractions are more closely relevant to the mixture fraction than a principal component well correlated with temperature. Regardless of the scaling method chosen, it is interesting to note that conditional fluctuations of CO_2 , H_2O and NO mass fractions using Ψ_1 are of the same order of magnitude as Z , if not better, particularly near the burner's tip and at the interval $x/d = 15-45$. This can be explained considering that all three species are homeomorphic to the temperature scalar. As expected, a similar behavior can be seen on the conditional fluctuations of temperature around Ψ_1 due to both PC1-scores being aligned with T .

Further downstream the burner (i.e., $x/d = 60-75$), all three conditioning variables promote the same tendency with reduced spatial gradient effects. While local averages of all conditional fluctuations are anchored at zero, and thus, suggesting that both PC1-scores can sufficiently well characterize the considered dataset, it is nearly impossible to evaluate the effectiveness of a single principal component score to accurately describe the thermo-chemical state-space and detach it from spatial coordinates and Reynolds number effects.

Fig. 7 shows the normalized RMS of all previously investigated scalars. Regardless of the conditioning variable investigated, the RMS

results of Y'_{CO_2} , Y'_{H_2O} , Y'_{NO} and T' seem to remain below 10% of their corresponding maximum value. A slight improvement can be seen with PC1-AS. Assuming that RMS of the order of 10% can be considered as “relatively small” [48], one can suppose using a single PC1-score as conditioning variable for this database and still get acceptable predictions of the considered reactive scalar for conditional moment closure models. However, the normalized RMS of intermediate species and CH_4 around Ψ_1 exhibit a strong dependence on the flow dynamics and the burner's geometry, highlighted by values exceeding the indicative 10% threshold used as a guideline within this study. This suggests that the conditional averages are different (changing in function of space and Reynolds numbers) and that a single principal component score selected as controlling variable is perhaps insufficient. This behavior is in good agreement with the results depicted with the PCA analysis (cf. Fig. 3), where two PC-scores were needed to adequately represent the thermo-chemical state-space of both flames. The most deplorable results are seen with the mass fractions of methane and carbon monoxide. As expected, using the mixture fraction as a single conditioning variable for the two previous species reduces their respective RMS' by a factor of 4 compared to both PC1-scores. Further downstream the axial position, the RMS of all the reactive quantities around both scaling options are improved or of the same order of magnitude as those calculated with Z . This is consistent with the tendency mentioned previously in Fig. 6.

The same analysis was performed for Flame F, illustrated in Fig. 8. Unlike the trends visualized on Flame C, the conditional fluctuations of CO , OH and H_2 around both Ψ_1 perform as well as those obtained around the mixture fraction. A slight increase can be identified in the axial position $x/d = 45$, corresponding (with $x/d = 30$) to the

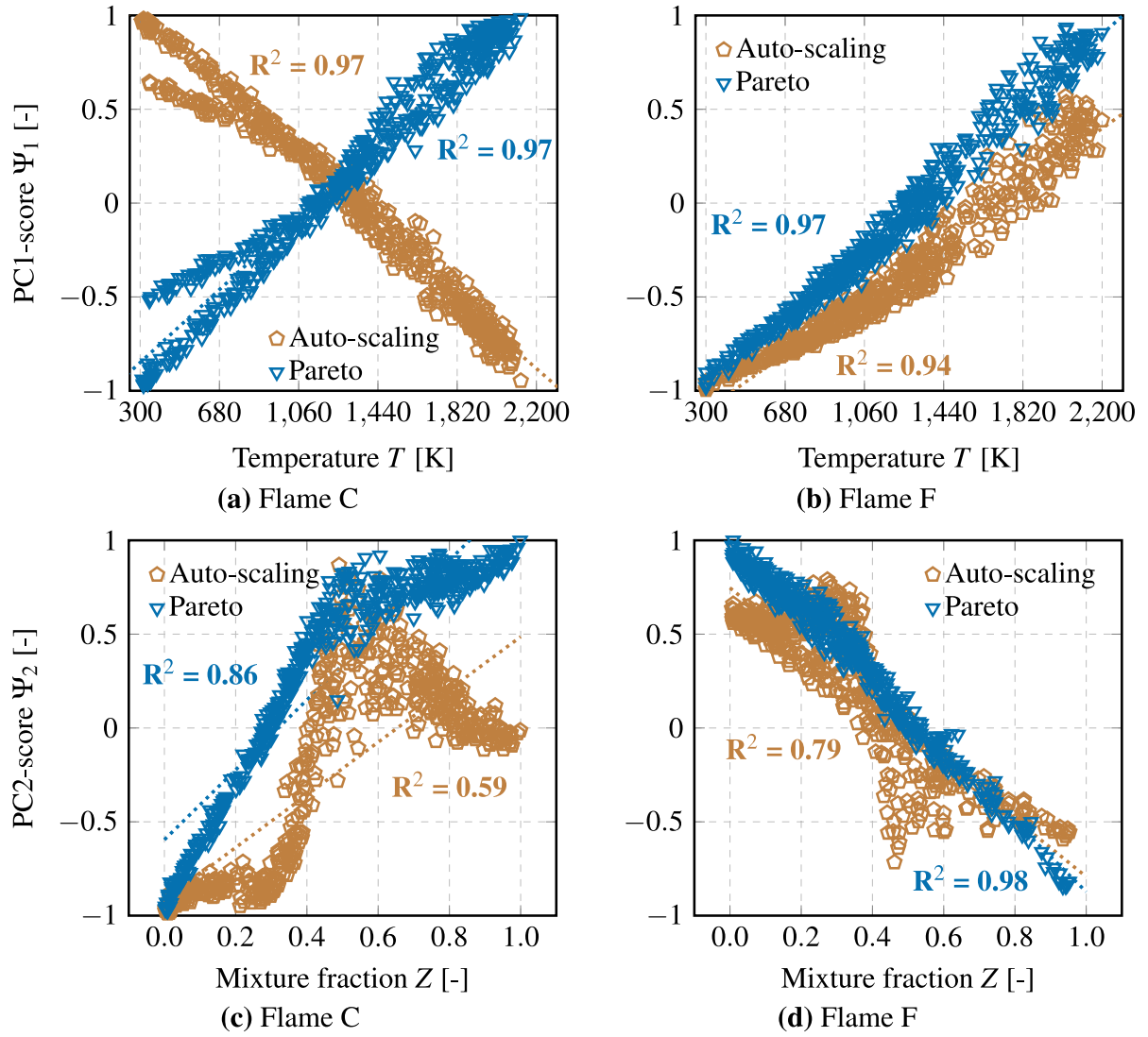


Fig. 5. Comparison of correlations obtained with Pareto (triangles) and Auto-scaling (pentagons) for the leading principal component with temperature for (a) Flame C and for (b) Flame F, and the second principal component with mixture fraction for (c) Flame C and for (d) Flame F. The markers illustrate 500 point-based measurements randomly selected within the flames' and PCs-scores' datasets.

location exhibiting the most intense levels of turbulence fluctuations and probabilities of localized extinction. The conditional fluctuations of CH_4 around both PC1-scores are more spread out across the domain (far from zero) than those obtained with Flame C. This is attributed to the flow dynamics and operating conditions that are more severe in Flame F. As expected, the results for Y'_{CH_4} around Z remain close to zero due to fuel mass fractions being more closely relevant to the mixture fraction than a PC1-score well aligned with temperature. Further downstream the burner (i.e., $x/d \geq 45$), where turbulent mixing rates a less intense, all three conditioning variables exhibit the same behavior. Despite the increased levels of turbulence intensity, the conditional fluctuations of CO_2 , H_2O and NO mass fractions using both scaling methods remain stable. The conditional fluctuations of the same three species around Z show a much more pronounced dependence with the physical domain than those calculated with Flame C, suggesting that a PC-score correlated with temperature is perhaps more suited for this flame. The conditional space fluctuations of temperature around the mixture fraction highlight the zones within the flow associated with extinction phenomena, which *in fine* increase the temperature fluctuations. As expected, both Ψ_1 are able to lower the conditional fluctuations of T' more effectively, with results remaining close to zero.

The local averages of all conditional fluctuations are fixed at zero throughout all the axial positions, suggesting that using either of the three conditioning variables should provide an accurate approximation of the chemical source-term, i.e. a closure utilizing only the first term of a Taylor expansion of the reaction rate.

Fig. 9 provides further insights. Regardless of the scaling method selected, adopting a single PC1-score appears to perform equally well as Z for CO and H_2 . Excluding the results observed at $x/d = 45$, where values deviate by a factor of ~ 2 compared to the ones obtained with the mixture fraction, one can suppose that a PC1-score might still give acceptable predictions of the two considered reactive scalars. This suggests that both PC1-Pareto and PC1-AS can more effectively decrease the conditional space fluctuations in Flame F than in Flame C. This is possibly due to the inherent nature of the PCA model, where PC-scores provide a rigorous mathematical formalism to reduce the dimensionality of the original data while retaining most of the variance induced by the turbulence. The normalized RMS of the mass fractions of CH_4 indicate that a principal component score well aligned with temperature is not suited to adequately characterize the state-space of the fuel near the burner's tip and in zones exhibiting important levels of turbulence intensity. At the burner's outlet, the RMS of Y'_{OH} around

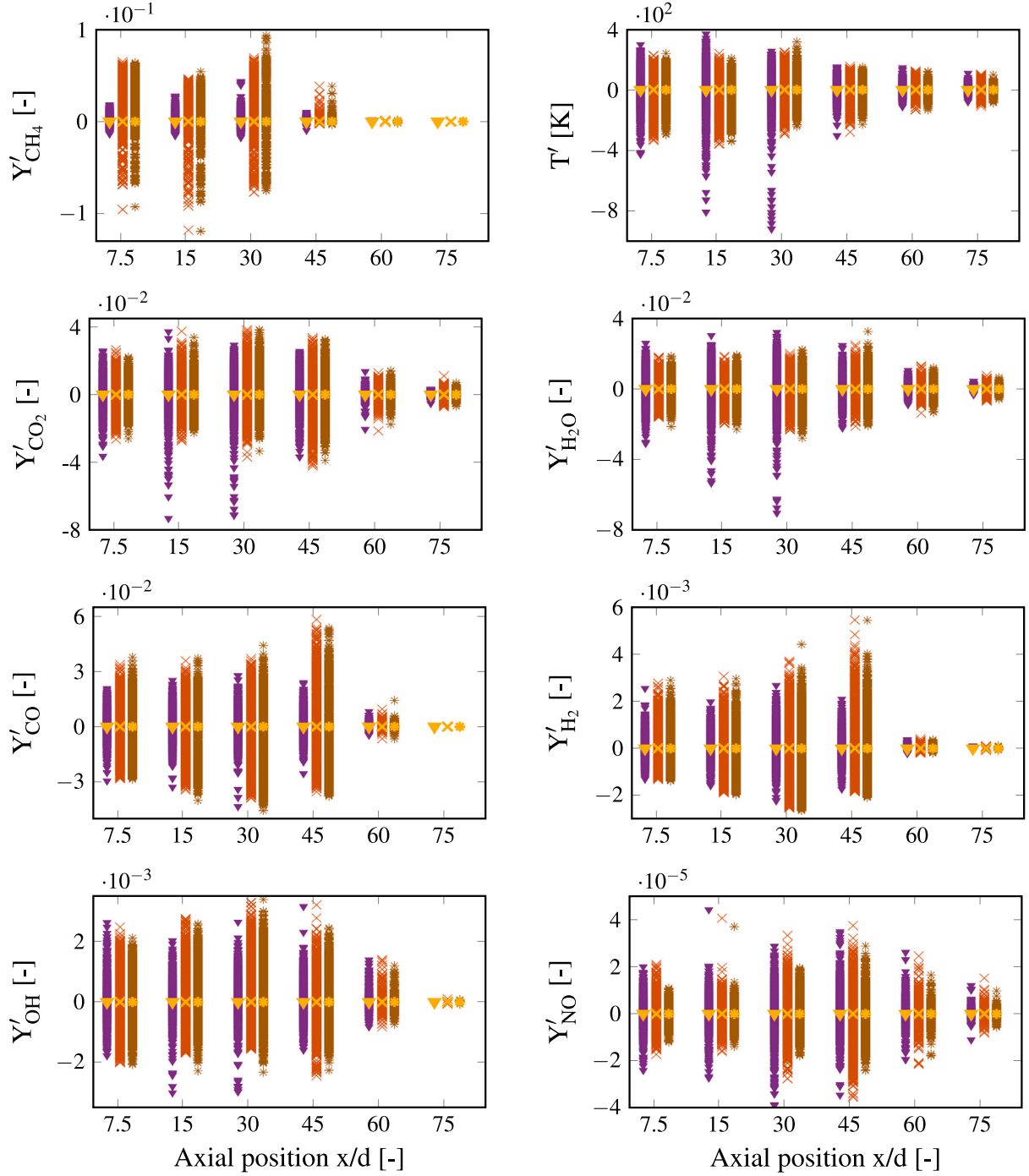


Fig. 6. Conditional fluctuations of temperature and species mass fractions for Sandia/TUD Flame C around the conditional average $\langle f | \eta = \Psi_1 \rangle(x)$ using the PC1-scores Ψ_1 of Pareto (crosses) and Auto-scaling (asterisks), and around the conditional average $\langle f | \eta = Z \rangle(x)$ using the mixture fraction as the single conditioning variable, and collecting all points at different radii together; are also shown the local average of these conditional fluctuations $\langle f'_{i,k} \rangle$ (golden markers).

Ψ_1 are reduced compared to Z , suggesting that under the operating conditions of Flame F, the correlation between OH and the temperature scalar (through the use of the PC1-score) is perhaps more relevant than it is with the mixture fraction. Throughout all the flame's axial locations, the normalized RMS of NO, CO₂, H₂O, and temperature show that conditional averages around both PC1-scores are less likely to be affected by the spatial coordinates and Reynolds numbers. On the contrary, the results calculated for the same four scalars with Z promote RMS values exceeding the indicative 10% threshold. Based on the presented results, as turbulence intensity increases, a single principal

component score appears to be a more rigorous/coherent choice to separate the conditional manifold from the real domain compared to a single control variable based on Z . Assuming a single conditioning variable is retained, one can suppose that filter-chemistry interaction effects could perhaps be better captured using a PC1-score well correlated with temperature. As expected, for non-premixed “quasi-laminar” flames, such as Flame C, using Z as a control variable seems more appropriate. Considering that neither the mixture fraction nor the two PC1-scores can sufficiently reduce the spatial gradient and Reynolds numbers effects of some intermediate species (i.e., CO and OH), con-

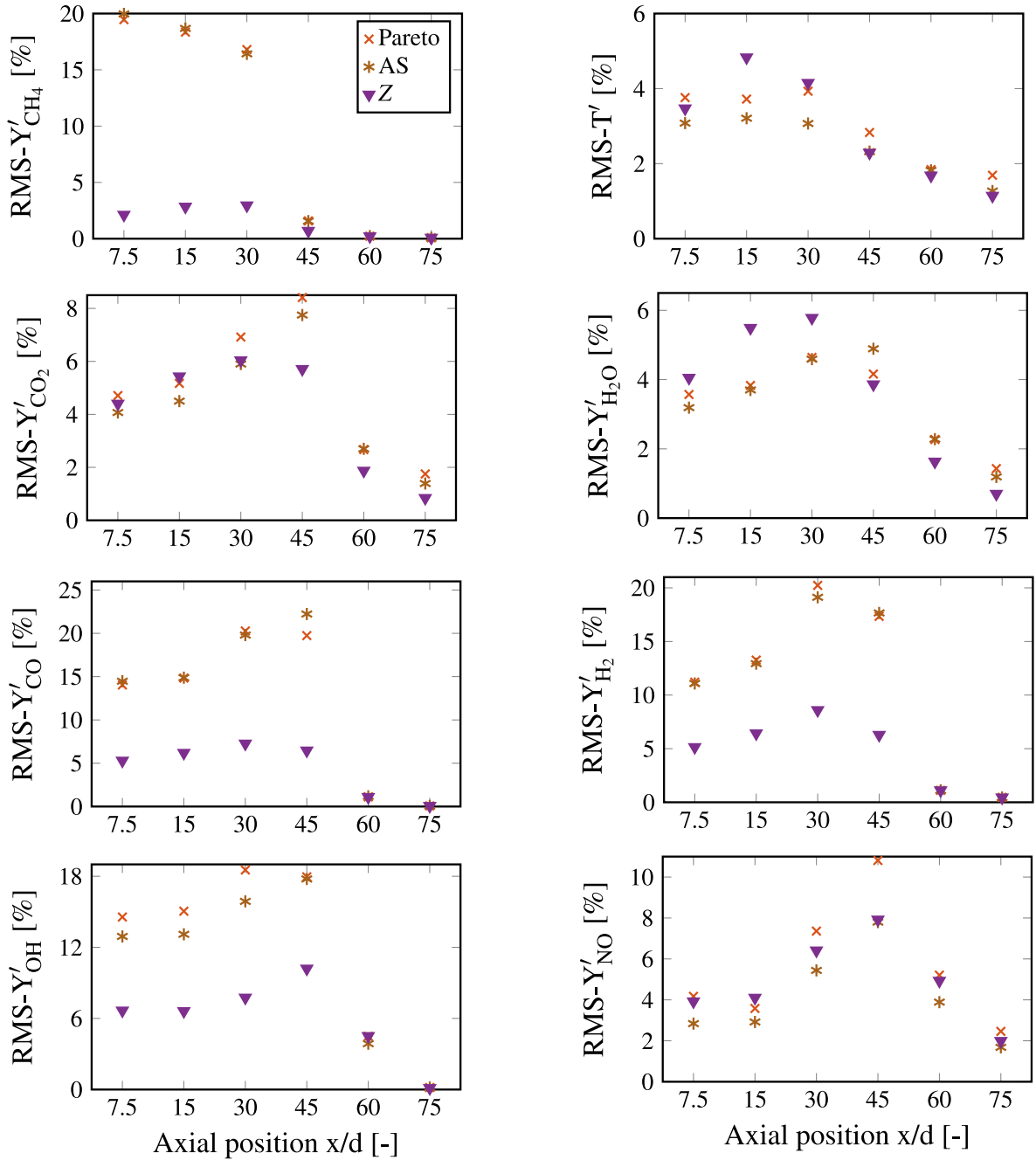


Fig. 7. Normalized RMS of the conditional fluctuations of temperature and species mass fractions for Sandia/TUD Flame C around the conditional average $\langle f | \eta = \Psi_1 \rangle(x)$ using the PC1-scores Ψ_1 of Pareto (crosses) and Auto-scaling (asterisks), and around the conditional average $\langle f | \eta = Z \rangle(x)$ (triangles) using the mixture fraction as the single conditioning variable, and collecting all points at different radii together.

ditional fluctuations around two-condition conditional averages are investigated to provide further insights on the feasibility of using CSE with PCA.

4.3. Two conditioning variables

The conditional fluctuations previously obtained using PC1-Pareto and PC1-AS are further examined and compared against the conditional fluctuations calculated around two-condition conditional averages using the mixture fraction and one of the four progress variable definitions previously introduced in the methodology section. The conditional fluctuations at each axial location are determined as

$$f'_{i,k} = f_i - \langle f | \eta = Z, \xi = c_k \rangle(x) \quad (15)$$

In this approach, the conditional averages are obtained by dividing each of the four progress variables' dimensions into 20 bins while preserving the initial 50 bins for the mixture fraction. The number of bins selected for the second conditioning variable is often based on user expertise. Here, the number of bins retained for the progress variables can be justified given that more bins increase the possibility of having intervals with an insufficient number of data points which imposes unrealistically small fluctuations around the mean value [47]. Moreover and within a DCSE context, if more bins are included then more computational time is needed during the inversion process with the matrix of the joint PDF. This implies that a much larger matrix needs to be inverted compared to CSE-PCA or previous implementations of CSE for premixed and non-premixed flames. Following the analysis depicted on the conditional space fluctuations of Flame C, (where

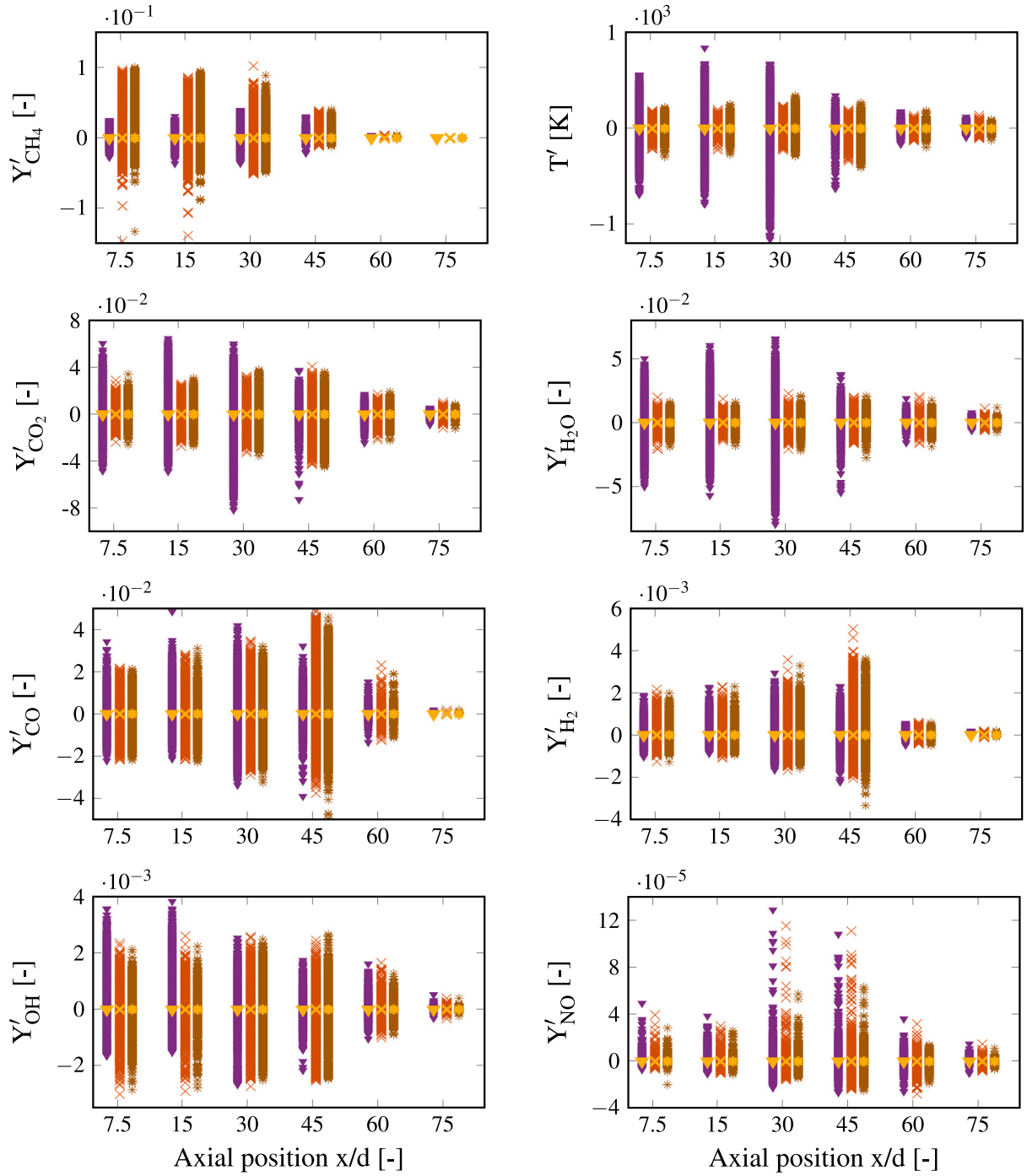


Fig. 8. Conditional fluctuations of temperature and species mass fractions for Sandia/TUD Flame F around the conditional average $\langle f|\eta = \Psi_1 \rangle(x)$ using the PC1-scores Ψ_1 of Pareto (crosses) and Auto-scaling (asterisks), and around the conditional average $\langle f|\eta = Z \rangle(x)$ (triangles) using the mixture fraction as the single conditioning variable, and collecting all points at different radii together; are also shown the local average of these conditional fluctuations $\langle f'_{i,k} \rangle$ (golden markers).

adopting the mixture fraction as a single control variable enabled to sufficiently well reduce the thermo-chemical state dependence on the burner's characteristics) only the conditional fluctuations of Flame F around Z , c_k will be presented. Doubly conditional averages will not significantly improve the scalars predictions of Flame C, but will however lead to unnecessary additional implementation complexities and computational costs. The labels and markers given in Table 3 will be used throughout this study.

Fig. 10 illustrates the conditional fluctuations of Flame F around two-condition conditional averages. Regardless of the progress variable definition adopted, it can be deduced that: (i) doubly conditioning is

of particular interest for intermediate species such as H_2 , in particular those believed to be highly correlated with Z (e.g., Y_{CO}), and more importantly (ii) conditional averages of all other investigated variables around both PC1-scores exhibit similar behavior as those obtained with Z , c_k . On the contrary, the results of CH_4 around all four combinations of doubly conditioning are approximately three times lower than those obtained with PC1-Pareto and PC1-AS. It is interesting to note that the Y'_{CH_4} around the mixture fraction (cf. Fig. 8) are of the same order of magnitude as those obtained with two-condition conditional averages. This suggests that adding any of the four c_k does not significantly affect/improve the fit for this quantity. Slightly better results with both

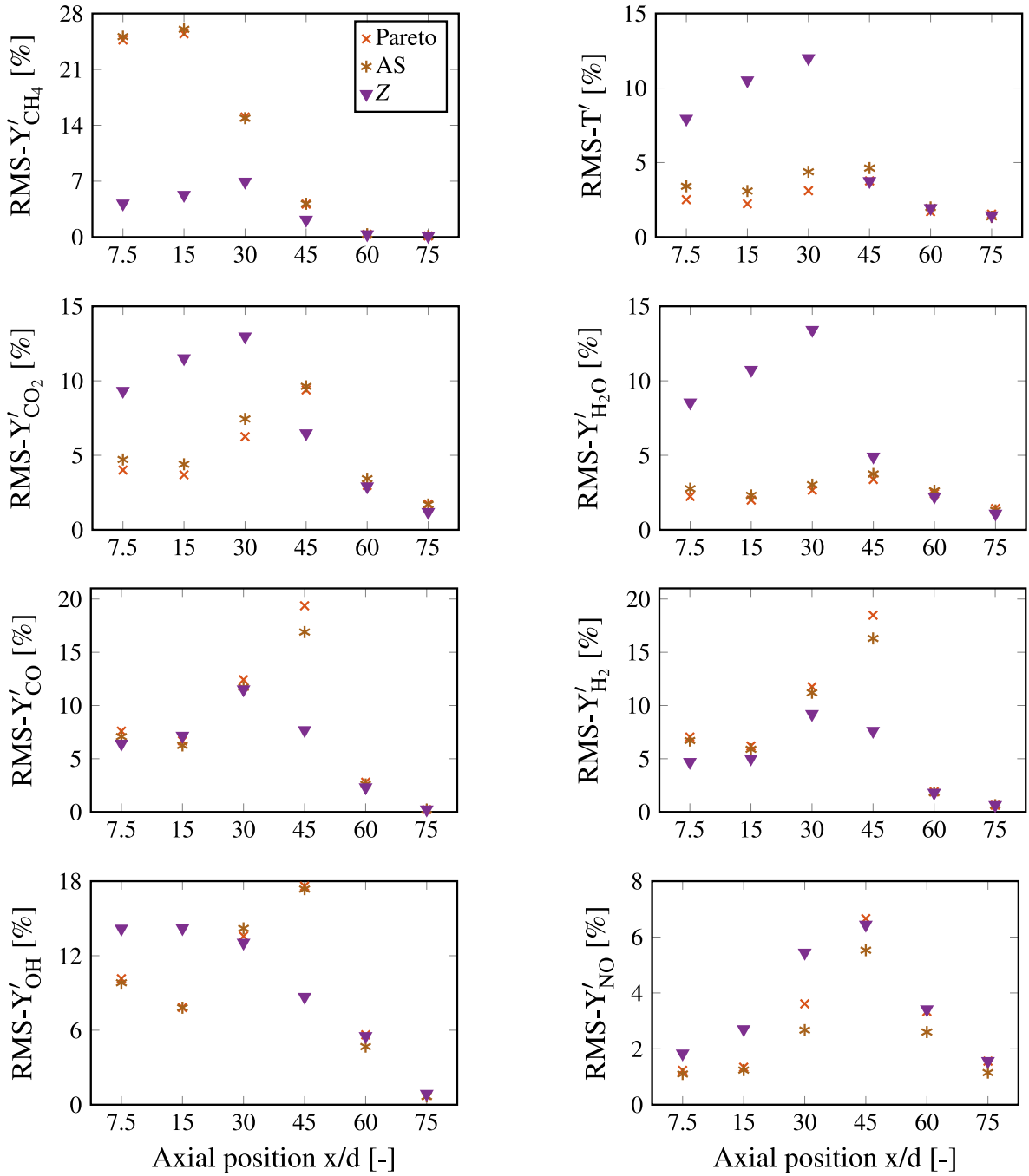


Fig. 9. Normalized RMS of the conditional fluctuations of temperature and species mass fractions for Sandia/TUD Flame F around the conditional average $\langle f|\eta = \Psi_1 \rangle(x)$ using the PC1-scores Ψ_1 of Pareto (crosses) and Auto-scaling (asterisks), and around the conditional average $\langle f|\eta = Z \rangle(x)$ (triangles) using the mixture fraction as the single conditioning variable, and collecting all points at different radii together.

Table 3

Summary of the four progress variables used in this study. The markers designate the four combinations of doubly conditional source-term estimation with mixture fraction and progress variable.

Label	Scalar(s) ϕ_k	Mark
c_1	Temperature	▲
c_2	Y_{CO_2}	■
c_3	$Y_{rmCO_2} + Y_{CO}$	●
c_4	$Y_{CO} + Y_{H_2} + Y_{H_2O} + Y_{CO_2}$	◆

Ψ_1 for the mass fractions of H_2O and the temperature can be identified at the axial interval $x/d = 7.5$ -30.

This tendency is illustrated more straightforwardly in Fig. 11, where across all axial positions, PC1-Pareto and PC1-AS have lower normalized RMS values for the conditional fluctuations of temperature and the mass fraction of H_2O . This suggests that a single PC-score can more effectively decrease conditional averages' functional dependence on spatial coordinates and Reynolds numbers than two-condition conditional averages. Surprisingly, it can also be deduced that using a single PC-score can equally well perform as the four combinations of Z , c_k for CO_2 and NO mass fractions. As anticipated, the normalized RMS of intermediate species, namely CO , OH and H_2 remain relatively stable with doubly conditioning and do not seem affected by the

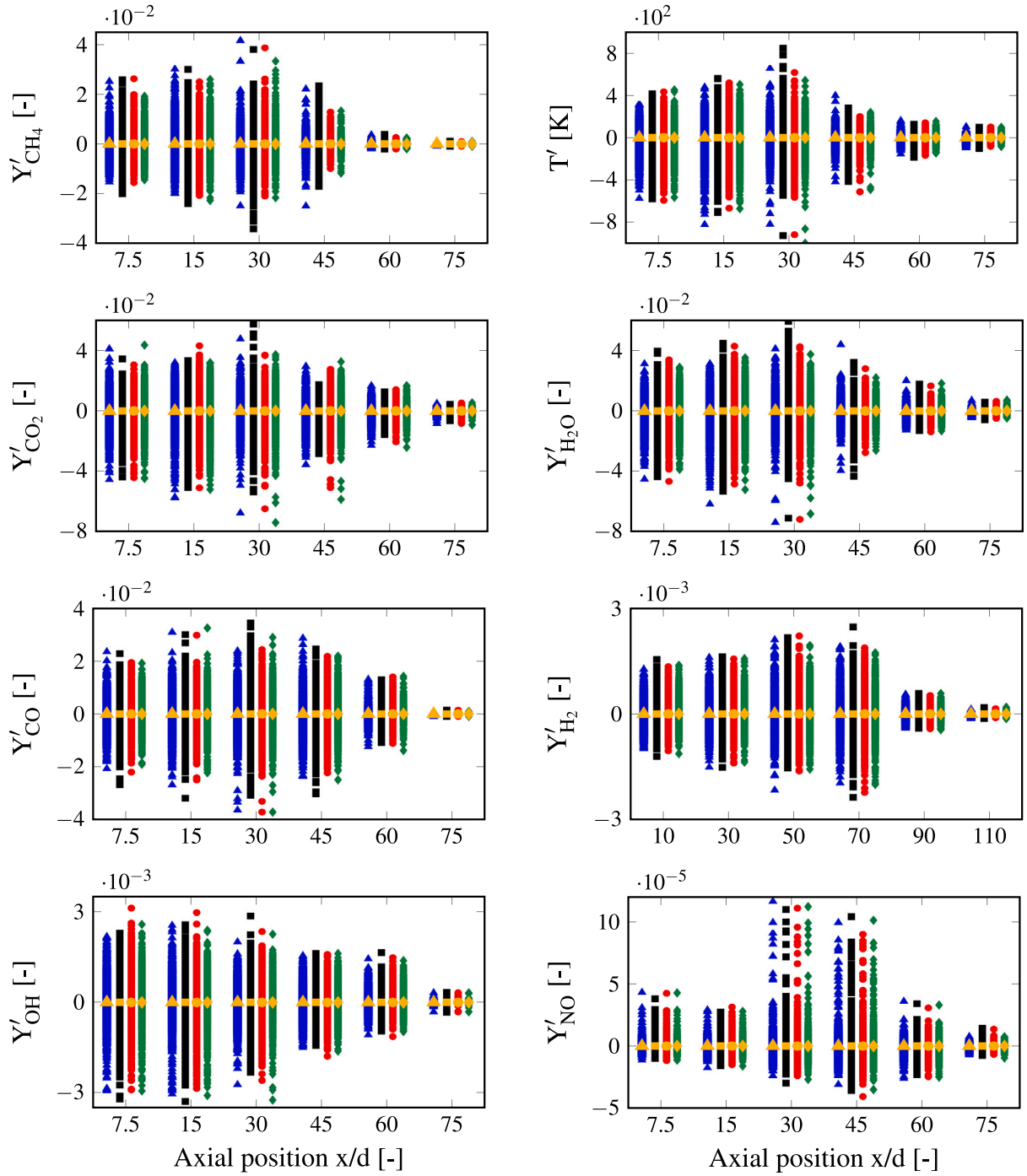


Fig. 10. Conditional fluctuations of temperature and species mass fractions for Sandia/TUD Flame F around the conditional average $\langle f | \eta = Z, \xi = c_k \rangle(x)$ (markers) using the mixture fraction and one of the four progress variables, and collecting all points at different radii together; are also shown the local average of these conditional fluctuations $\langle f'_{i,k} \rangle$ (golden markers).

flow's conditions. RMS results obtained around both PC1-scores for the same three species are generally slightly higher than two conditioning variables, however, below the indicative 10% if excluding the values corresponding to the axial positions equal to 30 and 45. The normalized RMS results of the fuel proved to be more effective around two-condition conditional averages than those obtained with Z .

Somewhat less relevant, it is interesting to note that all four sets of doubly conditioning exhibit similar behavior for all the investigated scalars, with conditional fluctuations and normalized RMS results of the same order of magnitude. The differences remain minor, suggesting that the choice of a particular progress variable definition may

not be important in CMC-based approaches, as deduced in [31]. The local averages throughout all downstream locations are equal to zero, suggesting that the definition attributed to c is perhaps less relevant in a closure context than in an accurate representation of the chemical state. However, it is believed that including the temperature scalar (suggested by the PCA analysis) as one of the controlling variables will provide a much more accurate representation of the Sandia/TUD flames' chemistry (such as in a manifold) compared to a species-based progress variable. One can suppose that diffusion effects play an essential role in describing the chemical states. As such, it is assumed that progress variable definitions based on single species (or a linear combination

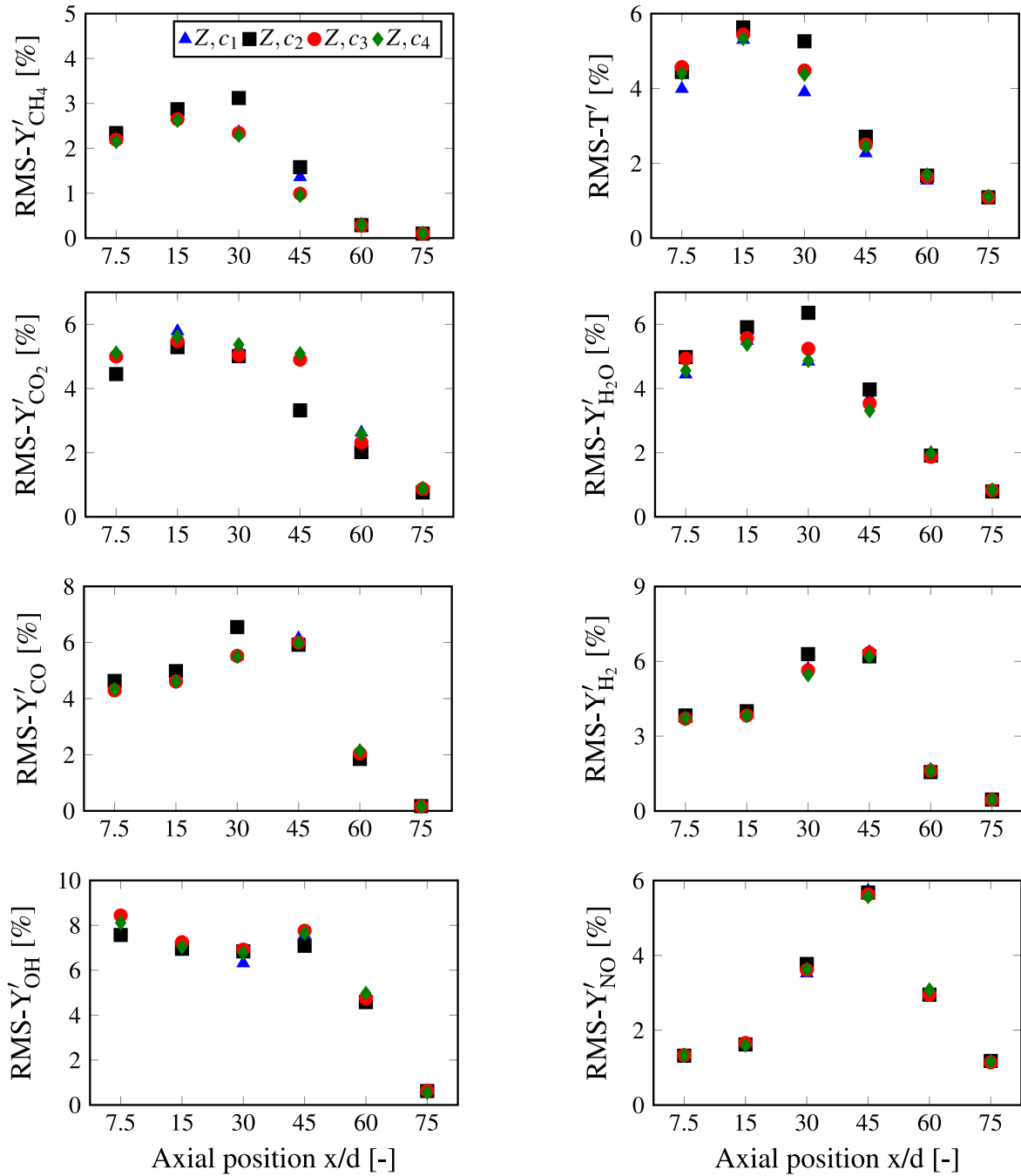


Fig. 11. Normalized RMS of the conditional fluctuations of temperature and species mass fractions for Sandia/TUD Flame F around the conditional average $\langle f | \eta = Z, \xi = c_k \rangle(x)$ (markers) using the mixture fraction and one of the four progress variables, and collecting all points at different radii together.

of species) would be a poor choice as the diffusion coefficient is often modeled using the unity Lewis number assumption. In contrast, the reduced temperature progress variable includes the thermal diffusivity by solving the diffusion flux term.

All in all, this work demonstrates that PC-based variables can capture the major physical processes and reduce conditional space fluctuations for reacting flows. Considering the results presented throughout this study, one can suppose that a CMC-based model such as CSE, coupled with PCA, can perhaps recover with satisfactory levels of approximation the reactive scalars of Sandia/TUD Flame F. Using such a method can potentially provide an alternative computationally efficient route for turbulent reactive flows combining both the advantages of the two models as highlighted before [13,24]. However, additional *a-priori* research would be required to evaluate the real capabilities of adopting

a single PC-score as a conditioning variable in CSE, particularly in defining the shape of the PC1-score PDF which remains an important unresolved issue. As quoted previously by the authors, presuming the PDF's shape of a single reactive PC1-score will need careful/special attention as its shape affects the results in an *a posteriori* fashion. This challenge can be seen as analogous to the progress variable one where a single distribution function accurately presuming its PDF shape has not been unanimously accepted by the community (as opposed to the mixture fraction Z and the beta-function) and remains an extensive field of research. As such, one can suppose that different alternatives and/or methods are worth exploring if classical parametric distributions (normal, beta, log-normal, etc.) do not agree with the experimental PC1-score PDF profiles. Recently, Ranade et al. [35] used a multi-dimensional kernel density estimation approach (KDE) to

construct the joint PDF in PC-space, where an *a-priori* validation of the PCA-KDE has already been carried out in [51] and showing good results using the same Sandia/TUD database. However, while this approach is suited for reconstructing a PDF profile from given measurements, the extrapolation to different cases and conditions, taking solely the mean and variance terms, seems unrealistic within a numerical framework. Facing that, other data-driven techniques such as deep neural network models (DNN) have gained considerable attention in the last few years. The recent studies of de Frahan et al. [53] and Chen et al. [54] have shown the possibilities of adopting DNN for presuming the joint PDF shape of Z and c using the mean, variance and co-variance terms of the latter as inputs. This approach can serve as a starting point that can possibly be extended to single-variable PDF problems.

5. Conclusions

In this work, a new *a-priori* study is undertaken to evaluate the effectiveness of adopting a single principal component (PC) as conditioning variable in the conditional source-term estimation (CSE) model. To this scope, the Sandia/TUD measurements of Flame C and Flame F are used in conjunction with principal component analysis (PCA) to obtain the new set of controlling variables, PC-scores. Two scaling methods for the PCA model have been adopted, namely Pareto and Auto-scaling (AS). For both datasets, it was found that: (i) the first principal component, regardless of the scaling method used, is well-aligned with temperature, and (ii) the second principal component was found to be highly correlated with the mixture fraction, Z , using Pareto scaling, whereas AS promotes a PC2-score describing the chemical reactivity. Considering that only the first principal component associated with temperature explains the largest amount of data variance, only the PC1-scores of Pareto and AS were tested as conditioning variables.

The conditional spaces of both flames are examined by investigating the conditional fluctuations and normalized RMS of temperature and various species mass fractions obtained with single-conditional averages around both PC1-scores and compared against the ones calculated with the mixture fraction. While adopting Z can more effectively decrease the spatial dependence of all the reactive scalars of Flame C, the normalized RMS of intermediate species and the fuel around both PC1-scores indicated a strong dependence on the flow dynamics and the burner's geometry. This suggests that the conditional averages are different and changing in function of space and Reynolds numbers, and as such, cannot sufficiently well separate the conditional manifold from the physical domain. Both scaling options equally well perform as Z for the conditional fluctuations of temperature, and the mass fractions of CO_2 , H_2O and NO , with slightly better results using AS. Interestingly, it was found that using a single principal component score (regardless of the scaling method) improves the tendency for Flame F. A single control variable based on mixture fraction is not able to detach the conditional averages from the real domain, with normalized RMS values exceeding the indicate 10% threshold used in this study. This indicates that a principal component well correlated with temperature can more effectively reduce the spatial gradient effects. The analysis was followed by comparing the conditional averages obtained with PCA against two-condition conditional averages around Z and four different progress variable definitions, c . For that purpose, only the dataset of Flame F was used. Here, it was shown that the conditional fluctuations using two-condition conditional averages are significantly improved, regardless of the progress variable definition. Using PC1-Pareto or PC1-AS promote higher normalized RMS than doubly conditioning. Surprisingly, the RMS values of certain scalars in Flame F do not deviate very much from two conditions. This suggests that a single principal component score, here correlated with temperature, can perhaps still get acceptable predictions for the reactive scalars of Flame F within a CSE framework. However, the normalized RMS of the mass fractions of the fuel, CO and H_2 still exhibit a clear functional dependence on the burner's characteristics, in particular for regions within the flow associated with

important levels of turbulence intensity and probabilities of localized extinction.

Overall, this study shows that coupling CSE and PCA can perhaps be a promising strategy for reduced-order modeling. Additional posterior research would however be required on premixed flames and combustion cases exhibiting complex flow structures to assess the modeling capabilities of this approach. It should also be mentioned that the modeling of the probability density function (PDF) of PC-scores remains an important unresolved issue that will require additional research.

Finally, among the four sets of doubly conditioning investigated, all combinations exhibited very similar behavior for all the scalars of Flame F, with conditional fluctuations and normalized RMS results of the same order of magnitude. The local averages of the conditional fluctuations were found to be equal to zero throughout all downstream locations, suggesting that the definition attributed to the progress variable is perhaps less relevant in a closure context than in an accurate representation of the chemical state.

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

Special thanks are given to Dr. Rob Barlow for the high-quality measurements of the Sandia/TUD burner and for making them available to the community. Dr. XiaoHao Fang gratefully acknowledges the financial support provided by the Natural Sciences and Engineering Research Council of Canada. This publication also arises from research funded by the John Fell Oxford University Press Research Fund.

References

- [1] M. Ihme, W.T. Chung, A.A. Mishra, Combustion machine learning: Principles, progress and prospects, *Prog. Energy Combust. Sci.* 91 (2022) 101010.
- [2] J.C. Sutherland, A. Parente, Combustion modeling using principal component analysis, *Proc. Combust. Inst.* 32 (1) (2009) 1563–1570.
- [3] A. Parente, J. Sutherland, L. Tognotti, P. Smith, Identification of low-dimensional manifolds in turbulent flames, *Proc. Combust. Inst.* 32 (1) (2009) 1579–1586.
- [4] B.J. Isaac, J.N. Thornock, J. Sutherland, P.J. Smith, A. Parente, Advanced regression methods for combustion modelling using principal components, *Combust. Flame* 162 (6) (2015) 2592–2601.
- [5] M.R. Malik, B.J. Isaac, A. Coussement, P.J. Smith, A. Parente, Principal component analysis coupled with nonlinear regression for chemistry reduction, *Combust. Flame* 187 (2018) 30–41.
- [6] H. Mirgolbabaei, T. Echehki, N. Smaoui, A nonlinear principal component analysis approach for turbulent combustion composition space, *Int. J. Hydrogen Energy* 39 (9) (2014) 4622–4633.
- [7] A. Biglari, J.C. Sutherland, An a-posteriori evaluation of principal component analysis-based models for turbulent combustion simulations, *Combust. Flame* 162 (10) (2015) 4025–4035.
- [8] A. Coussement, B.J. Isaac, O. Gicquel, A. Parente, Assessment of different chemistry reduction methods based on principal component analysis: Comparison of the MG-PCA and score-PCA approaches, *Combust. Flame* 168 (2016) 83–97.
- [9] O. Owoloye, T. Echehki, Toward computationally efficient combustion DNS with complex fuels via principal component transport, *Combust. Theory Model.* 21 (4) (2017) 770–798.
- [10] M.R. Malik, P. Obando 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 (2) (2021) 2635–2643.
- [11] R. Barlow, J. Frank, Effects of turbulence on species mass fractions in methane/air jet flames, *Symp. (Int.) Combust.* 27 (1) (1998) 1087–1095, Twenty-Seventh Symposium (International) on Combustion Volume One.
- [12] A. Biglari, J.C. Sutherland, A filter-independent model identification technique for turbulent combustion modeling, *Combust. Flame* 159 (5) (2012) 1960–1970.
- [13] M.R. Malik, A. Coussement, T. Echehki, A. Parente, Principal component analysis based combustion model in the context of a lifted methane/air flame: Sensitivity to the manifold parameters and subgrid closure, *Combust. Flame* 244 (2022) 112134.

- [14] R. Cabra, J.-Y. Chen, R. Dibble, A. Karpetis, R. Barlow, Lifted methane–air jet flames in a vitiated coflow, *Combust. Flame* 143 (4) (2005) 491–506, Special Issue to Honor Professor Robert W. Bilger on the Occasion of His Seventieth Birthday.
- [15] D. Veynante, L. Vervisch, Turbulent combustion modeling, *Prog. Energy Combust. Sci.* 28 (3) (2002) 193–266.
- [16] W.K. Bushe, H. Steiner, Conditional moment closure for large eddy simulation of nonpremixed turbulent reacting flows, *Phys. Fluids* 11 (7) (1999) 1896–1906.
- [17] J.W. Labahn, I. Stanković, C.B. Devaud, B. Merci, Comparative study between conditional moment closure (CMC) and conditional source-term estimation (CSE) applied to piloted jet flames, *Combust. Flame* 181 (2017) 172–187.
- [18] S.B. Pope, U. Maas, Simplifying chemical kinetics: Trajectory-generated low-dimensional manifolds, *Mech. Aerosp. Eng. Rep.: FDA* (1993) 93–11.
- [19] M. Wang, J. Huang, W. Bushe, Simulation of a turbulent non-premixed flame using conditional source-term estimation with trajectory generated low-dimensional manifold, *Proc. Combust. Inst.* 31 (2) (2007) 1701–1709.
- [20] M.M. Salehi, W.K. Bushe, K.J. Daun, Application of the conditional source-term estimation model for turbulence-chemistry interactions in a premixed flame, *Combust. Theory Model.* 16 (2) (2012) 301–320.
- [21] J.W. Labahn, C.B. Devaud, Species and temperature predictions in a semi-industrial MILD furnace using a non-adiabatic conditional source-term estimation formulation, *Combust. Theory Model.* 21 (3) (2017) 466–486.
- [22] A. Hussien, C. Devaud, Simulations of partially premixed turbulent ethanol spray flames using doubly conditional source term estimation (DCSE), *Combust. Flame* 239 (2022) 111651, A dedication to Professor Kenneth Noel Corbett Bray.
- [23] X. Fang, R. Ismail, N. Sekularac, M. Davy, On the prediction of Spray A end of injection phenomenon using conditional source-term estimation, in: *WCX SAE World Congress Experience*, SAE International, 2020.
- [24] X. Fang, R. Ismail, K. Bushe, M. Davy, Simulation of ECN diesel spray A using conditional source-term estimation, *Combust. Theory Model.* 24 (4) (2020) 725–760.
- [25] Z. Huang, M.J. Cleary, Z. Ren, H. Zhang, Large eddy simulation of a supersonic lifted hydrogen flame with sparse-Lagrangian multiple mapping conditioning approach, *Combust. Flame* (2021) 111756.
- [26] X. Fang, N. Sekularac, M.H. Davy, Parametric studies of a novel combustion modelling approach for low temperature diesel spray simulation, in: *Proceedings of the ASME 2020 Internal Combustion Engine Division Fall Technical Conference*, 2020, ASME 2020 Internal Combustion Engine Division Fall Technical Conference, V001T06A005.
- [27] A.H. Mahdipour, M.M. Salehi, Localized conditional source-term estimation model for turbulent combustion, *Combust. Flame* (2021) 111715.
- [28] N. Sekularac, X. Fang, W.K. Bushe, M.H. Davy, Conditional space evaluation of progress variable definitions for cambridge/sandia swirl flames, *Combust. Theory Model.* 27 (6) (2023) 736–767.
- [29] A. Hussien, C. Devaud, Simulations of turbulent acetone spray flames using the conditional source term estimation (CSE) approach, *Combust. Theory Model.* 25 (2) (2021) 269–292.
- [30] D. Dovizio, C. Devaud, Doubly conditional source-term estimation (DCSE) for the modelling of turbulent stratified V-shaped flame, *Combust. Flame* 172 (2016) 79–93.
- [31] C. Devaud, W.K. Bushe, J. Bellan, The modeling of the turbulent reaction rate under high-pressure conditions: A priori evaluation of the conditional source-term estimation concept, *Combust. Flame* 207 (2019) 205–221.
- [32] W.K. Bushe, C. Devaud, J. Bellan, A priori evaluation of the double-conditioned conditional source-term estimation model for high-pressure heptane turbulent combustion using dns data obtained with one-step chemistry, *Combust. Flame* 217 (2020) 131–151.
- [33] X. Fang, L. Shen, C. Willman, R. Magnanon, G. Virelli, M.H. Davy, R. Stone, Manifold reduction techniques for the comparison of crank angle-resolved particle image velocimetry (PIV) data and Reynolds-averaged Navier-Stokes (RANS) simulations in a spark ignition direct injection (SIDI) engine, *Int. J. Engine Res.* 23 (8) (2022) 1275–1294.
- [34] J. Shlens, A tutorial on principal component analysis, 2014, <http://dx.doi.org/10.48550/arXiv.1404.1100>.
- [35] R. Ranade, T. Echehki, A framework for data-based turbulent combustion closure: A posteriori validation, *Combust. Flame* 210 (2019) 279–291.
- [36] H. Mirgolbabaie, T. Echehki, Nonlinear reduction of combustion composition space with kernel principal component analysis, *Combust. Flame* 161 (1) (2014) 118–126.
- [37] T. Echehki, H. Mirgolbabaie, Principal component transport in turbulent combustion: A posteriori analysis, *Combust. Flame* 162 (5) (2015) 1919–1933.
- [38] T. Poinot, D. Veynante, *Theoretical and Numerical Combustion*, second ed., RT Edwards, Philadelphia PA, 2005.
- [39] K. Zdybał, J.C. Sutherland, A. Parente, Manifold-informed state vector subset for reduced-order modeling, *Proc. Combust. Inst.* 39 (4) (2023) 5145–5154.
- [40] K. Zdybał, E. Armstrong, A. Parente, J.C. Sutherland, PCAfold 2.0—Novel tools and algorithms for low-dimensional manifold assessment and optimization, *SoftwareX* 23 (2023) 101447.
- [41] A. Masri, R. Dibble, R. Barlow, The structure of turbulent nonpremixed flames revealed by Raman-Rayleigh-LIF measurements, *Prog. Energy Combust. Sci.* 22 (4) (1996) 307–362.
- [42] S.H. Starnes, R.W. Bilger, R.W. Dibble, R.S. Barlow, Piloted diffusion flames of diluted methane near extinction: Mean structure from Raman/Rayleigh fluorescence measurements, *Combust. Sci. Technol.* 70 (4–6) (1990) 111–133.
- [43] S.H. Starnes, R.W. Bilger, R.W. Dibble, R.S. Barlow, Piloted diffusion flames of diluted methane near extinction: Detailed structure from laser measurements, *Combust. Sci. Technol.* 72 (4–6) (1990) 255–269.
- [44] A. Karpetis, R. Barlow, Measurements of scalar dissipation in a turbulent piloted methane/air jet flame, *Proc. Combust. Inst.* 29 (2) (2002) 1929–1936.
- [45] R. Barlow, J. Frank, A. Karpetis, J.-Y. Chen, Piloted methane/air jet flames: Transport effects and aspects of scalar structure, *Combust. Flame* 143 (4) (2005) 433–449, Special Issue to Honor Professor Robert W. Bilger on the Occasion of His Seventieth Birthday.
- [46] A. Parente, J.C. Sutherland, Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity, *Combust. Flame* 160 (2) (2013) 340–350.
- [47] A. Mousemi, W.K. Bushe, S. Hochgreb, Evaluation of manifold representations of chemistry in stratified, swirl-stabilized flames, *Combust. Flame* 229 (2021) 111418.
- [48] W. Bushe, Spatial gradients of conditional averages in turbulent flames, *Combust. Flame* 192 (2018) 314–339.
- [49] P.A. Kalt, Y.M. Al-Abdell, A.R. Masri, R.S. Barlow, Swirling turbulent non-premixed flames of methane: Flow field and compositional structure, *Proc. Combust. Inst.* 29 (2) (2002) 1913–1919.
- [50] A. Masri, P. Kalt, R. Barlow, The compositional structure of swirl-stabilised turbulent nonpremixed flames, *Combust. Flame* 137 (1) (2004) 1–37.
- [51] R. Ranade, T. Echehki, A framework for data-based turbulent combustion closure: A priori validation, *Combust. Flame* 206 (2019) 490–505.
- [52] E. Armstrong, J.C. Sutherland, A technique for characterising feature size and quality of manifolds, *Combust. Theory Model.* 25 (4) (2021) 646–668.
- [53] 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.
- [54] 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.