Available online at www.sciencedirect.com

ScienceDirect

Proceedings of the Combustion Institute 000 (2020) 1–9

Proceedings of the Combustion Institute

www.elsevier.com/locate/proci

Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F

Mohammad Rafi Malik [a,b,*], Pedro Obando Vega [a,c], Axel Coussement [a], Alessandro Parente [a,b,*]

[a] *Université Libre de Bruxelles, Ecole Polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Bruxelles, Belgium*
[b] *Université Libre de Bruxelles and Vrije Universiteit Brussel, Combustion and Robust Optimization Group (BURN), Bruxelles, Belgium*
[c] *Institute of Energy and Power Plant Technology, Technische Universitaet Darmstadt, Darmstadt, Germany*

Received 7 November 2019; accepted 20 July 2020

Abstract

The present work shows the first application of the PC-transport approach in the context of Large Eddy Simulation (LES) of turbulent combustion. Detailed kinetic mechanisms, together with advanced computational tools, are needed to advance our knowledge of turbulent reacting systems. However, the cost related to high-fidelity simulations of turbulent reacting flows is still prohibitive for realistic configurations. Therefore, there is a need to reduce the complexity of the problem by identifying low-dimensional manifolds. To this end, the potential offered by Principal Component Analysis (PCA) in parameterizing the thermo-chemical state-space is very appealing. The present paper extends the PC-transport framework to three-dimensional Large Eddy Simulation (LES), coupling PCA with Gaussian Process Regression (GPR). To demonstrate the potential of the method, LES simulations of Sandia flames D, E and F are shown. Results show the great potential of the PC-GPR model, as indicated by the accuracy of the simulation results when compared with experimental data, using only 2 principal components. The sensitivity to the kinetic mechanism and subgrid closure model is also investigated.

Keywords: Combustion; Principal Component Analysis; Nonlinear regression; Large Eddy Simulation; Low-dimensional manifolds

* Corresponding author.
E-mail addresses: rafi.malik@ulb.ac.be (M.R. Malik), Alessandro.Parente@ulb.be (A. Parente).

https://doi.org/10.1016/j.proci.2020.07.014

1. Introduction

Technological breakthroughs in combustion technologies require a deep understanding of the different phenomena occurring in reacting systems. This can be achieved through the use of detailed kinetic mechanisms and advanced simulation tools. However, the simulation of turbulent reacting systems using detailed mechanisms is still very expensive. In order to lower the cost, one can parameterize the thermo-chemical state-space using a reduced number of optimal scalars using different approaches. Among them, Principal Component Analysis (PCA) [1] allows to represent a system using a reduced number of optimal variables by identifying low-dimensional manifolds. The potential of PCA for combustion has been investigated comprehensively by several groups in the past years [2–9] and has led in particular to the development of a combustion model based on PCA [3] and enhancements of the latter by combination of PCA with nonlinear regression techniques [8,9].

Starting from a data set containing the thermo-chemical state-space of a system of interest, PCA automatically reduces the size of the data set by identifying a new set of variables, the principal components (PCs), containing most of the variance present in the original data. Projecting the state-space on those PCs gives the PC scores, and transporting only a subset of those scores in a numerical simulation allows accelerating the simulation. Transport equations for the scores were introduced by Sutherland and Parente [3] (PC-score approach). The method was enhanced by combining PCA with nonlinear regression [7,10,11], to map the thermo-chemical source terms onto the new basis identified by the PCs, and thus maximize the reduction potential of the method. Isaac et al. [8] and Echekki and Mirgolbabaei [6] provided the first a posteriori studies on the use of the PC-score approach. In particular, Isaac et al. [8] demonstrated the potential of the PC-score approach coupled with Gaussian Process Regression (GPR) on an unsteady calculation of a perfectly stirred reactor (PSR) burning syngas. The method showed remarkable accuracy for the prediction of temperature and species, requiring only 2 transported variables instead of 11. Malik et al. [9] extended the study to methane and propane and showed its ability to produce very accurate representation of all state space variables using only 2 transported variables instead of 34 for methane, and 2 variables instead of 162 for propane.

Recently, the PC-score approach was employed for the simulation of Sandia flame F using one-dimensional turbulence (ODT) [6], the Direct Numerical Simulation (DNS) of premixed syngas [5] and methane-air combustion [12], and to develop a framework for closure models based on experimental data [13,14].

The aim of the present paper is to investigate the potential of PC-score coupled with nonlinear Gaussian Process Regression (GPR) in the framework of non-premixed turbulent combustion in a fully three-dimensional Large Eddy Simulation (LES). The use of GPR allows one to map the highly nonlinear source terms as well as other state-space variables (such as temperature, density, species mass fraction, viscosity) with a very low number of uncorrelated variables, identified using PCA. The database for model training is based on 1D counter diffusion methane flames. The approach is then validated using the experimental data available for Sandia flames D, E and F [15]. To the authors' knowledge, the current work is the first attempt to use such an approach.

2. PCA-based modeling

The main advantage of PCA in combustion applications is its ability to identify optimal scalars to parameterize low-dimensional manifolds. PCA offers the possibility of identifying patterns and correlations between variables in high dimension data sets. Once these correlations have been identified, the data set can be compressed by reducing the number of dimensions without much loss of information.

Following the methodology in [2,3,5,6,8,9,16], the principal component scores, $\mathbf{Z}$ $(n \times Q)$, are obtained by projecting the original data set $\mathbf{X}$ on the eigenvectors matrix $\mathbf{A}$ (also called the basis matrix):

$$\mathbf{Z} = \mathbf{XA}. \tag{1}$$

In combustion applications, $\mathbf{X} = [T, p, Y_1, Y_2, \ldots, Y_{ns}]$ is a matrix collecting n observations of a given thermo-chemical state-space, where Y_i is the mass fraction of species i, n_s the total number of species in the system, T its temperature and p the pressure. $\mathbf{X}$ is of dimension $(n \times Q)$ with n the number of observations and Q the number of independent variables. In the present work, it is assumed that $\mathbf{X}$ has been appropriately centered and scaled before PCA is carried out, to account for the different dimensions and units of the state variables, as discussed thoroughly in [17]. The reduction in dimensionality is obtained by truncating $\mathbf{A}$, i.e. by retaining only q PCs (with $q < Q$), noted $\mathbf{A_q}$. The original data set $\mathbf{X}$ can then be approximated using:

$$\mathbf{X} \approxeq \mathbf{X_q} = \mathbf{Z_q A_q^T}$$

where $\mathbf{X_q}$ is the approximation of $\mathbf{X}$ based on the first q eigenvectors of Q, and $\mathbf{Z_q}$ is the $(n \times q)$ matrix of the principal component scores. As the basis matrix contains orthonormal vectors, it can be noted that $\mathbf{A^{-1}} = \mathbf{A^T}$. In the present work, only the sample matrix of species mass fractions is employed to find the principal components

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014

($\mathbf{X} = [Y_i, ..., Y_{ns}]$), used in turn to parameterize the full termo-chemical state.

2.1. PC-score approach

Sutherland and Parente proposed a combustion model based on PCA and derived transport equations for the principal components [3]. Projecting the variables of interest, in this case the vector of species mass fractions $\mathbf{y} = [y_1, y_2, \ldots, y_{n_s}]$, onto the eigenvector matrix $\mathbf{A}$ gives:

$$\frac{\partial}{\partial t}(\rho \mathbf{z}) + \nabla(\rho u \mathbf{z}) = \nabla \cdot \mathbf{J_z} + \mathbf{S_z} \tag{2}$$

where $\mathbf{z} = \mathbf{Z_i^t}$ represents an individual score realization, ρ the density, $\mathbf{u}$ the velocity, $\mathbf{J_z}$ and $\mathbf{S_z}$ are the diffusive flux and chemical source terms of the principal components, respectively. Using Eq. (1), $\mathbf{J_z}$ and $\mathbf{S_z}$ can be linearly related to the diffusive fluxes and source terms of $\mathbf{y}$, $\mathbf{S}_z = \mathbf{A}^T\mathbf{S_y}$ and $\mathbf{J_z} = \mathbf{A}^T\mathbf{J_y}$, where $\mathbf{S_y}$ and $\mathbf{J_y}$ are the species source terms and diffusive fluxes, respectively. While the source terms can be directly retrieved from the PCA transformation, it was shown in [4,6,8,9,18] that the non-linearity of the source terms results in an error propagation that forces to increase significantly the number of components to be retained (and hence reduce the size reduction), thus justifying the use of non-linear regression approaches, as discussed in Section 2.2. As for the diffusive fluxes, expressing $\mathbf{J_y}$ and $\mathbf{J_z}$ as $\mathbf{J_y} = \rho \mathbf{D_y} \nabla \mathbf{y}$ and $\mathbf{J_z} = \rho \mathbf{D_z} \nabla \mathbf{z}$, and noticing that $\nabla \mathbf{y} = \mathbf{A} \nabla \mathbf{z}$, one can express the score matrix of diffusion coefficients as $\mathbf{D_z} = \mathbf{A}_q^T \mathbf{D_y} \mathbf{A}_q$, where $\mathbf{D_y}$ is the diagonal matrix of diffusion coefficients for species. The calculation of the score diffusion matrix $\mathbf{D_z}$ can be simplified relying on a unity Lewis number approximation [19]. This assumption was used in the present work, hence the matrix $\mathbf{D_z}$ is replaced by $\alpha = \mathbf{k}/(\rho\, \mathbf{c_p})$, where k is the thermal conductivity and c_p the specific heat capacity at constant pressure. Without the unity Lewis approximation, the score diffusion matrix can be directly related to the species one [5,20] and must be rotated to obtain a quasi-diagonal matrix of score diffusion coefficients. The final equation reads:

$$\frac{\partial}{\partial t}(\rho \mathbf{z}) + \nabla(\rho u \mathbf{z}) = \nabla \cdot \left(\frac{k}{c_p} \nabla \mathbf{z}\right) + \mathbf{S_z} \tag{3}$$

The number of score transport equations is reduced compared to the original set by taking a truncated matrix of eigenvectors $\mathbf{A}_q$ instead of $\mathbf{A}$.

2.2. Gaussian process regression

The state-space variables $\phi = (T, \rho, Y_1, Y_2, \ldots, Y_{n_s})$ and the PC source terms (S_{z_q}) are mapped to the PC basis using nonlinear regression in the form of:

$$\phi \approx f_\phi(\mathbf{Z}_q)$$

with f_ϕ being the nonlinear regression function and ϕ representing the dependent variables (i.e. Y_k, T, ρ and S_{z_q}). The function f_ϕ is obtained using Gaussian Processes (GPs) [21]. GPs have the advantage of not assuming beforehand a specific model for f_ϕ. The dependent variables are described by a gaussian distribution:

$$\phi \approx GP\,(m(x), \mathbf{K}(x, x^{'})$$

where m is a mean function and $\mathbf{K}$ is a covariance function (or kernel). The mean function is often assumed to be zero. The covariance function used in this work is the Squared Exponential:

$$\mathbf{K}(x, x^{'}) = \sigma_f^2\, exp\left[\frac{-(x - x^{'})^2}{2l^2}\right]$$

where σ_f^2 is the signal variance and l the characteristic length scale. σ_f^2 and l (the *hyper-parameters)* are first initialized and then optimized using a Gaussian likelihood function.

3. Training data and model generation

The global approach of the PC-GPR model can be summarized as follows: starting with a detailed kinetic mechanism and a canonical reactor, the reference data set is generated, with the same composition space as the system under study. PCA is then performed on the data, and the state-space variables are then regressed onto the PC basis. The model is then applied in a reactive flow simulation.

3.1. Experimental configuration

Flame D, E and F are three piloted methane-air diffusion flames with an axi-symmetric geometry. The burner consists of three coaxial jets. The main jet has a diameter $D = 7.2$ mm and the fuel consists of a mixture of CH_4 and air (25%/75% by volume). The fuel velocity is 49.6 m/s for flame D ($Re = 22,400$), 74.4 m/s for flame E ($Re = 33,600$) and 99.2 m/s for flame F ($Re = 44,800$), respectively. The fuel inlet temperature is 294 K. This main jet is surrounded by a pilot jet (⌀18.2 mm) at 1880 K, consisting of burnt gases (C_2H_2, H_2, air, CO_2 and N_2), and with a bulk velocity of 11.4 m/s (for flame D), 17.1 m/s (for flame E) and 22.8 m/s (for flame F, respectively). An air coflow with a velocity of 0.9 m/s and a temperature of 291 K surrounds the flame. The amount of local extinction increases from Flame D to F, with Flame F representing the most challenging test case, being close to global extinction. The flames have been experimentally investigated [15] through Rayleigh measurements for the temperature, and Raman and LIF measurements for mass fractions of chemical species. The

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014

available data consist of the mean and root mean square (rms) of temperature and mass fractions of major (CH_4, H_2, H_2O, CO_2, N_2 and O_2) and minor species (NO, CO and OH) at several axial locations. Laser Doppler Velocimetry (LDV) measurements of the velocity field are also available [22].

3.2. Reference data-set

High-fidelity data sets are required by PCA-based models in order to generate the PC basis and properly characterize the thermo-chemical state-space. Most of the time, the training data set is generated using a canonical configuration of interest for the system under investigation. In this study, an unsteady 1D laminar counter diffusion flame setup was used. Indeed, the most critical aspect when generating a training data-set is to make sure that the generated state-space includes all the possible states accessed during the actual simulation. Thus, a transient solver was preferred over a stationary one in order to cover all possible states from equilibrium to extinction (especially for flame F). The code used is part of the OpenSMOKE++ suite developed in Politecnico di Milano [23,24]. The GRI 3.0 [25] mechanism, involving 35 species and 253 reactions (excluding NO_x), was used. The inlet conditions, for the fuel on one side and air on the other, were set as in the experimental setup (see Section 3.1). The counterflow diffusion flames were pulsated with a sinusoidal profile, therefore allowing multiple simulations by varying the strain rate, from equilibrium to complete extinction. The unsteady solutions were saved on an uniform grid of 400 points over a 0.15 m domain. All of the unsteady data from the various simulations was used collectively for the PCA analysis. The final data set consisted of $\sim 80,000$ observations for each of the state-space variables.

3.3. Determination of the PCA basis

The PCA basis is generated using two approaches: (i) using the whole set of species (35) and (ii) using a subset of species. The latter has the advantage of removing certain scalars which may contribute to highly nonlinear source terms as shown in previous studies [7,8]. When a subset of species was used, the major species were adopted, namely CH_4, O_2, CO_2, H_2O and N_2 for the present case. The PCA analysis is carried out using PARETO scaling, which adopts the square root of the standard deviation as scaling factor. It was shown in [9] that PARETO scaling allows to obtain the greatest reduction for methane mechanisms and produces an easily regressible surface. A comparison of the results obtained with the full and reduced PCA basis is shown in Section 4.

The basis matrix weights obtained from the PCA analysis on the major species are as follows

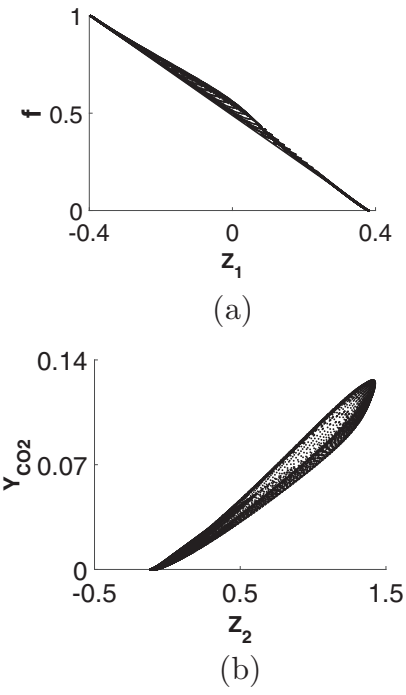


Fig. 1. Scatter plot; Z_1 is correlated with mixture fraction (a) and Z_2 with the progress of reaction (b).

for the first two PC's: $Z_1 = -0.02 \cdot Y_{H_2O} - 0.18 \cdot Y_{O_2} - 0.64 \cdot Y_{N_2} + 0.73 \cdot Y_{CH_4} - 0.02 \cdot Y_{CO_2}$ and $Z_2 = 0.51 \cdot Y_{H_2O} - 0.67 \cdot Y_{O_2} - 0.01 \cdot Y_{N_2} - 0.14 \cdot Y_{CH_4} + 0.5 \cdot Y_{CO_2}$. It can be seen that Z_1 has a large positive weight for CH_4 and a large negative value for the oxidizer (O_2 and N_2). This can be linked to the definition of Bilger's mixture fraction [26], f, as shown on Fig. 1(a). Therefore, in the numerical simulation, Z_1 is directly replaced by the mixture fraction, to avoid transporting a reactive scalar. The weights for Z_2 also show an interesting pattern: a positive correlation for H_2O and CO_2, and a negative correlation for CH_4, O_2 and N_2. This can be linked to a progress variable, where products have positive stoichiometric coefficients and reactants negative ones, as shown in Fig. 1(b). It is interesting to point out how PCA identifies these controlling variables without any prior assumptions or knowledge of the system of interest.

The nonlinear state-space variables (temperature, density, species mass fraction and PCs source terms) were regressed onto the linear PC basis using Gaussian Process regression (GPR). All variables were accurately regressed, with an $R^2 > 98.6\%$ for all source terms, species mass fraction, temperature and density. Fig. S1a shows source term S_{Z_2} as a function of Z_1 and Z_2, and Fig. S1b shows the regression of that manifold ($R^2 = 99.28\%$).

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014

4. Numerical setup

LES simulations were performed in OpenFOAM using a tabulated chemistry approach, in which the variables of interest (i.e. the PCs) are transported and the state-space (Y_k, T, ρ, S_{z_q}) is recovered from the nonlinear regression. The low-Mach Navier–Stokes equations were solved on an unstructured grid, together with the PCs transport equation (Eq. (2)). The state-space being accurately regressed using 2 PCs, the simulation was carried out using Z_1 and Z_2 as transported scalars. As Z_1 is highly correlated with the mixture fraction, the latter was transported instead. The boundary conditions for the PCs can be obtained using Eq. (1) [5]. A backward scheme was used for the time derivative and the Gauss linear scheme, with second order accuracy, was used for the divergence terms. The computational grid comprises 4 million hexahedra elements. The grid is conical, with a width of $7D$ at the inlet and $40D$ at the outlet, and a length of $80D$. The element size within the flame zone is 1.9×10^{-4} m. There is an injection pipe for the main jet, which extends $13D$ upstream the inlet. For the pilot, no inlet pipe is used. A turbulent inlet generator was used for both the fuel pipe (i.e. $13D$ upstream) and the pilot jets to provide the necessary turbulent fluctuations in the flow field. The turbulence generation is based on the digital filter method by Klein [27]. The grid was generated with an expansion ratio of 1.001 in the axial direction, and of 1.004 in the radial direction in order to obtain a fine resolution near the inlet nozzles. A mesh sensitivity analysis was carried out and the results proved to be mesh independent. On the current mesh, the Pope criterion is satisfied in the domain ($> 80\%$ in the flame region).

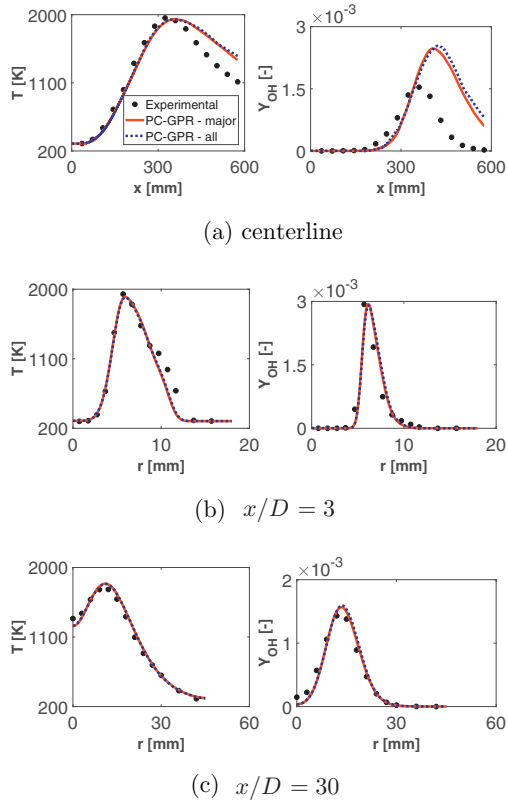
Three-dimensional Favre-averaged equations were solved for mass and momentum, together with Favre-averaged transport equations for the scores:

$$\rho \frac{D\widetilde{\mathbf{z}}}{Dt} = \nabla \cdot \left[\left(\frac{\mu}{Sc} + \frac{\mu_t}{Sc_t} \right) \nabla \widetilde{\mathbf{z}} \right] + \overline{\mathbf{S}}_{\mathbf{z}} \tag{4}$$

where the SGS Reynolds stresses were modelled using the WALE model [28] and the unity Lewis number was assumed for the species (thus the scores) The laminar and turbulent Schmidt number (respectively Sc and Sc_t) were set to 0.7. The laminar viscosity. μ, was tabulated in function of the PCs, while the turbulent viscosity, μ_t, was obtained through the WALE model. $\overline{\mathbf{S}}_{\mathbf{z}}$ is the filtered score source vector.

5. Results and discussion

The results of the PC approach on flames D–F are discussed in the present Section. The influence of the PC basis is first discussed, then the sensitivity to the mechanism used to generate the data-set is analyzed. Finally, the influence of a subgrid closure model on the thermo-chemical parameterization is assessed. The different simulations were run for at least 10 flow through periods, in order to have a sufficiently large averaging window.



(a) centerline

(b) $x/D = 3$

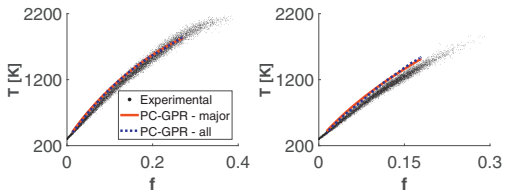
(c) $x/D = 30$

Fig. 2. Comparison between the PC basis calculated using the major species (PC-GPR – major) and the basis obtained using the full set of species (PC-GPR – all). Results show the axial (a) and radial profiles (b, c) for temperature (T) and OH mass fractions.

5.1. Full set vs reduced set

A comparison was made between the PCA basis containing the full set of species (35) and the basis computed on a reduced set of major species only (5). It can be observed in Fig. 2 that the PCA-GPR model is able to reconstruct all variables with great accuracy. Moreover, both PCA bases provided comparable results, at all locations. Whether looking at the centerline (Fig. 2a), close to the burner exit (Fig. 2b) or further downstream (Fig. 2c and d), the mean profiles obtained with the two bases do not show any significant discrepancy. It can be then argued that using only the major species in order to build the PC basis results in no major loss of information, supporting the findings in [6].

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014



(a) $x/D = 60$ (b) $x/D = 75$

Fig. 3. Conditional averages at different downstream positions for the PC basis using major species (PC-GPR – major) and the basis obtained using the full set of species (PC-GPR – all) plotted against the single shot experimental data.

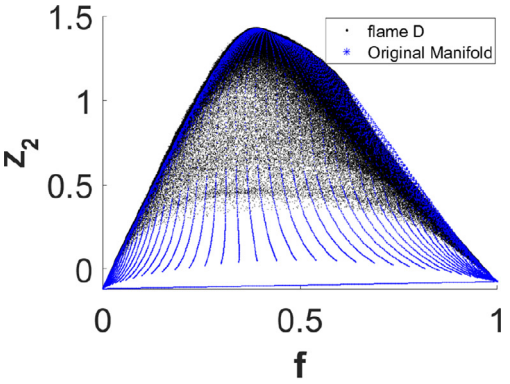
Fig. 2(a) shows that temperature is overpredicted on the centerline farther downstream. This can be due to an underestimation of the diffusion/mixing process at the outlet section. Fig. S2 in SM shows the same plots for mixture fraction and CO_2. A more significant comparison can be made looking at the profiles of temperature conditioned on mixture fraction at axial locations $x/D = 60$ and 75 ($x = 432$ and 540 mm) shown on Fig. 3. It can be observed that the predicted temperature lies well inside the single shot experimental data points.

At $x/D = 60$, the temperature agrees quite well with the experimental data, both on the lean side and near stoichiometry. At $x/D = 75$, the temperature lies slightly outside the single point data. Fig. S3 in Supplemental material shows the RMS centerline profiles for temperature, mixture fraction and species mass fraction.

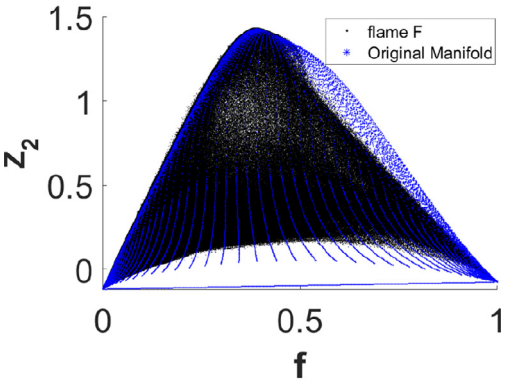
Fig. 4(a) shows the manifold accessed during the simulation with major species at $t = 1$ s, plotted against the original manifold obtained from the training data-set. It can be observed that the simulation did not leave the training manifold: all the points accessed are bounded inside the original training manifold. It is also apparent that most of the data is contained near the equilibrium solution, showing that for flame D the simulation did not experience significant extinction and re-ignition.

5.2. Model sensitivity to the chemical mechanism and subgrid closure

The impact of the kinetic mechanism was also assessed. The GRI 3.0 mechanism was compared to the KEE-58 mechanism [26]. The latter consists of 17 species and 58 reactions (excluding N containing species except N_2). The PCA basis was once again computed based on the same reduced set of species (CH_4, O_2, CO_2, H_2O and N_2). A GPR regression was carried out for the entire thermochemical state-space, and a table was generated using the same grid spacing. Fig. 5 shows a comparison of different axial and radial profiles using the



(a) flame D vs original manifold



(b) flame F vs original manifold

Fig. 4. Scatter plot of the PCA manifold using two PCs: the original manifold obtained for the training data-set plotted against the one represented during the simulation for flame D (a) and flame F (b). Points were downsampled for clarity.

GRI 3.0 mechanism and the KEE-58. It can be observed that overall the GRI performs better than the KEE, predicting the temperature and species mass fraction peaks more accurately. This suggests that the level of accuracy and detail in the kinetic mechanism is not lost during the construction of the PC-GPR model. Thus, a PC model trained on a more detailed mechanism will result in better a posteriori predictions.

The sensitivity to a subgrid closure was also investigated. A mean value closure for the filtered PC's source terms ($\overline{\mathbf{S}}_\mathbf{z}$) might not be sufficient, and the influence of small-scale turbulent fluctuations on the large scales must be assessed. Therefore, a *beta*-shaped probability density function ($\boldsymbol{\beta}$-PDF) was used to represent the necessary scalar fluctuations. A transport equation for the mixture fraction

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014

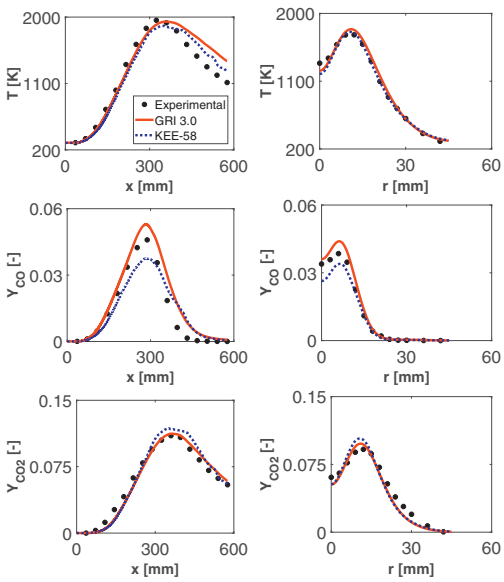
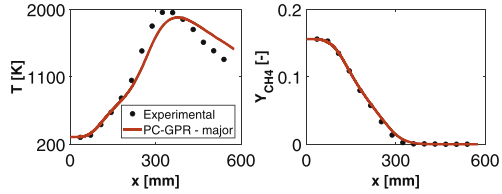
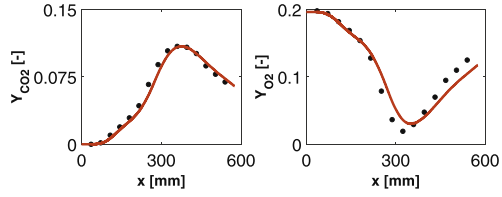


Fig. 5. Comparison between GRI 3.0 and KEE-58 mechanisms on the centerline (left) and at radial location $x/D = 30$ (right) for temperature, CO and CO_2 mass fractions.

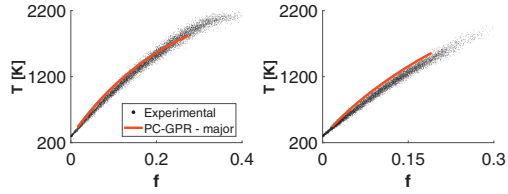


(a) Temperature (left) and CH_4 mass fraction (right)



(b) CO_2 (left) and O_2 mass fraction (right)

Fig. 6. Flame F: temperature and major species profiles plotted against the experiments – centerline.



(a) $x/D = 60$ (b) $x/D = 75$

Fig. 7. Flame F – conditional averages at different downstream positions plotted against the single shot experimental data.

variance was used following the approach in [29]. Results shown in Supplemental Material (Fig. S4) indicate no major influence of the subgrid model. While the filtered equations were resolved, the effect of the SGS terms appears to be negligible compared to the resolved part of the flow due to the high resolution. This suggests that z_1 and z_2 are well resolved by the relatively fine grid used. However, the effect of subgrid closure should be further investigated.

5.3. Flame F: results and discussion

To demonstrate the potential of the PC-GPR on more challenging cases, a simulation of Sandia flame E and F was carried out. The geometry and numerical setup were identical to the ones of flame D. The regression table based on the subset of species was used, and only 2 PCs were transported. The velocity boundary conditions were adapted to match the experimental setup. The simulation was run for at least 10 flow through periods. Only the results associated to flame F are shown in this Section, while Fig. S5 in SM shows the temperature and some species mass fraction profiles on the centerline and radial profiles at different axial locations for flame E, confirming the ability of the model to reconstruct all scalar variables with great accuracy. Fig. 6 shows the comparison between the experimental and numerical profiles of temperature and selected species mass fraction profiles on the centerline for flame F. It can be observed that the PC-GPR model can accurately predict the peak and the decay in temperature and species mass fraction profiles. Fig. 7 shows the profiles of temperature conditioned on mixture fraction at axial locations $x/D = 60$ and 75 plotted against the single shot data. It can be observed that the predicted temperature lies well inside the experimental data. The same conclusion can also be drawn looking at the profiles in Fig. S6 of the SM, showing the profiles of conditional mean of temperature and species mass fraction on mixture fraction. Furthermore, as expected from the experimental data, it can be observed from Fig. 4(b) that flame F experiences high levels of extinction and re-ignition. It is apparent that the region of the manifold accessed during the simulation is wider compared to the flame D one, and that the data is evenly distributed between the equilibrium solution and the extinction region of the manifold.

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014

Fig. S7 in SM shows radial statistics for flame F at different axial locations for temperature and selected species. It can be observed that the PC-GPR model accurately predicts the peak and decay in radial directions as well.

6. Conclusion

This study presents the first application of the PC-score approach coupled with nonlinear Gaussian Process Regression (GPR) on a 3D LES simulation of the Sandia flames D–F.

The PC-GPR model showed very good accuracy when compared with experimental data using only 2 components, instead of the 35 species present in the GRI 3.0 mechanism. The first PC was found to be highly correlated with mixture fraction, thus allowing to transport directly the latter, instead of a reacting scalar which would have required to model an additional source term. Moreover, results showed that the PCA basis can be constructed using only a subset of species, containing most of the information of the system. The PCs remained bounded to the training manifold during the simulation, indicating that the choice of an unsteady canonical reactor ensures to span all the potential chemical states accessed during the simulation.

The proposed model also showed very good accuracy for the prediction of flames E and F, despite the increasing complexity. The PC-GPR model was able to handle the extinction and reignition phenomenon properly, and thus showing the importance of including unsteady data in the training manifold. Indeed, as the counterflow diffusion flames were pulsated with a sinusoidal profile, the database includes flames that both ignite and extinguish.

The strength of the method resides in the fact that PCA does not require any prior selection of variables. Instead, it automatically extracts the most relevant variables to describe the system of interest. From this perspective, the PC-GPR method can be regarded as a generalization of tabulated chemistry approaches, particularly for complex systems requiring the definition of a larger number of progress variables. Future work will focus on this aspect, considering also the inclusion of additional canonical reactors in the database generation, including partially stirred reactors and pairwise mixing stirred reactor approaches [14].

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 project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 714605.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.proci.2020.07.014.

References

[1] I. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 2002.
[2] A. Parente, J.C. Sutherland, L. Tognotti, P.J. Smith, *Proc. Combust. Inst.* 32 (1) (2009) 1579–1586.
[3] J.C. Sutherland, A. Parente, *Proc. Combust. Inst.* 32 (2009) 1563–1570.
[4] A. Biglari, J.C. Sutherland, *Combust. Flame* 159 (5) (2012) 1960–1970.
[5] A. Coussement, B. Isaac, O. Gicquel, A. Parente, *Combust. Flame* 168 (2016).
[6] T. Echekki, H. Mirgolbabaei, *Combust. Flame* 162 (5) (2015) 1919–1933.
[7] H. Mirgolbabaei, T. Echekki, *Combust. Flame* 160 (2013) 898–908.
[8] B. Isaac, J. Thornock, J. Sutherland, P. Smith, A. Parente, *Combust. Flame* 162 (6) (2015) 2592–2601.
[9] M.R. Malik, B. Isaac, A. Coussement, P. Smith, A. Parente, *Combust. Flame* 187 (2018) 30–41.
[10] Y. Yang, S. Pope, J. Chen, *Combust. Flame* 160 (2013) 1967–1980.
[11] S.B. Pope, *Proc. Combust. Inst.* 34 (2013) 1–31.
[12] O. Owoyele, T. Echekki, *Combust. Theory Model.* 21 (2017) 770–798.
[13] R. Ranade, T. Echekki, *Combust. Flame* 206 (2019) 490–505.
[14] R. Ranade, T. Echekki, *Combust. Flame* 210 (2019) 279–291.
[15] R.S. Barlow, J.H. Frank, *Proc. Combust. Inst.* 27 (1998) 1087–1095.
[16] A. Biglari, J.C. Sutherland, *Combust. Flame* 162 (10) (2015).
[17] A. Parente, J.C. Sutherland, *Combust. Flame* 160 (2) (2013) 340–350.
[18] B. Isaac, A. Coussement, O. Gicquel, P. Smith, A. Parente, *Combust. Flame* 161 (2014).
[19] R. Barlow, J. Frank, A. Karpetis, J.-Y. Chen, *Combust. Flame* 143 (2005).
[20] H. Mirgolbabaei, T. Echekki, *Combust. Flame* 161 (2014) 118–126.
[21] C. Rasmussen, *Gaussian processes for machine learning, The MIT Press*, 2006.
[22] C. Schneider, A. Dreizler, J. Janicka, E. Hassel, *Combust. Flame* 135 (2003) 185–190.
[23] A. Cuoci, A. Frassoldati, T. Faravelli, E. Ranzi, *Comput. Phys. Commun.* 192 (2015) 237–264.
[24] A. Cuoci, A. Frassoldati, T. Faravelli, E. Ranzi, *Energy Fuels* 27 (12) (2013) 7730–7753.

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014

[25] G.P. Smith, D.M. Golden, M. Frenklach, N.W. Moriarty, B. Eiteneer, M. Goldenberg, C.T. Bowman, R.K. Hanson, S. Song, W.C.G. Jr., V.V. Lissianski, Z. Qin, 1999, [link]. http://www.me.berkeley.edu/grimech/.
[26] R. Bilger, S. Starner, R. Kee, *Combust. Flame* 80 (1990) 135–149.
[27] M. Klein, A. Sadiki, J. Janicka, *J. Comput. Phys.* 186 (2) (2003) 652–665.
[28] F. Nicoud, F. Ducros, *Flow Turbul. Combust.* 62 (3) (1999) 183–200.
[29] M. Ihme, H. Pitsch, *Combust. Flame* 155 (2008) 90–107.

Please cite this article as: M.R. Malik, P. Obando Vega and A. Coussement et al., Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.07.014