ÉCOLE POLYTECHNIQUE DE BRUXELLES
UNIVERSITÉ LIBRE DE BRUXELLES

Reduced-order Combustion Models for Innovative Energy Conversion Technologies

Thesis presented by Mohammad Rafi Malik
in fulfilment of the requirements of the PhD Degree in Engineering Sciences and Technology (”Docteur en Sciences de l’ingénieur et technologie”)

Academic year 2020-2021

Supervisor : Professor Alessandro PARENTE

Thesis jury :
Axel COUSSEMENT (Université libre de Bruxelles, Chair)
Gérard DEGREZ (Université libre de Bruxelles, Secretary)
Patrick HENDRICK (Université libre de Bruxelles)
Tarek ECHEKKI (North Carolina State University, USA)
Hong IM (King Abdullah University of Science and Technology, KSA)

DECLARATION

This thesis is submitted to the Université Libre de Bruxelles (ULB) for the degree of philosophy doctor. This doctoral work has been performed at the Université Libre de Bruxelles, École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Bruxelles, Belgium, with Professor Alessandro Parente.

M. Rafi Malik

December 2020

ABSTRACT

The present research seeks to advance the understanding and application of Principal Component Analysis (PCA)-based combustion modelling for practical systems application. This work is a consistent extension to the standard PC-transport model, and integrates the use of Gaussian Process Regression (GPR) in order to increase the accuracy and the potential of size reduction offered by PCA. This new model, labelled PC-GPR, is successively applied and validated in *a priori* and *a posteriori* studies.

In the first part of this dissertation, the PC-GPR model is validated in an *a priori* study based on steady and unsteady perfectly stirred reactor (PSR) calculations. The model showed its great accuracy in the predictions for methane and propane, using large kinetic mechanisms. In particular, for methane, the use of GPR allowed to model accurately the system with only 2 principal components (PCs) instead of the 34 variables in the original GRI-3.0 kinetic mechanism. For propane, the model was applied to two different mechanisms consisting of 50 species and 162 species respectively. The PC-GPR model was able to achieve a very significant reduction, and the thermo-chemical state-space was accurately predicted using only 2 PCs for both mechanisms.

The second part of this work is dedicated to the application of the PC-GPR model in the framework of non-premixed turbulent combustion in a fully three-dimensional Large Eddy Simulation (LES). To this end, an *a posteriori* validation is performed on the Sandia flames D, E and F. The PC-GPR model showed very good accuracy in the predictions of the three flames when compared with experimental data using only 2 PCs, instead of the 35 species originally present in the GRI 3.0 mechanism. Moreover, the PC-GPR model was also able to handle the extinction and re-ignition phenomena in flames E and F, thanks to the unsteady data in the training manifold. A comparison with the FPV model showed that the combination of the unsteady data set and the best controlling variables for the system defined by PCA provide an alternative to the use of steady flamelets parameterized by user-defined variables and combined with a PDF approach.

The last part of this research focuses on the application of the PC-GPR model in a more challenging case, a lifted methane/air flame. Several key features of the model are investigated: the sensitivty to the training data set, the influence of the scaling methods, the issue of data sampling and the potential of a subgrid scale (SGS) closure. In particular, it is shown that the training data set must contain the effects of diffusion in order to accurately predict the different properties of the lifted flame. Moreover, the kernel density weighting method, used to address the issue of non-homogenous data density usually found in numerical data sets, allowed to improve the predictions of the PC-GPR model. Finally, the integration of subgrid scale closure to the PC-GPR model allowed to significantly improve the simulations results using a presumed PDF closure. A qualitative comparison with the FPV model showed that the results provided by the PC-GPR model are overall very comparable to the FPV results, with a reduced numerical cost as PC-GPR requires a 4D lookup table, instead of a 5D in the case of FPV.

PUBLICATIONS

1. Mohammad Rafi Malik, Alessandro Parente, *Principal Component Analysis based combustion model in the context of a lifted methane/air flame: sensitivity to the manifold parameters and subgrid closure*, submitted to Combustion and Flame, 2021.

2. Mohammad Rafi Malik, Pedro Obando Vega, Axel Coussement, Alessandro Parente, *Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F*, Proceedings of the Combustion Institute, 38 (2021). In Press.

3. Mohammad Rafi Malik, Benjamin Isaac, Axel Coussement, Philip Smith, Alessandro Parente, *Principal component analysis coupled with nonlinear regression for chemistry reduction*, Combustion and Flame 187 (2018), 30-41.

4. Zhiyi Li, Mohammad Rafi Malik, Alberto Cuoci, Alessandro Parente, *Edcsmoke: A new combustion solver for stiff chemistry based on OpenFOAM®*, AIP Conference Proceedings, 1863 (2017).

5. Alessandro Parente, Mohammad Rafi Malik, Francesco Contino, Alberto Cuoci, Bassam Dally, *Extension of the Eddy Dissipation Concept for turbulence/chemistry interactions to MILD combustion*, Fuel 163 (2016), 98-111.

For my parents.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to take this opportunity to acknowledge and express my thanks to those who helped me with the various aspects of conducting this research and reaching this accomplishment.

First of all, I would like to express my deep gratitude to Alessandro for his guidance, patience, and for pushing me to pursue the path of academic research. He inspired me to work with dedication, and helped me building the confidence I needed to learn and grow as a researcher. I am very grateful for the patience he has had with me and for helping me to achieve my goals. I am grateful for the time that I have had and will continue to have under his leadership.

I wish to acknowledge Prof. Phil Smith and his research group at the University of Utah for welcoming me in their facilities and giving me the opportunity to work with them. I would especially like to acknowledge Ben Isaac for his great help and support during my stay in Utah. Ben helped me in the beginning of my research by teaching me everything he learned during his PhD, so that I could benefit from his work and experience and be able to carry on my research independently.

I wish to acknowledge my parents for their continious support and encouragement.

I would also like to thank my colleagues for their help during the various stages of my research.

CHAPTER 1

INTRODUCTION

1.1 Background

Energy generation through combustion of hydrocarbons is likely to continue as the dominant energy conversion technology for at least the next century. Its efficiency and emissions have thus an important impact on global resources, environmental quality and climate change.

At present, chemical energy derived from combustion of fossil fuels (coal, petroleum or natural gas) supplies a disproportionally large fraction of the total world energy needs. And this, despite the large variety of alternate energy sources available (such as nuclear, wind, solar ...). This monopoly of combustion can be explained by its convenience and high-energy density. Indeed, on one hand, the intermittent nature of renewable sources doesn't allow their direct use in some applications, such as air and ground transportation, where high energy density and continuous output are required. On the other hand, energy density is the essential feature that places fossils fuels at the center of highly demanding energy applications, such as transportation and industrial processes.

Our local environment shows us the importance of combustion in our daily life: from the big industries (such as blast furnaces) to small scale engines used in our every day vehicles, combustion is sustaining our modern lifestyle and also improving our quality of life.

Because of this high demand, there is a continued need for improvement, enhancement, and even understanding of the combustion process as the dual challenges of energy and climate change are highlighting the need for new and improved combustion technologies. It is therefore necessary to carry fundamental studies bringing together available experimental information with simulation approaches, to drive the development of modern

energy conversion technologies, and to make best use of existing resources while limiting their environmental impact.

1.2 Reduced-order modeling in turbulent combustion

1.2.1 Problem statement

Recently, combustion modeling has become an important tool in gaining deeper insights and understanding complex systems. The continuous increase in computational power during the last years has been driving research in new combustion models, providing combustion engineers with increasingly sophisticated modelling options that can take full advantage of the latest computational resources.

Considering the complex nature of the phenomena involved, the use of Computational Fluid Dynamics (CFD) is acknowledged to be essential for the development of such novel combustion technologies. In particular, CFD calculations can be applied directly at the industrial scale of interest, thus avoiding scaling-up the results from lab-scale experiments. However, the optimal integration of detailed kinetics within CFD calculations still requires major developments.

The numerical modeling of turbulent combustion is a very challenging task as it combines the complex phenomena of turbulence and chemical reactions. This study becomes even more challenging when large detailed kinetic mechanisms are used in order to understand some special features such as pollutant formation. A detailed combustion mechanism for a simple fuel such as methane involves 53 species and 325 chemical reactions [60]. Moreover, the number of species and reactions increases with increasing fuel complexity. The coupling of the kinetic equations with the set of Navier-Stokes equations results in a problem that is too complex to be solved by the current computational means.

In a CFD calculation, the number of species tracked impacts the memory usage and CPU time. It is thus important to minimize this number by the use of a simpler but representative set of variables. Therefore, there is a need for methods allowing to parameterize efficiently the thermo-chemical state of a reacting system with a reduced number of optimal reaction variables.

1.2.2 Methods of reduction of kinetic models

The development of methods for reduction of chemical mechanisms is driven by the demand for a speedup in computational time. For the most complex CFD simulations, a reduced mechanism is not an option but a necessity [37]. Therefore, a variety of reduction techniques have been developed over the past decades in order to reduce the overall CPU time and memory requirement.

A reduced kinetic model can be constructed based on two methods: build up the kinetic model starting from formal structures for a specific type of oxidation process, or start from a comprehensive model and then select only the most important species and reactions for the reduced scheme [25]. Detailed kinetic schemes have been extensively developed over the past decades by a number of different groups. With the advent of new computational tools and artificial intelligence, novel procedures for automatic generation of large schemes have also been proposed [25]. This work instead focuses on methods that lead to the reduction of large scale models, without major loss to its qualitative or quantitative potentials.

The remaining part of this Section presents an overview of various reduction techniques that are of most interest. One can distinguish two ways of simplifying the chemical kinetics. On one hand, the traditional methods such as the *Quasi Steady-State Approximation* (QSSA) can be identified. On the other hand, reduction algorithms based on mathematical methods can also be used, such as *sensitivity analysis,* the *Intrinsic Low-Dimensional Manifolds* (ILDM) algorithm, the *Computational Singular Perturbation* (CSP) algorithm, the *Rate-Controlled Constrained Equilibrium* (RCCE) methodology, and *Principal Component Analysis* (PCA).

1.2.2.1 Sensitivity analysis

Sensitivity analysis was one the prime methods used for the reduction of large kinetic schemes, and gained a lot of importance in the early 1980s [25]. The idea is to remove redundant species from the scheme through the sensitivity of the rate of production of important species to a change in concentration of another species i [9]:

$$B_i = \sum_{n=1}^{N} \left(\frac{c_i}{f_n} \frac{\partial f_n}{\partial c_i} \right)^2$$

where c_i is the concentration of species i, N is a group of necessary species and f_n the net rate of production of species n. The first step is to define a group of important species (including key products such as H_2O and CO_2). Then, all the species which fall above a defined tolerence value are also admitted to N. After that, the B_i index is iteratively calculated until no new species are admitted in N.

The magnitude of the sensitivity coefficient provides an indication of how strongly the behaviour of the system is affected by that species. If the production rate shows little or no sensitivity to a perturbation, that species can be identified as unimportant, and can thus be eliminated from the kinetic scheme for that particular system. It should also be pointed out that sensitivity analysis is a local method. Therefore, in order to accurately represent the dynamics of the full system using such method, it must be applied over a range of conditions and the necessary species and reactions must be either combined into a single reduced mechanism or applied adaptively [9].

Sensitivity analysis allows to create a skeleton mechanism where redundant species and reactions have been removed. However, in many cases the level of reduction achieved might still not be sufficient for an application within complex flow environments [9]. Further reduction may be achieved using time scale based methods, such as the *Quasi Steady-State Approximation* (QSSA), the *Intrinsic Low-Dimensional Manifolds* (ILDM) and the *Computational Singular Perturbation* (CSP).

1.2.2.2 QSSA

QSSA methods [Peters, [37] are the simplest of the time scale based methods, and are based on the idea that some species in the system react on very short time scales, thus allowing to assume that their concentrations have reached an equilibrium state compared to the other slow reacting species. Therefore, the source terms of those steady state species are simply set to zero, allowing to replace the differential equation describing the species conservation equations with an algebraic equation. The main challenge behind this method is to find a reliable procedure that identifies those short time scales species. A common way is to use the instantaneous QSSA error for a single species concentration (Δc_i), defined as [25]:

$$\Delta c_i = \frac{1}{J_{ii}} \frac{dc_i}{dt}$$

where J_{ii} is the diagonal element of the chemical Jacobian for species i. Calculating the QSSA errors for all the species at different times allows to sort them based on their maximum errors. Then, choosing a certain threshold leads to the selection of the suitable species.

When working with highly coupled systems, the fast time scales cannot be always directly associated with particular chemical species [9]. Therefore, geometric based methods such as the ILDM could provide with a greater reduction than QSSA. Moreover, in the traditional methods of QSSA or sensitivity analysis presented above, the selection of species must be carried out individually for every mechanism. This procedure can quickly become unmanageable when dealing with large mechanisms and complex fuels. Therefore, systematic methods based on time scales analysis were developed, such as the Computational Singular Perturbation (CSP) which is described subsequently.

1.2.2.3 ILDM

The Intrinsic Low-Dimensional Manifold method [38] has proved to be a useful method to simplify detailed reaction mechanisms. The slow and fast time scales are separated based on the dynamical systems approach, using a timescale analysis based on the Jacobian of the chemical source terms. The ILDM method is based on the chemical source term only since transport processes are not included in the manifold. Considering a species conservation equation of the form:

$$\rho \frac{d\mathbf{Y}}{dt} = \mathbf{s}(\mathbf{Y})$$

where ρ is the density, $\mathbf{Y}$ the species mass fraction vector and $\mathbf{s}$ the vector with the chemical source terms, a linearization of the system is performed around a reference point $\mathbf{Y^0}$, leading to:

$$\rho \frac{d}{dt}(\mathbf{Y} - \mathbf{Y^0}) \cong \mathbf{J}(\mathbf{Y} - \mathbf{Y^0}) = \mathbf{U}\Lambda\mathbf{U^{-1}}(\mathbf{Y} - \mathbf{Y^0})$$

where the elements of the Jacobian $\mathbf{J}$ are $J_{ij} = {}^{\partial s_i}/_{\partial Y_j}$. Λ and $\mathbf{U}$ represent the matrices of the eigenvalues and corresponding right eigenvectors, respectively. If subjected to a small perturbation, the system can react in three different ways [38]:

1. If the perturbation is in the direction of an eigenvector whose corresponding eigenvalue is zero, the perturbation will not change with time. This usually corresponds

to conserved variables, i.e. the element fractions.

2. If the perturbation in the direction of an eigenvector whose eigenvalue has a negative real part, the perturbation will dampen to zero. This corresponds to fast damping processes, which can be assumed to be in steady state.

3. If the perturbation is in the direction of an eigenvector whose corresponding eigenvalue has a positive real part, the perturbation will increase. This corresponds to the slow processes, which describe the evolution of the chemical system.

The ILDM method will then find the points in the state-space where the chemical source term has a component only in the direction of the slow processes. Compared to conventional reduction methods, such as the QSSA, the reduction is performed automatically and the only required inputs are the detailed kinetic mechanism and the chosen dimension of the manifold. However, as diffusion is omitted in the construction of the manifold, the latter will show reduced accuracy in regions where both chemistry and diffusion are significant, which is generally the case in a large part of the reaction zone [7]. Moreover, if the slow and fast processes are not well separated (which is usually the case in simulations with complex fuels), the dimension of the manifold must be increased in order to provide reasonable accuracy, which in turn leads to a less effective reduction. To tacke these problems, a flame prolongation of ILDM [22] can be used.

1.2.2.4 CSP

Lam and Goussis [36] presented the Computational Singular Perturbation (CSP) method for the reduction of complex mechanisms. Originally, the CSP method was developed for chemical kinetics equations, but can also be applied to other time scale problems.

A chemical kinetic system of equations can be written as a set of ordinary differential equation (ODE):

$$\frac{d\mathbf{y}}{dt} = \mathbf{g}(\mathbf{y})$$

where $\mathbf{y}$ is the vector of species concentrations and $\mathbf{g}(\mathbf{y})$ the global reaction rate vector. CSP calculates the eigenvalues of the Jacobian of the source term and separates them as fast and slow modes. It then uses a CSP pointer to identify the species mostly related to those time

scales. At any given time and space of the calculation domain, $\mathbf{g}(\mathbf{y})$ is expanded in terms of basis vectors and the pointer looks for the basis vector that shows the largest gradient. This basis vector can be associated with the current largest eigenvalue of the current Jacobian of $\mathbf{g}(\mathbf{y})$. The reactions contributing to this vector are classified into the *fast reaction group.* When the contribution of this fast reaction group to $\mathbf{g}(\mathbf{y})$ becomes numerically insignificant, it is discarded from $\mathbf{g}(\mathbf{y})$. The process is then repeated with the rest of the $\mathbf{g}(\mathbf{y})$ vector, and the next fast reaction group is identified and its contribution to $\mathbf{g}(\mathbf{y})$ is monitored and eventually discarded, etc [36]. Therefore, by this process of identification and discard of fast reaction groups, a simplified kinetic scheme for the complex chemical system under consideration is obtained.

The CSP method has become very attractive recently thanks to the availability of mathematical algorithms that are able to compute the basis vectors at any given time (on-the-fly). Moreover, they are able to do this without the need for experience or intuition from the user. The user can therefore include all the relevant elementary reactions for a particular system in the model, and use CSP to identify the set of reactions governing the system at any given time.

1.2.2.5 RCCE

Rate-Controlled Constrained Equilibrium (RCCE) allows to generate reduced mechanisms based on an alternative approach (when compared with traditional QSSA method): a subset of the major species is chosen to drive the ODE integration, while the minor species are recovered based on algebraic equations derived from the minimisation of the free energy. The choice of the major species to retain can be done by analyzing the magnitude of the species concentrations across the flame. Therefore, the dynamics of the chemical system are led by a subset of ODEs from the detailed mechanism, while the species exhibiting fast time scales (which can be regarded as in equilibrium compared to the major ones) are obtained from the minimisation of the free energy, which define a manifold of constrained equilibrium states and force the system to remain there [20].

RCCE was first proposed by Keck and Gillespie [33, 34] and applied to a variety of cases by several groups [20]: the coupling of RCCE with the In-Situ Adaptive Tabulation (ISAT) in the context of laminar non-premixed and premixed flames; in turbulent flames where

RCCE was coupled with Conditional Moment Closure (CMC) and Probability Density Function (PDF), and also with ISAT. RCCE has been applied in turbulent flames both through tabulation and direct integration methods [20].

RCCE employs the concept of the minimisation of free energy to determine the equilibrium concentration of a chemical mixture. The system of equations describing the constrained equilibrium states are derived by minimising the Gibbs free energy, subject to the conservation of the elements, enthalpy, mass and constrained species (N_c):

$$\mu_j^0 + RT \ln \frac{n_j}{n} + RT \ln \frac{P}{P_0} + \sum_{i=1}^{N_e} \lambda_i^e a_{ij}^e + \sum_{i=1}^{N_c} \lambda_i^c a_{ij}^c = 0 \tag{1.1}$$

where μ_j^0 and P_0 are respectively the pressure and chemical potential at the standard state, n_j represents the concentration of chemical species j, a_{ij}^e and a_{ij}^c are the matrices that relate the constrained species and elements with the overall concentrations, λ_i^e and λ_i^c are the Lagrange multipliers for the elements and constrained species respectively, and N_e represents the number of atomic elements. Equation 1.1 is solved together with the ODEs for the constraints using a detailed mechanism describing N species and N_r reactions:

$$\frac{dC_i}{dt} = \sum_{j=1}^{N} a_{ij}^c \left[\sum_{k=1}^{N_r} \nu_{jk} r_k(n_1, n_2, ..., n_N, T, p) \right] (i = 1, ..., N_c) \tag{1.2}$$

where ν_{jk} represents the stoichiometric factors and r_k the individual reaction rates. Equations 1.1 and 1.2 form a differential–algebraic (DAE) problem in which the dependent variables are the species, Lagrange multipliers, temperature and density.

1.2.3 Flamelet Generated Manifold and Flamelet Prolongation of ILDM

The chemical reduction methods presented above are based on the idea that most of the chemical time scales in the system are very small. Neglecting transport processes, a time-scale analysis allows to separate the fastest time scales from the slow ones, the fastest being assumed to be in steady-state.

The laminar flamelet model [52] on the other hand is based on the idea that flame structures are much thinner than most scales of the distortions in the flow. It also assumes that chemical reactions are very fast compared to the other flow time scales. Therefore, the flame is modelled as a thin flame front moving around in the flow in a frozen fashion. The flame is then described by a kinematic equation for the propagation of the flame front

and the mixture fraction equation for the mixing, coupled with the fluid dynamics (CFD) equations [52].

In 1999, van Oijen and de Goey proposed the Flamelet Generated Manifold (FGM) technique [63] and argued that the FGM approach is more accurate than ILDM or CSP methods in colder parts of the flame, because it also takes transport effects into account in the reduction of chemical kinetics. ILDM/CSP techniques are solely based on chemical kinetics, while FGM takes into account the balance between convection, diffusion and reaction [63]. The transport equation for a chosen progress variable $\mathcal{Y}(\mathbf{x}, t)$ is given by:

$$\frac{\partial}{\partial t}(\rho\mathcal{Y}) + \nabla \cdot (\rho\mathbf{v}\mathcal{Y}) - \nabla \cdot \left(\frac{1}{Le_{\mathcal{Y}}}\frac{\lambda}{c_p}\nabla\mathcal{Y}\right) = \omega_{\mathcal{Y}} \tag{1.3}$$

where λ is the thermal conductivity, c_p is the specific heat at constant pressure, $Le_{\mathcal{Y}}$ the Lewis number of $\mathcal{Y}$ and $\omega_{\mathcal{Y}}$ the source term of the progress variable. Equation 1.3 describes the convection-diffusion-reaction balance of $\mathcal{Y}$ along flamelet paths [63].

In the FMG technique, instead of solving all the transport equations at hand, only a subset of those equations describing the main progress in the flame are solved, and the rest of the detailed parameters are stored in a database which is prepared beforehand and accessed by the solver during the actual flame simulation. The key variable stored in the database is the chemical source term, which takes into account chemical reactions and molecular diffusion effects. Other physical effects impacting the flame structure are flame stretch, curvature and preferential diffusion. The chemical source term is not very sensitive to stretch and curvature, but can be influenced by changes in pressure, enthalpy and elemental composition [63]. The sensitivity of the source term to those effects must be taken into account in the database, which leads to a multiple-dimensional manifold, generated using flamelets experiencing the same perturbations. When applied in RANS or LES, closure to the FGM model is often obtained through the presumed PDF method, which is probably the most applied one [63].

Other methods were also developed in order to couple diffusion and chemistry in reduction methods, such as the Flamelet Prolongation of ILDM (FPI) technique [22]. FPI was originally developed in order to tackle an important limitation of the ILDM model, namely that the low-temperature zone of the flame cannot be predicted correctly when using a small number of coordinates. The reason is that the ILDM method neglects the

fast characteristic times. As a consequence, very high-dimensional manifolds are required by the ILDM method to correctly descibe the low-temperature region. The authors of the FPI method therefore extended the manifold in these regions by combining the one-dimensional ILDM manifold for high temperatures and the prolongation of the manifold in the low-temperature domain. This prolongation was based on the computation of laminar premixed one-dimensional free flames, instead of using a simple linear prolongation [22]. Using this approach, they were able to take into account differential diffusion in a straightforward way.

FGM and FPI were initially developed for premixed flames, and later extended to non-premixed conditions. On the other hand, the Flamelet/Progress Variable (FPV) approach developed by Pierce and Moin [53] was specifically designed for application in non-premixed flames.

1.2.3.1 Flamelet Progress Variable

In the FPV model, a turbulent diffusion flame is considered as an ensemble of laminar flamelets. The composition space is parameterized by the mixture fraction f and a progress variable C, whose transport equations are given by [53]:

$$\frac{\partial}{\partial t}(\overline{\rho}\widetilde{f}) + \nabla \cdot (\overline{\rho}\widetilde{\mathbf{v}}\widetilde{f}) = \nabla \cdot \left[\overline{\rho}(\widetilde{D_f} + D_t)\nabla\widetilde{f}\right]$$

$$\frac{\partial}{\partial t}(\overline{\rho}\widetilde{C}) + \nabla \cdot (\overline{\rho}\widetilde{\mathbf{v}}\widetilde{C}) = \nabla \cdot \left[\overline{\rho}(\widetilde{D_C} + D_t)\nabla\widetilde{C}\right] + \overline{\rho}\widetilde{\omega}_C$$

where $\widetilde{\ }$ denotes Favre-averaged values and D_t is the turbulent diffusivity. The Favre-averaged species mass fraction $\widetilde{y_i}$ and progress variable source term $\widetilde{\omega}_C$ are obtained by integrating the laminar composition state from the flamelet library over the joint Probability Density Function (PDF) of f and C:

$$\widetilde{y_i} = \int y_i(f,C)\widetilde{P}(f,C)dfdC$$

$$\widetilde{\omega_C} = \int \omega_C(f,C)\widetilde{P}(f,C)dfdC$$

The joint PDF $\widetilde{P}(f,C)$ is obtained through:

$$\widetilde{P}(f,C) = \widetilde{P}(C|f)\widetilde{P}(f)$$

where $\widetilde{P}(f)$ is a beta-PDF and $\widetilde{P}(C|f)$ is a delta function:

$$\widetilde{P}(C|f) = \delta(C - \widetilde{C|f})$$

1.2.4 Low-dimensional manifolds

Among the different methods found in the litterature, Principal Component Analysis (PCA) appears as an ideal candidate to fulfill the purpose of identifying low-dimensional manifolds [18, 19, 30, 47, 48, 54, 62]. PCA offers the possibility of automatically reducing the dimensionality of data sets consisting of a large number of correlated variables, while retaining most of the variation present in the original data. After reduction, the new set of variables, called principal components (PCs), are orthogonal, uncorrelated and linear combinations of the original variables. By retaining the PCs containing most of the variance and transporting them in a numerical simulation, the dimensionality of the system can be highly reduced. Another advantage of PCA resides in the fact that the PCs can be obtained through data sets based on simple systems (such as canonical reactors) and then applied to a similar, more complex system [4].

A methodology based on PCA was proposed [48] for the identification of the controlling dynamics in reacting systems and for the consistent reduction of very large kinetic mechanisms. Sutherland and Parente [62] proposed a combustion model based on the concepts from PCA (PC-score approach). They derived transport equations for the principal components (PCs), and proposed a model where the state-space variables are constructed directly from the PCs. The PCA-based modeling approach was enhanced [19, 41, 65] by combining PCA with nonlinear regression techniques, allowing a nonlinear mapping of the thermo-chemical state and the corresponding source terms onto the basis identified by the principal components. As a result, the nonlinear nature of chemical manifolds is better captured, thus, maximizing the potential size reduction provided by the method.

Isaac et al. [30] and Echekki and Mirgolbabaei [18] provided the first a posteriori studies on the use of the PC-score approach. In particular, Isaac et al. showed in [30] the potential of PC-transport based combustion models coupled with nonlinear regression techniques. The model was tested on an unsteady calculation of a perfectly stirred reactor (PSR) burning syngas. The authors showed that Gaussian Process Regression (GPR) technique produced the most accurate reconstruction, showing remarkable accuracy for the prediction of temperature and major and minor species with 2 transported variables instead of 11. The approach was also tested for the first time within a CFD solver.

1.3 Principal component analysis

Principal Component Analysis [31] is a useful statistical technique that has found application in combustion for its ability of identifying low-dimensional manifolds. In high dimension data sets, where graphical representation is not possible, PCA can be a powerful tool as it identifies correlations and patterns in a data set. Once these patterns have been identified, the data set can be compressed by reducing the number of dimensions without much loss of information. PCA analyzes the covariance between variables in a data set and identifies a linear representation of the system through orthogonal vectors, each one having a significance proportional to its eigenvalue.

In order to perform principal component analysis, a data-set $\mathbf{X}\,(n \times Q)$ consisting of n observations of Q independent variables is needed. Then, the data must be centered (by subtracting its mean) and scaled (using an appropriate scaling method): centering is used to convert observations into fluctuations over the mean, while scaling is done in order to compare the data evenly (if they have different units or order of magnitudes):

$$\mathbf{X_{SC}} = (\mathbf{X} - \overline{\mathbf{X}})\mathbf{D}^{-1} \tag{1.4}$$

where $\overline{\mathbf{X}}$ is $(1 \times Q)$ vector containing the mean of each variable and $\mathbf{D}$ is a $(1 \times Q)$ vector containing the scaling factor of each variable. Several scaling methods can be found in the litterature: *auto scaling*, *range scaling*, *pareto scaling*, *variable stability scaling* and *level scaling* [47].

Then, one can compute the covariance matrix $\mathbf{S}$ defined as (the notation $\mathbf{X}$ will be used in the following instead of $\mathbf{X_{SC}}$ for the sake of simplicity):

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X}$$

The diagonal elements of $\mathbf{S}$ represent the variance of each variable, while the off-diagonal values show the covariance between two variables. Since $\mathbf{S}$ is a square matrix (of size $(Q \times Q)$), an eigenvalue decomposition can be performed yielding the eigenvectors and eigenvalues of the system:

$$\mathbf{S} = \mathbf{ALA}^T$$

where $\mathbf{A}\,(Q \times Q)$ and $\mathbf{L}\,(Q \times Q)$ are respectively the eigenvectors of $\mathbf{S}$ (also called principal components, PCs) and the eigenvalues of $\mathbf{S}$, in decreasing order. The eigenvectors matrix

$\mathbf{A}$, also called the basis matrix, is used to obtain the principal component scores, $\mathbf{Z}$ $(n \times Q)$, by projecting the original data set $\mathbf{X}$ on that basis:

$$\mathbf{Z} = \mathbf{X}\mathbf{A} \tag{1.5}$$

Eq. 1.5 indicates that the original data set can be uniquely recovered using the PCs and their scores:

$$\mathbf{X} = \mathbf{Z}\mathbf{A}^{-1}$$

where $\mathbf{A}^{-1} = \mathbf{A}^T$. Then, using a subset of $\mathbf{A}$ by retaining only q PCs (with $q < Q$), noted $\mathbf{A_q}$, an approximation of $\mathbf{X}$ based on the first q eigenvectors ($\mathbf{X_q}$) is obtained:

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

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. In the PC analysis, the largest eigenvalues correspond to the first columns of $\mathbf{A}$. This means the largest amount of variance in the original variables is described by the first PCs. Thus, the truncation is made on the last eigenvectors (corresponding to the smallest eigenvalues). By removing the last PCs, the dimension of the system is reduced while retaining most of the variation in the system.

1.3.1 PC-score approach

In the work of Sutherland and Parente [62], a model based on transport equations for the PCs is proposed derived from the general species transport equation:

$$\frac{\partial}{\partial t}\left(\rho Y_k\right) + \nabla\left(\rho \bar{\mathbf{u}} Y_k\right) = \nabla\left(\rho D_k \nabla Y_k\right) + R_k \qquad k = 1, ..., n_s \tag{1.6}$$

where Y_k is the mass fraction of species k and R_k is its corresponding source term (with n_s the total number of species in the system), D_k the diffusion coefficent for species k, ρ the density and $\bar{\mathbf{u}}$ the velocity vector. Transport equations for the PC scores ($\mathbf{Z}$) can be formulated from Eq. 1.6 given the basis matrix $\mathbf{A}$ and the scaling factors d_k:

$$\frac{\partial}{\partial t}\left(\rho \mathbf{z}\right) + \nabla\left(\rho \bar{\mathbf{u}} \mathbf{z}\right) = \nabla\left(\rho \mathbf{D_z} \nabla \mathbf{z}\right) + \mathbf{s_z} \tag{1.7}$$

$$\mathbf{s_z} = \sum_{k=1}^{Q} \frac{R_k}{d_k} \mathbf{A_{kq}} \tag{1.8}$$

where $\mathbf{z} = \mathbf{Z_i^t}$ represents an individual score realization. One of the major weaknesses of classic PCA is that a multi-linear model is used to approximate a highly nonlinear manifold. The nonlinearity of chemical manifolds can be attributed to the high nonlinearity of

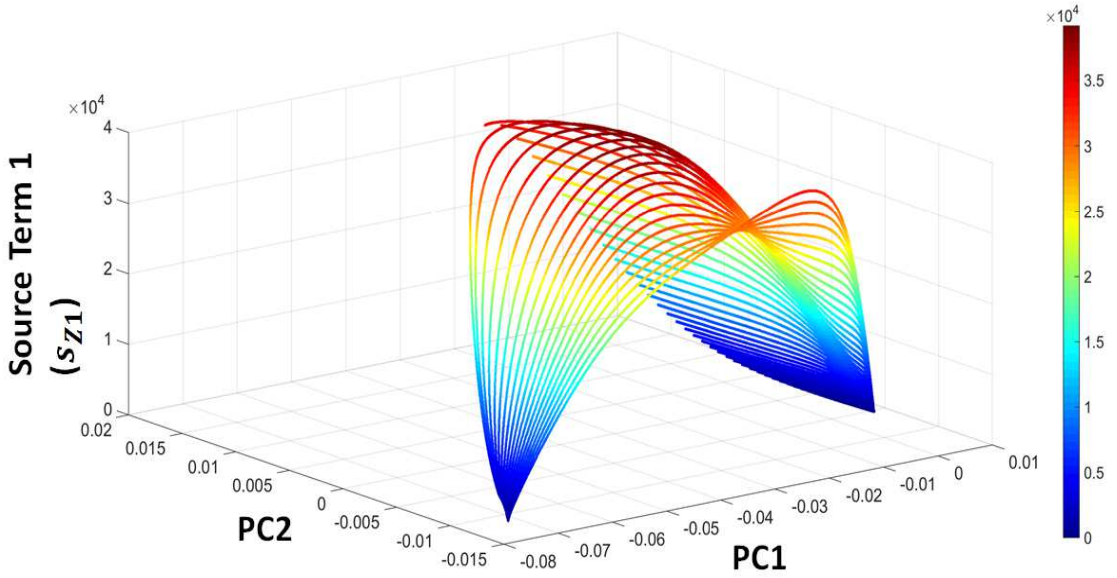


Figure 1.1: Manifold of Source Term 1 ($\mathbf{s_{z1}}$) in function of $PC1$ and $PC2$

chemical source terms (Arrhenius). This can be visualized in Fig. 1.1, showing the first principal component source term $\mathbf{s_{z1}}$, as a function of the first two principal components for a propane case.

1.4 Regression Models

In this study, the state-space variables $(Y_k, T, \rho, ...)$ and the PC source terms $(\mathbf{s_{z_q}})$ are mapped to the PC basis using nonlinear regression:

$$\phi \approx f_\phi \left(\mathbf{Z_q}\right)$$

where f_ϕ is the nonlinear regression function and ϕ represent the dependant variables (i.e. Y_k, T, ρ and $\mathbf{s_{z_q}}$). In a previous study [30], the authors compared different regression models in their ability to accurately map the highly nonlinear functions (such as the chemical source terms) on the plane PCA manifold. These models include:

- *Linear Regression Model* (LIN) in which the state-space is mapped to the PC using a linear function [11]

- *Mutivariate Adaptive Regression Splines* (MARS) where the model is build from product spline basis functions [21]

- *Artificial Neural Networks* (ANN) that uses the concept of networking various layers of estimation resulting in a highly accurate output layer [46]
- *Support Vector Regression* (SVR) which is a subset of support vector machines (SVM) and in which the idea is again to create a model which predicts s_Z given Z using learning machines which implement the structural risk minimization inductive principle [61]
- *Gaussian Process Regression* (GPR), which is based on the idea that dependent variables can be described by a gaussian distribution [43, 57]. In particular, it was shown that GPR produced the most accurate reconstruction of the state-space variable, using only 2 transport equations instead of 11 in the full system without regression.

1.4.1 Gaussian Process Regression

When dealing with high-dimensional space function approximation, GPR provides a very suitable and powerful approach. Starting from a set of n training data points $\{\mathbf{x}_i, \phi_i\}_{i=1}^n$, a functional expression $f(\mathbf{x}_i)$ can be found transforming the input vector $\mathbf{x}_i$ into the target value ϕ_i using:

$$\phi_i = f(\mathbf{x}_i) + \epsilon_i$$

where ϵ_i reprsents Gaussian noise with zero mean and variance σ_n^2 [57]. The dependant variables (i.e. the targets) can therefore be described by a Gaussian distribution:

$$\mathbf{Œ} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I})$$

where $\mathbf{X}$ represents all the input points $\mathbf{x}_i$ and $\mathbf{K}(\mathbf{X}, \mathbf{X})$ the covariance matrix. The covariance function used here is the Squared Exponential:

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

with σ_f^2 being the signal variance and l the characterictic lenght scale. These two parameters of the covariance function are called *hyper-parameters*. Given a query point $\mathbf{x}_*$, the predicted mean value ϕ_* can be obtained using:

$$\phi_* = \mathbf{K}_*^{\mathbf{T}} (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y}$$

where $\mathbf{K}_* = \mathbf{K}(\mathbf{X}, \mathbf{X}_*)$. The hyperparameters remain the only open parameters of a Gaussian process. After an initial guess, their optimal value for a given training data set

can be automatically estimated using a gradient-based marginal likelihood optimization. The GPR implementation for MATLAB, called the *gpml toolbox* [57], was used in this work for the regression of the different manifolds.

In the present work, the choice of GPR was motivated by its semi-parametric nature, which increases the generality of the approach. Moreover, Isaac et al. [30] performed a comparative study on the various regression methods presented above. Their analysis was based on a one-dimensional turbulence (ODT) data-set of a non-premixed synthesis/air jet, obtained using a 11-species mechanism. The different regression models were trained on $n = 5000$ sample points evenly distributed over PC space, using 2 and 3 scores ($q = 2$ or 3). The authors showed that the linear regression model (LIN) had difficulties mapping the highly nonlinear source terms, with $q = 2$ and 3. Polynomial methods such as MARS also struggled with the mapping given the high degree of non-linearity. Local tuning methods (ANN, SVR, GPR) were able to better approximate the challenging regions of the manifold. Overall, GPR was able to produce the most accurate reconstruction for the PC source terms with $q = 2$ and 3. It was also pointed out by the authors that the efficiency of GPR regression compared to the other techniques was due to the robust optimization of the hyper-parameters that the implementation utilizes. Indeed, GPR required the smallest amount of optimization work from the user due to the minimization functions, which automatically estimates the hyper-parameters.

1.5 Methodology

The present dissertation will articulate around the development of a reduced-order model based on PCA for multi-scale simulations. Therefore, this work is oraganized as follows:

– Chapter 2 will present an *a priori* validation of the reduced-order model using perfectly stirred reactors (PSR), for two different fuels (methane and propane) and three different kinetic mechanisms of increasing complexity (34, 50 and 162 species respectively). Detailed chemistry simulations of a PSR will be performed by varying the residence time in order to create the training database. Thereafter, the data set will be parameterized using PCA and validated using the PC-transport approach in the same PSR. After that, the potential of coupling PCA with nonlinear Gaussian

Process Regression (GPR) will also be demonstrated, with the aim of increasing the size reduction potential offered by PCA. Finally, a novel approach using local nonlinear regression will be presented as an enhancement of the model.

– Chapter 3 will show the first application of the PCA based reduced model *a posteriori* in the context of Large Eddy Simulation (LES) of turbulent combustion. The model will be validated against experimental data of Sandia flames D, E and F. The reduced model will be trained on $1D$ counterflow flamelets of varying strain rate, with the GRI 3.0 mechanism for the description of the chemistry. GPR will be used to parameterize the thermo-chemical state-space using the first two PCs. In LES, transport equations for the PCs will be solved, and a lookup table will be used to recover all the dependant variables. The extinction and re-ignition phenomena in flames E and F will be examined using scatter plots of temperature and CO mass fraction, and compared with the scatter data from the experimental measurements. Finally, a comparaison with another numerical study will be performed, mainly with the FPV model.

– Chapter 4 will advance the understanding and application of the proposed reduced-order model in the context of LES, by investigating some key features of the model and will also propose, for the first time, a subgrid scale closure model for the application of PCA within a CFD solver. Several aspects will be investigated: the sensitivty to the training data set where a comparison will be made between $1D$ counterflow and $0D$ PSR data sets, the influence of the scaling methods used in PCA, the issue of data sampling will be addressed using the kernel density weighting method and the potential of a subgrid scale closure will also be tested. A qualitative comparison with the results from another numerical study using the FPV model will also be performed.

CHAPTER 2

PRINCIPAL COMPONENT ANALYSIS COUPLED WITH NONLINEAR REGRESSION FOR CHEMISTRY REDUCTION

The objective of the present chapter is to advance the understanding and application of the PC-transport approach by applying this method to more complex fuels such as methane and propane.

First, 0D simulation of a PSR is used to generate the database for model training. Then, the solution of a steady and unsteady PSR calculation using the PC-transport approach for large kinetic mechanisms is compared with the full solution. Next, the PC-transport approach is coupled with nonlinear regression (PC-GPR) in order to increase the size reduction potential of PCA. Finally, the first study on an enhancement of the classical PC-transport approach by the use of local nonlinear regression (PC-L-GPR) is also shown. It should be pointed out that the objective of the present work is to demonstrate the applicability of GPR regression for accurate source term regression. To this purpose, the choice of a PSR is quite obvious as it allows to focus on such an aspect without the influence of transport processes.

2.1 Introduction

In the present work, PCA is used to identify the most appropriate basis to parameterize the empirical low-dimensional manifolds and define transport equations in the new space (see Eq. 4.1 and 1.8). Then, both the state space and the source terms are non-linearly regressed onto the new basis using several approaches, described in Section 1.4.

The non-linear regression of the chemical state space and of the corresponding source terms is intended to overcome the shortcomings associated to the multi-linear nature of PCA, and to reduce the number of components required for an accurate description of the state-space. The method belongs to the family of Empirical Low-Dimensional Manifolds

(eLDMs) [54], and it is based on the idea that compositions occurring in combustion systems lie close to a low-dimensional manifold. eLDMs require samples for the construction of reduced models, which might be seen as a limitation of the approach, as all system states are required before model reduction. However, although initial studies on PCA models involved DNS data of turbulent combustion [54, 62], recent studies have demonstrated [5, 14, 18] that PCA-based models can be trained on simple and inexpensive systems, such as 0D reactors and 1D flames, and then applied to model complex systems, such as flame-vortex interaction [13], flame-turbulence interactions [30] as well as turbulent premixed flames [14].

2.2 Local regression

In order to improve further the accuracy of the regression and increase PCA's potential for size reduction, a novel approach is proposed where the PC-score approach is coupled with locally regressed state-space (PC-L-GPR).

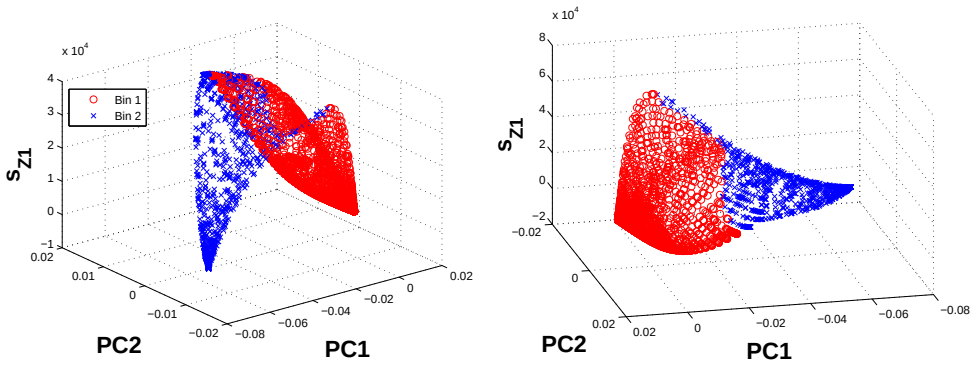
The idea is to divide the PC state-space into bins or clusters, and to perform a GP regression seperately in each of these bins. As a consequence, a better regression would be obtained (if each bin is chosen appropriately) and the computational time required for GPR will also be reduced. In order to define such bins, a conditioning variable has to be chosen. This variable should be able to capture the general characteristics of the state-space. Possible candidates are the PCs source terms, as the latter are highly nonlinear over the PC space.

Clustering the source terms manifolds such as they can be approximated by quasi-linear functions in each bin would simplify and accelerate the regression algorithm. As to the author's knowledge, this approach has not yet been tested previously in the context of PC-transport approach.

2.2.1 Effect of conditioning on the regression accuracy

As stated above, the PC source terms are appropriate candidates to define the clusters for the local regression, because they show the strongest nonlinearities in the PC space. A good candidate would be the first PC source term s_{Z_1}, as the latter is highly correlated with the major species and also contains most of the variance in the system. Figure 2.1 shows the

first source term's manifold in a 2D PC space. The bin borders are chosen to the extrema of s_{Z_1}. This results in two bins as shown on Figure 2.1 , the border being represented by the red line. It can be seen that in each bin, s_{Z_1} is a rather smoothly increasing (or decreasing)



(a) (b)

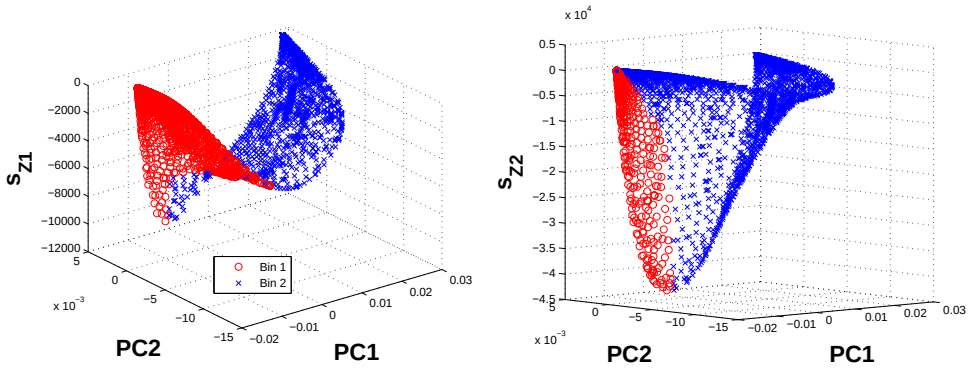
Figure 2.1: Clustering based on the extrema of s_{Z_1} for both s_{Z_1} (a) and s_{Z_2} (b) (propane case, Polimi mechanism)

function of Z_1 and Z_2. Regressing each of these two bins seperately is easier, more accurate and faster than regressing the whole manifold at once (i.e. global regression). It must be noted that the manifold of s_{Z_2} was also clustered based on the extrema of s_{Z_1} (Figure 2.1b)), but those extrema do not necessarily fall within the ones of s_{Z_2}. Although this approach leads to improved results compared to global regression (cfr. Section 2.4.2.2), they can be further improved even using the double conditioning method (cfr. Section 2.2.1).

Local regression provides better results when the bins and conditioning variable are chosen correctly (cfr. Section 2.4.2.2). In order to handle the discontinuties that could occur at the boundaries of the bins, the clusters were artificially extended across the bin border, by providing an overlap of 2% at the boundaries of the cluster region, to ensure smoothness of the soluton and avoid discontinuities.

In some cases, local regression with a single conditioning variable can still provide unsatisfactory results. In such cases, the accuracy of the results can be further improved by using a second conditioning variable. In the case of PC-transport where the first conditioning variable is s_{Z_1}, a natural choice for the second conditionig variable would be the second PC source term s_{Z_2}. Thus, s_{Z_1} is regressed locally based on clusters defined by its extrema, while s_{Z_2} is regressed locally in clusters defined by its own extrema (i.e. not

based on s_{Z_1} extrema). Figure 2.2 shows the first and second source terms' manifolds for the propane case, together with the clusters borders.



(a) (b)

Figure 2.2: Clustering based on the extrema of s_{Z_1}(a) and s_{Z_2}(b) (propane case, San Diego mechanism)

2.3 Perfectly stirred reactor and test cases

The objective of the present chapter was to extend previous investigation on syngas [30] to more complex fuels, with a significantly large number of species and reactions. In [30], the proposed PCA approach was demonstrated on the unsteady solution of a perfectly stirred reactor (PSR). The solution from the full set of equations was compared to the standard PC-transport approach, and the PC-transport approach using nonlinear regression.

In this work, the analysis of the proposed PC model in its ability to handle complex fuels and large kinetic mechanisms was done in a similar way. The data sets for PCA were generated by performing unsteady simulations by varying the residence time in the vessel from extinction to equilibrium. For each residence time, the temporal solution was saved until steady-state was reached. The vessel was initialized at equilibrium conditions (constant pressure and enthalpy) and the inlet conditions for the reactor were set at an equivalence ratio of 1. The initial conditions for the reactor are set at the equilibrium conditions of the inlet and the system is run until a steady-state solution is reached. The PSR is modeled assuming constant volume, residence time and pressure. The ideal gas

law was used to model the behaviour of the mixture. Thermodynamic properties were obtained through the Cantera software package [23]. Two different fuels were investigated:

- methane (CH_4), burned with pure oxygen. The mechanism used was the GRI 3.0 [60], without species containing nitrogen (resulting in 34 species). The inlet temperature was set to 300K. One hundred cases were run between residence time of $1e^{-4}\,s$ to $1e^{-6}\,s$. The PCA database generated in this way contained $\sim$ 100,000 points.

- propane (C_3H_8), burned with air. Two different kinetic schemes were used: the San Diego Mechanism [64] (subsequently referred as San Diego), without nitrogen species (50 species, 230 reactions) and the Primary Reference Fuels Polimi_PRF_PAH_HT_1412 kinetic mechanism [26] (subsequently referred as Polimi), without nitrogen species (162 species, $\sim$6,000 reactions). In both cases, N_2 stays inert. The inlet temperatures were set to 1300K for the San Diego scheme and to 1500K for the Polimi mechanisms. One hundred cases were run between residence time of $1e^{-1}\,s$ to $1e^{-7}\,s$. The PCA database consisted of $\sim$ 110,000 points for the San Diego scheme and of $\sim$ 420,000 points for the Polimi one.

The PCA process described in the previous section is then applied to the database to create the basis matrix A_q, and the regression functions f_ϕ for the state-space variables, ϕ. Gaussian Process Regression was done using 1, and 2 PC's as independant variables. The implementation of the PSR equations was done using MATLAB together with the Cvode toolbox and Cantera. The temporal solution to the equations is obtained using the Newton nonlinear solver, and the BDF multistep method. Governing equations for species transport and energy were implemented and solved:

$$\frac{\partial m_i}{\partial t} = \dot{m}_{i,in} - \dot{m}_i + \omega_i \cdot MW_i \cdot V \tag{2.1}$$

where m_i (kg) and ω_i $(kmol/m^3/s)$ are the mass and the net molar production rate of the i^{th} species, MW_i is the molecular weight of the i^{th} species and V (m^3) the volume of the reactor. or the mass flow rates (kg/s), $\dot{m}_{i,in}$ is the mass flow of the i^{th} species entering the reactor and $\dot{m}_{i,out}$ is the mass flow exiting the reactor. The residence time τ $(/s)$ in the reactor is defined as:

$$\tau = \frac{\rho V}{\dot{m}}$$

where ρ is the density of the mixture inside the reactor. For the energy equation:

$$\frac{\partial H}{\partial t} = \dot{m}_{in} h_{in} - \dot{m} h + V \frac{dP}{dt} \tag{2.2}$$

where H is the enthalpy of the system and h is the specific enthalpy (J/kg), $\dot{m}_{in}$ is the total mass flow entering the reactor and $\dot{m}$ the total mass flow rate leaving the system. The last term of Eq. 2.2 being zero as the PSR operates in constant pressure conditions. In this study, no accumulation of mass inside the reactor has been assumed, thus $\dot{m}_{in} = \dot{m}_{out} = \dot{m}$, but $\dot{m}$ can change due to a change in density.

2.4 Results and discussion

In this section, the proposed method is demonstrated in a PSR, comparing the calculations using the full set of equations to the standard PC-transport approach and to the PC-transport approach using nonlinear regression. This demonstration is done for two different fuels: a simple one, methane (CH_4), and a more complex, propane (C_3H_8). But first, an analysis is performed on the effects of several scaling methods used in PCA (Eq. 1.4).

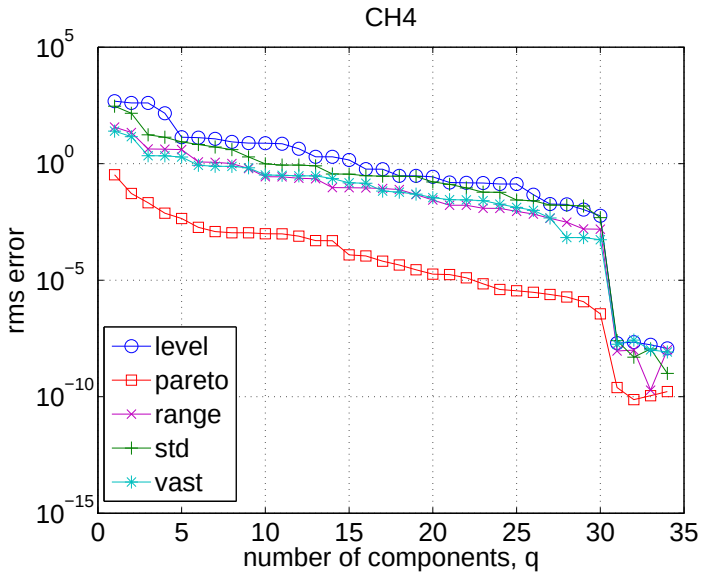
2.4.1 Scaling

As mentioned in [47], scaling has an important effect on the accuracy of the method. It can change the PCA structure by altering the relative importance of various species, and the choice of a particular scaling method is motivated by the goal of the resulting PCA to reconstruct specific variables. In [30], the authors showed that pareto scaling method is able to achieve the greatest reduction, and produces a highly regressible surface for a syngas mechanism. This was also consistently showed in other previous investigations [3, 5, 29, 47, 50].

In order to assess the accuracy of the various scaling methods presented in Section 1.3, a similar study was performed for the methane and propane cases on the species and PCs source terms. The *rms error* was used as a mean of quantifying the error in the reconstruction of species mass fractions and PCs source terms. The definition of the *rms error* used here is:

$$rms\,error = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{predicted,i} - x_i\right)^2}{n}}$$

Figure 2.3 shows the *rms error* for the mass fraction of CH_4 for the methane case, and for



(a)

Figure 2.3: *Rms error* values for CH_4 mass fraction while varying q, the number of PCs, and the scaling method (methane case).

the various scaling methods while varying the number of principal components, q. It is clear from Fig. 2.3 that pareto scaling provides the lowest error in the reconstruction for CH_4, and this for all the range of q, while all other methods show similar behaviour. It can also be seen that a significant decrease in the *rms error* is not achieved until $q = 31$, and this is observed with all the scaling methods. With $q = 31$, only a minor reduction is achieved. This is due to the linear nature of PCA based models, which try to model highly nonlinear reaction rates on a linear basis. An alternative approach to overcome this issue can be the use of nonlinear regression functions, which can be used to map the nonlinear reaction rates or nonlinear species concentrations to the lower dimensional representation given by the PCs. A similar analysis of the influence of scaling methods was also done for the propane cases, which led to the same conclusion, i.e. that pareto scaling provides the lowest error in the reconstruction of all species (major and minor).

2.4.2 Standard PC-score approach vs PC-score with Gaussian Process Regression

The standard PC-score approach based on Eq. 4.1 and 1.8 was tested for both methane and propane, and compared to the full solution, i.e. the solution based on the transport of all species (Eq. 2.1 and 2.2). Then, the non linear state-space variables were mapped to the linear PC basis using Gaussian Process Regression (GPR). GPR was performed on all variables (temperature, species and score source terms) using 5,000 sample points evenly distributed over the PC space. Error quantification is done through the coefficient of determination R^2:

$$R^2 = \frac{\sum_{i=1}^{n} \left(x_{predicted,i} - \bar{x}\right)^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

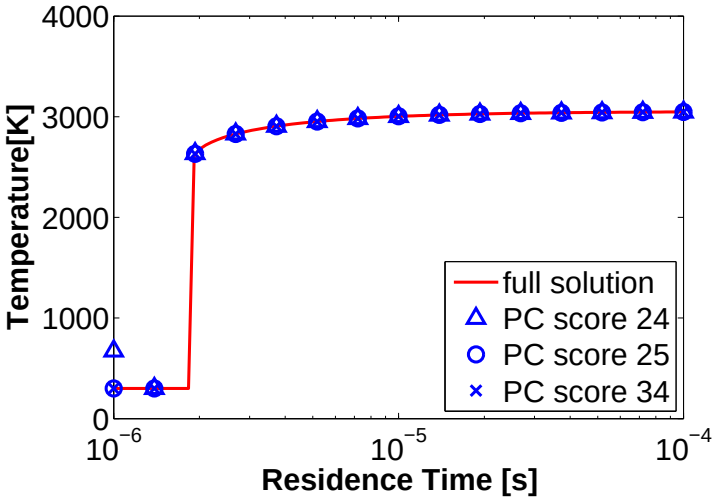
where $\bar{x}$ is the mean value of an observed variable.

2.4.2.1 Methane case

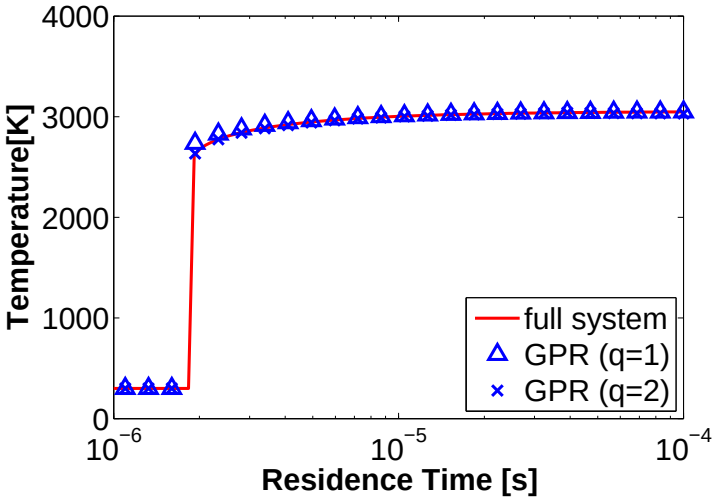
Figure 2.4 and 2.5 show the solution using the standard PC-score model (i.e. without regression) and the solution using the PC-score model together with GPR (PC-GPR) for the methane case. It can be seen that using the standard PC-score approach, at least 25 components out of 34 are required in order to obtain an accurate solution, which correspond to a model reduction of 26%. However, when using GPR, the reduction potential is highly increased: using only 2 PCs, the results show remarkable accuracy for the model with regression over the range of residence times for the predicted temperatures, and both major and minor species. A similar degree of accuracy is not observed in the model without regression until $q = 25$. Also, using PC-GPR with $q = 1$ does not provide sufficient accuracy in the ignition region, where the ignition delay is under-estimated. Moving to $q = 2$ allows to capture the ignition adequately. The regression of ϕ using Gaussian Process and pareto scaling yielded an R^2 of 0.999 for all variables using $q = 2$, and an R^2 of 0.986 or higher with $q = 1$.

2.4.2.2 Propane case - Polimi mechanism

Figure 2.6 shows the temperature profile for the combustion of propane and air using the Polimi mechanism. As far as the standard PC-score approach is concerned, it can be seen that at least 142 components out of 162 are required in order to get an accurate description using a reduced model, which represent a model reduction of 12%. When



(a)



(b)

Figure 2.4: PSR temperature as a function of the residence time, with the solid line representing the full solution. The markers represent the results for the standard PC-score model while varying q (a), and the PC-score with GPR regression (b) using q = 1 and 2 PCs

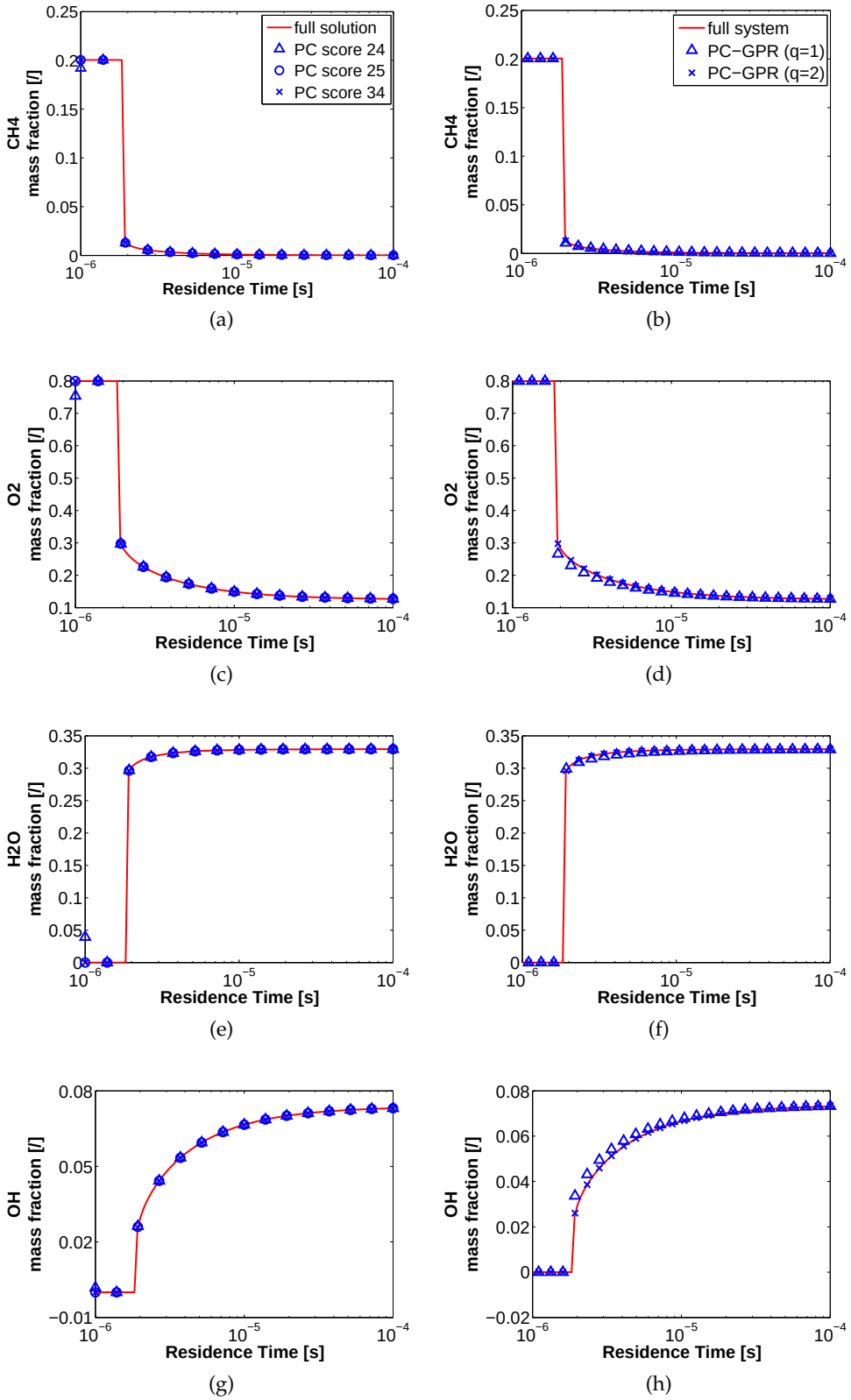
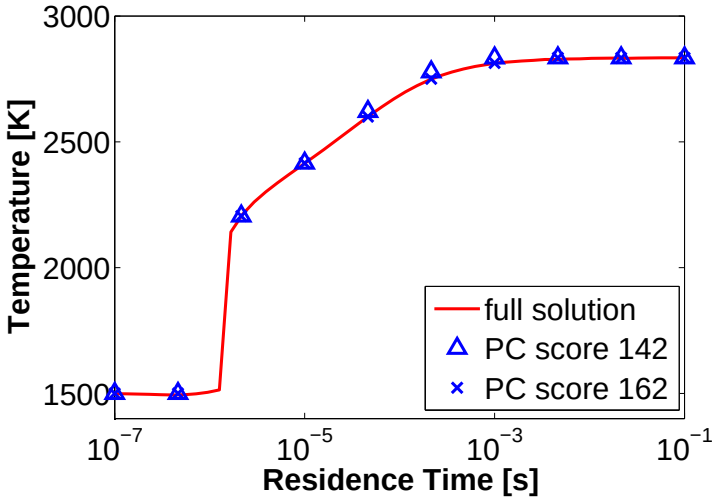


Figure 2.5: Species mass fraction as a function of the residence time, with the solid line representing the full solution. The markers represent the results for the standard PC-score model while varying q (left plots), and the PC-score with GPR regression (right plots) using q = 1 and 2 PCs

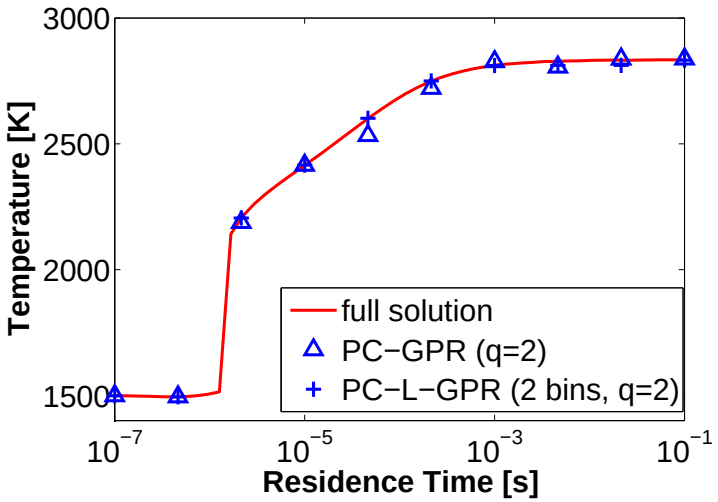
adding the potential of GPR (PC-GPR), this number can be reduced to 2, leading to fair solution, but not yet satisfying. Indeed, a significant deviation from the full solution can be observed in the ignition/extinction region. In order to improve the model even further, the potential of using GPR locally, together with the PC-score approach (PC-L-GPR), is assessed. In this study, the first principal component's source term, s_{Z1}, was chosen as the variable on which the clustering should be conditioned. The data-set was thus single conditioned on s_{Z1}. This choice can be justified knowing that the first PC's source term is highly correlated with the major species, thus containing most of the variance in the system, and also very nonlinear. The clustering algorithm used in this work searches for the extrema of the conditioning variable, and defines the borders of the bins at those extrema. This allows to have a monotonic increasing (or decreasing) variable in each bin, thus making the job easier for the regression algorithm. In the present analysis, 2 bins were identified (cfr. Figure 2.1). It can be observed on Figure 2.6b that using local regression with only 2 components instead of 162 (reduction of 98%) improves significantly the accuracy of the model, especially in the ignition/extinction region, leading to an almost perfect match. Figure 2.7 shows some of the species mass fraction. Again, it can be seen that using local regression allows to increase the accuracy in the predictions, both for major and minor species.

2.4.2.3 Propane case - San Diego mechanism

Figure 2.8 shows the temperature profile for the combustion of propane and air using the San Diego mechanism. It can be seen on Figure 2.8a that using the standard PC-score approach at least 36 components out of 50 are required in order to get an accurate description using a reduced model, which represent a model reduction of 28%. When coupling GPR with PC-score (PC-GPR), the solution obtained using only 2 components is accurate enough, except in the ignition/extinction region. In order to increase the accuracy in that region as well, the potential of PC-score with local GRP (PC-L-GPR) was assessed. Here again, the data-set was single conditioned based on s_{Z1}. Again, 2 bins were identified for the San Diego mechanism (cfr. Figure 2.2a). Figure 2.8b shows a significant improvement in the accuracy of the model in the ignition/extinction region while using only 2 components instead of 50 (reduction of 96%). Figure 2.9 shows some of the major



(a)



(b)

Figure 2.6: PSR temperature as a function of the residence time (Polimi), with the solid line representing the full solution. The markers represent the results for the standard PC-score model while varying q (a), and the PC-score with global and local GPR regression (b) using q=2 PCs and single conditioning

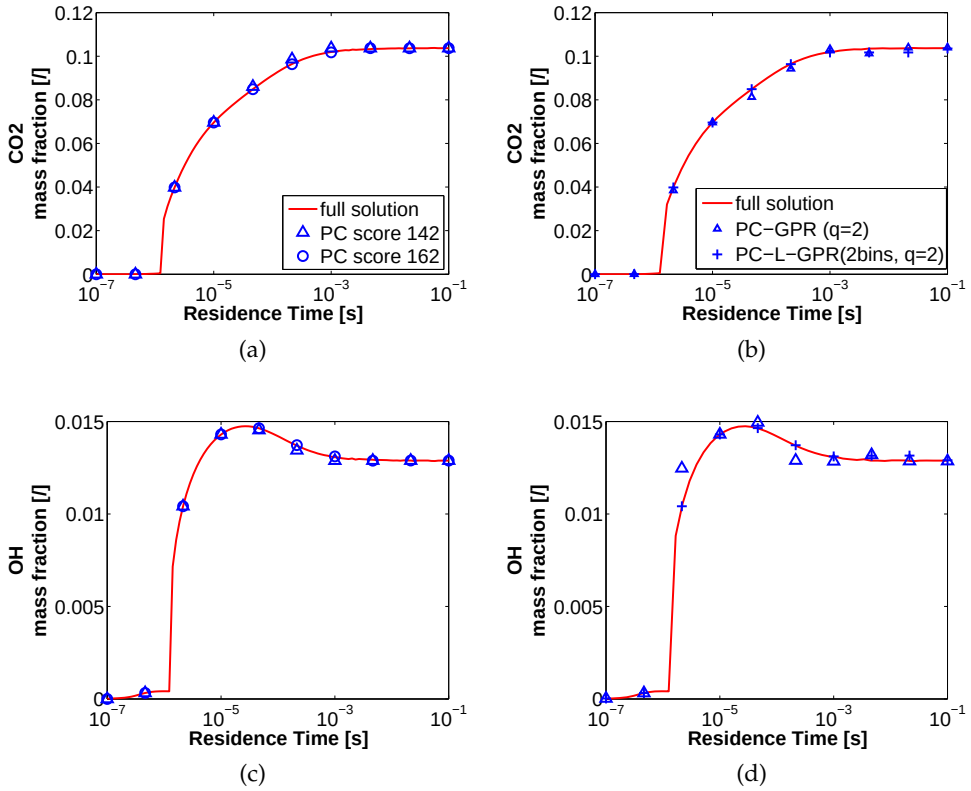


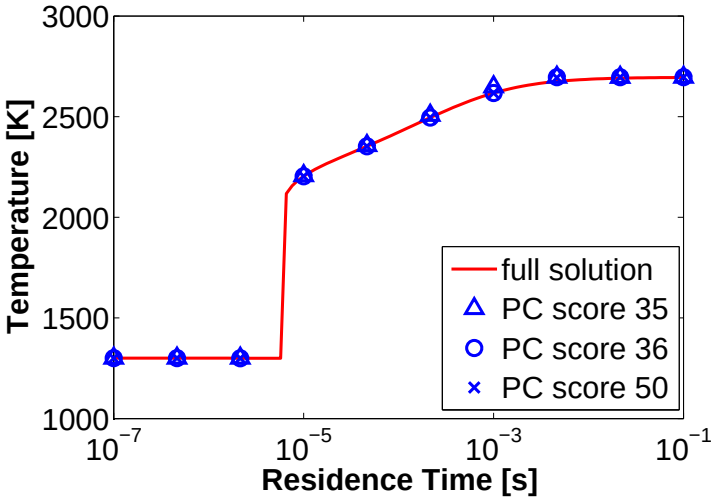
Figure 2.7: Species mass fraction as a function of the residence time (Polimi), with the solid line representing the full solution. The markers represent the results for the standard PC-score model while varying q (left plots), and the PC-score with GPR regression (right plots) using q = 2 PCs

and minor species mass fraction profiles, where it can be seen that using local regression allows to increase the accuracy of the predictions.

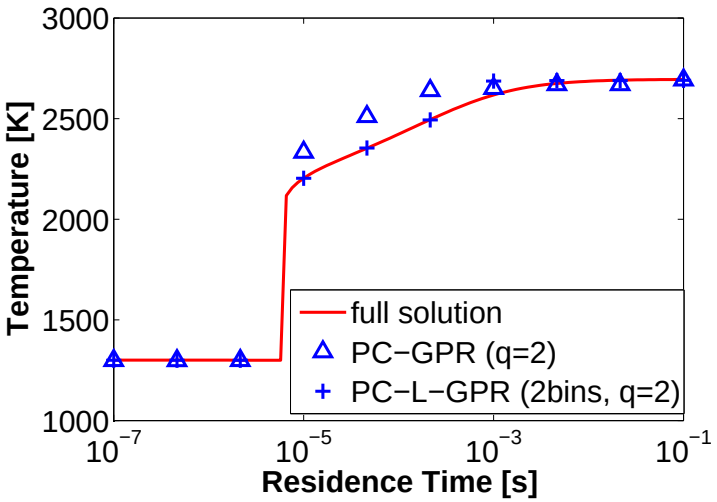
The single conditioned PC-L-GPR model gives quite satisfactory results, but these could be further improved by double conditioning the data set before using GPR. Indeed, clustering the first source term based on its own extrema increased the accuracy of the regression of s_{Z1}, but that clustering does not necessarily fall on the extrema of the second source term s_{Z2}(cfr. Figure 2.2). By clustering the s_{Z2} based on its own extrema, its subsequent regression can be strongly improved. Figure 2.10a shows the temperature profile with a comparison between single conditioned and double conditioned PC-L-GPR model. It can be seen that double conditioning the data set prior to applying the regression improves the accuracy of the result even further, leading to a perfect match between the reduced model and the full solution. The same conclusion can be drawn when looking at major and minor species profiles as shown on Figure 2.10b-d.

2.4.3 Reconstruction accuracy for transient simulations

The reduced model generation using the PC-GPR approach is now validated in a transient system. An accurate representation of the transient solution is also essential in order to guarantee reliable results. Figure 2.11 shows the temporal evolution of temperature and some species mass fraction for the methane case, with a residence time inside the reactor of $2 \cdot 10^{-5}s$. As previously, the reactor was initialized at the chemical equilibrium solution at constant enthalpy and pressure. It can be observed that temperature and species mass fractions are accurately predicted in time by the PC-GPR model, using only 2 PCs out of 35. Figure 2.12 shows the transient solution for the propane case, using the Polimi mechanism, with a residence time inside the reactor of $1 \cdot 10^{-5}s$. The temperature and species mass fraction profiles are shown for the full model and the PC-score with local GPR model, respectively. The reduced model is able to provide a very accurate representation of the transient evolution within the reactor, as for the methane case, using only 2 PCs out of 162. The ability of the reduced approach to reproduce the unsteday evolution of the chemical state using complex chemistry is very important towards its application in realistic turbulent combustion simulations.



(a)



(b)

Figure 2.8: PSR temperature as a function of the residence time (San Diego), with the solid line representing the full solution. The markers represent the results for the standard PC-score model while varying q (a), and the PC-score with global and local GPR regression (b) using q=2 PCs and single conditioning

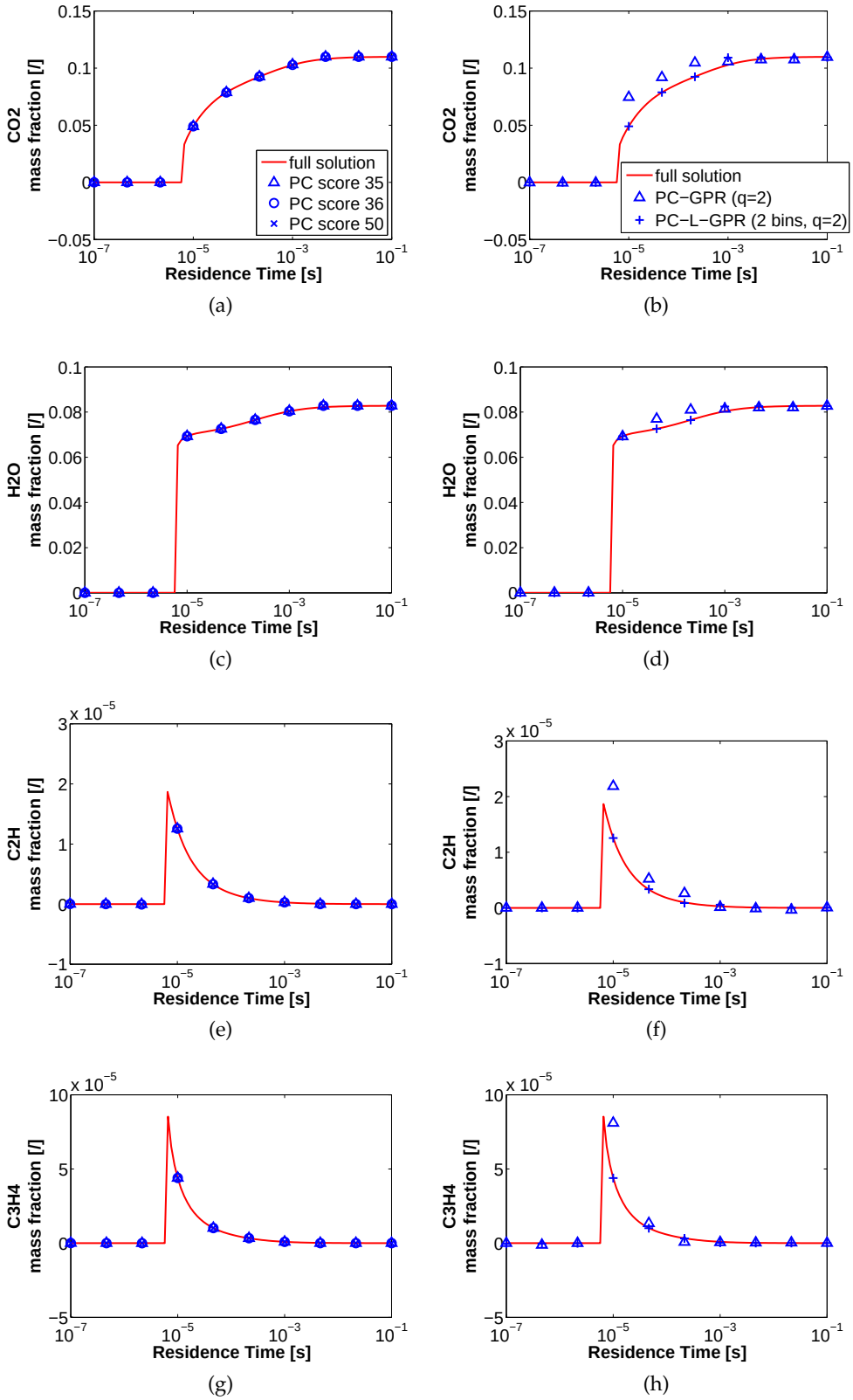


Figure 2.9: Species mass fraction as a function of the residence time (San Diego), with the solid line representing the full solution. The markers represent the results for the standard PC-score model while varying q (left plots), and the PC-score with GPR regression (right plots) using q = 2 PCs

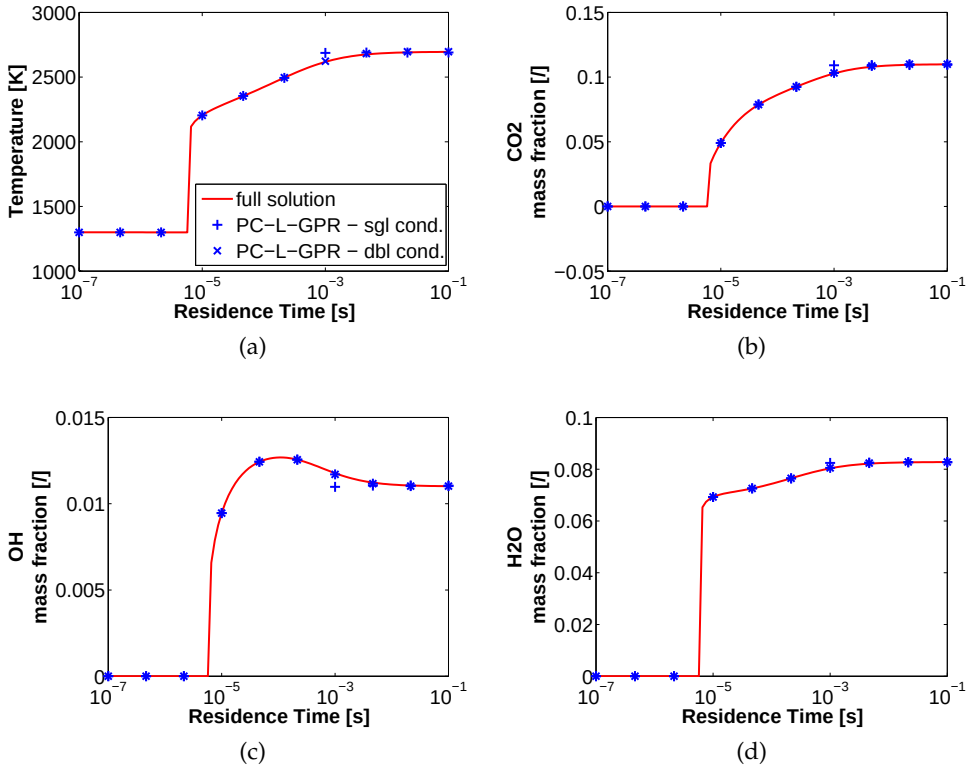


Figure 2.10: Temperature (a) and species mass fraction (b-d) as a function of the residence time (San Diego), with the solid line representing the full solution. The $'+'$ markers represent the results for the PC-L-GPR model with single conditionig and the $'*'$ markers show the solution using PC-L-GPR with double conditioning.

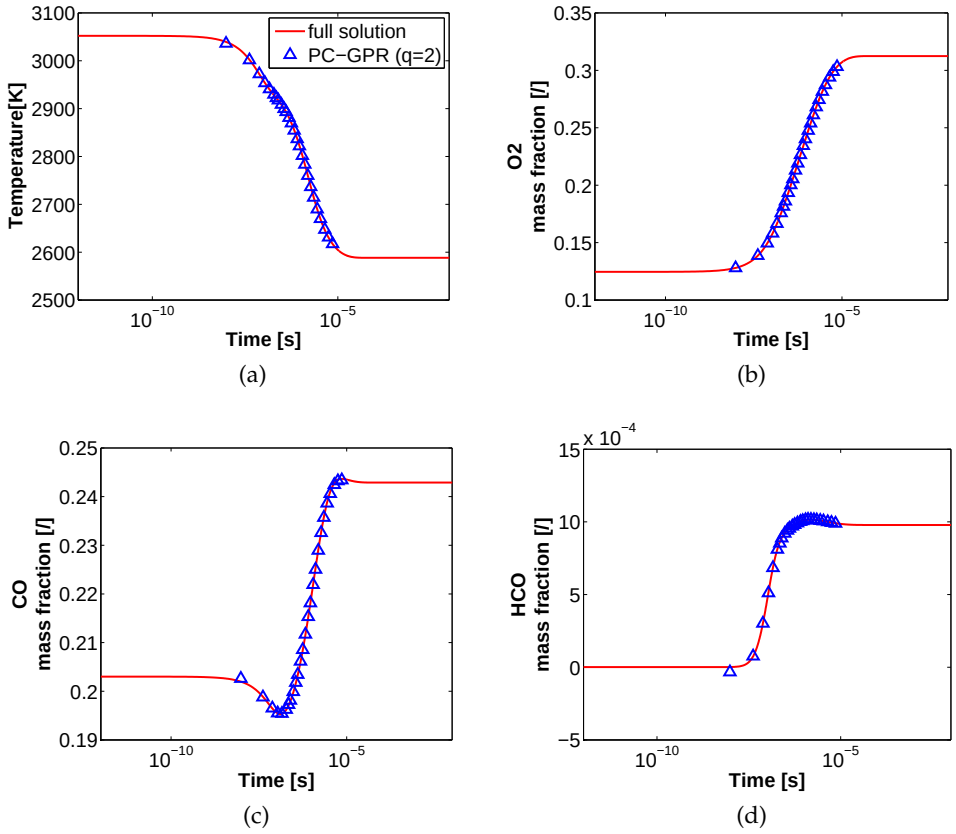


Figure 2.11: PSR temperature (a) and major and minor species (b-d) as a function of time (methane case), for a residence time of $2 \cdot 10^{-5}$s, with the solid line representing the full solution. The markers represent the results for PC-score with GPR regression using q = 2 PCs

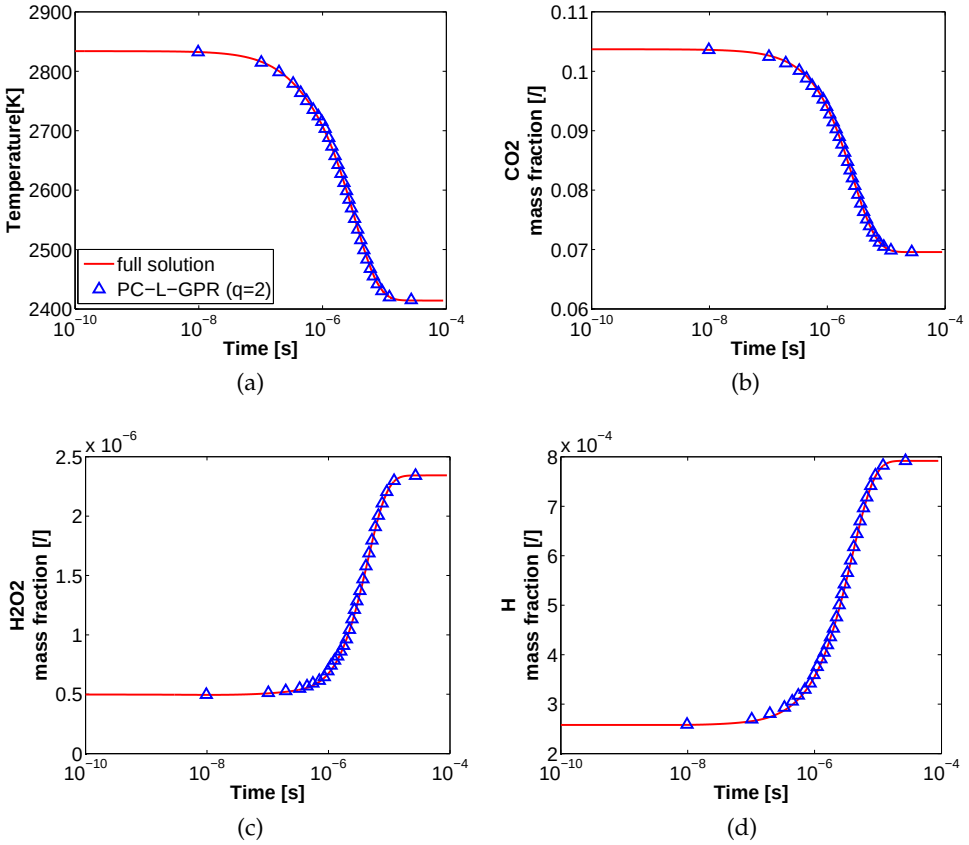


Figure 2.12: PSR temperature (a) and major and minor species (b-d) as a function of time (propane case, Polimi mechanism), for a residence time of $1 \cdot 10^{-5} s$, with the solid line representing the full solution. The markers represent the results for PC-score with local GPR regression using q = 2 PCs

2.5 Conclusion

The present chapter investigates the applicability of the PC-transport approach, focusing on the application of nonlinear regression to provide an accurate and compact parameterization of the thermo-chemical state. Steady and unsteady perfectly stirred reactor (PSR) calculations were carried out using the PC-transport approach, coupled to Gaussian Process Regression (GPR), for two different fuels (methane and propane) and three different kinetic mechanisms of increasing complexity.

The PC-GPR model showed its ability to produce very accurate representation of all state space variables, including temperature, major and minor species and source terms, using only a reduced number of principal components. In particular, for methane, the use of GPR allows to model accurately the system with only $q = 2$ principal components instead of the 34 variables in the original GRI-3.0 kinetic mechanism. For propane, the same approach lead to a very significant reduction, from 50 species, when using the San Diego mechanism, and 162 species, when using the Polimi mechanism, to only 2 PCs.

Moreover, the application of the PC-transport model using local nonlinear regression (PC-L-GPR) was demonstrated. The use of local regressions within bins improved the accuracy of the PC-GPR approach while decreasing the computational cost associated to the generation of the reduced model. In particular, the use of PC-L-GPR provided an optimized mapping of the thermo-chemical state and the corresponding source terms.

CHAPTER 3

COMBUSTION MODELING USING PRINCIPAL COMPONENT ANALYSIS: A POSTERIORI VALIDATION ON SANDIA FLAMES D, E AND F

After the *a priori* validation of the PC-GPR model using perfectly stirred reactors shown in Chapter 2, the aim of the present Chapter is to investigate, *a posteriori*, the potential of the PC-score approach coupled with nonlinear Gaussian Process Regression (GPR) in the framework of non-premixed turbulent combustion in a fully three-dimensional Large Eddy Simulation (LES).

3.1 Introduction

As shown in Chapter 2, 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. Transport equations for the scores were introduced by Sutherland and Parente [62] (PC-score approach). The method was later enhanced by combining PCA with nonlinear regression [41, 54, 65], 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. [30] and Echekki and Mirgolbabaei [18] provided the first a posteriori studies on the use of the PC-score approach. In particular, Isaac et al. [30] 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. [39] extended the study to methane and propane (cfr. Chapter 2), 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) [18], the Direct Numerical Simulation (DNS) of premixed syngas [14] and methane-air combustion [45], and to develop a framework for closure models based on experimental data [55, 56].

In this Chapter, the database for the PCA model training is based on 1D counter diffusion methane flames. The PC-GPR approach is then validated using the experimental data available for Sandia flames D, E and F [2]. To the authors' knowledge, the current work is the first attempt to use such an approach.

3.2 PC-score approach

Sutherland and Parente proposed a combustion model based on PCA and derived transport equations for the principal components [62]. Projecting the variables of interest, in this case the vector of species mass fractions $\mathbf{y} = [y_1, y_2, \cdots, y_{n_s}]$, onto the eigenvector matrix $\mathbf{A}$ gives:

$$\frac{\partial}{\partial t}(\rho \mathbf{z}) + \nabla(\rho u \mathbf{z}) = \nabla \cdot \mathbf{J_z} + \mathbf{S_z} \tag{3.1}$$

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. 1.5, $\mathbf{J_z}$ and $\mathbf{S_z}$ can be linearly related to the diffusive fluxes and source terms of $\mathbf{y}$:

$$\mathbf{S}_z = \mathbf{S_y A} \tag{3.2}$$

and

$$\mathbf{J_z} = \mathbf{J_y A}, \tag{3.3}$$

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, 18, 29, 30, 39] 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 1.4.

As for the diffusive fluxes, expressing $\mathbf{J_y}$ and $\mathbf{J_z}$ as:

$$\mathbf{J_y} = \rho \mathbf{D_y} \nabla \mathbf{y} \tag{3.4}$$

and

$$\mathbf{J_z} = \rho \mathbf{D_z} \nabla \mathbf{z}, \tag{3.5}$$

and noticing that $\nabla \mathbf{y} = \nabla \mathbf{z} \mathbf{A}^T$, one can express the score matrix of diffusion coefficients as

$$\mathbf{D_z} = \mathbf{A}_q^T \mathbf{D_y} \mathbf{A}_q, \tag{3.6}$$

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 [1]. This assumption was used in the present work, hence the matrix $\mathbf{D_z}$ is replaced by $\alpha = \frac{k}{(\rho\ c_p)}$, 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 [14, 42] and must be rotated to obtain a quasi-diagonal matrix of score diffusion coefficients. The final equation reads:

$$\frac{\partial}{\partial t}(\rho \mathbf{z}) + \nabla (\rho u \mathbf{z}) = \nabla \cdot \left(\frac{k}{c_p} \nabla \mathbf{z} \right) + \mathbf{S_z} \tag{3.7}$$

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 $\mathbf{A}$.

3.3 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.

3.3.1 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

Table 3.1: Conditions for Sandia flame experiments

	Flame D	Flame E	Flame F
Jet bulk velocity (m/s)	49.6	74.4	99.2
Pilot gas velocity (m/s)	11.4	17.1	22.8
Coflow velocity (m/s)	0.9	0.9	0.9
Reynolds number (–)	22,400	33,600	44,800
Fuel stream (–, K)	Y_{CH_4}=0.156, Y_{O_2}=0.196, Y_{N_2}=0.648, T=294		
Oxidizer stream (–, K)	Y_{O_2}=0.233, Y_{N_2}=0.767, T=291		
Pilot stream (–, K)	Y_{O_2}=0.054, Y_{N_2}=0.7342, Y_{CO_2}=0.1098, Y_{H_2O}=0.0942, Y_{OH}=0.0028, $Y_{CO} = 4.07 \times 10^{-3}$, $Y_O = 7.47 \times 10^{-4}$, $Y_{H_2} = 1.29 \times 10^{-4}$, $Y_H = 2.48 \times 10^{-5}$, T=1880		
Stoichiometric condition (–)	f_{st}=0.351		
Pressure (bar)	1.00616		
Inner jet diameter (mm)	7.2		
Inner pilot diameter (mm)	7.7		
Outer jet diameter (mm)	18.2		

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 294K. This main jet is surrounded by a pilot jet (∅18.2 mm) at 1880K, 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 291K 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 [2] through Rayleigh measurements for the temperature, and Raman and LIF measurements for mass fractions of chemical species. The available data consist of the mean and root mean square (rms) of temperature and mass fractions of major (CH_4, H_2, $H_{2,}O$, 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 [59]. The operation conditions and geometrical configuration of the Sandia flame experiments are summarized in Table 3.1.

3.3.2 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 [15, 16]. The GRI 3.0 [60] mechanism, involving 35 species and 253 reactions (excluding NOx), 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 3.3.1). 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.

3.3.3 Determination of the PCA basis

The PCA basis is generated using two approaches:

- (i) using the whole set of species (35)
- (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 [30, 41].

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 [39] and in Chapter 2 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 3.5.

The basis matrix weights obtained from the PCA analysis on the major species are as follows for the first two PC's:

$$Z_1 = -0.02 \cdot Y_{H_2O} - 0.18 \cdot Y_{O_2} - 0.64 \cdot Y_{N_2} + 0.73 \cdot Y_{CH_4} - 0.02 \cdot Y_{CO_2} \tag{3.8}$$

and

$$Z_2 = 0.51 \cdot Y_{H_2O} - 0.67 \cdot Y_{O_2} - 0.01 \cdot Y_{N_2} - 0.14 \cdot Y_{CH_4} + 0.5 \cdot Y_{CO_2}. \tag{3.9}$$

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 [6], f, as shown on Fig. 3.1a. Therefore, in the numerical simulation, Z_1 is directly replaced by the mixture fraction, to avoid transporting a reactive scalar.

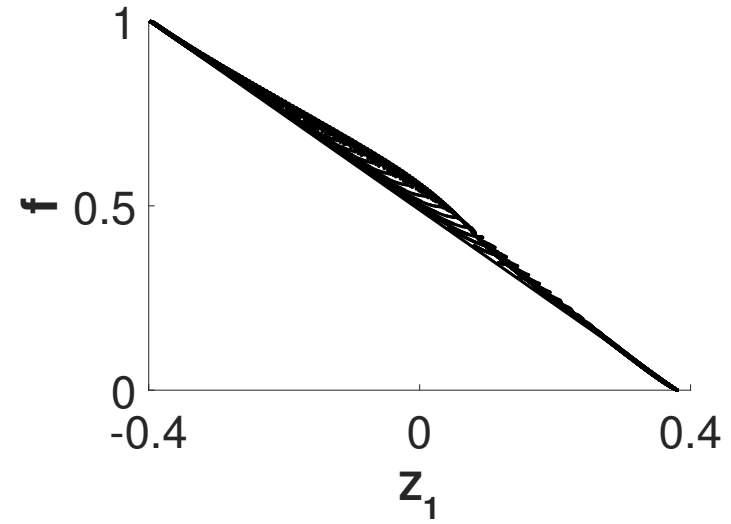
The weights for Z_2 also show an interesting pattern: a positive correlation for H_2O and CO_2, and a negative correlation for CH_4, O_2 and N_2. This can be linked to a progress variable, where products have positive stoichiometric coefficients and reactants negative ones, as shown in Fig. 3.1b. It 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. Figure 3.2a shows source term S_{z_2} as a function of Z_1 and Z_2, and Fig. 3.2b shows the regression of that manifold ($R^2 = 99.28\%$).

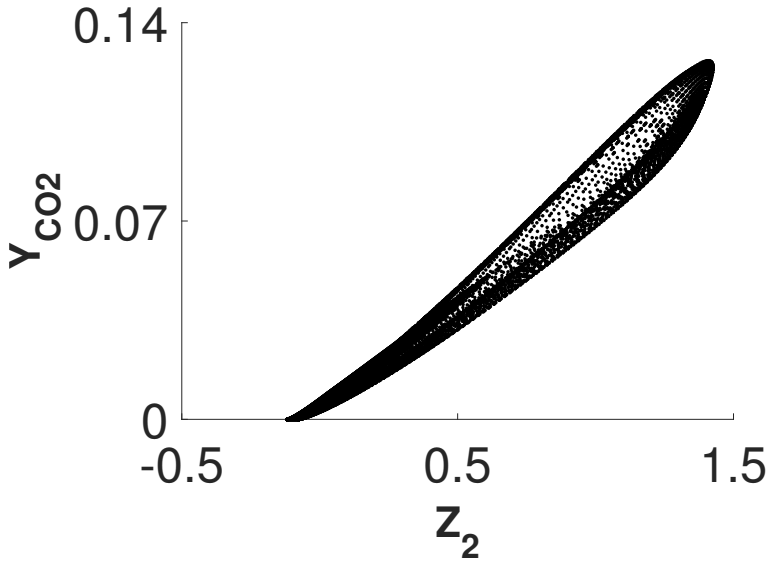
3.4 Elemental mass analysis

It was shown in Section 3.3.3 that the first PC is highly correlated with mixture fraction, i.e. with the elemental mass fractions of C, H, and O. The dependance of the principal components on the elemental mass fractions can be further analyzed by finding the overlap between conservative modes and the PCs of a system. To this end, the conservative modes can be computed for the chemical species in a system using [29]:

$$C_{i,j} = e_{i,j} d_i$$



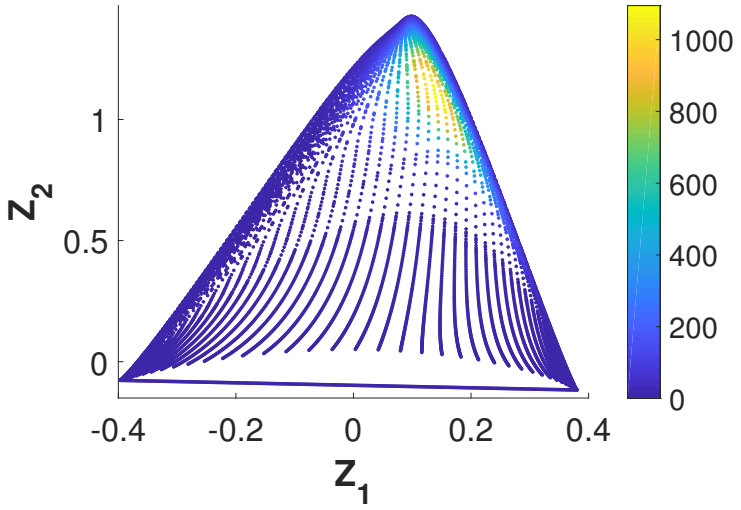
(a)



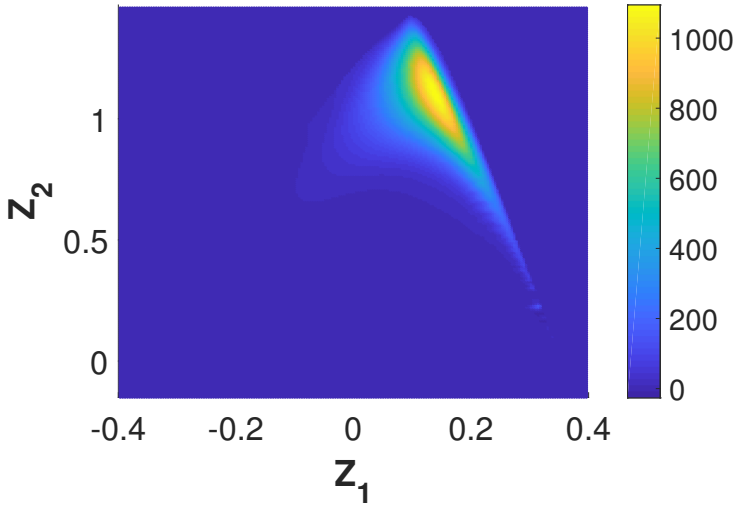
(b)

Figure 3.1: Scatter plot; Z_1 is correlated with mixture fraction (a) and Z_2 with the progress of reaction (b)

where $e_{i,j}$ represents the mass fraction of jth element in the ith species, and d_i is the scaling factor used for each of those species (which therefore depends on the scaling method used to compute the PC basis). If the PCs of the system are predominately a linear combination of conservative modes an approximation of the PCs can be given using a simple least



(a)



(b)

Figure 3.2: S_{z_2} in function of $Z_1$and Z_2, the original manifold from PCA (a) and the regressed manifold with GPR (b)

squares regression [29]:

$$\widetilde{\mathbf{A}} = \mathbf{C}\left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{A}$$

where $\mathbf{A}$ is the original PC basis matrix. In order to find the overlap between the conservative modes and the principal components, one can compute the dot product between $\widetilde{\mathbf{A}}$ and $\mathbf{A}$. If $\widetilde{\mathbf{A}} \cdot \mathbf{A} = 1$, the PCs fall completely in conservative space. If $\widetilde{\mathbf{A}} \cdot \mathbf{A} = 0$, the the PCs are not related to the conservative modes.

A calculation of the dot product was done on the reference data set using the whole set of species. The results are shown in Table 3.2, where a comparison is made between two different scaling methods, namely pareto and standard (auto scaling). It can be seen that using pareto, the first PC is highly correlated with the elemental mass fractions, which allows to recover the previously shown result that the first PC was highly correlated with mixture fraction (Figure 3.1a). On the other hand, standard scaling does not identify the first PC as falling in conservation space. With pareto, the PCs correlated with conservation laws are PCs #1, #4 and #31. With standard scaling, the PCs that correlate the most with the elements are PCs #5, #33 and #34.

Figure 3.3 shows contour plots for the dot product between $\widetilde{\mathbf{A}}$ and $\mathbf{A}$ for the first and second modes obtained from the flame D simulation. It is interesting to observe that the dot product is equal to 1 in the rection zone for the first mode (Fig. 3.3a) as the first PC is basically mixture fraction, whereas the second mode (Fig. 3.3b) is close to 0 in the reaction zone as the second PC is more correlated with a progress variable.

3.5 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. 3.1). 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):

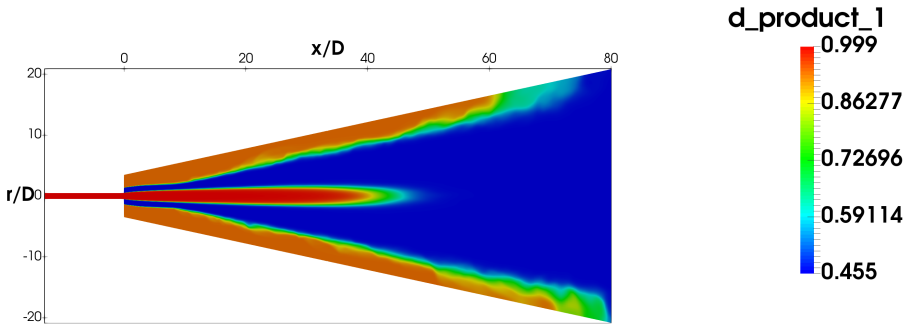
$$\mathbf{z}\,|_{\mathbf{boundary}} = \mathbf{X}\,|_{\mathbf{boundary}}\,\mathbf{A} \tag{3.10}$$

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

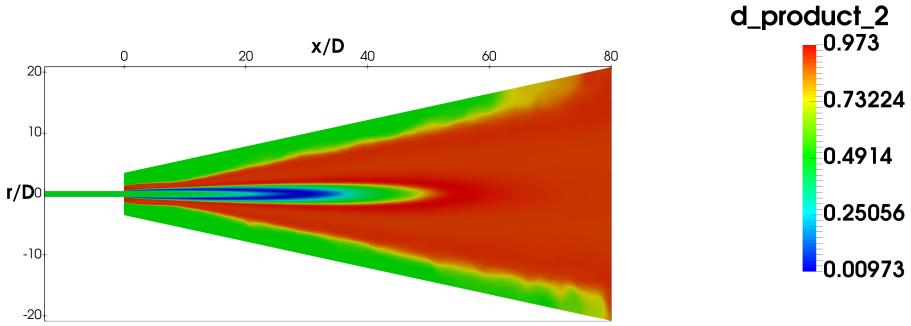
Table 3.2: Dot product between $\widetilde{\mathbf{A}}$ and $\mathbf{A}$ for the counterflow data set (35 species), using pareto and standard (auto) scalings

$\widetilde{\mathbf{A}} \cdot \mathbf{A}$	pareto	standard
1	0.9999	0.0017
2	0.0403	0.0023
3	0.0181	0.0059
4	0.5769	0.1125
5	0.1291	0.8067
6	0.1091	0.0224
7	0.0441	0.0471
8	0.0726	0.0015
9	0.0301	0.0013
10	0.0060	0.0017
11	0.0241	0.0027
12	0.0500	0.0201
13	0.0108	0.0118
14	0.0098	0.0178
15	0.1946	0.0024
16	0.0207	0.0002
17	0.0087	0.0240
18	0.4362	0.0133
19	0.0098	0.0028
20	0.1100	0.0047
21	0.0189	0.0196
22	0.0062	0.0056
23	0.0831	0.0054
24	0.0196	0.0819
25	0.0001	0.2351
26	0.0042	0.0143
27	0.0007	0.2015
28	0.0003	0.1864
29	0.0004	0.0020
30	0.0000	0.1316
31	0.9654	0.0163
32	0.0000	0.0019
33	0.0000	1.0000
34	0.0000	0.9750
35	0.0000	0.0203

(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 [35]. 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



(a) first mode



(b) second mode

Figure 3.3: Contour plots of $\widetilde{\mathbf{A}} \cdot \mathbf{A}$ for the first mode (a) and second mode (b) from flame D simulation

sensitivity analysis was carried out and the results proved to be mesh independent. The Pope criterion was used to verify the impact of the grid resolution. The criterion is defined as the following ratio:

$$M_E = \frac{k_{res}}{k_{sgs} + k_{res}}$$

where k_{sgs} represents the mean subgrid turbulent kinetic energy provided by the SGS model, and $k_{res} = \frac{1}{2} \left\langle (\widetilde{u_i} - \langle \widetilde{u_i} \rangle_T) (\widetilde{u_i} - \langle \widetilde{u_i} \rangle_T) \right\rangle_T$ is the mean resolved kinetic energy where $\langle \rangle_T$ indicates time-averaged quantities. 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:

$$\rho \frac{D\widetilde{\mathbf{z}}}{Dt} = \nabla \cdot \left[\left(\frac{\mu}{Sc} + \frac{\mu_t}{Sc_t} \right) \nabla \widetilde{\mathbf{z}} \right] + \overline{\mathbf{S}}_\mathbf{z} \tag{3.11}$$

where the SGS Reynolds stresses were modelled using the WALE model [44], 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, is tabulated in function of the PCs, while the dynamic viscosity, μ_t, was obtained through the WALE model. $\overline{\mathbf{S}}_\mathbf{z}$ is the filtered score source vector.

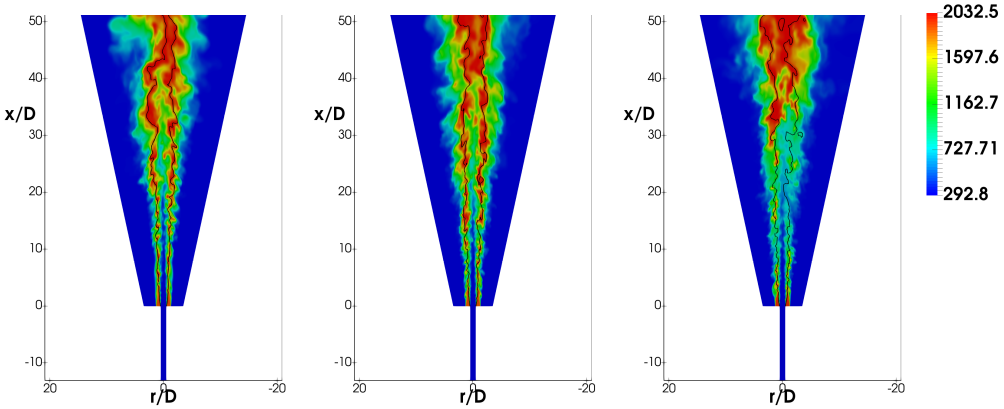
3.6 Results and discussion

The results of the PC approach on flames D-E-F are discussed in the present Section. First, a qualitative analysis is performed using the instantaneous and mean flow field structure (Section 3.6.1. After that, statistical flow field results are presented for the three flames (Section 3.6.2, 3.6.3 and 3.6.4). Flame D is used as a base case in order to verify the influence of several parameters: 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. In Section 3.6.5, the amount of extinction and reignition happening in flame E and F and the ability of the model to predict it is quantified in more detail. Finally, the performance of the PC-GPR model are compared against the FPV model in Section 3.6.6. For reference, experimental data are denoted by symbols and computational results are shown by lines. The different simulations were run for at least 10 flow through periods, in order to have a sufficiently large averaging window.

3.6.1 Instantaneous and mean flow field structure

Instantaneous temperature fields from the simulations of flames D, E and F are illustrated on Fig. 3.4. The solid line in these figures corresponds to the isocontour of the stoichiometric mixture fraction ($f_{st} = 0.351$). It can be observed on Fig. 3.4a for flame D that the shear layer in the near-burner region around the jet core is mainly composed of quasi-laminar structures, which can be attributed mainly to the increase in the molecular properties due to the heat release [27], whereas for flame E and F (Fig. 3.4b and 3.4c respectively), those near-burner structures appear to be more wrinkled. The structure of the flow in the $3 \leq x/D \leq 10$ region of the flame exhibits noticeable differences between the three flames: the amount of large scale oscillations are very limited for flame D, whereas flame E and F show more wrinkled structures around the contour of stoichiometric mixture fraction. Locally extinguished flame regions at stoichiometric condition are more frequentt

for flame E (Fig. 3.4b), and even more for flame F (Fig. 3.4c) for which those extinguished pockets extend further downstream. This can be explained by the increased levels of turbulence (therefore also mixing) in flame E and F, where the local diffusion time scales are much lower compared to flame D. As a consequence, non-equilibrium effects become visible. The diffusion of heat towards the surrounding is not always compensated by the production of heat due to the reactions, leading to the formation of extinction pockets where the fuel and oxidizer are in a premixed state. Those regions can reignite afterwards if sufficient heat and radicals are entrained from their surroundings [51, 52].



(a) flame D (b) flame E (c) flame F

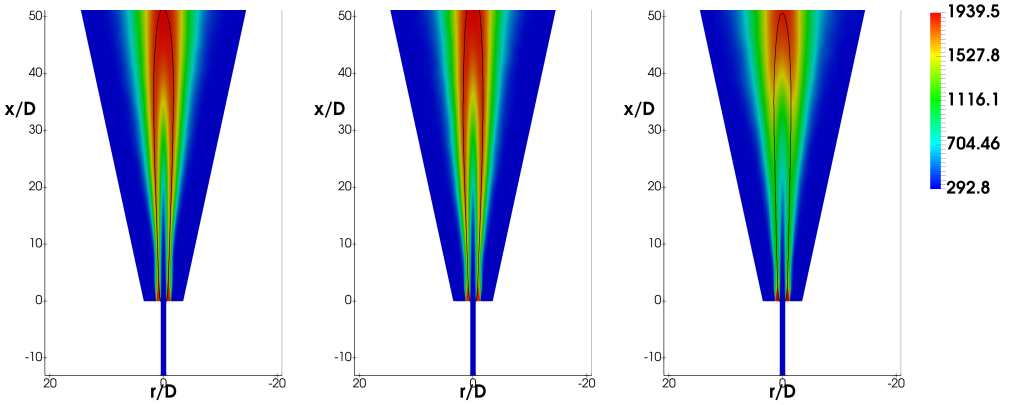
Figure 3.4: Instantaneous temperature fields for (a) flame D, (b) flame E and (c) flame F computed with the PC-GPR model. The solid line shows the location of stoichiometric mixture fraction.

Figure 3.5 shows the Favre averaged temperature fields from the simulations of flames D, E and F. Flame D and E exhibit very similar averaged temperature field, whereas for flame F, the transitional region of the flame extends further downstream. The temperature in the $3 \leq x/D \leq 10$ region of flame F is much lower than the two other flames due to the enhanced local extinction.

3.6.2 Results for flame D

3.6.2.1 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. 3.6 that the PCA-GPR model is able to reconstruct all variables with great accuracy.



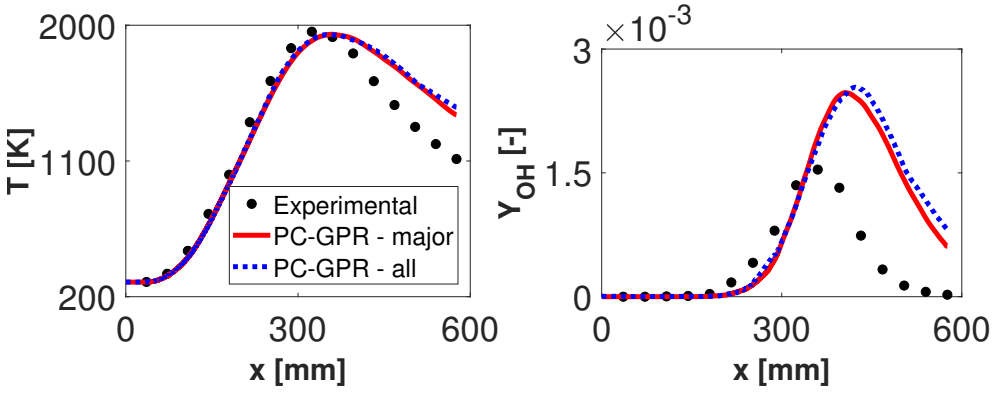
(a) flame D (b) flame E (c) flame F

Figure 3.5: Averaged temperature fields for (a) flame D, (b) flame E and (c) flame F computed with the PC-GPR model. The solid line shows the location of stoichiometric mixture fraction.

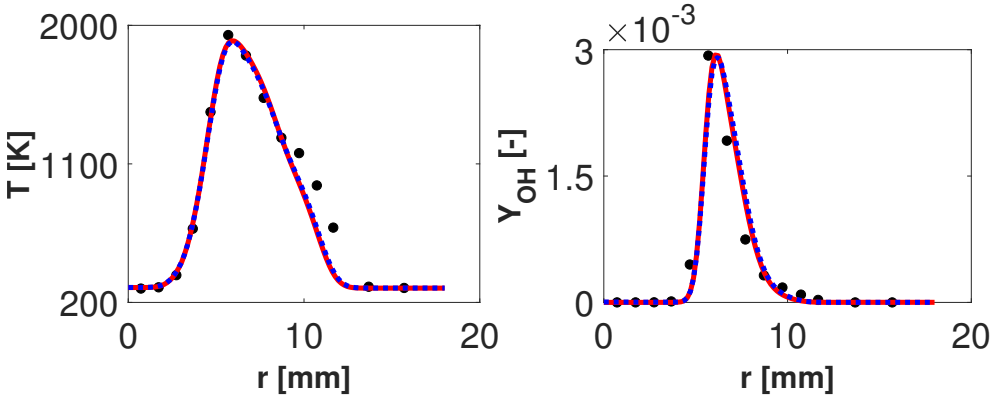
Moreover, both PCA bases provided comparable results, at all locations. Whether looking at the centerline (Fig. 3.6a), close to the burner exit (Fig. 3.6b) or further downstream (Fig. 3.6c and d), the mean profiles obtained with the two bases do not show any significant discrepancy. It can be then argued that using only the major species in order to build the PC basis results in no major loss of information, supporting the findings in [18].

Figure 3.6a 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. This overprediction is discussed in more details in Section 3.6.6. Figure 3.7 shows the same plots for mixture fraction and CO_2. It can be observed on Fig. 3.7a that the decay of mixture fraction is slightly underpredicted in the lean region of the flame, above the stoichiometric flame length. In the same way, the production of CO_2 is slightly overpredicted in that region. The radial profiles shown on Fig. 3.7 are in good agreement with the experimental data.

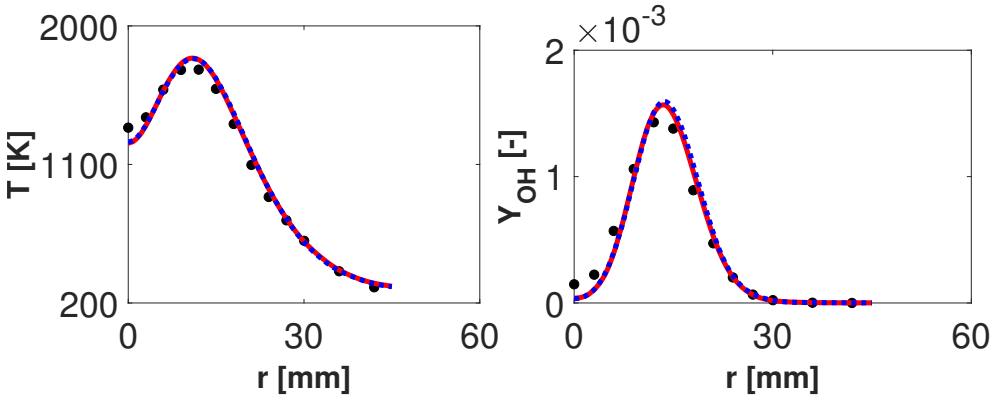
A comparison of radial profiles for CO and H_2 on the centerline and at $x/D = 7.5$ and 15 is also shown on Fig. 3.8. On the centerline, the intermediate CO is overpredicted in the lean part of the flame. As CO comes primarly from CH_4, this overprediction can be attributed by the overprediction of the fuel consumption in this region, which manifest itself by the overprediction of temperature as well. Radially, the profiles agree relatively well with the experiments. The peaks of H_2 are accurately predicted both at $x/D = 7.5$



(a) centerline

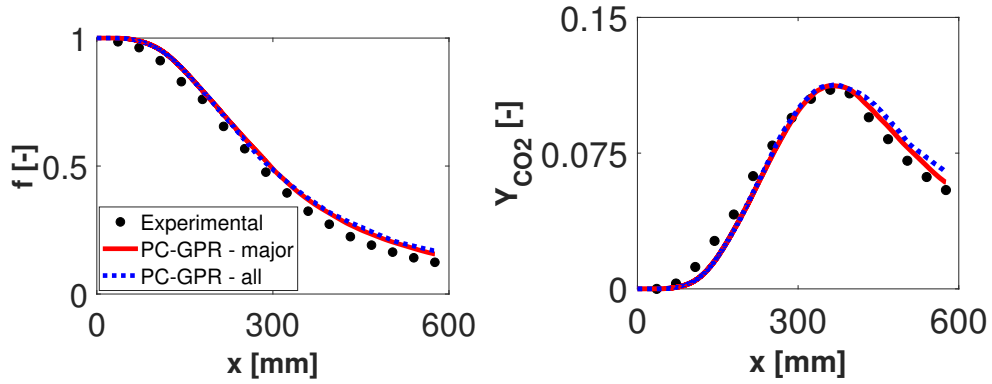


(b) $x/D = 3$

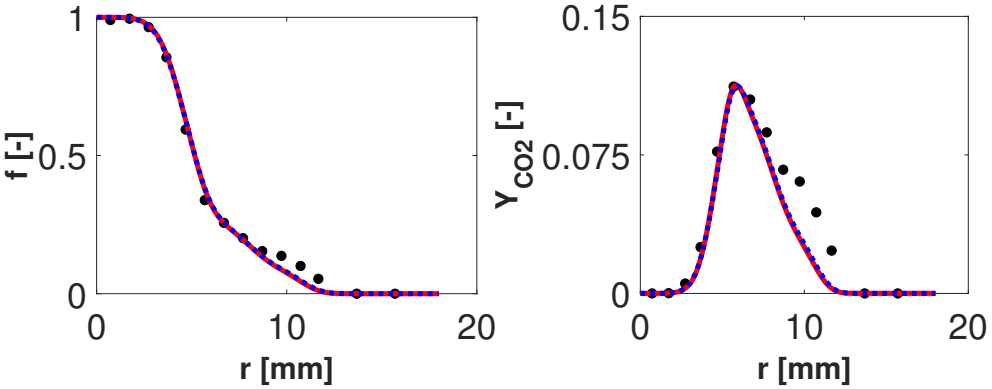


(c) $x/D = 30$

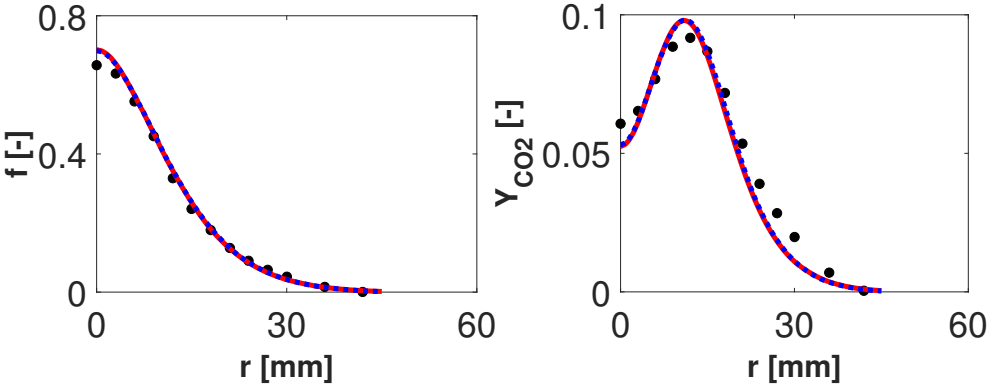
Figure 3.6: Comparison between the PC basis calculated using the major species (PC-GPR - major) and the basis obtained using the full set of species (PC-GPR - all). Results show the axial (a) and radial profiles (b-c) for temperature (T) and OH mass fractions.



(a) centerline



(b) $x/D = 3$



(c) $x/D = 30$

Figure 3.7: 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 mixture fraction (f) and CO_2 mass fractions.

and 15, while the centerline values are slightly overpredicted.

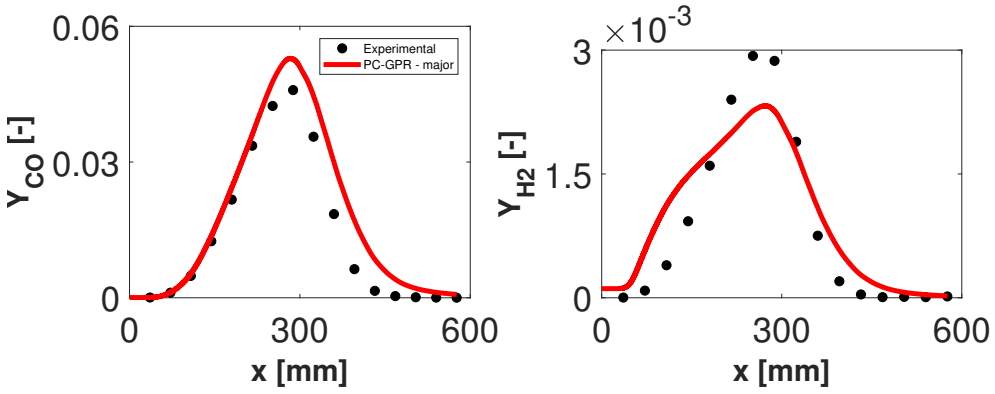
A more significant comparison can be made looking at the profiles of temperature conditioned on mixture fraction at axial locations $x/D = 60$ and 75 ($x = 432$ and 540 mm) shown on Fig. 3.9. It can be observed that the predicted temperature lies well inside the single shot experimental data points. 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. Figure 3.10 shows the RMS centerline profiles for temperature, mixture fraction and species mass fraction. It can be observed that the model is able to predict the peaks and the location of the peaks quite accurately, for all variables.

Figure 3.11a shows the manifold accessed during the simulation with major species at $t = 1s$, 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.

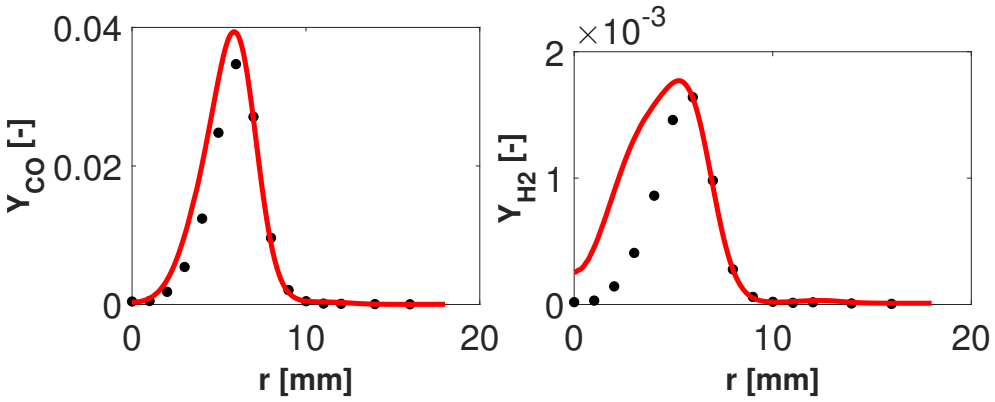
3.6.2.2 Sensitivity to kinetics and subgrid closure

The impact of the kinetic mechanism was also assessed. The GRI 3.0 mechanism was compared to the KEE-58 mechanism [6]. 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 thermo-chemical state-space, and a table was generated using the same grid spacing. Figure 3.12 shows a comparison of different axial and radial profiles using the 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.

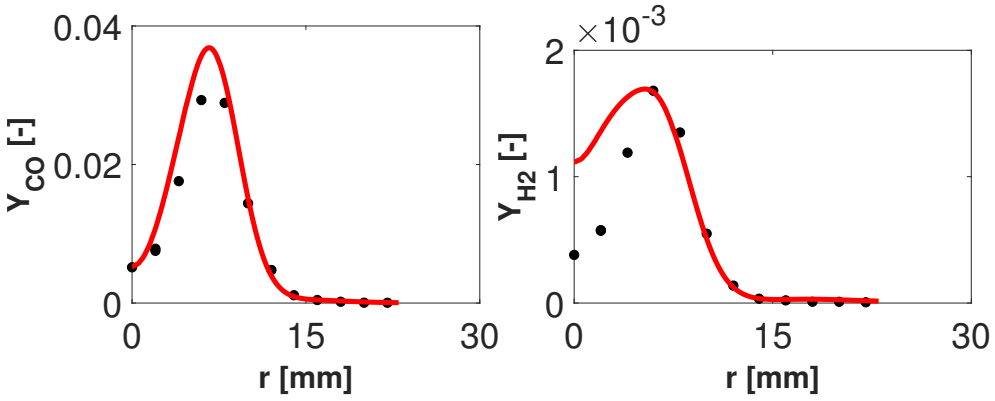
The sensitivity to a subgrid closure was also investigated. A mean value closure for the filtered PC's source terms ($\overline{\mathbf{S}_\mathbf{z}}$) might not be sufficient, and the influence of small-scale



(a) centerline

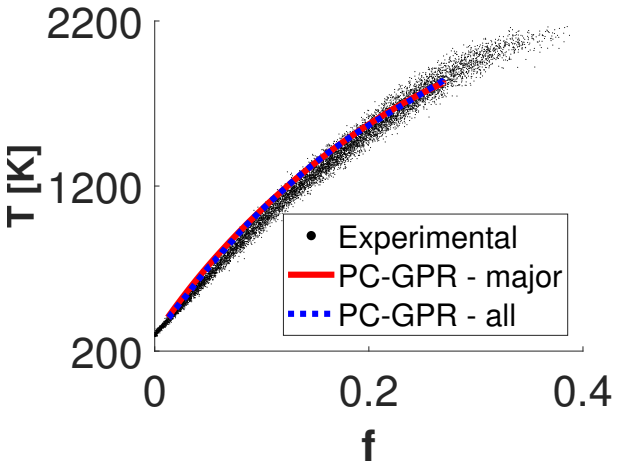


(b) $x/D = 7.5$

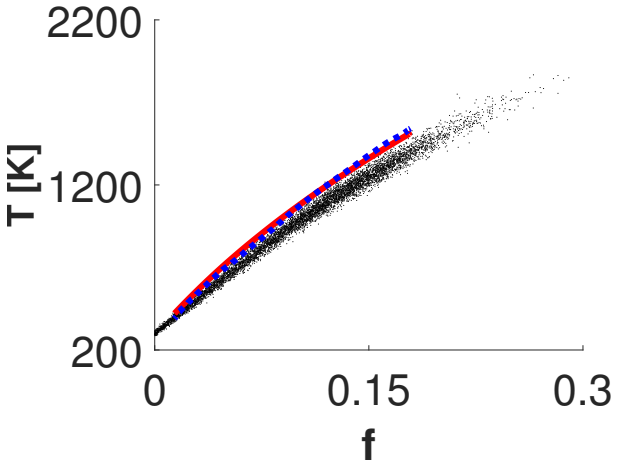


(c) $x/D = 15$

Figure 3.8: Flame D: CO and H_2 radial profiles plotted againt the experiments - (a) centerline, (b) $x/D = 7.5$ and (c) $x/D = 15$ (PC basis using the major species)

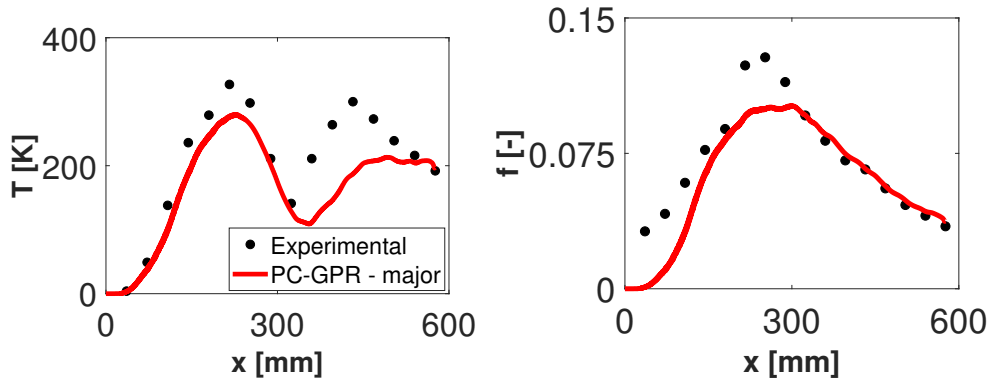


(a) $x/D = 60$

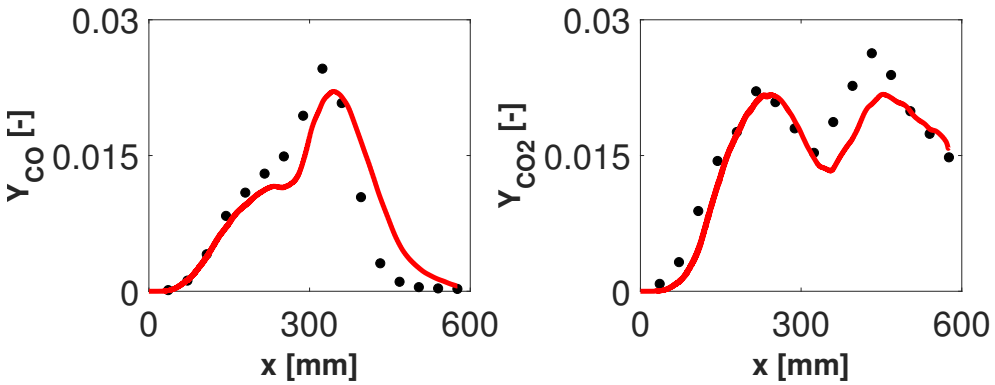


(b) $x/D = 75$

Figure 3.9: 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

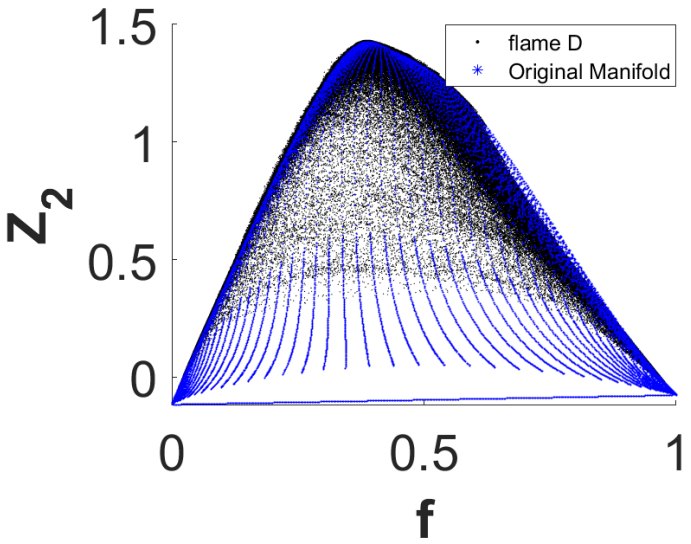


(a) Temperature (left) and mixture fraction (right)

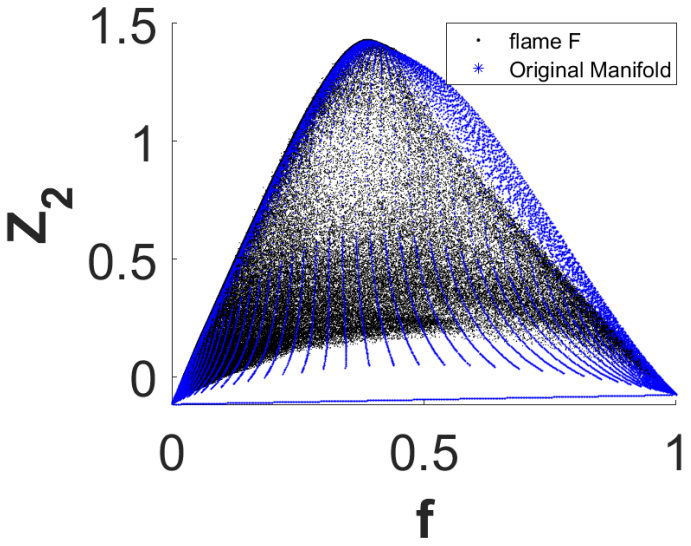


(b) CO (left) and CO2 (right) mass fraction

Figure 3.10: flame D - RMS centerline profile of temperature (T), mixture fraction (f), CO and CO_2 mass fractions plotted against the experiments.



(a) flame D vs original manifold



(b) flame F vs original manifold

Figure 3.11: 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.

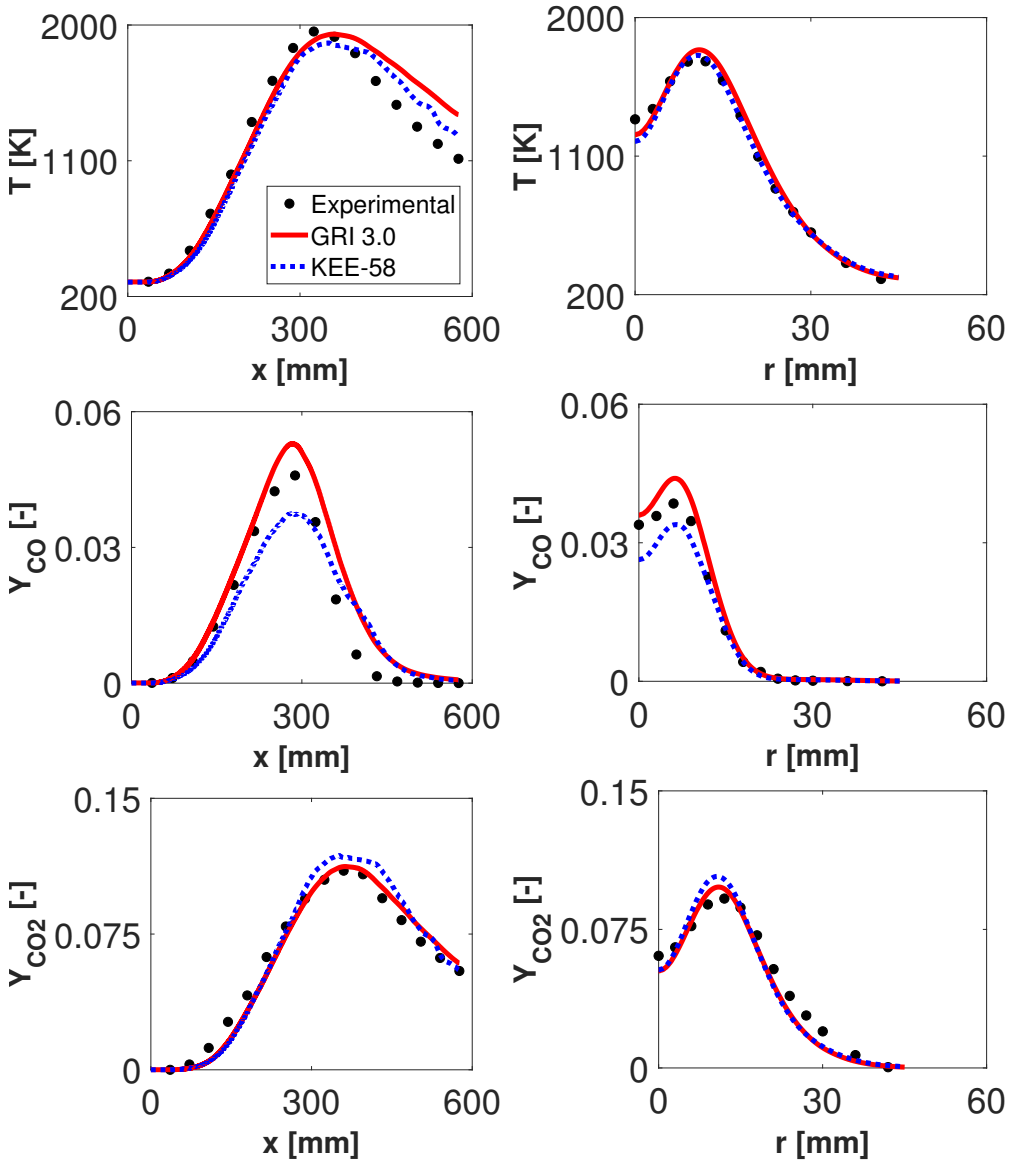


Figure 3.12: Comparison between GRI 3.0 and KEE-58 mechanisms on the centerline (left) and at radial location $x/D = 30$ (right) for temperature, CO and CO_2 mass fractions

turbulent fluctuations on the large scales must be assessed. Therefore, a *beta*-shaped probability density function (β-PDF) was used to represent the necessary scalar fluctuations. A transport equation for the mixture fraction variance was used following the approach in [27]. Results shown in Fig. 3.13 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.

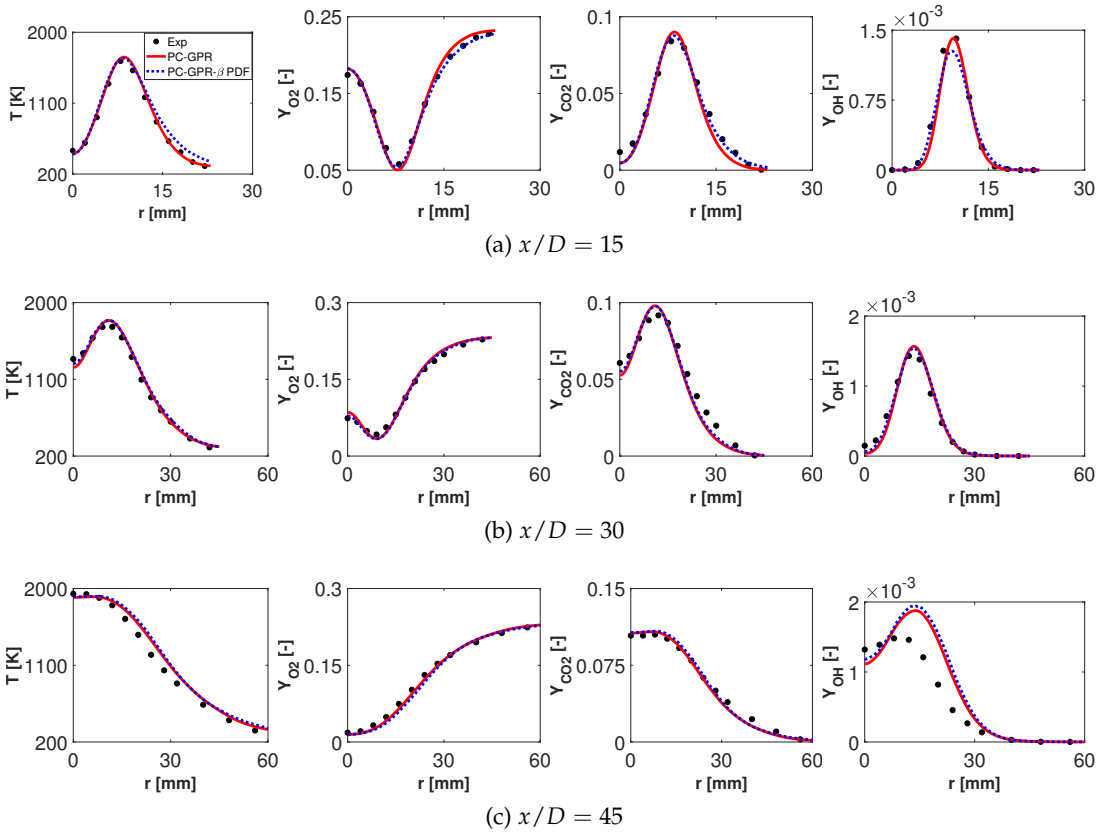
3.6.3 Results for flame E

To demonstrate the potential of the PC-GPR on more challenging cases, simulations of Sandia flame E and F were carried out. The results for flame E are presented in this Section, and the results for flame F are in the next Section.

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.

Figure 3.14 shows the temperature and some species mass fraction profiles (CH_4, CO_2 and OH) on the centerline and radial profiles at different axial locations ($x/D = 15$ and 30) for flame E, confirming the ability of the model to reconstruct all scalar variables with great accuracy. Similarily to flame D, the temperature is slightly overpredicted on the centerline near the outlet of the domain, which results in a slight overprediction of CO_2 as well. Radially, it can be observed that the profiles agree quiet well with the experiments. The peaks of temperature, as well as those of the species, are accurately captured by the PC-GPR model. The model predicts also remarkably well the dacay of those scalars in the lean part of the radial profiles.

Figure 3.15 shows the centerline and radial profiles of flame E ($x/D = 7.5$ and 15) for CO and H_2, obtained with the PC-GPR model using the major species basis. Similarily to flame D, the intermediate CO is overpredicted in the lean part of the flame. Again, this can be linked to the overprediction of temperature in that zone. The minor species H_2 is overpredicted close to the centerline, while its peaks are correctly captured at $x/D = 7.5$

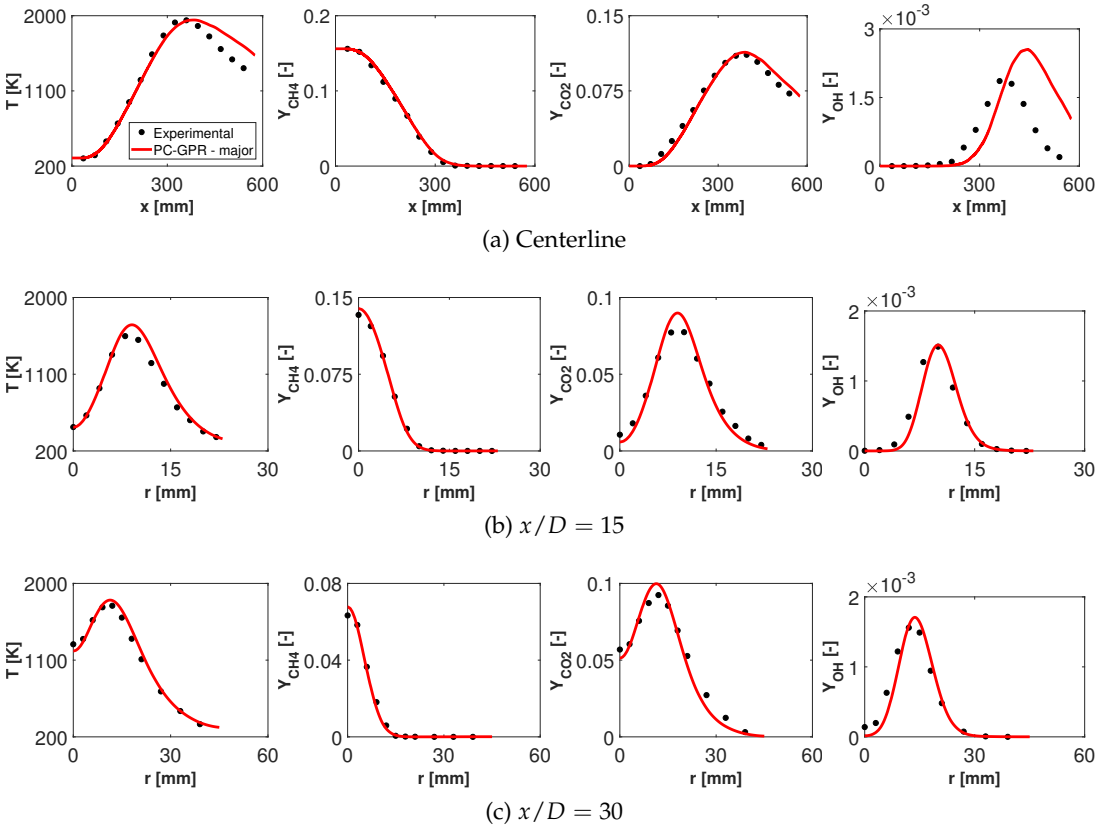


(a) $x/D = 15$

(b) $x/D = 30$

(c) $x/D = 45$

Figure 3.13: Comparison between the PC-GPR without subgrid closure and the PC-GPR with a beta-PDF closure. Results show the axial (a) and radial profiles (b-c) for temperature, O_2, CO_2 and OH mass fractions.



(a) Centerline

(b) $x/D = 15$

(c) $x/D = 30$

Figure 3.14: Flame E: temperature and major and minor species profiles plotted againt the experiments - centerline (a), $x/D = 15$ (b) and $x/D = 30$ (c)

and 15. The predictions of the model regarding strong local extinction and reingnition in flame E are further analyzed in Section 3.6.5 using scatter plots of temperature and CO mass fraction.

3.6.4 Results for flame F

Figure 3.16 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. Figure 3.17 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.

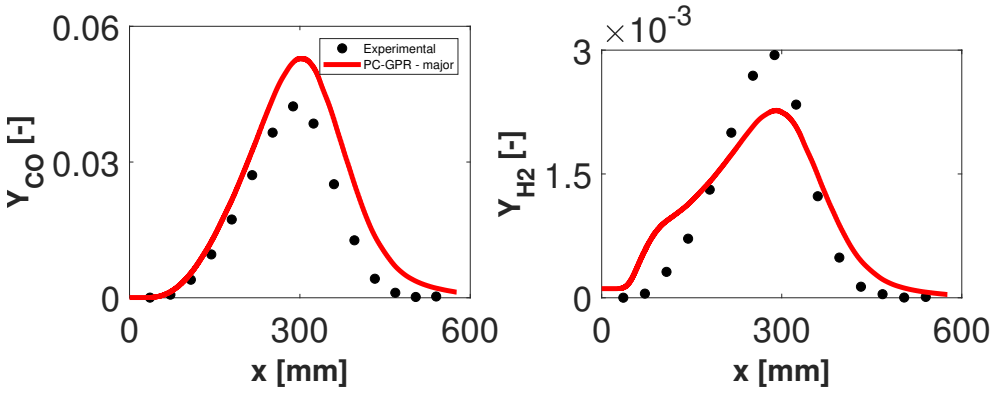
The same conclusion can also be drawn looking at the profiles in Fig. 3.18, showing the profiles of conditional mean of temperature and species mass fraction on mixture fraction.

Furthermore, as expected from the experimental data, it can be observed from Fig. 3.11b 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.

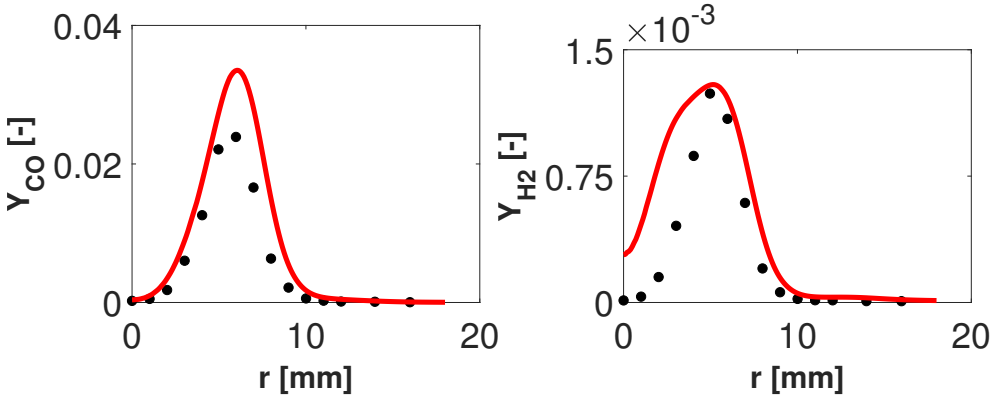
Figure 3.19 shows radial statistics for flame F at different axial locations for temperature and slelected species. It can be observed that the PC-GPR model accurately predicts the peak and decay in radial directions as well.

The predictions of the PC-GPR model regarding strong local extinction and reingnition phenomena in flame F are further analyzed in Section 3.6.5 using scatter plots of temperature and CO mass fraction.

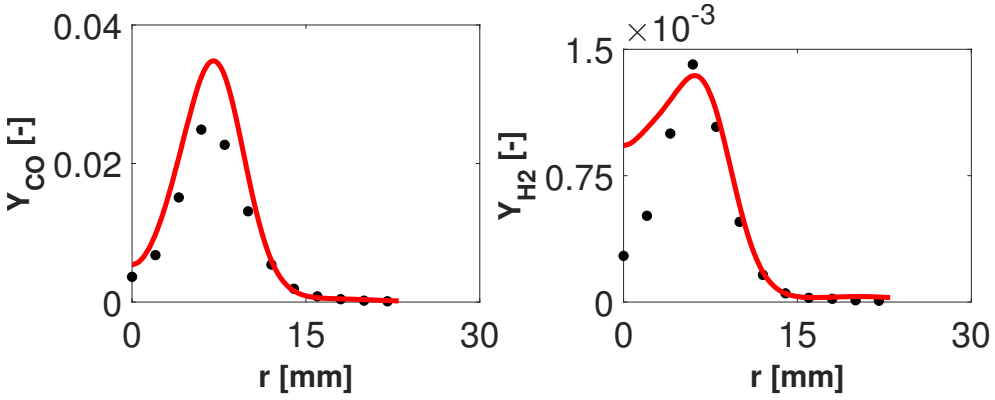
Figure 3.20 shows the centerline and radial profiles of flame F (at $x/D = 7.5$ and 15) for CO and H_2, obtained with the PC-GPR model using the major species basis. Similarily to flame D and E, the intermediate CO is overpredicted in the lean part of the flame for the same reason. The location of the peaks of CO are well captured both radially and on the centerline, while the value is slightly overpredicted at $x/D = 7.5$. The minor species H_2 is overpredicted close to the centerline at $x/D = 7.5$ and 15, and at 15 the peak value is not



(a) centerline

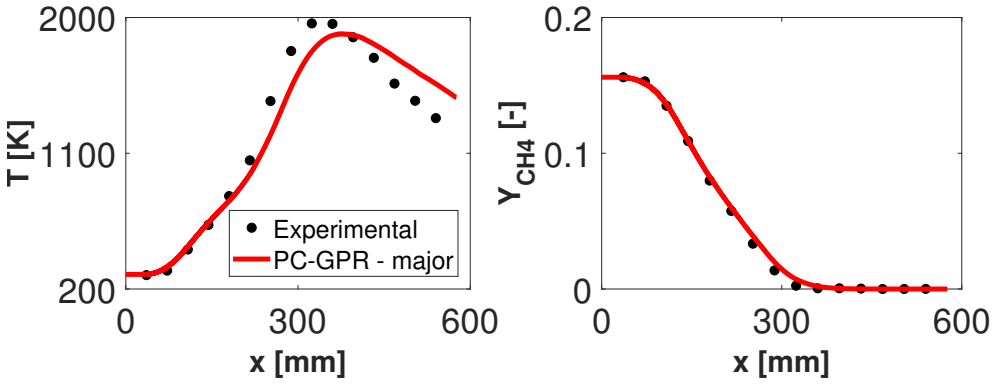


(b) $x/D = 7.5$

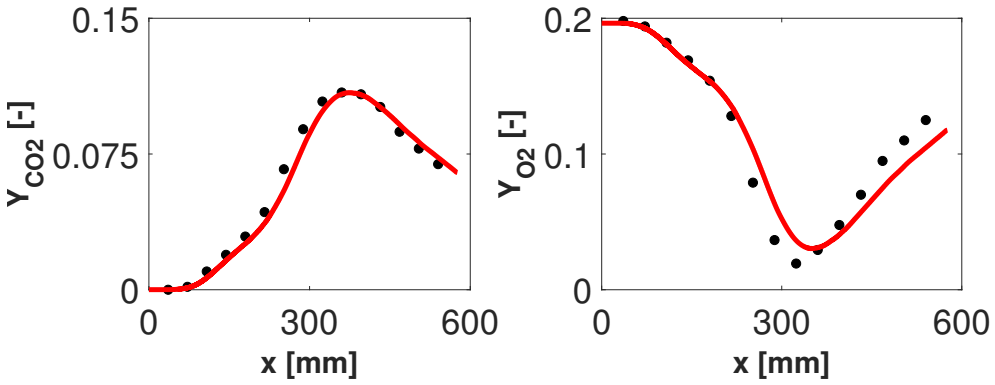


(c) $x/D = 15$

Figure 3.15: Flame E: CO and H_2 radial profiles plotted againt the experiments - (a) centerline, (b) $x/D = 7.5$ and (c) $x/D = 15$ (PC basis using the major species)

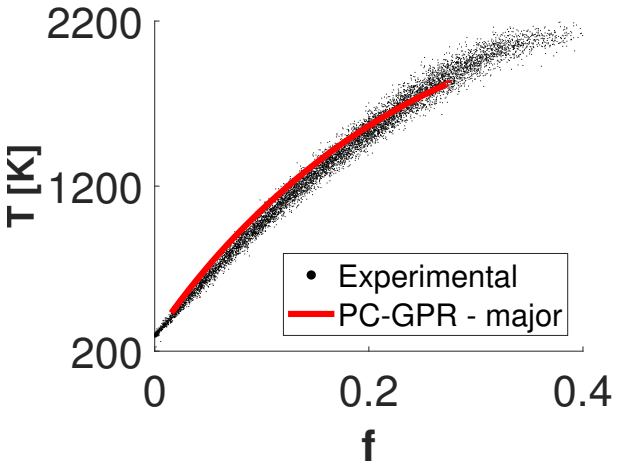


(a) Temperature (left) and CH_4 mass fraction (right)

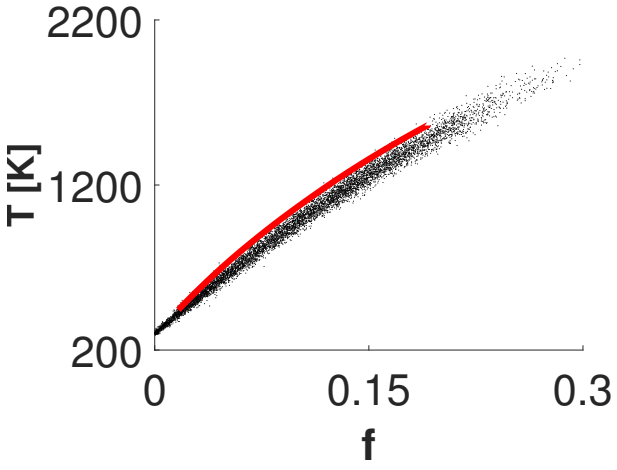


(b) CO_2 (left) and O_2 mass fraction (right)

Figure 3.16: Flame F: temperature and major species profiles plotted against the experiments - centerline

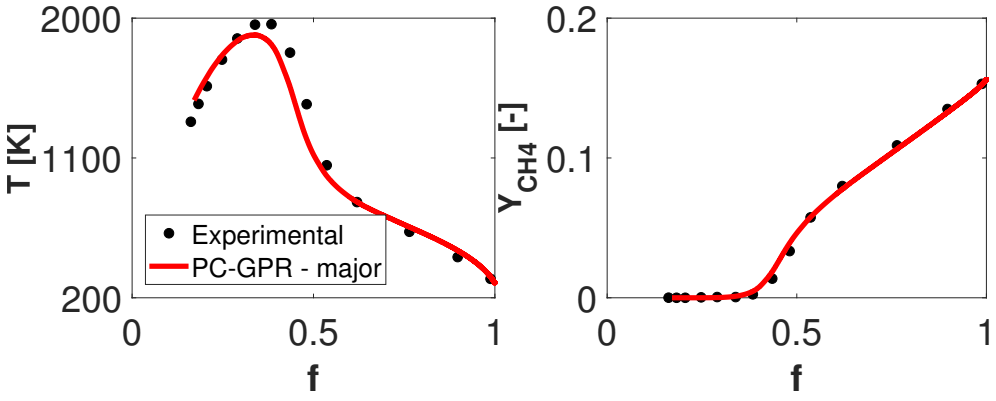


(a) $x/D = 60$

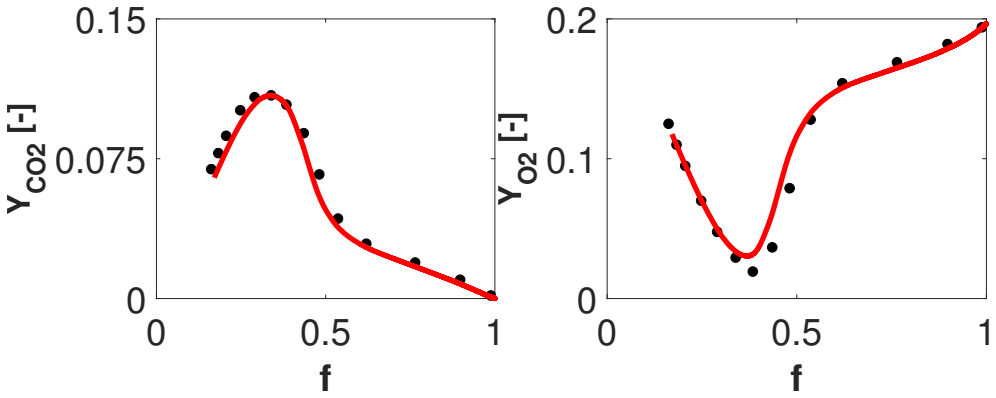


(b) $x/D = 75$

Figure 3.17: Flame F - conditional averages at different downstream positions plotted against the single shot experimental data

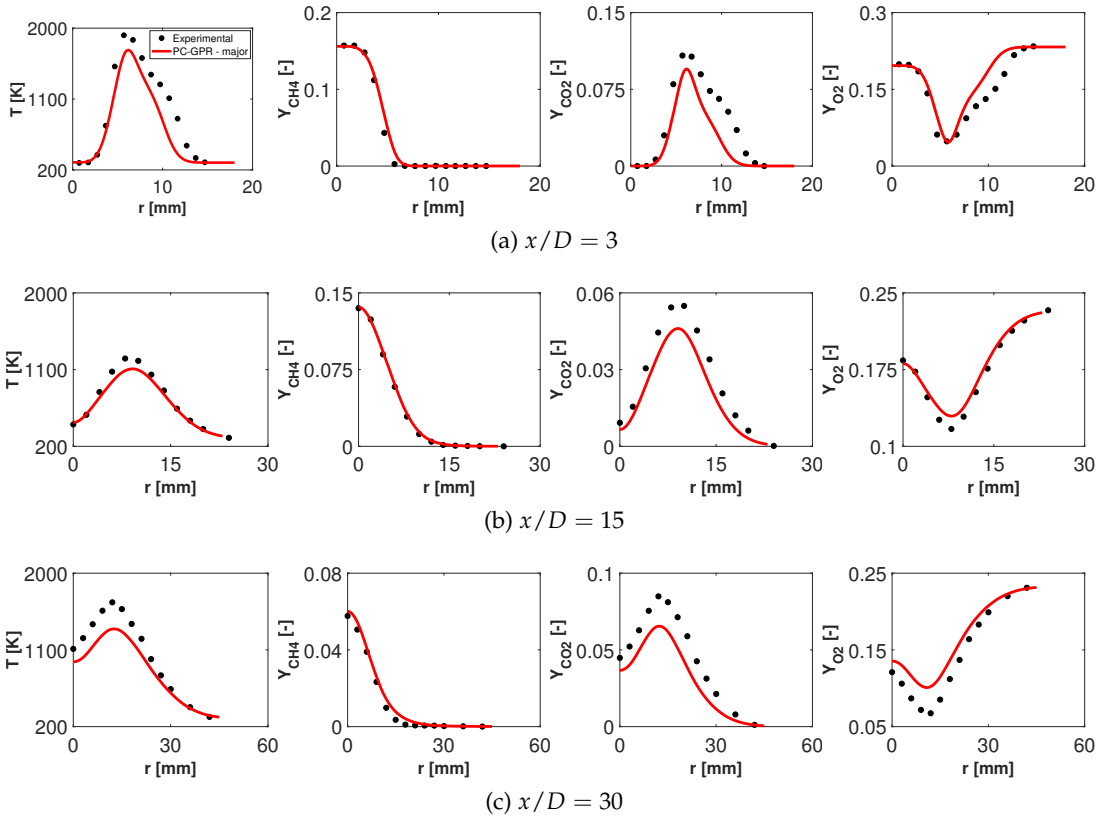


(a) Temperature (left) and CH_4 mass fraction (right)



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

Figure 3.18: Flame F: Conditional averages of temperature and species on the mixture fraction - centerline



(a) $x/D = 3$

(b) $x/D = 15$

(c) $x/D = 30$

Figure 3.19: Flame F: radial temperature and species profiles plotted againt the experiments - $x/D = 3$ (a), $x/D = 15$ (b) and $x/D = 30$ (c)

well captured.

3.6.5 Extinction and reignition quantification

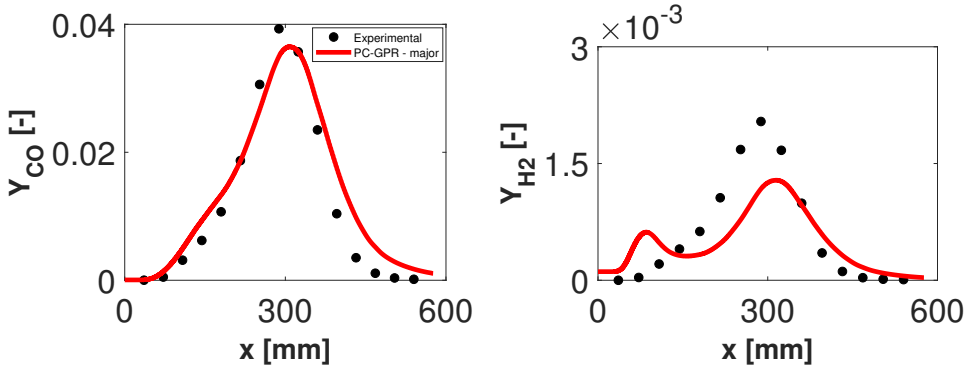
In order to qualitatively assess the degree of extinction/reignition happening in flame E and F, scatter plots for temperature versus mixture fraction at three different radial locations (i.e. $x/D = 7.5, 15$ and 30) are shown in Fig. 3.22 and 3.23 for flame E and F respectively. The scatter plots for flame D are also shown on Fig. 3.21 for comparison. It should be pointed out that these plots merely show instantaneous data from the simulations, but are nevertheless useful for a qualitative assessment of the instantaneous flame structure.

The scatter plots of temperature for flame D (Fig. 3.21) show a very good agreement of the calculations with the experimental data. Most of the points are located near the equilibrium solution at the three different locations. At $x/D = 15$, some amount of localized extinction can be seen on the experimental measurements, which is also captured by the simulation. The measurements also show a broad scattering of points for very rich mixtures, which cannot be observed in the calculations. This was also reported in [27], where the authors argued that this could be related to higher measurement uncertainties under fuel-rich conditions.

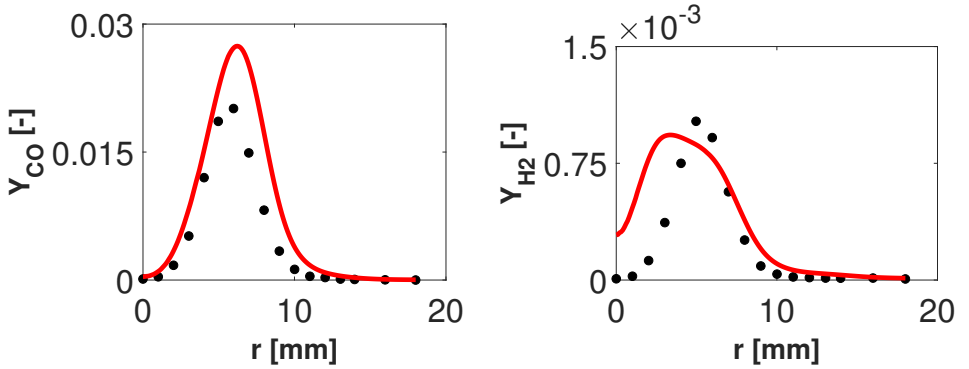
Scatter plots for flame E (Fig. 3.22) also compare well with the experiments. In this case, it can be seen that the flame experiences much higher levels of extinction and reignition compared to flame D. This is evident from the scatter plots at $x/D = 7.5$ (Fig. 3.22a) and at $x/D = 15$ (Fig. 3.22b). At these two locations, the simulation results are in very good agreement with the experiments. When moving further downstream (at $x/D = 30$), the flame is able to recover, resulting in a much narrow temperature distribution (Fig. 3.22c), which is also well captured by the model.

The scatter plots of temperature for flame F (Fig. 3.23) clearly show the broad scalar distribution around the stoichiometric condition for $x/D = 7.5$. This broad distribution is well captured by the simulation (Fig. 3.23a). At $x/D = 15$ and 30, the calculations are in good agreement with the measurements, although the region of high temperature fuel-rich mixtures is less populated at $x/D = 15$ compared to the experiments.

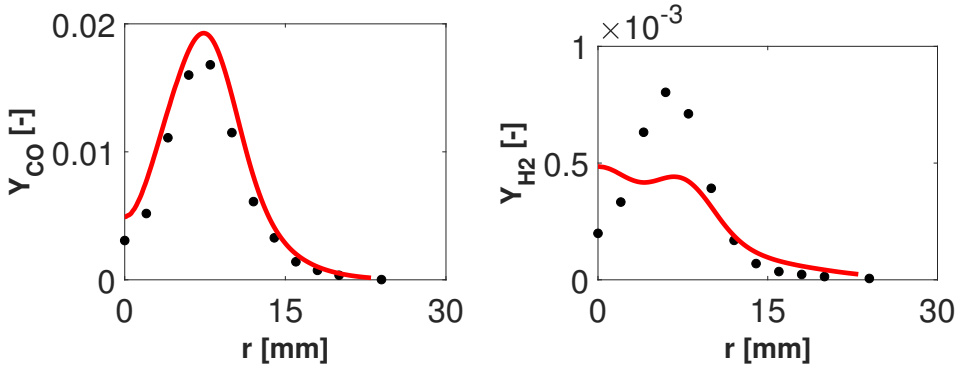
Scatter plots for the CO mass fraction are also shown on Fig. 3.24 for flame E and on



(a) centerline

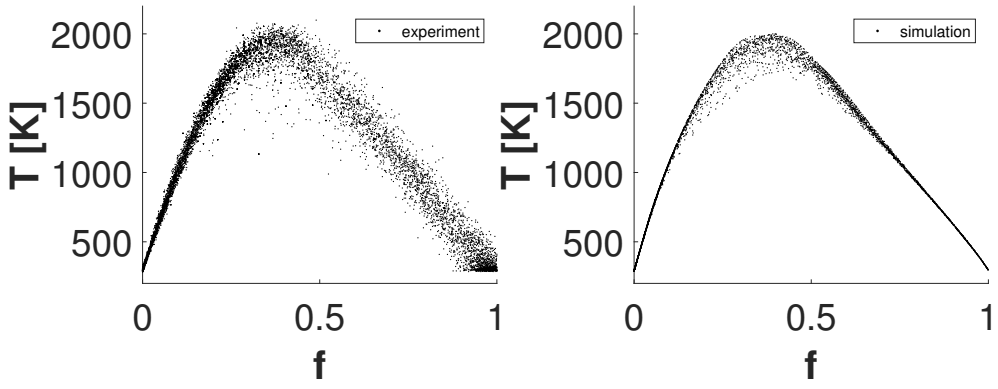


(b) $x/D = 7.5$

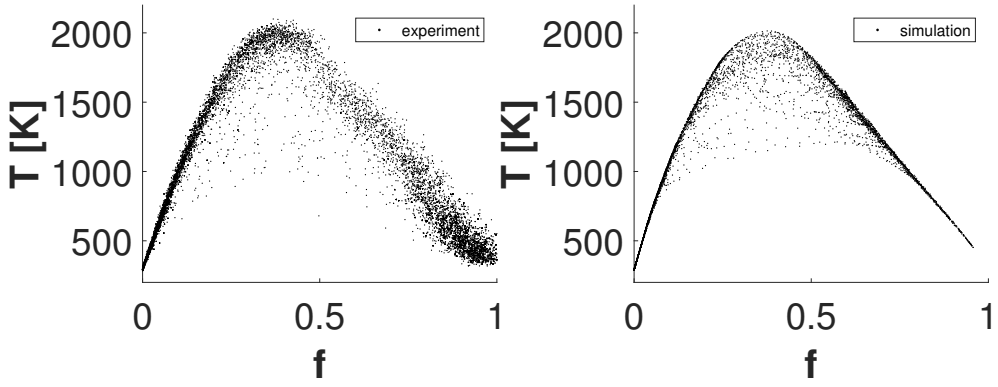


(c) $x/D = 15$

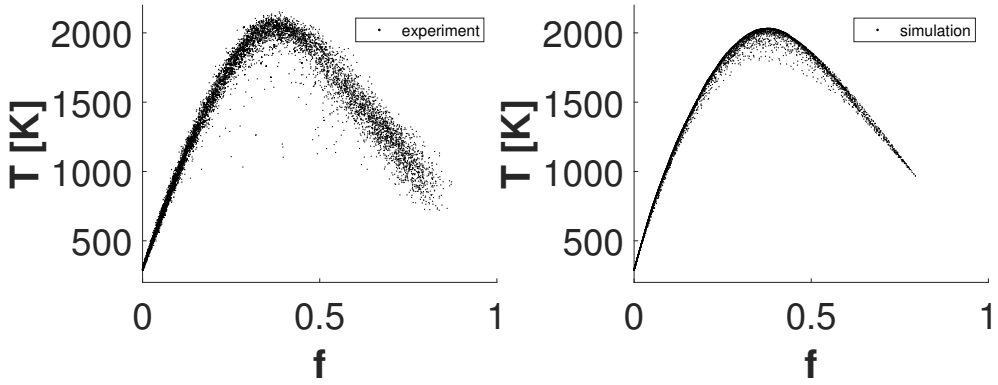
Figure 3.20: Flame F: CO and H_2 radial profiles plotted againt the experiments - (a) centerline, (b) $x/D = 7.5$ and (c) $x/D = 15$ (PC basis using the major species)



(a) $x/D = 7.5$

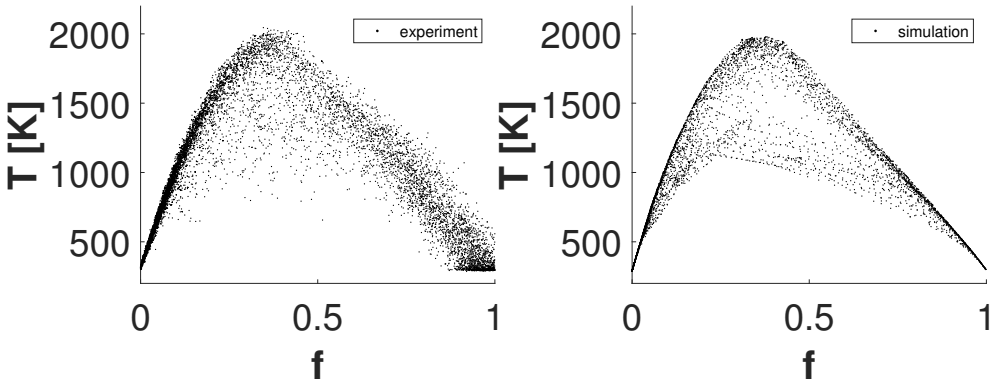


(b) $x/D = 15$

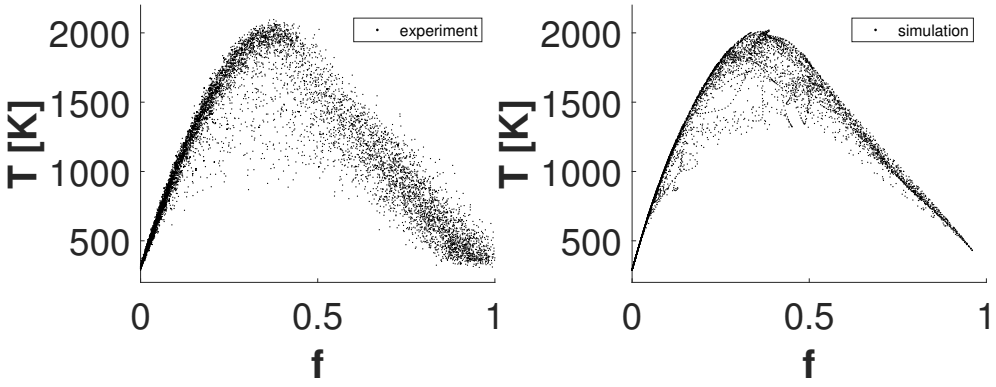


(c) $x/D = 30$

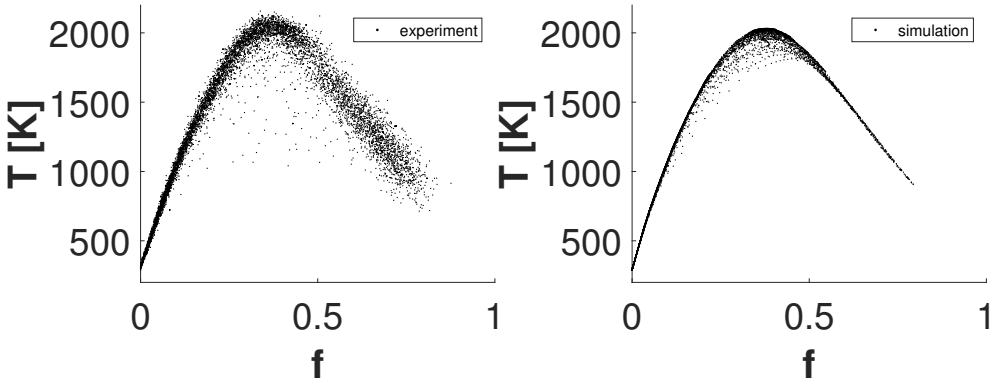
Figure 3.21: Flame D, scatterplots of the Favre-filtered temperature in function of mixture fraction: left, experimental data; right, numerical simulation



(a) $x/D = 7.5$

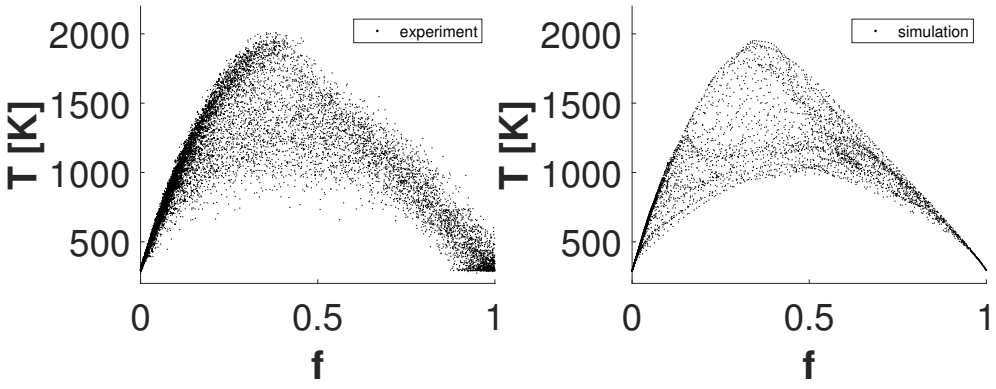


(b) $x/D = 15$

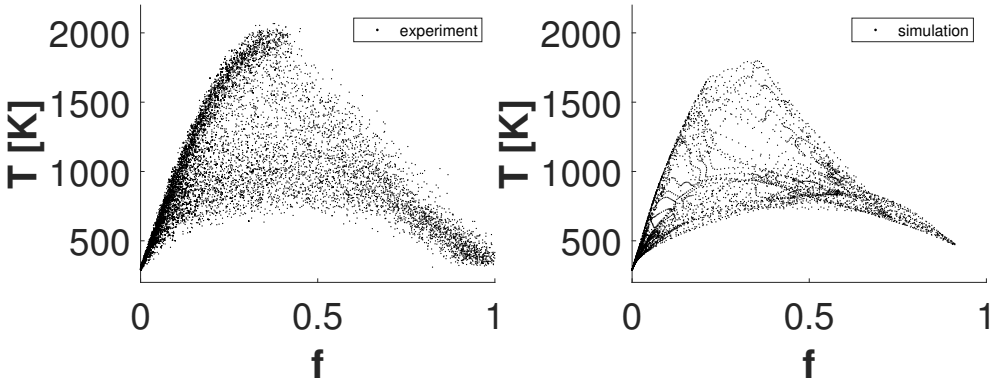


(c) $x/D = 30$

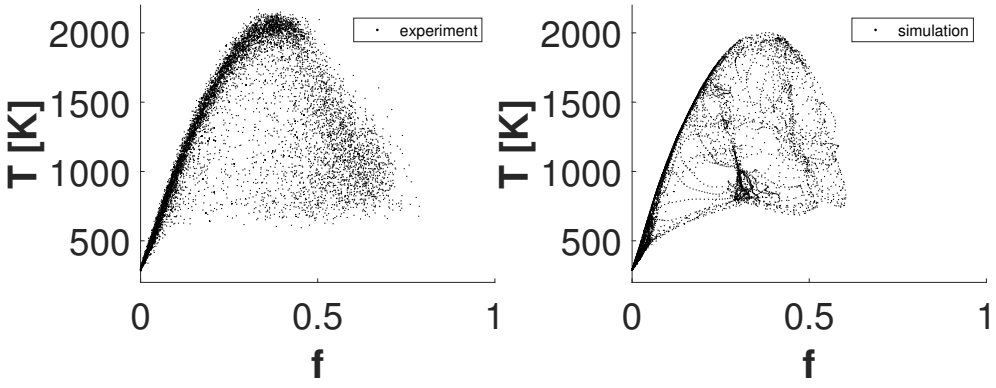
Figure 3.22: Flame E, scatterplots of the Favre-filtered temperature in function of mixture fraction: left, experimental data; right, numerical simulation



(a) $x/D = 7.5$



(b) $x/D = 15$



(c) $x/D = 30$

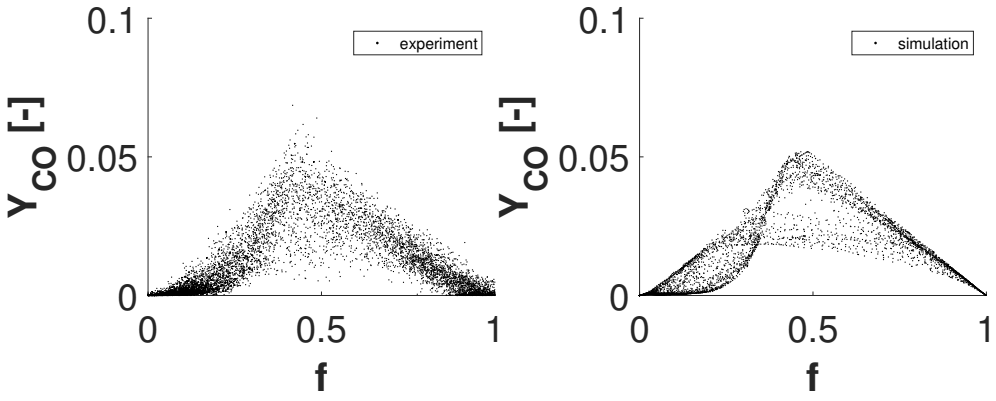
Figure 3.23: Flame F, scatterplots of the Favre-filtered temperature in function of mixture fraction: left, experimental data; right, numerical simulation

Fig. 3.25 for flame F. The broad distribution of CO mass fraction around the stoichiometric mixture fraction at $x/D = 7.5$ for flame E and at all three locations for flame F are qualitatively well predicted by the model. For flame E at $x/D = 30$, it can be observed on the experimental data that the width of the CO distribution has decreased around f_{st}, which indicates reignition and is also captured by the model. For flame F at $x/D = 15$ and 30, the model tends to predict lower values of Y_{CO} around f_{st}, showing nevertheless the broad scalar distribution characterizing local extinction.

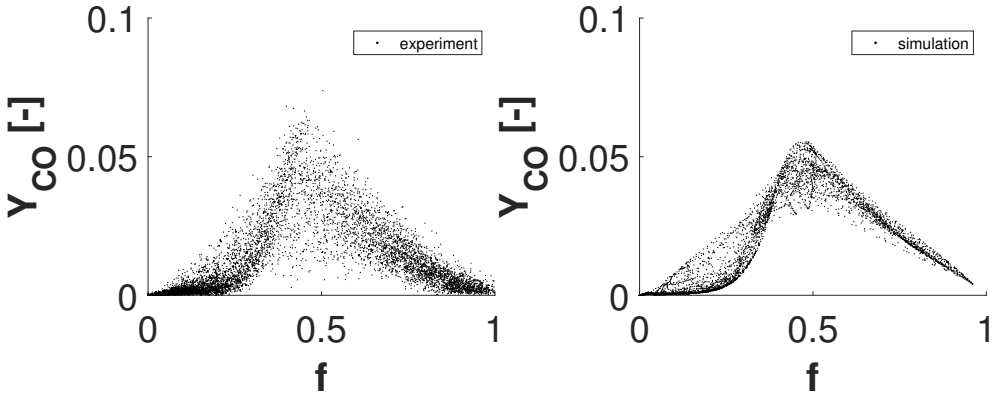
3.6.6 Comparison with the FPV model

In this Section, a discussion and a comparison of the proposed PC-GPR model with the existing and validated FPV model of Pierce and Moin [53] is proposed. As previously explained in Chapter 1, the FPV model was specifically developed for non-premixed combustion and adopts a tabulated chemistry appraoch, where the combustion process is parameterized using a functional manifold and a reduced number of controlling variables, namely the mixture fraction and a progess variable. In that sense, it shares many similarities with the PC-GPR model, except for the definition of the controlling variables (which are obtained after a PC analysis) and the use of nonlinear regression (which allows to reduce the size of the lookup table). Therefore, a comparison of the PC-GPR model with the FPV model is appropriate.

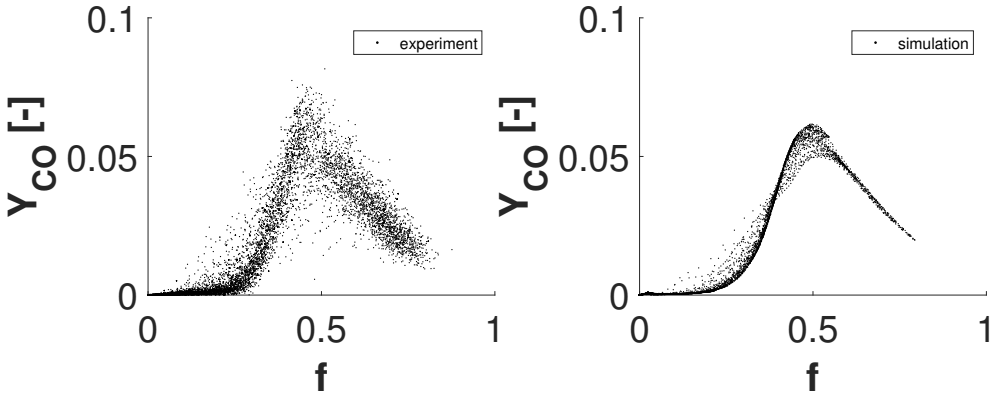
The study chosen here as a benchmark for the comparison is the study by Ihme and Pitsch [27], which shows very similar overall results as in the present work, and where the authors applied the FPV model in LES of Sandia flames D and E. They provided an extension to the classical FPV model, in which a presumed probability density function (PDF) of the beta-type is used for mixture fraction, and the PDF for the reactive scalar is modelled by a statistically most likely distribution (SMLD). This provides two advantages: first of all, the shape of the distribution depends on chemical and mixing time-scale information, and second, an arbitrary number of moments can be enforced [27]. The shape of this PDF was determined using the two moments of the progress variable and additional time scale information about chemistry and mixing. Therefore, the PDF approach required that additional transport equations be solved, namely for the residual scalar variance of mixture fraction and progress variable. The chemistry tabulation was based on the



(a) $x/D = 7.5$

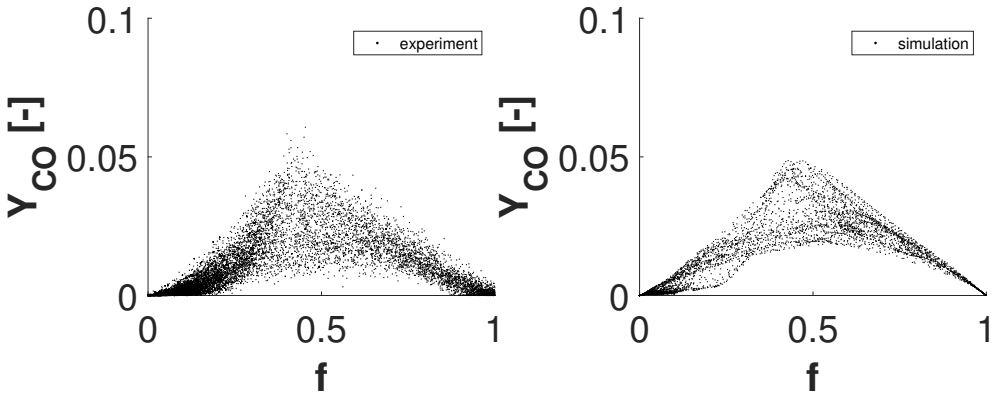


(b) $x/D = 15$

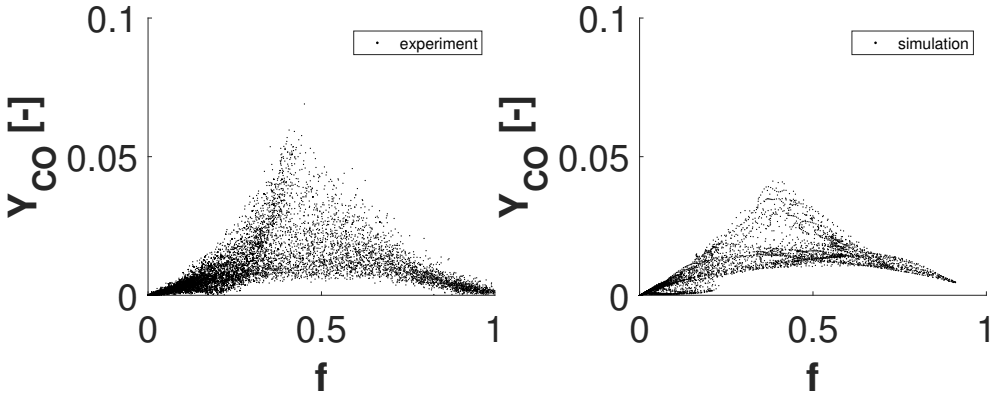


(c) $x/D = 30$

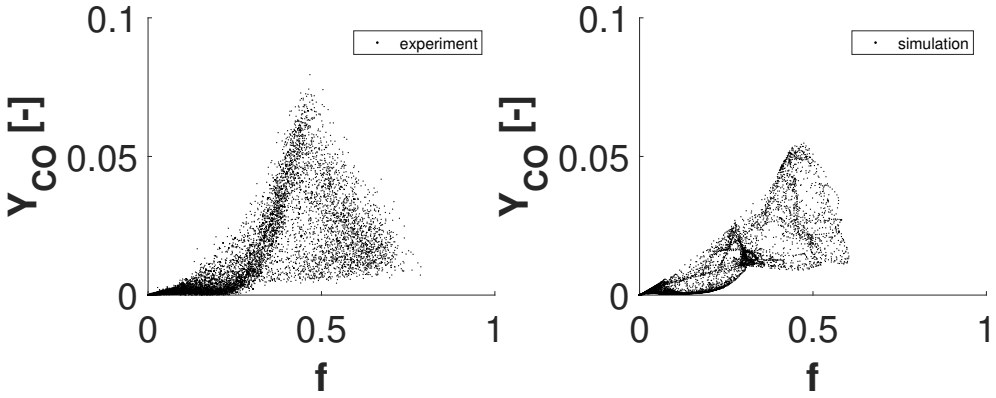
Figure 3.24: Flame E, scatterplots of the Favre-filtered CO mass fraction in function of mixture fraction: left, experimental data; right, numerical simulation



(a) $x/D = 7.5$



(b) $x/D = 15$



(c) $x/D = 30$

Figure 3.25: Flame F, scatterplots of the Favre-filtered CO mass fraction in function of mixture fraction: left, experimental data; right, numerical simulation

solution of the steady flamelet calculations and the GRI 2.11 mechanism [8], consisting of 279 reactions among 49 species, was used.

The comparison between the PC-GPR model and the FPV model can be done looking at several aspects. First of all, the FPV data set in [27] was based on steady laminar flamelets of the counterflow type, parameterized using the mixture fraction and a user-defined reaction progress variable, independent of mixture fraction, defined as $C = Y_{CO_2} + Y_{CO} + Y_{H_2O} + Y_{H_2}$. The authors argued that the classical steady laminar flamelet models, which usually parameterize the thermochemical quantities based on mixture fraction and the scalar dissipation rate of it, are unable to describe the transient states a flamelet experiences during the extinction and reignition processes [27]. According to the authors, the parametrization based on a reaction progress variable has the advantage that the table can also include flamelets with values of the progress variable that are lower than the value of the steady flamelet at the extinction scalar dissipation rate. In contrast, in the present work unsteady flamelets of the counterflow type were used, and the controlling variables were identified automatically by the PCA method. It turned out that the first PC, which contains most of the variance in the system, was highly correlated with mixture fraction. Moreover, the second PC was found to be very correlated with Y_{CO_2}. By doing so, PCA identified in fact the best progress variable for the given system, without the need for any user-defined definition. Another important observation is that the FPV model required the use of a PDF for the closure of the LES equations, whereas in the present work, the PC-GPR model was used without any subgrid closure, providing very similar overall results for the prediction of temperature, mixture fraction and species. It must be pointed out that the minimum LES filter width in the domain are very close for both studies (1.9×10^{-4} m here and 2.8×10^{-4} m in [27]). Therefore, it can be argued that the combination of the unsteady data set and the best controlling variables for the system defined by PCA provide an alternative to the use of steady flamelets parameterized by user-defined variables and combined with a PDF approach. The former method requires more work upfront compared to the latter one, but eventually saves CPU time and decreases memory cost during the simulation.

Results from the study of Ihme and Pitsch do share an interesting similarity with the results shown in this work: the overprediction of temperature in the lean region of the flame above the stoichiometric flame length. The results in [27] show an overprediction of

approximately 200 K, comparable to the level of overprediction obtained with the PC-GPR model. The authors stated that this overprediction is mainly due to the slight overprediction of the mixture fraction, which is caused by neglecting the interaction between turbulent mixing, chemistry, and radiative heat transfer [27]. It can be observed on Fig. 3.7a that the axial decay of mixture fraction is also underpredicted in this work in the lean region of the flame. To address this issue, the authors in [27] proposed to include thermal radiation effects in the lookup table, by introducing the enthalpy as an additional variable. Such an approach could also be applied in the context of the PC-GPR model, providing an interesting extension of the model to more complex systems where radiation plays an important role.

3.7 Conclusion

This chapter presented 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. Furthermore, 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 re-ignition 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 PC-GPR model was also compared with the FPV approach. The results of the

PC-GPR model are very similar to those of the FPV model. It was shown that the combination of the unsteady data set and the best controlling variables for the system defined by PCA provide an alternative to the use of steady flamelets parameterized by user-defined variables and combined with a PDF approach. Some limitations of the PC-GPR model are also shared by the FPV model, which can be addressed by extending the lookup table to a variable taking into account heat loss effects, such as the enthalpy.

CHAPTER 4

PRINCIPAL COMPONENT ANALYSIS BASED COMBUSTION MODEL IN THE CONTEXT OF A LIFTED METHANE/AIR FLAME: SENSITIVITY TO THE MANIFOLD PARAMETERS AND SUBGRID CLOSURE

The present Chapter extends the PC-transport approach, coupled with Gaussian Process Regression (GPR), to a lifted methane/air flame in Large Eddy Simulation (LES). Several key features of the model are investigated: the sensitivty to the training data set, the influence of the scaling methods, the issue of data sampling and the potential of a subgrid scale (SGS) closure.

4.1 Introduction

Recent developements in combustion systems are mainly driven by the need to increase fuel efficiency and reduce pollutant emissions. Among the novel stategies presented to tackle these issues, combustion of diluted fuels has gained attention due to its ability to reduce the combustion temperature by diluting the reaction zones, ensuring a decrease in the formation of thermal NO_x by avoiding high temperature peaks. Dilution, mainly achieved through the recirculation of hot combustion products, can also enhance flame stability. This recirculation of burned gases also introduces new challenges such as a reduction in the characteristic Damköhler number of the system. This phenomenon leads to an increased importance of the reaction kinetics in the combustion process, unlike conventional diffusion flames where the combustion process is primarily mixing-controlled. Thus, numerical prediction of turbulent diluted flames can represent a significant challenge for current combustion models, requiring them to be able to track a high number of variables while also adequately treating turbulence/chemistry interactions [28]. In

the context of Large Eddy Simulation (LES), this can become even more challenging as ignition mechanisms in such flames occur on scales that are computationally not resolved. Therefore, the thermo-chemical state must be parameterized using a reduced number of optimally chosen variables, coupled with subgrid scale closure models to describe the effects of unresolved scales and ignition kinetics.

The aim of the present chapter is to advance the understanding of the PC-GPR model in the context of LES simulations. The validation of the PC-GPR model is performed using the experimental burner described by Cabra et al. [10]. First, the influence of the canonical reactor used to generate the training dataset is investigated, comparing 1D counter-flow laminar flames (CFLF) and 0D perfectly-stirred reactor (PSR). Next, an analysis of the different scaling methods used in PCA is performed, showing the advantages and disadvantages of each of them. After that, the PC-GPR model is coupled with the kernel density method for the first time within a LES simulation in order to address the issue of data sampling. Fnally, a subgrid closure model is proposed for the PC-score approach within the context of LES simulations. To the authors' knowledge, the current work is the first attempt to use such an approach.

4.2 Experimental configuration and training data set

4.2.1 The vitiated coflow burner

The laboratory burner used to validate the PC-score model has been described by Cabra et al. [10]. It consists of a turbulent lifted jet flame within a hot environment. The central fuel jet is surrounded by a coflow of hot combustion products from a lean premixed flame. The coflow diameter is much larger than the central fuel jet diameter, allowing to isolate the fuel stream from the ambient air for a sufficiently long distance. The numerical treatment of such configuration can be thus treated as a two-stream flow. The central fuel pipe has a diameter of $D = 4.57$ mm, supplying a methane/air mixture (33% of CH_4 and 67% of air, by volume) at a temperature of 320 K. The vitiated coflow stream is obtained from the products of a hydrogen/air premixed combustion, and is mainly composed of H_2O and air at a temperature of 1350 K. The bulk velocites of the fuel jet and of the coflow are 100 m/s and 5.4 m/s, respectively. Details about the experimental conditions are shown in Table 4.1. A 2D schematic drawing of the Cabra lifted flame can be seen in Fig. 4.1.

Table 4.1: Conditions for the lifted methane–air jet flame in a vitiated coflow

	Jet	Coflow
Reynolds number	28,000	23,300
Diameter (mm)	4.57	201
Velocity (m/s)	100	5.4
T (K)	320	1350
Mach number	0.26	0.007
X_{O_2}	0.15	0.12
X_{N_2}	0.52	0.73
X_{H_2O}	0.0029	0.15
X_{CH_4}	0.33	0.0003
X_{H_2}	0.0001	0.0001
X_{OH}	–	0.0002
f_{st}	0.177	

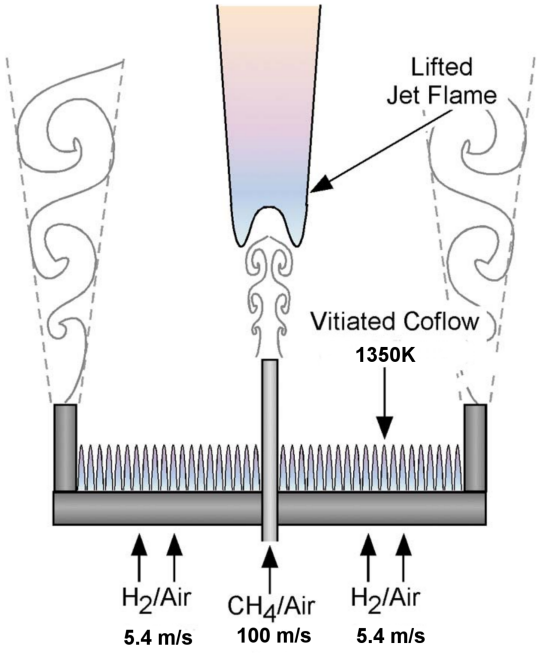


Figure 4.1: 2D schematic drawing of Cabra lifted flame (adapted from Cabra et al. [10]).

4.2.2 Reference data sets

PCA-based combustion models require high-fidelity data sets in order to generate the thermo-chemical state-space of interest, on which the PCA analysis is then applied. One of the advantages of PCA is that the training data set can be generated using simple configurations (such as canonical reactors), and then applied to more complex systems. Figure 4.2

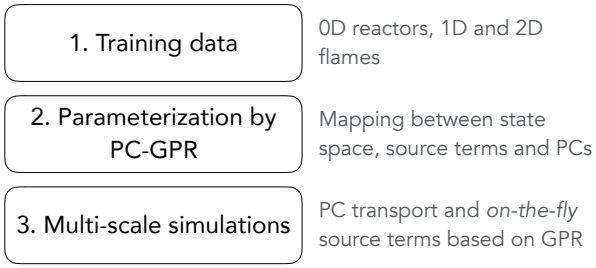


Figure 4.2: PC-GPR approach schematic

presents the idea behind the PC-GPR approach: using a detailed kinetic mechanism and a canonical reactor, the reference data set is generated. PCA is then applied on that data, and the state-space is regressed onto the PC basis. The model can then be applied in a CFD simulation.

In this study, two different types of canonical reactors are used and compared to each other: the *1D counter flow laminar flame* (CFLF) setup, and the *0D perfectly stirred reactor* (PSR). For both cases, unsteady simulations were performed in order to make sure that the generated state-space includes all the possible states accessed during the actual CFD simulation. The OpenSMOKE++ suite, developed in Politecnico di Milano [15, 16], was used to generate both data sets (CFLF and PSR), together with the GRI 3.0 kinetic mechanism [60], involving 35 species and 253 reactions (excluding NOx).

The first training data set was generated using an unsteady 1D laminar counter diffusion flame setup. The transient version of the CFLF solver was used in order to cover all possible states from equilibrium to extinction. The experimental conditions (Section 4.2.1) were used to set up the boundary conditions for the fuel and oxidizer sides. The counterflow flames were pulsated using a sinusoidal profile, allowing to vary the strain rate from equilibrium to extinction. A uniform grid of 400 points over a 0.15 m domain was used, with the final data set consisting of $\sim 90,000$ observations for each of the state-space variables.

The PSR data set was generated in a similar way: unsteady simulations were performed by varying the equivalence ratio inside the reactor while maintaining a sufficiently long residence time in order to reach convergence. The equivalence ratio was varied in order to cover all the possible mixtures from pure oxidizer to pure fuel inside the vessel. The

transient solutions were saved for all variables, reaching a total of $\sim 70,000$ observations for each of the state-space variables.

The PCA basis $\mathbf{A}$ is then generated using the approach described in Section 1.3 for both data sets. A subset of species is used for the $\mathbf{X}$ matrix, consisting of the major species present in the system, namely CH_4, CO, O_2, CO_2, H_2O and N_2. Using the major species has the advantage of removing certain scalars which may contribute to highly nonlinear source terms, without any loss in accuracy compared to the use of the whole set of species as shown in [18, 40]. If not stated, the PARETO scaling was used as the default scaling method throughout this work (except in Section 4.4.2 where the influence of different scaling methods is assessed), as it allows to obtain the greatest reduction for methane mechanisms, and produces at the same time an easily regressible surface, as shown in Chapter 3.

The nonlinear state-space variables (temperature, density, species mass fraction, PCs source terms and laminar viscosity) were regressed onto the linear PC basis using Gaussian Process regression (GPR). All variables were accurately regressed, with an $R^2 > 98\%$ for all source terms. Figure 4.3a shows the manifold for z_2's source term (s_{z_2}) as a function of z_1 and z_2, and Fig. 4.3b shows the regression of that manifold ($R^2 = 98.36\%$).

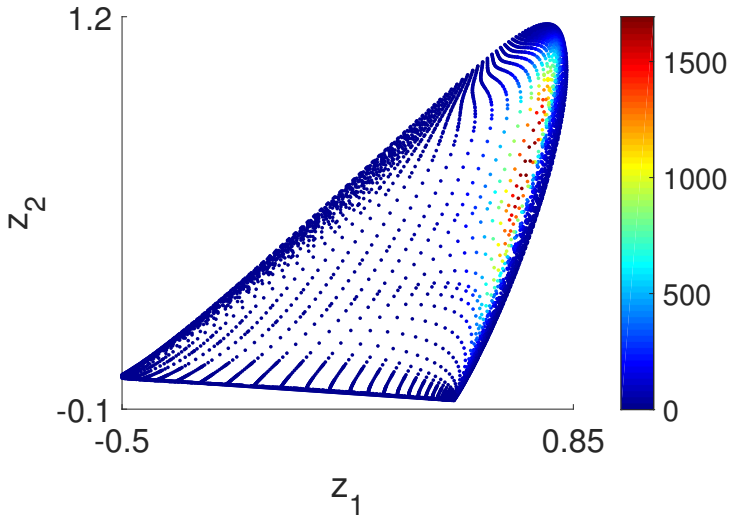
4.3 Computational setup

LES simaltions were performed in the open-source CFD software OpenFOAM using a tabulated chemistry approach. The variables of interest (i.e. the PCs) are transported, and the state-space (Y_k, T, ρ, S_z) is recovered from the non-linear regression. The low-Mach, three-dimensional, Favre-filtered governing equations for mass and momentum were solved on an unstructured grid, together with the filtered score transport equation:

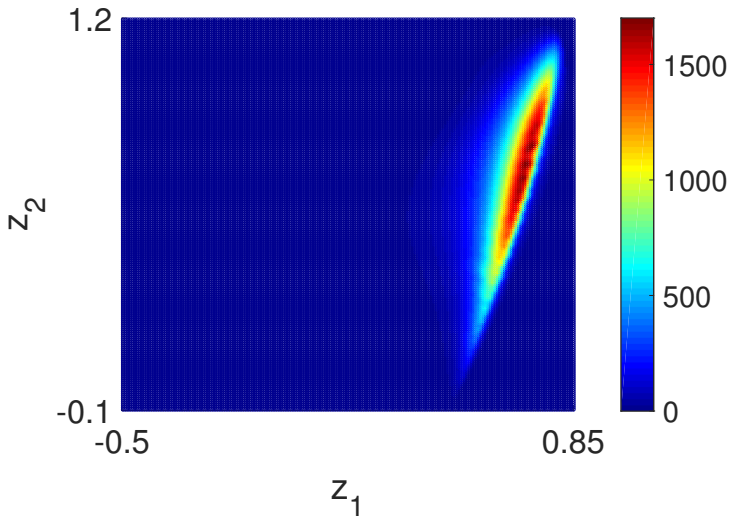
$$\frac{\partial \overline{\rho}\widetilde{\mathbf{z}}}{\partial t} + \nabla \cdot \overline{\rho}\widetilde{\mathbf{u}}\widetilde{\mathbf{z}} = \nabla \cdot (\overline{\rho}\mathbf{D}_\mathbf{z}\nabla\widetilde{\mathbf{z}}) + \nabla \cdot \overline{\tau}_\mathbf{z} + \overline{\mathbf{S}}_\mathbf{z} \tag{4.1}$$

where the SGS Reynolds stresses were modelled using the WALE model [44], and $\overline{\tau}_\mathbf{z} = \overline{\rho}\widetilde{\mathbf{u}}\widetilde{\mathbf{z}} - \overline{\rho\mathbf{u}\mathbf{z}}$ is the SGS turbulent flux, expressed using a gradient transport hypothesis with an SGS eddy viscosity obtained from the WALE model [44]. The final equation then reads:

$$\overline{\rho}\frac{D\widetilde{\mathbf{z}}}{Dt} = \nabla \cdot \left[\left(\frac{\mu}{Sc} + \frac{\mu_t}{Sc_t} \right) \nabla\widetilde{\mathbf{z}} \right] + \overline{\mathbf{S}}_\mathbf{z} \tag{4.2}$$



(a)



(b)

Figure 4.3: s_{z_2} in function of z_1 and z_2, the original manifold from PCA (a) and the regressed manifold with GPR (b) - pareto scaling

where μ is the laminar viscosity and is tabulated in function of the PCs, μ_t is the turbulent viscosity and is obtained through the WALE model. The laminar and turbulent Schmidt numbers (Sc and Sc_t respectively) were set to 0.7. Differential diffusion effects were reported to be negligible in this flame [10, 24], thus the unity Lewis number assumption was made. A backward scheme was used for the time derivative and the Gauss linear scheme, which has second order accuracy, was used for the divergence terms. The computational grid is conical, with a length of $100D$, a width of $45D$ at the inlet and $70D$ at the outlet, resulting in a total number of approximately 5.5 million hexahedra elements. An expansion ratio of 1.004 in the axial direction, and of 1.1 in the radial direction were used in order to have obtain finer cells near the inlets. The minimun element size in the domain is 1.6×10^{-4} m (near the nozzle exit). Turbulence was generated at the fuel and pilot inlets using the digital filter method by Klein et al. [35]. An injection pipe was used for the fuel inlet, extending $13D$ upstream the inlet. The regression of the state-space obtained with 2 PCs is accurate enough, therefore the simulations were carried out using only 2 PC transport equations (i.e. for z_1 and z_2). A look-up table was generated, having 200×200 points in $(\widetilde{z_1}, \widetilde{z_2})$ space, where all the state-space variables were tabulated. The boundary conditions for the PCs can be obtained using Eq. 1.5:

$$\mathbf{z}\,|_{\mathbf{boundary}} = \mathbf{X}\,|_{\mathbf{boundary}}\,\mathbf{A}$$

4.4 Canonical reactor and scaling method

The present section shows the results of the PC-score approach coupled with GPR (PC-GPR model) on the Cabra flame. The comparison between the data set generated using the *1D counter flow laminar flame* (CFLF) setup and the *0D perfectly stirred reactor* (PSR) is first presented. Then, the influence of the scaling method used to create the PCA basis is analyzed. All simulations were run for at least 10 flow through periods, in order to have a sufficiently large averaging window. Radial and axial profiles were averaged in both time and space using the axi-symmetric character of the flow and compared to experimental measurements.

4.4.1 LES results - CFLF vs PSR

Figure 4.4 shows the centerline profiles Favre-averaged temperature and its fluctuation, mixture fraction and its fluctuation, and mass fractions of O_2 and OH. The experimental results are compared with the numercial results from the CFLF and the PSR trained PCA models, both using pareto scaling. It can be observed that both data sets predicts an early ignition. The experimental results show a two stage flow: mixing without reaction taking place between the fuel jet and the vitiated coflow up to $x/D \sim 40$ and characterized by a slow temperature rise (primarily controlled by the scalar mixing process), followed by a flame stabilization region characterized by a rapid temperature rise, larger temperature fluctuations, and the rise of Y_{OH}.

The numerical results on the other hand show a significantly faster ignition process, with an ignition taking place at $x/D \sim 10$ for both data sets, a sharp increase in temperature fluctuations and a rapid consumption of O2 due to combustion. The mixture fraction decay in the near field is well predicted by both models, while deviating from the experiments in the flame stabilization region. It can also be seen that combustion occurs faster with the CFLF model, also depicting higher levels of temperature fluctuations and a more rapid O_2 consumption. The tempearture reaches its maximal value at $x/D \sim 40$ for the CFLF data set, while in the PSR case, the temperature shows a much slower increase, with the peak value reached at a location further downstream ($x/D \sim 80$). Thus, the slope of increase of the temperature in the transient region is better predicted by the CFLF data set. Moreover, the fully developed flame zone is also better captured by the CFLF data set ($x/D > 45$), whereas in the PSR simulation, the temperature keeps increasing until the outlet of the domain ($x/D = 90$).

Nevertheless, it is fair to say that with the standard PC-GPR approach both data sets do reproduce some of the qualitative characteristics of the experimental centerline profiles, while failing to reproduce some other, such as an under-prediction of the lift-off height or the early autoignition. This can be attributed to the absence of coupling between the turbulence of the flow field on one side, and the chemistry repesented by the PCA manifold on the other. Therefore, one can act either on the construction of the PCA basis (see Section 4.4.2), on the raw data set used to buid the PCA basis (Section 4.5), or on the treatment of turbulence/chemistry interactions (Section 4.6).

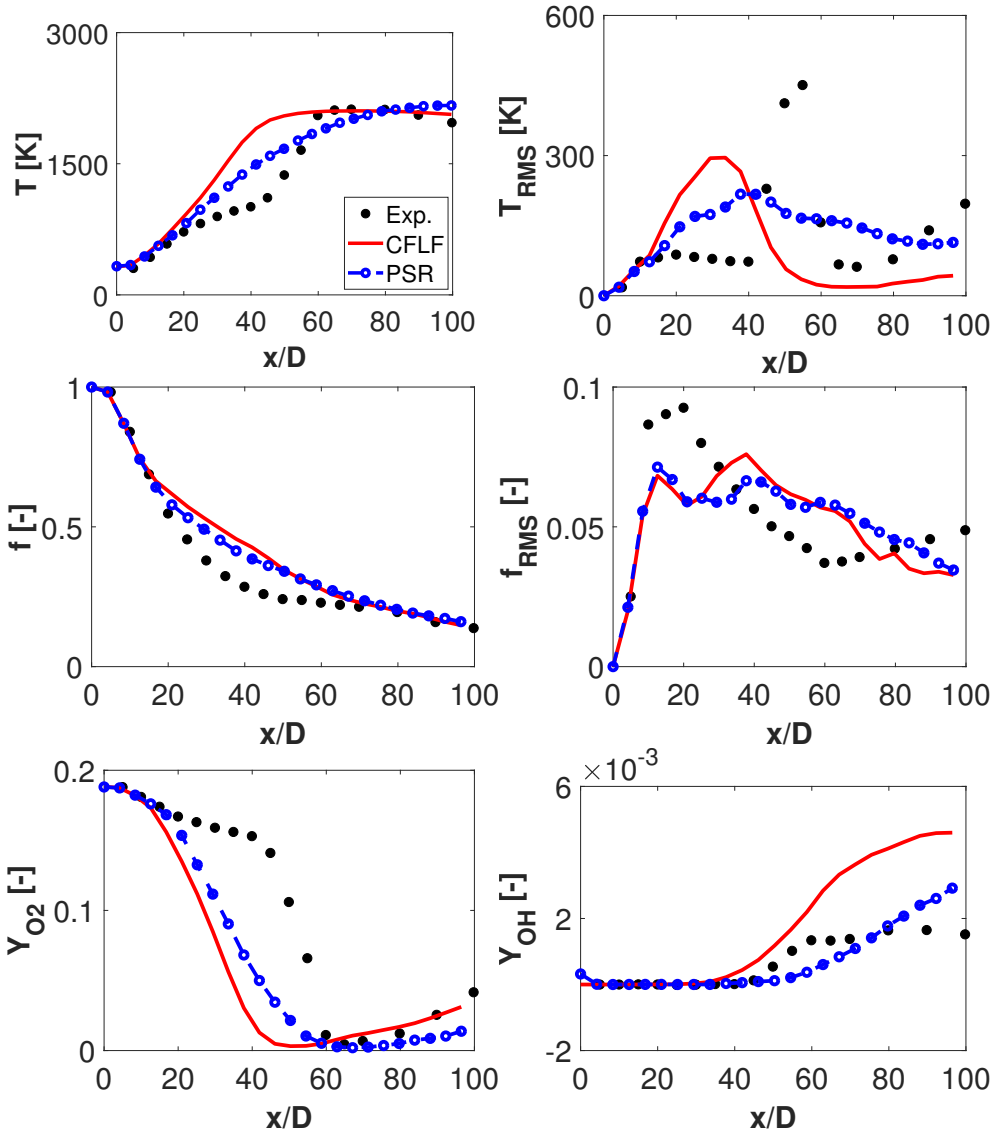


Figure 4.4: Comparison of centerline profiles of Favre-averaged mean and rms statistics of temperature (T), mixture fraction (f), O_2 and OH mass fractions. Numerical results (lines): counter flow laminar flame data set (CFLF), perfectly stirred reactor data set (PSR). Experimental results (symbols).

4.4.2 LES results - Influence of the PCA scaling method

It was mentionned in Section 1.3 that different scaling methods can be used during the construction of the PCA basis. The present Section will investigate the impact of various scaling methods found in the litterature on the numerical results of the Cabra flame. Before presenting the methods, it is useful to rewrite Eq. 1.4 in a scalar form for the sake of clarity:

$$x_{SC,j} = \frac{x_j - \overline{x}_j}{d_j} \; for \; j = 1, ..., Q$$

where $\overline{x}_j$ and d_j are the centering and scaling coefficients, respectively. The following scaling methods were tested in the present work [47]:

- *standard scaling*, which adopts the standard deviation as the scaling factor, $d_j = s_j$;
- *range scaling*, which adopts the dierence between the minimum and maximum variable value as the scaling factor, $d_j = \max(x_j) - \min(x_j)$;
- *pareto scaling*, which adopts the square root of the standard deviation as the scaling factor, $d_j = \sqrt{s_j}$;
- *vast (variable stability) scaling*, which adopts the product between the standard deviation and the coffecient of variation ($s_j/\overline{x}_j$) as the scaling factor, $d_j = s_j^2/\overline{x}_j$.

Figure 4.5 shows the manifold for z_2's source term (s_{z_2}) as a function of z_1 and z_2 for the various scaling methods presented above. It can be observed that the shapes of all manifolds are very similar, except for the standard scaling (Fig. 4.5a) which exhibits a more elongated manifold. Moreover, the order of magnitude of the source term also provides an insight on the accuracy of the subsequent regression. Pareto scaling provides the smallest maximum value for the source term, vast scaling leads to the highest variation, while standard and range scalings lie in between. Therefore, as the high source terms points are located in a very narrow region of the manifold, pareto scaling will provide the easiest surface to regress for the nonlinear regression method (GPR in this context) as the slope of the surface will be less steep in that region. This confirms the findings previously reported in Chapters 2 and 3.

The LES results obtained using the four scaling methods presented above are shown on Fig. 4.6 for the CFLF data set. Globally, all the scaling methods show similar trends,

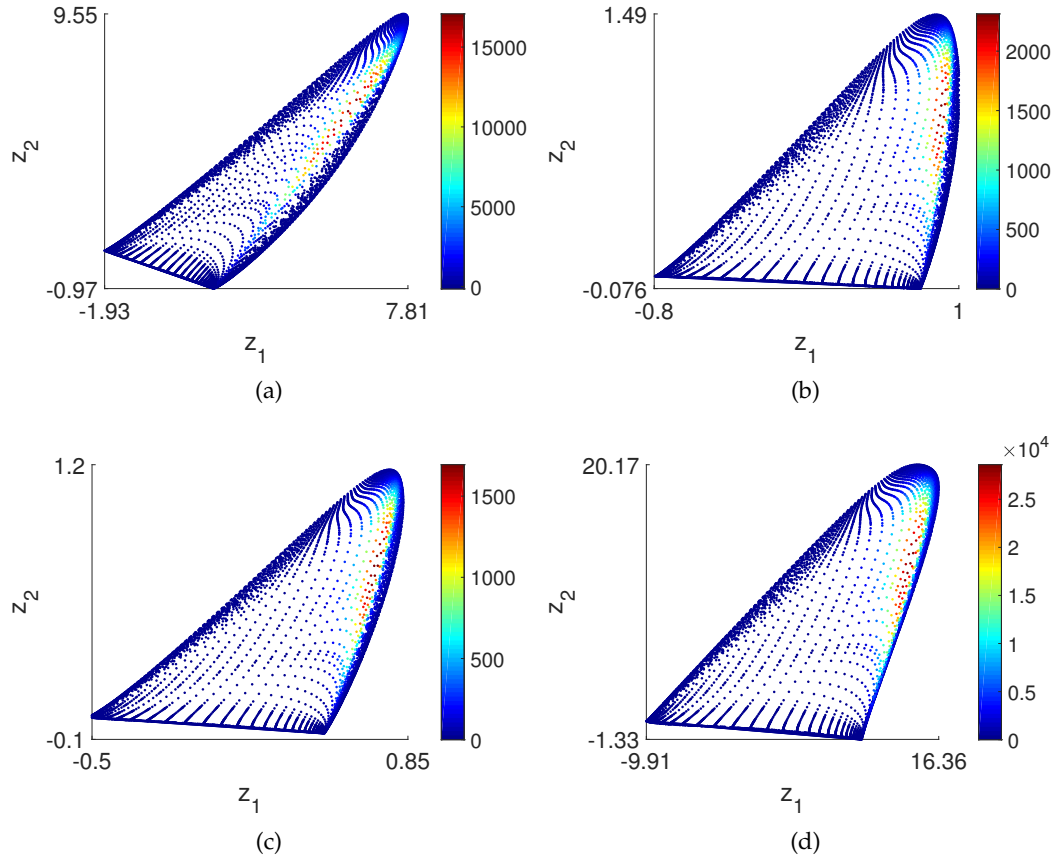
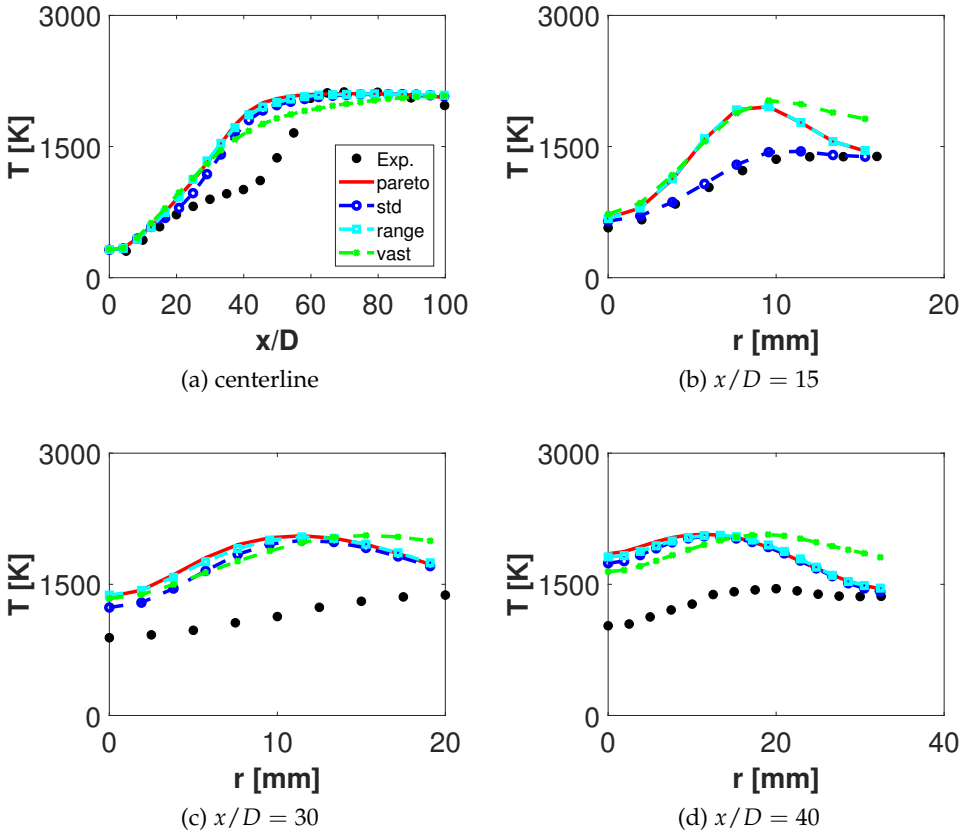


Figure 4.5: s_{z_2} in function of z_1 and z_2; scaling: (a) standard, (b) range, (c) pareto and (d) vast

namely an ignition starting at $x/D \sim 10$ and an under-estimation of the lift-off height, except for the standard scaling where ignition happens slightly after $x/D \sim 15$. At $x/D = 30$ and $x/D = 40$ (Fig. 4.6c and d), the numerical results show an already burning flame, while the experimental values suggest otherwise. Therefore, the choice of the scaling method does not influence the ability of the PC-GPR model to predict the correct lift-off height of the flame, the latter being under-estimated in all four cases. Therefore, the choice of the scaling method will be driven mainly by other parameters, such as the smoothness of the manifold and the easiness in the regression of it.

4.5 Kernel density weighted PCA

When applying PCA on a raw combustion data set, one may need to address the issue of data sampling. Indeed, numerical or experimental data sets usually show non-



(a) centerline (b) $x/D = 15$ (c) $x/D = 30$ (d) $x/D = 40$

Figure 4.6: Comparison of Favre-averaged mean temperature (T) profiles on the centerline (a) and radial profiles at various axial locations (b-d) using pareto, standard (std), range and vast scaling methods - CFLF data set

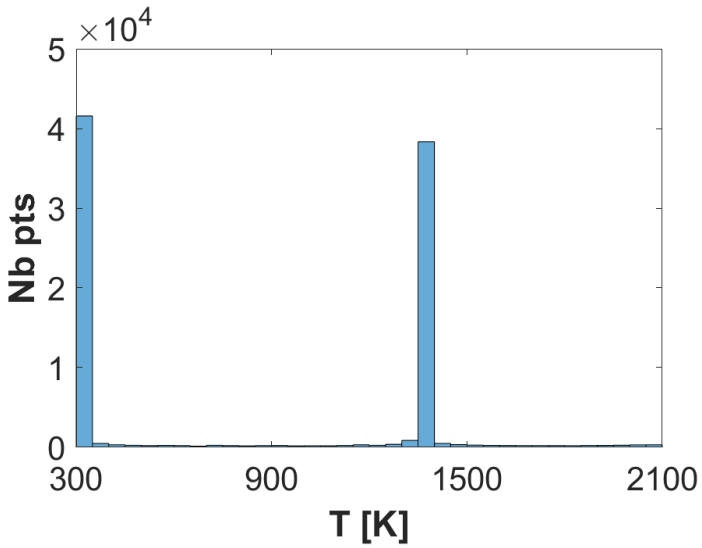


Figure 4.7: Density of the observed temperature for the CFLF data set

homogenous data density, hot and cold zones being generally over-represented. This can introduce bias in the PCA reconstruction, leading to a poor reconstruction of the flame region, usually under-represented in highly nonlinear multi-dimensional data sets, such as in combustion systems. This over-representation can be seen on Fig. 4.7, where the density of the observed temperature is shown for the counter-flow flames data set (see Section 4.2.2). It can be seen that most of the data points are located in the fuel ($T = 320$ K) and coflow ($T = 1350$ K) regions. Therefore, the hot and cold zones will be accurately represented by PCA, but not the flame front zone which is the main zone of interest. To tackle this problem, a combination of PCA with the kernel density method is used here. The kernel density method was introduced by Rosenblatt [58] and Parzen [49], and sucessfully applied by Coussement et al. [12] in the context of combustion. The general idea of this method is to apply a pre-treatment to the data set, by computing the density of the statistical sample in a distribution using a presumed normal distribution. For each observation in the data-set, the distance between the current observation c and an observation c' is computed:

$$d_{c,c'} = |x_{c'} - x_c|$$

with x_c the value of the variable at observation c. This allows to define a Gaussian kernel distribution $K_{c,c'}$ for each point c as:

$$K_{c,c'} = \sqrt{\frac{1}{2\pi h^2}} \exp\left(-\frac{d_{c,c'}^2}{2h^2}\right)$$

where h is the bandwidth of the weighting variable. Summing the Gaussian kernels provides with an estimation of the density at each point c:

$$\mathbf{K}_c = \sum_{c'=1}^{n} \frac{1}{n} K_{c,c'}$$

A normalized weighting can then be defined, for each observation, using the variable density:

$$\mathbf{W}_c = \frac{\frac{1}{\mathbf{K}_c}}{\max(\frac{1}{\mathbf{K}_c})}$$

The bandwidth, h, is computed using [12]:

$$h = \left(\frac{4\sigma}{3n}\right)^{\frac{1}{5}} \tag{4.3}$$

where σ is the standard deviation of the considered weighting variable within the data-set.

Coussement et al. [12] showed in their study that the combination of classical PCA with the kernel density method enhances the PCA accuracy, especially in the flame front zone. The approach was tested on a 1-D laminar flame, on a 2-D flame vortex interaction data-set and on a 3-D turbulent diffusion flame data set. They also showed that the mono-variable kernel density provides a better preconditioning of the data set and at a lower computational cost than the multi-variable method. Moreover, using the temperature to compute the kernel density was shown to be the most relevant choice, as this variable describes the progress of the combustion process efficiently.

4.5.1 Application

In the present work, as the PCA data set consists only of the species mass fractions, the conditioning variable chosen for the kernel density was CO_2. The bandwidth h was computed using Eq. 4.3 and the mono-variable kernel density was employed. Before applying the PCA, each centered and scaled variable is multiplied by the associated weighting:

$$\widehat{\mathbf{X}} = \mathbf{W}\,\mathbf{X}_{\mathbf{SC}}$$

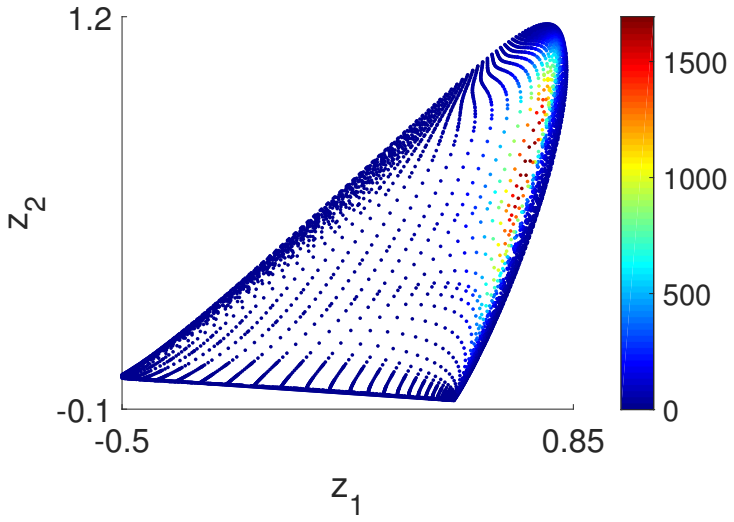
Table 4.2: R^2 statistics for the reconstruction of CFLF data set using 3 PCs, using classical PCA and kernel PCA (CFLF data set).

Variable	R^2 classical PCA	R^2 kernel PCA
Y_{CH_4}	0.968	0.9722
Y_{N_2}	0.9723	0.9966
Y_{H_2O}	1.0	0.9946
Y_{CO}	0.9982	1.0
Y_{CO_2}	0.9869	1.0
Y_{O_2}	0.9961	0.993

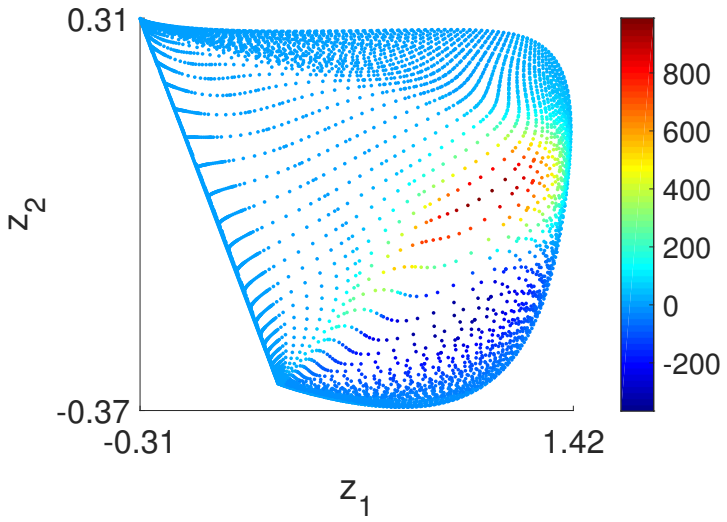
where $\mathbf{W}$ is the matrix containing the weighting for each observation. The weightings are computed using the centered and scaled data set, $\mathbf{X_{SC}}$, and that $\widehat{\mathbf{X}}$ is only used for the computation of the covariance matrix.

Table 4.2 shows the R^2 statistics for the species reconstruction, comparing the classical PCA reconstruction with the reconstruction obtained after applying the kernel density method for the CFLF data set. It should be pointed out that the R^2 statistic is strongly influenced by the over-represented regions (i.e. the hot and cold zones). Hence, the statistics shown here were computed only in the flame front zone (where the data density is lower), leaving out the hot and cold zone points in order to provide a much more faircomparison between the two methods. It can be observed that the R^2 statistics for all the species present in the PCA basis are improved after applying the kernel density, except for H_2O and O_2 where the R^2 shows a slight decrease.

The performances of the kernel density weighted PCA will now be shown on the simulation results of the Cabra flame, and compared with the results from classical PCA. The mono-variable kernel PCA method was applied to the CFLF data set, using CO_2 as the conditioning variable, as the latter is highly correlated with temperature. A sensitivity analysis was also performed on the choice of the conditioning variable (Section 4.5.2.2). All the state-space variables were regressed using GPR on the new manifold. Figure 4.8a shows the manifold for z_2's source term (s_{z_2}) as a function of z_1 and z_2 using classical PCA, and Fig. 4.8b shows the same manifold after applying the kernel PCA using CO_2. The numerical setup was the same as the one described in Section 4.3.



(a)



(b)

Figure 4.8: s_{z_2} in function of z_1 and z_2, the manifold from classical PCA (a) and the the manifold from kernel PCA (b) - pareto scaling

4.5.2 LES results - Kernel density weighted PCA

Results of the PC-GPR model coupled with the kernel density method are shown in this section. A comparison is made between the classical PC-GPR model and the kernel PC-GPR model. All simulations were run for at least 10 flow through periods, and radial and axial profiles were averaged in both time and space using the axi-symmetric character of the flow and compared to experimental data.

4.5.2.1 Instantaneous and mean flow field structure

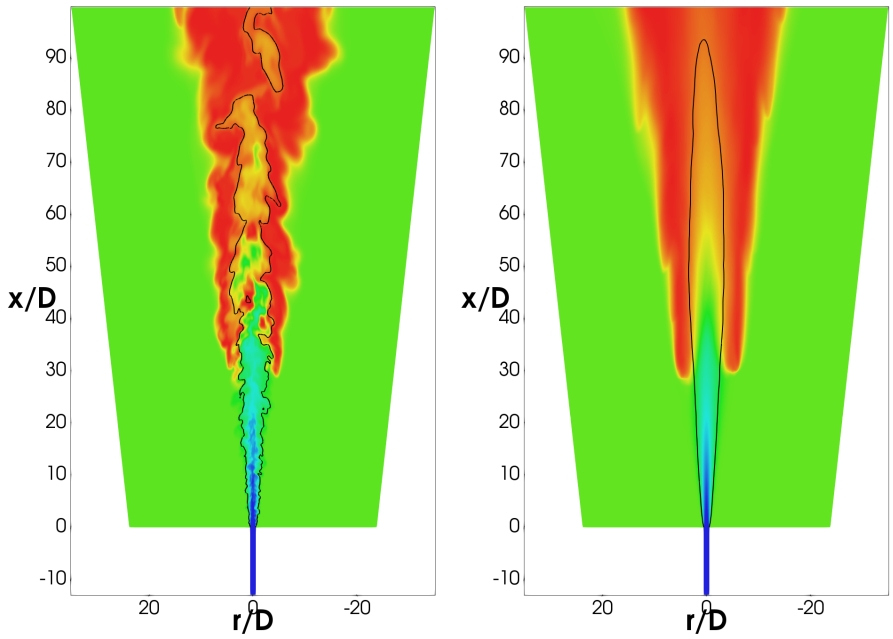
Instantaneous and averaged temperature fields obtained from both simulations (classical and kernel PCA) are illustrated on Fig. 4.9.The solid line in these figures corresponds to the isocontour of the stoichiometric mixture fraction ($f_{st} = 0.177$). Looking at the instantaneous temperature field obtained from the kernel PCA simulation (Fig. 4.9a), it can be observed that fuel and oxidizer mix together without significant heat release up to $\sim 30D$ above the jet exit. This inert mixing zone is then followed by a transient region (located approximately between $30 \leq x/D \leq 70$) in which the temperature increases due to autoignition. Beyond $x/D = 70$, the flame is continiously burning.

The classical PCA results (Fig. 4.9b) on the other hand show a very different flame behaviour. The inert mixing zone is significantly reduced and goes only up to $10D$ downstream of the jet nozzle. Beyond this location, the flame ignites and a very short transition region can be observed.

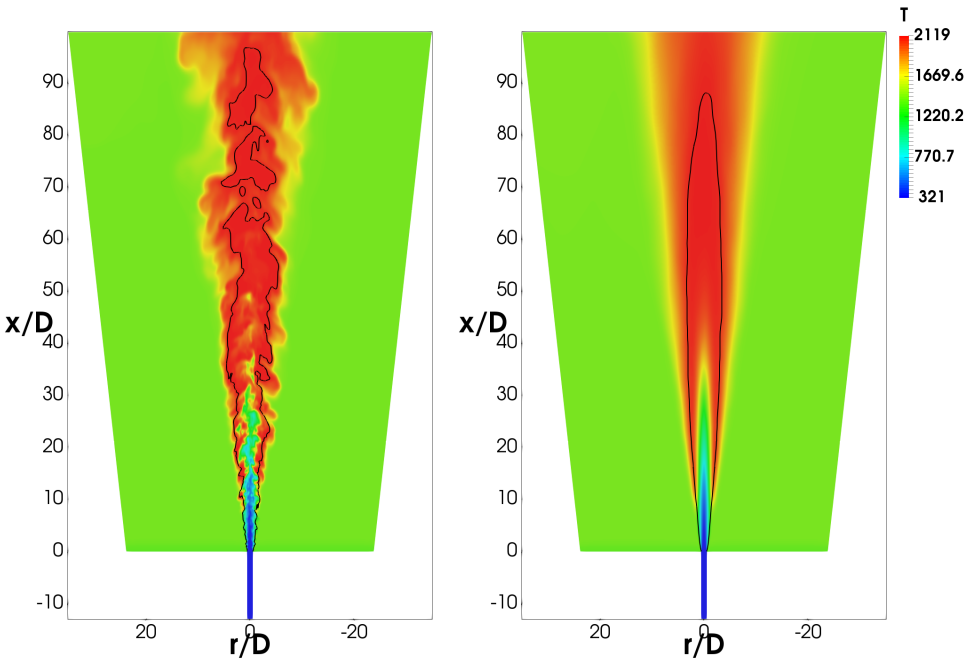
Another important observation that can be made when comparing both models is that the classical PCA results exhibit a smaller radial spreading rate of the flame compared to the kernel PCA model. This can in fact be linked to the shape of the manifolds: as shown on Fig. 4.8a, the manifold from classical PCA is much narrow on the lean part of the flame (large values of z_1), whereas for the kernel PCA manifold (Fig. 4.8b), that zone is much more widespread (low values of z_2). Therefore, the classical PCA simulation is confined to a narrow region on the manifold, resulting in a smaller radial spreding of the flame.

4.5.2.2 Statistical flow field results

Figure 4.10 shows a comparison of Favre-averaged results for temperature and mixture fraction along the jet centerline. Mean mixture fraction profile from the kernel PCA model is in very good agreement with the experimental results; the classical PCA over-predicts



(a) kernel PCA



(b) classical PCA

Figure 4.9: Instantaneous and averaged temperature fields obtained from (a) kernel PCA and (b) classical PCA model. The solid line shows the location of the stoichiometric mixture fraction, $f_{st} = 0.177$.

f in the region $15 \leq x/D \leq 65$. Mean and rms temperture profiles are shown on the bottom row of Fig. 4.10. The kernel PCA model captures well the first zone ($x/D \leq 45$) where the increase in centerline temperature is primarily controlled by the scalar mixing process. In the second zone ($x/D \geq 45$), where heat release effects become significant, the kernel PCA model deviates from the reported experimental results, showing an advanced ignition location and a lower temperature rise over the course of the autoignition process, and also a lower final temperature. This should be compared with the classical PCA model, which shows a significantly faster ignition and a flame reaching a steady-state condition at a distance of $x/D = 45$. Looking at the rms temperature fluctuations (top right of Fig. 4.10), it can be observed that the location of the peak is better captured by the kernel PCA model compared to classical PCA, but the peak value is under-predicted by both models by approximately 200 K.

A sensitivity analysis was also performed on the choice of the conditioning variable. Besides CO_2, different other species were used which have an importance describing the autoignition phenomena, namely OH, CH_2O and HO_2. A kernel weighting was performed using each one of them in the mono-variable method, which led to a different manifold for each of them. GPR regression was performed, and look-up tables were generated for each variable. Figure 4.11 shows the centerline temperature profiles from the simulations, comparing the kernels based on CO_2, OH, CH_2O and HO_2. It can be observed that the CO_2 kernel provides the best estimation of the lift-off height, where all the minor species based kernels provide an early ignition. This confirms the findings presented by Coussement et al. [12], that the kernel should be conditioned on a variable that describes efficiently the global progress of the combustion.

4.6 Subgrid scale closure

In turbulent flames, the accurate characterization of autoignition requires the consideration of turbulence/chemistry interactions and transient effects. While the latter are taken into account by the use of unsteady canonical reactors, the interactions between the flame structure and the turbulent environment are considered using a statistical description. Therefore, the unsteady data is integrated using presumed $\beta-$PDF functions of z_1 and z_2 to account for the trubulent fluctuations. The PC's being uncorrelated from each other

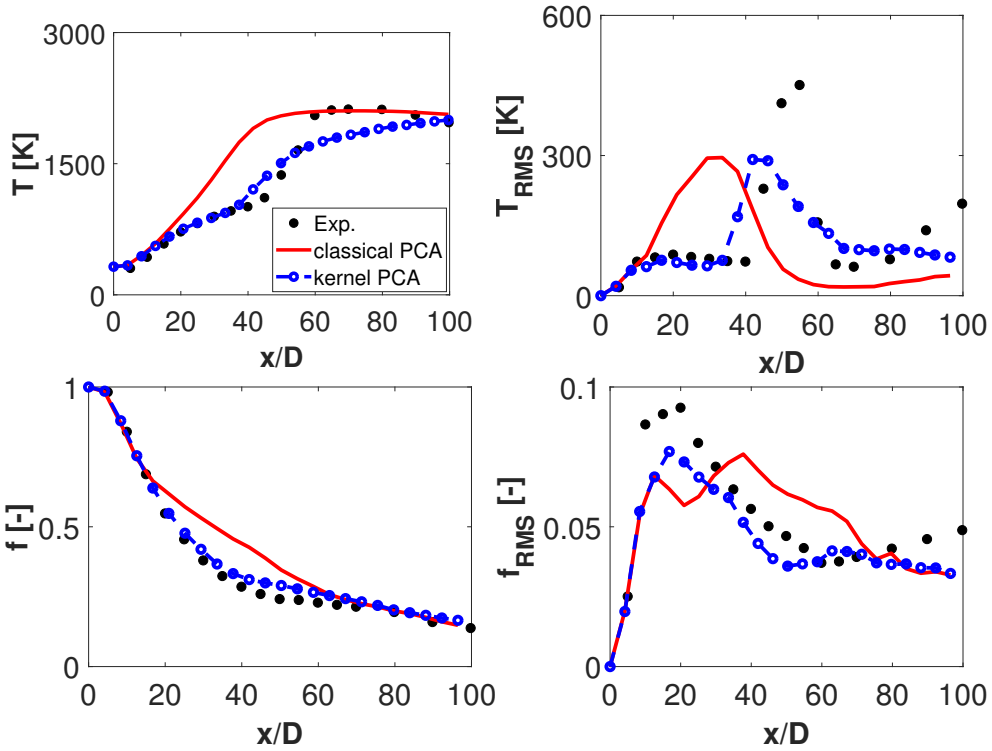


Figure 4.10: Comparison of centerline profiles of Favre-averaged mean and rms statistics of temperature (T) and mixture fraction (f). Numerical results (lines): classical PCA, kernel PCA based on CO_2. Experimental results (symbols).

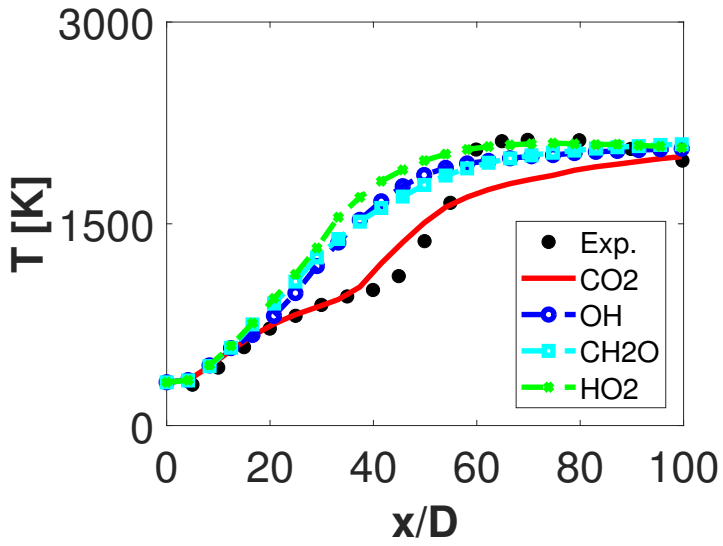


Figure 4.11: Comparison of centerline profile of Favre-averaged mean of temperature (T). Numerical results (lines): kernel PCA based on CO_2, OH, CH_2O and HO_2. Experimental results (symbols).

by definition, the filtered scalars and source terms are obtained from:

$$\widetilde{\phi} = \iint \phi(z_1, z_2)\, \widetilde{\beta}(z_1; \widetilde{z_1}, \widetilde{z_1''^2})\, \widetilde{\beta}(z_2; \widetilde{z_2}, \widetilde{z_2''^2})\, dz_1 dz_2 \tag{4.4}$$

where $\widetilde{z_1}$ and $\widetilde{z_1''^2}$ are the Favre-filtered mean and variance (respectively) of z_1 (the same reasoning applies to z_2). Two additional transport equations (for $\widetilde{z_1''^2}$ and $\widetilde{z_2''^2}$) are therefore required in order to close the system.

Following the procedure descibed by Domingo et al. [17], a transport equation is solved for $\widetilde{z^2}$; $\widetilde{z''^2}$ is then recovered noticing that $\widetilde{z''^2} = \widetilde{z^2} - \widetilde{z}\widetilde{z}$ (the indices have been omitted for the sake of notation simplicity). The equation for $\widetilde{z^2}$ reads:

$$\frac{\partial \overline{\rho}\widetilde{z^2}}{\partial t} + \nabla \cdot \overline{\rho}\widetilde{\mathbf{u}}\widetilde{z^2} = \nabla \cdot (\overline{\rho} D \nabla \widetilde{z^2}) + \nabla \cdot \overline{\tau}_{z^2} - 2\overline{\rho\chi}_z + 2\overline{zS}_z \tag{4.5}$$

where the SGS turbulent flux $\overline{\tau}_{z^2} = \overline{\rho}\widetilde{\mathbf{u}}\widetilde{z^2} - \overline{\rho\mathbf{u}z}$ is expressed using a gradient transport hypothesis with an SGS eddy viscosity obtained from the WALE model. The last term of the RHS of equation 4.5 (i.e. the filtered source term) is expressed from relation 4.4.

The SGS scalar dissipation rate of the PC $\overline{\rho\chi}_z$ deserves a careful treatment [17]. It can be decomposed into resolved and SGS parts:

$$\overline{\rho}\widetilde{\chi}_z = \overline{\rho D \left|\nabla z\right|^2} = \overline{\rho} D \left|\nabla \widetilde{z}\right|^2 + \overline{s}_{\chi_z}.$$

The unresolved part, $\overline{s}_{\chi_z}$, requires closure. Two different closures are tested in this work. The first one is the linear relaxation hypothesis (LRH), which assumes a linear relaxation of the variance within the subgrid [17]:

$$\overline{s}_{\chi_z} = \overline{\rho} \frac{\widetilde{z''^2}}{\Delta^2 / \nu_T} \tag{4.6}$$

where Δ is the characteristic filter size and ν_T is the SGS eddy viscosity given by the WALE model.

The second closure for $\overline{s}_{\chi_z}$ is an improved version of the scalar dissipation rate modeling which takes into account that the scores $\mathbf{z}$ are reactive scalars: they have source terms associated to them, which modify $\nabla \mathbf{z}$ and thus also $\widetilde{\chi}_z$. This closure, based on the bimodal-limit (BML) approach, reads [17]:

$$\overline{s}_{\chi_z} = (1 - S_c)\, \overline{\rho} \frac{\widetilde{z''^2}}{\Delta^2 / \nu_T} + S_c \left(-\overline{\rho} D \left|\nabla \widetilde{z}\right|^2 + \overline{zS}_z - \overline{S}_z / 2 \right) \tag{4.7}$$

where $S_c = \widetilde{z''^2} / \widetilde{z}(1 - \widetilde{z})$ is the normalized variance (the unmixedness) of the score z. The LRH hypotesis (Eq. 4.6) is more likely to be valid for small values of S_c, while the BML

model (second term of Eq. 4.7) holds for large values of S_c [17]. Eq. 4.7 combines both the LRH and the BML approaches, taking into account the local flame regime by automatically adjusting the SGS scalar dissipation rate.

Transport equations were solved for $\widetilde{z_1}$ and $\widetilde{z_2}$ (Eq. 4.1), and for $\widetilde{z_1^2}$ and $\widetilde{z_2^2}$ (Eq. 4.5). Filtered look-up tables were generated, where the state-space variables were evaluated using the presumed filtered probability density functions (relation 4.4) and tabulated as:

$$\widetilde{\phi} = \widetilde{\phi}(\widetilde{z_1}, \widetilde{z_1''^2}, \widetilde{z_2}, \widetilde{z_2''^2}).$$

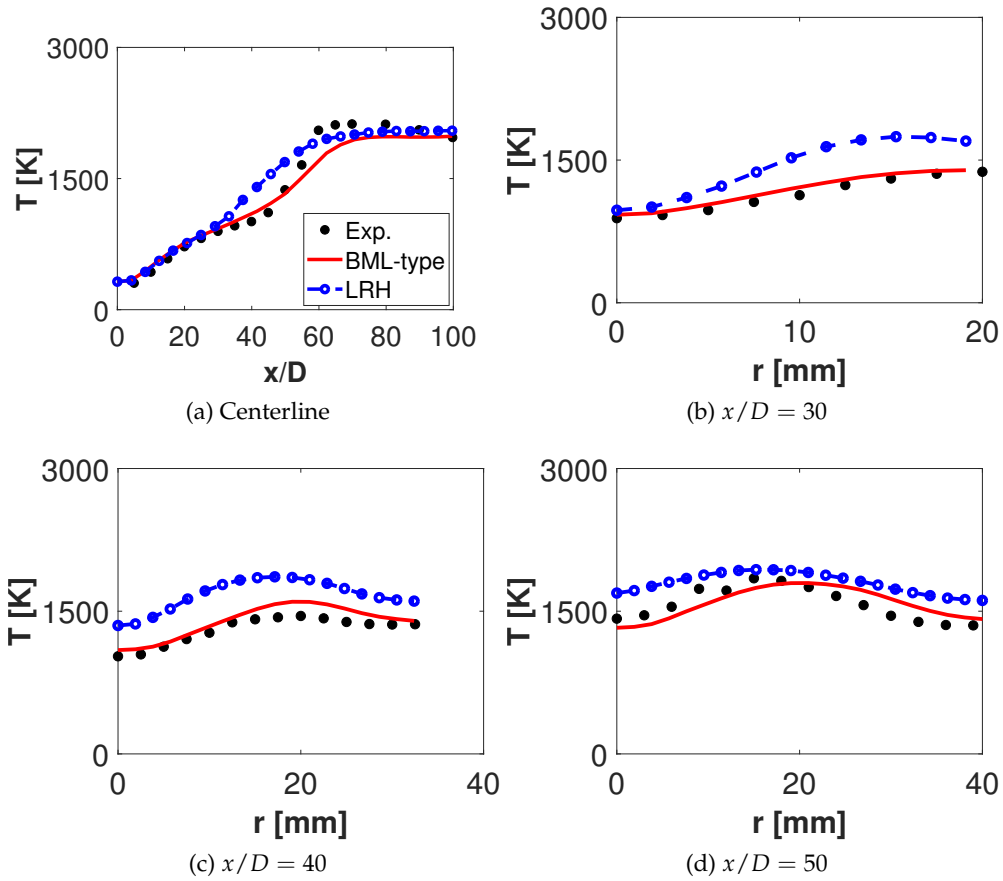
The dimensions of tables were $100 \times 11 \times 100 \times 11$ points in $(\widetilde{z_1}, \widetilde{z_1''^2}, \widetilde{z_2}, \widetilde{z_2''^2})$ space. The numerical setup was the same as the one described in Section 4.3.

4.6.1 LES results - SGS closure

The present section shows the numerical results obtained when applying a subgrid closure to the PC-GPR model. The simulations were done using the CFLF data set. First, the sensitivity to the flame to the modeling of $\overline{s}_{\chi_z}$ is assessed, comparing the closure provided by the linear relaxation hypothesis (LRH model, Eq. 4.6) and closure provided by Eq. 4.7 (BML-type model). Then, the sensitivity of the results to the PCA scaling methods (*standard, range, pareto and vast*, cfr. Section 4.4.2) is shown. Finally, the influence of the number of species in the PC basis is also shown.

4.6.1.1 Effect of the SGS closure model

The centerline average temperature is shown on Fig. 4.12a, and radial profiles at three different axial locations ($x/D = 30, 40, 50$) are shown on Fig. 4.12b-d, comparing the closure models provided by Eq. 4.6 and by Eq. 4.7. It can be observed that the centerline temperature distribution is well captured by the BML-type formulation of $\overline{s}_{\chi_z}$, the SGS scalar dissipation rate of z. The sensitivity of the flame to the modeling of $\overline{s}_{\chi_z}$ is well visible on the centerline and radial profiles as well. The BML-type model allows to better capture the ignition location, situated around $x/D = 40$, while the LRH formulation predicts an early ignition (around $x/D = 30$). The radial profiles (Fig. 4.12b-d) also show a better performance by the LRH-BML model, while the LRH model always over-predicts the temperature. Therefore, the BML-type closure will be used in all the subsequent simulations.



(a) Centerline (b) $x/D = 30$

(c) $x/D = 40$ (d) $x/D = 50$

Figure 4.12: Centerline profile (a) and radial profiles at different axial locations (b-d) of Favre-averaged mean temperature (T). Numerical results (lines): $\overline{s}_{\chi_z}$ given by Eq. 4.7 (LRH-BML), $\overline{s}_{\chi_z}$ given by Eq. 4.6 (LRH). Experimental results (symbols).

4.6.1.2 Effect of the PCA scaling method

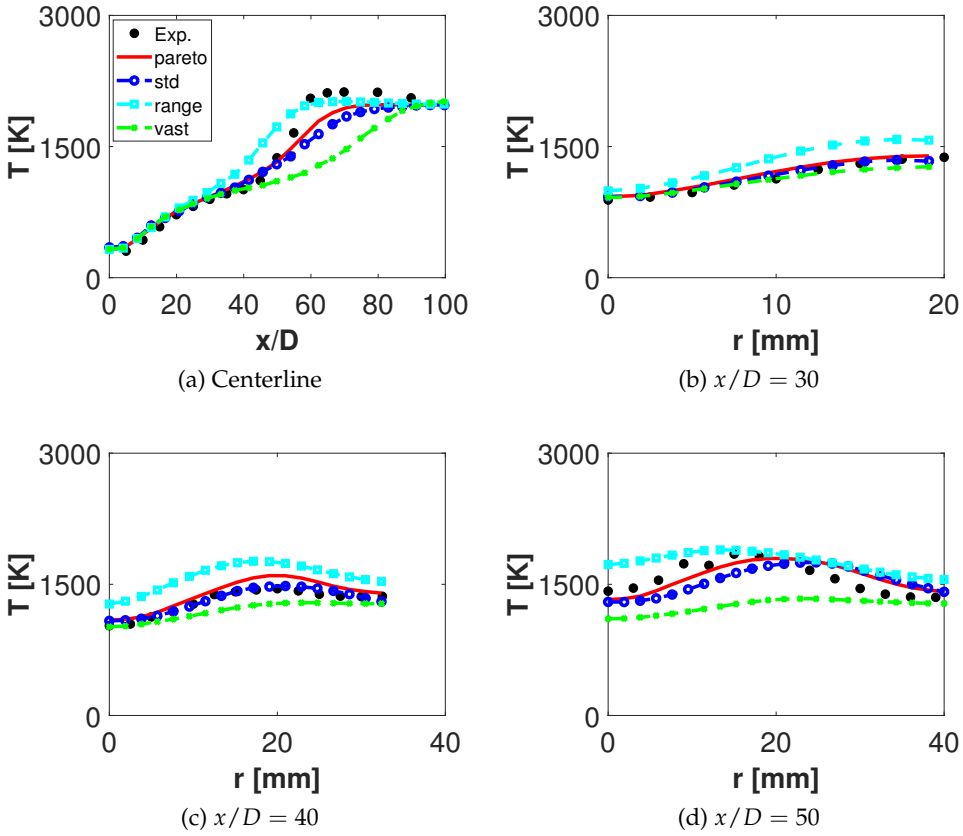
The LES results obtained using the four scaling methods presented in Section 4.4.2, together with the SGS closure, are shown on Fig. 4.13 for the temperature profile. From the centerline profile (Fig. 4.13a), it can be observed that pareto scaling provides the best overall results, accurately predicting the temperature increase due the turbulent mixing between the room temperature fuel jet and the hot vitiated coflow, from the jet exit plane up to $x/D = 40$. Range scaling gives an early ignition ($x/D = 30$), while using vast scaling, ignition occurs further downstream (around $x/D = 70$). Standard scaling provides similar results to pareto up to the ignition point, but after ignition the slope of mean temperature is smaller with standard scaling. The radial temperature profiles at $x/D = 30, 40, 50$ (Fig. 4.13b-d) show similar trends, with pareto scaling being closest to the experimental data.

4.6.1.3 Effect of the variables in the PCA basis

A comparison is made on the influence of the number of species used to create the PC basis. As stated in Section 4.2.2, a subset of the original species present in the kinetic mechanism (6 out of 35) was used for the **X** matrix, consisting of the major species present in the system (CH_4, CO, O_2, CO_2, H_2O and N_2). An analysis is done here in order to verify if building the PC basis on the major species does not lead to a lost of accuracy when the SGS effects are also taken into account. Besides the PC basis based on major species, another basis was constructed using all 35 species present in the mechanism. A new manifold was therefore generated, and a new table was created for the simulation.

Figure 4.14 shows the centerline and radial mean temperature profiles obtained using both set of species. It can be observed that using the major species does not lead to a loss of accuracy in general. Looking at the centerline profile (Fig. 4.14a), both bases yield similar results in the mixing zone ($x/D < 40$), while the major species basis provides a slightly better prediction after ignition ($x/D > 40$). At $x/D = 30$ (Fig. 4.14b), the difference between both bases is not visible; whereas at $x/D = 40$ and $x/D = 50$ (Fig. 4.14c and d), the results from both simulations are very close to each other.

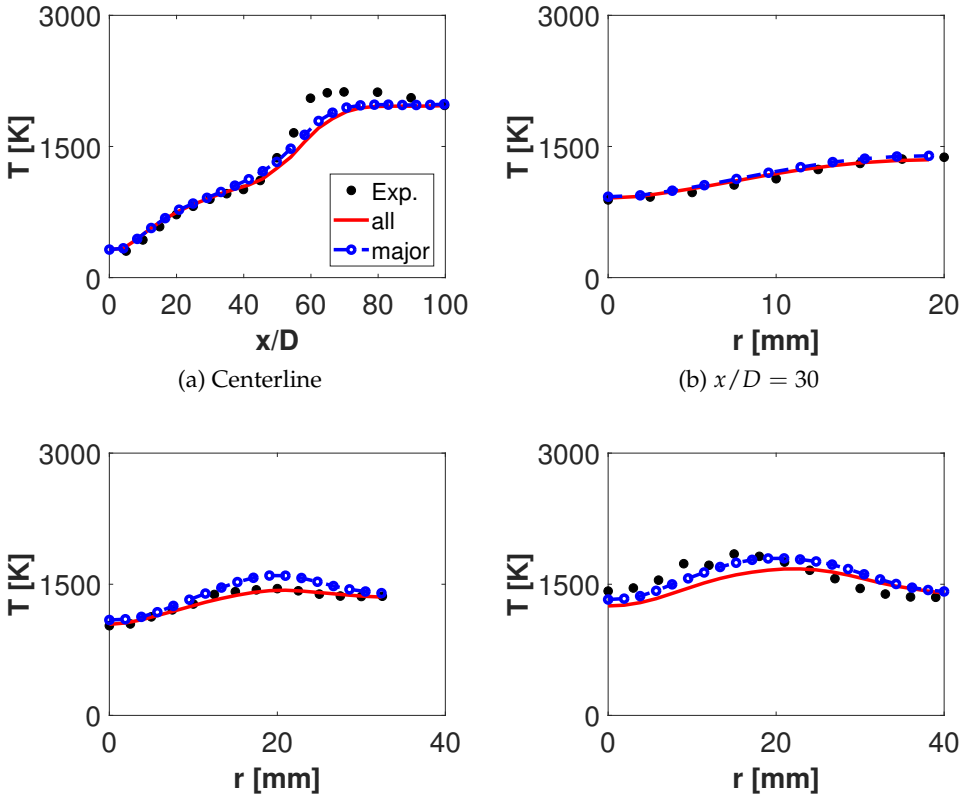
The RMS temperature fluctuations are shown on Fig. 4.15 for both set of species. The location of the peak is accurately predicted by the PC-GPR model, for both set of species.



(a) Centerline (b) $x/D = 30$

(c) $x/D = 40$ (d) $x/D = 50$

Figure 4.13: Comparison of Favre-averaged mean temperature (T) profile on the centerline (a) and radial profiles (b-d) using pareto, standard (std), range and vast scaling methods (lines). Experimental results (symbols).



(a) Centerline

(b) $x/D = 30$

(c) $x/D = 40$

(d) $x/D = 50$

Figure 4.14: Centerline profile (a) and radial profiles at different axial locations (b-d) of Favre-averaged mean temperature (T). Numerical results (lines): PC basis with all 35 species (all), PC basis with major 6 species (major). Experimental results (symbols). Pareto scaling.

The peak value, however, is underpredicted by approximately 200 K.

Centerline profiles of averaged mass fractions of CO_2, H_2O, O_2 and CH_4 are compared to experimental data in Fig. 4.16. These species profiles show some sensitivity to the PC basis, with the major basis leading to larger slopes in the profiles through the turbulent flame base.

Mixture fraction conditioned results for temperature and species mass fractions of CO_2, CH_4 and O_2 are shown on Fig. 4.17. The comparison of temperature profiles shows that the PC-GPR model coupled with SGS closure is able to well capture the autoignition process. The results for CH_4 show a very good agreement with the experimental data. Result for CO_2 and O_2 show quite good agreement, especially close to the flame base, while slightly deviating further downstream ($x/D = 50$).

Instantaneous temperature fields are illustrated on Fig. 4.18 for three different complexity level of the PC-GPR model: classical PC-GPR (results shown in Section 4.4), kernel PC-GPR (results shown in Section 4.5) and PC-GPR with SGS closure. Applying a SGS closure to the PC-GPR model improves the predictions obtained from the LES simulation. Indeed, it can be observed that the lift-off height is now better predicted, and that fuel and oxidizer mix together without significant heat release up to $\sim 40D$ above the jet exit. The inert mixing zone has increased, and the transient region (located approximately between $40 \leq x/D \leq 60$) in which the temperature increases due to autoignition has decreased compared to the kernel PC-GPR model. However, the model still exhibits an extended ignition region.

4.7 Comparison between the CFLF and PSR data sets

4.7.1 Autoignition vs premixed flame propagation

The laboratory-scale vitiated coflow burner presented in this Chapter is an excellent setup for studying autoignition and lifted flames, which are often required in practical applications such as gas turbines and industrial furnaces [24]. A lifted flame is detached from the burner and stabilized at some point downstream. Long lift-off heights are required to increase the rate of mixing and decrease emissions of nitrogen oxides and/or soot thanks to a greater amount of premixing before combustion [32]. Therefore, the understanding of the stabilisation mechanism is lifted-off flames is crucial. In the present flame, the two

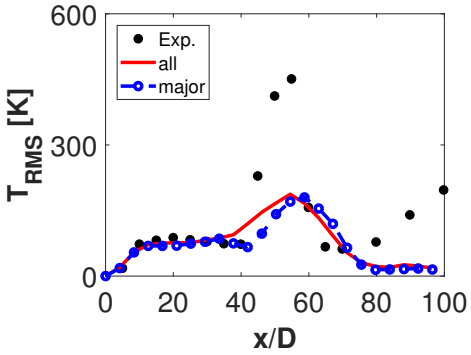


Figure 4.15: Centerline RMS temperature profile. Numerical results (lines): PC basis with all 35 species (all), PC basis with major 6 species (major). Experimental results (symbols). Pareto scaling.

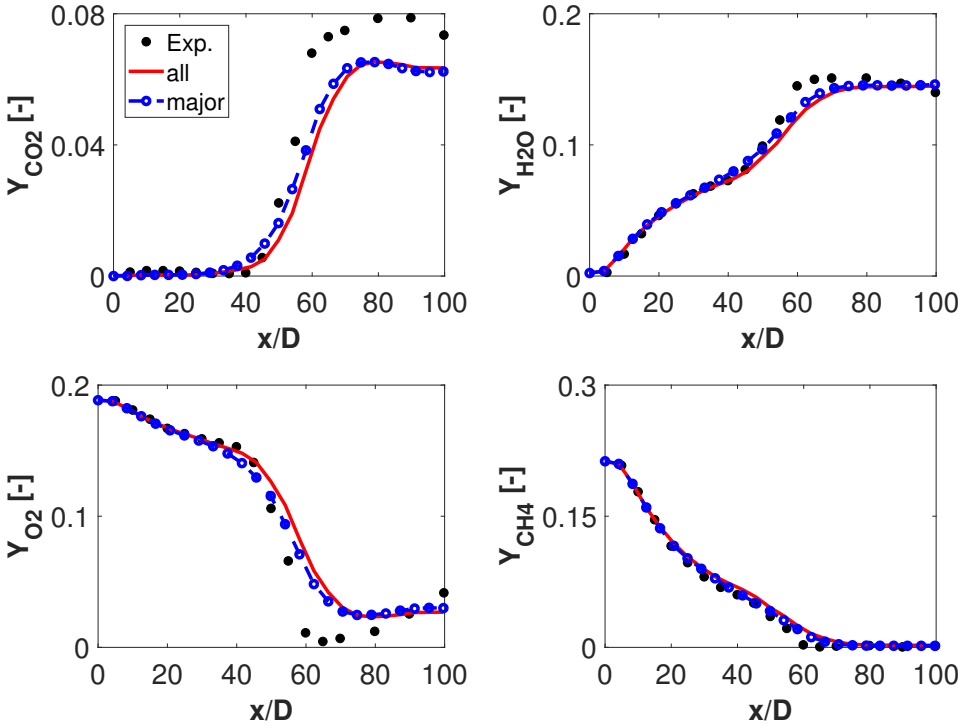
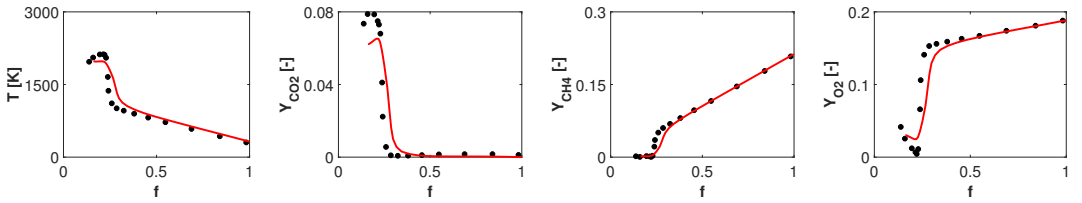
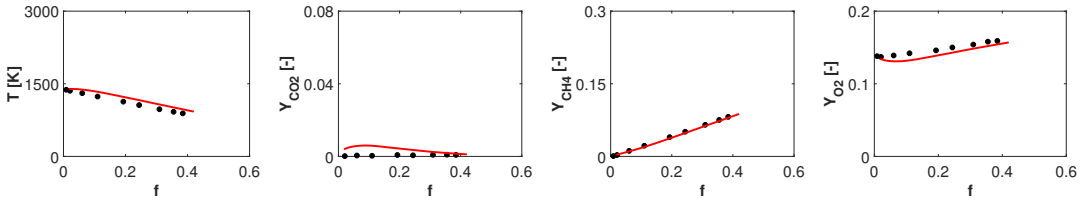


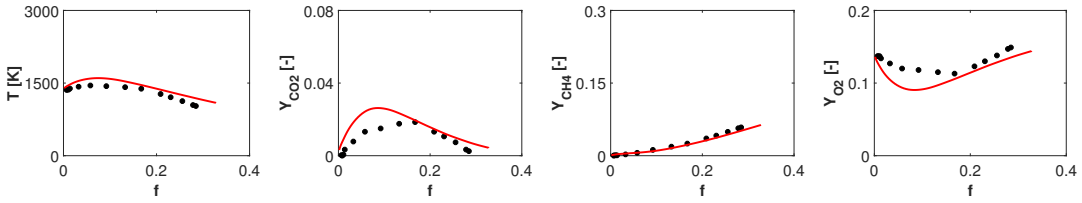
Figure 4.16: Centerline profiles of Favre-averaged mass fractions of CO_2, H_2O, O_2 and CH_4. Numerical results (lines): PC basis with all 35 species (all), PC basis with major 6 species (major). Experimental results (symbols). Pareto scaling.



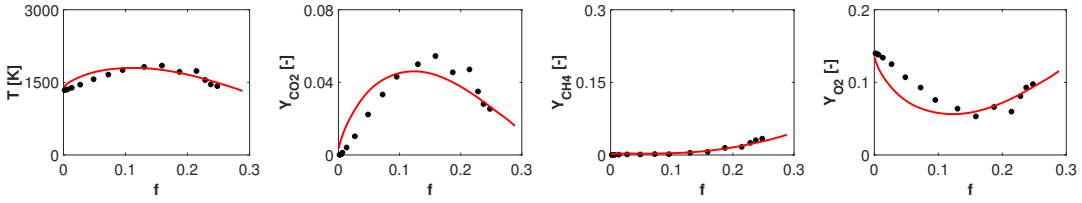
(a) Centerline



(b) $x/D = 30$



(c) $x/D = 40$



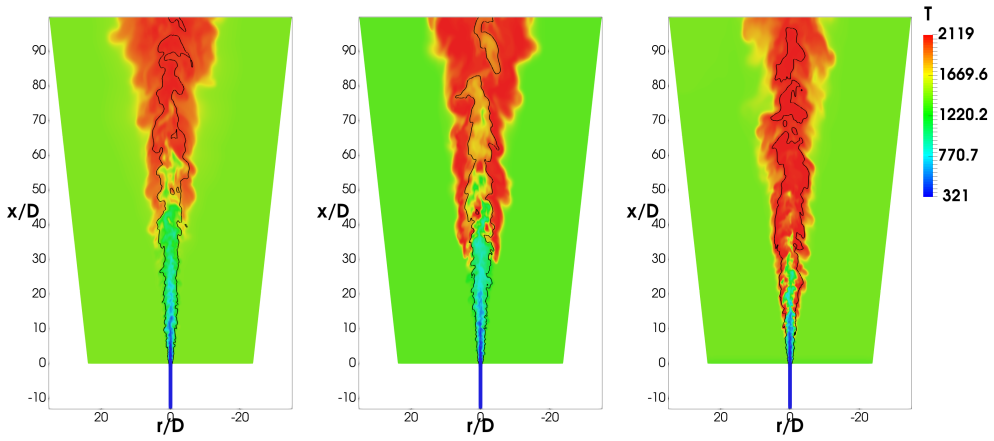
(d) $x/D = 50$

Figure 4.17: Conditional mean versus mixture fraction (f) of mean temperature and mean mass fractions of CO_2, CH_4 and O_2, on the centerline and at three axial locations ($x/D = 30, 40, 50$). Numerical results (line). Experimental results (symbols). Pareto scaling. Major species basis

most pausible mechanisms for flame stabilization are the phenomenon of autoignition and of premixed flame propagation [24]. Autoignition is characterized by a buildup of a radical pool of species prior to ignition and a balance is expected between reaction and convection, while in premixed flame stabilization the preheat zone is characterized by diffusion balancing convection followed by the reaction zone, where the dominant balance is between reaction and diffusion [24]. In a previous study of the Cabra flame by Gordon et al., the authors showed that autoignition is the dominant stabilization mechanism in this flame. They showed that a radical pool of species, consisting of CO, CH_2O, CH_3, C_2H_6, C_2H_4, C_2H_2, H_2O_2 and HO_2 starts forming and increasing at about $x/D = 30$, before ignition and the generation of the OH radical at about $x/D = 38$. They showed that the buildup of minor species such as CH_2O (which then leads to the production of CO through $CH_3 \rightarrow CH_2O \rightarrow CO$) and HO_2 upstream of the flamebase (which provides a source of OH through $HO_2 \rightarrow H_2O_2 \rightarrow 2OH$) prior to OH production is a relevant indicator of autoignition.

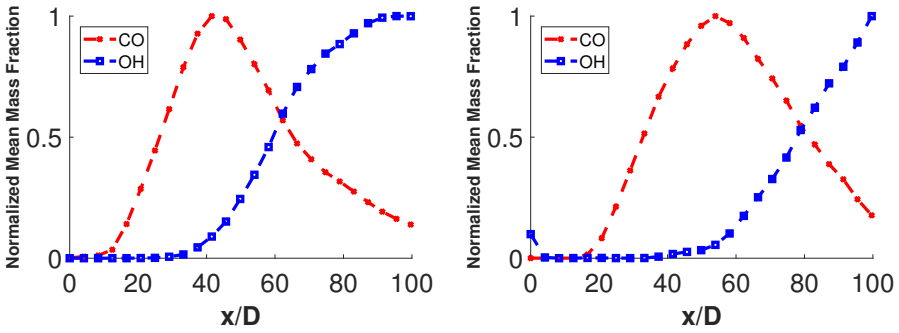
Figure 4.19 shows the axial profile of normalized CO and OH mass fractions, for both the simulation using the CFLF data set and the simulation with the PSR data set. For both cases, it can be seen that CO starts increasing before OH. The production of CO starts upstream of the autoignition zone, followed by fast production within the stabilization region. This is consistent with the findings in [24], showing that indeed autoignition is the primary flame stabilization mechanism at the flame base. Both data sets reached the same conclusion, showing that a $0D$ data set (i.e. the PSR) containing only information about reactions and a $1D$ data set (i.e. the CFLF) containing reactions and diffusion information will exhibit the same behaviour regarding the stabilization mechanism at the flame base. It is also noteworthy to point out that although the PC-GPR model is a physics-based, data-driven, reduced-order combustion model, it is able to reproduce some key features of elaborated physical models, given that the right training data set is fed into the model.

However, it was pointed out in a later study of the Cabra flame by Domingo et al. [17] that although combustion may start by autoignition at the turbulent flame base (which can be combined with premixed flame propagation to some extent), autoignition is less likely to be the major controlling phenomenon after the first ignited points, where premixed flame propagation becomes the dominant combustion mechanism. Also, when



(a) PC-GPR with SGS closure (b) kernel PC-GPR (c) classical PC-GPR

Figure 4.18: Instantaneous temperature fields obtained from (a) PC-GPR with SGS closure, (b) kernel PC-GPR and (c) classical PC-GPR model. The solid line shows the location of the stoichiometric mixture fraction, $f_{st} = 0.177$.



(a) CFLF data set (b) PSR data set

Figure 4.19: Axial profiles of normalized mean mass fractions of CO and OH for (a) the CFLF data set and (b) the PSR data set

autoignition cannot initially occur, flame propagation may be the main stabilizing mechanism. Much further downstream, diffusion combustion becomes the major mechanism controlling the flame. The authors also stated that a simulation of the Cabra flame based solely on an autoignition lookup table was only able to capture the very leading edge of the flame base, and that information about the diffusion process was necessary to fully reproduce the flame stabilization in LES [17]. Therefore, they proposed an approach where an autoignition and a premixed flamelet tables are combined into a single lookup table using a simple linear decomposition based on a user-defined progress variable. LES using this combination of chemistry tabulations was found to reproduce accurately the experimental properties of the Cabra flame. As the effects of diffusion were only included in the premixed flame table, the authors proposed that future work should also tabulate diffusion with an additional diffusion term, for instance using unsteady diffusion flamelets [17].

4.7.2 Scatter plots

Figure 4.20 presents the scatter data of temperature versus mixture fraction, for a comparison between the experimental data and the simulation results obtained using the CFLF and PSR data sets, both with SGS closure. The dashed lines on the numerical results show the pure mixing solution.

The experimental data clearly shows the progress from a predominantly mixing condition ($x/D = 30$, Fig. 4.20a) to a fully burning flame ($x/D = 70$, Fig. 4.20d). As the flame is detached from the nozzle, the central fuel jet entrains the hot coflow with it, creating a partially premixed flow (i.e. the mixing line). At $x/D = 40$, the fuel-rich boundary condition has decreased from $f = 1$ to $f \sim 0.45$, showing the dilution of the fuel samples. Figure 4.20b-c shows that the measurements at $x/D = 40$ and 50 are scattered over the entire manifold, between the mixing (lower) and fast chemistry (upper) limits of temperature. The distribution of points at those two locations is bimodal, i.e. the majority of the samples are either on the pure mixing line or close to the equilibrium curve. Those two locations are located around the experimentally observed lift-off height ($\sim 45D$).

The predicted scatter plots of temperature versus mixture fraction obtained using the CFLF data set are presented in the center column, and the result obtained with the PSR

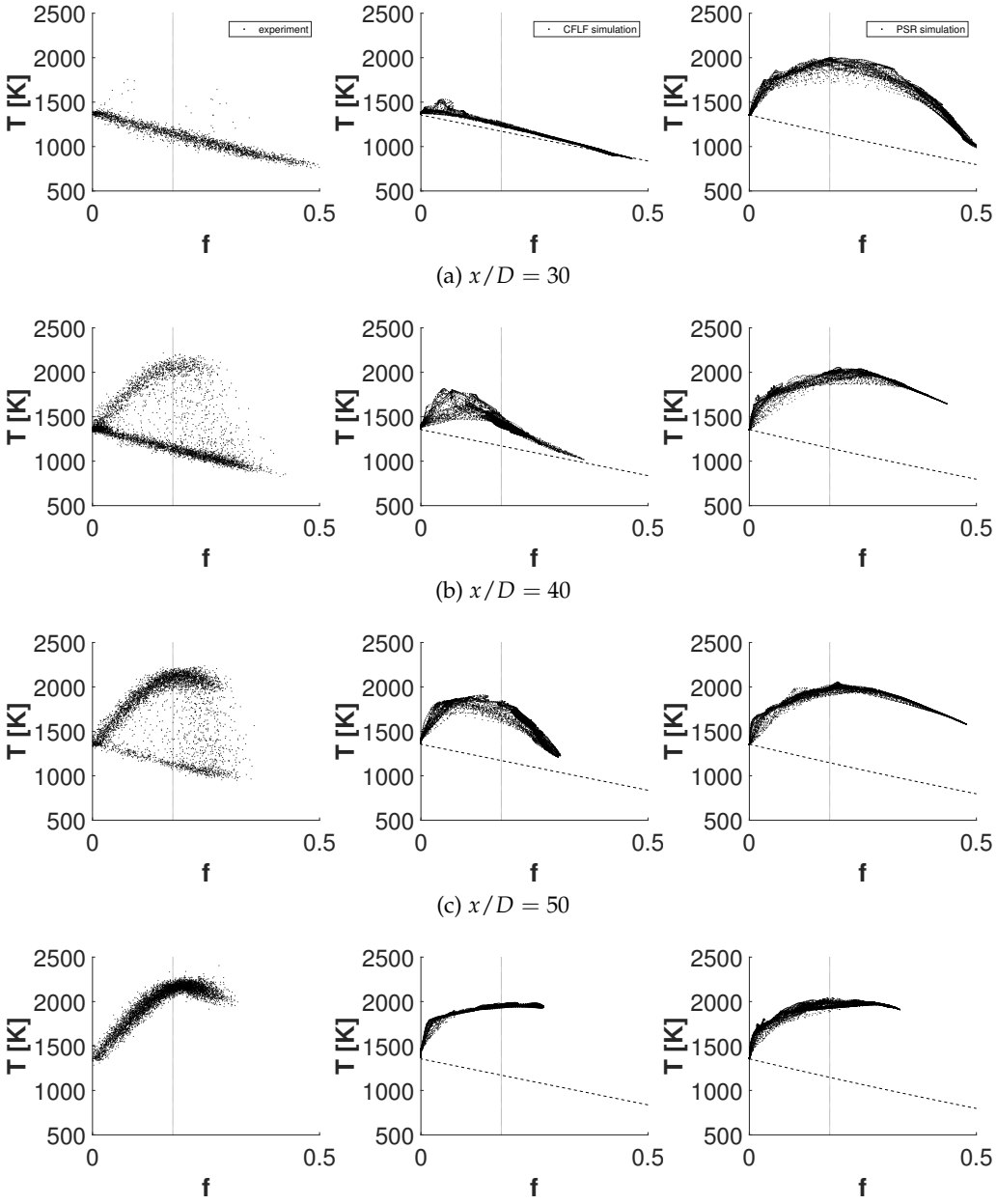
data set are located in the right column of Fig.4.20. Globally, the CFLF simulation is able to better match the experimental single shot measurements. At $x/D = 30$, all the points in the CFLF calculation lie close to the mixing line, while the PSR already predicts ignition around stoichiometric conditions. Moreover, the PSR does not predict the pure mixing on the fuel lean side on the the lower locations (at $x/D = 30$ and 40) due to a short lift-off height prediction. The bimodal shape of the measurements at $x/D = 40$ and 50 is not well captured by the CFLF simulation, which predicts a slightly shorter lift-off height than the experiments. Flame broadening at $x/D = 40$ is not well captured by both models. However, it was also reported that the scatter in the experimental results is due, in part, to experimental uncertainty [10].

4.7.3 Summary

In summary, it can be stated that the CFLF data set follows better the trends shown by the measurements, given that at the lowest axial location ($x/D = 30$) all the points are located on the lower band, and then gradually those points react to reach the fully burnt regime at the highest location ($x/D = 70$). Therefore, a data set representing both the effects of diffusion and reaction mechanisms is necessary for the PC-GPR model in order to accurately predict the initial pure mixing zone, the correct lift-off height and the subsequent flame stabilization region. A data set containing only reaction information ($0D$ PSR) for example) performs poorly in this case. Future work should also investigate the use of more complex data sets, such as those based on One-Dimensional Turbulence (ODT).

4.8 Comparison with the FPV model

In this Section, a comparison of the PC-GPR model with the FPV model [53] is proposed in the context of the lifted Cabra flame. The chosen study is the work of Ihme and See [28] in which the authors used an extended FPV model, called the unsteady flamelet/progress variable (UFPV) model, which extends the steady flamelet/progress variable (SFPV) approach to the prediction of autoignition in turbulent lifted flames. In their work, they used an unsteady flamelet formulation to describe the transient evolution of all thermochemical quantities during the flame ignition process. They explained that the steady flamelet



(a) $x/D = 30$

(b) $x/D = 40$

(c) $x/D = 50$

(d) $x/D = 70$

Figure 4.20: Scatter plots of temperature (T) versus mixture fraction (f) from LES simulations (CFLF data set - middle, PSR data set - right) compared to the experimental data (left), at four different axial locations: $x/D = 30$, 40, 50 and 70. Pure mixing line (dashed) is shown for reference for both data sets. Vertical line (dotted) shows the stoichiometric mixture fraction.

formulation restricts the flamelet solution to that of the S-shaped curve. Therefore, they used the unsteady flamelet equations to generate the flamelet library, and parameterized all thermochemical quantities in function of mixture fraction, a chosen reaction progress parameter, and the stoichiometric scalar dissipation rate. The chemistry was described using the GRI 2.11 mechanism [8]. The reaction progress variable was chosen to be a linear combination of reaction products, defined as $C = Y_{CO_2} + Y_{CO} + Y_{H_2O} + Y_{H_2}$. In order to populate the manifold (i.e. the solutions 'inside' the S-shaped curve), they start with the steady flamelet solution for a given value of the scalar dissipation rate, and then solve the unsteady flamelet equations until the stable solution of the upper branch is reached. The process is then repeated with a different value for the scalar dissipation until the complete state space is populated. This is very similar to the method used in the present study, where the manifold is populated by solving the unsteady counterflow equations for different values of the strain rate. Ihme and See also combined their UFPV model with a statistical approach in order to describe the turbulence/chemistry interactions. For this, they used a presumed PDF closure, in which a beta-distribution is used for the mixture fraction, a statistically most-likely distribution for the reaction progress parameter and a Dirac delta function for the stoichiometric scalar dissipation rate. Therefore, a $5D$ lookup table was generated, and four scalar transport equations were solved describing the conservation of the first two moments of mixture fraction and progress variable. The SFPV and the UFPV models were then applied in LES of the Cabra flame. The authors showed that the UFPV was able to better predict the flame structure, lift-off height and the evolution of the flow field compared to the SFPV model, which predicted a significantly faster ignition process that is not in agreement with the experimental data.

The PC-GPR model employed in this work is also based on unsteady flamelets. The difference with the UFPV lies in the definition of the controlling variables, which in the PC-GPR model are obtained from a PCA analysis on the data set. Moreover, in the PC-GPR model, the state-space is parameterized by two PCs instead of mixture fraction, a user-defined reaction progress parameter and scalar dissipation rate. The final lookup table is $4D$, instead of $5D$ for the UFPV, and nonlinear regression is used instead of linear interpolation.

A qualitative analysis between the UFPV and the PC-GPR models can be done based on

the results provided in [28]. Overall, the results from both models can be seen as similar, despite minor differences. The UFPV model overpredicts the temperature evolution on the centerline in the initial pure mixing region, while the PC-GPR model better captures this part of the flow field. The PC-GPR also predicts a lift-off height very close to the experimental measurement, while the UFPV shows a slightly shorter lift-off, thus predicting an early ignition as well. However, the increase of temperature in the ignition region is better captured by the UFPV model, which also matches well the measurements in the fully burning part of the flame. Radially, the UFPV always overpredicts the value of temperature on the centerline, while the values from the PC-GPR model lie closer to the experiments on the axis. The conditional statistics also show similar results.

4.9 Conclusion

This Chapter showed the application of the PC-score approach coupled with nonlinear Gaussian Process Regression (PC-GPR model) in the context of LES simulations of a lifted methane/air jet flame in a vitiated coflow (Cabra et al. [10]). It was an extension of the standard PC-GPR model used in Chapter 3.

The influence of the canonical reactor used to generate the PC basis has been investigated. The impact of the scaling method involved in the construction of the underlying manifold has also been shown. Moreover, a novel approach based on an extension of the PC-GPR model was developed, combining the kernel density weighting of the PCA basis with nonlinear regression. Finally, a presumed PDF closure model was employed for the first time in the context of the PC-score approach, to evaluate Favre-averaged thermochemical quantities, using beta-distributions for the PCs.

First, two different types of canonical reactors were used and compared to each other: the *1D counter flow laminar flame* (CFLF) setup, and the *0D perfectly stirred reactor* (PSR). Although the standard PC-GPR model was unable to provide satisfactory results in all parts of the domain, the CFLF data set was found to provide better trends, especially in the transient burning zone of the flame. Moreover, all the scaling methods showed similar trends.

The kernel density weighting method allowed to improve the standard PC-GPR model by addressing the issue of non-homogenous data density usually found in numerical

and experimental data sets. The mono-variable kernel PCA method was applied to the CFLF data set, using CO_2 as the conditioning variable. The kernel PC-GPR model led to improved predictions compared to the standard PC-GPR model, capturing well the scalar mixing zone and the autoignition phenomena. However, the temperature rise over the autoignition process was still underpredicted by the kernel PC-GPR model.

The integration of a presumed PDF closure to the PC-GPR model allowed to quantify the significance of turbulence/chemistry interactions in this flame. Simulations results showed significant improvements in the predictions of the flame structure, lift-off height, and spatio-temporal evolution of the flow field. In particular, it was shown that the BML-type closure for the scalar dissipation rate of the PCs was more suited than the LRH closure. Pareto scaling, combined with the subgrid closure, provided the closest solution to the experimental data. Furthermore, the PC basis trained on the major species led to similar results when compared to the full basis, confiming the fact that the PC basis can be trained on the major scalars without any significant loss of accuracy. Finally, it can be stated that neglecting the subgrid fluctuations of the PCs leads to an early ignition of the flame.

A closer analysis of the CO and OH mass fractions from the simulations showed that both data sets are able to predict that autoignition is the dominant flame stabilization mechanism at the flame base. Both data sets reached the same conclusion, showing that the $1D$ data set can also predict autoignition as accurately as the $0D$ data set. Moreover, although autoignition is the main flame stabilization mechanism at the flame base, premixed flame propagation quickly takes over downstream of the flame base, and is also the dominant mechanism when autoignition cannot initially occur. Therefore, it is necessary for the data set to also contain the effects of diffusion in order to accurately predict the different properties of this lifted flame. The PC-GPR model based on the CFLF data set provided more accurate results than the reduced model based on the PSR data. This was also confirmed by scatter plots of temperature, where the CFLF data set followed the trends shown by the measurements more accurately, while the PSR data set would predict a significantly faster ignition process.

Finally, the performances of the PC-GPR model were also qualitatively compared to the FPV model. The results were overall very comparable, but at a lower cost with the PC-GPR

model, which required a $4D$ lookup table, instead of a $5D$ in the case of the FPV model. Moreover, the controlling variables were defined automatically in the PC-GPR model (i.e. the PCs), whereas the FPV model is based upon user-defined parameterizing variables.

CHAPTER 5

CONCLUSION

The present Chapter provides a brief discussion of the main accomplishments and original contributions of the present doctoral thesis, and some perspectives for future reasearch.

A methodology has been proposed for the development of a reduced-order model for reacting flow applications based on Principal Component Analysis (PCA) and nonlinear regression, namely Gaussian Process Regression (GPR). The use of local nonlinear regression was also investigated. The work carried out in the present Thesis represented the first application of the PC-GPR model to complex fuels and combustion problems.

The strength of the method resides in the fact that PCA is able to reduce the dimensionality of a system, and the conjuction of PCA with nonlinear regression allows the user to achieve significant additional reduction while maintaining reasonable accuracy. In addition, it has been shown that the variables identified by PCA may have physical definitions, such as mixture fraction or extent of reaction.

The PC-GPR model was applied in the context of steady and unsteady perfectly stirred reactor (PSR) calculations (Chapter 2), in 3D Large Eddy Simulations (LES) of the Sandia flames D, E and F (Chapter 3) and in the context of a lifted turbulent methane/air flame (Chapter 4).

The proposed reduced-order model was first validated in an *a priori* study based on steady and unsteady perfectly stirred reactor (PSR) calculations, for two different fuels (methane and propane) and three different kinetic mechanisms of increasing complexity. In particular, the following results were highlighted:

- The comparison of the full PSR results to a PSR simulation with 2 principal components (PCs) was achieved using nonlinear regression. The PC-GPR model showed its ability to produce very accurate representation of all state space variables, including

temperature, major and minor species and source terms. For methane, a significant reduction was achieved from the 34 species in the original GRI-3.0 kinetic mechanism down to only 2 PCs. In the case of propane using the Polimi mechanism, the PC-GPR model was able to reduce the size of the system from 162 species down to only 2 PCs, while provinding very accurate steady and unsteady solutions.

- The application of the PC-GPR model using local nonlinear regression (PC-L-GPR) was also demonstrated. The local aspect allowed to improve the accuracy of the regression for complex manifolds, while at the same time decreasing the computational cost associated to the generation of the reduced model.

The PC-GPR model was then validated in an *a posteriori* study in the context of a non-premixed turbulent combustion in a fully three-dimensional Large Eddy Simulation (LES). The experimental Sandia flames D, E and F were chosen to assess the applicability and predictions of the reduced model. The model was trained on unsteady counterflow laminar flames, and the chemistry was descibed using the GRI 3.0 mechanism containing 35 species. The model showed very good accuracy when compared with experimental data using only 2 PCs, for all three flames. The following important results were obtained:

- The first PC, which contains most of the variance in the system, was identified by PCA as being highly correlated with mixture fraction. This result allowed to directly transport mixture fraction in the simulation instead of the first PC, therefore avoiding transporting a reactive scalar. Moreover, the second PC was found to be very correlated with Y_{CO_2}. By doing so, PCA identified in fact the best progress variable for the given system, without the need for any user-defined definition.

- For flame D, 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). The statistical flow field results showed that the PC-GPR model is able to reconstruct all variables with great accuracy, depicting negligible differences between both bases. However, the temperature was overpredicted on the centerline in the lean region of the flame above the stoichiometric flame length. This overprediction was mainly due to the slight overprediction of the mixture fraction, which is caused by neglecting the interaction between turbulent mixing, chemistry, and radiative heat transfer. A

comparison of the intermediate CO and H_2 species near the burner exit also showed good agreement with the experiments.

- The sensitivity to the kinetic mechanism on the predictions was also investigated for flame D. The results from the manifold obtained with the KEE-58 mechanism containing 17 species were compared to the GRI 3.0 mechanism. It was shown that overall, the GRI 3.0 performed better, suggesting that the level of accuracy and detail in the kinetic mechanism is not lost during the construction of the reduced model. A first attempt at providing a subgrid scale (SGS) closure for the PC-GPR model was also tested. A presumed *beta*-shaped probability density function (PDF) was used for mixture fraction. An additional transport equation for the mixture fraction variance was solved. The results showed that the effect of the SGS terms appears to be negligible compared to the resolved part of the flow due to the high resolution. However, the effects of a subgrid closure for the PC-GPR model were further investigated in Chapter 4.

- For flame E and F, the model showed its ability to handle the extinction and re-ignition phenomena characterizing these flames. In particular for flame E, the model was able to reconstruct all scalar variables with great accuracy. Similarily to flame D, the temperature was slightly overpredicted on the centerline near the outlet of the domain. For flame F, centerline and radial profiles for different scalars were accuratley predicted, in physical space as well as in mixture fraction space. Scatter plots of temperature and CO mass fraction were also compared with experimental data. Those plots clearly showed the broad scattering of temperature, espcecially at the first locations close to the burner exit. This broad scattering of the data was correctly predicted by the calculations. Morover, when moving further downstream, the ability of the flame to recover was also well captured by the model. The broad distribution of CO mass fraction at the early stages, as well as its narrow scattering further downstream where the flame has reingnited, are also captured by the reduced model.

- The manifold accessed during the simulation showed that for flame D, most of the data is contained near the equilibrium solution, confirming that the flame D simula-

tion did not experience significant extinction and reignition. For flame F however, the region of the manifold accessed during the simulation was wider compared to flame D, showing that the data was evenly distributed between the equilibrium solution and the extinction region of the manifold. For both flames, the simulations stayed bounded to the training manifold.

- A qualitative comparison of the PC-GPR model with the FPV model was also performed. The study of Ihme and Pitsch was used [27] as benchmark. Results from both models were very similar to each other, but the numerical cost associated to the PC-GPR model was shown to be lower than the FPV cost. Indeed, the FPV model formulation was based on steady counterflow flamelets and parameterized by user-defined variables, namely mixture fraction and a progress variable. Morover, the FPV model had to be combined with a PDF approach, resulting in additional transport equations for the variance of mixture fraction and progress variable. The PC-GPR model on the other hand was based on unsteady counterflow flamelets, and parameterized by the best controlling variables for the system defined by PCA (i.e. the PCs), without the need for SGS closure. The FPV model therfore required 4 scalar transport equations and a $4D$ lookup table, while the PC-GPR provided similar results with only 2 transported scalars and a $2D$ table, which saves CPU time and decreases memory cost during the simulation. The unsteady canonical reactor and the PCs therefore allowed to avoid using a SGS closure. The PC-GPR model also shared some limitations with the FPV model, which can be addressed by extending the lookup table to a variable taking into account heat loss effects, such as the enthalpy.

In Chapter 4, the PC-GPR model was applied *a posteriori* on a more challenging case, a lifted methane/air flame, and several key features of the model were investigated: the sensitivty to the training data set, the influence of the scaling methods, the issue of data sampling and the potential of a subgrid scale closure.

- The influence of the canonical reactor used to generate the PC basis has been investigated. To this end, two different data sets were tried: the 1D counterflow laminar flame (CFLF) setup, and the 0D perfectly stirred reactor (PSR). The CFLF data set

was found to provide better trends, especially in the transient burning zone of the flame. With the CFLF data set, combustion occured faster than with the PSR, and the slope of increase of the temperature in the transient region is better predicted by the CFLF data set. Moreover, the fully developed flame zone is also better captured by the CFLF data set.

A closer analysis of the autoignition process, which is the dominant flame stabilization mechanism at the base of this flame, showed that the 1D data set can also predict autoignition as accurately as the 0D data set. It was also shown that although autoignition is the main stabilization mechanism at the flame base, premixed flame propagation quickly takes over downstream of the flame base, and is also the dominant mechanism when autoignition cannot initially occur. Therefore, it is necessary for the data set to also contain the effects of diffusion in order to accurately predict the different properties of this lifted flame. This was also confirmed by scatter plots of temperature, where the CFLF data set followed the trends shown by the measurements more accurately, while the PSR data set would predict a significantly faster ignition process.

- The impact on the numerical predictions of the various scaling methods for PCA found in the litterature was also investigated. Four different scalings were tried: standard, range, pareto and vast. Globally, all the scaling methods provided similar trends in the predictions of ignition and lift-off height. Therefore, the choice of the scaling method was shown to be driven mainly by other parameters, such as the smoothness of the manifold and the easiness in the regression of it. The analysis confirmed that pareto scaling was the most appropriate method, as it ensured smoothness of the manifold and easiness in the regression.

- The issue of non-homogenous data density, typically found in large numerical or experimental data sets, was also addressed. The kernel density weighting method allowed to improve the predictions of the standard PC-GPR model. The mono-variable kernel PCA method was applied to the CFLF data set, with CO_2 as the conditioning variable. The lift-off height was better captured compared to the standard PC-GPR

model. However, the temperature rise over the transient region was still underpredicted by the kernel PC-GPR model.

- The integration of a subgrid scale closure to the PC-GPR model allowed to significantly improve the simulations results. A presumed PDF closure was used, and the BML (bimodal limit)-type closure for the scalar dissipation rate of the PCs was more suited than the LRH (linear relaxation hypothesis) closure. Simulations results showed significant improvements in the predictions of the flame structure, lift-off height, and spatio-temporal evolution of the flow field. Pareto scaling, combined with the subgrid closure, provided the closest solution to the experimental data. Furthermore, the PC basis trained on the major species led to similar results when compared to the full basis, confiming the fact that the PC basis can be trained on the major scalars without any significant loss of accuracy.

- A qualitative comparison of the PC-GPR model with the FPV model for was also performed for this flame. The study used was carried out by Ihme and See [28]. The results were overall very comparable, but at a lower cost with the PC-GPR model, which required a 4D lookup table, instead of a 5D in the case of the FPV model. The FPV model was based on the unsteady flamelet formulation, and the underlying manifold was parameterized in function of mixture fraction, a user-defined reaction progress parameter, and the stoichiometric scalar dissipation rate. Moreover, a presumed PDF closure was applied, in which a beta-distribution is used for the mixture fraction, a statistically most-likely distribution for the reaction progress parameter and a Dirac delta function for the stoichiometric scalar dissipation rate. Therefore, a 5D lookup table was generated, and four scalar transport equations were solved. The PC-GPR reduced model was based on unsteady flamelets, and parameterized by the first two PCs. A beta-PDF closure was applied for both PCs. The final lookup table was of dimension 4. The results from both models were similar, despite minor differences. The PC-GPR model better captured the initial pure mixing region and predicted a lift-off height closer to the experimental measurement. However, the increase of temperature in the ignition region was better captured by the UFPV model, which also matched well the measurements in the fully burning part of the

flame.

In conclusion, the strength of the PC-GPR 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 perspectives

The research presented in this dissertation has shown the great potential of Principal Component Analysis in the modeling and analysis of turbulent combustion systems. This research paves the way for future developments in the field of data-based, physics-driven combustion modeling.

Several concepts are of interest for future work. The use of additional canonical reactors with increasing complexity, such as the One-Dimensional Turbulence (ODT), in the training database would strenghten the model even further. The application of PCA on DNS and experimental data sets of reacting flows would also be very interesting. Moreover, the development of subgrid scale closure models specific to PCA and the inclusion of more complex methods for the treatment turbulence/chemistry interactions, such as transported PDF methods, would provide an interesting extension of the model to more complex systems.

REFERENCES

[1] R. Barlow, J. Frank, A. Karpetis, and J.-Y. Chen, *Piloted methane/air jet flames: Transport effects and aspects of scalar structure*, Combust. Flame, 143 (2005), pp. 433–449.

[2] R. S. Barlow and J. H. Frank, *Effects of turbulence on species mass fractions in methane/air jet flames*, Proc. Combust. Inst., 27 (1998), pp. 1087–1095.

[3] A. Bellemans, T. Magin, A. Munafo, G. Degrez, and A. Parente, *Reduction of a collisional-radiative mechanism for argon plasma based on principal component analysis*, Physics Plasmas, (2015).

[4] A. Biglari and J. C. Sutherland, *A filter-independent model identification technique for turbulent combustion modeling*, Combustion and Flame, 159 (2012), pp. 1960–1970.

[5] A. Biglari and J. C. Sutherland, *An a-posteriori evaluation of principal component analysis-based models for turbulent combustion simulations*, Combust. Flame, 162 (2015).

[6] R. Bilger, S. Starner, and R. Kee, *On reduced mechanisms for methane-air combustion in nonpremixed flames*, Combustion and Flame, 80 (1990), pp. 135–149.

[7] H. Bongers, J. A. V. Oijen, and L. P. H. D. Goey, *Intrinsic low-dimensional manifold method extended with diffusion*, Proceedings of the Combustion Institute, 29 (2002), pp. 1371–1378.

[8] C. Bowman, R. Hanson, D. Davidson, W. Gardiner, V. Lissianski, G. Smith, D. Golden, M. Frenklach, and M. Goldenberg, *Gri-mech 2.11*, 1997.

[9] R. Brad, A. Tomlin, M. Fairweather, and J. Griffiths, *The application of chemical reduction methods to a combustion system exhibiting complex dynamics*, Proceedings of the Combustion Institute, 31 (2007), pp. 455–463.

[10] R. Cabra, J.-Y. Chen, R. Dibble, A. Karpetis, and R. Barlow, *Lifted methane-air jet flames in a vitiated coflow*, Combust. Flame, 143 (2005), pp. 491–506.

[11] W. Cleveland, E. Grosse, and W. Shyu, *Local regression models*, Statistical models, 79 (1992), pp. 531–554.

[12] A. Coussement, O. Gicquel, and A. Parente, *Kernel density weighted principal component analysis of combustion processes*, Combust. Flame, 159 (2012), pp. 2844–2855.

[13] ——, *Mg-local-pca method for reduced order combustion modelling*, Proc. Combust. Inst., 34 (2013), pp. 1117–1123.

[14] A. Coussement, B. Isaac, O. Gicquel, and A. Parente, *Assessment of chemistry reduction methods based on pca: comparison of mg-pca and score-pca approaches*, Combust. Flame, 168 (2016).

[15] A. Cuoci, A. Frassoldati, T. Faravelli, and E. Ranzi, *Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method*, Energy & Fuels, 27 (2013), pp. 7730–7753.

[16] ——, *Opensmoke++: An object-oriented frame-work for the numerical modeling of reactive systems with detailed kinetic mechanisms*, Computer Physics Communications, 192 (2015), pp. 237–264.

[17] P. Domingo, L. Vervisch, and D. Veynante, *Large-eddy simulation of a lifted methane jet flame in a vitiated coflow*, Combust. Flame, 152 (2008), pp. 415–432.

[18] T. Echekki and H. Mirgolbabaei, *Principal component transport in turbulent combustion: a posteriori analysis*, Combust. Flame, 162 (2015), pp. 1919–1933.

[19] J. Einbeck, B. Isaac, L. Evers, and A. Parente, *Penalized regression on principal manifolds with application to combustion modeling*, in Proceedings of the 27th International Workshop on Statistical Modelling, 2012.

[20] S. Elbahloul and S. Rigopoulos, *Rate-controlled constrained equilibrium (rcce) simulations of turbulent partially premixed flames (sandia d/e/f) and comparison with detailed chemistry*, Combust. Flame, 162 (2015), pp. 2256–2271.

[21] J. Friedman, *Multivariate adaptive regression splines*, The Annals of Statictics, 19 (1991), pp. 1–67.

[22] O. Gicquel, N. Darabiha, and D. Thevenin, *Laminar premixed hydrogen/air counterflow flame simulations using flame prolongation of ildm with differential diffusion*, Proc. Combust. Inst., 28 (2000), pp. 1901–1908.

[23] D. Goodwin, *Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes*, 2009.

[24] R. L. Gordon, A. R. Masri, S. B. Pope, and G. M. Goldin, *Transport budgets in turbulent lifted flames of methane autoigniting in a vitiated co-flow*, Combust. Flame, 151 (2007), pp. 495–511.

[25] J. Griffiths, *Reduced kinetic models and their application to practical combustion systems*, Progress in Energy and Combustion Science, 21 (1995), pp. 25–107.

[26] S. Humer, A. Frassoldati, S. Granata, T. Faravelli, E. Ranzi, R. Seiser, and K. Seshadri, *Experimental and kinetic modeling study of combustion of jp-8, its surrogates and reference components in laminar nonpremixed flows*, Proceedings of the Combustion Institute, 31 (2007), pp. 393–400.

[27] M. Ihme and H. Pitsch, *Prediction of extinction and reignition in nonpremixed turbulent flames using a flamelet/progress variable model 2. application in les of sandia flames d and e*, Combust. Flame, 155 (2008), pp. 90–107.

[28] M. Ihme and Y. C. See, *Prediction of autoignition in a lifted methane/air flame using an unsteady flamelet/progress variable model*, Combust. Flame, 157 (2010), pp. 1850–1862.

[29] B. Isaac, A. Coussement, O. Gicquel, P. Smith, and A. Parente, *Reduced-order pca models for chemical reacting flows*, Combust. Flame, 161 (2014).

[30] B. Isaac, J. Thornock, J. Sutherland, P. Smith, and A. Parente, *Advanced regression methods for combustion modelling using principal components*, Combust. Flame, 162 (2015), pp. 2592–2601.

[31] I. Jolliffe, *Principal Component Analysis*, Springer-Verlag New York, 2002.

[32] S. Karami, E. R. Hawkes, M. Talei, and J. H. Chen, *Mechanisms of flame stabilisation at low lifted height in a turbulent lifted slot-jet flame*, Journal of Fluid Mechanics, 777 (2015), pp. 633–689.

[33] J. C. Keck, *Rate-controlled constrained-equilibrium theory of chemical reactions in complex systems*, Progress in Energy and Combustion Science, 16 (1990), pp. 125–154.

[34] J. C. Keck and D. Gillespie, *Rate-controlled partial-equilibrium method for treating reacting gas mixtures*, Combust. Flame, 17 (1971), pp. 237–241.

[35] M. Klein, A. Sadiki, and J. Janicka, *A digital filter based generation of inflow data for spatially developing direct numerical or large eddy simulation*, Journal of Computational Physics, 186 (2003), pp. 652–665.

[36] S. Lam and D. Goussis, *Understanding complex chemical kinetics with computational singular perturbation*, Proc. Combust. Inst., 22 (1988), pp. 931–941.

[37] T. Lovas, P. Amneus, F. Mauss, and E. Mastorakos, *Comparison of automatic reduction procedures for ignition chemistry*, Proceedings of the Combustion Institute, 29 (2002), pp. 1387–1393.

[38] U. Maas and S. Pope, *Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space*, Combust. Flame, 88 (1992), pp. 239–264.

[39] M. R. Malik, B. Isaac, A. Coussement, P. Smith, and A. Parente, *Principal component analysis coupled with nonlinear regression for chemistry reduction*, Combust. Flame, 187 (2018), pp. 30–41.

[40] M. R. Malik, P. O. Vega, A. Coussement, and A. Parente, *Combustion modeling using principal component analysis: a posteriori validation on sandia flames d, e and f*, Proceedings of the Combustion Institute, 38 (2021).

[41] H. Mirgolbabaei and T. Echekki, *A novel principal component analysis-based acceleration scheme for les-odt*, Combust. Flame, 160 (2013), pp. 898–908.

[42] ——, *Nonlinear reduction of combustion composition space with kernel principal component analysis*, Combust. Flame, 161 (2014), pp. 118–126.

[43] D. Nguyen-Tuong, M. Seeger, and J. Peters, *Model learning with local gaussian process regression*, Advanced Robotics, 23 (2009), pp. 2015–2034.

[44] F. Nicoud and F. Ducros, *Subgrid-scale stress modelling based on the square of the velocity gradient tensor*, Flow, Turbulence and Combustion, 62 (1999), pp. 183–200.

[45] O. Owoyele and T. Echekki, *Toward computationally efficient combustion dns with complex fuels via principal component transport*, Combustion Theory and Modelling, 21 (2017), pp. 770–798.

[46] H.-T. Pao, *A comparison of neural network and multiple regression analysis in modeling capital structure*, Expert Systems with Applications, 35 (2008), pp. 720–727.

[47] A. Parente and J. C. Sutherland, *Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity*, Combust. Flame, 160 (2013), pp. 340–350.

[48] A. Parente, J. C. Sutherland, L. Tognotti, and P. J. Smith, *Identification of low-dimensional manifolds in turbulent flames*, Proceedings of the Combustion Institute, 32 (2009), pp. 1579–1586.

[49] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist., 33 (1962), pp. 1065–1076.

[50] K. Peerenboom, A. Parente, T. Kozak, A. Bogaerts, and G. Degrez, *Dimension reduction of non-equilibrium plasma kinetic models using principal component analysis*, Plasma Sources Sci. T., (2015).

[51] N. Peters, *Laminar diffusion flamelet models in non-premixed turbulent combustion*, Progress in Energy and Combustion Science, 10 (1984), pp. 319–339.

[52] ——, *Laminar flamelet concepts in turbulent combustion*, Proc. Combust. Inst., 21 (1986), pp. 1231–1250.

[53] C. D. Pierce and P. Moin, *Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion*, Journal of Fluid Mechanics, 504 (2004), pp. 73–97.

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

[55] R. Ranade and T. Echekki, *A framework for data-based turbulent combustion closure: A posteriori validation*, Combust. Flame, 210 (2019), pp. 279–291.

[56] ——, *A framework for data-based turbulent combustion closure: A priori validation*, Combust. Flame, 206 (2019), pp. 490–505.

[57] C. Rasmussen, *Gaussian processes for machine learning*, 2006.

[58] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist., 27 (1956), pp. 832–837.

[59] C. Schneider, A. Dreizler, J. Janicka, and E. Hassel, *Flow field measurements of stable and locally extinguishing hydrocarbon-fueled jet flames*, Combust. Flame, 135 (2003), pp. 185–190.

[60] 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, and Z. Qin, 1999.

[61] A. Smola and B. Scholkopf, *A tutorial on support vector regression*, Statistics and Computing, 14 (2004), pp. 199–222.

[62] J. C. Sutherland and A. Parente, *Combustion modeling using principal component analysis*, Proceedings of the Combustion Institute, 32 (2009), pp. 1563–1570.

[63] J. van Oijen, A. Donini, R. Bastiaans, J. ten Thije Boonkkamp, and L. de Goey, *State-of-the-art in premixed combustion modeling using flamelet generated manifolds*, Progress in Energy and Combustion Science, 57 (2016), pp. 30–74.

[64] F. A. Williams, *Chemical-kinetic mechanisms for combustion applications*, 2010.

[65] Y. Yang, S. Pope, and J. Chen, *Empirical low-dimensional manifolds in composition space*, Combust. Flame, 160 (2013), pp. 1967–1980.