POLITECNICO DI MILANO

Scuola di Ingegneria dei Processi Industriali

Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta"

Corso di Laurea Magistrale in Ingegneria Chimica

MANIFOLD GENERATED FROM LOCAL PRINCIPAL COMPONENT ANALYSIS (MG-L-PCA): APPLICATIONS TO COMBUSTION MODELING

Relatore: Ing. Alberto CUOCI

Correlatore: Ing. Alessandro PARENTE

Tesi di Laurea di:

Mafalda CILIBERTI

Matr. 764308

Anno Accademico 2011-2012

Contents

List of Figures

List of Tables

Abstract

Development of accurate combustion models able of predict the full thermo-chemical state of combustion systems with acceptable costs remains a challenge. Recently, Principal Component Analysis (PCA) was applied to the study of reacting systems, and its ability to reduce their dimensionality with minimum reconstruction error was demonstrated. The present work introduces a new approach, called Manifold Generated from Local PCA (MG-L-PCA), which fully couples the manifold identified by a PCA and a CFD solver.

In the first chapters, the laminarSMOKE solver adopted to simulate several laminar flames will be introduced and the results obtained with a flame feed with hydrogen and three methane coflowing flames will be proposed, bearing out the possibility to represent the involved physical-chemical variables through this tool. Then, the MG-L-PCA technique will be illustrated, showing its ability to find a low-dimensional, attracting thermo-chemical manifold through an *a priori* computation of some basis matrices, made up by the eigenvectors of the system covariance matrices. These matrices will be then exploited to simulate reacting systems, by solving transport equations only for a subset of the original state-space variables. Results show the possibility to reduce the number of the variables to be transported, while maintaining a low reconstruction error for state variables and corresponding source terms.

Abstract

Lo sviluppo di modelli di combustione accurati, capaci di predirre in modo esaustivo lo stato termo-chimico di sistemi interessati da combustione con costi accettabili, rimane un'attività impegnativa e complessa. Recentemente, l'analisi delle componenti principali (PCA) è stata applicata allo studio di sistemi reattivi ed è stata dimostrata la sua capacità di ridurre la loro dimensionalità con un errore di ricostruzione di minimo. Il presente lavoro di tesi introduce un nuovo approccio, chiamato Manifold Generated from Local PCA (MG-L-PCA), che accoppia completamente lo spazio a più dimensioni identificato con la PCA e il solver CFD.

Nei primi capitoli sarà introdotto il solver laminarSMOKE adottato per simulare diverse fiamme laminari e saranno presentati i risultati ottenuti, confermando la possibilità di rappresentare le grandezze fisico-chimiche in gioco con questo strumento. Poi, sarà illustrata la tecnica MG-L-PCA, mostrando la sua abilità di trovare un interessante spazio a più dimensioni attraverso un calcolo a priori di alcune matrici base, costituite dagli autovettori delle matrici covarianza del sistema. Queste matrici saranno poi sfruttate per simulare sistemi reagenti, risolvendo le equazioni di trasporto solo per un sottoinsieme delle variabili di stato originarie. I risultati mostrano la possibilità di ridurre del numero delle variabili da trasportare, pur mantenendo un errore di ricostruzione basso per le variabili di stato e per i termini

sorgente corrispondenti.

Introduction

Combustion modeling

The knowledge of the combustion phenomena is a key point in a large variety of applications, that includes furnaces, burners and engines design, study of the combustion emissions and fire safety analysis. Many physical and chemical processes have to be considered, like issues related to fluid mechanics, thermodynamics, combustion kinetics and heat exchange, involving a wide range of time and length scales. This makes hard the description of combustion phenomena and a lot of models have been adopted to model combustion. Thanks to the development of modern computers and the continuous improvement of the numerical techniques, it's possible to describe the evaluation of a flame through the *Computational Fluid Mechanics (CFD)*.

Computational Fluid Dynamics (CFD)

Computational Fluid Dynamics (CFD) includes all the computational tools for the study and the solutions of problems that involve fluids flow. Mass, energy and momentum conservation are considered through the transport equations (of chemical species, total enthalpy, etc...) and constitutive equations that describes phisical

properties of the fluids, as the Newton's Law, the Fourier's Law, the Fick's Law and the equations of state that express the relations between temperature, pression and the density of a fluid [10].

A wide variety of numerical methods, free or commercial, has been employed, but three basic steps are common to all CFD methods [10][1]:

1. **Pre-Processing**. In this step, all the information necessary for the numeric solver are definited and the input information is converted in an suitable one for the solver. The following operations are usually included:

   - definition of the c*omputational domain;*
   - discretization of the flow domain into cells: a set of grid lines or curves define the *mesh* and a set of nodes at which the flow variables are to be calculated;
   - phisical and chemical phenomena modeling;
   - definition of thermodynacis and transport fluids properties;
   - evaluation of proper bondary conditions.

2. **Solver**. This is the core of the fluid dynamics code: discretization of the governing equations is carried out in this step. The exact partial differential equations to be solved are replaced by approximate algebraic equations written in terms of the nodal values of the dependent variables. Among the numerous discretization methods, the following ones are the most common.

   (a) The *finite difference method* estimates spatial derivatives in terms of the nodal values and spacing between nodes. The governing equations are then written in terms of the nodal unknowns at each interior node.

Finite volume methods, related to finite difference methods, may be derived by a volume integration of the equations of motion, with application of the divergence theorem, reducing by one the order of the differential equations. Equivalently, macroscopic balance equations are written on each cell of the domain. This ensures the conservations of the variables.

(b) *Finite elements methods* are weighted residual techniques in which the unknown dependent variables are expressed in terms of basis functions interpolating among the nodal values. The basis functions are substituted into the equations of motion, resulting in error residuals which are multiplied by the weighting functions, integrated over the control volume, and set to zero to produce algebraic equations in terms of the nodal unknowns. Selection of the weighting functions defines the various finite element methods.

(c) Spectral methods approximate unknown variables through Fourier series or Chebyshev polinomies. The main difference with the previous methods is that spectral methods take on a global approach, with approximations effective in all the domain, while the others use a local approach.

After discretization of the equations and application of the boundary conditions, the result is a set of algebraic equations for the nodal unknown variables. Discretization in time is also required for the $\frac{\partial}{\partial t}$ time derivative terms in unsteady flow. The descretized equations represent an approximation of the exact equations, and their solution gives an approximation for the flow variables. The accuracy of the solution improves as the grid is refined, that is, as the number of nodal points is increased.

When the discretization step is carried out, the solver accounts for the solution

of the algebraic equations.

3. **Post-Processing**. All the data collected through the solution of transport equation by the solver are organized and visualized. For example, there are specific tools that allow to visualize in a graphic form the grid, the vector plots, the contour plots, pathlines or streamlines, etc...

CFD advantages In comparison with the traditional experimental activities, the main benefit of the computational fluid dynamics is the possibility to get data and information saving costs and time. In general the more data are required, the higher is the cost of an experiment, in terms of equipment, time spent by the operators, and so on. Once the simulation is developed for a particular system, CFD instead allows to get all the data required (for example, for sensitivity analysis) with costs almost indipendent from the quantity of the desired data. Last but not least, CFD allows to study systems under condition of risk (like fires symulations) without danger for people or things.

CFD limits Computational Fluid Dynamics is a powerfull tool but experimental validations cannot be put aside due to semplifications in adopted models and approximations:

- adopted models and corresponding differential equations have semplications and approximations (model errors);
- the discretization of the partial differential equations in space and time introduce approximations (dicretization errors);
- the translation of the partial differential equations into the computerized model may contain errors (programming errors);

- during the resolution step, an iterative algorithm is applied and there can be convergence problems (iterations errors).

Moreover, case studies through a CFD code require time, too. The time required for the preprocessing is not always neglectable; for istance, the grid definition for actual cases is a sensitive step and can employ half of the time spent by an industry for a CFD project [1]. Time required by the solver to achieve accurate results can exponentially grow due to the number of equations to be solved and numerical issues involved in the resolution.

As regards combustion modeling, it represents a challenging task due to the computational limitations determined by the broad range of overlapping fluid dynamic and chemical scales that characterize actual systems and that lead to *stiff* systems during the numerical solution of the transport equations. This becomes particularly relevant for systems requiring an accurate description of finite-rate chemistry effects, considering that detailed combustion mechanisms for fuel as simple as methane involve 53 species and 325 reactions and that the number of species and reactions dramatically increases with the molecular weight of the hydrocarbon fuel [4].

Overview of the thesis

Aim of this thesis is introducing the Manifold Generated from Local Principal Analysis (MG-L-PCA), a new methodology to reduce the computational cost of numerical simulations describing reactive systems, as the ones involved in combustion. This technique is the development of the Principal Component Analysis (PCA), that allows the reduction of the transport equations to be solved, through a linear projection onto a lower-dimensional space; variables to be transported, the Principal Variables (PV), are linear correlations of all the variables describing the data-set.

In the first Chapter, the solver chosen to be coupled with the MG-L-PCA will be presented. It is the laminarSMOKE solver, a new framework for the modeling of laminar flames with detailed gas-phase chemistry. Several test-cases will be proposed and the comparison between computed results and experimental values will be shown.

In the second Chapter, the MG-L-PCA approach will be described, starting from the rigorous mathematical formalism of the PCA that allows to reduce the dimensionality of the data-set, elencating the methods to select the PV and the optimal criteria to obtain all the necessary information from the preprocessing step and, eventually, showing how to couple the MG-L-PCA with the solver.

In the third Chapter, two data-sets provided by the simulation of a hydrogen flame will be considered and results obtained by the employ of the MG-L-PCA to reconstruct the two data-sets, showing the reconstruction error for the state variables and source terms, will be displayed.

In the end, conclusions of the present work will be shown.

Chapter 1

Laminar coflow flames

1.1 Introduction

Laminar flames have been widely studied to improve the design and the optimization of industrial and domestic equipment (i.e., furnaces, domestic gas burners, etc.) and for a better understanding and modeling of more complex flows (e.g., turbulent flames). However, modeling of multi-dimensional laminar flows with realistic chemical mechanisms places several demands on computational resources, mainly because of the large number of chemical species involved, the high stiffness of the governing equations and the presence of high gradient regions (especially close to the flame front) [7]. Therefore, a technique able to reduce the computational cost of numerical simulations for these flames would be very appealing and the MG-L-PCA techinique, that will be described in the Chapter 2, could be a very usefull tool.

Several laminar coflow flames have been simulated with the solver laminarSMOKE, whose validation was previously performed [7]. In particular, a flame feed with hydrogen and nytrogen and three methane flames have been considered. In this Chapter,

the governing equations of the reactive system, the numerical method applied to solve them and the kinetic schemes adopted by the solver will be illustrated. Then, the numerical results will be compared with the ones get by experimental measurements.

1.2 OpenFOAM

OpenFOAM® (Open Field Operation And Manipulation) is a free, open source CFD software package and it is the chosen CFD tool for my thesys. The standard OpenFOAM® solver for compressible, unsteady, non-reacting flows (pisoFoam) was modified in order to make possible the introduction of detailed kinetic mechanisms: the result is a new framework, called laminarSMOKE, for the numerical modeling of laminar reactive flows.

1.2.1 Governing equations

The reactive, laminar flows under investigations in the present thesys are mathematically described by the conservative equations for continuous, multicomponent, compressible, thermally-perfect mixtures of gases [7] The conservation equations of total mass, mixture momentum, individual species mass fractions and mixture energy, are reported in the following (assuming a Newtonian fluid):

$$\frac{\partial \rho}{\partial t} + \nabla \left(\rho \mathbf{u} \right) = 0 \tag{1.1}$$

$$\frac{\partial}{\partial t} \left(\rho \mathbf{u} \right) + \nabla \left(\rho \mathbf{u}\mathbf{u} + p\mathbf{I} \right) = \nabla \mathbf{T} + \rho \mathbf{g} \tag{1.2}$$

$$\frac{\partial}{\partial t}\left(\rho Y_k\right) + \left(\rho Y_k \mathbf{u}\right) = -\nabla\left(\rho Y_k \mathbf{u}\right) + \dot{Y}_k \qquad k = 1, ..., NC \tag{1.3}$$

$$\rho c_p \frac{\partial T}{\partial t} + \rho c_p \mathbf{u} \nabla T = -\nabla \mathbf{q} - \rho \sum_{k=1}^{NC} c_{p,k} Y_k \mathbf{u_k} - \sum_{k=1}^{NC} h_k \dot{Y}_k \tag{1.4}$$

where t is the time, ρ is the mixture density, p the pressure, $\mathbf{u}$ the mixture velocity vector, $\mathbf{T}$ the fluid stress tensor, $\mathbf{g}$ the acceleration vector due to gravity, Y_k the mass fraction of the species k, $\mathbf{u}_k$ is the diffusion velocity velocity of species k, $\dot{Y}_k$ the formation rate of species k, T the temperature, c_p and $c_{p,k}$ are respectively the specific heat at costant pressure of the mixture and of the individual species k, $\mathbf{q}$ the heat flux vector and h_k the individual species enthalpy.

The density of the mixture is calculated using the equation of state of ideal gases. The heat flux vector accounts for conduction and radiation:

$$\mathbf{q} = -\lambda \nabla T + \mathbf{q}_{rad} \tag{1.5}$$

where λ is the mixture thermal conductivity and $\mathbf{q}_{rad}$ the radiative heat flux. Both Fickian and thermal diffusion are taken into account for evaluating the diffusion velocities:

$$\mathbf{u_k} = -\frac{\mathscr{D}_k}{Y_k}\nabla Y_k - \frac{\mathscr{D}_k \Theta_k}{X_k}\frac{1}{T}\nabla T \tag{1.6}$$

where $\mathscr{D}_k$ is the individual species mixture averaged diffusion coefficient, X_k the mole fraction and Θ_k the thermal diffusion ratio of species k. An optically thin radiation model is included in the calculations. If self-absorption of radiation is neglected, the divergence of the net radiative flux can be written as:

$$\nabla \mathbf{q}_{rad} = -4\sigma a_p \left(T^4 - T_{env}^4\right) \tag{1.7}$$

where σ is the Stefan-Boltzmann constant, T_{env} the environmental temperature. Considering as the only radiating species H_2O, CO, CO_2 and CH_4, the Planck mean absorption coefficient, a_p, is evaluated according to the following expression:

$$a_p = p_{H_2O} a_{p,H_2O} + p_{CO_2} a_{p,CO_2} + p_{CO} a_{p,CO} + p_{CH_4} a_{p,CH_4} \tag{1.8}$$

where p_k is the partial pressure of species k and $a_{p,k}$ is the extinction coefficient of species k.

1.2.2 Numerical method: Strang splitting scheme

In order to solve the gas-phase transport equations of continuity, momentum, energy and species mass fractions, the OpenFOAM® framework has been employed. In particular, the standard OpenFOAM® solver for compressible, unsteady, nonreacting flows (pisoFoam) was modified to take into account a reacting system.

Segregated algorithms based on operator-splitting methods were used for the resolution of the system of partial differential equations, thus separating the stiff chemical reaction processes from the non-stiff transport processes.

As described in [6], for a general transport/reaction system like a laminar flame described by a set of partial differential equations (PDEs), the governing PDEs can be transformed into a set of ODEs by the spatial discretization and the application of the method of lines:

$$\frac{d\mathbf{\Psi}}{dt} = \mathbf{S}\left(\mathbf{\Psi}\right) + \mathbf{M}\left(\mathbf{\Psi},\, t\right) \tag{1.9}$$

where t is the time,$\mathbf{\Psi}$ the dependent (or orimary) variables (mass fractions and enthalpy), $\mathbf{S}\left(\mathbf{\Psi}\right)$ is the rate of change of $\mathbf{\Psi}$ due to the chemical reactions and $\mathbf{M}\left(\mathbf{\Psi}, t\right)$

the rate of change of $\mathbf{\Psi}$ due to transport processes, such as diffusion, convection, heat loss, inflow/outflow, etc. In order to solve Eq.1.9 numerically, the time is dicretized in increments $\triangle t$, and the integration time is performed using the Strang splitting scheme. According to this apprach, reaction is separated from the transport process and the numerical integration is performed in three sub-steps:

1. The reaction terms are integrated over a time interval $\frac{\triangle t}{2}$ through the solution of an ODE system:

$$\frac{d\mathbf{\Psi}^a}{dt} = \mathbf{S}\left(\mathbf{\Psi}^a\right) \tag{1.10}$$

   The initial condition $\mathbf{\Psi}^a(0)$ is taken equal to the final state $\mathbf{\Psi}$ from the previous time step and the solution of Eq.1.10 is indicated as $\mathbf{\Psi}^a\left(\frac{\triangle t}{2}\right)$.

2. The transport terms (convection and diffusion) are integrated over a time interval $\triangle t$ by solving:

$$\frac{d\mathbf{\Psi}^b}{dt} = \mathbf{M}\left(\mathbf{\Psi}^b,\, t\right) \tag{1.11}$$

   The initial condition $\mathbf{\Psi}^b(0)$ corresponds to the final state of the system from Sub-step 1, $\mathbf{\Psi}^a(\frac{\triangle t}{2})$, and the solution of Eq.1.11 is denotated by $\mathbf{\Psi}^b(\triangle t)$.

3. This step is identical to Sub-step 1, with the exception that the initial condition corresponds to the final state of the system from Sub-step 2, $\mathbf{\Psi}^b(\triangle t)$. The solution is used as the initial condition for the next time step.

The reaction operator $\mathbf{S}\left(\mathbf{\Psi}\right)$ is indipendent of time and does not involve any discretization operation in space, which means that reaction process is local (separate

for each grid point). On the contrary, the transport operator $\mathbf{M}(\mathbf{\Psi}, t)$ may be dipendent on time if boundary conditions are time dipendent; moreover $\mathbf{M}(\mathbf{\Psi}, t)$ is not separate for different grid points, which are coupled by the convective term and the Laplacian in the diffusion term. As a consequence, Sub-step 1 and 3 correspond to N indipendent stiff ODE systems, in $NC + 1$ unknowns (species mass fractions and temperatures), where N is the total number of computational cells and NC the number of species. Such ODE systems are conveniently integrated over the requested time step using the BzzOde solver, specifically conceived for very stiff ODE systems arising from the numerical modeling of reactive systems with detailed kinetics. Sub-step 2 correspond to an ODE system of $N \cdot (NC + 1)$ coupled equations. However, as these equations are not stiff (the chemical reactions are considered in Sub-steps 1 and 3), the solution is performed in a segregated approach: instead of solving the whole ODE system, $NC + 1$ ODE systems are solved, each having dimension equal to N. The OpenFOAM® framework is used to manage the spatial discretization of transport terms and to solve the Eq.1.11 using the implicit Euler method. The linear systems involved in this process are solved through iterative techniques, i.e. reducing the equation residual over a succession of solutions.

The splitting procedure described above is applied only to species and enegy equations. The continuity and momentum equations are solved in a segregated manner using the well known PISO approach, already available in OpenFOAM®. The whole numerical procedure is summerized in Figure 1.1. The time-step is automatically adapted to keep the Courant number sufficiently low in every computational cell, in order to avoid stability issues.

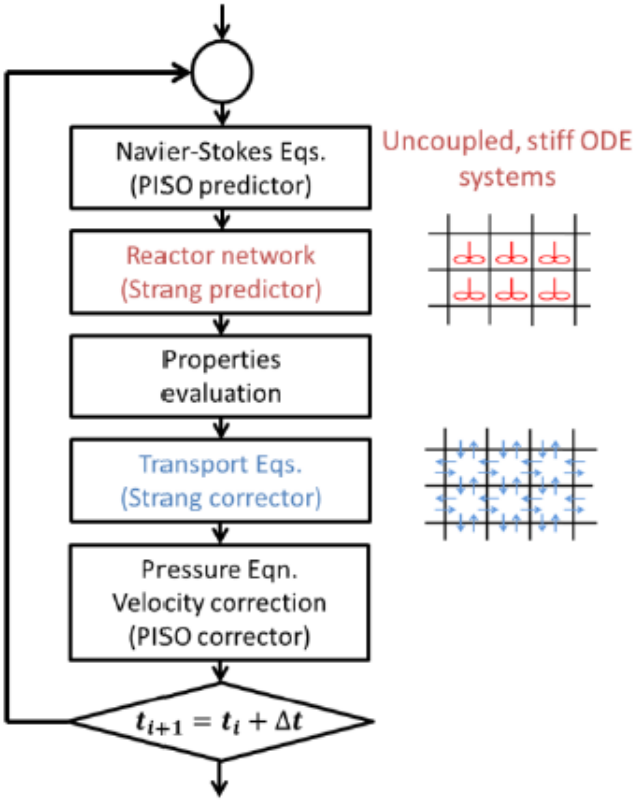

Figure 1.1: Numerical algorithm used in laminarSMOKE

1.2.3 Kinetic Schemes

In order to simulate the coflow laminar flames, two detailed kinetic mechanisms were adopted: the POLIMI_H21201 (11 chemical species, 21 reactions), for the flame feed by hydrogen and nitrogen, and POLIMI_HT1201 (187 chemical species and 6086 reactions) for the methane flame.

As regards the methane flame, a skeletal mechanism was taken into account; in particular, the

PolimiSkeletal_CH4 (23 species, 118 reactions) was applied.

All the kinetic mechanisms are freely available in CHEMKIN format (together with thermodynamic data and transport properties) at http://creckmodeling.chem.polimi.it/.

1.3 H_2/N_2 coflow flames

The laminar coflow flame feed with hydrogen, experimentally and computationally studied by Toro et al. [11], has been numerically modeled through the laminarSMOKE solver. In particular, an axisymmetric H_2/air flame in which a cylindric fuel stream is surrouned by coflowing air was considered, and the numerical results were compared with the experimental values get through two techniques, a spontaneous Raman scattering and Coherent Anti-Stokes Raman Scattering (CARS), described in [11].

In the burner arrangement, an upright stainless tube of 45.5 cm (i.d 0.9 cm), carrying the H_2/N_2 fuel (1:1 mole ractio), is surrounded by an air-coflow annulus (i.d. 9.5 cm) at ambient temperature. The fuel is diluted with nitrogen to reduce heat transfer to the burner; this move the reaction zone downstream so that the gradients in temperature and concentration at the exit of the burner are small. One average fuel exit velocity (50 cm/s) was considered in the present work.

Numerical simulations were performed on a computational domain with a width

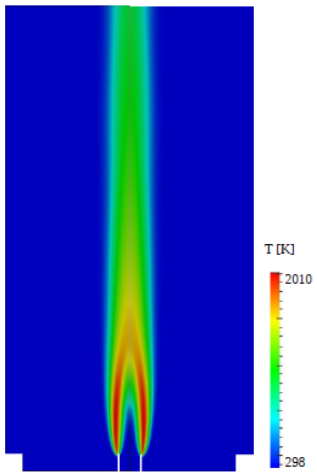

Figure 1.2: H_2 coflow flame: 2D maps of temperature.

of 5.5 cm and length of 20 cm, exploiting the axial simmetry of the flame; the adoopted mesh costists of 76800 cells. The chemistry was described by the the detailed mechanism POLIMI_H21201.

In order to illustrate the overall flame structure, Fig. 1.2 shows the two-dimensional flase-color plot od the computed distribution of temperature, while maps of the computed values of the mass fractions of H_2, O_2, H_2O, and N_2 are displayed in Fig. 1.3 .

In Fig. 1.4-1.9 the radial profiles of temperature and mole fractions of the major species at 3, 10 and 30 mm are presented, to follow the development of the flame. The computed results reproduce the experimental profiles well: the results are nearly quantitative.

Fig. 1.10 shows the axial centerline profiles of temperature. The shape of the profile obtained from experimental measurements is the same of the one get from the numerical simulation and temperature is perfectly predicted until the axial distance of 40 mm, but there are some discrepances between the experimental and the

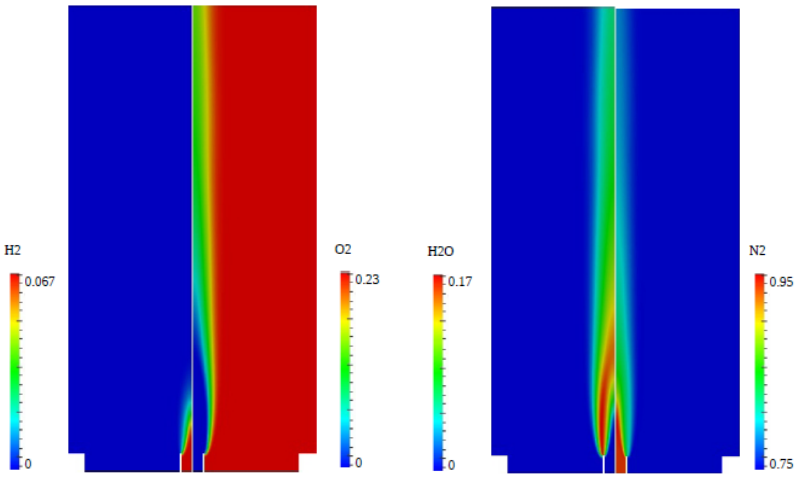

Figure 1.3: H_2 coflow flame: 2D maps of mass fractions of H_2, O_2, H_2O, and N_2.

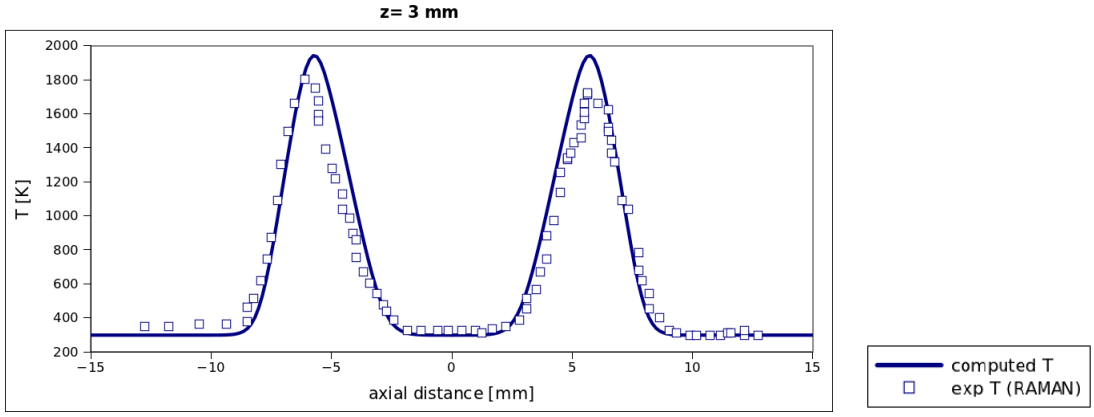

Figure 1.4: Radial temperature at z= 3 mm

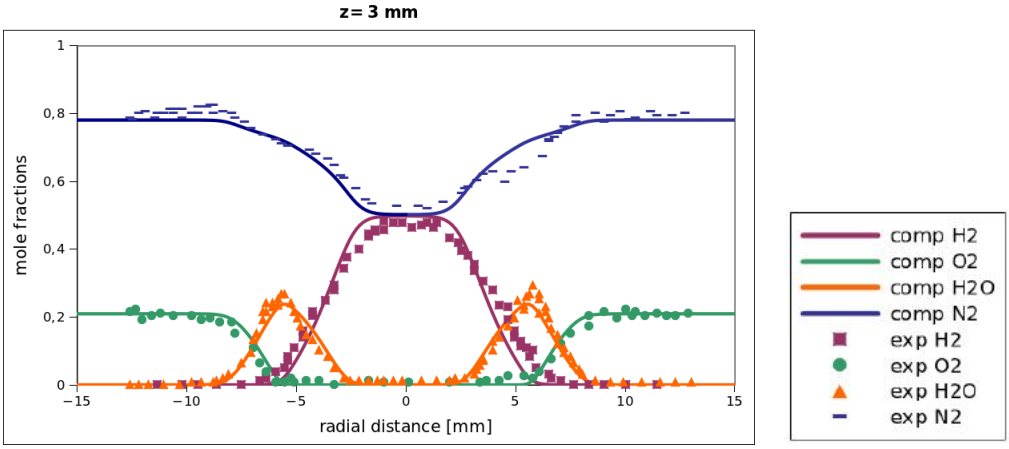

Figure 1.5: Radial mole fractions of H_2, O_2, H_2O, and N_2 at z= 3 mm

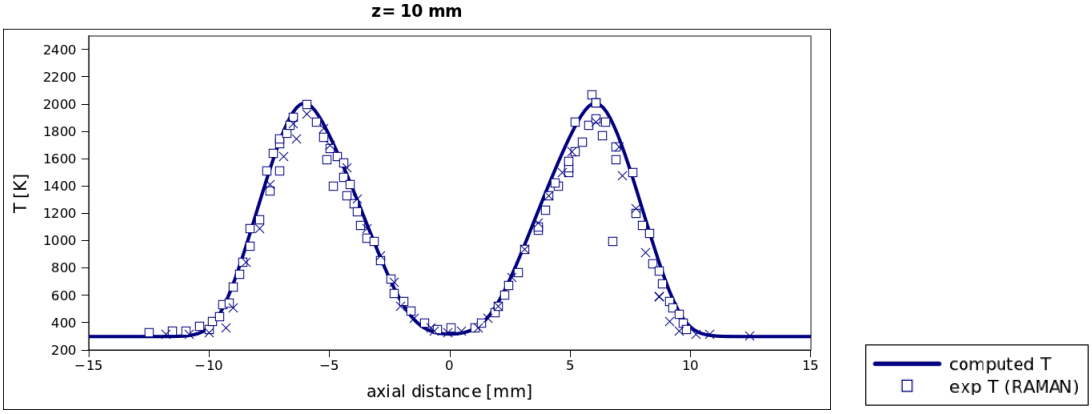

Figure 1.6: Radial temperature at z= 10 mm

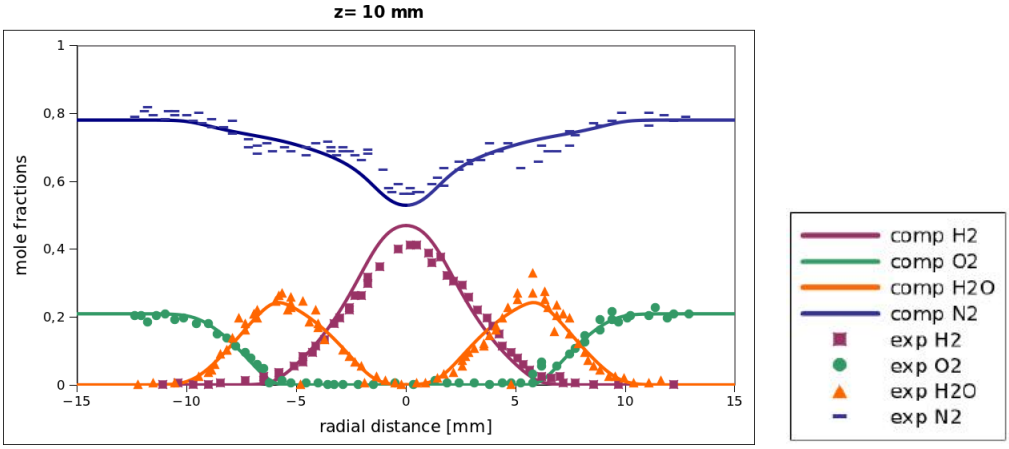

Figure 1.7: Radial mole fractions of H_2, O_2, H_2O, and N_2 at z= 10 mm

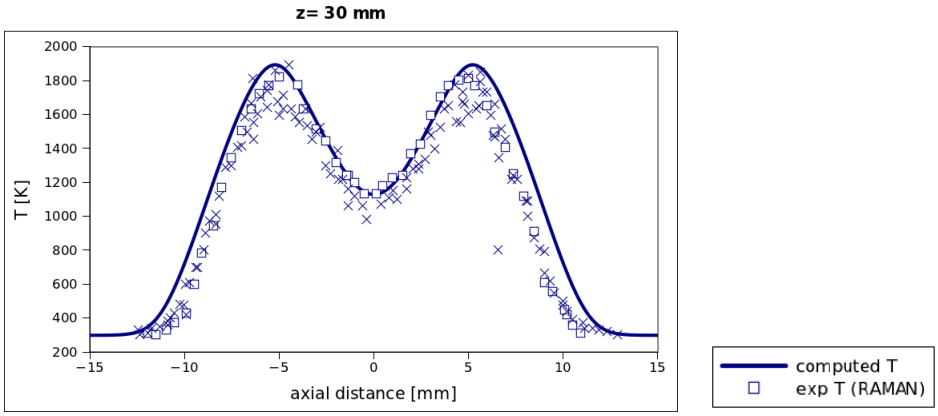

Figure 1.8: Radial temperature at z= 30 mm

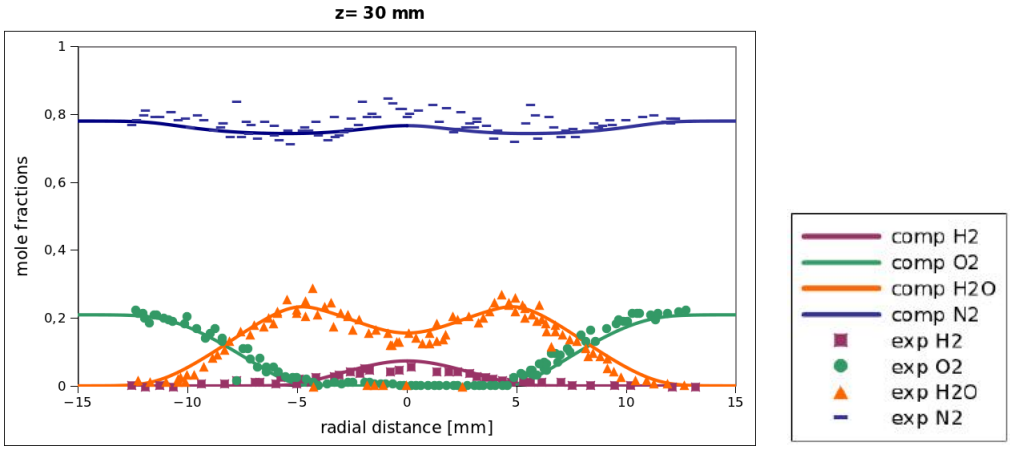

Figure 1.9: Radial mole fractions of H_2, O_2, H_2O, and N_2 at z= 30 mm

computed values starting from the this point. Anyway, the position of the peak of temperature is well predicted and the relative discrepance is never higher than 15%.

In Fig. 1.11, profiles of mole fractions of the major species are displayed. The computed profiles agree quite well with the experimental ones.

1.4 CH_4 coflow flames

Three coflowing laminar methane/air flames, varying in primary equivalence ratio, have been numerically studied. These flames were studied by Bennet et al. [3] and a schematic of the burner utilized is shown in Fig. 1.12. The inner fuel nozzle internal radius is r_I = 0.555 cm and the tube through which it flows has a wall thickness of w_{JET}= 0.080 cm. The coflow inner radius is r_O= 4.76 cm and the inner radius of the cylindrical shield is r_{max}= 5.10 cm. The primary equivalence ratio Φ (primary air flowrate required for complete combustion, divided by the actual primary air flowrate) for each flame appears in Table 1.1. The methane and primary air flowrates in the inner jet are given in Q_{CH_4} and Q_{air} columns, respectively. The coflow stream, reported in Table 1.2, is held fixed for all three cases. The primary air is oxygen-enriched (25% O_2 by volume), whereas the secondary air is regular

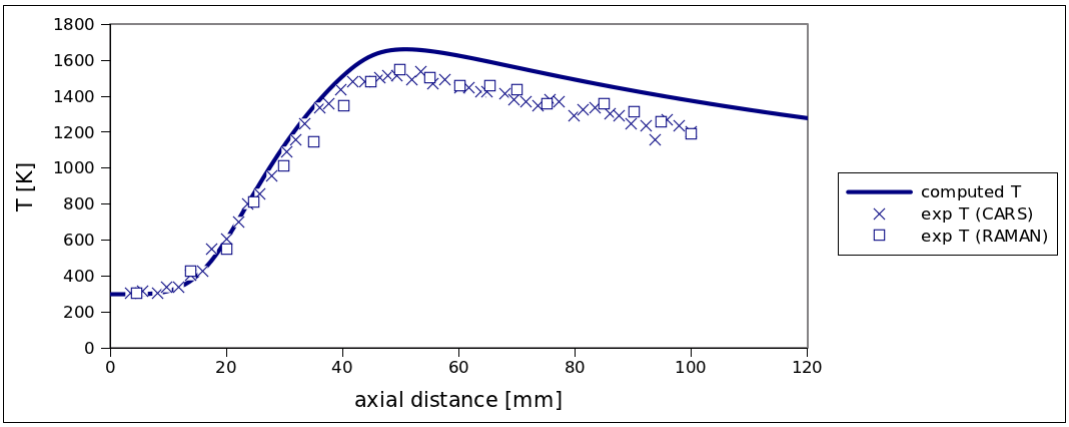

Figure 1.10: Temperature profiles along centerline

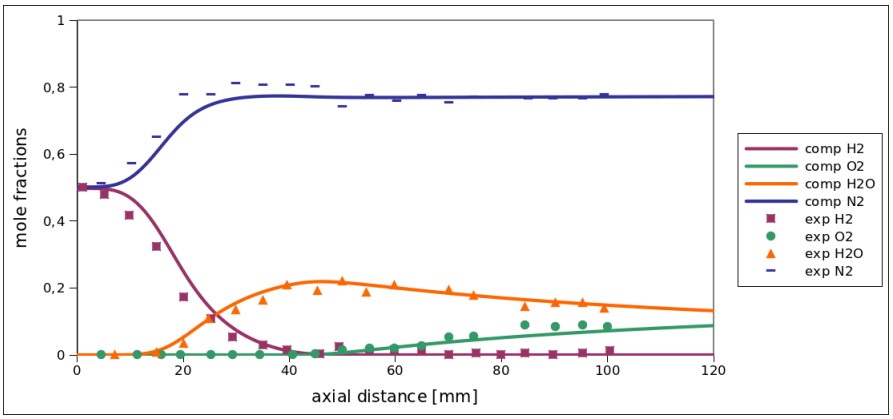

Figure 1.11: Mole fractions profiles of H_2, O_2, H_2O, and N_2 along centerline.

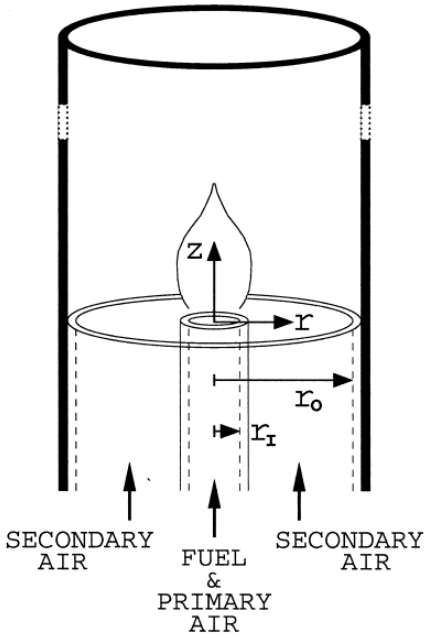

Figure 1.12: Schematic of the burner used for th[3]e tree flames

Φ	Q_{CH_4} $\left[\frac{cm^3}{min}\right]$	Q_{air} $\left[\frac{cm^3}{min}\right]$	v_z $\left[\frac{cm}{s}\right]$	Y_{CH_4}	Y_{O_2}	Y_{N_2}
∞	330	0	5.67	1.00000	0.00000	0.00000
6.160	330	420	12.89	0.30226	0.19627	0.50147
2.464	330	1050	23.71	0.14769	0.23975	0.61256

Table 1.1: Inner jet - Flow rate and composition (mass fractions)

air (20.9% O_2). Both the streams are fed at ambient temperature and atmospheric pressure.

The numerical calculations were performed on a non-equispaced, structured mesh of 23664 cells (with finest spacing in the region immediatly above the burner surface) using the POLIMI_HT1201 mechanism.

Fig. 1.13 reports an example of the two-dimensional maps of temperature amd mass fractions of CH_4, O_2, CO_2, H_2O, OH and C_2H_2 for the not premixed flame, to better show the main structural properties of the investigated flames.

Q_{CH_4} $\left[\frac{cm^3}{min}\right]$	Q_{air} $\left[\frac{cm^3}{min}\right]$	v_z $\left[\frac{cm}{s}\right]$	Y_{CH_4}	Y_{O_2}	Y_{N_2}
0	44000	10.48	0	0.23200	0.76800

Table 1.2: Outer jet - Flow rate and compositions (mass fractions)

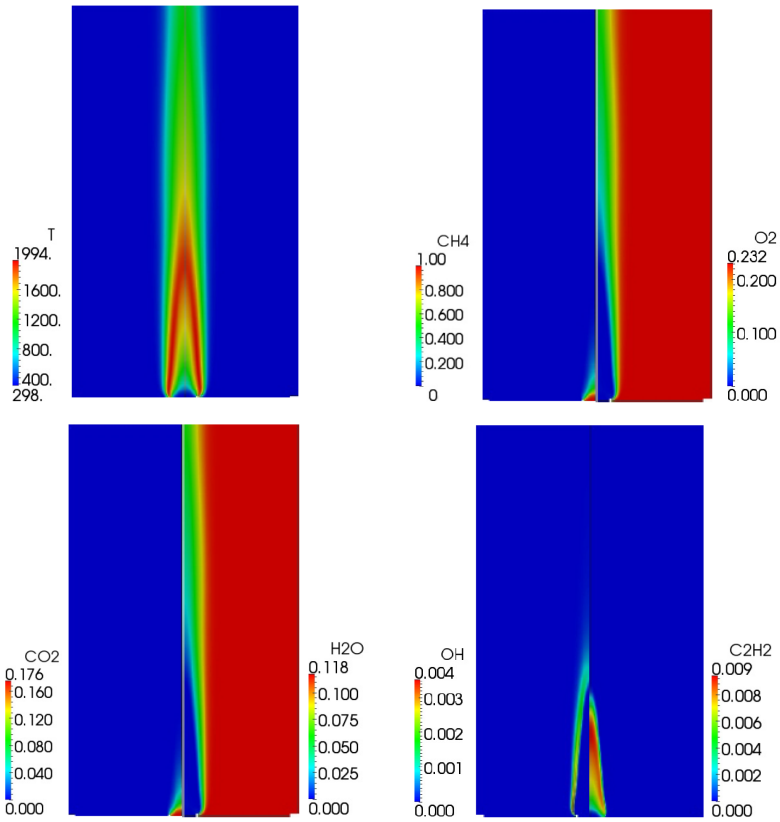

Figure 1.13: CH_4 coflow flame: 2D maps of temperature and mass fractions of CH_4, O_2, CO_2, H_2O, OH and C_2H_2 for the not premixed flame.

Φ	H_T [cm]	
	Exp.	Comp.
∞	5.70	5.83
6.160	4.90	5.13
2.464	3.80	3.72

Table 1.3: Experimental and computed flame heights

1.4.0.1 Flame Heights

The more partial premixing occurs, the smaller is the amount of the secondary oxygen that must diffuse inward to create a stoichiometric mixture, and thus the smaller is the axial distance required for this necessary diffusion to occur. This argument implies that the flame height will shrink with increased premixing, when the CH_4 flowrate is held constant [3]. For each of the three flames studied, Table 1.3 contains the experimental values of the height of the flame (H_T), defined as the axial location where the maximum of temperature occurs, and the computed values. In order to compare experimental results with numerical model predictions, the nondimensionalization of the axial coordinate z through division by H_T was applied.

1.4.0.2 Double flame structure

It has been shown [3] that each flame contains at least one non premixed flame front. In addition, in the partially premixed flames, evidence supporting the existence of an inner premixed flame front abounds. The latter is approximately in corripondence of $\frac{z}{H_T} = 0.5$ and the presence of an inner rich premixed flame and an outer nonpremixed flame is denoted as double flame structure. As partial premixing occurs, heat released decreases at the outer flame front and increases at the inner flame front, showing a "shoulder" near $\frac{z}{H_T} = 0.5$ in profiles of heat release. As the amount of partial

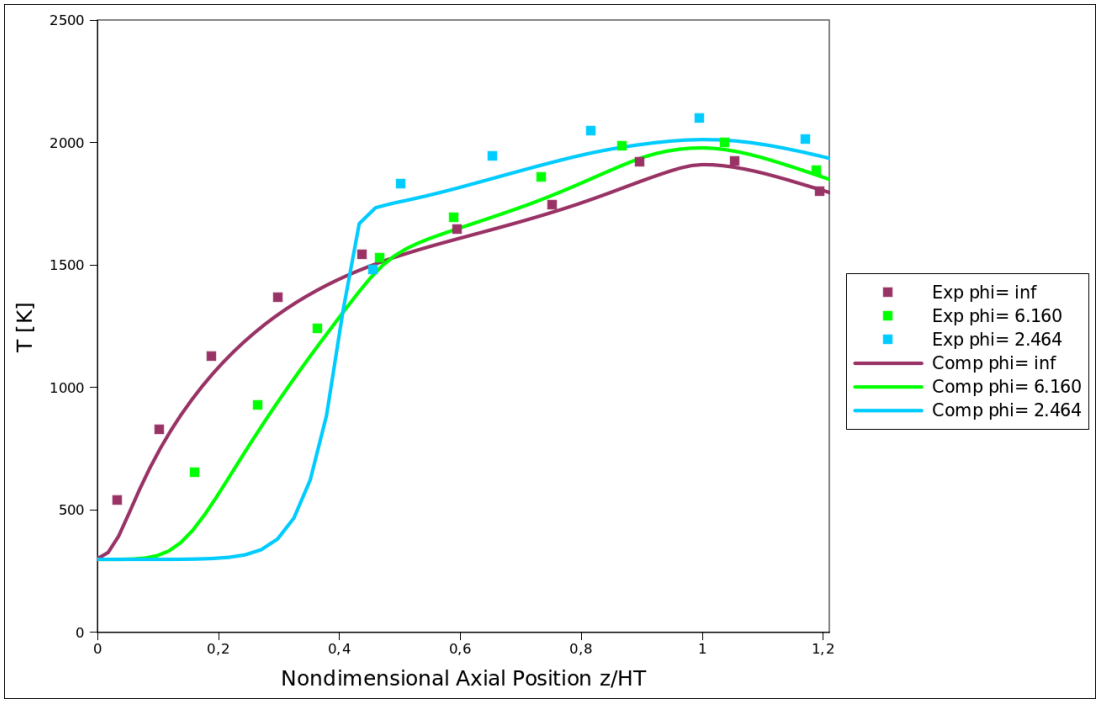

Figure 1.14: Profiles of temperature along the flame centerline. The symbols for experimentals results do not corrispond to specific data points; rather, they help to distinguish among the various plotted curves.

premixing increases, the increasing strenght of the inner flame front is accompanied by other effects, especially the increase of the overall flame temperatures caused by the decrease of radial mass and heat transport (due to the reduction in the flow radially inward).

These effect of premixing will be showed in the results obtained through laminarSMOKE, too.

1.4.0.3 Effect of premixing

Temperature Gas temperature along the centerline of each of the three flames are shown in Fig 1.14. The nondimensional axial position $\frac{z}{H_T}$ is plotted in abscissa.

Thermal boundary conditions in the numeric model sets the inlet temperature to 298 K. This value is chosen primarily because the experimental technique does not allow temperature measurement within a certain distance of the burner, so the true value is not known. This means that the numerical model includes neither heating of the burner lip, nor, consequently, preheating of the reactants. Therefore, the computed steady-state temperature are initially lower than the experimental ones.

As described in [3], three effects of partial premixing are present in both the computational and experimental results.

First, the peak centerline temperature increases from nonpremixed flame to the $\Phi = 2.464$ flame for both the computational and the experimental data: this trend is very well predicted by the model. This is likely caused by the increasing strenght of the inner flame front, leading to an increased heat release near the centerline. Another reason for the increase of the maximum temperature along the centerline may be that the less of the (cooler) coflowing air is entrained as partial premixing occurs.

The second effect of partial premixing present in the temperature profiles is the "shoulder" that forms near $\frac{z}{H_T} = 0.5$ due to heat release at the inner flame front, a feature present in both the computational and experimental results.

The third effect is the reduction in centerline temperatures near the burner surface ($\frac{z}{H_T} \leq 0.4$), an experimentally trend very well predicted by the computational results. This effect is best explained by the decrease of flow radially inward, in the region $\frac{z}{H_T} \leq 0.4$. This latter behavior subsequently reduces the heat transfer from the outer flame front toward the centerline. The same phenomenon also decreases the inward radial transport of chemical species formed off-axis (H_2O, CO_2, CO, etc.), thus delaying the rise of their centerline profiles.

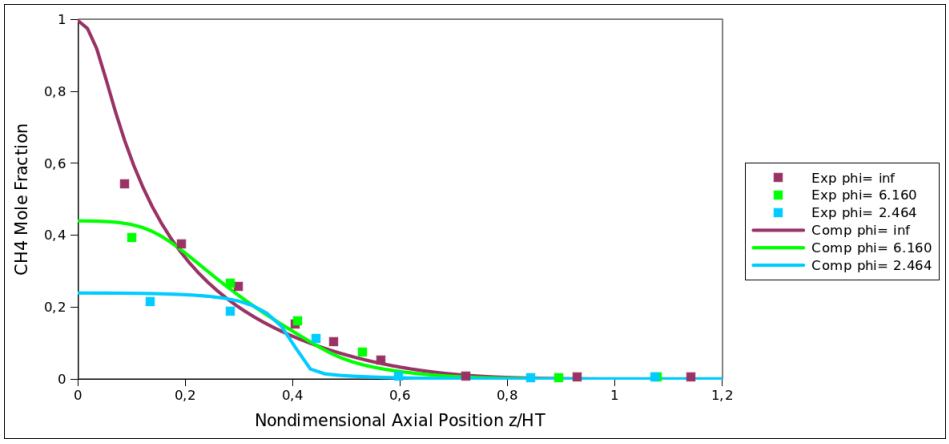

Figure 1.15: Profiles of CH_4 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

CH_4 mole fraction Examinations of the mole fractions of major species begins with those of methane, depicted along the centerline in Fig. 1.15.

Even if some discrepancies can be observed, in general, there is good agreement between computational and experimental results and the values of $\frac{z}{H_T}$ at which the CH_4 disappears are well predicted by the numerical model. Three trends are observed as partially premixing occurs.

First, the initial concentrations of CH_4 decrease, because the CH_4 is being diluted by primary air.

Second, the initial concentrations of CH_4 persist to larger heights above the burner surface, as evidenced by the longer flat region at the start of each profile. This behaviour is consistent with the decrease in coflow entraiment seen above, responsable for decreased dilution of the centerline reactants, and it also indicates a decrease in radial transport. Moreover, the axial component of velocity increases as partial premixing occurs (due to the increased flowrate in the inner jet), further decreasing the relative impact of radial transport.

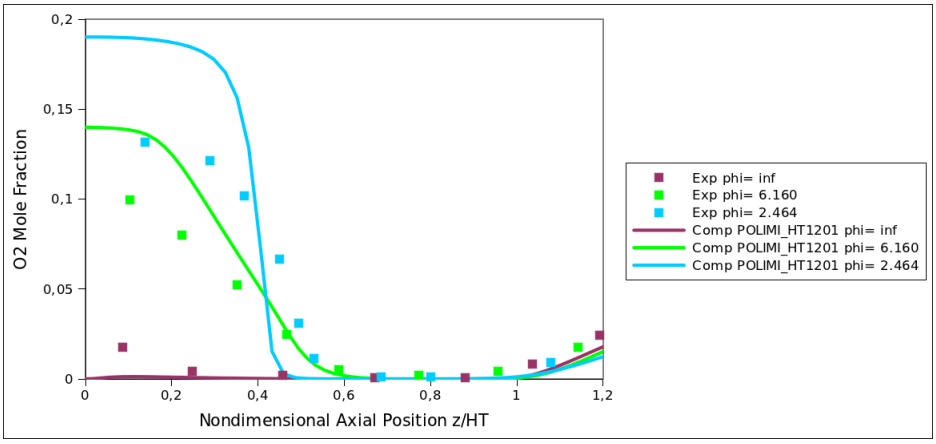

Figure 1.16: Profiles of O_2 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

The third trend apparent in both the compuational and experimental CH_4 centerline data is the increase in the rapidity with wich the concentrations vanish.

Some computational/experimental discrepancies appear near the burner surface, in particular the experimental CH_4 mole fractions for the nonpremixed flame do not appear to match the computed values when extrapolated to the burner surface. This behavior is believed to be caused by argon concentrations that are slightly lower than those assumed (species calibrations depend upon an assumed argon concentration of 1% through the measurement domain) [3].

O_2 mole fraction Oxygen mole fractions along the centerline are illustrated in Fig. 1.16.

As described in the previous paragraph, the argon calibration procedure for the species measurements relies on the assumption of a constant (known) argon concentration. The fact that the absolute experimental O_2 mole fractions near the burner surface are lower than those found numerically indicates that this assumption is

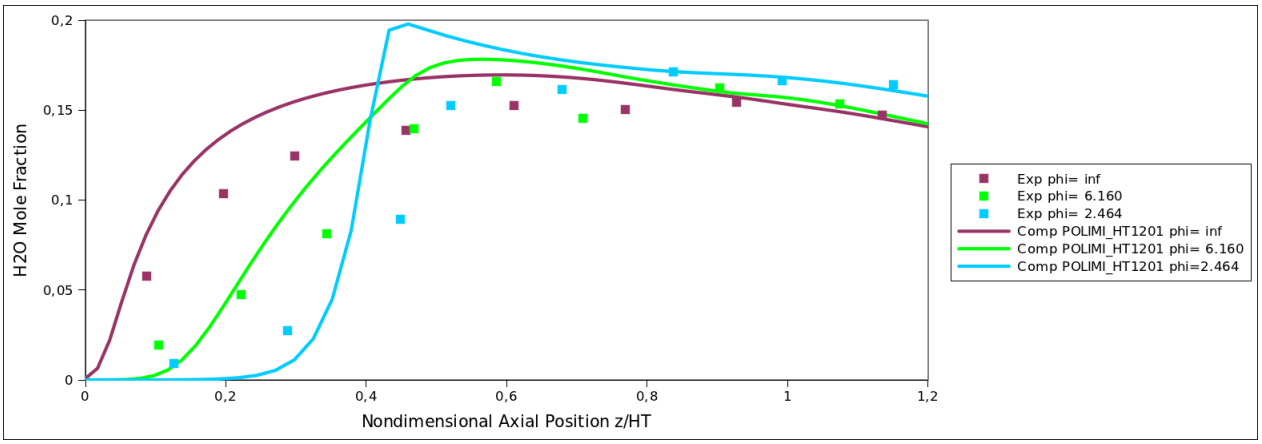

Figure 1.17: Profiles of H_2O mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

probably a poor one near the burner surface [3].

Computational and experimental results are in fairly good agreement at the values of $\frac{z}{H_T}$ at which O_2 first disappears and then reappears.

The compuational and experimental O_2 profiles both show three effects of partial premixing. First, the initial concentrations of O_2 increase, because the air concentration in the inner jet increases.

Second, the initial concentrations persist further into the flame (i.e., the profiles remain flatter further above the burner), which is a consequence of the decreasesed coflow entrainment and decreasedd inward radial transport previously discussed.

Third, O_2 disappears mich more sharply as partial premixing occurs, indicating the increasing strenght of the inner flame front.

H_2O mole fraction Centerline mole fraction of H_2O are displayed in Fig. 1.17.

As the low vapour pressure of H_2O (at nonflame temperature) makes experimental calibration difficult, the experimental values are multiplied by a scale factor

chosen such that the nonpremixed flame's H_2O concentration at $\frac{z}{H_T} = 1$ agrees with the computed value. Thus, no statements can be made regarding the agreement of absolute concentrations, but the good agreement observed in profile shape is meaningful [3].

Both the computational and experimental data show two features resulting from partial premixing.

First, the absence of H_2O at $\frac{z}{H_T} = 0$ persists further above the burner surface as partial premixing increases; in other words, the increase in H_2O concentration occurs later downstream. This behalvior results from decreased radial transport (of H_2O toward the centerline, in this case), as mentioned above in conjuction with related trends in the centerline T profiles.

Second, the increase in concentrations gets sharper as Φ decreases from ∞ to 2.464, in accordance with the strengthening of the inner flame front. After peaking near the the inner flame front, both the computational and experimental profiles are roughly flat for $\frac{z}{H_T} =\gtrsim 0.8$. At $\frac{z}{H_T} = 1$, H_2O concentrations are very close to those at equilibrium [3].

CO_2 mole fraction Mole fractions of CO_2 along the centerline are plotted in Fig. 1.18.

The computational and experimental results several discrepances. The first one is the overprediction of the CO_2 mole fraction in the region with $0.2 < \frac{z}{H_T} < 0.8$ for the not premixed flame and another important one is the underprediction of CO_2 mole fraction in the region with $\frac{z}{H_T} < 0.4$ for the partially premixed flames. However, both the computational and experimental results display some similar features. Indeed both the results show the primary effect of partial premixing, that is the increase of CO_2 (from its initial value of zero) occurs further downstream. The phe-

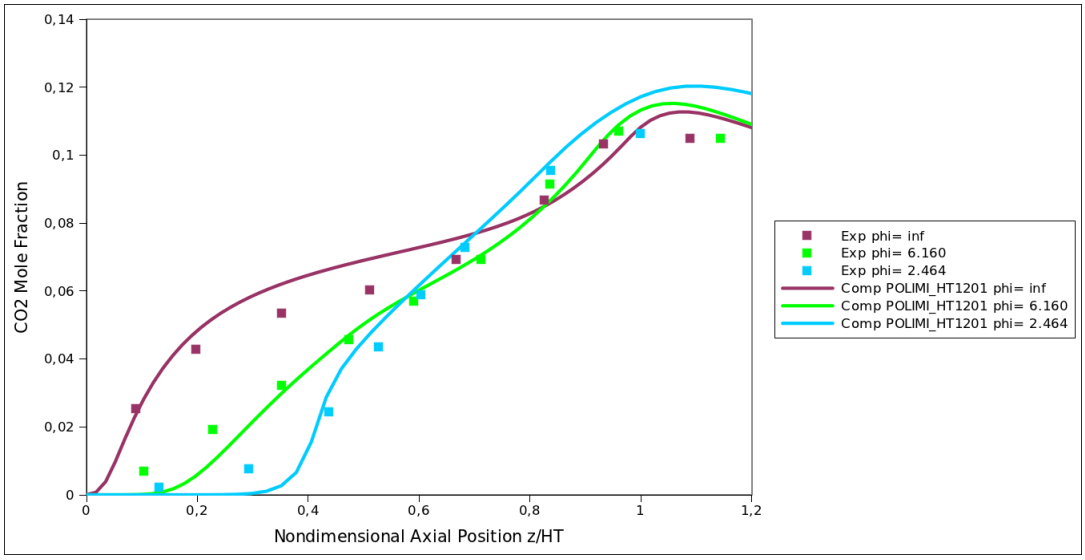

Figure 1.18: Profiles of CO_2 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

nomenon responsable for this effect is the reduced radial transport (of CO_2 toward the centerline, in this case), which similarly affects the other species and T.

Indipendently of partial premixing, the experimental CO_2 datasets peak at about the same concentrations ($X_{CO_2} \approx 0.1$), values very close to equilibrium values [3]. This trend for the peaks is well predicted by the model.

OH mole fraction OH mole fractions along the centerline are illustrated in Fig. 1.19, plotted in log scale in Figure 1.19(b).

Because of the instability of OH at reference conditions makes experimental calibration difficult, the experimental values were multiplied by a scale factor such that $\Phi = 2.464$ flame's maximum OH concentration agrees with the computed values. Thus, no statements can be made regarding the agreement of absolute concentrations, but only qualitative evaluations.

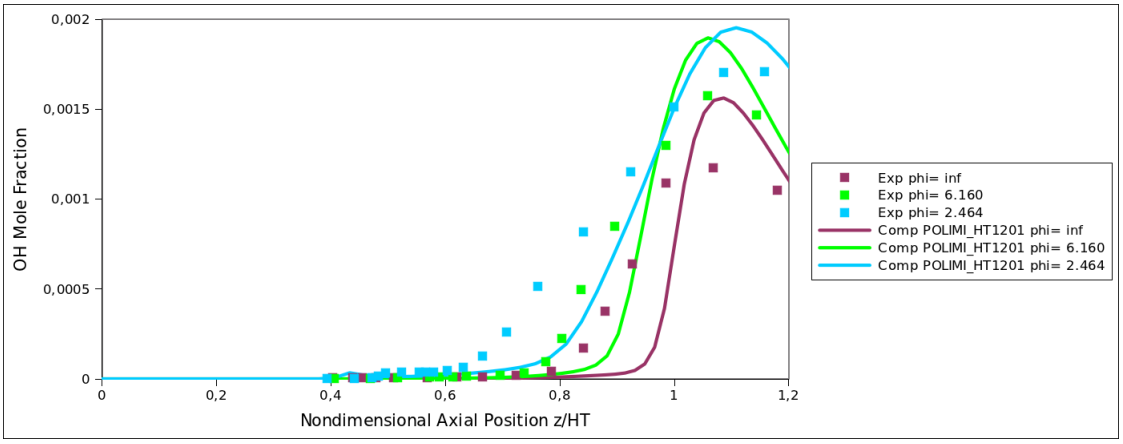

Figure 1.19: Profiles of OH mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

In the region $\frac{z}{H_T} \leq 0.4$ any existing OH will quickly react with CH_4, H_2 and other hydrocarbons, so OH is undetectable and experimental data start from $\frac{z}{H_T} = 0.4$, where it rises in concentration bacause of inward transport from off-axis peaks. Like the experimental results, the predicted peaks of OH concentrations increase with decresing Φ.

C_2H_2 mole fraction Mole fractions of acetylene (C_2H_2) along the centerline are illustrated in Fig. 1.20.

In general, an overprediction of C_2H_2 is evident, but the same trends are present in both computational and experimental results. First, the partial premixing delays the initial increase of C_2H_2 concentrations, a trend along the centerline similar to that observed in the previous paragraphs for T, X_{H_2O} and X_{CO_2}. A secondary effect of premixing present in both datasets is the earlier disappearance of C_2H_2.

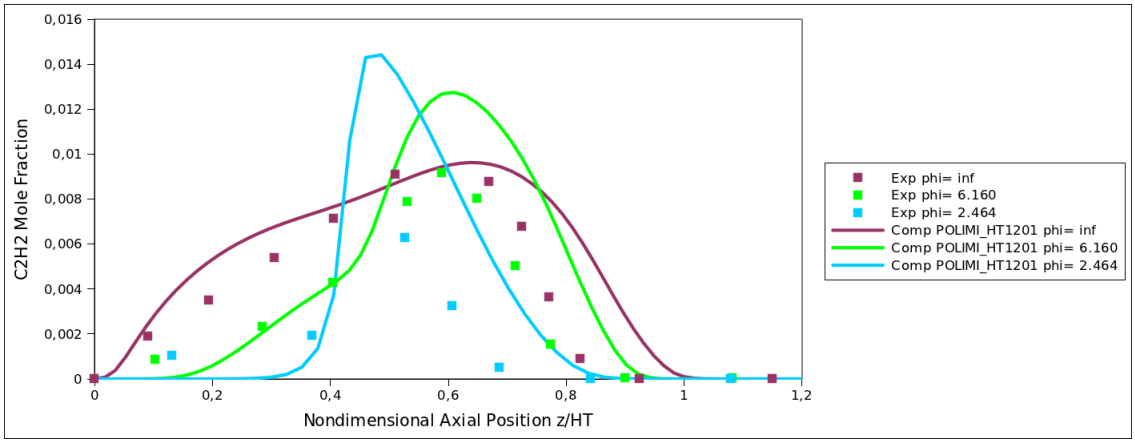

Figure 1.20: Profiles of C_2H_2 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

1.4.0.4 Skeletal mechanism

As mentioned above, a skeletal mechanism was applied before using the detailed one. in particular, the PolimiSkeletal_CH4 was used and a mesh of 1402 cells was adopted to dicretize the domain. All the plots showing the comparison between the results obtained with the two kinetic mechanisms are displayed in the Appendix A, where it's also possible to observe the effect of the mesh on computed results.

In general, the profiles obtained with the detailed mechanism have the same shape of the ones get through the skeletal mechanism and they often overlap with the latter ones. This can be observed in profiles of temperature, methane, oxygen and water.

The detailed mechanism is supposed to better predict the combustion process and, indeed, an improved agreement with experimental results can be observed, for example, in OH and C_2H_2 profiles. As it is shown in Fig. 1.21-1.22 for the three flames, in OH profiles the trends are more similar to the experimental ones and, in

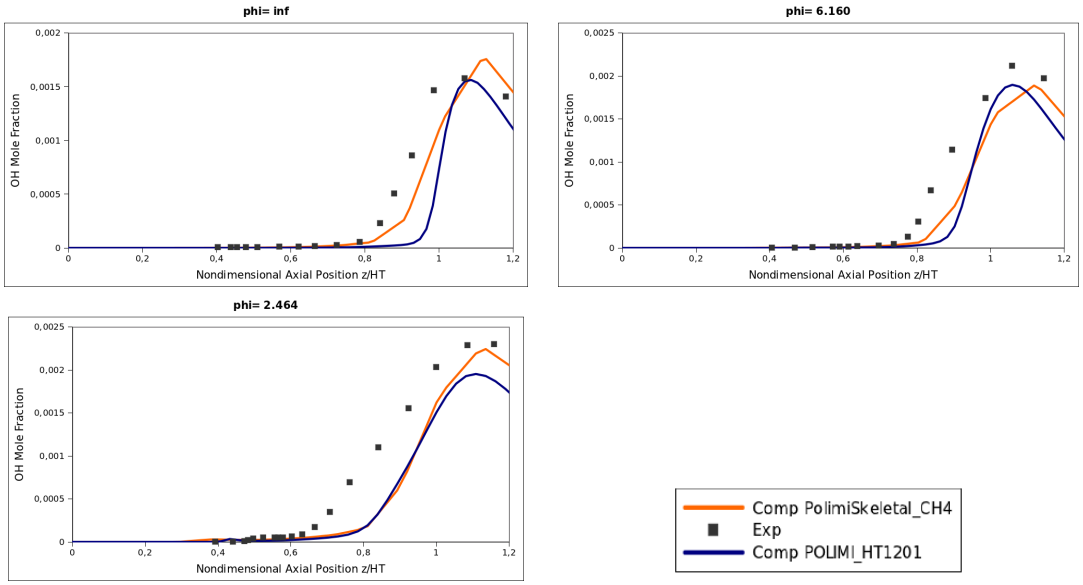

Figure 1.21: Profiles of OH mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

C_2H_2 profiles, the same shapes get through the skelatal mechanism are obtained but the overprediction is considerably reduced, especially for not premixed flame.

Effect of mesh In general, a finer mesh allows an improved prection of variables computed through the numerical simulations. Thus, a better agreement with experimental results is obtained, especially for variables sensitive to the grid.

This can be observed, for example, in the results get with different grids and the same kinetic mechanism (the skeletal mechanism) for the not premixed flame. In Fig. 1.23, temperature profiles obtained with the basic mesh consisting of 1402 cells are compared with the profiles obtained with other two finer ones, respectively composed of 4172 cells and 11329 cells. Only slight differences can be observed between the computational results obtained with different grids, but these differences increase,

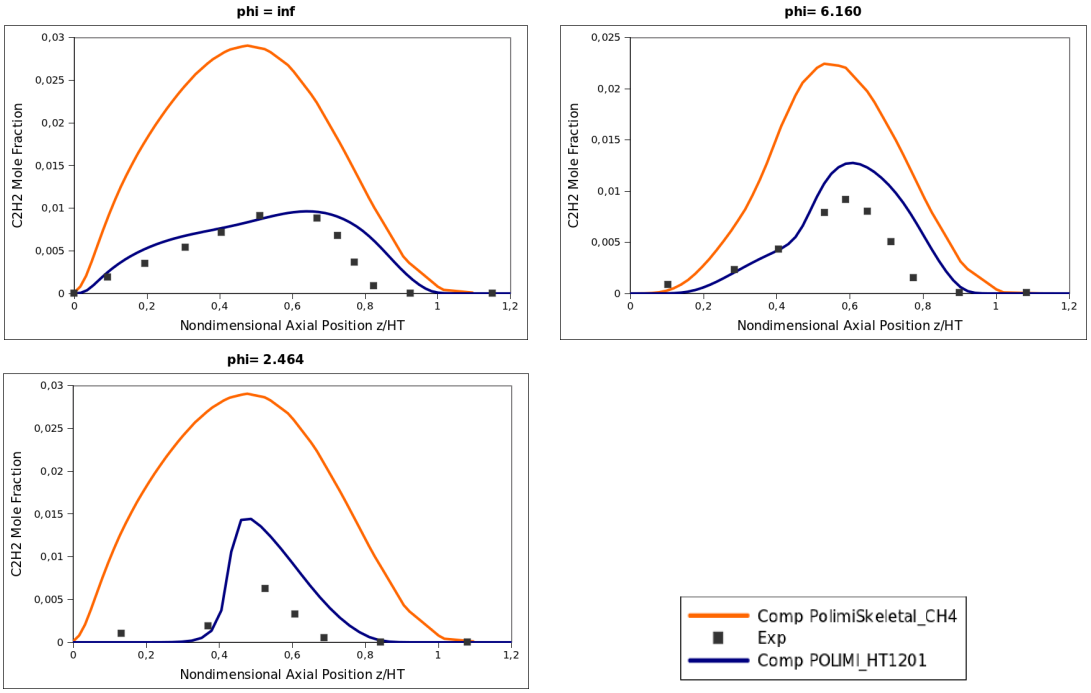

Figure 1.22: Profiles of C_2H_2 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

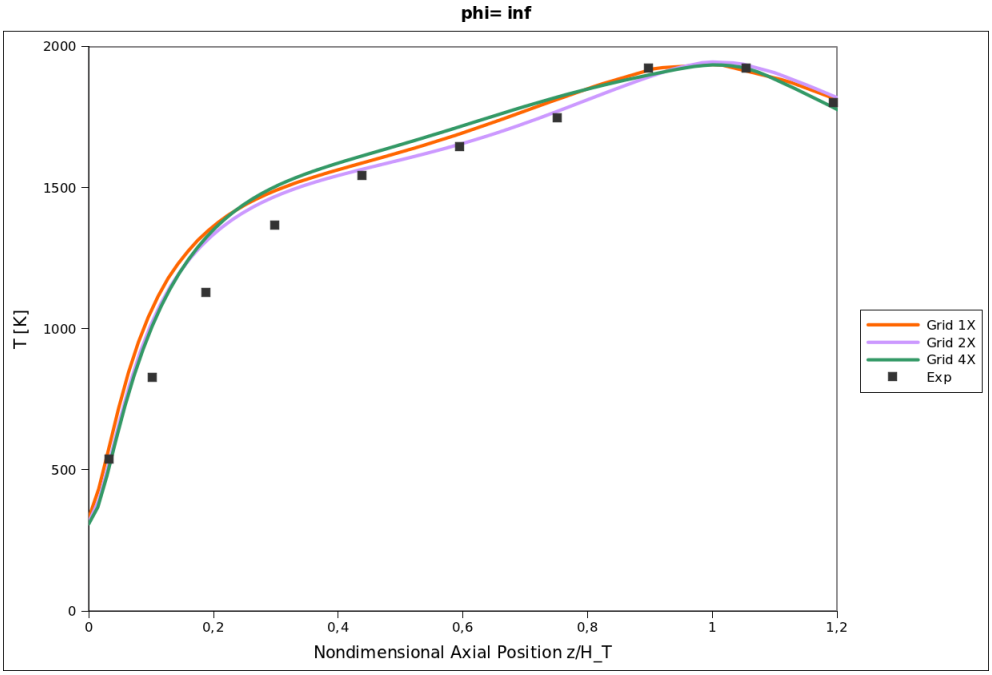

Figure 1.23: Effect of the mesh on temperature, adopting the PolimiSkeletal_CH4 mechanism

for example, with profiles of OH mole fractions, shown in Fig. 1.24, pointing out the importance of a mesh enough fine to simulate the recting flow, as the one adopted with the detailed mechanism and illustred in the previous Section.

1.5 Conclusions

In this Chapter, the solver laminarSMOKE was presented as a new laminar framework for the modeling of laminar gas flames with detailed gas-phase chemistry. The numerical method adopted by the solver is based on the operator-splitting approach to manage large and stiff kinetic schemes. The new code was performed on several laminar coflow flames, employing fine grids and describing the chemistry with de-

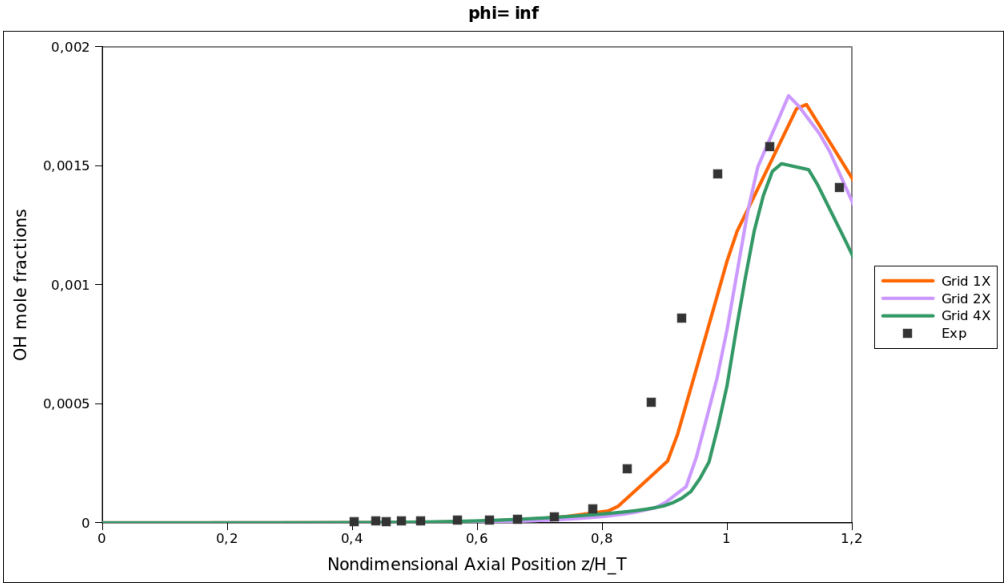

Figure 1.24: Effect of the mesh on OH mole fractions, adopting the PolimiSkeletal_CH4 mechanism

tailed kinetic schemes. The agreement between and experimental measurements was satisfactory for the flame feed with hydrogen as well the methane flames, demonstrating the feasibility and the accuracy of this new framework for the modeling of laminar, reacting flows.

Chapter 2

MG-L-PCA

2.1 Principal Component Analysis (PCA)

Solution of the species transport equations for reacting system can be very computationally intensive if no semplification is made.

As described in [2], the reduction of the number of species equations to be solved can be achieved in two ways:

- Reduction of the kinetic mechanism. This approach is based on the analysis of the dominant reaction rates at the conditions of interest and proceeds through the elimination of species and reactions in the original kinetic mechanism, ultimately leading to a reduced set of species equations to be solved.

- State space parametrization. This approach relies on the assumption that the thermodynamic state of a reacting system relaxes onto a low-dimensional manifold in chemical state space . The thermodynamic state of a single-phase reacting fluid having N_s species is uniquely determinated by $N_s + 1$ parameters (T, P and $N_s - 1$ species mass fractions). However, if a set of "optimal" varibles is identified, the whole thermochemical state can be re-parametrized with a lower number of variables.

This Chapter focuses on the second approach. **Principal Component Analysis (PCA)** provides a rigorous mathematical formalism to reduce the dimensionality of a data set consisting of a large number of correlated variables, retaining most of the variation present in the original data. The reduction is achieved by trasforming the original data set into a new set of variables, called Principal Components (PCs), which are uncorrelated and ordered so that the first few account for most of the variation present in all the original variables. PCA provides an optimal representation of the system based on q optimal variables, the PCs, which are linear combination of the $N_s + 1$ primitive variables T, P and Y_i .

2.1.0.5 Definition and derivation of the PCs

Consider a vector $\mathbf{X}$ of Q random variables, i.e. $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_Q})$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X}$. The (i, j)-th element of $\boldsymbol{\Sigma}$ represents the covariance between i-th and j-th variables of $\mathbf{X}$ if $i \neq j$, or the variance of the j-th element of $\mathbf{X}$ if $i = j$. PCA goal is to find $q\,(<< Q)$ variables, called Pricipal Components (PCs), which preserve most of the information present in the original data. The PCs are linear combinations of the original variables; moreover they are uncorrelated (i.e. orthogonal) and derived so that the variance of the j-th component is maximum [2].

The first PC of $\mathbf{X}$ is defined as the linear combination

$$\mathbf{z}_1 = \mathbf{X}\,\mathbf{a}_1 \tag{2.1}$$

where the vector $\mathbf{a_1}$ is constructed so that $var(\mathbf{z_1}) = \mathbf{a}_1^T\,\boldsymbol{\Sigma}\,\mathbf{a_1}$is maximized, subject to the constraint $\mathbf{a_1^T a_1} = 1$. In order to solve this constrained problem, the standard approach of Lagrange multipliers is adopted and the function

$$\mathbf{a_1^T}\,\boldsymbol{\Sigma}\,\mathbf{a_1} - \lambda\left(\mathbf{a_1^T a_1} - 1\right) \tag{2.2}$$

must be maximized, where λ is a Lagrange multiplier. Differentiating with respect to $\mathbf{a_1}$ and equating to zero, it gives $\boldsymbol{\Sigma}\,\mathbf{a_1} - \lambda\mathbf{a_1} = 0$, that is

$$(\mathbf{S} - \lambda \mathbf{I}_Q)\, \mathbf{a}_1 = 0 \tag{2.3}$$

where $\mathbf{I}_Q$ is the $[Q\, x\, Q]$ identity matrix. Thus λ is an eigenvalue of $\mathbf{\Sigma}$ and $\mathbf{a_1}$ the corresponding eigenvector. The eigenvector $\mathbf{a_1}$ which maximizes the variance of $\mathbf{z_1}$ is the one corresponding to the largest eigenvalue of $\mathbf{\Sigma}$,λ_1, being:

$$\mathbf{a_1^T}\, \mathbf{\Sigma}\, \mathbf{a_1} = \mathbf{a_1^T}\, \lambda\, \mathbf{a_1} = \lambda \mathbf{a_1^T}\, \mathbf{a_1} = \lambda \tag{2.4}$$

The second PC, $\mathbf{z_2} = \mathbf{X}\, \mathbf{a_2}$, maximizes the variance of $var(\mathbf{z_2}) = \mathbf{a_2^T}\, \mathbf{\Sigma}\, \mathbf{a_2}$, subject to the constraints $cov(\mathbf{X}\, \mathbf{a}_1, \mathbf{X}\, \mathbf{a}_2) = 0$ ($\mathbf{z_1}$ and $\mathbf{z}_2$uncorrelated) and $\mathbf{a_2^T a_2} = 1$. As

$$cov(\mathbf{X}\, \mathbf{a_1}, \mathbf{X}\, \mathbf{a_2}) = \mathbf{a_1^T}\, \mathbf{\Sigma}\, \mathbf{a_2} = \mathbf{a_2^T}\, \mathbf{\Sigma}\, \mathbf{a_1} = \lambda_1 \mathbf{a_1^T}\, \mathbf{a_2} = \lambda_1 \mathbf{a_2^T}\, \mathbf{a_1}, \tag{2.5}$$

each of the following equations

$$\mathbf{a_1^T}\, \mathbf{\Sigma}\, \mathbf{a_2} = 0 \qquad \mathbf{a_2^T}\, \mathbf{\Sigma}\, \mathbf{a_1} = 0 \qquad \lambda_1 \mathbf{a_1^T}\, \mathbf{a_2} = 0 \qquad \lambda_1 \mathbf{a_2^T}\, \mathbf{a_1} = 0 \tag{2.6}$$

could be used to specify no correlation between $\mathbf{z_1}$ and $\mathbf{z_2}$. Choosing arbitrarily the last expression in Eq. (2.6), the quantity to be maximized becomes

$$\mathbf{a_2^T}\, \mathbf{\Sigma}\, \mathbf{a_2} - \lambda\big(\mathbf{a_2^T}\, \mathbf{a_2} - 1\big) - \phi \mathbf{a_2^T}\, \mathbf{a_1} \tag{2.7}$$

where λ and ϕ are Lagrange multipliers. Differentiating with respect to $\mathbf{a}_2$ and equaling to zero, the expression (2.7) becomes

$$2\mathbf{\Sigma}\, \mathbf{a_2} - 2\lambda \mathbf{a_2} - \phi \mathbf{a_1} = 0 \tag{2.8}$$

Pre-multiplying by $\mathbf{a}_1^T$, it could be transformed as

$$2\left(\mathbf{a_1^T}\, \mathbf{\Sigma}\, \mathbf{a_2} - \lambda \mathbf{a}_1^T\, \mathbf{a_2}\right) - \phi \mathbf{a_1} \mathbf{a_1}^T = 0 \tag{2.9}$$

which reduces to

$$\phi = 0, \tag{2.10}$$

because

$$\mathbf{a_1^T}\,\mathbf{\Sigma}\,\mathbf{a_2} = \lambda \mathbf{a_1^T}\,\mathbf{a_2} \tag{2.11}$$

due to the constraint of $\mathbf{z_1}$ and $\mathbf{z_2}$ being uncorrelated. Then Eq. (2.9)reduces to

$$\mathbf{\Sigma}\,\mathbf{a_2} - \lambda\,\mathbf{a_2} = 0 \tag{2.12}$$

Therefore, λ is once more an eigenvalue of $\mathbf{\Sigma}$ and $\mathbf{a_2}$ is the corresponding eigenvector. Again, $\mathbf{a_2^T}\,\mathbf{\Sigma}\,\mathbf{a_1} = \lambda$. Assuming that $\mathbf{\Sigma}$ has all different eigenvalues, λ is the second largest eigenvalue of $\mathbf{\Sigma}$, λ_2, and $\mathbf{a_2}$ is the corresponding eigenvector.

In general, the k-th PC of $\mathbf{X}$ is $\mathbf{z_k} = \mathbf{X}\,\mathbf{a_k}$ and $var(\mathbf{z_k}) = \mathbf{a_k^T}\,\mathbf{\Sigma}\,\mathbf{a_k} = \lambda_k$, where λ_k is the k-th largest eigenvalue of $\mathbf{\Sigma}$ and $\mathbf{a_k}$is the corresponding eigenvector.

2.1.0.6 PCA approach

In the subsection (2.1.0.5) the definition and derivation of PCs have been discussed for an infinite population of measures, as described in [2], but in actual cases a random sample of n observations of the Q variables is available, i.e.

$$\mathbf{X} = [\mathbf{x_1}\,\mathbf{x_2}\,...\,\mathbf{x_Q}] = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1Q} \\ x_{21} & x_{22} & ... & x_{2Q} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & ... & x_{nQ} \end{bmatrix} \tag{2.13}$$

so that $\mathbf{X}_i = (x_{i1}, x_{i2}, ..., x_{iQ})$ represents the i-th observation from the data set. Thus, the data available for PCA is a [n x Q] matrix and an unbiased estimator of $\mathbf{\Sigma}$, $\mathbf{S}$[1], is employed [2].

Proceeding like described in the subsection (2.1.0.5), it's possible to construct $\mathbf{Z}$, the [n x Q] matrix of PCs scores, with (i,k)-th element equal to $z_{ik} = \mathbf{X_i a_k}$. Thus $\mathbf{Z}$ and $\mathbf{X}$ are related by

$$\mathbf{Z} = \mathbf{X}\,\mathbf{A} \tag{2.14}$$

[1]The matrix $\mathbf{S}$ represents the approximation of $\mathbf{\Sigma}$ for a finite population, i.e. the random sample consisting of n observations for Q variables

where **A** is the [QxQ] orthogonal matrix whose columns are the eigenvectors of **S**. The covariance matrix can be defined as

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X^T X}. \tag{2.15}$$

As **S** is a symmetric, non-singular matrix, it can be decomposed as

$$\mathbf{S} = \mathbf{A^T\, L\, A} \tag{2.16}$$

where **L** is the [QxQ] diagonal matrix containing the eigenvalues of **S** in descending order $(l_1 > l_2 > ... > l_Q)$.

The linear trasformation given by Eq. (2.14) simply recast the original variables into a set of new uncorrelated variables, whose coordinate axes are described by **A**. Since **A** is orthonormal, that is $\mathbf{A^{-1}} = \mathbf{A^T}$, the original variables can be stated as a function of the PCs as

$$\mathbf{X} = \mathbf{Z\, A^T}. \tag{2.17}$$

Therefore the original variables can be uniquely recovered through **Z**. However, the main goal of PCA is to replace the Q elements of **X** with q PCs (q<<Q), discarding only a small fraction of the variance originally contained in the data. The truncated PCs are defined as

$$\mathbf{Z_q} = \mathbf{X\, A_q}. \tag{2.18}$$

Eq. (2.18) can be inverted to obtain

$$\mathbf{X_q} = \mathbf{Z_q\, A_q^T} \tag{2.19}$$

where the subscrit q in the [n x Q] matrix $\mathbf{X_q}$ indicates that the data-set is reconstructed through q variables. A schematic illustration of the size reduction process carried out by PCA is showed in Figure 2.1.

2.1.0.7 Data preprocessing: centering and scaling

Data are usually centered before PCA is carried out: the variables means are substracted from the data sample, thus keeping only the relevant variation for analysis.

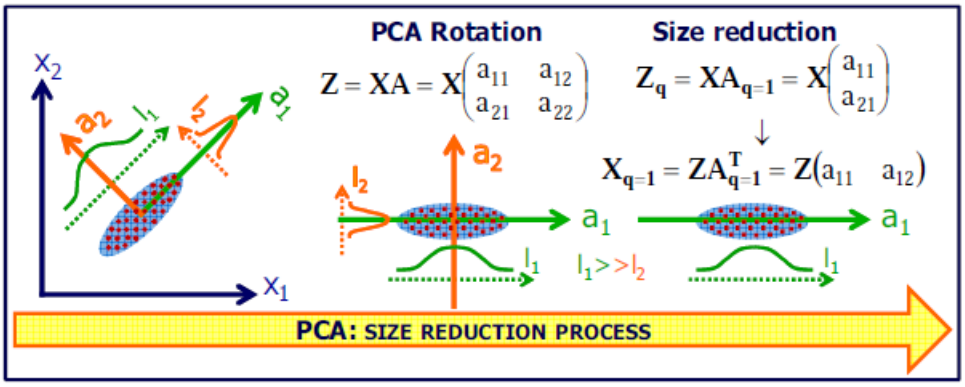

Figure 2.1: Schematic illustration of PCA size reduction process

So centered PCs are obtained. Centering is usually adopted with scaling.

Scaling is an essential operation when the elements of $\mathbf{X}$ are in different units or when they have different variances. The orders of magnitude are very different: temperature may range from ambient conditions to thousand of degrees while species mass fractions vary between zero and one. Moreover, even among species mass fractions there may be need for scaling: for example radicals appear in small concentrations and their mass fractions may range from zero to something far less than one (i.e. $10^{-3} - 10^{-6}$), while major species mass fractions range from 0 to 1. Considering also the centering, it's possible to define a scaled variable, $\widetilde{\mathbf{x}}_\mathbf{j}$, as:

$$\widetilde{\mathbf{x}}_\mathbf{j} = \frac{\mathbf{x_j} - \overline{\mathrm{x}}_\mathrm{j}}{d_j} \tag{2.20}$$

where $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ (j=1,2,..,Q), while d_j is the scaling parameter relative to the variable $\mathbf{x_j}$ [2]. In matrix form, Eq. 2.20 becomes:

$$\widetilde{\mathbf{X}} = (\mathbf{X} - \overline{\mathbf{X}}) \cdot \mathbf{D}^{-1} \tag{2.21}$$

where $\mathbf{D}$ is the diagonal matrix containing the scaling parameters. When scaling is applied, Eq. 2.15-2.18 are modified as:

$$\mathbf{S} = \frac{1}{n-1}\mathbf{D}^{-1}\left(\mathbf{X} - \overline{\mathbf{X}}\right)^T\left(\mathbf{X} - \overline{\mathbf{X}}\right)\mathbf{D}^{-1} \tag{2.22}$$

$$\mathbf{Z} = (\mathbf{X} - \overline{\mathbf{X}})\mathbf{D}^{-1}\mathbf{A} \tag{2.23}$$

$$\mathbf{Z_q} = (\mathbf{X} - \overline{\mathbf{X}})\mathbf{D}^{-1}\mathbf{A_q} \tag{2.24}$$

The choise of scaling parameters has a strong impact on the resulting eigenvectors. The following criteria are available[8]:

- *auto scaling*, which adopts the standard deviation as scaling factor, $d_j = s_j$;
- *pareto scaling*, which adopts the square root of the standard deviation as scaling factor, $d_j = \sqrt{s_j}$;
- *range scaling*, which adopts the difference between the minimal and the maximal variable value as scaling factor, $d_j = max\,(x_j) - min\,(x_j)$;
- *VAST (VAriable STability scaling)*, which focuses on stable variables and adopts the product between the standard deviation and the so-called coefficient of variation, $\frac{s_j}{\overline{x}_j}$, as scaling factor, $d_j = \frac{s_j^2}{\overline{x}_j}$.

2.1.0.8 Criteria to select Principal Variables

Principal Components (PC) are linear combinations of all the variables defining the data-set. However, these varibles are not necessarily equally important to the formation of PC: some of the variables might be critical, but some might be redun-

dant. Therefore, it's reasonable to try to link the PC back to a subset of the original variables, selecting critical variables or eliminating irrelevant ones. There are several methods to choose a subset of q original variables which preserve most of the variation in $\mathbf{X}$ [8].

- *B4 Forward method.* PCA is performed on the original matrix of Q variables and n observation. Thus, the eigenvalues of the covariance matrix are computed and it's possible to compute l^* ($l^* = 0.7\bar{l}$, where $\bar{l}$ is the average egeinvalue size). If q is the number of PV to be choosen and Q_1 is the number of components that have eigenvalues smaller than l^*, the eigenvectors associated with the remaining $Q - Q_1$ eigenvalues are evaluated starting with the first component. The variable associated with the highest eigenvector coefficient is then retained, as it is highly correlated with an important PC. PCA is the repeated until all the components have eigenvalues larger than l^*.

- *B2 Backward method.* PCA is performed on the original matrix of Q variables and n observation. Thus, the eigenvalues of the covariance matrix are computed and it's possible to compute l^* ($l^* = 0.7\bar{l}$, where $\bar{l}$ is the average egeinvalue size). If q is the number of PV to be choosen and Q_1 is the number of components that have eigenvalues larger than l^*, the eigenvectors associated with the remaining $Q-Q_1$ eigenvalues are evaluated starting with the last component. The variable associated with the highest eigenvector coefficient is then discarded, as it is highly correlated with a component not carrying any useful information. PCA is the repeated until all the components have eigenvalues larger than l^*.

- *M2 backward method.* PCA is performed on the original matrix of Q variables and n observation, so the eigenvalues of the covariance matrix are computed. If q is the number of PV to be choosen, the [n x q] matrix of PCs scores, $\mathbf{Z_q}$,

is then evaluated. The goal is to select m ($m < Q$ and $m \geq q$) variables from $\mathbf{X}$ which preserve most of the data variation. In the M2 method, q is the true dimensionality of the data as it is determined with PCA and $\mathbf{Z_q}$ is indicated as the true configuration, whereas the corresponding approximation based on m variables is indicated $\hat{\mathbf{Z}}$. The discrepancy between the two configurations is evaluated with a Procrustes Analysis. The idea is to compare the shape of $\mathbf{Z_q}$ and $\hat{\mathbf{Z}}$, to establish wich set of variables better reproduces the true configuration $\mathbf{Z_q}$, that is the following sum of squared differences between the configurations $\mathbf{Z_q}$ and $\hat{\mathbf{Z}}$ is minimized:

$$M_2 = tr\left(\mathbf{Z_q}^T\mathbf{Z_q} + \hat{\mathbf{Z}}^T\hat{\mathbf{Z}} - 2\mathbf{\Sigma}\right) \tag{2.25}$$

where $\mathbf{\Sigma}$ is the matrix of singular value decomposition (SVD) of $\hat{\mathbf{Z}}^T\mathbf{Z_q}$.

- *McCabe criteria method.* These approaches originate from the observation that the principal components satisfy a certain number of optimality criteria. A subset of the original variables that optimizes one of these criteria is called a set of principal variables by McCabe. Assuming that the set of variables of $\mathbf{X}$ is partitioned into subsets $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$, the covariance matrix of $\mathbf{X}$ can be decomposed as:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \tag{2.26}$$

Then, the partial covariance matrix for $\mathbf{X}_{(2)}$ given $\mathbf{X}_{(1)}$ is $\mathbf{S}_{22,1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$. The criteria proposed by McCabe for the definition of the prin-

cipal variables are:

$$MC1 \quad max\,|\mathbf{S}_{11}| = min\,|\mathbf{S}_{22,1}| = min \prod_{k=1}^{m} \delta_k \tag{2.27}$$

$$MC2 \qquad min\,tr\,(\mathbf{S}_{22,1}) = min \prod_{k=1}^{m} \delta_k \tag{2.28}$$

$$MC3 \qquad min\,\|\mathbf{S}_{22,1}\|^2 = min \prod_{k=1}^{m} \delta_k^2 \tag{2.29}$$

where ρ_k are the eigenvalues of $\mathbf{S}_{22,1}$. The matrix $\mathbf{S}_{22,1}$ represents the information left in the remaining unselected variables and, therefore, it is used to evaluate the optimality criteria to choose the PV. MC1 maximizes the variance taken into account by the subset of variables, while both MC2 and MC3 minimize the reconstruction error.

- *Principal Features.* Adopting this method, the dimension reduction is accomplished by choosing a subset of the original variables that contains most of the essential information, both in in the sense of maximum variability of the variables in the lower dimensional space and in the sense of minimizing the reconstruction error. The rows of the eigenvector matrix, $\mathbf{A_q}$, represent the projection of each variable onto the lower dimensional space, that is, the q elements of each row correspond to the weights of the corresponding variable on each axis of the subspace. The key observation of the method is that variables that are highly correlated or have high mutual information will have similar weight vectors (i.e., similar rows of the eigenvector matrix $\mathbf{A_q}$). On the two extreme sides, two fully correlated variables have identical weight vectors (up

to a change of sign) while two indipendent variables have completely different ones. Therefore, the structure of the rows of $\mathbf{A_q}$ is first analyzed by k-mean algorithm to choose high correlated variables, then a variable is extracted from each subset. The chosen variables represent each group optimally in terms of variability, recontruction accuracy and insensitivity to noise.

2.1.0.9 Transport equations for the PCs

The selection of optimal variable for reacting systems could be very appealing due to the possibility of trasporting in a numerical simulation only few linear combinations of the original variables: the PCs [2].

The mass fraction balance for a reacting chemical species is considered:

$$\frac{\partial \rho Y_k}{\partial t} + \frac{\partial \rho u_j Y_k}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\rho \mathscr{D}_k \frac{\partial Y_k}{\partial x_j}\right) + \dot{\omega}_k \tag{2.30}$$

where ρ is the density, Y_k is the mass fraction of the k-th chemical species, u_j is the velocity component along the direction x_j, $\mathscr{D}_k$ is the material diffusivity for the k-th chemical species and $\dot{\omega}_k$ is the mass reaction rate per unit volume of the k-th chemical species. Introducing the material derivative the equation becomes:

$$\rho \frac{DY_k}{Dt} = \frac{\partial}{\partial x_j}\left(\rho \mathscr{D}_k \frac{\partial Y_k}{\partial x_j}\right) + \dot{\omega}_k \tag{2.31}$$

Considering the Lewis number for the k-th chemical species ($Le_k = \frac{Sc_k}{Pr} = \frac{\lambda}{\rho c_p \mathscr{D}_k}$, where λ is the thermal diffusivity and c_p is the specific heat), it becomes:

$$\rho \frac{DY_k}{Dt} = \frac{\partial}{\partial x_j}\left(\frac{\lambda}{c_p Le_k} \frac{\partial Y_k}{\partial x_j}\right) + \dot{\omega}_k \tag{2.32}$$

If the species mean, $\overline{Y}_k$, is subtracted and a scaling factor, d_k, is applied to the centered variable, the following equation is obtained:

$$\rho \frac{D\left[\frac{\left(Y_k-\overline{Y}_k\right)}{d_k}\right]}{Dt} = \frac{\partial}{\partial x_j}\left[\frac{\lambda}{c_p Le_k}\frac{\partial}{\partial x_j}\left(\frac{Y_k-\overline{Y}_k}{d_k}\right)\right] + \frac{\dot{\omega}_k}{d_k} \tag{2.33}$$

Considering the weight of the k-th variable on the i-th PC, a_{ki}, it turns into:

$$\rho \frac{D\left[\frac{\left(Y_k-\overline{Y}_k\right)}{d_k}a_{ki}\right]}{Dt} = \frac{\partial}{\partial x_j}\left[\frac{\lambda}{c_p Le_k}\frac{\partial}{\partial x_j}\left(\frac{Y_k-\overline{Y}_k}{d_k}a_{ki}\right)\right] + \frac{\dot{\omega}_k a_{ki}}{d_k} \tag{2.34}$$

Summing over all the variables, the obtained equation is:

$$\rho \frac{D\left[\sum_{k=1}^{Q}\frac{\left(Y_k-\overline{Y}_k\right)}{d_k}a_{ki}\right]}{Dt} = \frac{\partial}{\partial x_j}\left[\frac{\lambda}{c_p}\sum_{k=1}^{Q}\frac{1}{Le_k}\frac{\partial}{\partial x_j}\left(\frac{Y_k-\overline{Y}_k}{d_k}a_{ki}\right)\right] + \sum_{k=1}^{Q}\frac{\dot{\omega}_k a_{ki}}{d_k} \tag{2.35}$$

As $\sum_{k=1}^{Q}\frac{\left(Y_k-\overline{Y}_k\right)}{d_k}a_{ki}$ is the definition of the i-th PC, $\mathbf{z}_i$, if the Lewis number is assumed equal for all the species, the Eq. 2.35 can be rewritten as:

$$\rho\frac{D\mathbf{z_i}}{Dt} = \frac{\partial}{\partial x_j}\left(\frac{\lambda}{c_p Le}\frac{\partial \mathbf{z_i}}{\partial x_j}\right) + \dot{\omega}_{\mathbf{z_i}} \tag{2.36}$$

where $\dot{\omega}_{\mathbf{z_i}}$is the source terms for $\mathbf{z}_i$

$$\dot{\omega}_{\mathbf{z_i}} = \sum_{k=1}^{Q}\frac{\dot{\omega}_k a_{ki}}{d_k} \tag{2.37}$$

If temperature is included in PCs, it's possible to show how to include it in transport equations for the PCs. Starting from the transport equation for temperature,

$$\rho c_p \frac{\partial T}{\partial t} + \rho c_p \frac{\partial u_j T}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\lambda \frac{\partial T}{\partial x_j}\right) + \dot{Q}_R \tag{2.38}$$

where $\dot{Q}_R$ is the heat released by reaction, and introducing the material derivative, Eq. 2.38 becomes

$$\rho \frac{DT}{Dt} = \frac{\partial}{\partial x_j}\left(\frac{\lambda}{c_p}\frac{\partial T}{\partial x_j}\right) + \frac{\dot{Q}_R}{c_p} \tag{2.39}$$

If the mean temperature, $\overline{T}$, is subtracted, a scaling factor, d_T, is applied to the centered variable and the weight of temperature on the i-th PC, a_{Ti}, is considered, the following equation is obtained:

$$\rho \frac{D\left[\frac{(T-\overline{T})}{d_T} a_{Ti}\right]}{Dt} = \frac{\partial}{\partial x_j}\left[\frac{\lambda}{c_p}\frac{\partial}{\partial x_j}\left(\frac{T-\overline{T}}{d_T} a_{Ti}\right)\right] + \frac{\dot{Q}_R a_{Ti}}{c_p d_T} \tag{2.40}$$

When all variables are summed over in Eq.2.35, the source terms for $\mathbf{z}_i$ becomes:

$$\dot{\omega}_{\mathbf{z_i}} = \frac{\dot{Q}_R a_{Ti}}{c_p d_T} + \sum_{k=1}^{Q} \frac{\dot{\omega}_k a_{ki}}{d_k} \tag{2.41}$$

It's possible to express Eq.2.36 in a more compact form:

$$\rho \frac{D}{Dt}(\mathbf{Z}) = -\nabla(\mathbf{j_Z}) + \dot{\omega}_{\mathbf{Z}} \tag{2.42}$$

where $\mathbf{j_Z}$ is the mass diffusive flux of $\mathbf{Z}$.

A prerequisite of the previous equation is that the matrix of PCs coefficients, $\mathbf{A}$, is constant. As the PCA analysis is based on the processing of a multitude of observations in both space and time, $\mathbf{A}$ is constant by construction.

2.2 Local PCA

The PCs are uncorrelated by construction but they can be highly statistically dependent, especially when the relationship among the correlated variables is non-linear, as it usually happens for a reacting system. In order to find a more compact description of the data in a low-dimensional space, a new approach was proposed, known as **Local Principal Component Analysis (LPCA)** and largely applied in the field of image processing[9]. According to LPCA, a vector quantization (VQ) algorithm first partitions tha data space into disjoint regions (clusters) and then PCA is performed in each of them, relying on the observation that if the local regions are small enough, the data manifold will not curve much over the extent of the region and a linear model will be a good fit.

For the LPCA to be effective, the VQ algorithm should not be independent of the PCA analysis. Thus, it was introduced a VQ algorithm based on a reconstruction error metric. The global reconstruction error for each observation $\mathbf{X_i}$, from the sample $\mathbf{X}$, can be defined as:

$$\varepsilon\left(\mathbf{X_i}-\overline{\mathbf{X}}^{(\mathrm{k})}\right)=\|\mathbf{X_i}-\mathbf{X_{i,q}}\|=\left\|\mathbf{X_i}-\left(\overline{\mathbf{X}}^{(\mathrm{k})}+\mathbf{Z_{i,q}}\mathbf{A}_{\mathbf{q}}^{\mathrm{T}(\mathrm{k})}\right)\right\| \tag{2.43}$$

where $\overline{\mathbf{X}}^{(\mathrm{k})}$ is the k-th cluster centroid, $\mathbf{X_{i,q}}$ is the rank q approximation of $\mathbf{X_i}$, $\mathbf{Z_{i,q}}$ is the i-th value of the truncated set of PCs, $\mathbf{Z_q}$, and $\mathbf{A}_{\mathbf{q}}^{(\mathrm{k})}$ is the matrix obtained by retaining only the first q eigenvectors of the covariance matrix, $\mathbf{S}^{(\mathrm{k})}$, associated with the k-th cluster. The original LPCA algorithm was modified to include data pre-processing (Section 2.1.0.7) in the quantization scheme and a very stable algorithm was obtained by using a global scaled reconstruction error metric, ε_{GSRE}, defined as [2]:

$$\varepsilon\left(\widetilde{\mathbf{X}}_\mathbf{i}-\widetilde{\mathbf{X}}^{(\mathrm{k})},\,\mathbf{D}\right)=\left\|\widetilde{\mathbf{X}}_\mathbf{i}-\widetilde{\mathbf{X}}_\mathbf{i,q}\right\| \tag{2.44}$$

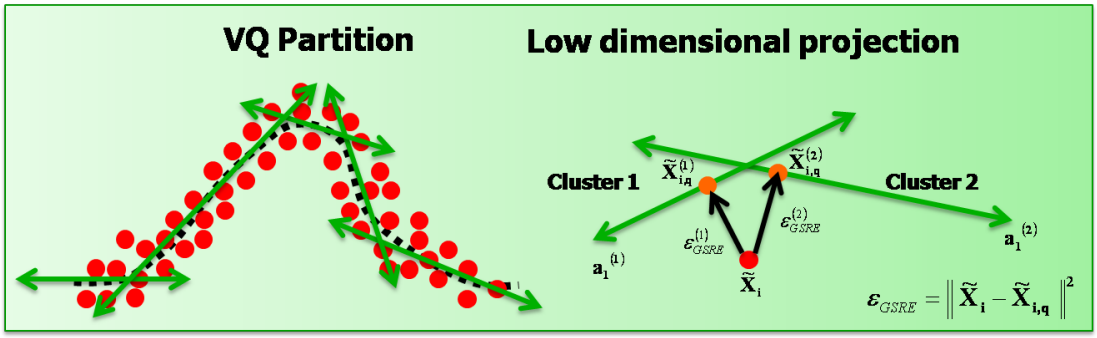

Figure 2.2: Schematic illustration of the VQPCA algorithm [2]

where $\widetilde{\mathbf{X}}_\mathbf{i}$ is the observation of the sample scaled by $\mathbf{D}$, the diagonal matrix whose j-th diagonal element is the scaling factor d_j associated to $\mathbf{x_j}$.

The proposed LPCA algorithm, briefly referred as **VQPCA**, can be summerized as follows:

1. *Initialization*: the cluster centroid, $\overline{\mathbf{X}}^{(\mathrm{k})}$, are randomly chosen from the data set and $\mathbf{S}^{(\mathrm{k})}$ is inialized to the identity matrix for each cluster.

2. *Partition*: each observation from the sample is assigned to a cluster using the metric ε_{GSRE}.

3. *Update*: the clusters' centroids are updated on the basis of partitioning.

4. *Local PCA*: PCA is performed in each disjoint region of the sample.

5. Steps 2-4 are iterated until convergence is reached.

The VQPCA algorithm is illustrated in Figure 2.2.

The reconstruction quality given by VQPCA is measured with respect to the mean variance in the data as

$$\varepsilon_{GSRE,n} = \frac{E\left(\varepsilon_{GSRE}\right)}{E\left[var\left(\widetilde{\mathbf{x}}_\mathbf{j}\right)\right]} \tag{2.45}$$

where E denotes the expectation operator and $\widetilde{\mathbf{x}}_\mathbf{j}$ is the scaled j-th variable from $\widetilde{\mathbf{X}}$. If auto scaling is employed in data preprocessing, Eq. 2.45 reduces to

$$\varepsilon_{GSRE,n} = E\left(\varepsilon_{GSRE}\right) \tag{2.46}$$

In order to obtain the partition of the data into clusters, the optimal approach based on the minimization of the reconstruction error is the VQPCA algorithm, but this approach can result computationally intensive for very large data sets (i.e. data from DNS) [2]. Therefore, it was proposed to split the sample in bins of mixture fraction and to perform the PCA analysis in each cluster. This algorithm, called **FPCA** (as mixture fraction is indicated with the letter F), can be summerized as follows:

1. *Partition*. The data are partitioned into bins of mixture fractions.

2. *Local PCA*. PCA is performed in each bin of the mixture fraction.

The advantage of this approach is twofold: first, the a priori knoledge of the partitioning variable simplifies the LPCA model implementation, moreover the use of mixture fraction is particularly appealing due to its widespread use in the combustion field [9].

A schematic representation of the FPCA method with 2 clusters is showed in Figure 2.3. The data are partitioned into mixture fractions bins, i.e. rich and lean regions, and a one-dimensional coordinate system is identified in each cluster. With respect to the VQPCA method, FPCA allows a very fast clustering, but it is not possible to state a priori that the choice of mixture fractions as conditioning variable is the best choice.

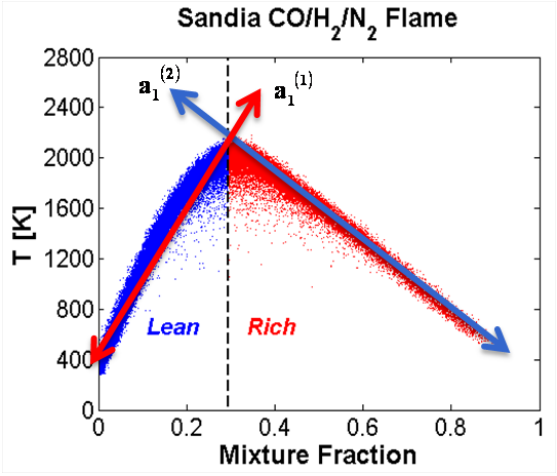

Figure 2.3: Schematic illustration of the FPCA algorithm for a CO/H2 flame [2]

It has been showed [9] that the two local PCA models, VQPCA and FPCA, outperform the global approach, **GPCA**. Indeed, GPCA is unable to provide a compact representation of the data in a low-dimensional space due the highly non-linear relationship existing among the state variables.

2.3 MGPCA

Scores transport equations require the computation of diffusive fluxes and source terms for the PC scores. As the PCs may be linear combination of all state variables (temperature and species mass fractions), the evaluation of source and diffusive terms is not straightforward. For systems showing significant diffusion effects, the evaluation of the diffusive fluxes in the PCA-score approach framework appear particularly challenging. Moreover, it can be shown that the error associated to the PCA reconstruction strongly affects the calculation of source terms, whose accuracy degrades very quickly when reducing the number of parameters defining the manifold. Indeed, the reconstruction of the temperature source term from the approximated

state-space rapidly deteriorates when reducing the number of components, where the same reduction does not affect temperature reconstruction [4].

The main reason for this observed behavior lies in the nature of the score-PCA approach. The scores are linear combinations of centered and scaled variables and this means that the PCA reconstruction error distributes among variables with very different variances, without taking into account the absolute size of the variables. As a consequence, radical species present in very small amounts are affected by reconstruction errors of the same order of magnitude as the major variables, leading to an uncontrolated propagation of error when the source terms are calculated from the approximated state-space [4],[8].

Therefore, a new methodology to employ PCA in the construction of reduced-order combustion models was developed. This model, called **Manifold Generated from PCA (MG-PCA)**, only requires classical transport equations for temperature and species: only a subset of the state varibles is transported while the remaining ones are recovered using the PCA transformation matrix [4].

2.4 MG-L-PCA

Limitations of PCA for the analysis of highly non-linear systems as the reactive ones have been showed (Section 2.2): PCA tries to approximate the non-linear chemical manifold by superimposing several linear effects. To avoid this problem, a local PCA approach was proposed to carry out the partition of the data into clusters. However, as such approach is based on the resolution of transport equations for the scores, the number of PC equations increases with the number of clusters.

To address this problem and the major drawback of PCA related to the source terms described in the previous Section, the current approach is the **Manifold Generated from Local PCA (MG-L-PCA)**, in which the local PCA is applied in

the MG-PCA context [5].

2.4.1 MG-L-PCA method

Considering a $\mathbf{X}$ data-set [n x Q] of n observations of Q variables, in MG-L-PCA method the classic transport equations are solved for a fixed subset of q variables ($q < Q$), in contrast to the PCA discussed in Section 2.1.0.6 where q scores are transported. The domain is divided in clusters according to a proper conditioning variable but the q transported variables are the same in all the clusters. These are the Principal Variables (PV) and they are selected during the pre-processing according to one of the methods described in Section 2.1.0.8.

As PCA, the mean idea of MG-L-PCA is the identification of the q mostly active directions for the data-set $\mathbf{X}$ that allow to approximate the state-space of the system with a reduced number of optimal parameters:

$$\widetilde{\mathbf{X}}_{\mathbf{q_j}} = \mathbf{Z}_{\mathbf{q_j}} \mathbf{A}_{\mathbf{q_j}}^{\mathrm{T}} \tag{2.47}$$

where $j = 1, 2, ..., NCL$ (NCL is the number of all the clusters), $\widetilde{\mathbf{X}}_{\mathbf{q_j}}$ [n_j x q] is the q-dimensional approximation of $\widetilde{\mathbf{X}}_{\mathbf{j}}$ [n_j x Q](the centered and scaled data of the j-th cluster with n_j points), $\mathbf{Z}_{\mathbf{q_j}}[n_j$ x $q]$ is the matrix with the first q principal component scores and $\mathbf{A}_{\mathbf{q_j}}$is the rectangular matrix containing the first q eigenvectors of the covariance matrix $\mathbf{S_j}$ of the data-set of the j-th cluster, $\mathbf{S_j} = \frac{1}{n_j - 1} {\mathbf{X_j}}^T \mathbf{X_j}$.

The MG-L-PCA approach is based on computing the q scores of the Principal Variables in each cluster and recovering the others $Q - q$ variables using the retained PCs. If the q Principal Variables are considered in each cluster ($\widetilde{\mathbf{X}}_{\mathbf{q_j}}$, with $j = 1, 2, ..., NCL$), the corresponding scores $\mathbf{Z}_{\mathbf{q_j}}$ can be approximated through a subset

of the eigenvector matrix $\mathbf{A_{q_j}}$ [Q x q][8]:

$$\mathbf{Z_{q_j}} = \widetilde{\mathbf{X}}_{\mathbf{q_j}} \left(\mathbf{A}^{\mathrm{T}}_{\mathbf{qq_j}} \right)^{-1} \tag{2.48}$$

where $\mathbf{A_{qq_j}}$ is the [q x q] matrix containing only the coefficients related to the q retained variables. Depending on the number of the retained principal variables contributing to the definition of $\mathbf{A_{qq_j}}$, the approximated $\mathbf{Z_{q_j}}$ scores can result in a weak representation of the state space trough Eq.2.47. This can constrain the possible achievable reduction if Eq.2.48 is employed. However, the MG-L-PCA is built a priori, allowing to simply overcome the limitation related to Eq.2.47 by considering the actual first j scores,

$$\mathbf{Z_{q_j}} = \widetilde{\mathbf{X}}_{\mathbf{j}} \mathbf{A_{q_j}} \tag{2.49}$$

and computing an optimized subset of the projection matrix, which also includes the effect of the discarded variables [8]:

$$\mathbf{B_j} = \left(\mathbf{A}^{\mathrm{T}}_{\mathbf{qq_j}} \right)^{\mathbf{-1}}_{opt} = \widetilde{\mathbf{X}}^{-1}_{\mathbf{q_j}} \mathbf{Z_{q_j}} \tag{2.50}$$

Once the $\mathbf{B_j}$ is available, the non-transported thermo-chemical variables can be approximated by

$$\mathbf{X_{Q_j}} = \widetilde{\mathbf{X}}_{\mathbf{q_j}} \mathbf{B_j} \cdot \mathbf{A}^{\mathrm{T}}_{\mathbf{q_j}}. \tag{2.51}$$

The MG-L-PCA approach consists of the two steps [8],[5]: *pre-processing* and *prediction.*

- A database is first generated, using canonical systems with the same chemical composition of the systems to be simulated.

- A classic PCA analysis is performed to identify the q Principal Variables (see Section 2.1.0.8).

- A conditioning variable is chosen: in the context of non-premixed flames it correspond quite well to mixture fraction, while for premixed ones temperature appears to provide an optimal choice [8]. According to the value of this variable each observation will be assigned to the corresponding cluster.

- An iterative algorithm is used to find the minimum number of clusters, NCL, required to obtain a quasi-exact reconstruction of the training dataset (i.e., an R^2 statistics $> 99.99\%$, see Section 2.4.1.1) In particular, the data-set is divided in clusters and, after centering and scaling, PCA is applied to find the approximate variables reconstructed trough the scores of the principal variables and the source terms obtained from these variables.

- The data-set is divided in NCL clusters and a PCA is applied in each one to identify the low-dimensional manifold, represented by all the projection matrices $\mathbf{A_{q_j}}$. The corresponding $\mathbf{B_j}$ matrices are computed too and the centering and scaling factors (see Section 2.1.0.7) used to scale and center the variables in each cluster are stored to be used in the prediction step.

Prediction Transport equations are set to be solved for the q principal variables identified in the pre-processing step.

At the end of each temporal iteration m, the following operations are performed:

- *Conditioning.* According to the conditioning variable, the domain is divided in NCL cluster

- *Centering and scaling.* Variables in each cluster are centered and scaled.
- *Scores computing.* PCs are computed in each cluster:
  $\mathbf{Z}_{\mathbf{q_j}}^{(m)} = \widetilde{\mathbf{X}}_{\mathbf{q_j}}^{(m)} \cdot \mathbf{B_j} \quad [\mathrm{n_j\,x\,q}] = [\mathrm{n_j\,x\,q}][\mathrm{q\,x\,q}]$
- The $Q - q$ missing variables are computed in each cluster:
  $\widetilde{\mathbf{X}}_{\mathbf{Q_j}}^{(m)} = \mathbf{Z}_{\mathbf{q_j}}^{(m)} \cdot \mathbf{A}_{\mathbf{q_j}}^{T} \quad [\mathrm{n_j\,x\,Q}] = [\mathrm{n_j\,x\,q}][\mathrm{q\,x\,Q}]$
  where PV are subsequently replaced by the values of $\widetilde{\mathbf{X}}_{\mathbf{q}}^{(m)}$
- *Uncentering and unscaling.* Variables in each cluster are uncentered and unscaled.
- *Cluster fitting.* To avoid discontinuities in the domain, the clusters are extended by the 10% in each direction and in the overlapping zones a linear interpolation between the reconstruction given by the two clusters are performed.
- Unconditioning. Data are recollected again in a unique matrix.
- *Source terms computing.* Source terms are computed with all the Q variables
  $\mathbf{S}\left(\mathbf{X}_{\mathbf{Q}}^{(m)}\right) = \left[\frac{Q_R}{\rho_{mix} \cdot c_{p_{mix}}}, \frac{R_1}{\rho_1}, \frac{R_2}{\rho_2}, ..., \frac{\mathrm{R_{NC}}}{\rho_{\mathrm{NC}}}\right]$
- *Transport equation solving.* Transport equations are solved for the q principal variables.

2.4.1.1 Constrains of the method

In order to obtain an effective approach, two main conditions must be fullfilled [8]:

- The manifold $\mathbf{A}_{\mathbf{q_j}}$ for each cluster should be identified from high-fidelity data-sets sharing similarities with the systems to be modelled. Ideally, the data-sets should be fast and easy to compute, to generate combustion models of tailored-accuracy with affordable computational resources.

- The principal varibles should allow an exact recontruction of the scores and of the variables defining thermo-chemical state. This is crucial to ensure that the source terms are accurately reconstructed for the reduced set of scalar transport equations. To this purpose, the R^2 statistics is used to quantify how well the species mass fractions and source terms are approximated. For the j-th species or source term, it's possible to compute

$$R_j^2 = 1 - \frac{\sum_{i=1}^{n} \left(\phi_{ij} - \widetilde{\phi}_{ij}\right)^2}{\sum_{i=1}^{n} \left(\phi_{ij} - \overline{\boldsymbol{\phi}}_j\right)^2} \tag{2.52}$$

where ϕ_{ij} refers to the i-th observation of the j-th variable or source term, $\widetilde{\phi}_{ij}$ represents its reconstructed value and $\overline{\boldsymbol{\phi}}_j$ is the average of $\boldsymbol{\phi}_j$.

Chapter 3

MG-L-PCA Results

3.1 MG-L-PCA approach

In Chapter 2, PCA technique was showed and the motivations to adopt an improvement of this method, the MG-L-PCA technique, were explained. In particular, the preprocessing step necessary to include MG-L-PCA in a CFD solver was illustrated and conditions to be fullfilled, in order to achieve a good reconstruction of involved variables, were pointed out. In this Chapter, results obtained in the preprocessing will be displayed, showing the information obtained for the employ of the MG-L-PCA by the solver.

3.2 Preprocessing

As described in Chapter 2, preprocessing starts with the generation of a database, using systems with the same chemical composition of the system to be simulated. In the present work, data get through the laminarSMOKE solver for the flames analyzed

in Chapter 1 were utilized.

3.2.1 Evaluation of the number of PV and the number of clusters.

An iterative algorithm was applied to find the major reduction obtained by the MG-L-PCA technique. In particular, if q is the number of transported variables and NCL is the number of clusters, the MG-L-PCA is applied q x NCL times: every time the dataset is divided in several clusters, variables are centered and scaled, scores are computed in each cluster and the not trasported variables are reconstructed, unscaled and uncentered. After the recollecting of all the data in one single matrix, source terms obtained with the reconstructed variables are compared with the ones get by the original data. Thus, it is possible to compute the R^2 statistics that quantify the reconstruction error for all the state variables and source terms. This procedure is repeated for $q = 1, ..., Q$ (if Q is the number of the variables involved in the simulation) and for a an increasing number of clusters (for example for $NCL = 1, ..., 20$), in order to cover all the combinations offered by several numbers of transported variables and several numbers of clusters. In order to summerize the information get through this analysis, it's possible to report the minimum recontruction error obtained for every iteration in two matrices, one for the minimum R^2 statistics for the reconstruction of state variables and one for the minimum R^2 statistics for the reconstruction of the source terms. To better illustrate these results, it is possible to show these matrices in a graphical form.

When the optimal variables of principal variables and cluster has been established, the MG-L-PCA is applied to get the information necessary for the prediction step that will be performed by the solver, that is the matrices $\mathbf{B_j}$ and

$\mathbf{A_{q_j}}(j = 1, ..., NCL)$, and the centering and scaling factors for each cluster. It should be underlined that the reconstruction error both for the state variables and the source terms must be low, in order to have a correct prediction of involved variables during the numerical simulation.

3.2.2 Results

In the following paragraphs, the summary plots obtained with the hydrogen flame analyzed in Chapter 1 will be displayed. The standard procedure to include MG-L-PCA in numerical simulations starts with the study of one-dimensional flame, in order to find the principal variables to be transported and the preprocessing information previously elencated; these informations will be then adapted for higher-dimensional flames. Therefore, the first reported summery plots will show results get with a 1D flame with the same composition of the mentioned flame. Then, summary plots obtained by the MG-L-PCA technique applied directly to the 2D hydrogen flame will be illustrated.

In all the considerated cases, the criterion chosen to select the principal variables is the B2, that appears as the method that provides the most accurate description on the state space [8]. Moreover, as the flame is not premixed, mixture fraction is chosen as conditioning variable.

H_2 flame (1D) The analyzed flame has the same composition of the one shown in Chapter 1. A detailed kinetic scheme was employed (POLIMI_H21201) to describe the chemistry and MG-L-PCA was performed with a data-set of 1o variables: T, H_2,

Plot showing the minimum R^2 statistics for state variables are displeyed in Fig.

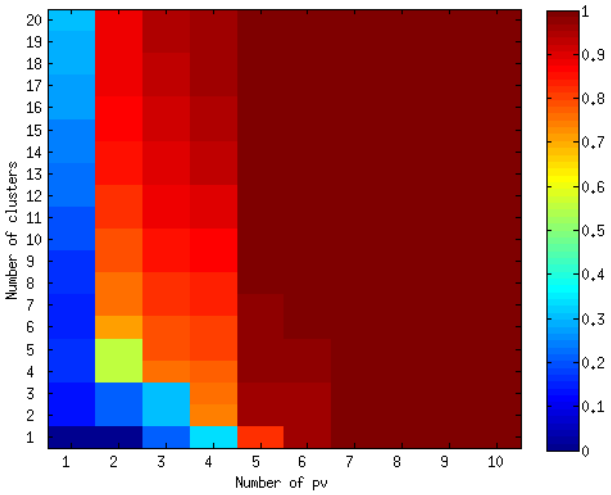

Figure 3.1: H_2 flame (1D): plots showing the minimum R^2 statistics for state variables

3.1, while the plot showing the minimum R^2 statistics for source terms is illustrated in Fig 3.2.

To obtain the previous results, the standard (STD) scaling criterion was employed. Other scaling criteria were utilized to apply the MG-L-PCA technique, but they show lower R^2 statistics, leading to the conclusion that the STD method is the best scaling criterion for these type of flame; Fig. 3.3-3.8 show plots get through the adoption of the other scaling criteria: pareto, VAST, range.

H_2 flame (2D) MG-L-PCA was applied to the same flame considered in the previous paragraph, adopting the same detailed kinetic scheme, but the dataset was generated by the numerical simulation performed on a 2D grid; the numerical results obtained for this flame have been shown in Chapter 1. The minimum R^2 statistics for state variables can be observed in Fig. 3.9, while the minimum R^2 statistics for source terms is reported in Fig. 3.10.

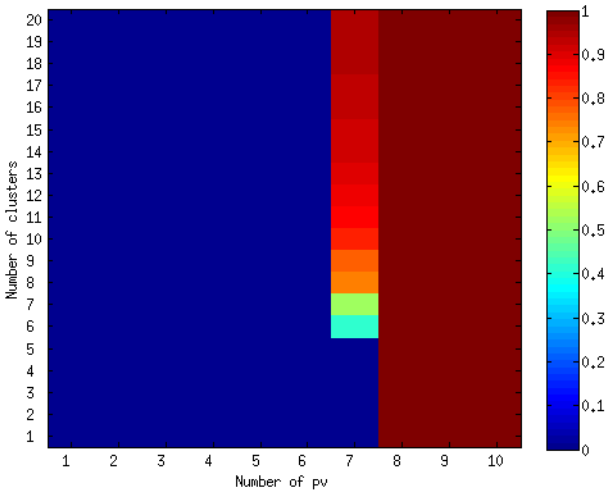

Figure 3.2: H_2 flame (1D): plots showing the minimum R^2 statistics for source terms

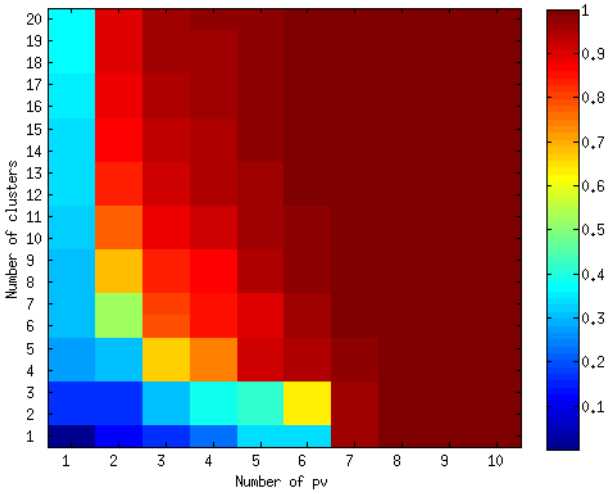

Figure 3.3: H_2 flame (1D): plots showing the minimum R^2 statistics for state variables adopting the pareto scaling.

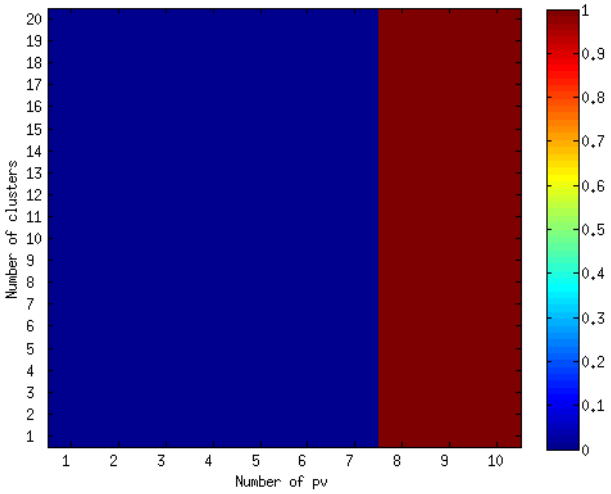

Figure 3.4: H_2 flame (1D): plots showing the minimum R^2 statistics for source terms adopting the pareto scaling

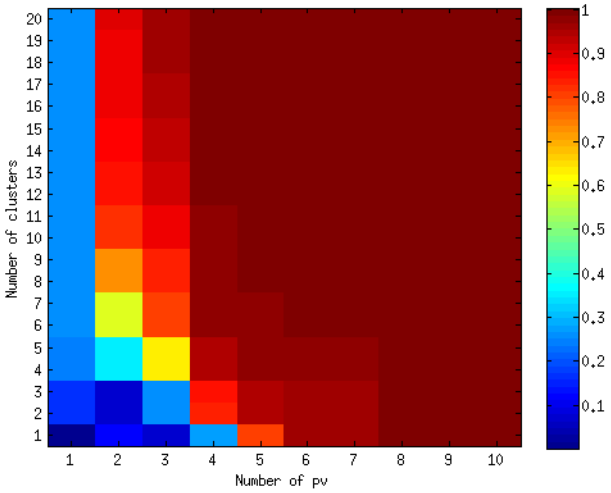

Figure 3.5: H_2 flame (1D): plots showing the minimum R^2 statistics for state variables adopting the VAST scaling.

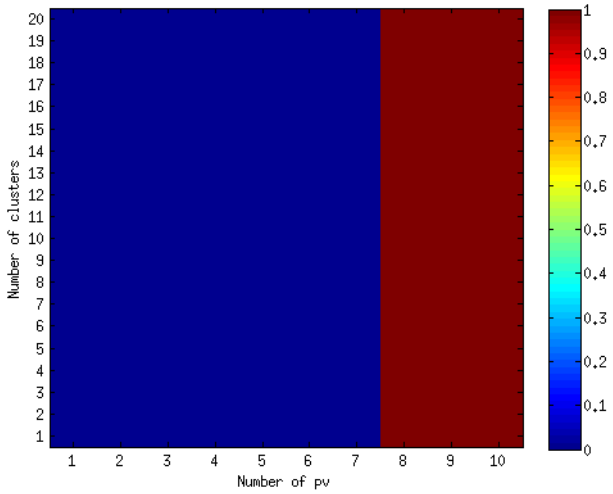

Figure 3.6: H_2 flame (1D): plots showing the minimum R^2 statistics for source terms adopting the VAST scaling

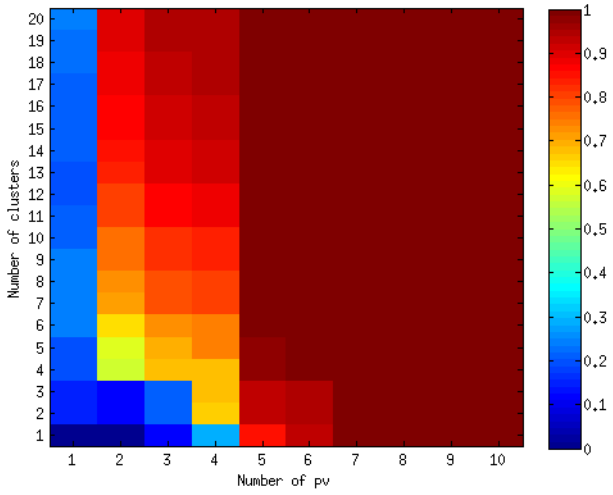

Figure 3.7: H_2 flame (1D): plots showing the minimum R^2 statistics for state variables adopting the range scaling.

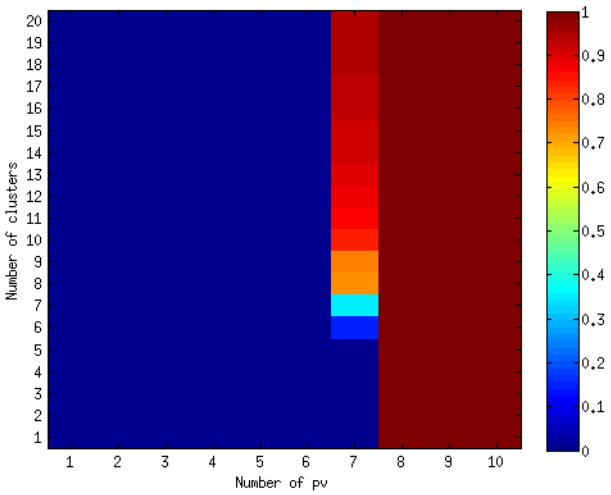

Figure 3.8: H_2 flame (1D): plots showing the minimum R^2 statistics for source terms adopting the range scaling

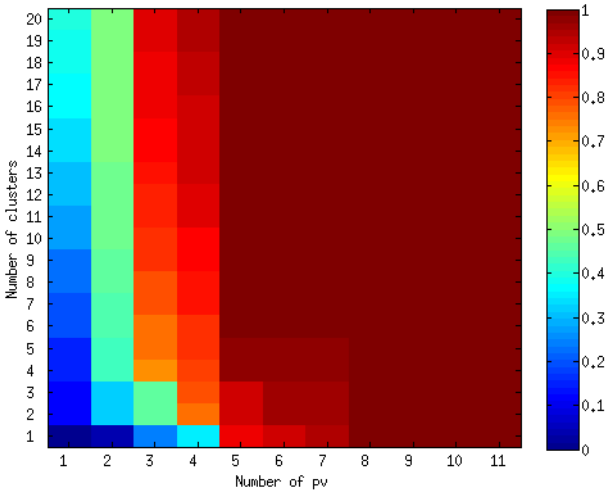

Figure 3.9: H_2 flame (2D): plots showing the minimum R^2 statistics for state variables

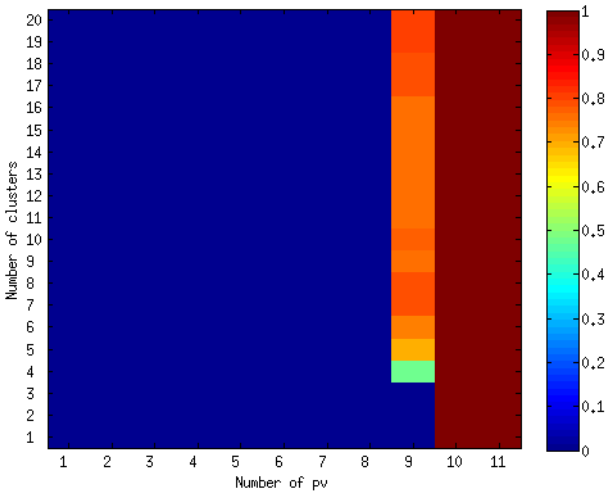

Figure 3.10: H_2 flame (2D): plots showing the minimum R^2 statistics for source terms

Also in this case, the standard scaling criterion seems to be the best one to achieve the minimum recontruction error, as it can be deduced observing Fig. 3.11-3.16, that display the minimum R^2 statistic for the variables reconstructed with other scaling criteria and the corresponding R^2 statistic for the source terms.

Summary plots obtained with hydrogen flame, both in the 1D flame and in the 2D flame, indicate that a satisfactory reconstruction of the state variable and the corresponding source terms can be achieved only with seven variables and several clusters. For example, if the 1D flame is considered, the minimum R^2 statistics with 20 clusters for source terms is 0.95 (corresponding to the reconstruction of the source term for H_2O_2); in Fig. 3.17, all the R^2statistics for the reconstruction of state variables and source terms, obteined with the choose of 7 PV (OH, H_2O_2, H_2, HO_2, O, H_2O, H) by the B2 method, adopting the stardand scaling criterion and splitting the domain in 20 cluster, are elencated.

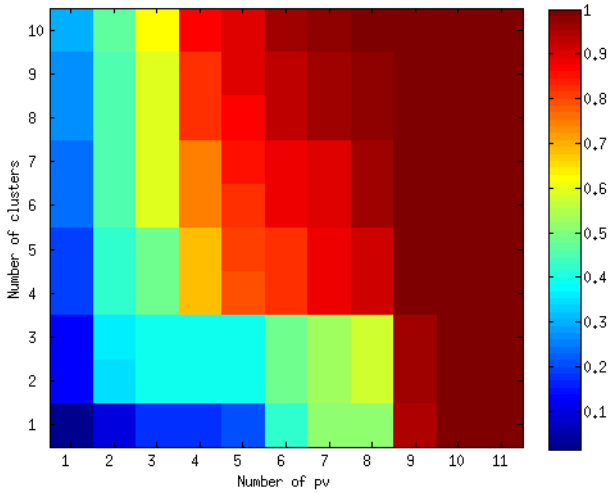

Figure 3.11: H_2 flame (2D): plots showing the minimum R^2 statistics for state variables adopting the pareto scaling.

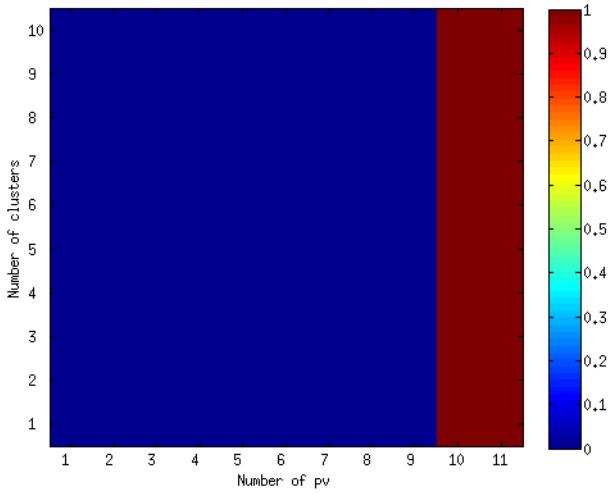

Figure 3.12: H_2 flame (2D): plots showing the minimum R^2 statistics for source terms adopting the pareto scaling

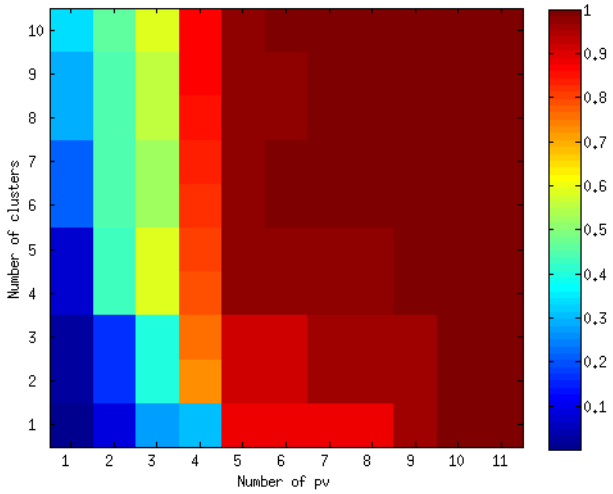

Figure 3.13: H_2 flame (2D): plots showing the minimum R^2 statistics for state variables adopting the VAST scaling.

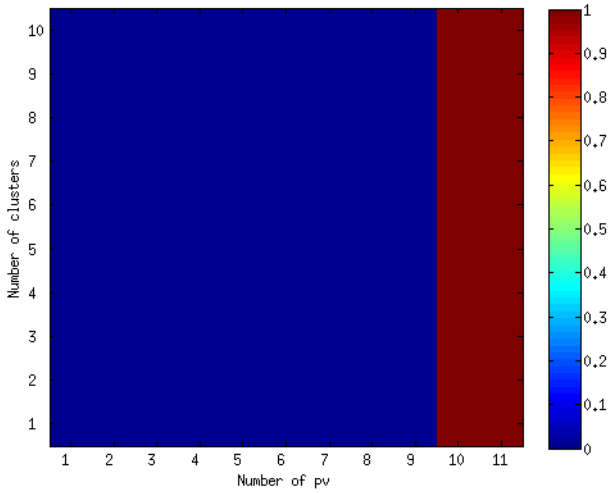

Figure 3.14: H_2 flame (2D): plots showing the minimum R^2 statistics for source terms adopting the VAST scaling

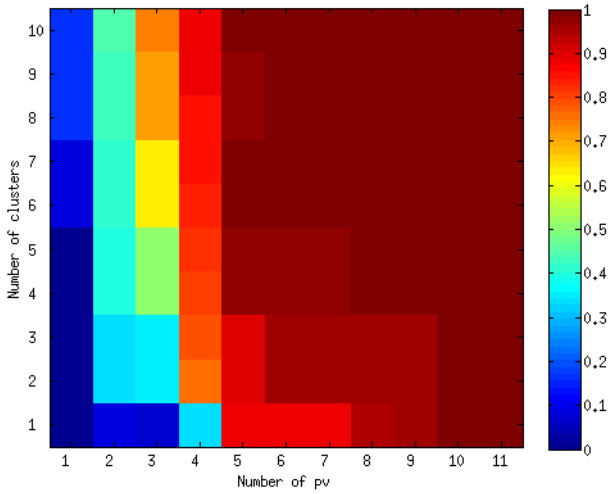

Figure 3.15: H_2 flame (2D): plots showing the minimum R^2 statistics for state variables adopting the range scaling.

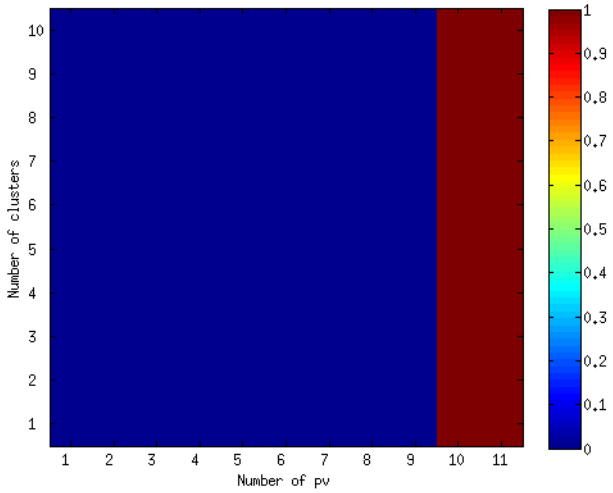

Figure 3.16: H_2 flame (2D): plots showing the minimum R^2 statistics for source terms adopting the range scaling

chemical species	R^2 state variables	chemical species	R^2 source terms
T	0.999999	T	0.999990
OH	1.000000	OH	0.996999
N_2	1.000000	N_2	1.000000
H_2O_2	1.000000	H_2O_2	0.948889
O_2	1.000000	O_2	0.999987
H_2	1.000000	H_2	0.999985
HO_2	1.000000	HO_2	0.975866
O	1.000000	O	0.999875
H_2O	1.000000	H_2O	0.999967
H	1.000000	H	0.999771

Figure 3.17: 1D hydrogen flame: R^2statistics for the reconstruction of state variables and source terms. Method to select PV: B2. Scaling criterion: STD. Number of PV: 7. Number of clusters: 20.

3.3 Solver

In the previous subsection, results obtained from the preprocessing were illustrated, showing the possibility to transport only a subset of the original state-space variables, in particular 7 PV for an hydrogen flame. Therefore, the matrices $\mathbf{B_j}$ and $\mathbf{A_{q_j}}(j = 1, ..., NCL)$, and the centering and scaling factors for each cluster, used to reconstruct the not transported variables by the MG-L-PCA in the preprocessing, were stored both for the 1D flame and the 2D flame.

Next step should be the modification of the laminarSMOKE solver to include the MG-L-PCA technique for the transport equations resolution. Goal of future works could be the simulation of the same flames with the modified solver and the comparison of the results obtained with the new methodology.

Moreover, hydrogen flames were chosen to the validation of the MG-L-PCA because chemistry involved in H_2 combustion is well-know and can be described with a small number of species, but a technique able to reduce the computational cost of

numerical simulation would be more appealing for complex systems and, the more complex the system is, the higher the reduction is supposed to be. Thus, when the hydrogen test-case will be tuned, the MG-L-PCA could be applied to simulate more complex systems, as flames feed with methane.

The MG-L-PCA technique could lead to a sensible reduction of the computational costs, due to a twofold effect: the former is the reduction of the number of transport equations to be solved, the latter is the possibility to obtain a reduce stiffness of the numerical systems, because of the smaller differences in orders of magnitude of the involved variables.

Conclusions

In this thesis, the Manifold Generated from Local Principal Component Analysis (MG-L-PCA) was presented, illustrating a new methodology to reduce the computational cost of numerical simulations of complex reactive flows, by the resolution of only a subset of the transport equations.

In particular, it has been shown the development of this technique from the Principal Component Analysis (PCA), that transforms the data through a linear projection onto a lower-dimensional space, retaining as much as possible of the variation present in the data-set. An overview of the coupling of this technique with a CFD solver was proposed; methods to select the variables to be transported, the Principal Variables (PV), have been elencated and the optimal criteria to obtain all the necessary information from the preprocessing step have been showed. Required data-sets for the preprocessing employed in this work have been get from a new solver able to simulate laminar flames (laminarSMOKE), the same that could be modified in order to apply the MG-L-PCA to predict the behavior of these flames. Therefore, the laminarSMOKE framework was previously presented, showing the governing equations, the numerical methods employed to solve them and the kinetic schemes adopted to describe the chemistry. Several laminar flames, feed with hydrogen and methane, were simulated with this new solver and computed results have been compared to the experimental values. As the agreement was satisfactory, the feasibility and accuracy of this tool was beared out.

The MG-L-PCA technique was eventually applied to two datasets of the same hydrogen/air flame, the former modelled in one-dimension, the latter two-dimensional. Summary plots displaying the reconstruction errors for the not transported values and source terms pointed out the possible to achieve a reduction of the variables to be transported, while preserving a fair accuracy. Additional work is required to include the MG-L-PCA in the solver, in order to simulate also more complex systems with lower computational costs.

Bibliography

[1] Cuoci A. Fluidodinamica degli incendi, 2011. notes lectures of the homonym course.

[2] Parente A. *Experimental and Numerical Investigation of Advanced Systems for Hydrogen-based Fuel Combustion.* PhD thesis, UniversitÃ di Pisa. Ingegneria chimica e dei materiali., 2008.

[3] Pfefferle L. D. Smooke M. D. Bennet B. A. V., McEnally C. Computational and experimental study of axisymmetric coflow partially premixed methane/air flames. *Combustion and flame*, 123:522–546, 2000.

[4] Giacquel O. Coussement A., Parente A. Reduced-order combustion modelling using principal components: the mg-pca approach. 2012.

[5] Parente A. Coussement A., Giacquel O. Mg-local-pca method for reduced order combustion modeling. *Proceedings of the Combustion Instutute*, xx:xxx–xxx (In Press), 2012.

[6] Faravelli T. Ranzi E. Cuoci A., Frassoldati A. Numerical modeling of laminar coflow flames with detailed kinetic schemes. *XXXIV Meeting of the Italian Section of the Combustion Institute*, 2011.

[7] Faravelli T. Ranzi E. Cuoci A. Frassoldati A. Cuoci A., Frassoldati A. A computational tool for the detailed kinetic modeling of laminar flames: application to c2h4/ch4 coflow flames.

[8] Isaac B. Smith P. J. Giacquel O. Parente A., Coussement A. Optimal selection of principal variables for the development of reduced-order combustion models. 2012.

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

[10] Green D. W. Perry R. H. *Perry's Chemical Engineers' Handbook*. McGraw-Hill, 1999.

[11] Levinsky H. B. Smooke M. D. Toro V. V., Mokhov A. V. Combined experimental and computational study of laminar, axisymmetric hydrogen-air diffusion flames. *Proceedings of the Combustion Institute*, 30:485–492, 2005.

Appendix A

Skeletal mechanism

Profiles obtained for CH_4 flames shown in Chapter 1 applying the skeletal mechanism are plotted with the values obtained with the detailed mechanism and the experimental values in Fig. A.1-A.7 .

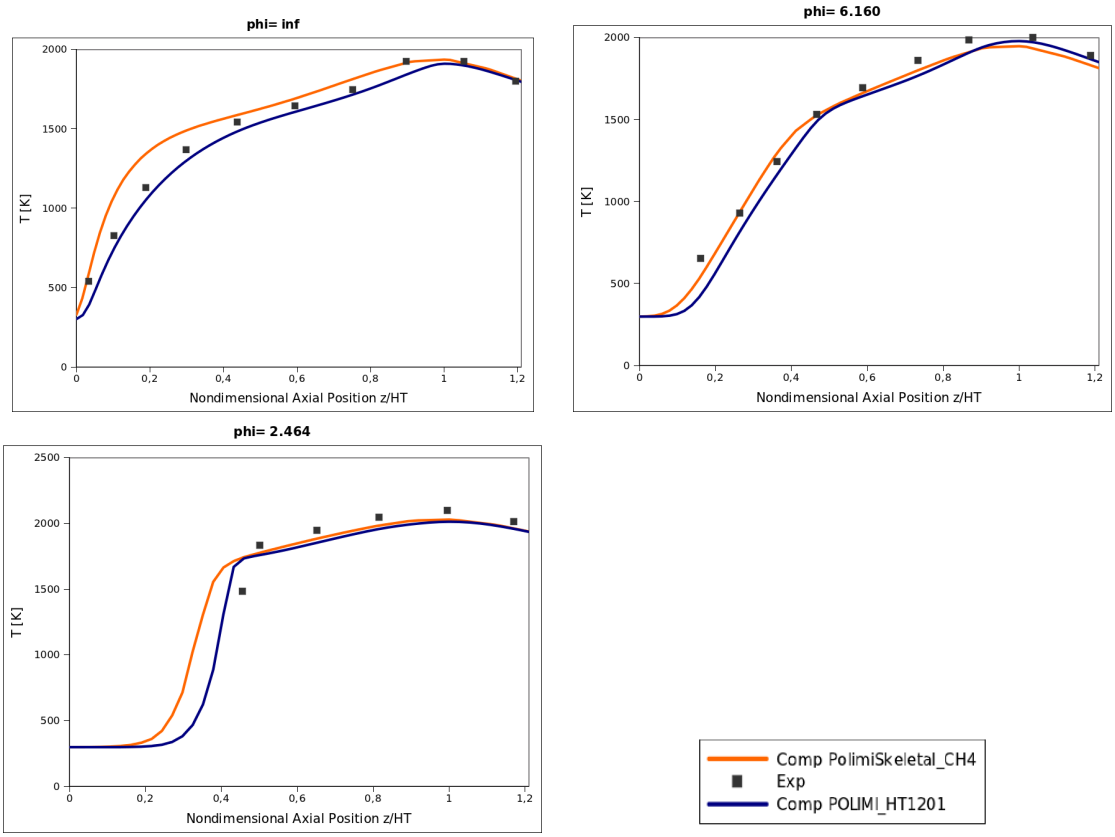

Figure A.1: Profiles of temperature along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

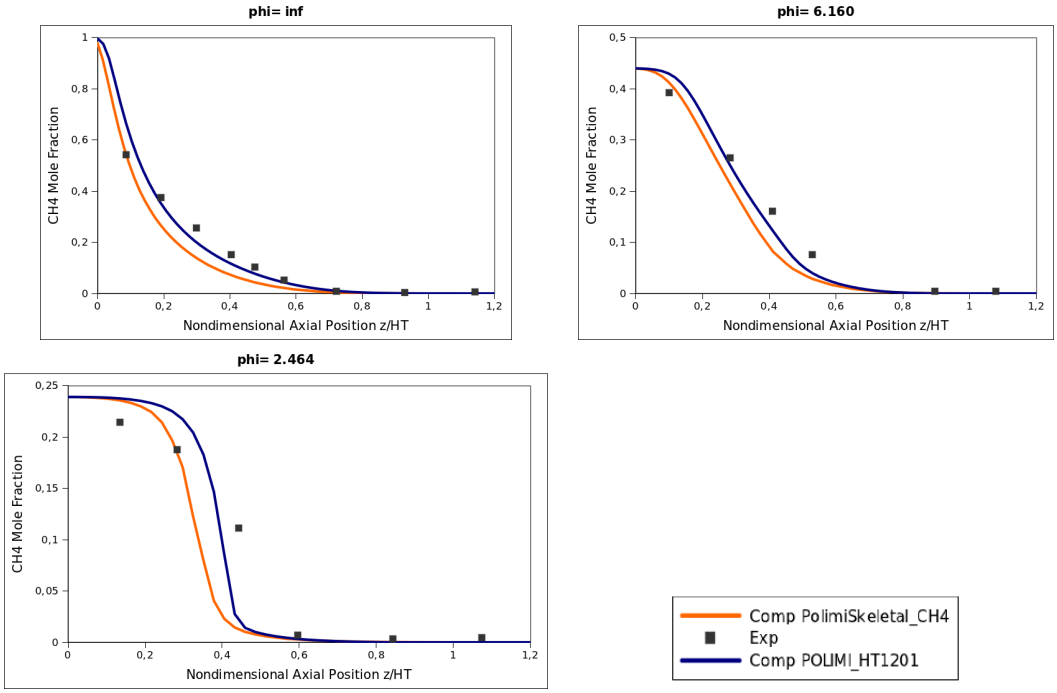

Figure A.2: Profiles of CH_4 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

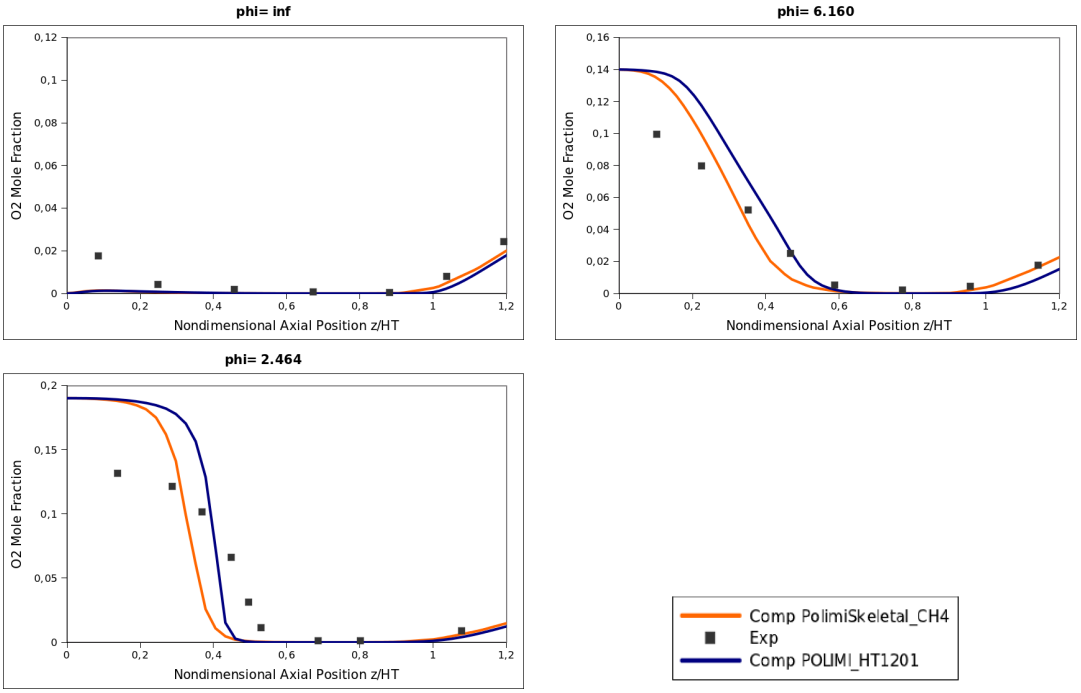

Figure A.3: Profiles of O_2 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

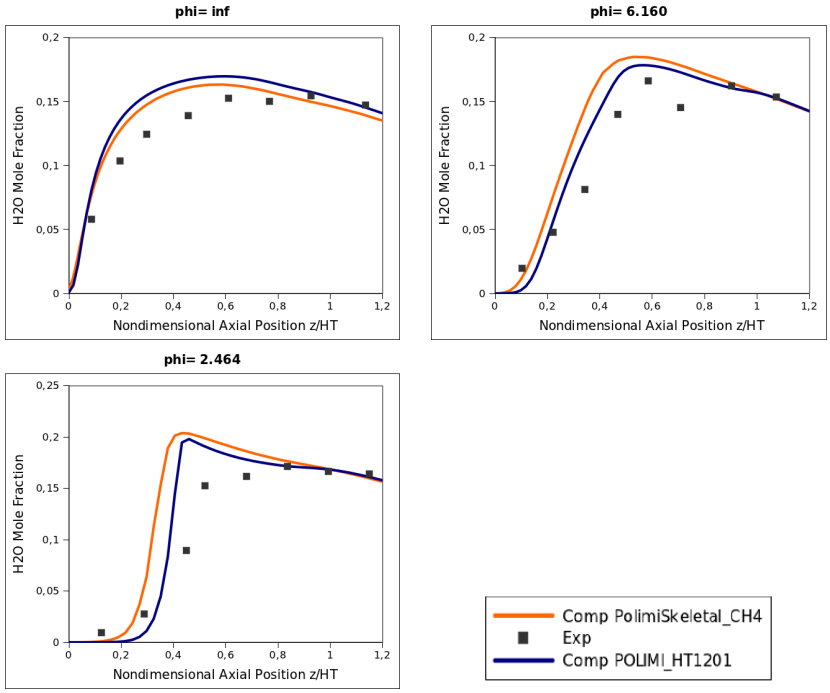

Figure A.4: Profiles of H_2O mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

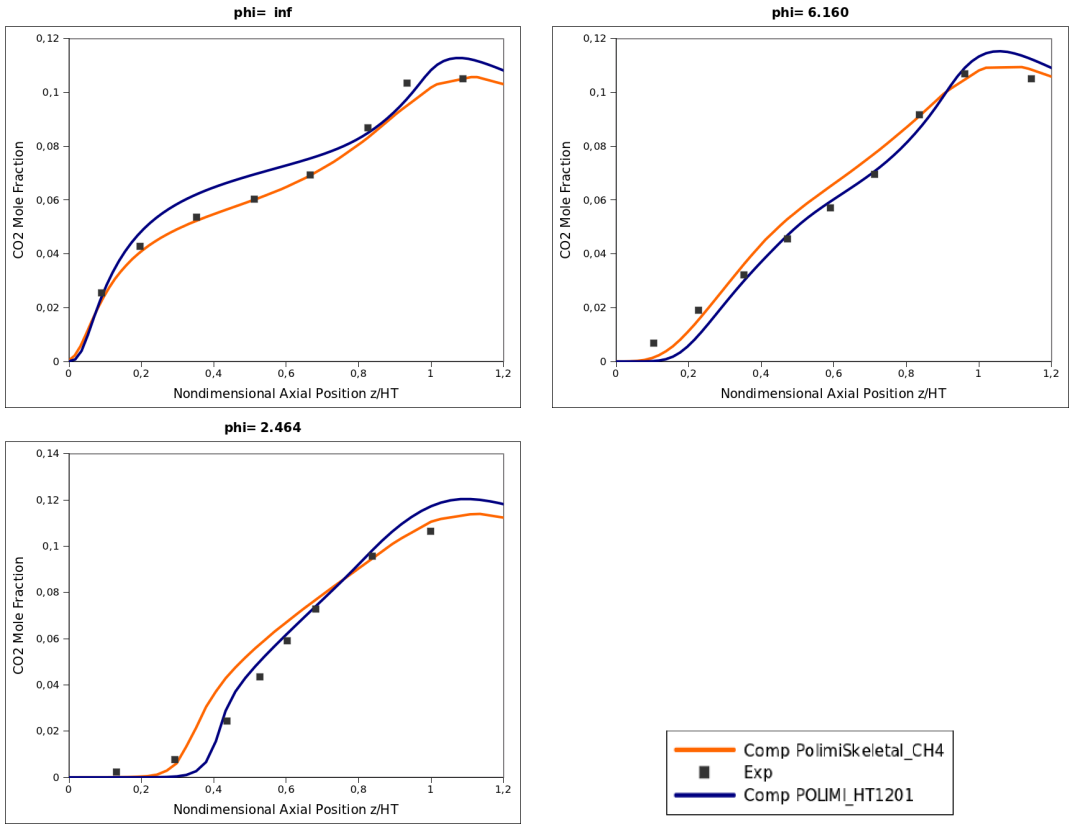

Figure A.5: Profiles of CO_2 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

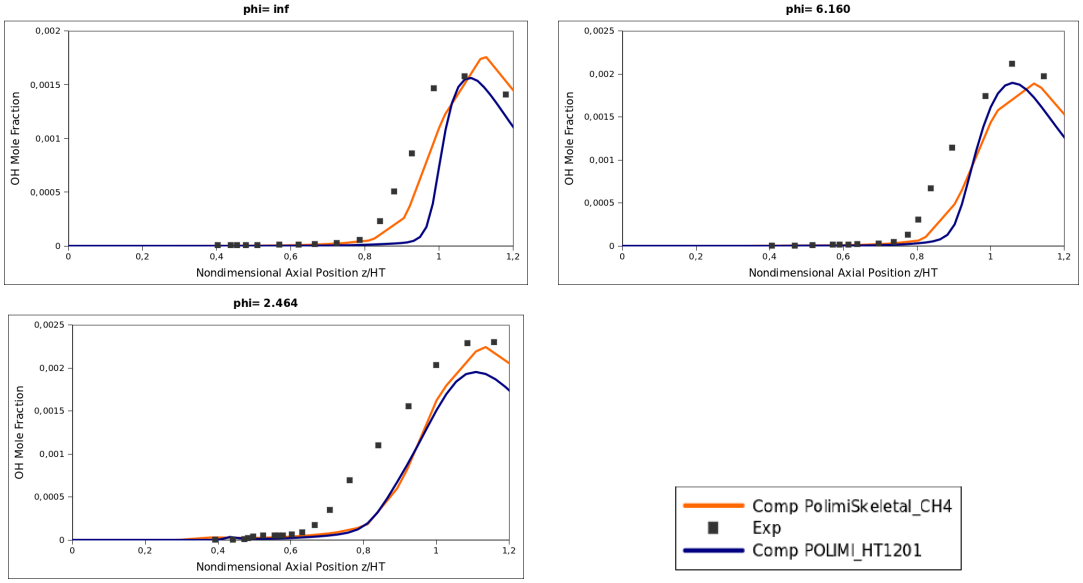

Figure A.6: Profiles of OH mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.

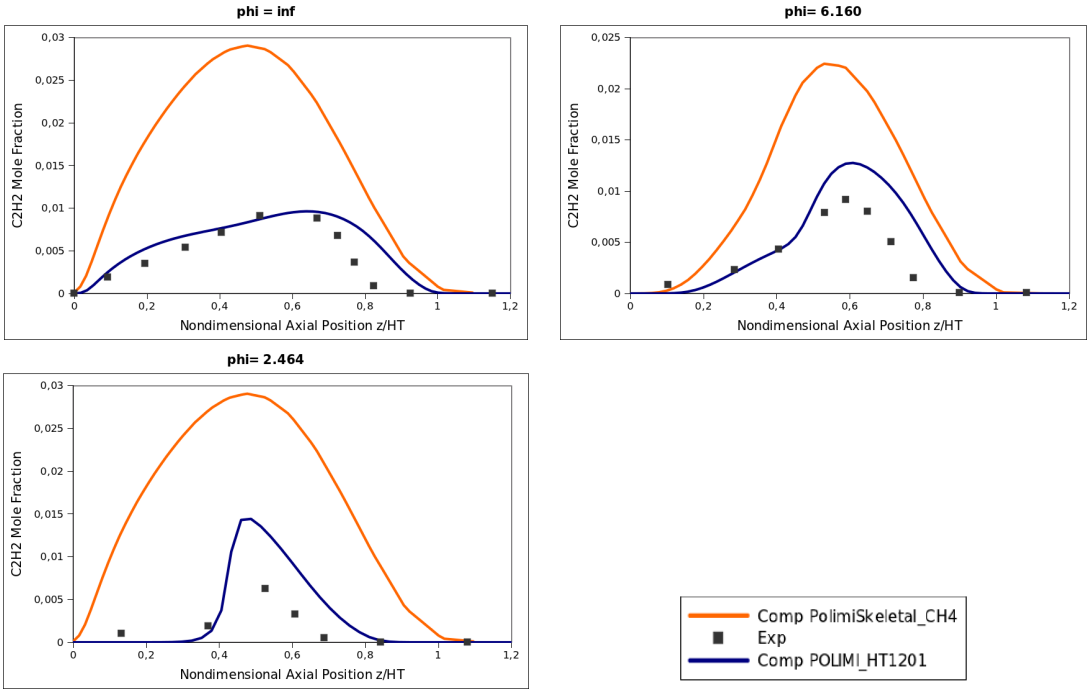

Figure A.7: Profiles of C_2H_2 mole fractions along the flame centerline, as functions of nondimensional axial position ($\frac{z}{H_T}$). Symbols for experimental data do not indicate specific data points.