A novel methodology based on Principal Component Analysis (PCA) is proposed for the investigation of the main features characterizing the MILD combustion regimes. High fidelity experimental datasets of CH4/H2 flames propagating into a hot and diluted co-flow are employed to this purpose. Results indicate that the proposed methodology can provide progress variables which coherently follow the modification of the flame structure experimentally observed at different levels of oxygen dilution in the co-flow. A local PCA approach is also proposed, based on the partition of the data sets into cluster, followed by the local application of PCA. Such approach ensures that the low-dimensional projection is characterized by the minimum parameterization error. Results indicate that local PCA leads to the identification of regions of the flow characterized by different physical processes, thus allowing the determination of optimal progress variables in each of them.
Investigation of the MILD combustion regime via Principal Component Analysis.

A. Parente\textsuperscript{1,2}, J.C. Sutherland\textsuperscript{3}, B.B. Dally\textsuperscript{4}, L. Tognotti\textsuperscript{2} and P.J. Smith\textsuperscript{3}

\textsuperscript{1} Environmental and Applied Fluid Dynamics Department, Von Karman Institute for Fluid Dynamics, Bruxelles, Belgium
\textsuperscript{2} Department of Chemical Engineering, University of Pisa, Pisa, Italy
\textsuperscript{3} Department of Chemical Engineering, University of Utah, Salt Lake City, UT, USA
\textsuperscript{4} School of Mechanical Engineering, The University of Adelaide, South Australia, Australia

\textbf{Corresponding author:}

James Sutherland.

Tel: +1 801 585 1246

e-mail: James.Sutherland@utah.edu

\textbf{Name of colloquium that describes the research}

\#12: NEW TECHNOLOGY CONCEPTS, REACTING FLOWS AND FUEL TECHNOLOGY

\textbf{Total length of paper and method of determination}

5749 words evaluated with method 1

\textbf{Word equivalent lengths for}

Main text: 3469 words
Nomenclature: 0 words
Equations: 0 words
References: 402 words
Figures with captions:
  Figure 1) 76
  Figure 2) 181
  Figure 3) 459
  Figure 4) 459
  Figure 5) 222
  Figure 6) 140
Tables:
  Table I) 153
  Table II) 70
  Table III) 117
1. Introduction

Moderate or Intense Low oxygen Dilution (MILD) combustion [1-3] has recently received great attention due to its potential of ensuring very high combustion efficiencies with very low pollutant emissions. MILD combustion is characterized by a complete modification of the reaction zone structure with respect to a conventional flame. Such a combustion regime requires the reactants to be preheated above the self-ignition temperature of the mixture and enough inert combustion products to be entrained in the reaction region. These two requirements ensure a distributed reaction spanning a large volume of the combustion chamber and resulting in a reduced and relatively uniform temperature field. By avoiding temperature peaks, NO\textsubscript{x} formation is substantially reduced. Experimental and numerical studies have been devoted to MILD combustion. However, the fundamentals of such combustion regime are still not completely understood and this generates uncertainties, especially when trying to model the process numerically.

The present paper proposes a novel approach, Principal Component Analysis (PCA) [4-5], to identify the dominant features of MILD combustion. PCA has been recently applied to the analysis of turbulent reacting systems [6-7], to investigate the existence of low-dimensional manifolds [8-9] in well-documented reference flames [10-11]. In particular, the PCA method was able to determine the dimensionality of such low-dimensional manifolds and to identify the dominant progress variables for the description of the systems. Such findings are extremely appealing from the modeling point of view, as they open a completely new perspective in the development of turbulence/chemistry interaction models. In the present paper, PCA is applied to a combustion regime different from the ones considered earlier [6-7]. The main purpose of the analysis is to help improve the fundamental understanding of MILD regime and provide optimal progress variables for its mathematical modeling.
2. **Theory**

Principal Component Analysis (PCA) is a well-established statistical technique for size reduction in data mining. PCA is concerned with detecting the directions which are most active in a multivariate data set, thus providing an optimal low-dimensional projection of the system. For a data set, $X$, consisting of $n$ observations of $p$ variables, the Principal Components (PCs), $Z$, are defined by the projection of the original data onto the eigenvectors, $A$, of the covariance matrix, $S$, $Z = XA$. The eigenvalue matrix, $L$, associated to $S$ quantifies the relative importance of the PCs, thus allowing defining a truncated subset of PCs of size $q$, $Z_q = XA_q$, which minimizes the amount of information loss in the low dimensional projection. The total variance accounted by a reduced set of PCs of size $q$ can be defined as $t_q = \frac{\sum_{k=1}^{q} l_k}{\sum_{k=1}^{p} l_k}$, where $l_k$ is the $k^{th}$ eigenvalue of $S$. It is also possible to define a measure of the individual variance, the contribution of the retained eigenvectors to the variance of each variable, as $t_{q,j} = \sum_{i=1}^{q} (L_{iz_j})^2$, where $L_{iz_j}$ is the correlation between the $i^{th}$ eigenvector of $A$ and the $j^{th}$ variable, i.e. loading, and it is calculated as $L_{iz_j} = \frac{\alpha_{ji} \sqrt{l_i}}{s_j}$, where $\alpha_{ji}$ is the weighting of the $j^{th}$ variable of $X$ on $x_i$, and $s_j$ is the variance of the $j^{th}$ variable of $X$. Variables are generally centered before PCA is carried out, to convert observations to fluctuations around the mean; moreover, scaling is applied when the elements of $X$ are in different units or when they have different variances. Several scaling options are available [12]: standard deviation is often used as scaling parameter for exploratory analysis. A scaled variable can be then defined, $\tilde{x}_j = x_j / d_j$, where $d_j$ is the scaling parameter adopted for variable $x_j$.

More details about PCA and its application to turbulent reacting systems may be found in [6-7,12].
2.1. Interpretation of Principal Components

The principal components are, by construction, linear combinations of all the measured variables of $X$. Therefore, their physical interpretation is usually not straightforward. Various attempts have been made to overcome this difficulty; among them, PCs rotation [13] represents a very common solution. Through rotation, the weights can be redefined to meet alternative criteria. In particular, rotation is aimed at attaining a simple structure for $A_q$, so that weights on a PC are either close to unity or close to zero and, thus, variables have large weights on only few or (ideally) one PC. The most common orthogonal rotation is based on the maximization of the VARIMAX criterion [13]. Verbally, such rotation concentrates the amount of variance explained for any of the original variables on single PCs. After VARIMAX rotation, $A_q$ will generally have fewer large coefficients in its columns, thereby making the columns more easily interpretable.

When rotation is applied, the total variance within the rotated $q$-dimensional subspace is redistributed amongst the rotated components more evenly than before rotation; however it remains the maximum that can be achieved. Moreover, the choice of $q$ can have a large effect on determination of the rotated set of PCs. Therefore, it is advised to apply rotation only after $q$ has been carefully determined, after a well defined variance criterion has been satisfied.

2.2. Local Principal Component Analysis

The PCA transformation cannot handle non-linear correlations. Therefore, when applied to turbulent reacting systems, global PCA (GPCA) fails to find the most compact representation of the system. Such realization has prompted the development of an alternative PCA approaches [14, 6], which employs a partition of the multivariate data set into clusters (Vector Quantization - VQ), followed by the local application of PCA into each region. The peculiarity of the proposed VQPCA algorithm, schematized in Figure 1, is that the partition step is not independent of the PCA analysis; on the contrary, it is based on
the minimization of the error associated to the PCA reconstruction, \( \varepsilon_{GSRE,n} = \| \hat{X}_i - \hat{X}_i \| \), where \( \hat{X}_i \) is the \( i \)th observation of the sample scaled by \( D \), the diagonal matrix whose \( j \)th diagonal element is the scaling factor \( d_j \) associated to \( x_j \). The VQ step partitions the data into clusters, following the curvature of the manifold in the low-dimensional space, as indicated by the double-headed arrows in Figure 1. Then, the points are assigned to the clusters depending on their low-dimensional projection on each cluster, as illustrated by the red dot (original data point) moved the orange ones (low-dimensional projection) in Figure 1. The reconstruction error is then measured with respect to the mean data variance as \( \varepsilon_{GSRE,n} = E(\varepsilon_{GSRE})/E[\text{var}(X)] \), where \( E(\phi) \) indicates the expectation of \( \phi \). From a modeling perspective, the translation of the VQPCA approach into a combustion model would follow a philosophy similar to the ones adopted in Conditional Moment Closure (CMC) and Multiple Mapping Conditioning (MMC), especially when a supervised conditioning on mixture fraction is employed [15-16].

3. Experimental data

The PCA analysis is carried out on a jet in hot co-flow (JHC) burner [18-20], designed to emulate MILD conditions. It consists of a central fuel jet (80% CH\(_4\) and 20% H\(_2\)) within an annular co-flow of hot exhaust products from a secondary burner mounted upstream of the jet exit plane. O\(_2\) in the co-flow is controlled at three different levels, 3, 6 and 9% (by vol.), while the temperature and exit velocity are kept constant. The \( X \) matrix to be processed with PCA is filled with high fidelity experimental measurements of temperature and species mass fractions (N\(_2\), O\(_2\), H\(_2\)O, H\(_2\), CH\(_4\), CO, CO\(_2\), OH, NO), available from the experiments [18]. Data are arranged so that each row is a different observation of state variables, each occupying a different column. The number of realizations available for the 3, 6 and 9% O\(_2\) cases are approximately 56000, 55000 and 61000, respectively.
Experimental data have been preprocessed to remove possible outlier observations. The multivariate nature of PCA itself allows defining proper metrics for the identification and removal of observations which but do not conform to the structure of the data. In particular, the first PCs are strongly affected by variables with large variances and, thus, they can detect outliers on the original variables. On the other hand, the last few PCs are sensitive to observations that are inconsistent with the correlation structure of the sample and they can trace multivariate outliers. Based on these considerations, a detection scheme developed in the field of image processing [17] has been extended and modified for the analysis of thermo-chemical samples. The approach is based on the definition of two classifiers, one from major, 

\[ PCC_1 = \sum_{k=1}^{q} z_{ik}^3 / l_k, \]  

and one from minor, 

\[ PCC_2 = \sum_{k=p-r+1}^{p} z_{ik}^3 / l_k, \]  

PCs. Then, an observation \( X_i \) is considered as outlier if \( PCC_1 > c_1 \) and \( PCC_2 > c_2 \), where \( c_1 \) and \( c_2 \) are chosen as the 0.99 quantile of the empirical distributions of the classifiers. The algorithm for outliers removal is iterative and it stops when the covariance structure of the sample is unaffected by the elimination of additional data points. Typically, a fraction of ~2-4% of original observations is removed during preprocessing, thus leaving plenty of statistics for the PCA analysis.

4. Results

Figure 2 shows the eigenvalue size distribution and the contribution of the \( q \) largest eigenvalues to the total variance, \( t_q \), for the complete JHC data sets with different \( O_2 \) in the co-flow, i.e. 3\% (a) and 9\% (b). It can be observed that four components are needed in all cases to account for more than 90\% of the variance contained in the original data. However, some differences can be highlighted regarding the eigenvalue distribution for the two investigated conditions. As the oxygen mass fraction in the co-flow increases, the eigenvalue associated to the second PC increases and the difference between the 3\textsuperscript{rd} and 4\textsuperscript{th} eigenvalues decreases. Moreover, the size of the 5\textsuperscript{th} PC is significantly reduced going from the 3\% to
the 9% case. Such difference denote that the structure of the covariance matrix, and, therefore, the flame features are strongly affected by the oxygen content, as also indicated by the experimental evidence [18]. The purpose of the PCA analysis is, then, to investigate how the change in flame structure affects the definition of the main progress variables which describe the system.

The total, $t_q$, and individual variance, $t_{q;j}$, accounted by the first three PCs for the JHC data sets with 3% and 9% $O_2$ in the co-flow are listed in Table 1. It can be observed that, by choosing $q=3$, it is possible to capture more than 90% of the individual variances of all the main species and temperature, while minor species such as CO and OH and NO require an additional PC, $q=4$, to reach level of approximation comparable to other state variables. Two major differences can be highlighted between the two different $O_2$ dilution levels by comparing the values of $t_{q,j}$ obtained for CO, OH and NO with $q=3$. In particular, the variance accounted for CO and NO species increases when increasing the $O_2$ content in the co-flow, whereas an opposite trend is observed for OH radical. This is further confirmed by Table 2, which shows the contribution of the first four PCs to the variance accounted for CO, OH and NO, when increasing the oxygen content in the co-flow. The major contribution to the explained variance of OH shifts from the 3rd to the 4th PC, when going to 3% to 9 % $O_2$. Thus, for the 9% case, the inclusion of the 4th PC appears fundamental to capture OH distribution. A possible explanation to this can be provided by taking into account that in MILD combustion regime, the reduction in reaction rates at low $O_2$ levels leads to the weakening of the OH layer and, more importantly, to a reduction of OH gradients in the reaction region [18-20]. As a consequence of the higher homogeneity, it is then possible to capture OH variation with a reduced number of PCs. As far as CO and NO are concerned, the increase of $O_2$ levels in the co-flow leads to a structure of the loading matrix (Table 2) which shows a more homogeneous distribution of the variance explained on the first three PCs. If we consider CO, a major part of its variance is explained by the 3rd PC for the 3% case, whereas for the 9% case, similar contributions are
provided by the 1st and 2nd PCs. Again, this indicates that the addition of oxygen to the co-flow determines a departure from MILD conditions towards a conventional flame structure, where finite rate chemistry effects are less important and the oxidation process can be captured by mixing parameters.

4.1. PCs interpretation

Figure 3 shows histograms of the structure of the eigenvector matrix $A_q$ with $q=4$, for the 3% and 9% O$_2$ cases. It can be easily seen how the first two components are characterized by non negligible weights on most of the original variables. Therefore, it is not straightforward to provide a physical interpretation of the PCs and to describe how they follow the modification of the flame structure. However, this is a known issue in PCA, as it was pointed out in Section 2.1, explaining the use of rotation to attain a simpler PC structure. Still, useful information can be obtained by the analysis of Figure 3. In particular, the sign of the coefficients within a PC specify its orientation in the original multidimensional space, thus clarifying the relation between the original variables of the sample. Then, if $a_{il}>0$ and $a_{ik}<0$, the trajectory of the PC will move towards increasing values of $x_l$ and decreasing values of $x_k$ or vice versa.

In fact, the absolute sign of a PC is unimportant, as reversing the sign on all coefficients of a given PC leaves the variance captured unchanged, as well as its orthogonality with the other PCs. Given this remark, it becomes clear that how the major difference between the 3% and 9% cases become visible for the 3rd and 4th PCs. The comparison between the weights of the CO and OH species on the 4th PC reveals, for the 3% O$_2$ case, the existence of regions where the variance is directed towards concordant (positive or negative) values of CO and OH, in opposition to what observed for the 9% case, where the weight relative to CO disappears. Correspondingly, the weight of CO$_2$ on the 3rd PC significantly decreases with respect to the 9% O$_2$ case. Such result appears of particular significance and confirms the existence of a more distributed reaction zone in MILD regime (3% O$_2$ case), characterized by slower reaction rates and resulting in a major relevance of minor and intermediate species. As far as NO is
concerned, minor contributions to the 3rd and 4th PCs are observed for the 3% O\textsubscript{2} case with respect to the 9% O\textsubscript{2} case. This can be explained by taking into account that NO formation in MILD combustion is more homogeneous than in traditional non-premixed combustion, due to the smaller temperature gradients [1-13]; therefore, NO is characterized by less variability and its contribution to the PCs reduces.

A clearer interpretation of the PCs physical meaning can be achieved by applying the VARIMAX rotation to the original unrotated eigenvector matrix (Figure 3). Following the discussion in Section 2.1, it was decided to apply rotation to the set of PCs determined to account for at determined amount of original variance. A level of 98% was chosen, to ensure that most of the information were recovered before rotation was applied, thus limiting the influence of \(q\) on the selected subset of PCs. Figure 4 shows histograms of the structure of the first four rotated eigenvectors for the 3 and 9% O\textsubscript{2} cases. Starting from the 3% O\textsubscript{2} case, it can be observed how the first (rotated) PC is an ensemble of temperature, oxidizer and product species. This component has the effect of capturing as much as possible of the original data variance and it is therefore generally referred as “size PC” within the context of PCA. The second PC is characterized by large weights on reacting and oxidizer species. Such component is very highly correlated to the system mixture fraction (>90%) and, thus, it can be regarded as a measure of the local stoichiometry of the system. The third PC is characterized by a dominant contribution of CO, which underlines the major role of intermediate combustion products in the investigated combustion regimes. Finally, the last component is exclusively the radical species OH, which is needed for capturing the location and extension of the reaction region. When the O\textsubscript{2} fraction in the co-flow is increased, some important changes occur. First, the weight of the O\textsubscript{2} species on the first rotated PC significantly decreases. This is not surprising, since O\textsubscript{2} control on the overall oxidation process becomes less important, due to the increased oxidizer availability. Another important difference
is observed for the 3\textsuperscript{rd} PC, whose structure completely changes, being the dominant contribution of the intermediate species CO replaced by a large NO weight. The modification of the PC reflects the evolution of the system towards a “conventional” diffusion flame, where NO emissions are significantly higher than for the 3\% O\textsubscript{2} case \cite{18} and they characterized by a larger variance, thus requiring a specific PC. The structure of the 2\textsuperscript{nd} (stoichiometry) and 4\textsuperscript{th} (OH) PCs does not significantly change when increasing the O\textsubscript{2} dilution level.

4.2. **Parameterization error: GPCA and Local PCA**

The quality of the low-dimensional reconstruction given by GPCA can be judged qualitatively plotting the original state variables versus the recovered ones after the PCA reduction is applied. Figure 5 shows the parity plots of temperature, CO and OH mass fraction given by GPCA with $q=3$ for the 3\% O\textsubscript{2} case. Although it is possible to recover around 90\% of the original data variance with 3PCs, Figure 5 points out the existence of non linear deviations in the recovered data, which can be probably ascribed to non linear dependencies among the original variables. It can be argued \cite{6} that the manifold dimensionality given by GPCA can be somehow biased, whereas the local linear model presented in Section 2.2 can be effectively adopted to overcome this issue. This can be supported quantitatively by means of the error metric introduced in Section 2.2, $\varepsilon_{\text{GSRE};n}$, which evaluate the reconstruction error on a scaled basis. Differently from the definition of $t_q$, such metric ensures that also minor variables are taken into account when judging the effectiveness of the low-dimensional projection. When VQPCA is applied (Figure 5) with $k=6$ and $q=3$, the parity plots of recovered versus original variables show a much better agreement, confirmed by a value of $\varepsilon_{\text{GSRE};n}$ equal to 0.08, almost 12 times smaller than that given by GPCA with $q=3$, i.e. $\varepsilon_{\text{GSRE};n}=1.0$. Interestingly, the parity plots for the state variables in Figure 5 show how, after partitioning, the relationships between the original and reconstructed data are mainly linear.
To better understand the mechanism underlying the VQ partitioning in the VQPCA algorithm; it is possible to analyze the structure of the rotated eigenvectors in the two clusters identified by VQPCA. Differently from the case shown in Figure 5, the PCs are extracted by fixing the number of clusters and by imposing a desired amount of variance, i.e. 95%, to be accounted for each of them. Such procedure appears optimal to extract only the relevant features within each cluster. As a result, 3 PCs are extracted in the first cluster and two in the second. The first eigenvectors in each of the two clusters are listed in Table 3. Moreover, Figure 6 shows how the partitioned temperature is projected back into the mixture fraction space, to help the physical interpretation of the VQPCA algorithm. It is very interesting to observe that VQPCA is able to separate two major regions of the system: the first consisting of the lean and rich branches (where mixing is dominant), and the second covering the distributed reaction region (when finite-rate chemistry effects play a major role). As a consequence of the partition, the PCs identified in the two clusters (Table 3) are quite different. In the first cluster, two PCs which resemble the “mixture fraction” and “size” PCs discussed in Section 4.1 can be recognized; moreover no major influence of intermediate and minor species is observed. As far as the second cluster is concerned, the first two PCs are very strongly correlated to the radical species OH (1st) and the intermediate species CO (2nd), thus confirming the relevance of finite-rate chemistry effects for such region.

5. Concluding remarks

A novel methodology based on Principal Component Analysis has been applied to experimental datasets of CH₄/H₂ flames propagating into a hot and diluted co-flow. Results indicate that PCA can capture the modification of the flame structure as a function of the oxygen dilution in the co-flow, through the modification of the PC structure. Interestingly, the PCA analysis confirms the relevance of finite-rate chemistry effects in MILD combustion, thus supporting the findings of previous work in this field. A local PCA variant has been compared to classic PCA, by partitioning the data into clusters with an
unsupervised minimization error algorithm. The local approach yields substantially improved parameterization of the experimental data and appears to be more suited for the identification of optimal progress variables for the description of the system.
References


Table 1 - Total, $t_q$, and individual variance, $t_{q,i}$, accounted for the JHC data sets at 3 and 9% O$_2$ by the GPCA reduction, as a function of the number of retained PCs, $q$.

<table>
<thead>
<tr>
<th></th>
<th>$t_{q,i}$ (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3% O$_2$</td>
<td>9% O$_2$</td>
<td></td>
</tr>
<tr>
<td>$q$</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>T</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>Y$_{O_2}$</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Y$_{N_2}$</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Y$_{H_2}$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Y$_{H_2O}$</td>
<td>0.97</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>Y$_{CH_4}$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Y$_{CO}$</td>
<td>0.76</td>
<td>1.00</td>
<td>0.98</td>
</tr>
<tr>
<td>Y$_{CO_2}$</td>
<td>0.91</td>
<td>0.96</td>
<td>0.89</td>
</tr>
<tr>
<td>Y$_{OH}$</td>
<td>0.71</td>
<td>1.00</td>
<td>0.38</td>
</tr>
<tr>
<td>Y$_{NO}$</td>
<td>0.78</td>
<td>0.78</td>
<td>0.89</td>
</tr>
</tbody>
</table>

$t_q$ (%) 90 96 90 98

Table 2 – Squared loadings for the CO, OH and NO species, as a function of the oxygen dilution level.

<table>
<thead>
<tr>
<th></th>
<th>$L^2_{z_1x_j}$</th>
<th>$L^2_{z_2x_j}$</th>
<th>$L^2_{z_3x_j}$</th>
<th>$L^2_{z_4x_j}$</th>
<th>$L^2_{z_1x_j}$</th>
<th>$L^2_{z_2x_j}$</th>
<th>$L^2_{z_3x_j}$</th>
<th>$L^2_{z_4x_j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y$_{CO}$</td>
<td>0.13</td>
<td>0.10</td>
<td>0.52</td>
<td>0.24</td>
<td>0.39</td>
<td>0.24</td>
<td>0.36</td>
<td>0.00</td>
</tr>
<tr>
<td>Y$_{OH}$</td>
<td>0.19</td>
<td>0.01</td>
<td>0.52</td>
<td>0.28</td>
<td>0.26</td>
<td>0.01</td>
<td>0.12</td>
<td>0.61</td>
</tr>
<tr>
<td>Y$_{NO}$</td>
<td>0.68</td>
<td>0.08</td>
<td>0.01</td>
<td>0.00</td>
<td>0.55</td>
<td>0.16</td>
<td>0.18</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Table 3- Rotated eigenvector in the two clusters identified by VQPCA with $q=3$.

<table>
<thead>
<tr>
<th></th>
<th>cluster 1</th>
<th>cluster 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_{1r}$</td>
<td>$a_{2r}$</td>
</tr>
<tr>
<td>$T$</td>
<td>-0.10</td>
<td>-0.40</td>
</tr>
<tr>
<td>$Y_{O_2}$</td>
<td>-0.39</td>
<td>0.52</td>
</tr>
<tr>
<td>$Y_{N_2}$</td>
<td>-0.51</td>
<td>-0.13</td>
</tr>
<tr>
<td>$Y_{H_2}$</td>
<td>0.53</td>
<td>0.04</td>
</tr>
<tr>
<td>$Y_{H_2O}$</td>
<td>-0.06</td>
<td>-0.44</td>
</tr>
<tr>
<td>$Y_{CH_4}$</td>
<td>0.53</td>
<td>0.04</td>
</tr>
<tr>
<td>$Y_{CO}$</td>
<td>-0.07</td>
<td>-0.38</td>
</tr>
<tr>
<td>$Y_{CO_2}$</td>
<td>-0.01</td>
<td>-0.43</td>
</tr>
<tr>
<td>$Y_{OH}$</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>$Y_{NO}$</td>
<td>-0.02</td>
<td>-0.18</td>
</tr>
</tbody>
</table>
List of Figure Captions

Figure 1 – Schematization of the VQPCA algorithm.

Figure 2 – Eigenvalue size distribution and contributions to the total variance for the GPCA reduction of the JHC data set with 3 (a) and 9% (b) O₂.

Figure 3 – Structure of the first four eigenvectors provided by the GPCA reduction of the JHC data sets with 3 (a) and 9% (b) O₂.

Figure 4 - Structure of the first four rotated eigenvectors provided by the GPCA reduction of the JHC data sets with 3 (a) and 9% (b) O₂.

Figure 5 - Parity plots of temperature (a), and CO (b) mass fraction

Figure 6 - Temperature as a function of mixture fraction in the two clusters selected by VQPCA for the JHC flame with 3% O₂.
Figure 1 – Schematization of the VQPCA algorithm.
Figure 2 – Eigenvalue size distribution and contributions to the total variance for the GPCA reduction of the JHC data set with 3 (a) and 9% (b) O₂.
Figure 3 – Structure of the first four eigenvectors provided by the GPCA reduction of the JHC data sets with 3% (a) and 9% (b) \( O_2 \).
Figure 4 - Structure of the first four rotated eigenvectors provided by the GPCA reduction of the JHC data sets with 3 (a) and 9% (b) $O_2$. 
Figure 5 - Parity plots of temperature (a), and CO (b) mass fraction, showing the GPCA ($q=4$) and VQPCA ($q=4$ and $k=6$) reductions.
Figure 6 - Temperature as a function of mixture fraction in the two clusters selected by VQPCA for the JHC flame with 3% O$_2$. 