



ELSEVIER

Available online at www.sciencedirect.com ScienceDirect

Proceedings of the Combustion Institute 32 (2009) 1579–1586

**Proceedings
of the
Combustion
Institute**

www.elsevier.com/locate/proci

Identification of low-dimensional manifolds in turbulent flames

A. Parente^{a,*}, J.C. Sutherland^b, L. Tognotti^a, P.J. Smith^b^a *Department of Chemical Engineering, University of Pisa, Via Diotisalvi 2, Pisa I-56126, Italy*^b *Department of Chemical Engineering, University of Utah, Salt Lake City, UT 84112-9203, USA*

Abstract

A novel methodology based on principal component analysis (PCA) was developed for the identification of the low-dimensional manifold of a chemical reacting system, the determination of its dimensionality and the selection of optimal manifold variables. Results are presented for a simple CO/H₂/N₂ jet flame and for a CH₄ piloted flame (TNF Flame F). Results of the global PCA analysis (GPCA) on the whole data sets show that, in all cases, the number of modes required to reproduce the different state variables is a strong function of the physical processes by which the variables themselves are affected. As a result, a greater dimensionality is observed for the piloted flame, characterized by significant extinction. A local PCA (LPCA) algorithm, VQPCA, was also employed to reduce the effect of non-linear dependencies among state variables on the manifold dimension determination. By applying VQPCA, the data were divided into separated clusters, by using an unsupervised partitioning algorithm based on reconstruction error minimization. Also, results from VQPCA were compared with those obtained with an approach based on mixture fraction conditioning (FPCA). Both approaches are comparable for the CO/H₂/N₂ flame, while VQPCA performances are superior to FPCA for the CH₄ flame, due to its higher complexity.

© 2009 The Combustion Institute. Published by Elsevier Inc. All rights reserved.

Keywords: Manifold; Principal component analysis; PCA; Turbulent flames; Modeling

1. Introduction

A primary challenge in turbulent combustion modeling lies in the broad range of scales which are inherently coupled, in space and time, through thermo-chemical and fluid dynamic interactions. However, in some applications, it is possible to achieve some degree of decoupling, relying on the observation that many chemical timescales are significantly smaller than the fluid dynamic

scales and that the thermodynamic state of a reacting system relaxes onto a low dimensional, strongly attracting, manifold in chemical state space [1,2]. The thermo-chemical state of a single-phase reacting fluid having N_s species is uniquely characterized by a point in $N_s + 1$ dimensional space (T , p , and $N_s - 1$ species mass fractions). Yet, if a set of “optimal” variables is identified, the whole thermo-chemical state can be re-parameterized with a lower number, $q \ll N_s + 1$, of variables, which nevertheless must provide a satisfactory approximation of the system in a lower dimensional space [1].

In the present paper, a novel methodology based on principal component analysis (PCA)

* Corresponding author. Fax: +39 0502217866.

E-mail address: alessandro.parente@ing.unipi.it (A. Parente).

[3,4] is proposed with three main purposes: (i) to investigate the existence of low-dimensional manifolds in turbulent flames, (ii) to find the most compact representation for them and (iii) to guide the selection of “optimal” reaction variables able to accurately reproduce the state space of a reacting system. PCA has been previously applied to combustion. Frouzakis et al. [5] applied PCA for data reduction of two-dimensional DNS data of opposed jet flames. The analysis was aimed at identifying the number of components required to accurately approximate the original data. To this purpose, the correlations among velocities, pressure and species concentrations at different times were taken into account, thus leading to eigenvectors which are linear combination of the temporal snapshots considered. Similarly, Danby and Echehki [6] implemented PCA for the analysis of an unsteady two-dimensional direct numerical simulation of auto ignition in inhomogeneous hydrogen air mixtures, with the main purpose of determining the requirements to reproduce passive and reactive scalars during the process of auto ignition. The approach presented here is quite different from the ones described above. The main purpose of the developed PCA methodology is to find correlations among the state variables (temperature and species concentration) to allow an optimal approximation of the system in a low-dimensional space. Such an approach leads to the determination of eigenvectors which are linear combinations of the original variables in a way that allows reducing the dimension of the system. A similar method was proposed by Maas and Thévenin [7] for the analysis of DNS data. However, they only considered a very small sampling in state space. The current study provides significantly more depth in its analysis, and applies PCA to experimental datasets of laboratory-scale flames.

To demonstrate the methodology, two experimental data sets, provided under the framework of the TNF workshop [8], are considered. The first flame is a CO/H₂/N₂ turbulent jet flame [9], selected as base case for the PCA approach due to its simplicity. In fact, the flame does not experience any liftoff or localized extinction and retains the simple flow geometry of the hydrogen jet flames [8], while adding a modest level of chemical kinetic complexity. The second flame is Flame F, part of a series of four CH₄ piloted jet flames investigated by Barlow and Frank [10]. It was selected to judge PCA capabilities in flames close to global extinction.

The experimental data to be analyzed with PCA are organized in a $(n \times p)$ matrix, X . Each column of X represents a state variable while each row denotes an instantaneous spatial snapshot of the reacting species' concentrations and temperature. For the CO/H₂/N₂ flame, X consists of 66,275 observations of nine state variables (T, N₂, O₂, H₂O, H₂, CO, CO₂, OH, and NO) while,

for Flame F, 62,766 measurements of 10 state variables (T, N₂, O₂, H₂O, H₂, CH₄, CO, CO₂, OH, and NO) are available.

Three different PCA approaches are proposed. A global PCA analysis is compared to two local PCA (LPCA) models, vector quantization PCA (VQPCA) and mixture fraction PCA (FPCA). Both VQPCA and FPCA partition the data into separated clusters and then perform PCA locally. However, VQPCA partitioning is based on a reconstruction error minimization algorithm, while FPCA is based on a supervised conditioning into bins of mixture fraction. The performances of the three models are compared and discussed.

2. Principal component analysis

The central idea of principal component analysis is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. The reduction is achieved by transforming to a new set of variables, called the 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. The actual computation of the PCs reduces to an eigenvalue–eigenvector problem applied to the covariance matrix of the data set.

Consider a random sample, X , of n observations of p variables, $X = (x_1, x_2, \dots, x_p)$, so that $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ represents the i th observation from the data set. Here we assume, for simplicity, that the variables have zero mean, otherwise the mean of each variable is subtracted from the columns of X before PCA is applied. Then, the sample covariance matrix, S , of X can be defined as $S = 1/(n - 1) X'X$. Recalling the eigenvector decomposition of a symmetric, non-singular matrix, S can be decomposed as $S = ALA'$, where A is the $(p \times p)$ matrix whose columns are the eigenvectors of S , and L is a $(p \times p)$ diagonal matrix containing the eigenvalues of S in descending order, $l_1 > l_2 > \dots > l_p$.

The PCs, Z , are defined by the linear transformation $Z = XA$ which recast the original variables into a set of new uncorrelated variables, whose coordinate axes are described by A . Then, the original variables can be stated as a function of the PCs as $X = ZA'$, being A orthonormal and, hence, $A^{-1} = A'$. This means that, given Z , the values of the original variables can be uniquely recovered. However, the main objective of PCA is to replace the p elements of X with a much smaller number, q , of PCs, which nevertheless discard a small fraction of the variance originally contained in the data. If a subset of size $q \ll p$ is used, the truncated PCs are defined as $Z_q = X_q A_q$. This equation can be inverted to obtain:

$$\mathbf{X}_q = \mathbf{Z}_q \mathbf{A}'_q \quad (1)$$

where \mathbf{A}_q is the matrix obtained by retaining only the first q columns of \mathbf{A} .

The linear transformation provided by Eq. (1) is particularly appealing for size reduction in multivariate data analysis due to some optimal properties. First, the new coordinate axes (the eigenvectors of \mathbf{S}) are orthogonal and, when ordered by the descending values of their corresponding eigenvalues, provide independent and decreasing contributions to the amount of original variance explained by the PCs. Thus, if only the subset \mathbf{A}_q of \mathbf{A} is retained, \mathbf{X}_q represents the best q -dimensional approximation of \mathbf{X} in terms of squared prediction error.

It is crucial to determine how small q can be taken without serious information loss. The most used criterion for choosing q is to select a cumulative fraction of the total variance that the PCs have to account for, i.e. 0.8 or 0.9. The required number of PCs, q , is then the smallest value of q for which this chosen percentage is exceeded. It can be easily shown that the variance of \mathbf{X} , $\sum_{j=1}^p s_j^2$, is given by $\sum_{k=1}^p l_k$, where s_j is the standard deviation of x_j . Then, the desired fraction of the total variance, t_q , can be defined as $t_q = \sum_{k=1}^q l_k / \sum_{k=1}^p l_k$. It is also possible to define a measure of the individual variance, the variance accounted for each variable by the retained eigenvectors, as $t_{q,j} = \sum_{i=1}^q (a_{ji} \sqrt{l_i} / s_j)^2$ where a_{ji} is the weight of the j th variable on the i th eigenvector.

2.1. Data preprocessing strategies: centering and scaling

Data preprocessing strategies like centering and scaling may have a strong impact on PCA. Centering converts all the observations to fluctuations, thus leaving only the relevant variation for analysis. In the following, it is implicitly assumed that the data have been centered before the analysis. Scaling is essential when the elements of \mathbf{X} are in different units or when they have very different variances, i.e. temperature and species' concentrations (main species and radicals).

It is possible to define a scaled variable $\tilde{x}_j = x_j / d_j$ where d_j is the scaling parameter adopted for variable x_j . Several choices of d_j are available. The most common, denoted as "auto scaling", adopts the standard deviation of x_j as the scaling factor. After auto scaling, all the elements of \mathbf{X} have a standard deviation equal to one and, therefore, the data are analyzed on the basis of correlations instead of covariances, thus giving all the variables similar relevance. However, other scaling options are available. Gower and Hand [11] note the possibility of dividing the variables by their ranges or means. In the present study, on the basis of scoping simulations, it

was chosen to adopt the auto scaling criterion being the one providing the best results.

3. Local principal component analysis

The PCA transformation described in Section 2 can suffer from its reliance on second-order statistics. In fact, the PCs are uncorrelated, i.e. their second-order product moment is zero, but they can still be highly statistically dependent. This is particularly important when the relationships among the correlated variables are non-linear, as it usually happens for a reacting system. In this case, PCA fails to find the most compact description of the data and it usually requires a larger number of components to model the low-dimensional hyper plane embedded in the original space with respect to a non-linear technique.

This simple realization has prompted the development of non-linear alternatives to PCA, especially in the context of neural networks. Nevertheless, here we are more interested in a different approach, introduced by Kambhatla and Leen [12] in the field of images processing and known as local principal component analysis (LPCA). LPCA employs a local linear approach to reduce the statistical dependency between the variables of a sample and to achieve the desired optimal dimension reduction. According to LPCA, a vector quantization (VQ) algorithm first partitions the data space into disjoint regions and then PCA is performed in each cluster, 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. For this reason, Kambhatla and Leen [12] introduce a VQ algorithm based on a reconstruction error metric. Given an observation from the sample \mathbf{X} , \mathbf{X}_i , a global reconstruction error for each observation can be defined as:

$$\begin{aligned} \varepsilon_{\text{GRE}}(\mathbf{X}_i, \bar{\mathbf{X}}^{(k)}) &= \|\mathbf{X}_i - \mathbf{X}_{i,q}\| \\ &= \left\| \mathbf{X}_i - \left(\bar{\mathbf{X}}^{(k)} + \mathbf{Z}_{i,q} \mathbf{A}'^{(k)}_q \right) \right\| \quad (2) \end{aligned}$$

where $\bar{\mathbf{X}}^{(k)}$ 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 $\mathbf{A}'^{(k)}_q$ is the matrix obtained by retaining only the first q eigenvectors of the covariance matrix, $\mathbf{S}^{(k)}$, associated with the k th cluster. In the context of reacting systems, Eq. (2) needs to be modified to take into account the differences in size and units of the state variables. In fact, a clustering based on ε_{GRE} would lead to an optimization with respect to temperature only. Therefore, the original LPCA algorithm from Kambhatla and Leen [12] was modified to

include data preprocessing (Section 2.1) in the quantization scheme. A very stable algorithm was obtained by using a global scaled reconstruction error metric, $\varepsilon_{\text{GSRE}}$, defined as:

$$\varepsilon_{\text{GSRE}}(\tilde{\mathbf{X}}_i, \bar{\mathbf{X}}^{(k)}, \mathbf{D}) = \left\| \tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_{i,q} \right\|. \quad (3)$$

where $\tilde{\mathbf{X}}_i$ is the i th 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 summarized as follows:

1. *Initialization*: the cluster centroids, $\bar{\mathbf{X}}^{(k)}$, are randomly chosen from the data set and $\mathbf{S}^{(k)}$ is initialized to the identity matrix for each cluster.
2. *Partition*: each observation from the sample is assigned to a cluster using the metric $\varepsilon_{\text{GSRE}}$.
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 reconstruction quality given by VQPCA is measured with respect to the mean variance in the data as $\varepsilon_{\text{GSRE},m} = E(\varepsilon_{\text{GSRE}})/E[\text{var}(\tilde{\mathbf{x}}_j)]$, where E denotes the expectation operator and $\tilde{\mathbf{x}}_j$ is the scaled j th variable from $\tilde{\mathbf{X}}$.

The performances of VQPCA were compared with another LPCA algorithm, based on a supervised partitioning of the data into separated clusters. The sample is divided into bins of mixture fraction and a PCA analysis is performed in each cluster. The advantage of this approach, here denoted with FPCA, is twofold. First, the a priori knowledge of the partitioning variable simplifies the LPCA model implementation. Furthermore, the use of mixture fraction is particularly appealing due to its widespread use in the combustion field. In the following, the local PCA algorithms, VQPCA and FPCA, will be compared to the PCA analysis carried out on the whole data sets, i.e. taking $k = 1$, and denoted as GPCA.

4. Results and discussion

The results of the analysis performed on the CO/H₂/N₂ and CH₄ flames are presented here. There are two main objectives for the discussion. The first is to compare the performances of GPCA and VQPCA algorithms, in terms of the error metric, $\varepsilon_{\text{GSRE},m}$, defined in Section 3. Then, VQPCA is compared to FPCA to assess the feasibility of mixture fraction partitioning to optimize the PCA reduction.

4.1. Comparison of GPCA and VQPCA

In this section, the performances of GPCA and VQPCA are compared. The first two columns in Table 1 show the values of the total (t_q) and individual ($t_{q,j}$) variance accounted for the CO/H₂/N₂ flame with the GPCA reduction by keeping only the first two or three PCs. It is clear that the first two eigenvalues alone account for more than 92% of the total variance in the data. Moreover, Table 1 shows that, by choosing $q = 2$, it is possible to capture more than 90% of the individual variances of all the main species and temperature while the minor species (OH and NO), require an additional component, $q = 3$, to reach levels of approximation comparable to the other state variables. Thus, a strong size reduction, from 9 to 2 or 3, seems to be achievable with GPCA.

However, some questions arise from the analysis of the parity plots of Fig. 1a–c, which illustrate the GPCA reconstruction of temperature, OH and NO, respectively, by choosing $q = 2$. Significant deviations are observed for minor species like OH (Fig. 1b) and NO (Fig. 1c). In addition, even variables like temperature (Fig. 1a), which are well characterized in terms of $t_{q,j}$, show non-linear deviations in the reconstructed data. If the number of PCs retained is increased to three, some improvements are observed for OH and NO (Table 1) while the other species and temperature are relatively unaffected, as indicated by the corresponding values of $t_{q,j}$ (Table 1). Moreover, the same deviations observed with $q = 2$ remain. This result can be explained by ascribing the non-linearities in the GPCA reconstruction to non-linear dependencies among the original variables, thus confirming the need of local algorithms able to cluster the data into locally linear regions.

Table 1

Total, t_q , and individual variance, $t_{q,j}$, accounted for the CO/H₂/N₂ and CH₄ flames by the GPCA reduction, as a function of the number of retained PCs, q

	$t_{q,j}$ (%)			
	CO/H ₂ /N ₂		CH ₄	
	$q = 2$	$q = 3$	$q = 3$	$q = 4$
T	97.1	97.2	96.7	97.1
Y_{O_2}	98.7	98.7	97.8	97.9
Y_{N_2}	98.6	98.6	97.9	98.0
Y_{H_2}	96.8	96.9	96.9	97.0
$Y_{\text{H}_2\text{O}}$	93.1	93.6	98.3	98.4
Y_{CH_4}	–	–	98.4	98.4
Y_{CO}	99.4	99.4	96.1	96.9
Y_{CO_2}	97.3	97.7	96.9	97.4
Y_{OH}	73.8	94.0	71.1	97.8
Y_{NO}	77.2	93.0	79.2	89.2
t_q (%)	92.4	96.6	92.9	96.8

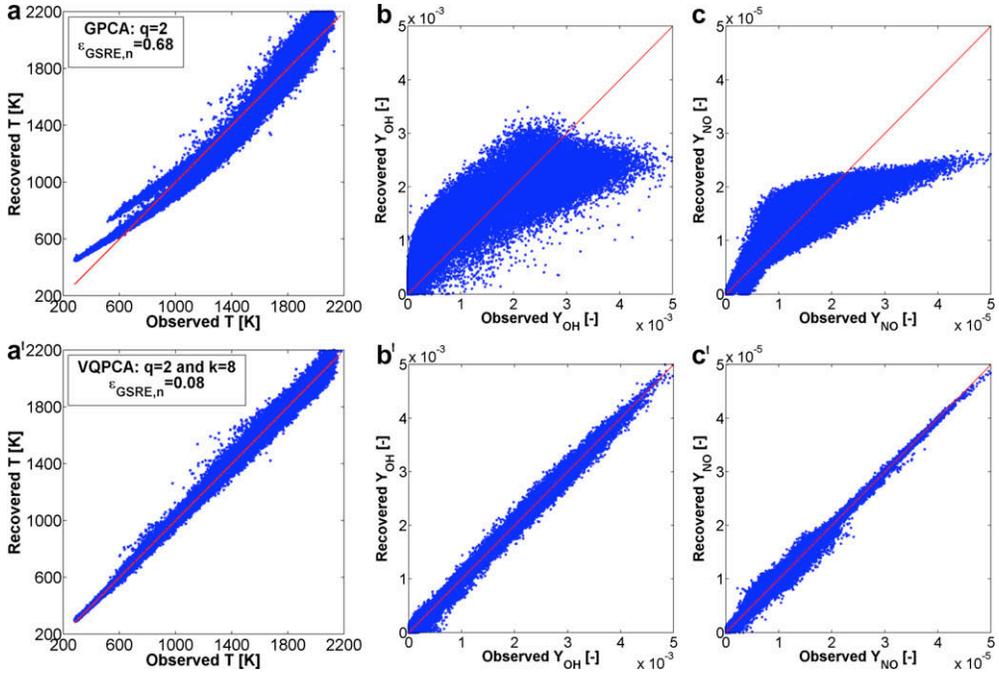


Fig. 1. Parity plots of temperature (a, a'), OH (b, b') and NO (c, c') mass fractions illustrating the GPCA ($q = 2$) and VQPCA ($q = 2$ and $k = 8$) reductions of the CO/H₂/N₂ flame, respectively.

Table 2

Values of $\epsilon_{GSRE,n}$, associated with the GPCA, VQPCA and FPCA reconstructions of the CO/H₂/N₂ and CH₄ flames, as a function of the number of clusters, k , and retained PCs, q

	k	CO/H ₂ /N ₂				CH ₄			
		$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 3$	$q = 4$	$q = 5$	$q = 6$
GPCA	1	0.6805	0.3085	0.0975	0.0459	0.7068	0.3196	0.1243	0.0662
VQPCA	2	0.2079	0.1062	0.0492	0.0254	0.2046	0.1189	0.0574	0.0228
	4	0.1122	0.0556	0.0246	0.0091	0.1313	0.0756	0.0311	0.0115
	6	0.0907	0.0464	0.0217	0.0083	0.0949	0.0524	0.0257	0.0091
	8	0.0790	0.0338	0.0160	0.0059	0.0898	0.0398	0.0190	0.0069
FPCA	2	0.2138	0.0841	0.0467	0.0199	0.2632	0.1466	0.0670	0.0335
	4	0.1205	0.0659	0.0324	0.0121	0.1583	0.0865	0.0409	0.0179
	6	0.1032	0.0512	0.0270	0.0096	0.1335	0.0673	0.0347	0.0146
	8	0.0921	0.0448	0.0244	0.0077	0.1217	0.0631	0.0323	0.0133

Table 2 lists the values of $\epsilon_{GSRE,n}$ given by GPCA, VQPCA and FPCA for the CO/H₂/N₂ and CH₄ flames, as a function of the number of clusters, k , and retained PCs, q .

It is interesting to observe (Table 2) that, when the reconstruction error is evaluated on a scaled basis, all the state variables become relevant and the quality of the reconstruction can be properly evaluated. We should recall here that $\epsilon_{GSRE,n}$ is a global scaled reconstruction error, normalized by the mean variance present in the original data. So, for example, the value of $\epsilon_{GSRE,n}$ associated with the GPCA reduction of the CO/H₂/N₂ flame

with $q = 2$ is fairly large, $\epsilon_{GSRE,n} = 0.68$, thus reflecting the large deviations observed in Fig. 1 for some state variables. Even when q is increased to three, a significant error, $\epsilon_{GSRE,n} = 0.31$, is obtained, confirming the persistence of mainly non-linear departures from the original data.

However, when VQPCA is employed, much higher performances in terms of $\epsilon_{GSRE,n}$ are obtained when increasing k , even for smaller values of q , i.e. $q = 2$. The parity plots in Fig. 1a'–c' shows the VQPCA reconstruction of temperature, OH and NO, respectively, with $k = 8$ and $q = 2$. A much better agreement between original

and reconstructed data are observed, as it is confirmed by the value of 0.08 obtained for $\varepsilon_{\text{GSRE},n}$. A similar value of $\varepsilon_{\text{GSRE},n}$ would require $q = 5$ if GPCA is applied (Table 2). Moreover, Fig. 1a'–c' shows how, after partitioning, the relationships between the original and reconstructed data are mostly linear.

The third and fourth columns in Table 1 list the values of t_q and $t_{q,j}$ accounted for the CH_4 flame with the GPCA reduction by keeping the first three or four eigenvalues. It can be observed that the first three eigenvalues explain 93% of the variance in the original variables. Also, the higher requirements of minor species, such as OH and NO, in terms of number of PCs is confirmed by the corresponding values of $t_{q,j}$. Moreover, similarly to the $\text{CO}/\text{H}_2/\text{N}_2$ flame, strong non-linear deviations in the reconstructed data are observed for the case $q = 3$ (Fig. 2a–c). The reconstruction error, $\varepsilon_{\text{GSRE},n}$, associated with the GPCA reduction equals 0.71, thus confirming the inability of GPCA to determine the most compact description of the data in a low-dimensional space. Figure 2a'–c' illustrates the VQPCA reconstruction of temperature, OH and NO, respectively, with $k = 8$ and $q = 3$. The value of $\varepsilon_{\text{GSRE},n}$ obtained is 0.09, almost eight times smaller than the values given by GPCA with $q = 3$. Again, GPCA would require 6 PCs to achieve a value of $\varepsilon_{\text{GSRE},n}$ as small as 0.09 (Table 2).

4.2. Comparison of VQPCA and FPCA

Rows 6–9 in Table 2 list the values of $\varepsilon_{\text{GSRE},n}$ given by FPCA for the $\text{CO}/\text{H}_2/\text{N}_2$ and CH_4 flames, as a function of the number of clusters, k , and retained PCs, q . For $k = 1$, the value of $\varepsilon_{\text{GSRE},n}$ associated with the GPCA reconstruction is obtained. These values can be weighed against those obtained with the VQPCA algorithm (rows 2–5 in Table 2) for the same values of q and k . The comparison is illustrated graphically in Fig. 3.

For the $\text{CO}/\text{H}_2/\text{N}_2$ flame, VQPCA generally performs better than FPCA. In most cases, the values of $\varepsilon_{\text{GSRE},n}$ given by VQPCA are 3–25% lower than those provided by FPCA with the exception of the cases corresponding to $k = 2$ and $q = 3$. However, the performances of FPCA are quite promising, especially from the point of view of model implementation. In fact, FPCA partitioning is much simpler and straightforward than the one underlying VQPCA. Moreover, it should be underlined here how the CPU time associated with VQPCA increases with k , reaching values of the order of minutes when $k = 8$ on an Intel Core 2 Duo 2.2 GHz with 2GB RAM. On the same machine, the GPCA and FPCA reconstruction of the same data requires CPU times of the order of seconds. This result suggests that, for the $\text{CO}/\text{H}_2/\text{N}_2$ flame, the mixture fraction can be considered an “optimal” variable for the

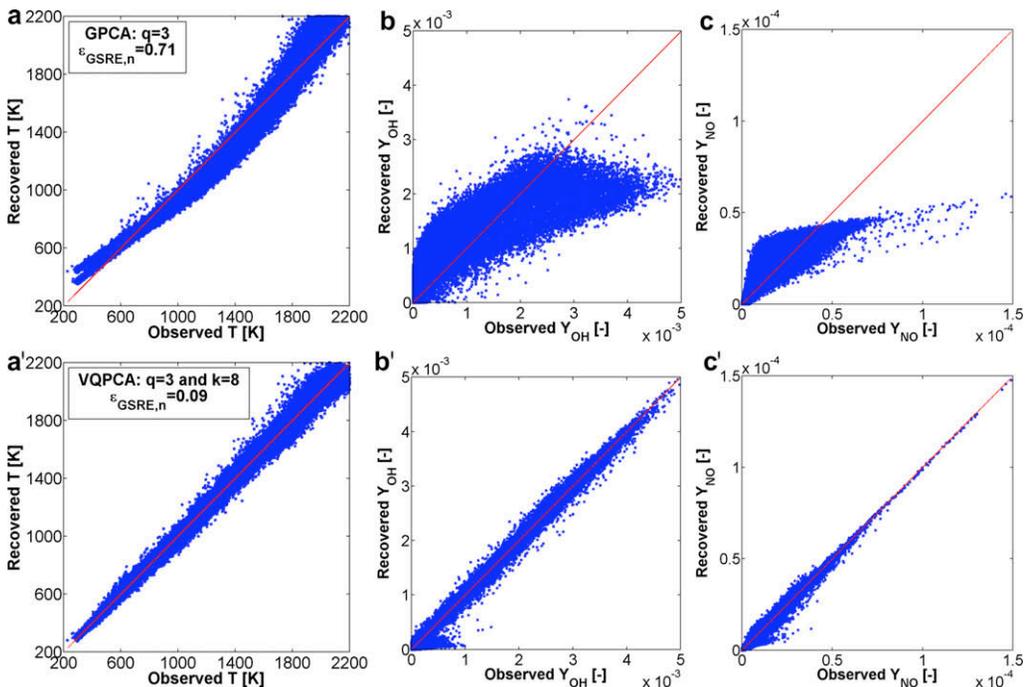


Fig. 2. Parity plots of temperature (a, a'), OH (b, b') and NO (c, c') mass fractions illustrating the GPCA ($q = 3$) and VQPCA ($q = 3$ and $k = 8$) reductions of the CH_4 flame, respectively.

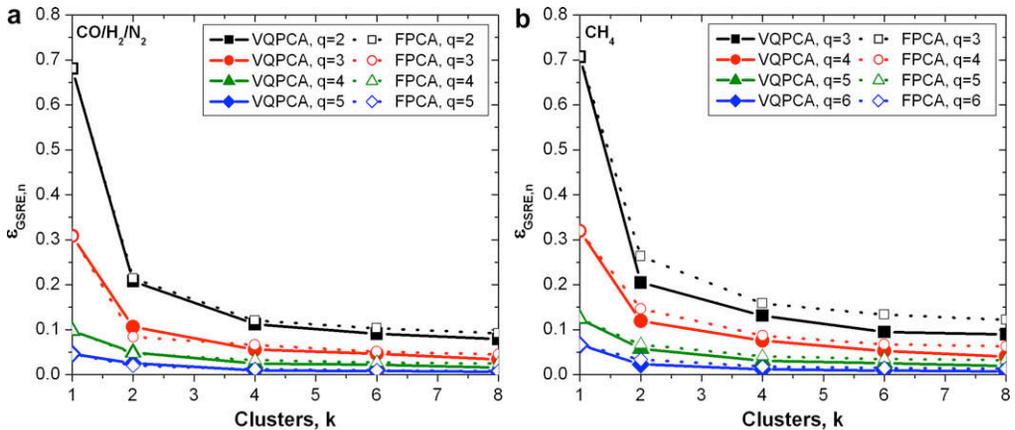


Fig. 3. Comparison of the values of $\epsilon_{GSRE,n}$ obtained with VQPCA and FPCA for the $CO/H_2/N_2$ (a) and CH_4 (b) flames as a function of the number of clusters, k , and retained PCs, q .

parameterization of the thermo-chemical state of the system, as it is generally assumed in many models for non-premixed combustion. The information added here is that mixture fraction is an “optimal” variable from the point of view of reconstruction error minimization.

As far as the CH_4 flame is concerned, Table 2 and Fig. 3 point out that VQPCA outperforms FPCA in all cases, providing values of $\epsilon_{GSRE,n}$ 13–40% lower than those given by FPCA. These results confirm the notorious complexity of this flame, characterized by significant local extinction and re-ignition. In the context of the Conditional Moment Closure method [13], for example, it has been recognized [14] that conditioning on mixture fraction is not sufficient for Flame F and a second conditioning variable should be used. Then, it can be concluded that, for Flame F, mixture fraction does not represent an “optimal” reaction variable. Therefore, VQPCA could provide an appealing tool to guide the selection of the most compact subset of reaction variables needed to properly describe the thermo-chemical state of a reacting system.

5. Concluding remarks

In the present paper, a novel methodology based on principal component analysis (PCA) is proposed for the identification of low-dimensional manifolds in turbulent flames, the estimation of their dimensionality and the selection of optimal reaction variables. To this purpose, two high fidelity experimental data sets, corresponding to a $CO/H_2/N_2$ turbulent jet flame and a piloted CH_4 flame were investigated.

Three different PCA approaches are proposed. A global PCA analysis, GPCA, is compared to

two local PCA models, VQPCA and FPCA, based on the partitioning of the experimental data into separate clusters where PCA is performed locally. However, the partitioning algorithm used by VQPCA is unsupervised and based on reconstruction error minimization while FPCA conditions the data a priori on the mixture fraction.

Results show that the local PCA approaches (VQPCA and FPCA) outperform the global approach in all cases. Indeed, GPCA is unable to provide a compact representation of the data in a low-dimensional space due to the highly non-linear relationships existing among the state variables.

Regarding the local approaches, the performances of VQPCA and FPCA are comparable for the $CO/H_2/N_2$ flame, while for Flame F, mixture fraction proves unable to capture some important features of the system, resulting in higher reconstruction error with respect to VQPCA.

The reduced representation given by PCA has great potential, especially in its local formulations, i.e. VQPCA and FPCA. In fact, the selection of “optimal” variables for turbulent reacting systems could be exploited for the development of turbulence–chemistry interactions models. In this context, the linearity of the PCA method is extremely appealing. In fact, if a set of reaction variables is selected, only few linear combinations of the original variables need to be transported in a numerical simulation. Nevertheless, since the reaction variables provided by PCA are not conserved scalars, the ability of the PCs to parameterize their source should be investigated a priori, to assess the feasibility of the generated manifold method. To this end, the availability of the source terms for the state variables becomes crucial, thus forcing to move from experimental to numerical, i.e. DNS, data analysis.

References

- [1] U. Maas, S.B. Pope, *Proc. Combust. Inst.* 24 (1992) 103–112.
- [2] J.C. Sutherland, P.J. Smith, J.H. Chen, *Combust. Theor. Model.* 11 (2007) 287–303.
- [3] I.T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, 1986.
- [4] J.E. Jackson, *A User's Guide to Principal Components*, Wiley, New York, NY, 1991.
- [5] C.E. Frouzakis, Y.G. Kevrekidis, J. Lee, K. Boulouchos, A.A. Alonso, *Proc. Combust. Inst.* 28 (2000) 75–81.
- [6] S.J. Danby, T. Echekeki, *Combust. Flame* 144 (2006) 126–138.
- [7] U. Maas, D. Thévenin, *Proc. Comb. Inst.* 27 (1998) 1183–1189.
- [8] TNF Workshop, available at <http://www.ca.sandia.gov/>.
- [9] R.S. Barlow, G.J. Fiechtner, C.D. Carter, J.-Y. Chen, *Combust. Flame* 120 (2000) 549–569.
- [10] R.S. Barlow, J.H. Frank, *Proc. Combust. Inst.* 27 (1998) 1087–1095.
- [11] J.C. Gower, D.J. Hand, *Biplots*, Chapman & Hall, London, 1996.
- [12] N. Kambhatla, T.K. Leen, *Neural Comput.* 9 (1997) 1493–1516.
- [13] A.Y. Klimenko, R.W. Bilger, *Prog. Energy Combust. Sci.* 25 (1999) 595–687.
- [14] M.R. Roomina, R.W. Bilger, *Combust. Flame* 125 (2001) 1176–1195.