Combustion and Flame 161 (2014) 2785–2800

Contents lists available at ScienceDirect

Combustion and Flame

journal homepage: www.elsevier.com/locate/combustflame

Reduced-order PCA models for chemical reacting flows

Benjamin J. Isaac [b,a,1], Axel Coussement [b,c,d], Olivier Gicquel [c,d], Philip J. Smith [a], Alessandro Parente [b,*]

[a] Department of Chemical Engineering, University of Utah, Salt Lake City, UT 84112, USA
[b] Service d'Aéro-Thermo-Mécanique, Université Libre de Bruxelles, Bruxelles, Belgium
[c] Ecole Centrale Paris, Grande Voie des Vignes, 92295 Chatenay-Malabry, France
[d] CNRS, UPR 288 Laboratoire d'energetique moleculaire et macroscopique, combustion, Grande Voie des Vignes, 92295 Chatenay-Malabry, France

ARTICLE INFO

Article history:
Received 23 October 2013
Received in revised form 7 February 2014
Accepted 7 May 2014
Available online 13 June 2014

Keywords:
Dimensionality reduction
Low-dimensional manifolds
Principal component analysis
Reduced-order models
Turbulent combustion modeling

ABSTRACT

One of the most challenging aspects of turbulent combustion research is the development of reduced-order combustion models which can accurately reproduce the physics of real systems. The identification and utilization of the low dimensional manifolds in these system is paramount to understand and develop robust models which can account for turbulence–chemistry interactions. Recently, principal components analysis (PCA) has been given notable attention in its analysis of reacting systems, and its potential in reducing the number of dimensions with minimum reconstruction error. The present work provides a methodology which has the ability of exploiting the information obtained from PCA. Two formulations of the approach are shown: Manifold Generated from PCA (MG-PCA), based on a global analysis, and Manifold Generated from Local PCA (MG-L-PCA), based on performing the PCA analysis locally. The models are created using the co-variance matrix of a data-set which is representative of the system of interest. The reduced models are then used as a predictive tool for the reacting system of interest by transporting only a subset of the original state-space variables on the computational grid and using the PCA basis to reconstruct the non-transported variables. The present study first looks into the optimal selection of the subset of transported variables and analyzes the effect of this selection on the approximation of the state space and chemical species source terms. Then, a demonstration of various *a posteriori* cases is presented.

1. Introduction

It is well established that the ability to model industrial combustion systems is dependent on the ability to represent the reaction system with a reduced number of parameters. Literature enumerates numerous approaches to reduce the computational cost associated to turbulent combustion problems. Several methods are based on the parameterization of the state-space with a reduced number of optimal variables. This leads to fewer transport equations, and provides a reduction in computation time. Several examples include: Steady Laminar Flamelet Method (SLFM) [1,2], Flamelet-Generated Manifold (FGM) [3,4] and Flamelet-Prolongation of ILDM model (FPI) [5,6]. Alternatively, many models attempt to simplify the chemistry in the system using equilibrim assumptions, such as quasi steady state approximation (QSSA) [7], intrinsic low-dimensional manifolds (ILDM) [8], or rate controlled constrained-equilibrium (RCCE) [9,10].

Recent work has been pushing for the development of a new class of models which are entirely based on empirical data-sets. The concept is to use principal components analysis (PCA) on empirical data-sets to identify a low dimensional representation of the reacting system. Previous work by Maas and Thévenin [11] applied PCA to premixed DNS cases, to identify correlations between species concentrations. In the work by Parente et al. [12,13], PCA was used to identify the best linear representation of the underlying manifold contained in these highly coupled reacting systems. Biglari and Sutherland [14] and Pope [15] extended such a concept using the PCA basis in conjunction with non-linear regression, maximizing the size reduction for a given accuracy. Mirgolbabaei and Echekki [16] extended such an analysis to the application of artificial neural networks, showing also the effect of minor species on the accuracy in the reconstruction. Finally, Mirgolbabaei and Echekki [17] investigated the potential of kernel PCA, showing the high compression potential derived by transforming the initial problem into a non-linear featured

* Corresponding author. Address: Avenue F.D. Roosevelt 50, 1050 Bruxelles, Belgium.
E-mail addresses: Benjamin.J.Isaac@utah.edu (B.J. Isaac), axcousse@ulb.ac.be (A. Coussement), Olivier.Gicquel@ecp.fr (O. Gicquel), Philip.Smith@utah.edu (P.J. Smith), Alessandro.Parente@ulb.ac.be (A. Parente).
[1] Address: 155 South 1452 East, Room 350, Salt Lake City, UT 84112, USA.

http://dx.doi.org/10.1016/j.combustflame.2014.05.011
0010-2180/

space where linear PCA is carried out. These works show the capability of PCA to recover a highly accurate reconstruction of the state space variables with a significant dimension reduction. This indicates that a lower dimensional manifold exists in turbulent reacting systems, and that PCA is well suited for identifying the manifold.

Various approaches have been developed in order to use the manifold identified by PCA. The PC-score approach was first described by Sutherland and Parente [18] as a model which transports the principal components (PCs) directly. Several other groups have proposed transporting a subset of state space variables and reconstructing the non-transported variables from the PC basis [19,20]. In particular, the Manifold Generated from PCA model (MG-PCA) by Coussement et al. [19] was the first PCA model for which *a posteriori* validation was provided, by computing a hydrogen flame-vortex interaction using a DNS solver. Finally, in the work by Najafi-Yazdi et al. [21], PCA was used to identify optimal progress variables in the context of the flamelet generated manifold approach.

The present paper focuses on the use of PCA for combustion models. Primarily, the MG-PCA model [19] is examined. In this model transport equations are solved for a subset of the originally transported variables which contain most of the variance of a reacting system. The remaining variables are reconstructed from the PCA basis which has been calculated *a priori*. The current work aims at extending the analysis of the MG-PCA method, investigating various *a priori* aspects of the MG-PCA models, and showing several *a posteriori* demonstrations of the model. The *a priori* investigation shows the improvement in accuracy provided by the new model formulation, the effect of the various transported variables selection methods and pre-processing techniques on size reduction, as well as on reconstruction of state space variables and source terms. Next the MG-PCA method is demonstrated through several different configurations including: auto-ignition delay times, laminar flame speed, flame-vortex interaction, and flame-turbulence interaction.

2. Principal component analysis

For a data-set $\mathbf{X}$ $(n \times Q)$, containing n samples of Q original variables, PCA provides an approximation of the original data-set using only q $(q < Q)$ linear correlations between the Q variables [12,18]. PCA starts with the computation of the sample co-variance matrix $\mathbf{S}$:

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

where the superscript T indicates the transpose matrix. Using the spectral decomposition, $\mathbf{S}$ is then decomposed:

$$\mathbf{S} = \mathbf{A}\,\mathbf{L}\,\mathbf{A}^T \tag{2}$$

where $\mathbf{A}(Q \times Q)$ and $\mathbf{L}(Q \times Q)$ are respectively the Q eigenvectors of $\mathbf{S}$, called principal components (PCs), and the eigenvalues of $\mathbf{S}$, in decreasing order. The principal component scores, $\mathbf{Z}(n \times Q)$, are then computed using the eigenvector matrix as:

$$\mathbf{Z} = \mathbf{XA}. \tag{3}$$

One of the main advantage of PCA is that the original set of data $(\mathbf{X})$ can be uniquely recovered using the PCs and their associated scores:

$$\mathbf{X} = \mathbf{ZA}^{-1} \tag{4}$$

where it should be noted that $\mathbf{A}^{-1} = \mathbf{A}^T$. However, the main objective of PCA is dimension reduction. Indeed, if one only uses the first q PCs $(q < Q)$, an approximation of $\mathbf{X}$ based on the first q eigenvectors $(\mathbf{X}_q)$ is obtained:

$$\mathbf{X} \approx \mathbf{X}_q = \mathbf{Z}_q\mathbf{A}_q^T \tag{5}$$

where $\mathbf{X}_q$ is the approximation of $\mathbf{X}$ based on the first q eigenvectors of $\mathbf{S}$, and $\mathbf{Z}_q$ is the $n \times q$ matrix of the principal component scores.

Finally, it should be stressed that the data is pre-processed prior to performing PCA. In particular, each variable of the original data-set $\mathbf{X}$ is centered and scaled in order to increase the accuracy of the method [12,22]. Applying centering and scaling on the data-set reads:

$$\mathbf{X}^s = (\mathbf{X} - \overline{\mathbf{X}})\mathbf{D}^{-1} \tag{6}$$

where $\overline{\mathbf{X}}$ is a $n \times Q$ matrix containing the mean of each variable and $\mathbf{D}$ is a $Q \times Q$ matrix containing the standard deviation of each variable (see [22] for details).

Using the previous analysis two general classes of PCA based combustion models have been identified:

- First, one can directly transport the principal components, as proposed in the work by Sutherland and Parente [18]. The thermo-chemical state space is then recovered using Eq. (5). While the approach is straightforward, it suffers from a major drawback related to the PC source terms. In particular, the error associated to the PCA reconstruction strongly affects the calculation of the source terms, whose accuracy degrades quickly when reducing the number of parameters defining the manifold. This is due to the fact that PCA evenly distributes the reconstruction error on the state variables, 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 uncontrolled propagation of error when the source term are calculated from the approximated state-space. Therefore, non-linear regression techniques such as MARS are being used to parameterize the full thermo-chemical state [14,15].
- Second, one can transport a subset of the original variables and recover the remaining variables using the information from PCA (MG-PCA). The MG-PCA approach was developed [19] to better control the propagation of the reconstruction error. Such approach is based on the resolution of classic transport equations for the system principal variables. Indeed, Eq. (5) indicates that $\mathbf{X}_q$ can be obtained from $\mathbf{Z}_q$. Moreover, those scores can be approximated from a subset of the original variables $\mathbf{X}(q)$ of size $n \times q$ composed of only q variables:

$$\widetilde{\mathbf{Z}}_q = \mathbf{X}(q)\,\left(\mathbf{A}(q)_q^T\right)^{-1} \tag{7}$$

where $\mathbf{A}(q)_q$ is a $(q \times q)$ matrix containing only the coefficients related to the q retained variables. Combining Eqs. (5) and (7) one finds:

$$\mathbf{X}_q = \mathbf{X}(q)\left(\mathbf{A}(q)_q^T\right)^{-1}\mathbf{A}_\mathbf{q}^\mathbf{T} \tag{8}$$

or

$$\mathbf{X}_q = \mathbf{X}(q)\mathbf{B} \tag{9}$$

defining the matrix $\mathbf{B}$ $(q \times Q)$ as:

$$\mathbf{B} = \left(\mathbf{A}(q)_q^T\right)^{-1}\mathbf{A}_\mathbf{q}^\mathbf{T}. \tag{10}$$

Therefore by transporting q variables (which can be temperature or species mass fractions) it is possible to recover the $(Q - q)$ remaining variables by retaining the appropriate $(Q - q)$ columns of the $\mathbf{B}$

matrix in Eq. (9), corresponding to the non-transported $(Q-q)$ state variables:

$$\mathbf{X}_q(Q-q) = \mathbf{X}(q)\mathbf{B}(Q-q) \tag{11}$$

By comparison with the score approach, this method requires the *a priori* selection of q transported variables. This method allows to better control the propagation of the error linked to the model, which is the major advantage of MG-PCA.

MG-PCA also allows the use of PCA locally [23]. Parente et al. [12] first applied the local PCA formulation to turbulent combustion data, identifying the limitations of global PCA for the analysis of highly non-linear systems such as the ones observed in combustion. In fact, PCA tries to approximate the non-linear chemical manifold by superimposing several linear effects, resulting in a manifold size higher than the actual problem dimensionality. To avoid such a problem the local PCA approach was proposed to optimally partition the data into clusters, based on an iterative algorithm which minimized the reconstruction error of the state space. However, the implementation of local PCA in terms of a combustion model does not appear straightforward for two main reasons: first, the approach is based on the resolution of transport equations for the scores, implying a modification of the PC definition with the cluster, and, second, the conditioning variable is not known *a priori* and it is not guaranteed that it could be somehow related any state variable.[2] On the other hand, the use of local PCA appears well suited in the MG-PCA context, as indicated in Coussement et al. [19]. The main steps of the approach, briefly indicated as MG-L-PCA, are:

- The principal variables are extracted from the full data-set, to define the transport equations which need to be resolved in all identified clusters.
- Then, the matrices $\mathbf{A}_q$ and $\mathbf{B}$ are computed in each cluster, allowing an optimal local reconstruction of the non-transported variables using Eq. (5).

Differently from Parente et al. [12], the conditioning variables are chosen *a priori* to build continuous clusters on the basis of a progress variable displaying a monotonic increase throughout the flame. For premixed cases, as the ones described in the present paper, temperature represents an optimal choice[3] [19].

The MG-PCA algorithm can be divided into two parts:

- First, the data-set $\mathbf{X}$ is generated using a "canonical reactor" with the same chemical composition of the system to be simulated. Obviously, the data-sets should be simple to compute, in order to generate combustion models of tailored-accuracy with affordable computational resources. In the present work, one-dimensional premixed flames are used for several of the *a posteriori* examples; however, for non-premixed systems, steady laminar flamelets [24] with varying strain-rate could be used.
- A principal component analysis is then performed to identify the manifold. The $\mathbf{B}$ matrix is computed and the subset of q retained variables is identified. Note that if the local formulation is used one has to identify the clusters and compute their corresponding $\mathbf{B}$ matrices.

Then the database is used in the flow solver:

- Transport equations are solved for the q non-conserved scalars.
- At the end of each temporal (or pseudo-temporal) iteration, the missing $(Q-q)$ variables are reconstructed using the $\mathbf{B}$ matrix (Eq. (11)) at every grid point.
- All the species are then available for the next temporal (or pseudo temporal) iteration. The diffusion and source terms appearing in the conservation equations of the q retained variables, are computed using a CHEMKIN like [25] formalism for all the species (retained and recovered).

The $\mathbf{B}$ matrix and the coefficients used to center and scale the data are constants.[4] Therefore, this algorithm requires only one matrix–vector multiplication. The computational cost to recover non-transported species using the MG-PCA technique is low.

3. Challenges of the MG-PCA model

Before applying the MG-PCA model to actual computations several issues must be carefully addressed:

- A common issue in PCA based models is the need for a data-set which represents the system of interest. The data-set also needs to be easy to compute (e.g. 1-D flames). By generating a data-set which is representative of the system PCA can indeed construct a model which accounts for non-intrinsic and intrinsic physics in the system. However, PCA does not distinguish the non-intrinsic effects, accordingly caution is required in selecting or generating a data-set for the system of interest.
- The accuracy of the model will obviously depend on the accuracy of the reconstruction of the missing variables. Since the method relies on the matrix $\mathbf{B}$ for the reconstruction of the non-transported variables, its computation must be as accurate as possible. The use of Eq. (10) does not provide satisfactory results, as it will be shown below, and an alternative optimal estimation of $\mathbf{B}$ must be provided for the success of the method.
- The selection of the transported variables is crucial requiring that the most informative variables in the data-sets are selected. Moreover, the number of transported variables, q, must be appropriately selected to ensure that the source terms are accurately reconstructed for the reduced set of scalar transport equations.
- The scaling and centering coefficients (see Eq. (6)) have a great impact on the accuracy of the method and they must be chosen with care.

3.1. Reference data-set

In order to address the issues presented above, an *a priori* demonstration is now provided to cover the aforementioned aspects using a DNS flame-turbulence data-set. The compressible flow solver YWC, developed at the EM2C Laboratory by Coussement et al. [26,27] is used to generated this data-set. The 2-D flame turbulence field is initialized using a 1-D flame extended along the y-axis and by super-imposing a turbulence field. The turbulence field is initialized using the Passot–Poquet spectrum [28] with $Re_t = 1423$ giving a Kolmogorov length scale of $l_k = 8.2 \cdot 10^{-6}$. The initial turbulence field is shown in Fig. 1. The computational domain extends $8 \cdot 10^{-3}$ m in both x and y directions, and mesh spacing is $1.25 \cdot 10^{-6}$ m. The fuel considered is syngas (CO/H_2 mixture, 50/50 molar basis), the oxidizer is air and the equivalence ratio is $\phi = 0.88$. Boundary conditions consist of an inlet at $x = x_{min}$ which imposes the turbulent field generated with a convection velocity of 4 m/s along the x-axis. At $x = x_{max}$ an outlet boundary

[2] In the context of non-premixed flames, it was shown that the conditioning variable corresponded quite well to mixture fraction [12].

[3] For non-premixed flames, mixture fraction would represent an optimal choice, as indicated in [12].

[4] Note that if local-PCA is used the coefficient for centering and scaling along with the $\mathbf{B}$ matrix are dependent on the cluster which must first be identified [19].

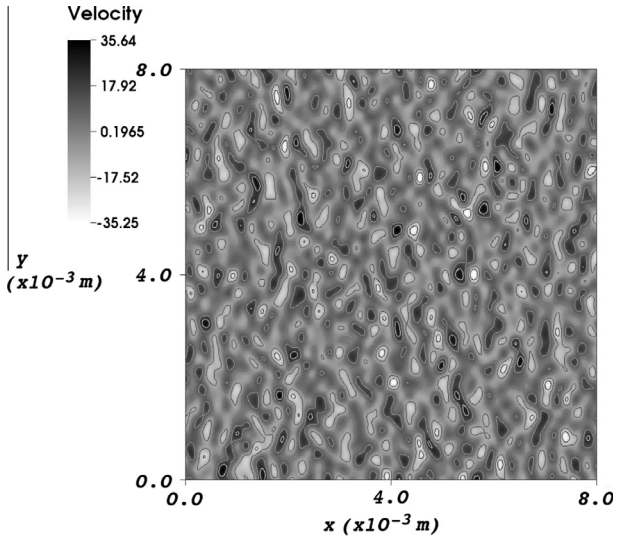

Fig. 1. Initial turbulence field for the 2-D DNS, x-component velocity (m/s).

is imposed with $p = 101,325$ Pa. Remaining boundary conditions ensure periodicity along the y-axis. The chemical scheme is the one from Davis et al. [29] including 12 species ($N_s = 12$) : $N_2, O_2, H_2, H_2O, H_2O_2, CO, CO_2, O, H, OH, HO_2$ and HCO. Thermo-chemical and transport properties are computed using a CHEMKIN-like formalism [25]. The DNS field used to perform the analysis below is taken at $t = 7.1241 \cdot 10^{-4}$ s. Mass fraction of H_2O (Y_{H_2O}) and HCO (Y_{HCO}) are given in Fig. 2.

It is important to note that differential diffusion is used. With $N_s = 12$ and only a constraint on mass conservation (here energy and elemental mass fractions are not constant) there are N_s degrees of freedom, because pressure is also constant ($p = p_{atm}$). In principle a reduction of the state-space dimensionality is achieved if the proposed methods allow to transport less than $q = 12$ variables (if energy is included in the PCA analysis), this can be further verified by analyzing the conservative modes of the system, which is shown later in Section 4.

3.2. Computation of the **B** matrix

The computation of the **B** matrix using Eq. (10) is not optimal. To illustrate this we revisit the definition of the scores computed using Eq. (7):

$$\widetilde{\mathbf{Z}}_q = \mathbf{X}(q)\ \left(\mathbf{A}(q)_q^T\right)^{-1}$$

Depending on the number of retained principal variables contributing to the definition of $\mathbf{A}(q)_q$, the approximated scores ($\widetilde{\mathbf{Z}}_q$) can result in a weak representation of the state space. This constrains the achievable reduction when Eq. (8) is employed. However, being that the MG-PCA model is built *a priori*, this limitation can be overcome by considering the real q principal component scores, obtained using the first q components of Eq. (3):

$$\mathbf{Z}_q = \mathbf{X}\mathbf{A}_q. \tag{12}$$

Here all of the state space variables in **X** as well as all of the weights in the first q eigenvectors are used instead of a subset. Thus, by using $\mathbf{Z}_q$ instead of $\widetilde{\mathbf{Z}}_q$ there should be a general decrease in reconstruction error in the non-transported species. In order to reflect this change, the calculation of **B** is performed using:

$$\mathbf{B} = \mathbf{X}(\mathbf{q})^{+}\mathbf{Z}_q\mathbf{A}_\mathbf{q}^\mathbf{T} \tag{13}$$

which represents the solution in the least squares sense to the system given by Eq. (12). The matrix $\mathbf{X}(\mathbf{q})^+$ is the pseudo inverse of $\mathbf{X}(\mathbf{q})$ of size $q \times n$, which is given by

$$\mathbf{X}(\mathbf{q})^{+} = \left(\mathbf{X}(\mathbf{q})^T\ \mathbf{X}(\mathbf{q})\right)^{-1}\mathbf{X}(\mathbf{q})^T. \tag{14}$$

The two methods for the calculation of **B** are compared in Fig. 3 with reference to the 2-D DNS data-set. The approximation error is reported using a normalized root mean square distance identical to that used by Pope [15], defined as:

$$nrms_j = \frac{1}{std(\phi_{ij})}\left(\left(\frac{1}{N}\sum_{i=1}^{n}(\phi_{ij} - \tilde{\phi}_{ij})\right)^2\right)^{1/2} \tag{15}$$

The *nrms* error statistic for the non-transported state variables are calculated, and the largest *nrms* value is plotted. The figure illustrates the improvement in accuracy.

3.3. Selection of the transported variables

In the current approach, the method used to select the q transported variables relies on the principal variables concept [30]. Principal components (PC) are linear combinations of all the variables defining the data-set. However, these variables are not necessarily equally important to the formation of the PCs. Some of the variables may be critical whereas others may be redundant. Motivated by this fact, one can try linking the PC back to a subset of the original variables, which satisfy one or more optimal properties

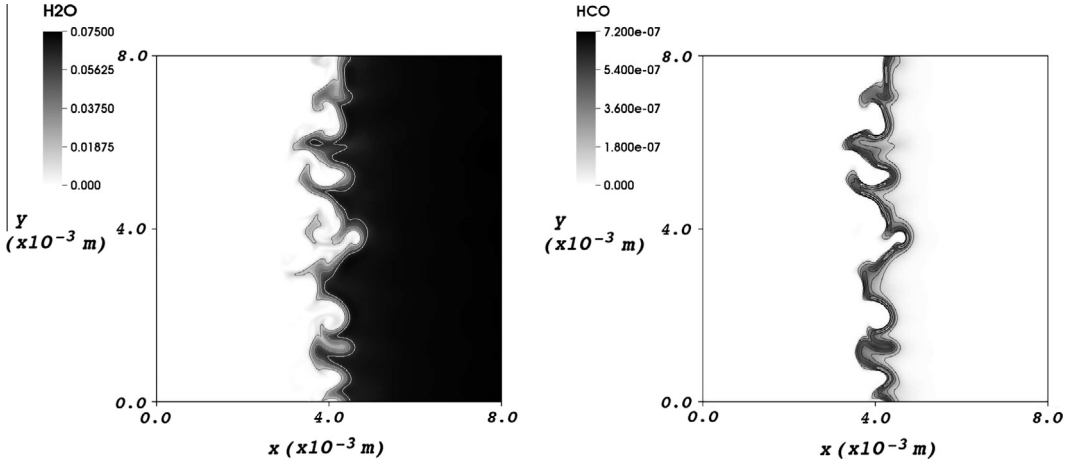

Fig. 2. Y_{H_2O}, Y_{HCO} field for the 2-D DNS, $t = 7.1241 \cdot 10^{-4}$ s.

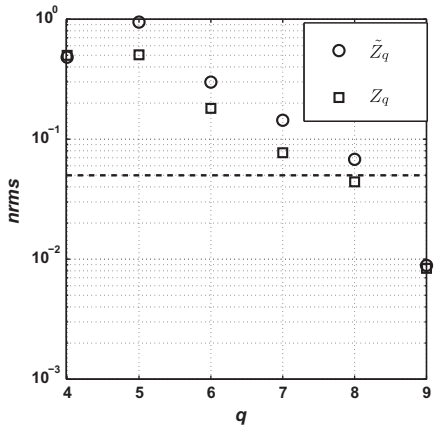

Fig. 3. Maximum *nrms* distance (*y*-axis) for the reconstruction of the state space variables as a function of the number of variables (q) (*x*-axis), while using the $\tilde{\mathbf{Z}}_q$ or $\mathbf{Z}_q$ in the construction of the **B** matrix. The dashed line shows 5% *nrms*. Scaling: standard deviation. PV selection: B2.

of PCA, such as the maximization of the variance of the original data **X**. A number of methods exist for selecting a subset of q original variables. The following methods are considered here:

- *B2 backward and B4 forward methods* [31]. Variables are determined by analyzing the principal component weights ($\mathbf{A}_q$). The *B2 backward method* removes variables associated to the PC with the smallest eigenvalues. In contrast the *B4 forward method* identifies variables which are associated with the PC with the largest eigenvalues.
- *M2 backward method* [32]. Variables are determined by comparing **Z** (Eq. (3)) with an approximate **Z**. Here $\widetilde{\mathbf{Z}}$ is constructed from a subset of the original variables, and the comparison is made to the original scores using a Procrustes analysis. Upon selection of appropriate variables the underlying structure of the data is preserved and the PVs are identified as the subset of variables which have been used for the approximation.
- *McCabe criteria* [33]. McCabe identified that the PCs satisfy a certain number of optimality criterion. The criterion are based on partial co-variance matrices calculated by selecting subsets of variables. In the current study the *MC1* and *MC2* criteria are used, which rely on the determinant (*MC1*) or trace (*MC2*) of the partial co-variance matrices.
- *Principal features* [34]. Variables are identified by analyzing the correlation between variable weights in $\mathbf{A}_q$. Variables are grouped according to their correlation, and the *k-means* algorithm [35] is used to extract a given number of variables from each subset, thus providing a representation of each of the groups.

Next, the influence of the different PV approaches on the reconstruction of the state variables is assessed and discussed. A useful metric for this task is one that describes the amount of variance lost by the selection of q (the number of transported variables) [30]. One can partition the variables into 2 groups, one for retained variables and the second for the remaining variables. The sample co-variance matrix can then be calculated:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \tag{16}$$

and the partial co-variance can then be given as:

$$\mathbf{S}_{22,1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}. \tag{17}$$

Taking the trace of the partial co-variance gives a quantitative value for the amount of variance lost (λ) for a given number of principal variables:

$$\lambda = \frac{trace\left(S_{22,1}(q)\right)}{trace(S)} \quad for \quad q = 1, 2, \ldots, Q \tag{18}$$

For example if $q = 0$ then $\lambda = 1$, meaning all variance is lost, or when $q = Q$ then $\lambda = 0$, which means all of the original variance in the system is explained.

Figure 4 shows the resultant percentage of lost variance calculated from Eq. (18), for the various PV selection methods. The figure shows that for $q = 6$ only 1% of the variance is lost while using the B4, B2, MC1 and MC2 methods, the PF, and M2 methods do not achieve such a degree of lost variance until $q = 7$. It is also interesting to note that several of the PV selection methods (M2, and PF) can indeed lose variance upon addition of PVs. The B2, MC1, and MC2, method appear to show a consistent increase in explained variance while adding principal variables. Table 1 shows with the acronym *pv* for the principal variables selected by the different approaches, while using auto scaling (see Section 3.4), and $q = 7$. It can be observed that different sets of PV are identified, depending on the selection method. However, a common logic seems to hold for all cases: all methods tend to select the majority of the principal variables among the radical species (HCO, HO_2, OH, and H_2O_2) which identify ignition or reaction regions, whereas temperature (which is forced to be a principal variable) and one other major variable are in general sufficient for capturing slower changes in the system. The first benchmark among the sets of PV must be carried out with respect to their ability of accurately reproducing the non-transported state space variables. Table 1 lists the *nrms* values for all the state variables reconstructed using MG-PCA and the different PV selection approaches. It can be observed that almost all methods provide a good approximation of the state space, with an exception for M2 and PF methods, which cannot properly recover some of the radical species (OH, H, HO_2). This can be explained by the fact that the best performing methods only select one of the major species, while keeping several radical species, which are usually associated with increased non-linearity. It is interesting to note that the forward and backward methods (B4 and B2) both select either of the major reaction products H_2O and CO_2, whereas the McCabe methods keep either H_2O or O_2. The PV and M2 methods performed the worst yielding *nrms* values greater than or equal to 10^{-1} for two or more of the variables. These methods identified two or three major species as PVs, thus decreasing the model's ability to represent the minor species.

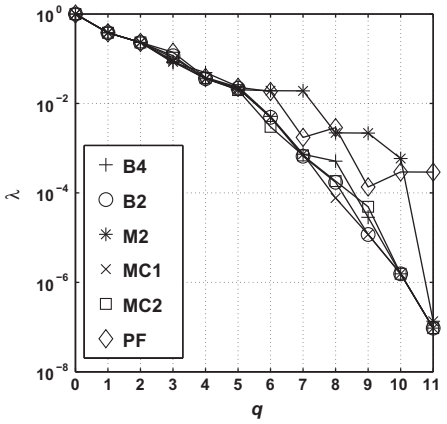

Fig. 4. Lost variance (*y*-axis) while adding principal variables (*x*-axis). The trace is given for the various PV selection methods highlighted in Section 3. Scaling: standard deviation.

Table 1
nrms distance for the reconstructed state space with $q = 7$ while testing various PV selection methods. The acronym *pv* indicates if the variable is a principal variable. Scaling method: auto scaling.

		B4	B2	M2	MC1	MC2	PF
T	1	*pv*	*pv*	*pv*	*pv*	*pv*	*pv*
O_2	2	$10^{-1.9}$	$10^{-2.1}$	*pv*	$10^{-1.9}$	*pv*	*pv*
H_2	3	$10^{-1.3}$	$10^{-1.3}$	*pv*	$10^{-1.3}$	$10^{-1.4}$	$10^{-1.4}$
H_2O	4	*pv*	$10^{-1.9}$	$10^{-1.8}$	*pv*	$10^{-1.9}$	*pv*
H_2O_2	5	*pv*	*pv*	*pv*	*pv*	*pv*	*pv*
CO	6	$10^{-1.8}$	$10^{-1.9}$	$10^{-2.0}$	$10^{-1.8}$	$10^{-1.8}$	$10^{-2.1}$
CO_2	7	$10^{-1.7}$	*pv*	*pv*	$10^{-1.7}$	$10^{-1.7}$	$10^{-2.0}$
O	8	$10^{-1.1}$	$10^{-1.1}$	*pv*	$10^{-1.1}$	$10^{-1.1}$	*pv*
H	9	*pv*	*pv*	$10^{-0.9}$	*pv*	*pv*	$10^{-1.0}$
OH	10	*pv*	*pv*	$10^{-0.9}$	*pv*	*pv*	$10^{-1.0}$
HO_2	11	*pv*	*pv*	$10^{-0.3}$	*pv*	*pv*	*pv*
HCO	12	*pv*	*pv*	*pv*	*pv*	*pv*	*pv*

3.4. Centering and scaling

In order to optimally choose the scaling and centering coefficient several methods will be investigated. Before presenting the methods, it is useful to rewrite Eq. (6) in a scalar form for the sake of clarity:

$$x_j^s = \frac{x_j - \bar{x}_j}{d_j} \quad \text{for } j = 1, \ldots, Q \tag{19}$$

The centering and scaling coefficients, $\bar{x}_j$ and d_j are stored to be used in the prediction step. The following scaling methods are adopted in the present work [22]:

- *auto scaling*, which adopts the standard deviation as the scaling factor, $d_j = s_j$;
- *range scaling*, which adopts the difference between the minimum and maximum variable value as scaling the factor, $d_j = \max(x_j) - \min(x_j)$;
- *pareto scaling* [36], which adopts the square root of the standard deviation as the scaling factor, $d_j = \sqrt{s_j}$;
- *vast* (variable stability) *scaling* [37], which adopts the product between the standard deviation and the coefficient of variation $(s_j/\bar{x}_j)$ as the scaling factor, $d_j = s_j^2/\bar{x}_j$;
- *level scaling*, which adopts the mean as scaling factor, $d_j = \frac{1}{N}\sum_{i=1}^{N} x_j$.

The B2 selection method presented some of the most promising results for the selection of PVs, accordingly an analysis is now given on the effects of scaling given that the transported variable are selected using the B2 method. Table 2 shows the effect of scaling on the reconstruction of the state variables for the DNS case. It is clear from Table 2 that the scaling methods have a significant effect on the ability to reconstruct the non-transported variables. The scaling results for the DNS case suggest that auto, range, and pareto scaling provide the most accurate results.

While analyzing the trace using Eq. (18), Fig. 5 shows nearly consistent decay in energy upon addition of variables over the various scaling methods, except for pareto scaling. This is due to the very large weight given by such a scaling to temperature [22] with respect to other variables. Thus the scaled data (under pareto) attribute almost all of the variance in the system to temperature. Temperature is the first PV so the lost variance description is consistent with what is observed in Fig. 5 for the other scaling methods.

3.5. Comparison of the PC-score approach, MG-PCA, and MG-L-PCA

Now a comparison of the three PCA based modeling techniques is made *a priori* on the 2-D DNS field in order to assess their performances. The classic PCA approach presented by [18] has the advantage of not requiring the selection of the systems principal variables, and simply transports the PCs of the system (Eq. (12)). This gives an equal distribution of error among all of the state-variables. Table 3 shows *nrms* and R^2 error ($R_j^2 = 1 - \sum_{i=1}^{n}(\phi_{ij} - \tilde{\phi}_{ij})^2 / \sum_{i=1}^{n}(\phi_{ij} - \bar{\phi}_j)^2$) statistics for the PC-score approach and the MG-PCA approach while attempting to reconstruct the state space variables with $q = 7$. MG-PCA has difficulty in reconstructing the O and H_2 radicals, whereas the PC-score reconstruction is much better overall for the state variables. Table 4 lists the R^2 and *nrms* values for the species source terms, which are calculated using the approximate state space while retaining 7 PVs (MG-PCA) or PCs (PC-score) of the co-variance matrix, i.e. $q = 7$. The relatively small errors in the state-space variables, lead to even larger inaccuracies in the approximation of the source terms. This is due to the high sensitivities of the reaction rates to minor changes in species concentrations (including radicals) [22,18]. The source terms for the PCs $\left(s_z = \frac{\omega_i A_z}{\gamma_i}\right)$ are also shown in Table 4. Even though the PC-score state space analysis shows an accurate reconstruction of all state-space variables (R^2 of nearly 1 for all of the variables), the error statistics for the source terms shows very inaccurate approximations. This condition is complicated by the fact that in the PC-score approach all the source terms are needed and they should all be computed with great precision. However, this requirement can never be fulfilled as a reconstruction error is always present, without any distinction between the transported and non-transported variables. The PC-score approach has a potential for

Table 2
nrms distance for the reconstructed state space using $q = 7$ variables, while testing various scaling methods. S = *auto*, R = *range*, P = *pareto*, V = *vast*, L = *level*.

	State variables				
	S	R	P	V	L
T	*pv*	*pv*	*pv*	*pv*	*pv*
O_2	$10^{-2.1}$	$10^{-2.1}$	$10^{-2.8}$	*pv*	$10^{-2.0}$
H_2	$10^{-1.3}$	$10^{-1.4}$	*pv*	$10^{-1.4}$	$10^{-1.3}$
H_2O	$10^{-1.9}$	$10^{-1.8}$	*pv*	$10^{-1.9}$	$10^{-1.9}$
H_2O_2	*pv*	*pv*	$10^{-0.3}$	*pv*	*pv*
CO	$10^{-1.9}$	$10^{-2.1}$	$10^{-2.4}$	$10^{-1.9}$	$10^{-1.9}$
CO_2	*pv*	*pv*	*pv*	$10^{-1.8}$	*pv*
O	$10^{-1.1}$	$10^{-1.1}$	*pv*	$10^{-1.1}$	*pv*
H	*pv*	*pv*	$10^{-1.0}$	*pv*	*pv*
OH	*pv*	*pv*	*pv*	*pv*	$10^{-1.1}$
HO_2	*pv*	*pv*	*pv*	*pv*	*pv*
HCO	*pv*	*pv*	$10^{-0.4}$	*pv*	*pv*

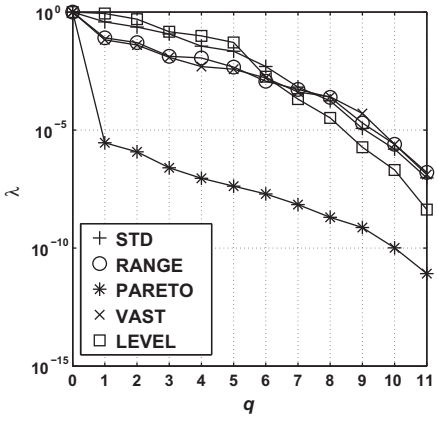

Fig. 5. Lost variance (*y*-axis) while adding principal variables (*x*-axis). The trace is given for the various scaling methods highlighted in Section 3. PV selection: B2.

Table 3
nrms distance and R^2 statistics for the reconstructed state space with $q = 7$. Scaling: standard deviation. PV selection: B2.

MG-PCA			PC-Score		
Variable	*nrms*	R^2	Variable	*nrms*	R^2
T	*pv*	*pv*	O_2	$10^{-2.0}$	1
O_2	$10^{-2.1}$	1	H_2	$10^{-1.5}$	0.999
H_2	$10^{-1.3}$	0.998	H_2O	$10^{-1.7}$	1
H_2O	$10^{-1.9}$	1	H_2O_2	$10^{-2.6}$	1
H_2O_2	*pv*	*pv*	CO	$10^{-1.8}$	1
CO	$10^{-1.9}$	1	CO_2	$10^{-1.8}$	1
CO_2	*pv*	*pv*	O	$10^{-1.7}$	1
O	$10^{-1.1}$	0.994	H	$10^{-2.0}$	1
H	*pv*	*pv*	OH	$10^{-1.8}$	1
OH	*pv*	*pv*	HO_2	$10^{-3.5}$	1
HO_2	*pv*	*pv*	HCO	$10^{-2.2}$	1
HCO	*pv*	*pv*			

Table 4
nrms distance and R^2 statistics for the reconstructed source terms with $q = 7$. Scaling: standard deviation. PV selection: B2.

MG-PCA			PC-Score		
PV	*nrms*	R^2	PC	*nrms*	R^2
T	$10^{-0.4}$	0.87	*PC* 1	$10^{-0.3}$	0.72
H_2O_2	$10^{-0.8}$	0.98	*PC* 2	$10^{0.7}$	< 0
CO_2	$10^{-1.3}$	0.99	*PC* 3	$10^{0.4}$	< 0
H	$10^{-0.2}$	0.61	*PC* 4	$10^{-0.2}$	0.65
OH	$10^{-0.9}$	0.98	*PC* 5	$10^{1.3}$	< 0
HO_2	$10^{0.3}$	< 0	*PC* 6	$10^{-0.3}$	0.80
HCO	$10^{-0.5}$	0.90	*PC* 7	$10^{0.3}$	< 0

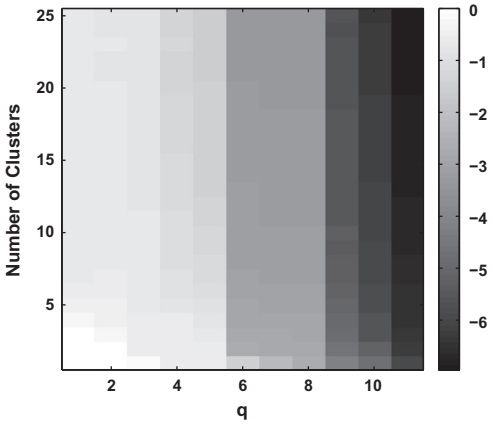

Fig. 6. Minimum R^2 statistic for the state variables as a function of the number of retained variables (x-axis) and clusters (y-axis). The error is plotted as $log_{10}(1 - R^2)$. Scaling: Pareto. Fuel: Syngas.

higher compression, because the variance explained by a linear combination of all of the variables is greater than that provided by a subset of optimal variables. However, without a method to resolve the error propagation in the source terms the approach can be inaccurate.

3.6. MG-L-PCA

The initial analysis of the premixed syngas case indicates that MG-PCA can provide the required precision if $q = 7$. Now the MG-L-PCA method is tested against MG-PCA to investigate the potential of the local PCA formulation. By clustering the data

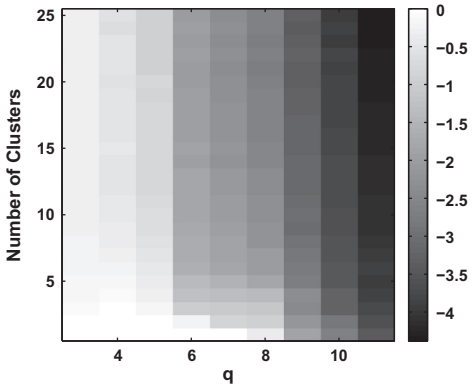

Fig. 7. Minimum R^2 statistic for the principal variable source terms as a function of the number of retained variables (x-axis) and clusters (y-axis). The error is plotted as $log_{10}(1 - R^2)$. Scaling: Pareto. Fuel: Syngas.

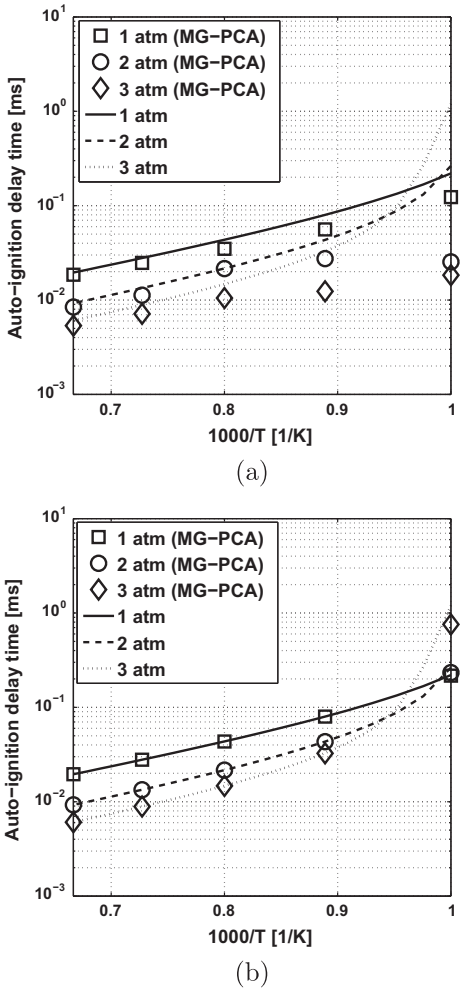

Fig. 8. Auto-ignition delay time using MG-PCA as a function of temperature for various system pressures, while $q = 6$ (a) or $q = 7$ (b).

according to temperature the MG-L-PCA method may now be used to attempt to create a better local basis for the reconstruction of the non-transported variables. Figure 6 provides a visualization of the error produced while varying q (x-axis) and c, the number of clusters (y-axis). The gray scale in the figure represents the lowest R^2 statistic for the reconstructed state space variables and is shown as $log_{10}(1 - R^2)$. The analysis confirms the earlier findings

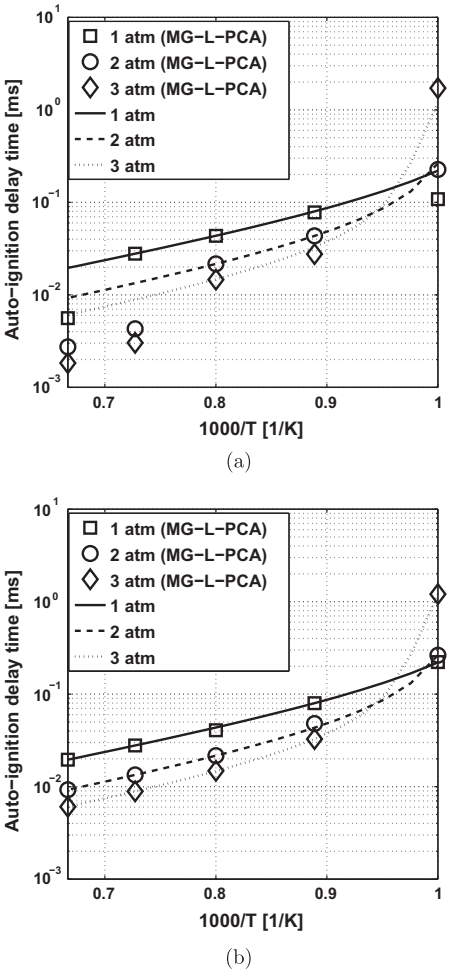

Fig. 9. Auto-ignition delay time using MG-L-PCA as a function of temperature for various system pressures, while using $q = 4$ (a) or $q = 5$ (b).

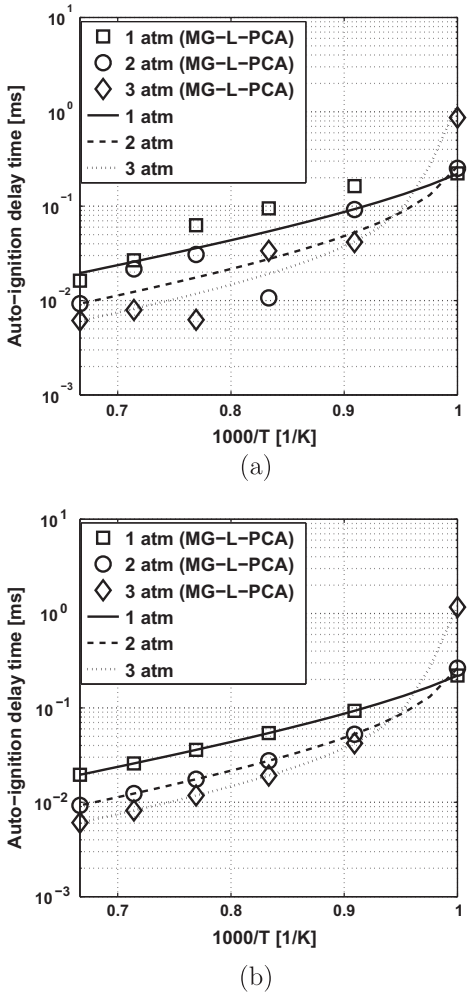

Fig. 10. Auto-ignition delay time using MG-L-PCA as a function of temperature for various system pressures, while using $q = 4$ (a) or $q = 5$ (b). Here the MG-L-PCA model has been trained for each case using data which ranged over several initial nearby temperatures and pressures.

for MG-PCA, with a minimum R^2 of 0.994 while using $q = 7$ and $c = 1$ (this is in fact the MG-PCA model). By adding clusters one can better identify the local **B** matrix and get better reconstruction of the state space.

Figure 7 also confirms that a higher value of q and c are needed to capture the source term with respect to the state space, due to the error propagation in the non-linear source terms. Optimization figures such as Figs. 6 and 7, are very useful for deciding the parameters of the reduced model to be implemented in a CFD solver, as it provides the minimum number of cluster needed to achieve a desired reduction.

From the above analysis it can be concluded that the application of MG-L-PCA has the potential of providing accurate results while achieving a significant reduction state space variables.

4. Results

The present section shows actual computations using the MG-PCA and MG-L-PCA methods. First an auto-ignition (0-D) case is considered with a new technique for cluster identification in the framework of MG-L-PCA. Then, two *a posteriori* DNS simulations are presented for a flame-turbulence and a flame-vortex interaction, to allow an assessment of the model with a more complex chemistry i.e. syngas, than in [19] and in the presence of non-unity Lewis number effects.

4.1. Auto-ignition delay time

The auto-ignition delay time is a rigorous test that demonstrates the ability of a model or a chemical kinetics mechanism to capture complex physical characteristic of the ignition process. The auto-ignition delay time is often tested at various temperatures and pressures, in order to asses the robustness of the model being applied. Accurate prediction of the auto-ignition delay time is particularly important in systems with pre-heated, and premixed mixtures, such as gas turbines. In the current study a simplified case is examined with the following assumptions: constant pressure, homogeneous, stagnant premixed fuel, mixture temperature above the auto-ignition temperature of the fuel, and adiabatic conditions. The process is modeled using the differential equations for the chemical species in the system, and the temperature:

$$\frac{dY_i}{dt} = \frac{1}{\rho} W_i \omega_i \tag{20}$$

$$\frac{dT}{dt} = \frac{1}{\rho c_{P,mix}} \left(-\sum_{i=1}^{N} h_i W_i \omega_i \right) \tag{21}$$

where Y_i are the species mass fractions, W_i is the ith species molecular weight, ω_i is the molar source term for the ith species, h_i is the molar enthalpy for species i, and $c_{P,mix}$ is the heat capacity of the

mixture. The equations are solved using an in-house implementation of the batch reactor.

In contrast with the DNS case, which exhibits differential diffusion, the degrees of freedom in this case are different. The auto-ignition cases solve 13 transport equations and are constrained by the elemental balances (C, H, O), conservation of mass, and energy, leaving 8 degrees of freedom. The degrees of freedom can be further analyzed by finding the overlap between conservative modes and the principal components identified in a system. The conservative modes for the chemical species of a system can be given by:

$$C_{i,j} = e_{i,j} d_i \tag{22}$$

where $e_{i,j}$ is the mass fraction of jth element on the ith species, and d_i is the scaling used for each of the species in the system. If the PCs of the system are predominately a linear combination of conservative modes an approximation of the PCs can be given using a simple least squares regression:

$$\widetilde{\mathbf{A}} = \mathbf{C}\left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{A} \tag{23}$$

If $\widetilde{\mathbf{A}} \cdot \mathbf{A} = 1$, the PCs fall completely in conservative space, or if $\widetilde{\mathbf{A}} \cdot \mathbf{A} = 0$ the PC is not related to the conservative laws. A simple calculation of the dot product for the auto-ignition results showed, as expected, the last 5 PCs aligned with the conservative modes, leaving 8 degrees of freedom. Accordingly any reduction leaving less than 8 variables is a true reduction achieved by the model.

As described in Section 3, a representative data-set is needed in order to determine the number of variables (q) required for a desired accuracy, the variables to transport, and the matrix/matrices **B** (MG-PCA/MG-L-PCA). Initially the full system of equations is solved at various running conditions in order to generate the data-sets for the *a priori* construction of the model. The number of transported variables q, the transported variables and the **B** matrix/matrices are calculated from each case involving different initial conditions. In the current study the B2 selection method is used with pareto scaling.

The MG-PCA method shows good accuracy when using 7 (13% reduction), of the original 12 variables (see Fig. 8b), which is consistent with the results found in the *a priori* analysis. This is by no means a significant reduction. Minor differences are observed at smaller initial temperatures where the ignition event takes much longer. It is observed that with 6 transported variables that a considerable loss in accuracy occurs over the entire range of pressures and initial temperatures (Fig. 8a).

While using the MG-L-PCA method, proper clustering of the data is crucial in order to achieve an accurate local reconstruction of the data. As temperature is directly transported in this system, it is an optimal variable for identifying clusters, and at run time providing the local **B** matrix which gives the most accurate reconstruction of the local state space. As one would suspect, problems may arise near cluster boundaries or when a given cluster contains a highly non-linear peak from one of the transported radical species. Because of this a clustering algorithm was devel-

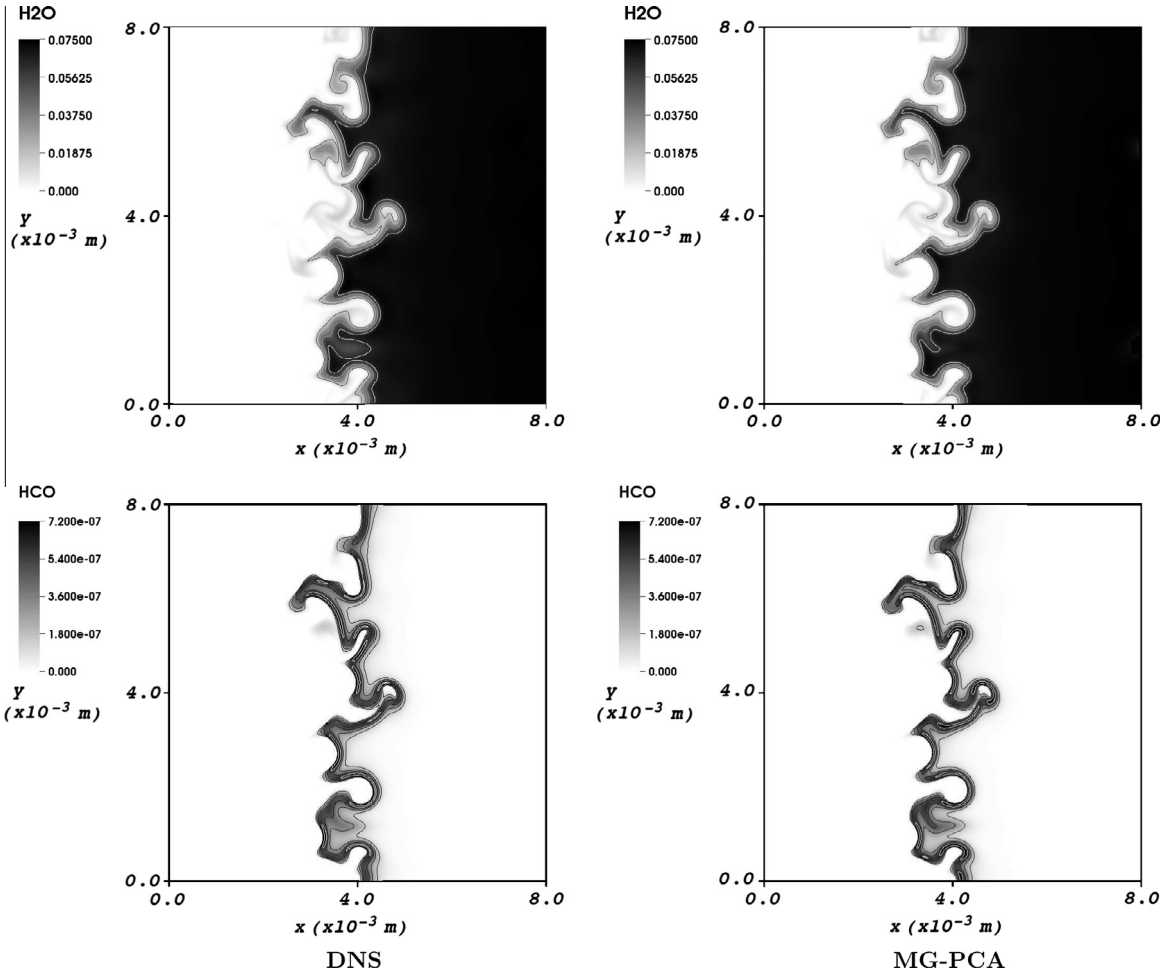

Fig. 11. Fields of Y_{H_2O} (top) and Y_{HCO} (bottom) using DNS (left) and MG-PCA (right) $t = 1.064 \cdot 10^{-3}$ s.

oped which looks at the *a priori* data and finds local extrema in the radical species profiles, and creates new cluster boundaries at these locations in order to increase the accuracy and provide smoother transition between clusters.

In reference to Fig. 9b, the results for MG-L-PCA show a much better approximation while transporting as few as 5 variables (38% reduction). However, when moving to 4 variables (50% reduction) reasonable accuracy is observed with moderate initial temperatures, with higher discrepancies at both higher and lower initial temperatures due to the error from the model (see Fig. 9a).

In order to further assess the ability of the model to predict the auto-ignition delay time, MG-L-PCA is again tested in predictive fashion. Multiple simulations are run over a range of initial temperatures, allowing fluctuations of ±15 K and ±0.5 bar with respect to the target temperature and pressure. Then, the model is used at the target initial condition, which is not used in computing the

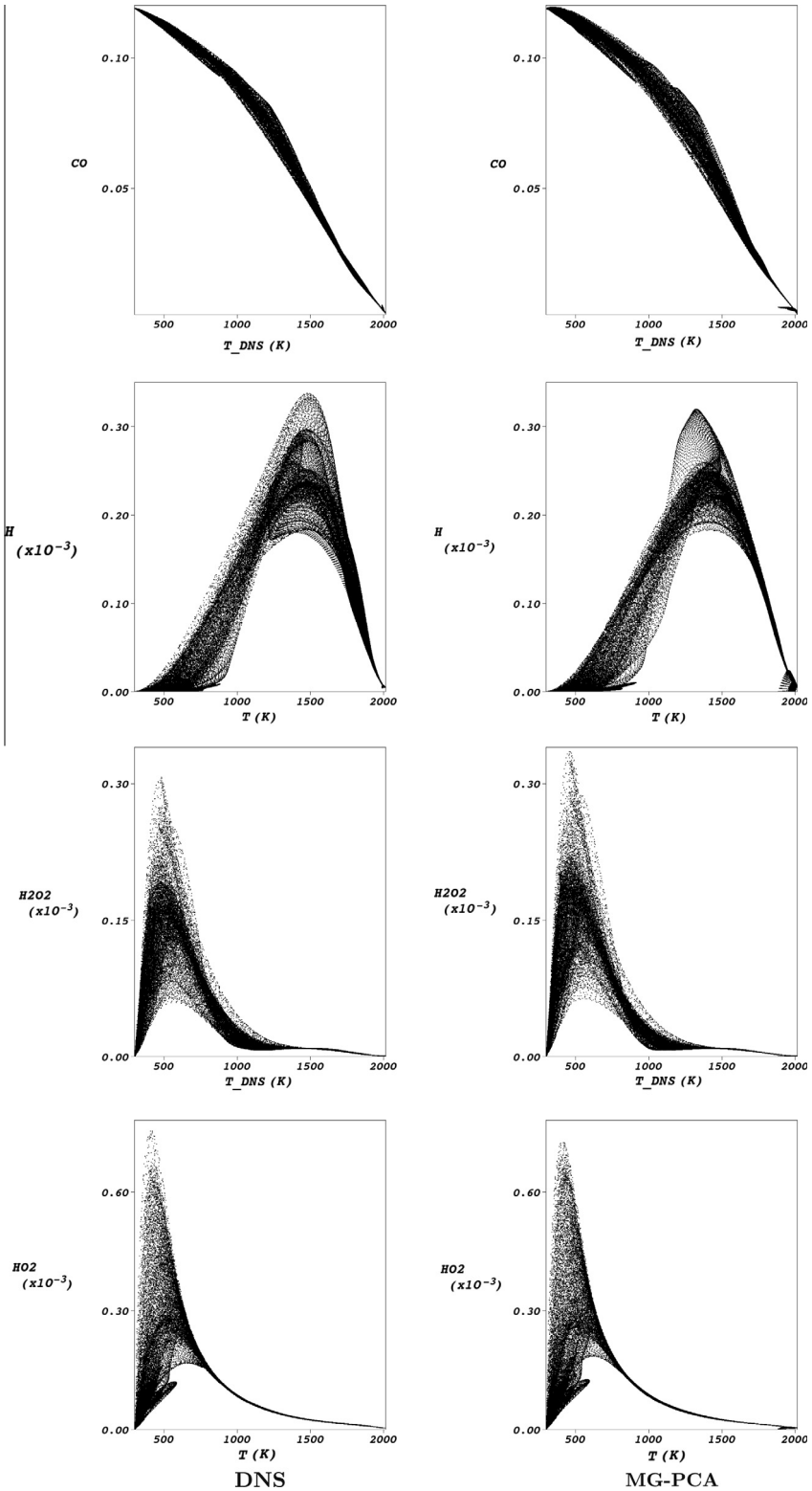

Fig. 12. Scatter plot of (from top to bottom) Y_{CO}, Y_H, $Y_{H_2O_2}$, Y_{HO_2} *vs* temperature, using DNS (left) and MG-PCA (right) $t = 1.064 \cdot 10^{-3}$ s.

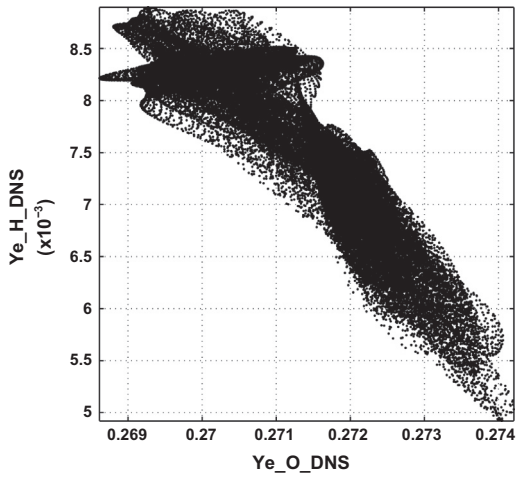

Fig. 13. Elemental mass fraction of H as a function of the elemental mass fraction of O at $t = 1.064 \cdot 10^{-3}$ s.

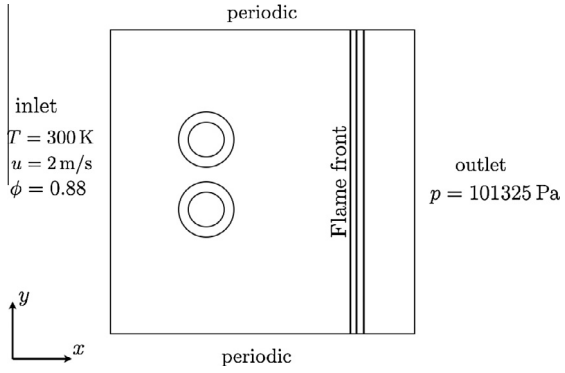

Fig. 14. Boundary conditions for the flame vortex ring.

model parameters, yielding a true prediction. Figure 10 shows the predicted auto-ignition times given the new training data-sets, demonstrating relatively good predictions. As one would expect, the prediction for the auto-ignition delay times are not quite as accurate (in the case with 4 variables) as those found in Fig. 9. This is due to the fact that the approximation in each of the clusters was slightly less accurate as a result of the small differences in the species profiles due to the range in initial conditions. However when moving to 5 variables the solution is again accurate. It is also interesting to note that with a change in initial temperatures and pressures, the enthalpy and pressure of the resulting system are not exactly constant, yielding a more challenging problem. The results indicate that, for a small fluctuation around the conditions of interest, the reduced PCA model is able to provide accurate results without requiring additional dimensions. For stronger variations, the model dimensionality is likely to increase, due to the increased degrees of freedom of the system.

4.2. Flame-turbulence interaction

The present section reports the result of a DNS calculation of a flame-turbulence interaction case using the MG-PCA method. Numerical setup is exactly the same as in Section 3.1 and is not recalled. The simulation is performed using $q = 8$ (33% reduction) in order to reconstruct quasi-exactly the missing variables leaving $R^2 > 0.9999$ for all variables. As indicated in Section 3.1, the theoretical degrees of freedom for the system are 12 indicating that with $q = 8$, a supposed reduction of 4 degrees of freedom is achieved. However, by repeating the procedure in Eqs. (22) and (23) it was found that 2 of the last 5 modes are correlated with the conservative modes, implying that the true degrees of freedom for this system are actually 11. With $q = 8$, a reduction of 3 degrees is still achieved.

Following the conclusions of the previous sections, pareto scaling is used in combination with the B2 selection method, giving the following variables to be transported inside the solver: $T, Y_{H_2}, Y_{H_2O}, Y_{CO_2}, Y_O, Y_{OH}, Y_{HO_2}, Y_{HCO}$.

It should be stressed that the database used to train the model is the 1-D laminar flame which was extended along the y-axis to initialize the computation (see Section 3.1). This one dimensional laminar flame is a syngas–air flame at $\phi = 0.88$. The computational grid consists of 2501 points with a spacing of $1.25 \cdot 10^{-5}$m. Inlet velocity and temperature are set to 0.2 m/s and 300 K respectively and pressure is set to 101, 325 Pa.

Before analyzing the results of the DNS simulation, the impact of the reduction on the laminar flame speed is calculated for the 1-D steady laminar flame which is used to train the model parameters. The laminar flame speed for a system is estimated using the following equation:

$$S_l = -\frac{1}{\rho Y_F}\int_{mesh} \dot{\omega}_F dx \tag{24}$$

where Y_F is the mass fraction of the fuel, $\dot{\omega}_F$ is the net production rate of the fuel ($kg/m^3/s$), and ρ is the density of the mixture. The modeling error is reported using the following equation:

$$error_{S_l} = 1 - \frac{|S_l - S_{l,MG-PCA}|}{S_l} \tag{25}$$

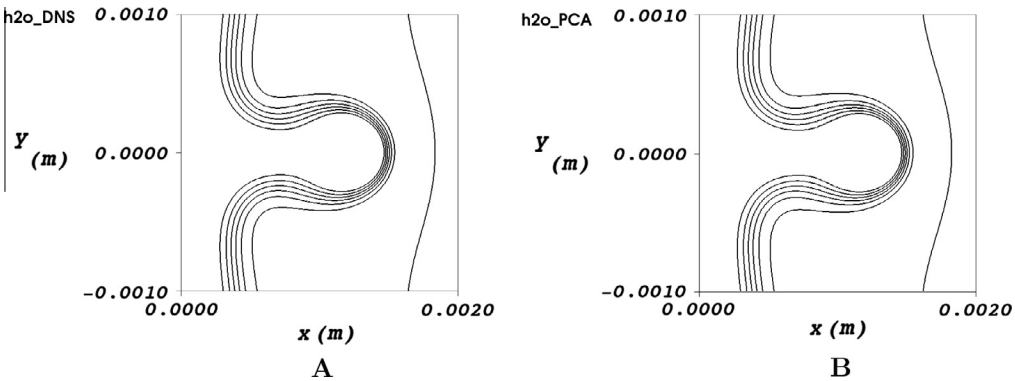

Fig. 15. Y_{H_2O} field for the flame-vortex interaction for the DNS (**A**) and MG-PCA (**B**), at $t = 56.7$ µs.

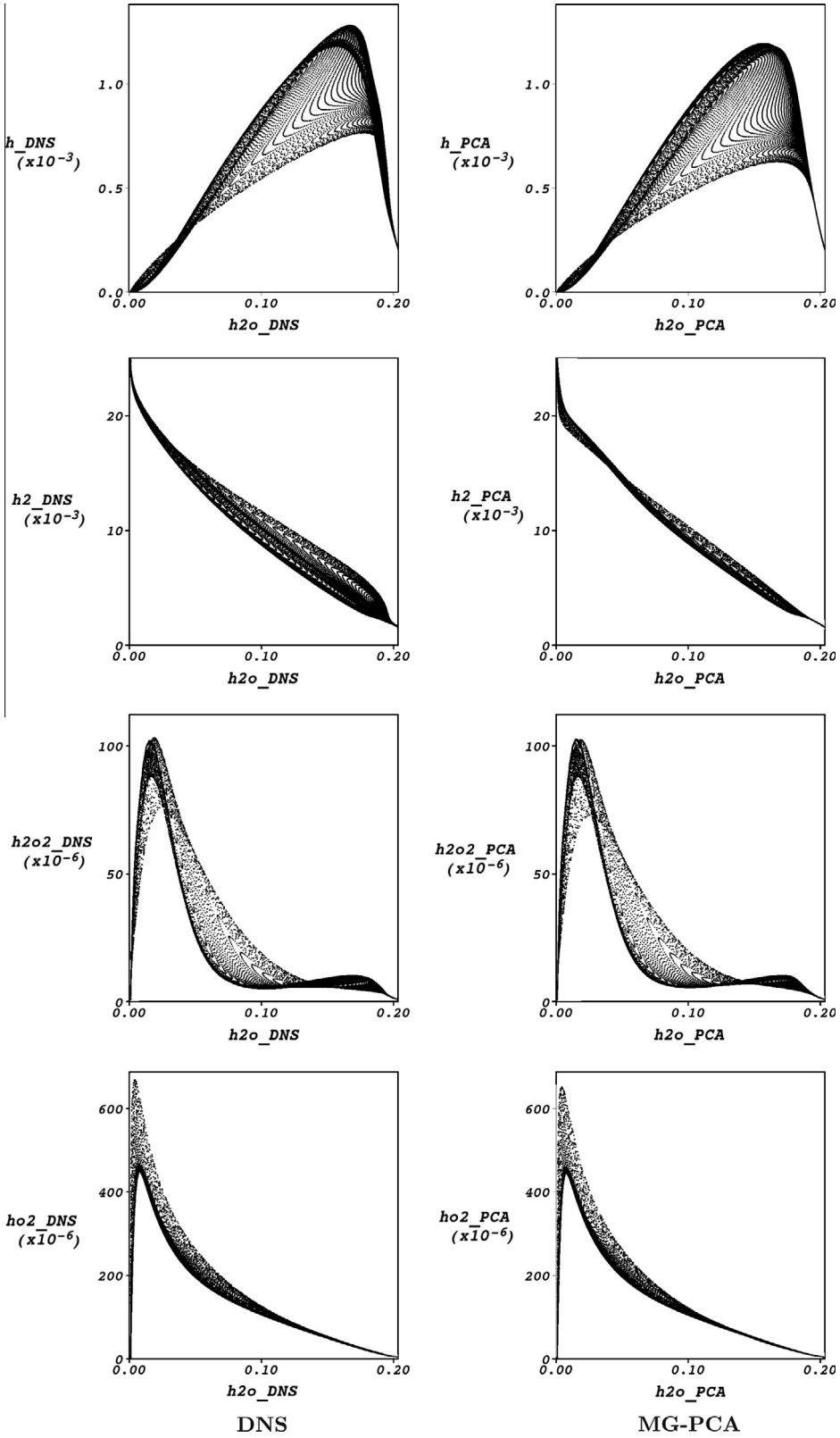

Fig. 16. Comparisons of DNS (left) and MG-PCA (right) results for (from top to bottom): $Y_{H}, Y_{H_2}, Y_{H_2O_2}$ and Y_{HO_2} at $t = 56.7$ μs.

The MG-PCA method resulted in a modeling error for S_l of 0.14% when compared to the full solution. This small amount of error in the flame speed calculation gives a better picture of the accuracy of the model in representing the chemistry occurring in the system without the effects of turbulence.

Now an analysis of the results to the flame-turbulence is presented. First a comparison of Y_{H_2O} and Y_{HCO} fields from the DNS (left) and MG-PCA (right) simulations are given in Fig. 11, which shows a very good agreement between the full and reduced solutions. Similarly, Fig. 12 shows scatter plots of $Y_{CO}, Y_{H}, Y_{H_2O_2}, Y_{HO_2}$

vs temperature for DNS and MG-PCA computations, indicating again a good agreement between the two solutions. Figure 13 gives the elemental mass fractions of H as a function of O, emphasizing that the differential diffusion effect are significant in this configuration. It appears that the model can naturally account for the turbulence–chemistry interactions of the system, thanks to the higher number of degrees of freedom available with respect to other methods such as ILDM or FPI and the appropriate selection of the key variables to be transported in the code. The current study shows that the single 1-D laminar flame is sufficient to train the model for a more complex flame-turbulence simulation. This indicates that the applicability of the model is not entirely limited to the system used to generate the database (such as SLFM or FPI), but indeed the model can account for small perturbations around

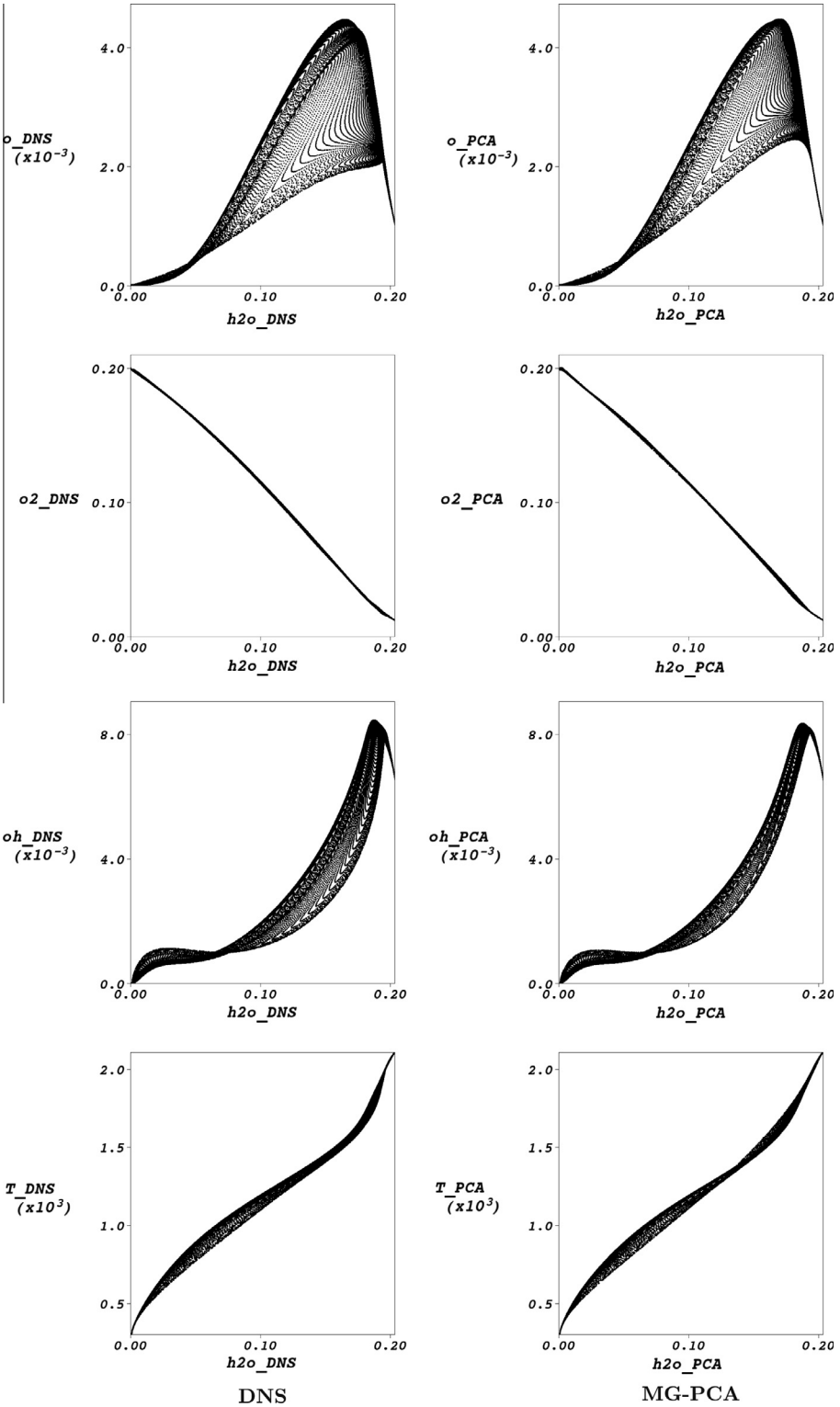

Fig. 17. Comparisons of DNS (left) and MG-PCA (right) results for (from top to bottom): Y_O, Y_{O_2}, Y_{OH} and T at $t = 56.7$ μs.

the original manifold. It is important to clarify that the method does not explicitly account for the non-intrinsic effects in a system, meaning that the model used for the current case may not be reliable when applied to a case with a much higher turbulence intensity.

4.3. Flame-vortex interaction

To further investigate the ability of the proposed MG-PCA approach in dealing with more complex flame structures, a hydrogen flame-vortex interaction is now computed. Indeed while dif-

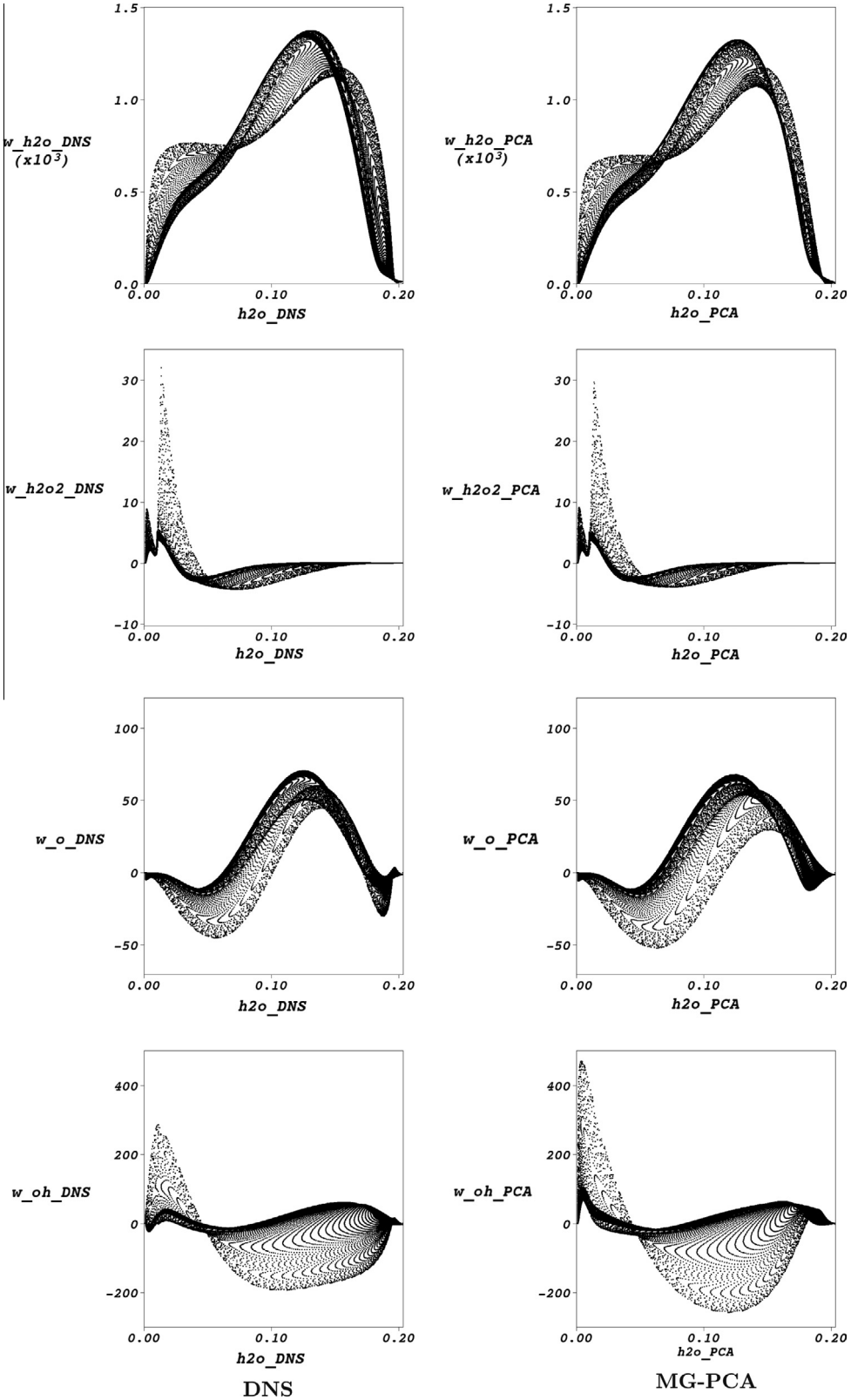

Fig. 18. Comparisons of the chemical source terms per unit of mass for the DNS (left) and MG-PCA (right) for (from top to bottom): Y_{H_2O}, $Y_{H_2O_2}$, Y_O and Y_{OH} at $t = 56.7$ μs.

ferential diffusion was accounted for in the previous section, it is of common knowledge that differential diffusion effects are stronger in pure hydrogen flames. Moreover, this flow configuration was chosen because the vortex distorts the flame front yielding variations of the species mass fraction in the flame front [38], generating a complex flame structure with significant non-equilibrium phenomena, i.e. extinction.

The data-set used to train MG-PCA is obtained from a premixed freely propagating hydrogen–air 1-D laminar flame at an equivalence ratio of $\phi = 0.88$ and a fresh gas temperature of 300 K. The pressure is set to 101,325 Pa, velocity of the unburned gases is 1.25 m/s, and the chemical reaction mechanism [39] is employed. The domain consists of 2541 points with a constant grid spacing of $\Delta x = 6.25 \cdot 10^{-6}$ m. There are 9 species in the mechanism: Y_H, $Y_{H_2}, Y_O, Y_{O_2}, Y_{OH}, Y_{N_2}, Y_{H_2O}, Y_{HO_2}, Y_{H_2O_2}$. As for the flame-turbulence interaction, pareto scaling is used in combination with the B2 selection method yielding the following set of 6 transported variables ($q = 6$): Y_H, Y_O, Y_{OH}, Y_{HO_2}, Y_{H_2O} and $Y_{H_2O_2}$. Since differential diffusion is taken into account a reduction of 3 degrees of freedom is achieved. This is also confirmed by the *a posteriori* analysis of the conservative modes of the system. For the present case, the PCA reduction on the laminar flame speed leads to an error of 1.6%.

Next, the description and results for the flame-vortex case are presented. The flow is initialized as proposed by Renard et al. [38]: the 1-D laminar flame presented above is extended along the y-axis and two counter-rotating vortices are added in the unburned gases. Each of those vortices are initialized by:

$$u_1(x,y) = u_0 + \frac{\tau}{R_c^2} y \exp\left(-\frac{(x^2+y^2)}{2R_c^2}\right) \tag{26}$$

$$u_2(x,y) = \frac{\tau}{R_c^2} x \exp\left(-\frac{(x^2+y^2)}{2R_c^2}\right) \tag{27}$$

$$p(x,y) = p_0 \; exp\left(-\frac{\gamma}{2}\left(\frac{\tau}{cR_c}\right)^2 exp\left(-\frac{(x^2+y^2)}{R_c^2}\right)\right) \tag{28}$$

where u_1 is the velocity along the x-axis, u_2 is the y-axis velocity and p is the pressure. Parameters are set as follows: $u_0 = 1.25$ m/s and $p_0 = 101,325$ Pa. The distance between the vortices is equal to R_c. The grid is composed of 641×641 nodes with a constant grid spacing in both directions: $\Delta x = \Delta y = 6.25\ 10^{-6}$ m. The boundary conditions are summarized in Fig. 14. An inlet condition is imposed at $x = x_{min}$ which enforces $T = 300$ K,$u_1 = 1.25$ m/s, $u_2 = 0$ m/s and a composition corresponding to an equivalence ratio of $\phi = 0.88$. An outlet boundary condition is imposed at at $x = x_{max}$ which sets $p = 101,325$ Pa. The remaining two boundaries use periodic conditions. Finally, τ and R_c are set to $\tau = 0.02$ and $R_c = 2 \cdot 10^{-4}$ m yielding a maximum velocity of 122 m/s and a vortex convection speed of 61 m/s.

Figure 15 shows a close view of the Y_{H_2O} field in the interaction zone for the DNS and MG-PCA method at $t = 56.7$ μs when the maximum strain rate is observed in the flame. It already indicates a very good agreement between the DNS and MG-PCA results. However to allow a better comparison, scatter plots of the reactive species and temperature *vs* Y_{H_2O} are shown for DNS and MG-PCA at $t = 56.7$ μs in Figs. 16 and 17. From those figures, it can be observed that the vortex interaction with the flow leads to a dispersion of the scalars from the 1-D laminar flame used to train MG-PCA; moreover, a partial, extinction of the flame is observed.

As for the flame-turbulence interactions, the MG-PCA method is able to access regions of the state-space that were not included in its training data-set. To further shows the quality of the MG-PCA method, the source terms per unit of mass of $Y_{H_2O}, Y_{H_2O_2}, Y_O, Y_{H_2O}$, Y_{OH} are plotted as a function of Y_{H_2O} in Fig. 18. As for the species

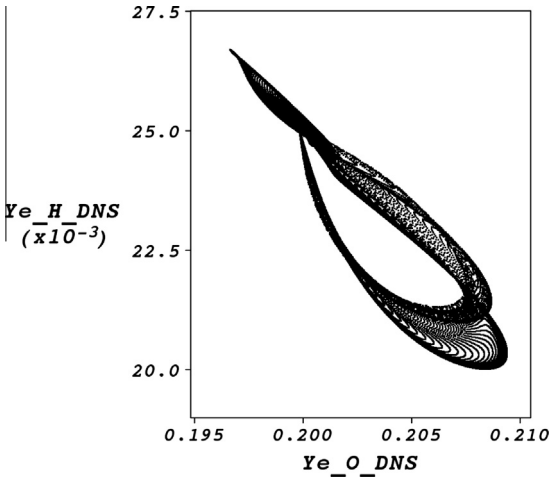

Fig. 19. Elemental mass fraction of H as a function of the elemental mass fraction of O for the DNS computation at $t = 56.7$ μs.

mass fractions, the MG-PCA method allows an accurate computation of the source terms and therefore of the flame behavior.

Finally, Fig. 19 gives the elemental mass fractions of H as a function of O for the DNS computation at $t = 56.7$ μs, emphasizing that the differential diffusion effect are of great importance in this configuration.

5. Conclusions

PCA has demonstrated capability in identifying the low-dimensional manifold which can accurately describe a chemically-reacting system with a reduced number of optimal parameters. The MG-PCA approach provides a realistic application of PCA for combustion systems through the use of principal variables. A comparison with the classic PC-score approach indicated the strong potential of MG-PCA for the development of reduced-order combustion models of tailored accuracy. The present paper provides an easy and effective tool for the development of reduced models, whose implementation requires only minor modifications of existing CFD codes, and for which the accuracy of the model can be evaluated and tailored *a priori*. The main findings of the present paper can be summarized as follows:

- The calculation of the **B** matrix (Eq. (13)) using the actual $\mathbf{Z}_q$, gives the MG-PCA methods an increased accuracy in reconstruction of non-transported state space variables.
- The selection of the transported variables using the various principal variable selection methods greatly affects the reliability and accuracy of MG-PCA models, thus justifying the need for optimal selection techniques, as the ones outlined here. As far as the PV selection techniques are concerned the B2 method was found to be the most robust and the one providing the best approximation of the state space.
- Scaling methods play a major role in the identification of the optimal projection matrix $\mathbf{A}_q$ and subsets of transported variables. In particular, it was shown that scaling methods other than the standard auto-scaling can also provide increased accuracy in reproducing state space variables and principal variable source terms. Among them, auto, range, and pareto scaling provided the better results.
- The global MG-PCA approach can be effectively employed for simple fuels such as hydrogen [19]. However, for larger mechanisms such as syngas or methane, the MG-L-PCA formulation

must be employed, in order to achieve a significant reduction and to capture the non-linear features of the manifold underlying the chemically reacting system.
- The laminar flame speed calculations helped in showing that the MG-PCA method is indeed accurately describing the reaction system in the absence of large velocity gradients.
- The flame-turbulence and flame-vortex interactions computations demonstrated the ability of the MG-PCA method to perform an actual computation. Moreover, the results also indicate that the model can account for a small perturbation around the manifold, allowing to account for some of the turbulent effects of the system.

Future work will attempt to further automate the manifold generation procedure within CFD codes and to validate the overall approach on a broad range of combustion systems. Also, the coupling of the proposed MG-L-PCA clustering technique with a flow solver is currently under investigation.

Acknowledgments

This research was sponsored by the National Nuclear Security Administration under the Accelerating Development of Retrofittable CO_2 Capture Technologies through Predictivity program through DOE Cooperative Agreement DE-NA0000740 and the Fonds National de la Recherche Scientifique, FRS-FNRS (Communauté Française de Belgique). Computing resources were provided by the IDRIS under the allocation 2009-i2009020164 made by GENCI (Grand Equipement National de Calcul Intensif).

References

[1] N. Peters, Proc. Combust. Inst. 24 (1986) 1231–1250.
[2] N. Peters, Prog. Energy Combust. Sci. 10 (1984) 319–339.
[3] J. Van Oijen, L. De Goey, Combust. Sci. Technol. 161 (2000) 113–137.
[4] J.A. van Oijen, Eindhoven University of Technology, 2002.
[5] O. Gicquel, N. Darabiha, D. Thevenin, Proc. Combust. Inst. 28 (2000) 1901–1908.
[6] B. Fiorina, R. Baron, O. Gicquel, D.T. Evenin, S. Carpentier, N. Darabiha, D. Thevenin, Combust. Theory Modell. 7 (2003) 449–470.
[7] K. Kuo, Principles of Combustion, A Wiley-Interscience publication, Wiley, 1986.
[8] U. Maas, S.B. Pope, in: Symposium (International) on Combustion, vol. 24. Elsevier, pp. 103–112.
[9] W. Jones, R. Stelios, Combustion and Flame 142 (2005) 223–234.
[10] S. Rigopoulos, Int. J. Multiscale Comput. Eng. 5 (2007) 11–18.
[11] U. Maas, D. Thévenin, Proc. Combust. Inst. 27 (1998) 1183–1189.
[12] A. Parente, J.C. Sutherland, P.J. Smith, L. Tognotti, Proc. Combust. Inst. 32 (2009) 1579–1586.
[13] A. Parente, J.C. Sutherland, B.B. Dally, L. Tognotti, P.J. Smith, Proc. Combust. Inst. 33 (2011) 3333–3341.
[14] A. Biglari, J.C. Sutherland, Combust. Flame 159 (2012) 1960–1970.
[15] S.B. Pope, Proc. Combust. Inst. 34 (2013) 1–31.
[16] H. Mirgolbabaei, T. Echekki, Combust. Flame 160 (2013) 898–908.
[17] H. Mirgolbabaei, T. Echekki, Combust. Flame 160 (2014) 118–126.
[18] J. Sutherland, A. Parente, Proc. Combust. Inst. 32 (2009) 1563–1570.
[19] A. Coussement, O. Gicquel, A. Parente, Proc. Combust. Inst. 34 (2013) 1117–1123.
[20] Y. Yang, S.B. Pope, J.H. Chen, Combust. Flame 160 (2013) 1967–1980.
[21] A. Najafi-Yazdi, B. Cuenot, L. Mongeau, Combust. Flame 159 (2012) 1197–1204.
[22] A. Parente, J.C. Sutherland, Combust. Flame 160 (2013) 340–350.
[23] N. Kambhatla, T.K. Leen, Neural Comput. 9 (1997) 1493–1516.
[24] N. Peters, Prog. Energy Combust. Sci. 10 (1984) 319–339.
[25] R. Kee, F. Rupley, E. Meeks, J. Miller, Release 3 (1996).
[26] A. Coussement, O. Gicquel, J. Caudal, B. Fiorina, G. Degrez, J. Comput. Phys. 231 (2012) 5571–5611.
[27] A. Coussement, O. Gicquel, B. Fiorina, G. Degrez, N. Darabiha, J. Comput. Phys. 245 (2013) 259–280.
[28] T. Passot, A. Pouquet, J. Fluid Mech. 181 (1987) 441–466.
[29] S.G. Davis, A.V. Joshi, H. Wang, F. Egolfopoulos, Proc. Combust. Inst. 30 (2005) 1283–1292.
[30] B.J. Isaac, A. Parente, C. Galletti, J.N. Thornock, P.J. Smith, L. Tognotti, Energy Fuels 27 (2013) 2255–2265.
[31] I.T. Jolliffe, Appl. Stat. 21 (1972) 160–173.
[32] W. Krzanowski, Appl. Stat. 36 (1987) 22–33.
[33] G.P. McCabe, Technometrics 26 (1984) 137–144.
[34] I. Cohen, Q. Tian, X. Sean, Z. Thomas, S. Huang, in: Proc. of the 15th Int. Conf. on Multimed., vol. 15, 2007, pp. 301–304.
[35] H. Steinhaus, Bulletin L'Acadmie Polonaise des Science 4 (1956) 801–804.
[36] I. Noda, J. Molec. Struct. 883–884 (2008) 216–227.
[37] H.C. Keun, T.M.D. Ebbels, H. Antti, M.B. Bollard, O. Beckonert, B. Holmes, J.C. Lindon, J.K. Nicholson, Anal. Chim. Acta 490 (2003) 265–276.
[38] P. Renard, D. Thévenin, J. Rolon, S. Candel, Progr. Energy Combust. Sci. 26 (2000) 225–282.
[39] M. Ó Conaire, H. Curran, J. Simmie, W. Pitz, C. Westbrook, Int. J. Chem. Kinet. 36 (2004) 603–622.