Dimensionality reduction and unsupervised classification for high-fidelity reacting flow simulations

Mohammad Rafi Malik[a,*], Ruslan Khamedov[a], Francisco E. Hernández Pérez[a], Axel Coussement[b], Alessandro Parente[b], Hong G. Im[a]

[a] *Clean Combustion Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia*

[b] *Université Libre de Bruxelles, Ecole Polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Brussels, Belgium*

Abstract

The development of reduced-order combustion models able to accurately reproduce the physics of reactive systems has been a highly challenging aspect of numerical combustion research in the recent past. The complexity of the problem can be reduced by identifying and using low-dimensional manifolds able to account for turbulence-chemistry interactions. Recently, Principal Components Analysis (PCA) has shown its potential in reducing the dimensionality of a chemically reactive system while minimizing the reconstruction error. The present work demonstrates the application of the Manifold Generated by Local PCA (MG-L-PCA) approach in direct numerical simulation (DNS) of turbulent flames. The approach is enhanced with an unsupervised clustering based on Vector Quantization PCA (VQPCA) and an *on-the-fly* PCA-based classification technique. The reduced model is then applied on a three-dimensional (3D) turbulent premixed NH_3/air flame 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. Results are compared with both a detailed reaction mechanism and a Computational Singular Perturbation (CSP) reduced skeletal mechanism. A comparison between training the reduced model using one-dimensional (1D) and 3D data sets is also included. Overall, the MG-L-PCA allows not only for a reduction in the number of transport equations, but also a significant reduction in the stiffness of the system, while providing highly accurate results.

Keywords: Numerical combustion; Principal Component Analysis; Low-dimensional manifolds; Direct numerical simulation

*Corresponding author.

Email address: mohammadrafi.malik@kaust.edu.sa (Mohammad Rafi Malik)

1. Introduction

Direct Numerical Simulation (DNS) using detailed chemistry is still an overly demanding task for most computational facilities due to the large number of species and reactions in the chemical system. To lower the computational cost associated with detailed kinetics and make the simulations more affordable, various strategies have been proposed in the combustion literature, which can be classified into two categories: reduction of the chemical mechanism and parameterization of the state-space with a reduced number of optimal variables.

The reduction of a chemical mechanism can be achieved using various techniques. Among those, Computational Singular Perturbation (CSP) theory [1] has gained attention in recent years for its ability to reduce the number of elementary reactions in a complex reaction system by grouping reactions that are characterized by a single characteristic time scale, and subsequently, discarding those reactions which are associated with fast time scales. By the process of identifying and discarding fast reaction groups, a simplified kinetic scheme for the complex chemical system is obtained.

Among the various state-space parameterization methods, Principal Component Analysis (PCA) [2] has proven to be a particularly elegant tool for this purpose. In this approach, empirical data sets are analyzed using PCA to identify a low-dimensional representation of the reacting system, where the low-dimensional manifold is parameterized in terms of a reduced number of optimal variables [3–9]. If the original thermo-chemical state-space consists of Q variables, PCA will provide a truncated set of q ($q < Q$) Principal Components (PCs) – which are linear combinations of the original variables – containing most of the information present in the original data. Therefore, PCA offers the ability to reduce the dimensionality of the system, accelerating the computation accordingly.

Two families of methods exist to use the manifold identified by PCA. In the PC-transport approach [4], the Principal Component scores are directly transported in the flow solver. However, this approach requires significant code modifications and special attention should be given to the PC source terms accuracy, requiring nonlinear regression techniques to reduce the number of components to transport [7–10].

The second method consists in solving the Navier–Stokes equations for a subset of species mass fractions and reconstructing the non-transported variables from the PCA basis. This is referred to as the Manifold Generated from PCA (MG-PCA) model [11, 12], which relies on classical transport equations for a reduced set of principal variables (PV). In particular, in [11] the model was coupled with local PCA [3], allowing a local application of the MG-PCA method in clusters, thus further reducing the number of transported species. The MG-L-PCA model was applied to a laminar two-dimensional (2D) hydrogen flame-vortex interaction case in a DNS solver, with the training data being generated using a one-dimensional (1D) freely propagating hydrogen-air laminar flame.

The present work focuses on MG-L-PCA in the context of DNS of three-dimensional (3D) turbulent flows. The objective is to demonstrate the approach *a posteriori*, coupling the method with an unsupervised classification of the state-space, for the definition and retrieval of optimal local parametrisations. Vector quantization PCA (VQPCA) is used both *a priori*, for the generation of local basis, and during the simulations, for the *on-the-fly* classification of the state-space. The training data set is generated using a collection of 1D strained premixed counterflow flames. The MG-L-PCA results are benchmarked against detailed DNS results and the DNS results obtained from a CSP reduced mechanism. The test case is a 3D premixed NH_3/air flame interacting with turbulence. To the authors' knowledge, the present work represents the first attempt to use the MG-L-PCA model in such a comprehensive and unsupervised manner.

2. Methodology

2.1. Principal Component Analysis (PCA)

The methodology behind PCA and MG-PCA is briefly introduced here. For detailed information, the reader may refer to [4–11].

For a data set $\mathbf{X}$ of dimension $(n \times Q)$, containing n samples of Q independent variables, PCA provides an approximation based on q ($q < Q$) linear correlations:

$$\mathbf{X} \approxeq \mathbf{X}_\mathrm{q} = \mathbf{Z}_\mathrm{q}\mathbf{A}_\mathrm{q}^\mathrm{t}, \tag{1}$$

where $\mathbf{X}_\mathrm{q}$ $(n \times Q)$ is the approximation of $\mathbf{X}$ based on the first q eigenvectors, $\mathbf{Z}_\mathrm{q}$ $(n \times q)$ are the principal components scores and $\mathbf{A}_\mathrm{q}^\mathrm{t}$ $(q \times Q)$ is the matrix of the first q eigenvectors, i.e. the truncated basis matrix (the superscript t denotes the transposed). In the present work, $\mathbf{X}$ contains the species mass fractions ($\mathbf{X} = [Y_i, ..., Y_{ns}]$) and it is assumed that $\mathbf{X}$ has been appropriately centered and scaled before PCA is carried out, as discussed thoroughly in [13].

In the MG-PCA approach, a subset of the original variables is solved using classical transport equations, and the remaining variables are recovered using PCA. Indeed, the scores $\mathbf{Z}_\mathrm{q}$ can be approximated from a subset q of the original variables:

$$\widetilde{\mathbf{Z}_\mathrm{q}} = \mathbf{X}(q)\,(\mathbf{A}(q)_\mathrm{q}^\mathrm{t})^{-1}, \tag{2}$$

where $\mathbf{X}(q)$ $(n \times q)$ contains only q variables, and $\mathbf{A}(q)_\mathrm{q}$ is a $(q \times q)$ matrix containing only the coefficients related to the q retained variables. Combining Eqs. (1) and (2), the $Q - q$ species mass fractions can be directly recovered from the q transported ones:

$$\mathbf{X}_\mathrm{q} = \mathbf{X}(q)\,(\mathbf{A}(q)_\mathrm{q}^\mathrm{t})^{-1}\,\mathbf{A}_\mathrm{q}^\mathrm{t} = \mathbf{X}(q)\mathbf{B}. \tag{3}$$

In the present work, the projection matrix $\mathbf{B}$ is retrieved following Isaac et al. [12]:

$$\mathbf{B} = \mathbf{X}(q)^{+}\mathbf{Z}_{\mathrm{q}}\mathbf{A}_{\mathrm{q}}^{\mathrm{t}}, \tag{4}$$

where $\mathbf{X}(q)^{+}$ is the pseudo-inverse of $\mathbf{X}(q)$. In the flow solver, the MG-PCA method is implemented as follows: (1) transport equations are solved for the q retained variables, (2) at the end of each temporal (or pseudo-temporal) iteration, the remaining $Q-q$ variables are reconstructed using Eq. (3), (3) all the species mass fractions are then available for the next iteration, and their source terms, diffusion coefficients and other mixture properties can be computed.

2.2. Local PCA (L-PCA)

For reacting flows, a larger number of PCs may be required to properly describe the system due to the intrinsically multi-linear nature of the technique [12]. To overcome this limit, a local PCA approach was introduced by Parente et al. [3], where the data set is divided into k clusters using an unsupervised algorithm based on Vector Quantization (VQ). The VQPCA algorithm partitions the data into separate clusters by minimizing the reconstruction error. Subsequently, PCA is performed in each cluster to find a local reduced representation. A global scaled reconstruction error (ε_{GSRE}) for each observation can be defined as [3]:

$$\varepsilon_{GSRE}\left(\widetilde{\mathbf{X}}_i, \overline{\mathbf{X}}^{(k)}\right) = \| \widetilde{\mathbf{X}}_i - \widetilde{\mathbf{X}}_{i,q} \|, \tag{5}$$

where $\widetilde{\mathbf{X}}_i$ is the scaled ith observation of the sample, $\overline{\mathbf{X}}^{(k)}$ is the kth cluster centroid and $\widetilde{\mathbf{X}}_{i,q}$ is the rank q approximation of $\widetilde{\mathbf{X}}_i$. The iterative VQPCA algorithm can be summarized as follows:

1. *Initialization*: the number of clusters k is set by the user and the clusters centroids $\overline{\mathbf{X}}^{(k)}$ are initialized;
2. *Partition*: each sample observation is assigned to the cluster satisfying Eq. 5;
3. *Update*: the new clusters centroids are computed on the basis of the partitioning from the previous step;
4. *Local PCA*: PCA is performed in each cluster;
5. Steps 2-4 are iterated until convergence is reached.

The use of local PCA is well-suited in the context of MG-L-PCA. Indeed, as the q retained variables are chosen based on the global data set, only the $\mathbf{B}$ matrices change between clusters. Therefore, the transport equations will not change from one cluster to another, but rather the local basis matrices $\mathbf{A}_{\mathrm{q,k}}^{\mathrm{t}}$, used to recover the missing $Q-q$ variables, will differ [11].

2.3. Classification via PCA reconstruction error

The PCA reconstruction error-based metric, defined as the difference between the original ($\mathbf{x}_i$) and the reconstructed ($\widetilde{\mathbf{x}}_i$) observation (i.e. $\varepsilon_r = \|\widetilde{\mathbf{x}}_i - \mathbf{x}_i\|$) can also be adapted for the *on-the-fly* classification of a new, unobserved, vector $\mathbf{y} \in \mathbb{R}^q$. During the simulation, in each cell, the vector $\mathbf{y}$ representing the state-space is projected onto the k local manifolds described by the pre-computed local basis matrices $\mathbf{A}_{\mathrm{q,k}}^{\mathrm{t}}$. Each cell is then classified to the cluster for which the reconstruction error is minimized [14]. Then, the missing $Q-q$ variables are computed by selecting the $\mathbf{B}$ matrix associated to each cluster. This procedure allows for a fast and efficient classification thanks to the efficiency of the on-the-fly PCA algorithm. Figure 1 shows the procedure graphically: starting from an existing clustering solution, the new observation to be classified is projected onto each local PCA bases, and the reconstruction error is computed for each projection. The observation is then assigned to the cluster that minimizes ε_r.

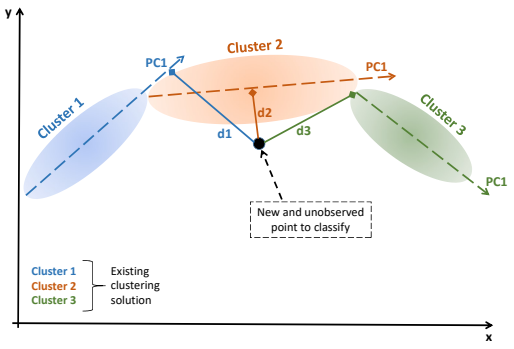


Figure 1: Classification procedure via PCA reconstruction error (adapted from [14]).

2.4. Variables selection via PCA

The selection of the subset of variables to be transported is a key step for the quality of the MG-L-PCA results [12]. Indeed, the retained subset of species should allow for a quasi-exact reconstruction of the entire state-space to avoid numerical errors in the source terms computation. In the present work, the B2 backward method is used [15]. The B2 algorithm eliminates variables by analyzing the weights distribution on the PCs: the weights on the last $Q-q$ PCs are inspected (those associated with small eigenvalues), and the variables with the highest weights on those PCs are discarded. In a previous study, the B2 method was shown to be the most robust and providing the best approximation of the state-space [12].

3. Numerical setup

MG-L-PCA was implemented and applied in the simulation of turbulent reacting flows. The 3D DNS computations were carried out using the KARFS code [16, 17] developed at KAUST, which solves the compressible Navier–Stokes equations fully explicitly. A fourth-order, six-stage explicit Runge–Kutta scheme is used for time integration, while the spatial discretization is done using an eighth-order, non-dissipative, central difference operator. For stiff chemistry problems, a stiff-ODE solver and a second-order operator splitting algorithm are also available. The code also includes an implementation of the Navier–Stokes characteristic boundary conditions (NSCBC) [18].

The chosen case is a premixed, rich, NH_3/air *flame in a box* configuration, schematically illustrated in Fig. 2. The equivalence ratio is $\phi = 1.2$, the pressure is $P = 1$ atm and the unburned temperature is $T_u = 500$ K. The computational domain is discretized with (168, 64, 64) uniform grid points in the (x, y, z) directions, respectively. The grid resolution is chosen such that $\Delta x = \Delta y = \Delta z = 1.09 \times 10^{-4}$ m $\leq 2.1\eta_k$, where η_k is the Kolmogorov length scale, and the thermal flame thickness is resolved with more than 12 grid points. The turbulent Reynolds, Damkhöler, Karlovitz, and Lewis numbers are respectively 72, 0.1, 85, and 1.12. Periodic boundary conditions are used in the y and z directions, while Navier–Stokes characteristic boundary conditions are applied for the outflow.

The simulation is initialized using the solution from a 1D premixed flame (at the same equivalence ratio, pressure and temperature), mapped onto the 3D domain. An initial isotropic turbulence field was generated with a spectrum function [19], which was then fed at the inlet ($x = 0$). The turbulent forcing described by Bassenne et al. [20] is applied in the unburned region to retain the initial turbulence level. Moreover, the mean inflow velocity is properly adjusted based on the fuel consumption speed, to anchor the flame at a specified location within the computational domain [21]. Thermo-chemical and transport properties were computed using the Cantera library [22]. Ammonia combustion was described with a detailed kinetic mechanism involving 38 species and 263 reactions, developed by Zhang et al. [23]. In addition, a skeletal mechanism consisting of 25 species and 175 reactions, which was obtained using CSP reduction, was also employed for comparison with the MG-L-PCA approach.

4. Training data and manifold

The training data set for the MG-L-PCA model was generated using a 1D laminar premixed counterflow flame setup. Simulations were performed using the Cantera software package [22]. This particular setup simulates two counterflow jets of reactants impinging onto each other. The initial thermo-chemical

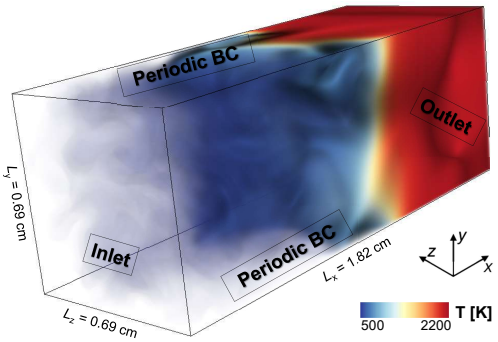


Figure 2: *Flame in a box* configuration used in the current DNS work.

conditions were the same as for the 3D DNS case described in Section 3, and the detailed kinetic mechanism by Zhang et al. [23] was used. The inlet axial velocity was varied in order to generate a collection of strained flame solutions, which were saved on a grid of 60 points over a 25 mm domain. All of the data points were used collectively for the PCA analysis. The final data set consisted of $\sim 30,000$ observations for each of the state-space variables.

In the MG-L-PCA approach, a subset of variables are transported and, therefore, the latter must allow a quasi-exact reconstruction of the non-transported variables in order to avoid nonphysical behavior of the flame. Using the B2 backward selection method [15] with pareto scaling [13], 21 species were identified. The VQPCA analysis showed that by using 3 clusters, the retained 21 species allowed for a quasi-exact reconstruction ($R^2 > 0.999$) of the other non-transported species. The minimum number of retained species is constrained by the reconstruction accuracy of the discarded species, while the maximum number of clusters is constrained by the accuracy of the *on-the-fly* classification. It was found that the 21 retained species (using the B2 backward method) and 3 clusters (identified in an unsupervised manner using VQPCA) provided the best compromise between accuracy in the reconstruction of the remaining species and accuracy in the on the fly classification algorithm. The different sets of species used for the detailed, CSP-reduced and MG-L-PCA simulations, respectively, are reported in Table 1.

Two different MG-L-PCA models were built using two different data sets of increasing complexity: one using the 1D strained counterflow data set described above (hereafter called MG-L-PCA - 1D training), and a second one using the data obtained from the detailed DNS simulation (hereafter called MG-L-PCA - 3D training). The number of transport equations for the different cases are summarized in Table 2 (for all simulations Y_{N_2} was computed using mass balance).

Table 1: List of transported species used for the detailed, CSP-reduced and MG-L-PCA simulations.

Method	No. of species	Transported species
Detailed mechanism by Zhang et al. [23]	38	$AR, H, H_2, HE, O, OH, H_2O, HO_2, H_2O_2, O_2, OH^*, N, NH, NH_2, NH_3, NNH, N_2H_2, N_2H_3, N_2H_4, H_2NN, NO, N_2O, NO_2, HNO, HON, HONO, HNO, HNOH, NH_2OH, HNO_2, HONO_2, NO_3, HNO_3, CO, CO_2, CH_4, C_2H_6, N_2$
CSP-reduced skeletal mechanism	25	$H, H_2, O, OH, H_2O, HO_2, H_2O_2, O_2, N, NH, NH_2, NH_3, NNH, N_2H_2, N_2H_4, H_2NN, NO, N_2O, NO_2, HNO, HONO, H_2NO, HNOH, HNO_2, N_2$
MG-L-PCA reduced model	21	$H, H_2, O, OH, H_2O, HO_2, H_2O_2, O_2, N, NH, NH_2, NH_3, N_2H_2, N_2H_3, N_2H_4, NO, N_2O, NO_2, HNO, H_2NO, HNOH$

Table 2: Number of transport equations for the different DNS cases.

Chemistry description	No. of equations
Detailed mechanism	$5+37$
CSP-reduced	$5+24$
MG-L-PCA - 1D training	$5+21$
MG-L-PCA - 3D training	$5+21$

5. Results

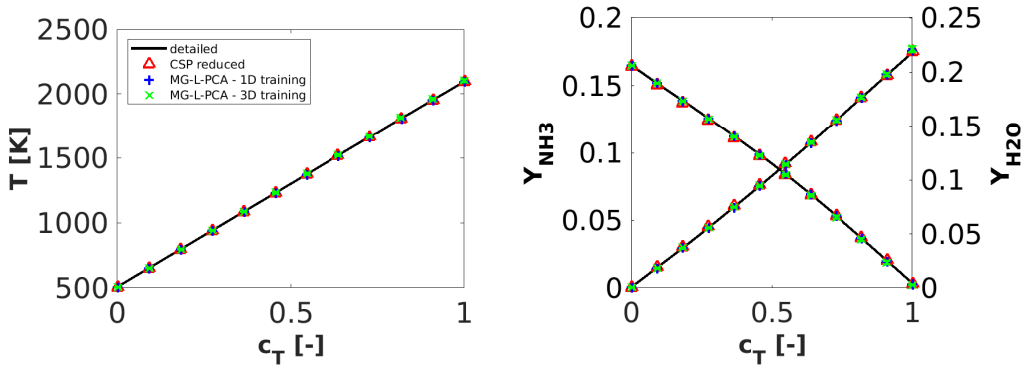
In this section, results obtained from the DNS computations using the detailed mechanism, CSP-reduced skeletal mechanism and MG-L-PCA reduced model are all compared for a 3D ammonia/air turbulent premixed flame.

Figures 3 and 4 show the conditionally-averaged profiles of temperature along with major species (Fig. 3), and intermediate and minor species (Fig. 4), plotted as a function of the temperature-based progress variable c_T, for each of the four cases described in Table 2. The solution from the detailed DNS is represented by solid lines (mean) and shaded regions (standard deviation to the mean), while the symbols and their associated error bars show the mean and standard deviation for the three reduced models.

It is observed in Fig. 3 that the three DNS using reduced models show nearly identical results to the detailed DNS for the predictions of major state-space variables. In particular, the MG-L-PCA model is able to predict the temperature and major species mass fractions as accurately as the CSP method over the entire range of the progress variable, but at a lower computational cost. This decrease in CPU cost is due to two reasons: 1) a smaller number of species transport equations are required (21 instead of 37, representing a 43% reduction), and 2) the stiffness of the system is also considerably reduced as the DNS simulations with the MG-L-PCA models were able to run using a timestep five times larger than the timestep used for the detailed and skeletal (from CSP) mechanisms. To support this, a more rigorous timescale analysis is presented in Section 5.1. Figure 3 also indicates that there are no significant differences between the MG-L-PCA models trained on 1D or 3D data sets for major state-space variables.

Regarding Figure 4, when referring to the MG-L-PCA models, intermediate and minor species are divided into two categories: transported and non-transported species. Figures 4a-b show the results for two transported species: Y_{N_2O} and $Y_{N_2H_4}$. It is seen that those species are accurately captured by both MG-L-PCA models, showing good agreement between the 1D and 3D trainings. In particular, in the case of $Y_{N_2H_4}$, the MG-L-PCA models show improved accuracy compared to the CSP method. For non-transported species (Fig. 4c-d), a clear advantage of using a 3D training data set is evident. Indeed, those non-transported minor species are poorly predicted by the 1D trained MG-L-PCA, whereas the 3D trained model captures the evolution of those species better. This suggests that the complex transport phenomena present in the 3D DNS data set allows to build a more accurate **B** matrix for the recovery of non-transported variables, compared to the laminar strained 1D flames. After all, utilization of the 1D data set relies on the flamelet assumption, implying an additional level of approximation and potential inaccuracy. It is noted that, although Y_{H_2NN} is not transported, it is nevertheless better predicted by the 3D trained MG-L-PCA model as well as, and slightly better than, the CSP-reduced model, in which Y_{H_2NN} is explicitly transported and solved.

Figure 5 shows scatter plots of the chemical source terms of one major transported species ($\dot{\omega}_{NH_3}$, top row) and a minor one ($\dot{\omega}_{N_2H_4}$, bottom row), comparing the detailed, CSP reduced and MG-L-PCA results respectively. The source terms were extracted over the entire duration of the simulations. It can be observed that the MG-L-PCA model is able to accurately predict the chemical source terms when compared to the detailed results, even outperforming the CSP reduced model in the case of $\dot{\omega}_{N_2H_4}$. Therefore, the observed level of inaccuracy in the reconstruction of minor species for the MG-L-PCA model does not significantly affect the calculations of the source



(a) Temperature

(b) Y_{NH_3} and Y_{H_2O}

Figure 3: Comparison of conditional average of temperature (a) and major species (b) as a function of the temperature-based progress variable c_T. The solid line indicates the mean for the detailed case, and shaded region is the standard deviation to the mean. Symbols and associated error bars indicate the mean and standard deviation for CSP and MG-L-PCA cases, respectively.

terms of transported variables.

Table 3 shows the normalized root-mean-square (*nrms*) error values as a quantitative assessment of the reduced models performances. For some scalars, the MG-L-PCA trained on 1D flames is able to provide more accurate results than both the CSP and the MG-L-PCA trained on a 3D DNS data set, exhibiting 2 to 4 times lower *nrms* values. This suggests that MG-L-PCA model can be trained on simple configurations using detailed mechanisms. Therefore, a 1D trained model represents a feasible and generalizable approach providing very accurate results for major variables. The 1D flame based training yielding a lower error for the major variables (compared to the 3D DNS training) could be due to the increased complexity and curvature of the DNS data, as the latter inherently contains the effect of turbulence and mixing. This in turn leads to a more challenging manifold, requiring a higher number of clusters to accurately compute the **B** matrix in each of them. The major variables could be more sensitive to this effect as they are more strongly affected by large scale turbulence. Regarding the minor non-transported species, it should be pointed out that a classical mechanism reduction technique completely removes those species from the system, whereas MG-PCA retains them and still provides a reasonable prediction of those without solving for their transport equations. Given that the minor species are only estimated, it is possible that a 3D data set provides better results. However, the accuracy of the minor non-transported species, although essential for the source terms computations, does not seem to affect the predictions of the major species and temperature.

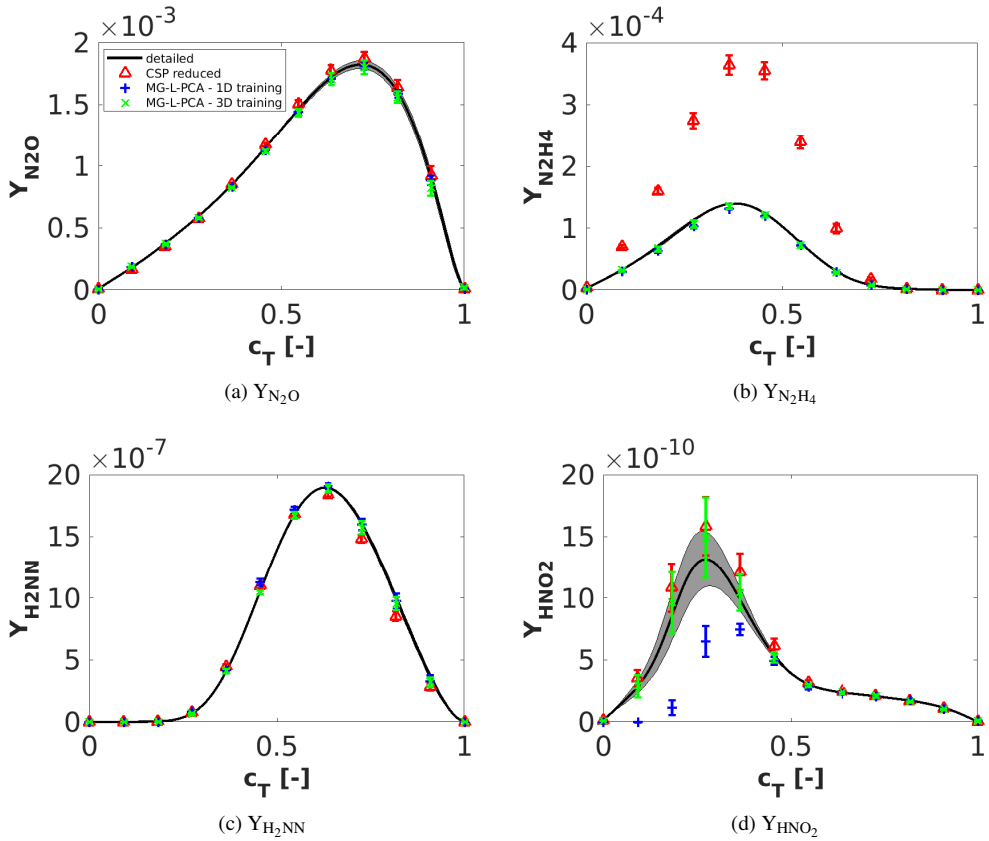
Table 3: Normalized root-mean-square (*nrms*) error values for the variables plotted in Figs. 3 and 4.

Variable	CSP	1D training	3D training
T	0.47%	0.22%	1.30%
Y_{NH_3}	1.79%	0.47%	1.73%
Y_{H_2O}	1.71%	0.84%	1.14%
Y_{N_2O}	5.57%	2.22%	5.74%
$Y_{N_2H_4}$	247.5%	9.40%	5.95%
Y_{H_2NN}	9.72%	4.41%	3.34%
Y_{HNO_2}	32.9%	79.56%	18.42%

5.1. Timescales analysis

A chemical timescale analysis is performed here using the approach for complex kinetic schemes proposed by Fox [24], where the chemical timescales are computed based on the inverse of the eigenvalues of the chemical source term Jacobian matrix. A comparison is conducted for the detailed, CSP and MG-L-PCA cases for a 0D case and for the 3D DNS cases. The 0D simulations were carried out using a constant-volume reactor in KARFS. The Jacobian matrix and timescales calculations were performed using the *in-house* MATLAB code JACOBEN [25]. The MG-L-PCA case was run using the same number of species as the CSP case, to allow a more consistent comparison.

Figure 6a shows the minimum chemical time scales ($\tau_{c,min}$) evolution with each time-step associated with the 0D reactor for the three cases under consideration, namely the detailed mechanism (38 species), the CSP-reduced mechanism (25 species) and the MG-L-PCA model (1D training, 25 species, 3 clusters). For the cases under study, the MG-L-



(a) Y_{N_2O} (b) $Y_{N_2H_4}$

(c) Y_{H_2NN} (d) Y_{HNO_2}

Figure 4: Comparison of conditional average of minor species as a function of the temperature-based progress variable c_T. The solid line indicates the mean for the detailed case, and shaded region is the standard deviation to the mean. Symbols and associated error bars indicate the mean and standard deviation for CSP and MG-L-PCA cases, respectively.

PCA model has an advantage when examining the limiting chemical timescales. Indeed, the stiffness of the system is considerably reduced, indicating an increase in the minimum chemical timescale of almost 2 orders of magnitude, compared to the detailed and CSP systems. This reduction in stiffness offered by MG-L-PCA is particularly appealing in the context of DNS, especially when an explicit time integrator is used. The same analysis is also shown for the 3D DNS cases in Fig. 6b, where the averaged minimum chemical time scales (for all eddy turnover times of the simulation) are plotted as a function of temperature. Again, a significant decrease in the stiffness of the system can be observed for the MG-L-PCA case, whereas the detailed and CSP-reduced cases exhibit similar behaviour except in the mid-range temperature region where the CSP shows a slight increase in timescales.

6. Conclusions

The present work demonstrated the application of the MG-L-PCA model, coupled with an unsupervised clustering based on Vector Quantization PCA (VQPCA) and with an *on-the-fly* PCA-based classification. The reduced model was trained using two different data sets of increasing complexity (1D laminar strained flames and 3D DNS data). The model was tested on a 3D DNS case of a premixed NH_3/air flame in a turbulence field (*flame in a box* configuration), and compared to DNS results using a detailed mechanism (38 species) and a CSP-reduced skeletal mechanism (25 species).

The MG-L-PCA showed a remarkable accuracy for the predictions of temperature and major and minor species mass fractions. Using 21 transported species out of the original 38, the model was able to provide highly accurate results when compared to the detailed and the CSP simulations. For the predictions of temperature and major species, the 1D- and 3D-trained models showed similar accuracies, with

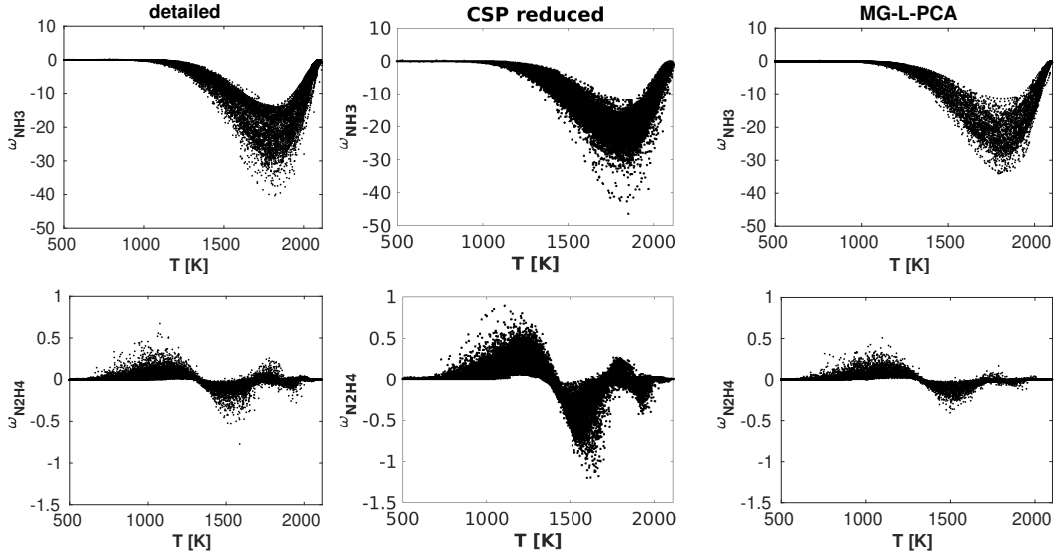


Figure 5: Comparison of $\dot{\omega}_{\mathrm{NH_3}}$ (top row) and $\dot{\omega}_{\mathrm{N_2H_4}}$ (bottom row) in $\mathrm{kg}/(\mathrm{m}^3\mathrm{s})$ for the detailed, CSP and MG-L-PCA cases, respectively. Points were sampled down for clarity.

the 1D-trained model slightly outperforming the 3D-trained one. On the other hand, for some intermediate and minor species the 3D-trained model outperformed the 1D-trained one, especially in the case of non-transported species. This suggests that a data set with a higher complexity used in an *a priori* manner leads to a more accurate reconstruction matrix.

A chemical timescales analysis also showed that the chemistry stiffness of the system is considerably reduced with the MG-L-PCA model, compared to the detailed and CSP-reduced mechanisms.

Future work will focus on the generalisation of the non-transported species representation using nonlinear regression techniques, with the objective of increasing the accuracy and reducing the cost of numerical simulations.

Acknowledgments

This work was sponsored by King Abdullah University of Science and Technology (KAUST). Computational resources were provided by the KAUST Supercomputing Laboratory (KSL). The work of A. Parente has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 714605.

References

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

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

[3] A. Parente, J. C. Sutherland, L. Tognotti, P. J. Smith, Identification of low-dimensional manifolds in turbulent flames, Proceedings of the Combustion Institute 32 (1) (2009) 1579–1586.

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

[5] A. Coussement, B. Isaac, O. Gicquel, A. Parente, Assessment of chemistry reduction methods based on PCA: comparison of MG-PCA and score-PCA approaches, Combust. Flame 168 (2016) 83–97.

[6] T. Echekki, H. Mirgolbabaei, Principal component transport in turbulent combustion: a posteriori analysis, Combust. Flame 162 (5) (2015) 1919–1933.

[7] B. Isaac, J. Thornock, J. Sutherland, P. Smith, A. Parente, Advanced regression methods for combustion modelling using principal components, Combust. Flame 162 (6) (2015) 2592–2601.

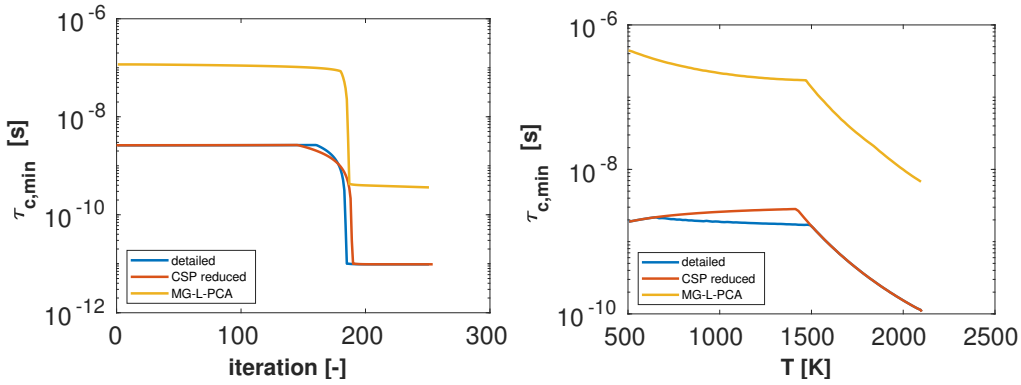
[8] M. R. Malik, B. Isaac, A. Coussement, P. Smith, A. Parente, Principal component analysis coupled with nonlinear regression for chemistry reduction, Combust. Flame 187 (2018) 30–41.

[9] M. R. Malik, P. O. Vega, A. Coussement, A. Parente, Combustion modeling using principal component analysis: a posteriori validation on Sandia flames D, E and F, Proceedings of the Combustion Institute 38 (2) (2021) 2635–2643.

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

[11] A. Coussement, O. Gicquel, A. Parente, MG-local-PCA method for reduced order combustion modelling, Proc. Combust. Inst. 34 (2013) 1117–1123.

[12] B. Isaac, A. Coussement, O. Gicquel, P. Smith, A. Par-



(a) 0D reactor (b) 3D DNS

Figure 6: Minimum chemical timescales ($\tau_{c,min}$) for the 0D cases (left) and 3D cases (right), utilizing the detailed mechanism, CSP-reduced mechanism and MG-L-PCA reduced model.

ente, Reduced-order PCA models for chemical reacting flows, Combust. Flame 161 (11) (2014) 2785–2800.

[13] A. Parente, J. C. Sutherland, Principal component analysis of turbulent combustion data: Data preprocessing and manifold sensitivity, Combust. Flame 160 (2) (2013) 340–350.

[14] G. D'Alessio, A. Parente, A. Stagni, A. Cuoci, Adaptive chemistry via pre-partitioning of composition space and mechanism reduction, Combust. Flame 211 (2020) 68–82.

[15] I. Jolliffe, Discarding variables in a principal component analysis. i: Artificial data, Journal of the Royal Statistical Society Series C 21 (2) (1972) 160–173.

[16] F. E. Hernández Pérez, N. Mukhadiyev, X. Xu, A. Sow, B. J. Lee, R. Sankaran, H. G. Im, Direct numerical simulations of reacting flows with detailed chemistry using many-core/GPU acceleration, Computers & Fluids 173 (2018) 73–79.

[17] S. Desai, Y. J. Kim, W. Song, M. B. Luong, F. E. Hernández Pérez, R. Sankaran, H. G. Im, Direct numerical simulations of turbulent reacting flows with shock waves and stiff chemistry using many-core/GPU acceleration, Comput. Fluids 215 (2020) 104787.

[18] C. S. Yoo, H. G. Im, Characteristic boundary conditions for simulations of compressible reacting flows with multi-dimensional, viscous and reaction effects, Combustion Theory and Modelling 11 (2) (2007) 259–286.

[19] T. Passot, A. Pouquet, Numerical simulation of compressible homogeneous flows in the turbulent regime, Journal of Fluid Mechanics 181 (1987) 441–466.

[20] M. Bassenne, J. Urzay, G. I. Park, P. Moin, Constant-energetics physical-space forcing methods for improved convergence to homogeneous-isotropic turbulence with application to particle-laden flows, Physics of Fluids 28 (3) (2016) 1–16.

[21] J. B. Bell, M. S. Day, J. F. Grcar, M. J. Lijewski, Active control for statistically stationary turbulent premixed-flame simulations, C.A.M.C.S 1 (1) (2006) 29–51.

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

[23] X. Zhang, S. P. Moosakutty, R. P. Rajan, M. Younes, S. M. Sarathy, Combustion chemistry of ammonia/hydrogen mixtures: Jet-stirred reactor measurements and comprehensive kinetic modeling, Combust. Flame 234 (2021) 111653.

[24] R. Fox, Computational Models for Turbulent Reacting Flows, Cambridge University Press, 2003.

[25] B. J. Isaac, A. Parente, C. Galletti, J. Thornock, P. J. Smith, L. Tognotti, A novel methodology for chemical time scale evaluation with detailed chemical reaction kinetics, Energy & Fuels 27 (2013) 2255–2265.