https://doi.org/10.1016/j.combustflame.2019.09.010

Adaptive chemistry via pre-partitioning of composition space and mechanism reduction

Giuseppe D'Alessio[a,b], Alessandro Parente[*a,c], Alessandro Stagni[b], Alberto Cuoci[†b]

[a] *Université Libre de Bruxelles, École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Bruxelles, Belgium*

[b] *CRECK Modeling Lab, Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20131 Milano, Italy*

[c] *Université Libre de Bruxelles and Vrije Universiteit Brussel, Combustion and Robust Optimization Group (BURN), Brussels, Belgium*

Abstract

Numerical simulations of multi-dimensional laminar flames with complex kinetic mechanisms are computationally very demanding, because of the large number of species and the strong non-linearity and stiffness of governing equations. In this work, we present and apply a novel adaptive chemistry methodology for mitigating the computational cost of such simulations, based on machine-learning algorithms which automatically classify the composition space via *a priori* defined classifiers.

The methodology, called SPARC (Sample-Partitioning Adaptive Reduced Chemistry), is based on four steps: generation of data sets covering the temperature and composition space which is expected to be visited by the multi-dimensional flame; partitioning of the composition space in a prescribed number of clusters with similar composition via Local Principal Component Analysis (LPCA); generation of reduced kinetic mechanisms for each cluster via Directed Relation Graph with Error Propagation (DRGEP); CFD simulation of a multi-dimensional flame based on locally reduced mechanisms.

The approach has been firstly demonstrated for the CFD simulations of steady and transient laminar coflow diffusion flames fed with a mixture of CH_4 and N_2 burning in air. The transient behaviour was artificially induced by different sinusoidal perturbations in the velocity profiles. Several numerical tests were carried out to explore the impact of partitioning parameters and the degree of mechanism reduction on the flame simulation, and very satisfactory results were observed, both in terms of accuracy and computational

[*]Corresponding author: alessandro.parente@ulb.be

[†]Corresponding author: alberto.cuoci@polimi.it

efficiency. In particular, even if a relatively small mechanism (84 species) was adopted, speed-up factors of ~ 4 were observed. Larger speed-up factors can be achieved if more complex fuels are considered, since also larger and stiffer kinetic mechanisms are required and a higher level of reduction via DRGEP can be reached.

Keywords: Adaptive-chemistry, Principal Component Analysis, Mechanism reduction, DRGEP, Laminar flame

---

1. Introduction

Nowadays, detailed kinetic mechanisms are today recognized as a key element for the predictive simulation of laminar and turbulent flames, aimed at a more effective design of combustion devices (such as internal combustion engines and gas turbines). However, the coupling of Computational Fluid Dynamics (CFD) and detailed chemistry is computationally demanding, mainly because of the large number of species and the wide range of chemical times typically involved in complex chemistry [1]. During the last years, several techniques have been proposed for reducing the computational cost associated to detailed kinetic mechanisms. Some of the most commonly adopted include skeletal reduction approaches to generate mechanisms with a reduced number of both species and reactions [2]. However, they are not always sufficient to bring the computational cost at an affordable level, especially in case of complex fuels.

When inhomogeneous reactive flows are solved by numerical algorithms based on the operator-splitting strategy [3], more effective techniques can be considered for reducing the computational cost of detailed chemistry. In operator-splitting approaches, most of the CPU time is tipically spent in the chemical step, which corresponds to the solution of N (with N being the total number of computational cells) independent, stiff, non linear systems of ordinary differential equations (ODEs):

$$\frac{d\boldsymbol{\psi}}{dt} = \boldsymbol{S}(\boldsymbol{\psi}) \tag{1}$$

where $\boldsymbol{\psi}$ is the thermo-chemical state vector (composition and temperature) in a single cell and $\boldsymbol{S}(\boldsymbol{\psi})$ the corresponding rate of change due to chemical reactions. Since the ODE systems above are independent, *adaptive chemistry* techniques are very effective and easy

to implement. They are based on the observation that in any small range of temperature and compositions many species have negligible concentrations and only a limited number of them is chemically active [4]. This means that the kinetic mechanism locally needed to describe the chemical evolution may require fewer species and reactions than the complete mechanism applicable over the whole thermo-chemical space. In principle, the maximum level of adaptivity (i.e. maximum level of reduction) can be achieved through the Dynamic Adaptive Chemistry (DAC) approach [5, 6]: at each time step of the simulation, the local thermo-chemical state is analyzed, the important reactions and species are recognized, and only the corresponding equations are solved. This means that the evolution in each cell is described through a local reduced mechanism. Unfortunately, the computational overhead needed for the on-the-fly reduction of the mechanism can be significant (especially for large mechanisms).

In order to reduce the cost associated to the adaptive reduction, it is possible to build a library of reduced mechanisms in a pre-processing step, covering the composition space which is expected to be visited by the flame of interest [7, 8]. The generated composition space can be pre-partitioned in a prescribed number of *clusters*, for which a library of reduced kinetic mechanisms is created. The difficulty with this approach is to find the optimal way to partition the derivation of reduced kinetic mechanisms, ensuring that each individual cluster is sufficiently homogeneous from the kinetic point of view, i.e. that a single reduced kinetic mechanism is able to correctly describe all the samples belonging to it [4].

In this paper, a new methodology for the efficient inclusion of detailed chemistry in multidimensional CFD simulations, based on the coupling between Local Principal Component Analysis (LPCA) [9] and mechanism reduction via DRGEP (Directed Relation Graph with Error Propagation) [10], is presented. Instead of performing *on-the-fly* chemical reduction, a classification of the composition space which will be visited during the CFD simulation is carried out in a pre-processing step. Then, the generated composition space is pre-partitioned in a prescribed number of clusters via LPCA and for each cluster a reduced kinetic mechanism is generated via DRGEP. The CFD simulation is thus carried out by transporting the whole set of species in the detailed mechanism. Before carrying out the chemical step, each cell is *classified* using the same unsupervised-learning

algorithm adopted for pre-partitioning, in order to identify the cluster to which the cell belongs and select the corresponding reduced mechanism.

The paper is organized as reported in the following. In Section 2 the overall procedure used for the adaptive chemistry is described step-by-step. Then, in Section 3 the results obtained from the dataset partitioning, the mechanism reduction and the adaptive CFD simulations are shown and discussed. Lastly, in section 4 the possibility to use data obtained from lower dimensional simulations (0D and 1D) to perform multidimensional simulations by means of an adaptive-chemistry approach is investigated.

2. Approach

The overall procedure, called SPARC (Sample-Partitioning Adaptive Reduced Chemistry), consists of three training stages, followed by the final step, i.e. the multi-dimensional CFD simulation:

1. Dataset generation;
2. Clustering via Local Principal Component Analysis (LPCA);
3. Generation of reduced kinetic mechanisms via DRGEP for each cluster resulting from the LPCA partitioning;
4. CFD simulation of a case of interest.

2.1. Dataset generation

The dataset generation is a step of crucial importance for successfully applying the LPCA and the reduction procedures. In order to be valuable for the proposed strategy, the selected dataset must adequately cover the composition space which is expected to be visited during the simulation of the problem of interest. The training of LPCA algorithm and the chemical reduction can be done either using a dataset composed by previous detailed simulations on the system of interest, or simple 0D/1D simulations.

Because of the multivariate nature of the original dataset, preprocessing in the form of centering and scaling the variables before applying LPCA is a mandatory operation to get reliable results [11]. By means of these two operations, the original set of thermo-chemical variables $\boldsymbol{\xi}$ is standardized, and a new set of variables $\boldsymbol{\phi}$ is obtained.

2.2. Clustering via Local Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique used to reduce a large number of interdependent variables to a smaller number of uncorrelated variables, while retaining as much as possible of the original data variance [12].

Starting from a dataset $\mathbf{X}$ of n rows (observations) and p columns (variables), it is possible to compute the covariance matrix $\mathbf{S} = \frac{1}{(n-1)}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, and then to decompose it: $\mathbf{S} = \mathbf{ALA^T}$. The eigenvectors of the covariance matrix, which are the columns of the $\mathbf{A}$ matrix, are called Principal Components (PCs) while the eigenvalues, which are located on the principal diagonal of the $\mathbf{L}$ matrix, represent the portion of variance they account for. The dimensionality reduction comes by considering only a subset of the PCs, whose size is chosen depending on the variance accounted for. Since each PC is associated to a certain percentage of variance by the magnitude of its associated eigenvalue, it is possible to measure the global variance explained by a subset of PCs defining the ratio: $t_q = \sum_{j=1}^{q} l_j / \sum_{j=1}^{p} l_j$, where q is the chosen size of the subset and l_j is the jth eigenvalue obtained from the decomposition of the covariance matrix $\mathbf{S}$.

Although the effectiveness of PCA-based techniques in combustion applications has been already proved both *a priori* and *a posteriori* [13, 14, 11, 15], it has to be mentioned that for strongly non-linear systems, such as reacting flows, a large number of PCs may be required in order to properly describe the system. The motivation is mainly due to the intrinsically multi-linear nature of the technique [13, 16]. In order to overcome this limit, it is possible to adopt a locally linear approach, introduced by Kambhatla for image-processing applications [9] and previously used in many scientific fields [17, 18, 19], called Local Principal Component Analysis (LPCA). LPCA is an unsupervised-learning algorithm able to partition the data-space in different regions (clusters), where the variables share common features. Then, in each cluster PCA is performed to find a reduced representation. The higher effectiveness of LPCA in dimensionality reduction, compared to PCA, has already been proved in the field of combustion [13, 20].

The partitioning of the data-space into different clusters is achieved by means of an iterative algorithm minimizing the reconstruction error, $\epsilon_{\mathrm{GRSE,k}} = \|\tilde{\boldsymbol{x}}_{\mathrm{i}} - \tilde{\boldsymbol{x}}_{\mathrm{i,q}}\|$. Vectors $\tilde{\boldsymbol{x}}_{\mathrm{i}}$ and $\tilde{\boldsymbol{x}}_{\mathrm{i,q}}$ are, respectively, the original and the reconstructed, via PCA, ith centered and scaled observation of the dataset [20]. The LPCA algorithm can be summarized as

follows:

1. *Initialization:* the initial clusters centroids $\bar{\boldsymbol{x}}_{\boldsymbol{k}}$ are chosen from a k-means [21] solution. The eigenvectors matrix in each cluster, $\boldsymbol{s}_{\boldsymbol{k}}$, is initialized as the identity matrix.
2. *Partition:* each observation is assigned to a cluster by means of the calculation of the $\epsilon_{GRSE,k}$.
3. *Update:* the cluster centroids are updated on the basis of the partitioning carried out at step 2.
4. *Local-PCA:* Principal Component Analysis is performed in each cluster found at step 2.
5. Steps from 2 to 4 are iterated until one of the convergence criteria is met.

Convergence is reached if one of the following criteria is met:

1. The global mean reconstruction error, defined as: $\epsilon_{GRSE,K} = \frac{1}{K}\sum_{k=1}^{K}\epsilon_{GRSE,k}$, is below a fixed threshold: $\epsilon_{GRSE,K} < \epsilon^{*}_{GRSE,K}$, with K equal to the total number of clusters.
2. The variation of the position of all the cluster centroids between two consecutive iterations is below a fixed threshold (δ): $\Delta\bar{\boldsymbol{x}}_{\boldsymbol{k}} < \delta$.
3. The $\Delta\epsilon_{GRSE,K}$ between two consecutive iterations is below a fixed threshold (ρ): $\Delta\epsilon_{GRSE,K} < \rho$.

The initialization of the LPCA algorithm from a partition based on k-Means can drastically reduce the CPU-time required to reach convergence, compared to a random or uniform initialization of the centroids. The k-Means algorithm is only used for the initialization task because the Euclidean distance, used as objective function, is not optimal as the reconstruction error used in LPCA to define groups of similar points in a high-dimensional space. Moreover, k-Means can often provide clusters with lower a number of observations than variables, not ensuring the statistical meaningfulness of the defined clusters.

2.3. Reduction via DRGEP

For each cluster identified through the partitioning procedure described above, a reduced mechanism is generated via DRGEP (Directed Relation Graph with Error Propagation)

[10]. In particular for each cluster k:

1. After defining the target tolerance ϵ and the list of target species, a reduced mechanism is generated for each sample point belonging to the cluster k;
2. A single reduced mechanism for cluster k is generated from the union of the reduced mechanisms of individual samples belonging to the cluster k.

If the partitioning by means of LPCA is correctly carried out, the individual mechanisms generated for each sample are expected to be very similar, because their compositions are similar too. Thus, their union is expected to produce a final mechanism that is not significantly larger, in terms of number of species and reactions, than the individual mechanisms.

Considering an ideal case in which the clusters are perfectly uniform from a reacting point of view, all the individual generated mechanisms would be identical. Because this almost never happens, a measure of non-uniformity must be introduced. If n_{it} is the number of observations in a given cluster and n_{sp} is the number of species resulting from the union of individual mechanisms, it is possible to define $x_i = \frac{1}{n_{it}} \sum_{j=1}^{n_{it}} \delta_{i,j}$, where $\delta_{i,j}$ is equal to 1 if species i is contained in the reduced mechanism for sample j, or equal to 0 if not. Then, for each group a similarity coefficient can be defined as: $\lambda = \frac{1}{n_{sp}} \sum_{i=1}^{n_{sp}} (x_i - 1)^2$. The coefficient λ is equal to 0 in case of perfect uniformity, and its upper bound is 1.

2.4. Adaptive simulation

The last step of the procedure is the CFD simulation of the reacting system of interest. The strategy is particularly effective if operator-splitting (or fractional) algorithms are adopted. In particular, at the beginning of the chemical step, each computational cell is classified, i.e. the cluster to which it belongs is recognized.

The classification of each cell is performed by means of an on-the-fly PCA classification, which consists in assigning each grid point to the pre-computed cluster for which the reconstruction error is minimized as described in [9]. Then, the evolution of composition and temperature is computed by adopting the reduced mechanism associated to each cluster, solving a lower number of ODEs for the chemistry step in each cluster. The key point in the procedure is the fast classification of cells with a negligible overhead thanks to the efficiency of the on-the-fly PCA algorithm.

3. Results

The simulations were carried out with the `laminarSMOKE` code, a CFD solver based on `OpenFOAM` and specifically conceived for laminar reacting flows with detailed kinetic mechanisms [22]. The `laminarSMOKE` solver is based on the operator splitting technique, which means that advancement in time is carried out through two successive sub-steps: the transport sub-step (convection and diffusion) and the chemical sub-step.

The configuration chosen to test the adaptive chemistry methodology here presented is an axisymmetric, time-varying, non-premixed laminar coflow flame numerically and experimentally studied by Mohammed et al. [23]. The fuel is nitrogen-diluted consisting of 65% CH_4 and 35% N_2 (molar basis), while the oxidizer stream is regular air. Both streams are fed at ambient temperature and atmospheric pressure. The `POLIMI_C1C3_HT_1412` kinetic mechanism (84 species and 1698 reactions) [24] was used for the simulations. This mechanism was chosen because despite its relative small size it also accounts for complex heavy species whose chemistry is relatively slow (soot precursors such as benzene and naphtalene). Extensive validation and detailed description of such mechanism can be found in [25].

A CH_4/N_2 fuel stream enters through a circular nozzle (internal diameter of 4 mm and thickness of 0.38 mm), while the coflow air stream enters through an annular region (internal diameter of 50 mm). The transient behavior is induced by a sinusoidal perturbation in the velocity profile of the fuel stream with frequency f and amplitude A: $v_r(r) = v_{max}(1 - \frac{r^2}{R^2})(1 + Asin(2\pi ft))$, where r is the radial coordinate, R the internal radius, t the time and v_{max} equal to 70 cm/s. The coflow air is injected at 35 cm/s. The 2D computational domain (with lengths of 54 mm and 120 mm in the radial and axial directions) was discretized through a Cartesian mesh whose size was set around $\sim 25,000$ cells after a mesh sensitivity study.

The pre-partitioning of the composition space and the generation of the reduced kinetic mechanisms was performed on a dataset consisting of several timesteps of an unsteady 2D simulation with a sinusoidal perturbation ($f = 10$ Hz and $A = 0.25$, which will be referred to as f10A25), previously carried out with the detailed mechanism. In Table 1 all the frequencies and amplitudes which were used for the adaptive chemistry simulations are summarized, as well as the ones used for the detailed simulation to generate the training

dataset, which consisted of $\sim 131,000$ observations.

The availability of a large number of points does not ensure an adequate description of the problem: the chosen data shall indeed cover a range of conditions representative of the statistical variability of the system. The choice of the training data-set for machine-learning tasks cannot be validated *a priori*, but the accuracy of the trained model should be evaluated on new, unobserved data. For the present case, since the training dataset was built merging several timesteps of an unsteady 2D simulation with sinusoidal perturbation, the only available parameter to ensure a good statistical description was the time interval used to sample the time-steps.

Following the work of Mohammed et al. [23], the steady-state conditions ($A = 0$) were simulated first, then the approach was tested on four transient flames: two with a sinusoidal perturbation obtained using different frequencies and amplitudes, and the remaining two with non-sinusoidal perturbations. The first, sinusoidal, one was characterized by a perturbation with frequency $f = 10$ Hz and amplitude $A = 0.25$, while for the second one a more complex perturbation was used, obtained as a linear combination of several frequencies and amplitudes ($f = 10, 40, 80$ Hz and $A = 0.90, 0.50, 0.75$); the two non-sinusoidal perturbations consisted in a squared wave and a sawtooth wave, respectively. The steady-state adaptive simulation was also compared with two steady-state simulations using globally reduced mechanisms, to assess the true impact of clustering and of on-the-fly classification steps.

The latter were both generated from the same training dataset used for the adaptive case, the f10A25.

Before applying the LPCA partitioning algorithm the variables were centered with their mean values and scaled with their standard deviations (Auto scaling criterion).

Table 1: Amplitudes and frequencies of the sinusoidal perturbation adopted for the adaptive simulations and the detailed simulations used as training dataset.

Adaptive simulation	**Detailed simulation for the training**
Steady-state ($A = 0$)	f10A25
f10A25	f10A25
$f = 10, 40, 80$ Hz; $A = 0.90, 0.50, 0.75$	f10A25
Squared wave	f10A25
Sawtooth wave	f10A25

3.1. LPCA training

The parameters to set to perform LPCA are the number of PCs (n_{PCs}) to retain and the number of clusters (k). As mentioned in the previous paragraph, the criterion to choose the number of PCs consists in calculating the cumulative variance $t_{q,k}$ explained in the kth cluster, with $t_{q,k} = \sum_{j=1}^{q} l_{\mathrm{j,k}} / \sum_{j=1}^{p} l_{\mathrm{j,k}}$, ensuring that in all the clusters the condition $t_{q,k} > 0.99$ is verified in order to not lose any amount of information after the dimensionality reduction. If an inadequate number of PCs is considered for the dimensionality reduction during the clustering algorithm, the information loss can be deduced observing the behaviour of the mean coefficient of determination and the mean normalized root mean squared error (NRMSE), averaged for all the variables, between the vectors representing the original and the reconstructed data (from the reduced-dimensionality space) for an increasing number of retained PCs in each cluster, reported in Figure 1.

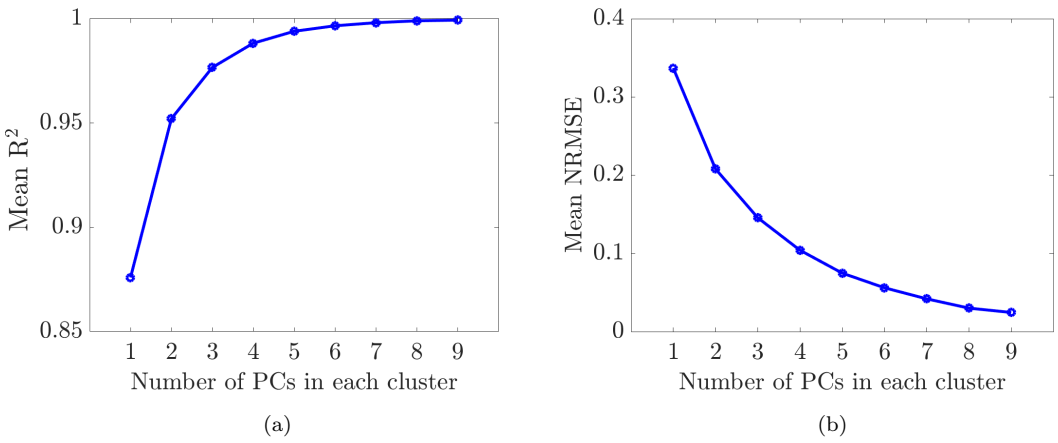


Figure 1: (a). Mean coefficient of determination R^2 for the reconstruction of the original profiles of the variables using an increasing number of PCs in each cluster and (b). Mean normalized root mean squared error for the reconstruction of the original profiles of the variables using an increasing number of PCs in each cluster.

The mean values for both the coefficient of determination and the NRMSE in Figure 1, for the reconstruction of the original variables, resulted to be not acceptable when the number of PCs was lower than 8. Considering $n_{PCs} = 8$, instead, all the original variables could be recovered with acceptable accuracy ($R^2 > 0.99$ and NRMSE < 0.1) after the dimensionality reduction, so all the variables of the system were correctly considered when the state-space was partitioned by means of LPCA. Moreover, increasing the number of eigenvectors taken into account did not improve considerably the quality of the

reconstruction, as an asymptotic value was already achieved for $n_{PCs} = 8$.

The normalized root mean squared error between two vectors $\hat{y}$ and y, both consisting of N elements, was defined as: NMRSE $= \frac{1}{\bar{y}} \cdot \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}-y)^2}{N}}$.

To choose the number of clusters, several LPCA classifications were carried out on the selected dataset increasing k. Whilst an increase in the number of clusters led to the reduction of the global reconstruction error $\epsilon_{GRSE,K}$, it also resulted in less populated clusters. To ensure statistically meaningful results, k was chosen as the largest one for which the condition: $n_{obs} > n_{var}$ was still verified in each cluster. If this condition is not satisfied, indeed, the efficiency of the classification algorithm is undermined and it can determine an increase of the errors in the adaptive simulation. This aspect is clearly observable looking at the non-monotonic relation between the error and the number of clusters used to partition the composition space, reported in the supplemental material. At first, the error decreases as the number of clusters is increased, because a larger k leads to smaller groups of points which are also more uniform from a chemical point of view, as also observable from the behaviour of the mean non-uniformity coefficient λ_{mean}. As soon as the condition $n_{obs} > n_{var}$ is violated, the error starts to grow again despite λ_{mean} keeps decreasing, due to the wrong assignment of the points to the clusters. With this into account, the number of PCs was set to 8 and the samples of the training dataset were divided in 12 groups. The large variance associated with the distribution of the observations in the clusters, reported in Table 2, could be explained by considering the presence of one cluster (cluster number 2) grouping all the reactive points located on the tail of the flame for each time-step, i.e. $\sim 50,000$ in total. Other clusters obtained after the partitioning, instead, were relatively small as they grouped very few points of the reactive layer.

Table 2: Cluster population statistics for the partitioning of the composition space spanned by the unsteady 2D detailed simulation obtained by means of a sinusoidal perturbation with $f = 10$ Hz and $A = 0.25$ with LPCA ($k = 12$).

Total number of clusters	12
Total number of observations of the dataset	$\sim 131,000$
Maximum number of observations in a cluster	$\sim 50,000$
Minimum number of observations in a cluster	128
Mean number of observations in a cluster	$\sim 11,000$
Variance of the observations distribution	$2.73 \cdot 10^8$

3.2. Mechanism reduction via DRGEP

After the partitioning algorithm, DRGEP was applied in each cluster to generate the reduced mechanisms.

For each sample point a reduced mechanism was generated. A single reduced mechanism for each cluster was created as the union of species and reactions corresponding to the individual reduced mechanisms of each sample point for that cluster. Fuel and oxidizer were assumed as target species for DRGEP, and several tolerance thresholds were tested (ϵ_{DRGEP} from 0.005 to 0.3) to examine the impact of the reduction on the reliability of the simulation and the CPU time. Table 3 shows the mean and the maximum uniformity coefficients (λ_{mean} and λ_{max}, respectively), resulting from the DRGEP procedure, as a function of the DRGEP thresholds. The table also reports the maximum and mean numbers of species of reduced mechanisms (respectively n_{sp}^{max} and n_{sp}^{mean}). The mean and the maximum number of species obviously increase with decreasing in the DRGEP tolerance, but it is possible to notice that the λ_{mean} coefficient also decreases, i.e. the mechanisms in each cluster start to be more and more uniform.

Table 3: Partitioning of composition space by LPCA with $k = 12$: number of species and uniformity coefficients for prescribed tolerances ϵ in DRGEP

ϵ_{DRGEP}	n_{sp}^{mean}	n_{sp}^{max}	λ_{mean}	λ_{max}
0.03	31	38	0.08	0.11
0.02	34	42	0.079	0.11
0.01	39	44	0.075	0.11
0.005	43	50	0.071	0.11

3.3. CFD results: steady-state flame

As explained in the previous paragraph, the steady-state conditions ($A = 0$) were simulated first, adopting both the complete mechanism and the proposed adaptive-chemistry methodology. In order to assess the quality of all the simulations carried out with the reduced mechanisms, two approaches were proposed: firstly a qualitative comparison of the contours between detailed and adaptive was carried out; then, a quantitative analysis based on the evaluation of the parity plots, as well as the normalized root mean squared error for selected variables, was pursued.

From the analysis of the maps in Figures 2 and 3, it is possible to recognize an excellent agreement between the steady-state simulations carried out using the detailed

mechanism and the adaptive-chemistry. Negligible discrepancies can only be observed in terms of lift-off height, and therefore localized to a very small region of the domain, just for $\epsilon_{DRGEP} = 0.03$ and 0.02.

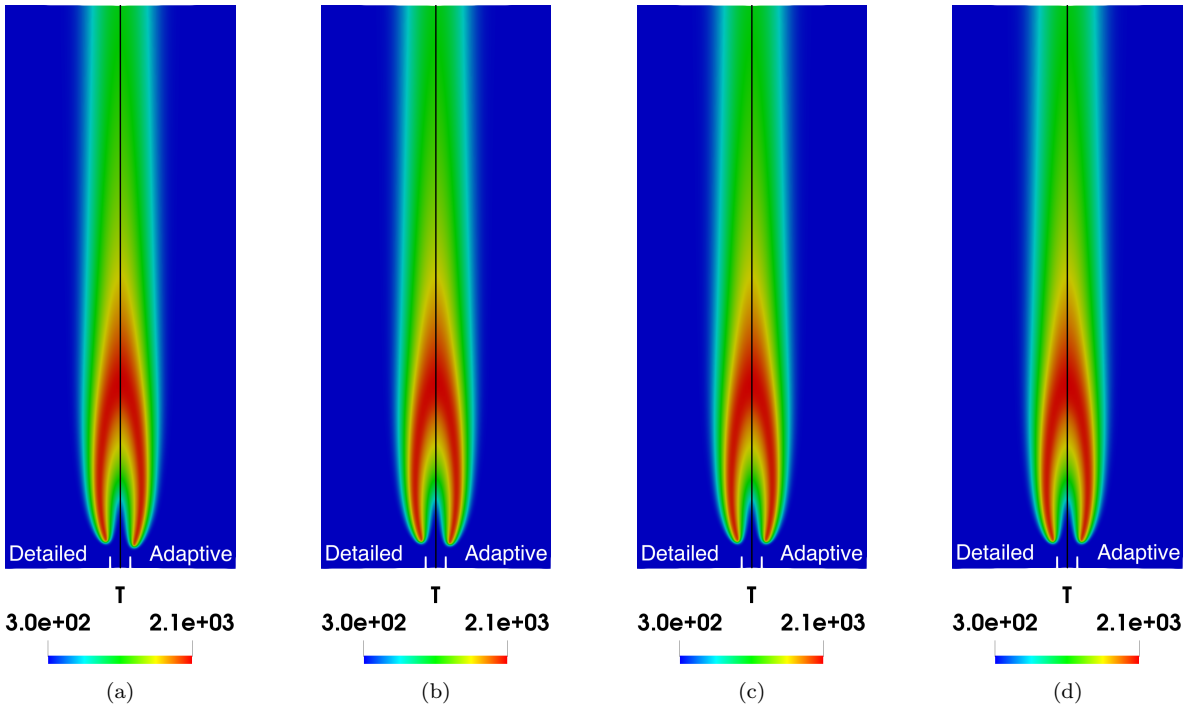


Figure 2: Comparison of temperature maps (in Kelvin) for detailed (left part of each contour) and adaptive (right part of each contour) steady-state ($A = 0$) simulations with prescribed DRGEP tolerances. (a) $\epsilon_{DRGEP} = 0.03$; (b) $\epsilon_{DRGEP} = 0.02$; (c) $\epsilon_{DRGEP} = 0.01$; (d) $\epsilon_{DRGEP} = 0.005$.

For a more extensive error analysis, Figure 4 reports the parity plots, while in Table 4 the relative NRMSE for the selected variables (temperature and main species) are reported for the simulation with $\epsilon_{DRGEP} = 0.005$.

Table 4: Normalized root mean squared error for temperature and selected species for the steady-state adaptive simulation carried out with $\epsilon_{DRGEP} = 0.005$.

Variable	**NRMSE**
T	$1.15 \cdot 10^{-3}$
O_2	$9.50 \cdot 10^{-4}$
CO	$1.77 \cdot 10^{-2}$
CO_2	$2.22 \cdot 10^{-3}$
CH_4	$1.77 \cdot 10^{-3}$
O	$9.04 \cdot 10^{-3}$
H_2O	$2.03 \cdot 10^{-3}$
OH	$7.78 \cdot 10^{-3}$

Parity plots in Figure 4 and the accuracy statistics in Table 4 confirm the excellent

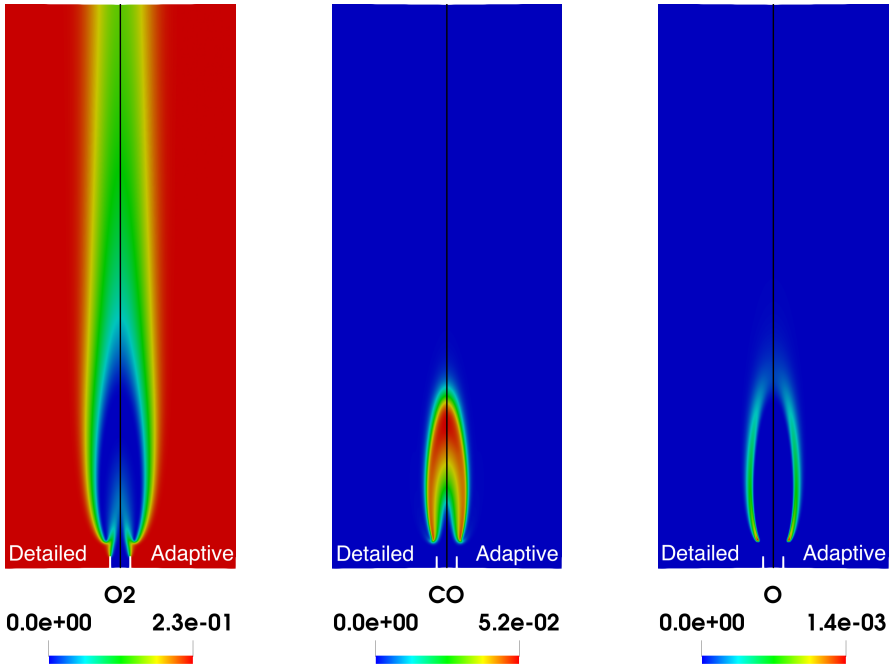


Figure 3: Comparison of selected species maps (from left to right: O_2, CO, O) for detailed (left part of each contour) and adaptive (right part of each contour) steady-state ($A = 0$) simulations, with $\epsilon_{DRGEP} = 0.005$.

agreement between the detailed and the steady-state adaptive simulation observed in the qualitative analysis. As expected, the lowest errors were registered for temperature and the target species (fuel and oxidizer). Nevertheless, also for the main radical species the errors can be considered small, since they are of the same order of magnitude of the target species and the fully oxidized products (CO_2 and H_2O). The same analysis with parity plots for all the selected values of ϵ_{DRGEP} is presented in the supplemental material.

The relation between the overall accuracy for the temperature and the selected chemical species of the reduced model and the degree of reduction (i.e. the tolerance used for DRGEP) is shown in the boxplots in Figure 5. From the latter it is possible to notice that the error behaviour is globally not linear with the reduction threshold and, in this particular case, the results for large ϵ_{DRGEP} ($\epsilon_{DRGEP} = 0.03$ and $\epsilon_{DRGEP} = 0.02$) seem not to be acceptable in terms of accuracy, as the error for some species can reach up to 40% and 30%, respectively. The normalized root mean square error decreases if the DRGEP tolerance is decreased, since more species are included locally in the mechanism. Analyzing the behaviour of the error with the reduction for the individual species (Figure 5b), it is possible to notice that non-radical species (CH_4, O_2, CO) and temperature exhibit

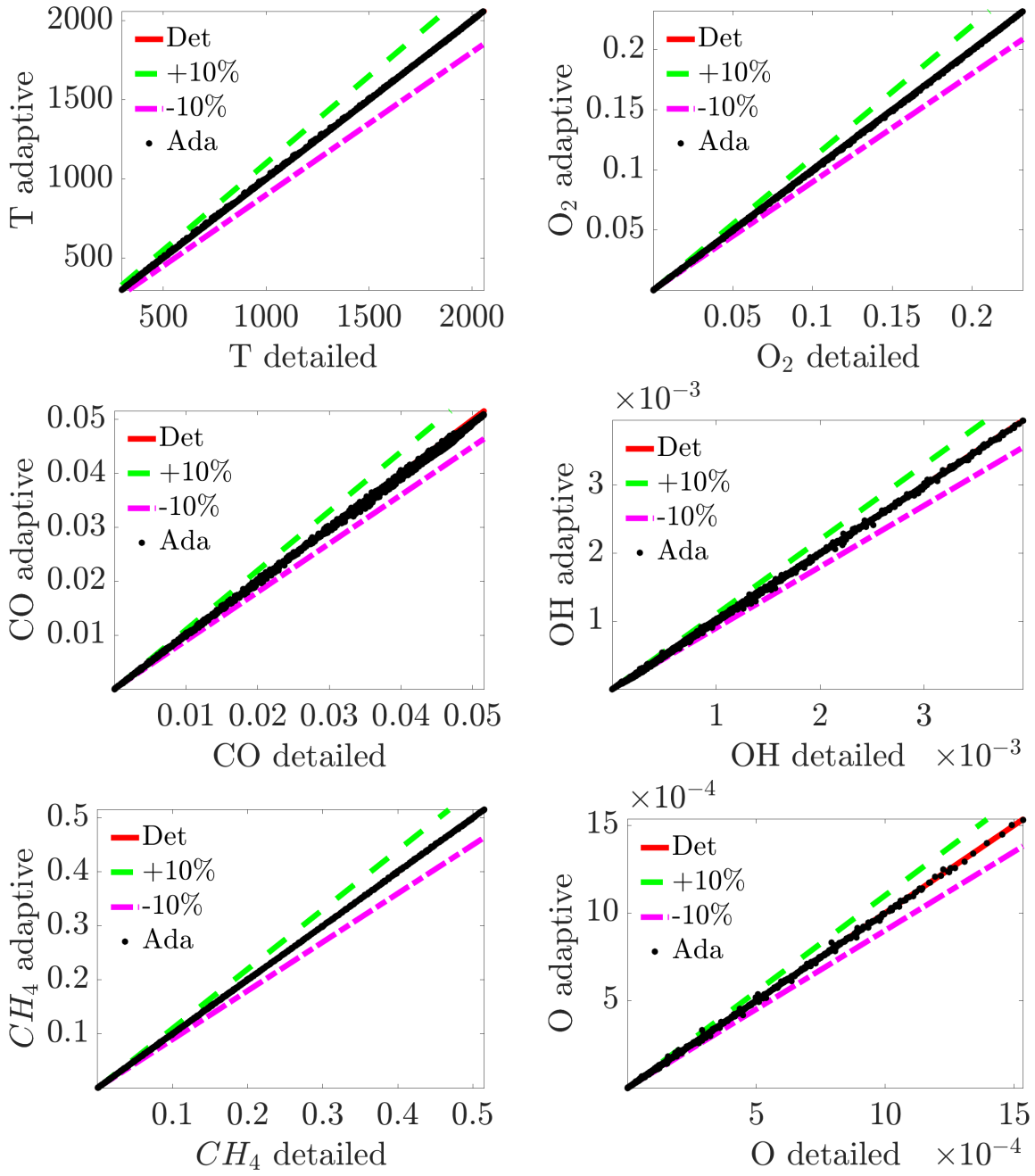
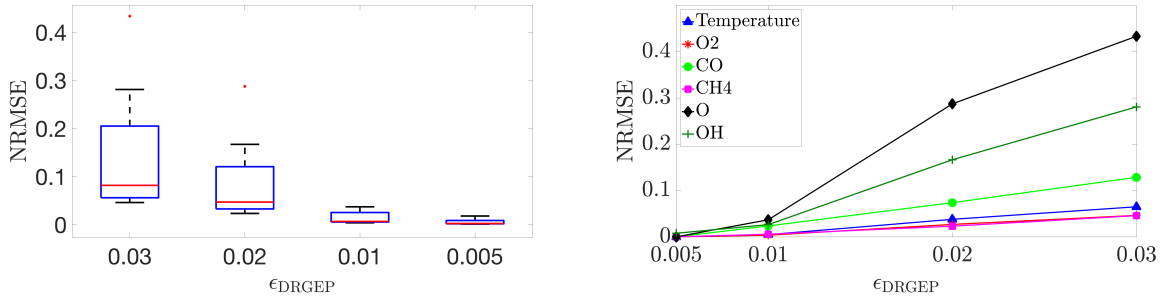


Figure 4: Parity plots of temperature and selected species for $\epsilon_{DRGEP} = 0.005$, steady-state conditions ($A = 0$).

a distinct linear trend, while the error for radical species like O and OH has a non-linear relation with the reduction, for $0.01 < \epsilon_{DRGEP} < 0.03$. The non-linear behaviour of the error related to the radicals in this range could be explained by the lift-off height shift for these values of ϵ_{DRGEP} observed in Figure 2. Thus, since these species are formed and react in a less extended region of space, if compared to all the reported stable species and

temperature, the mean errors between the adaptive and the detailed simulation are more pronounced.



(a) (b)

Figure 5: (a). Boxplots of the normalized root mean squared error for selected species calculated for several reduction tolerances; (b). Relation between the error and the reduction tolerance for temperature and several species.

The details about the computational performances reported in Table 5 show that the mean CPU-time for carrying out the chemical step ($\bar{\tau}_{chem}$) is strongly dependent on the DRGEP tolerance, obviously increasing when decreasing the tolerance ϵ_{DRGEP}. The corresponding chemical speed-up S_{chem} is large (~ 4) even adopting the largest number of species ($\epsilon_{DRGEP} = 0.005$). The chemical step usually has a massive impact on the CPU-time of a numerical simulation of a reacting flow with a detailed chemistry. In this specific case, it accounted for about the 85% of the total CPU-time, while the remaining 15% was taken by the transport step.

Table 5: Performances of adaptive-chemistry algorithm: average CPU-time per cell (in ms) for chemical step integration ($\bar{\tau}_{chem}$) and relative mean speed-up factor (S_{chem}) for the steady flame with $k = 12$.

ϵ_{DRGEP}	$\bar{\tau}_{chem}$	S_{chem}
0.03	2.78	5.39
0.02	3.03	4.94
0.01	3.25	4.61
0.005	3.78	3.97
detailed	15.02	-

As mentioned in Section 3, two additional steady-state simulations were implemented using two globally reduced mechanisms, to assess the impact of clustering and on-the-fly classification steps on the results. The first, globally reduced, mechanism (Global 1) consisted of 59 species, and it was generated with a $\epsilon_{DRGEP} = 0.005$. The second one (Global 2), instead, was generated using a DRGEP tolerance $\epsilon_{DRGEP} = 0.05$, to

approximately retrieve the same average number of species ($n_{sp}^{Global2} = 44$) of the adaptive approach in the case of reduction using $\epsilon_{DRGEP} = 0.005$. The boxplots in Figure 6 show the NRMSE for the aforementioned three cases, clearly showing the superior performances associated to a local and adaptive chemical mechanism reduction.

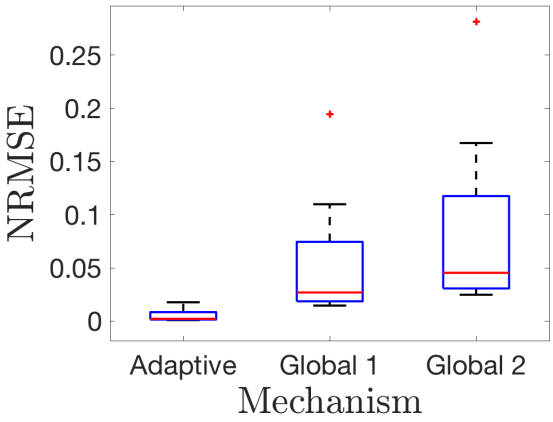


Figure 6: Boxplots of the normalized root mean squared error for selected species calculated for: steady-state adaptive simulation with SPARC approach (Adaptive), globally reduced mechanism using $\epsilon_{DRGEP} = 0.005$ (Global 1) and globally reduced mechanism with $\epsilon_{DRGEP} = 0.05$ (Global 2).

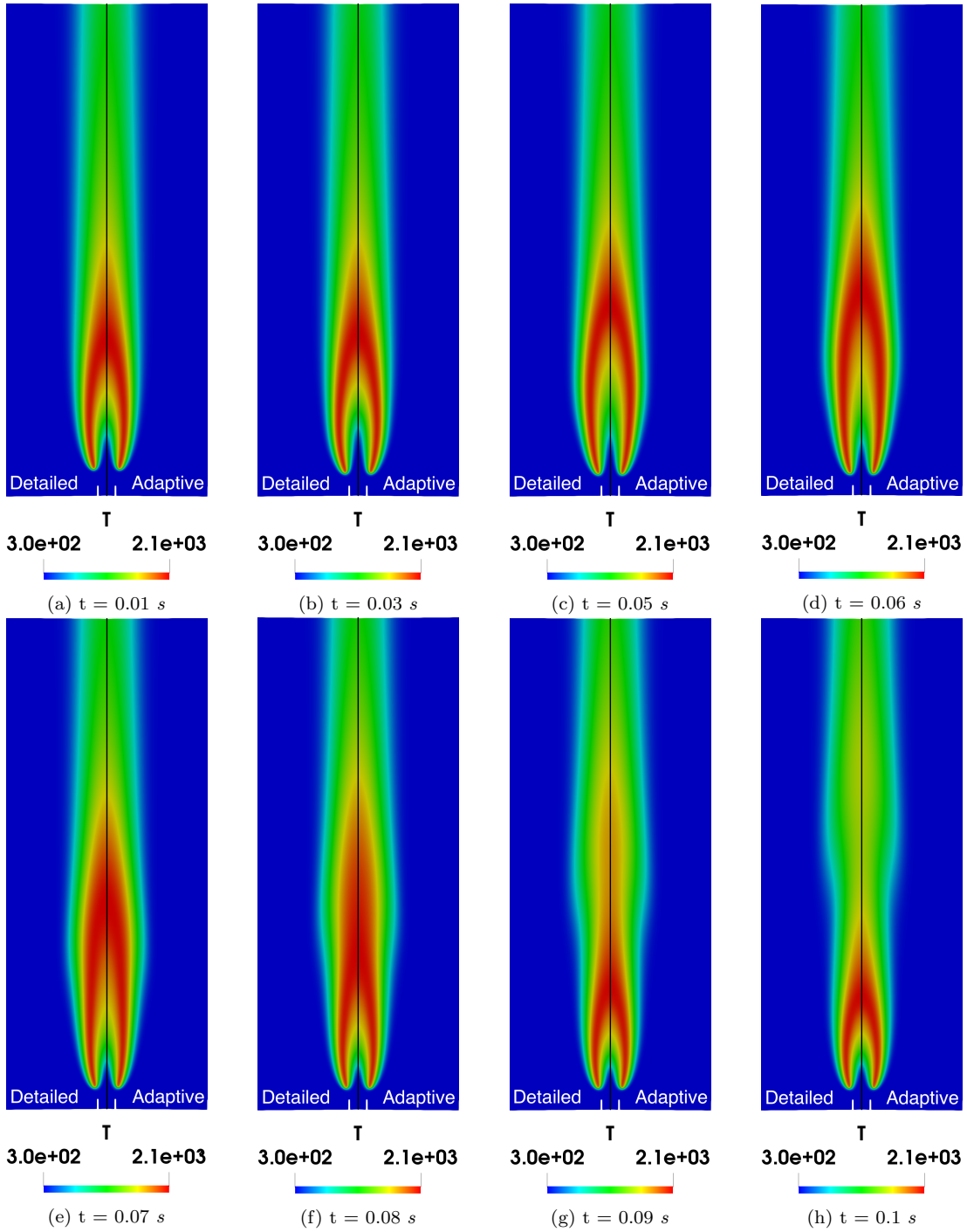
The higher accuracy of the adaptive simulation can be easily explained considering that the chemistry reduction is performed on disjoint regions of the thermochemical space, leading to locally optimized mechanisms and accounting for *local* phenomena which are usually overlooked by global reduction approaches. Finally, the adaptive simulations also ensure the highest speed-up factor $S_{chem} = 3.97$, compared to the one obtained with Global 1, $S_{chem} = 2.11$, and Global 2, $S_{chem} = 3.46$.

3.4. CFD results: unsteady flames

In order to analyze in greater depth the accuracy and the reliability of the proposed procedure, the adaptive-chemistry approach was then applied to two unsteady methane flames with a sinusoidal perturbation of the fuel velocity profile. As also described in the previous paragraph, the first unsteady flame was obtained imposing a continuous simple sinusoidal perturbation with fixed frequency $f = 10$ Hz and amplitude $A = 0.25$ to the parabolic fuel velocity profile (f10A25), while for the second flame a sinusoidal perturbation was still imposed, but it was obtained as linear combination of different frequencies and amplitudes ($f = 10, 40, 80$ Hz and $A = 0.90, 0.50, 0.75$).

Also for the f10A25 case a qualitative and a quantitative assessment of the model was carried out. The qualitative assessment for this simulation is proposed in Figures 7 and 8, where the maps of temperature and species of the detailed simulation (left side of each contour) are compared with the maps obtained by means of an adaptive-chemistry approach using $\epsilon_{DRGEP} = 0.005$ (right side of each contour). No lift-off height shift between the detailed and the adaptive simulation was observed, and the shapes of the examined profiles have a satisfactory agreement. In Figure 9 and in Table 6 a quantitative analysis of the error is proposed for the time-step $t = 0.09$ s for the f10A25 simulation carried out with the reduced mechanisms using $\epsilon_{DRGEP} = 0.005$. Also in this case, as for the steady-state simulation, the lowest NRMSE were registered for the target species, but the errors for the radicals had the same order of magnitude so they could be again considered satisfactory.

Since this is not a steady-state case, it is not sufficient to examine the error statistics for just one timestep, as previously done: the accuracy of the results, as well as the number of species (n_{sp}), reactions (n_r) and, consequently, the speed-up of the simulation, can vary at each step. For this reason, the NRMSE was averaged, taking into account all the previously selected variables, and monitored at each timestep to observe its behaviour. By looking at Figure 10a, where the mean NRMSE is reported for several timestep of the f10A25 simulation, it is possible to conclude that, even if a minimal oscillatory behaviour is observed, the error does not amplify with time: after reaching a maximum at $t = 0.07$ s it starts to decay again, staying always bounded in the order of magnitude of $\sim 10^{-3}$. In the supplemental material it is possible to find the parity plots for the selected variables for several timesteps of the considered interval of time. Comparing the behaviour of the mean error with time and the mean number of species and reactions reported in Figure 10b, it is possible to observe that the error does not depend exclusively on the degree of reduction adopted for the prescribed timestep. Indeed, even if the timestep $t = 0.12$ s was characterized by the lowest number of species and reactions, the error was lower than the one registered for the timestep $t = 0.07$ s, where a larger number of species was considered. Comparing Figures 10b and 10c, instead, the direct relation between the mean number of species and reactions and the speed-up over time can be explicitely observed, since the mean-CPU time (and, consequently, the speed-up) only depends on



(a) t = 0.01 s (b) t = 0.03 s (c) t = 0.05 s (d) t = 0.06 s

(e) t = 0.07 s (f) t = 0.08 s (g) t = 0.09 s (h) t = 0.1 s

Figure 7: Comparison of temperature maps (in Kelvin) for detailed (left part of each contour) and adaptive (right part of each contour) unsteady simulations obtained with a sinusoidal perturbation in the velocity profile ($f = 10$ Hz and $A = 0.25$), with $\epsilon_{DRGEP} = 0.005$, for several timesteps.

the number of species to consider at each time-step.

The same quantitative assessment, analyzing the behaviour of the mean NRMSE over time, was applied to the second unsteady flame, obtained with a complex sinusoidal perturbation which linearly combines several frequencies and amplitudes. The good results

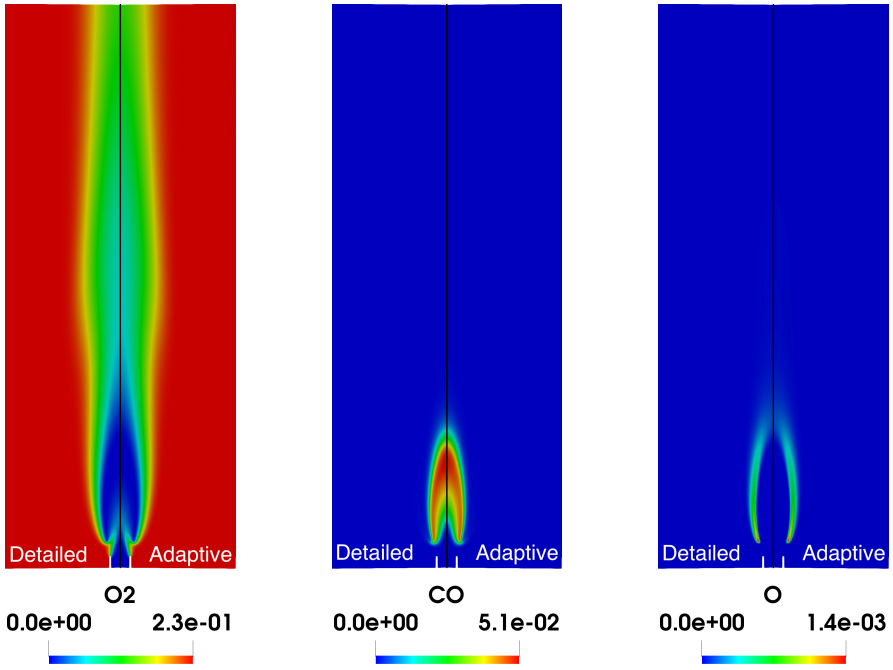


Figure 8: Comparison of selected species maps (from left to right: O_2, CO, O) for detailed (left part of each contour) and adaptive (right part of each contour) unsteady simulations obtained with a sinusoidal perturbation in the velocity profile ($f = 10$ Hz and $A = 0.25$), with $\epsilon_{DRGEP} = 0.005$, at time t = 0.09 s.

obtained for the steady-state and the f10A25 flame represented a verification of the approach, since the partitioning and the mechanism reduction steps were carried out on the data obtained from the detailed simulation of the same flame. Testing the approach on other flames (i.e. changing the flame dynamics), then, was necessary to fully prove the effectiveness of the method and extend its applicability. It is possible to see from Figure 11a that the errors show the same oscillatory behaviour, even if they are slightly larger. Also in this case, the same considerations about the nature of the error can be made: analyzing Figures 11a and 11b it is possible to see that the lowest accuracy for a timestep does not correspond to the lowest number of species adopted: therefore the contribute of a time-varying efficiency of the classification algorithm must be also considered. Anyway, despite the moderately larger errors, these results can once again be considered acceptable if it is taken into account that for this last simulation the training phase (the partitioning and the generation of reduced kinetic mechanisms) was made with data obtained from a different flame. The parity plots and the error statistics for several timesteps of the complex simulation are available in the supplemental material.

The adaptive-chemistry approach was finally tested on two unsteady flames obtained

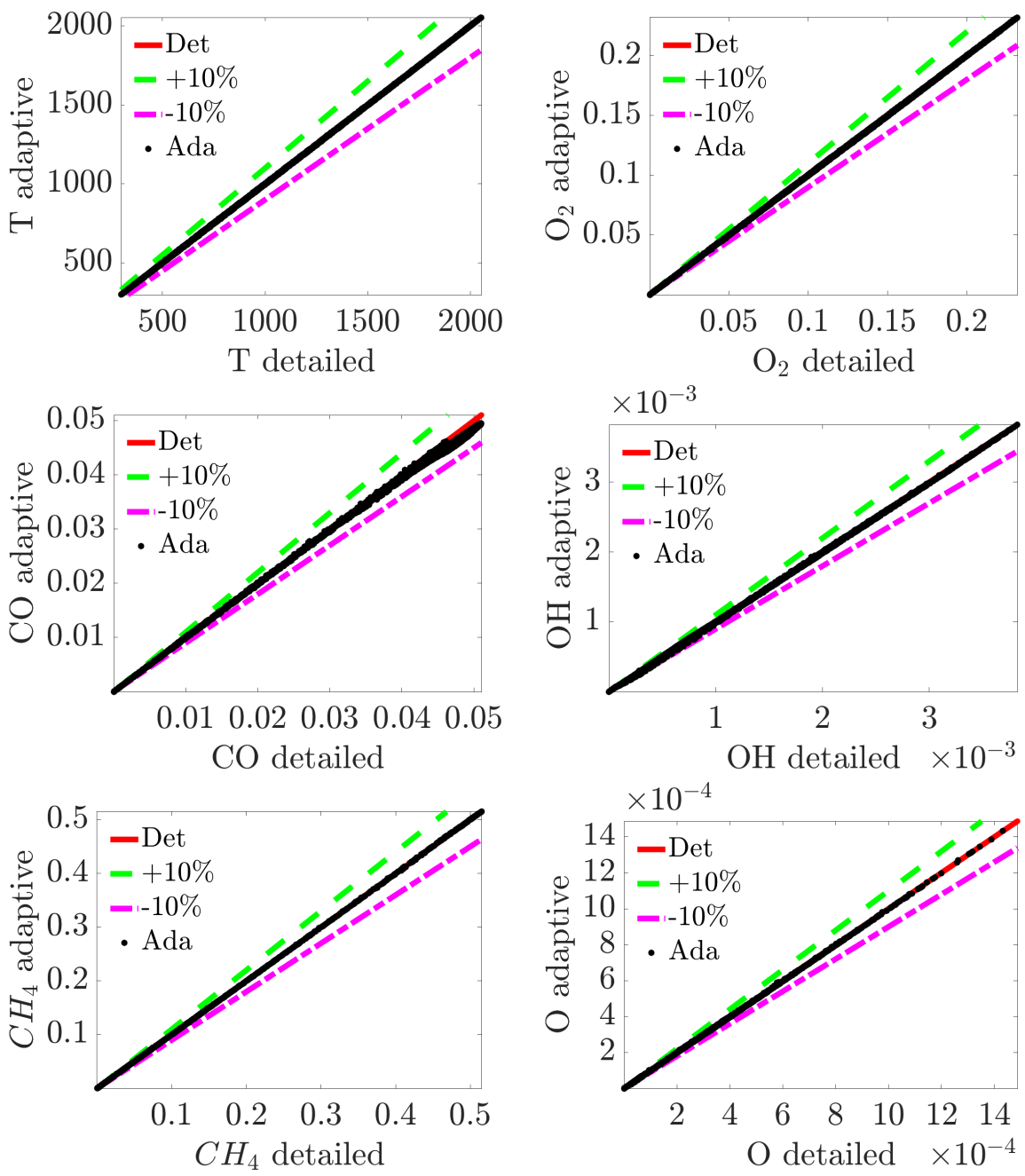


Figure 9: Parity plots for temperature and selected species for an unsteady adaptive simulation obtained with a sinusoidal perturbation in the velocity profile (f = 10 Hz and A = 0.25), with ϵ_{DRGEP} = 0.005, at time t = 0.09 s.

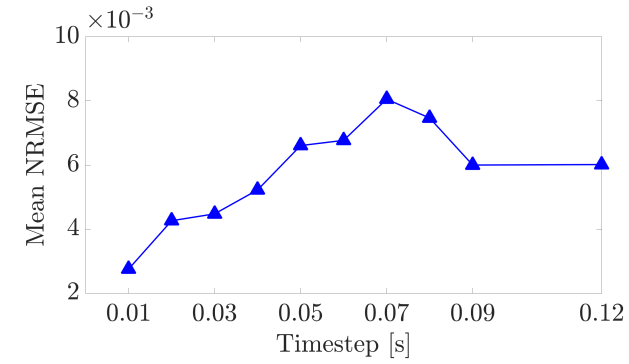
by means of non-sinusoidal perturbations to the fuel parabolic velocity profile: a squared wave and a sawtooth wave, whose dynamics is reported in Figure 12. The CFD simulations were carried out using the reduced mechanisms previously obtained from the f10A25 dataset. Analyzing the behaviour of the mean NRMSE over time for these two additional adaptive simulations, reported in Figure 13, one can observe that the accuracy is in line

Table 6: Normalized root mean squared error for temperature and selected species for an unsteady adaptive simulation obtained with a sinusoidal perturbation in the velocity profile ($f = 10$ Hz and $A = 0.25$), with $\epsilon_{DRGEP} = 0.005$, at time t = 0.09 s.

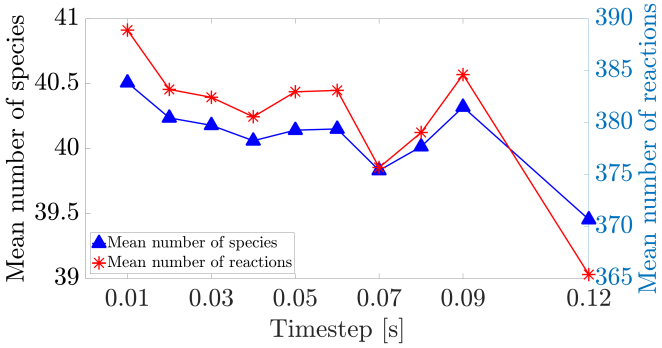
Variable	**NRMSE**
T	$1.24 \cdot 10^{-3}$
O_2	$1.07 \cdot 10^{-3}$
CO	$2.50 \cdot 10^{-2}$
CO_2	$2.94 \cdot 10^{-3}$
CH_4	$1.37 \cdot 10^{-3}$
O	$6.44 \cdot 10^{-3}$
H_2O	$2.97 \cdot 10^{-3}$
OH	$6.99 \cdot 10^{-3}$

with the one obtained using sinusoidal perturbations, and the error does not amplify over time.

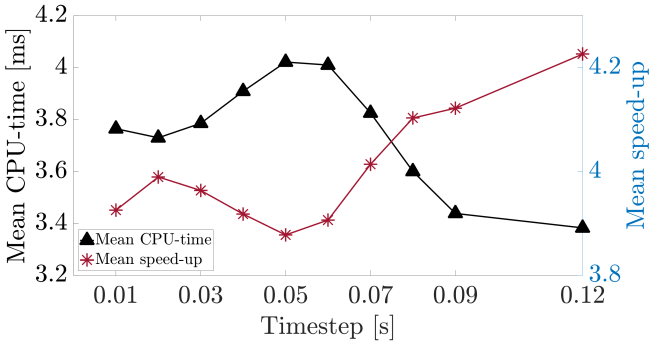
The results obtained with the adaptive-chemistry approach for both the steady and unsteady simulations can be considered satisfying, as it was possible to obtain accurate results using not only reduced mechanisms generated from previous 2D simulations of the same chemical system, but also from previous simulations of similar chemical systems. Moreover, PCA appears to be a very promising tool either for the partitioning phase by means of its local algorithm, as well as for the on-the-fly computation because, although its classification efficiency does not have a constant value in time, it does not alter the overall accuracy of the simulation significantly.



(a)

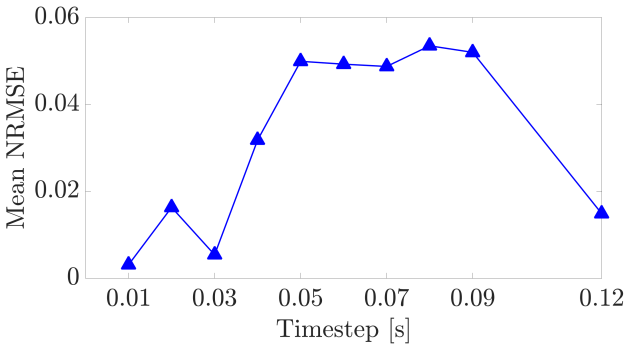


(b)

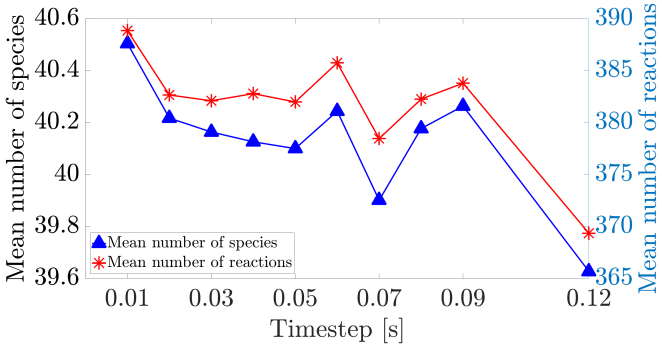


(c)

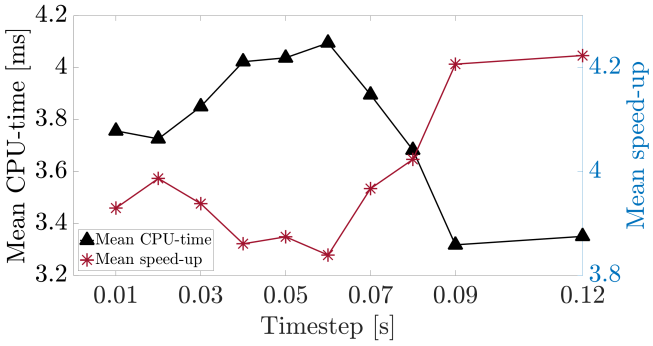
Figure 10: Statistics for several timesteps of the unsteady adaptive simulation obtained with a sinusoidal perturbation in the velocity profile ($f = 10$ Hz and $A = 0.25$), using the reduced mechanisms ($\epsilon_{DRGEP} = 0.005$) obtained from the detailed unsteady simulation with $f = 10$ Hz and $A = 0.25$. (a). Averaged normalized root mean squared error over time; (b). Mean number of species and reactions over time; (c). Mean CPU-time for chemistry step and relative speed-up factor compared to detailed simulation over time.



(a)



(b)



(c)

Figure 11: Statistics for several timesteps of the unsteady adaptive simulation obtained with a complex sinusoidal perturbation ($f = 10, 40, 80$ Hz and $A = 0.90, 0.50, 0.75$), using the reduced mechanisms ($\epsilon_{DRGEP} = 0.005$) obtained from the detailed unsteady simulation with $f = 10$ Hz and $A = 0.25$. (a). Averaged normalized root mean squared error over time; (b). Mean number of species and reactions over time; (c). Mean CPU-time for chemistry step and relative speed-up factor compared to detailed simulation over time.

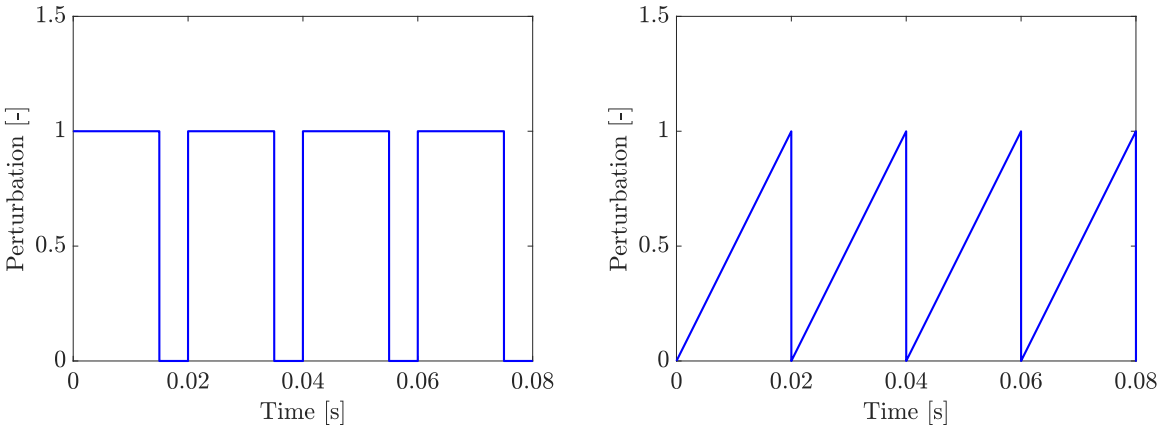


Figure 12: (a). Squared wave perturbation imposed to the fuel parabolic velocity profile in time; (b). Sawtooth wave perturbation imposed to the fuel parabolic velocity profile in time.

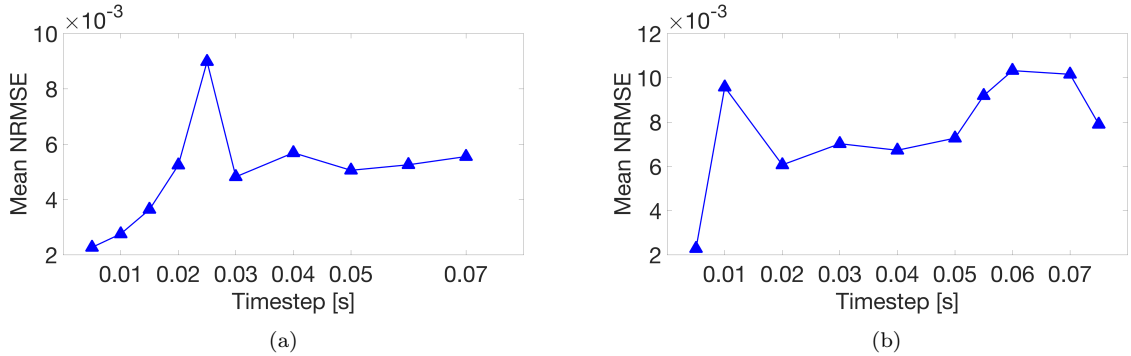


Figure 13: (a). Averaged normalized root mean squared error over time for several timesteps of the unsteady adaptive simulation obtained with a squared wave perturbation imposed to the fuel parabolic velocity profile; (b). Averaged normalized root mean squared error over time for several timesteps of the unsteady adaptive simulation obtained with a sawtooth wave perturbation imposed to the fuel parabolic velocity profile.

4. Effects of the training dataset on adaptive-chemistry simulations

According to the results showed in the previous paragraphs, it is possible to accurately simulate steady or unsteady reactive flows using reduced mechanisms obtained from similar chemical systems, as long as the latter adequately cover the composition space which is expected to be visited during the simulation of interest. This section explores the possibility of using also lower dimensional detailed simulations (0D and 1D) of the same chemical system for the generation of the training data, the classification and the generation of reduced mechanisms to use in a multidimensional simulation (2D or 3D) carried out by means of an adaptive-chemistry approach. In fact, given the low computational cost associated with 0D and 1D simulations, the possibility to perform the aforementioned training steps of the adaptive-chemistry approach on this data can constitute a viable option in case no multidimensional simulation previously carried out with detailed mechanism is available.

A dataset consisting of observations generated by means of steady-state counter-flow simulations (CFDF) was considered. A range of strain rates from 10 s^{-1} to 330 s^{-1} (corresponding to extinction conditions) was simulated, using the same inlet conditions in terms of temperature and compositions used for the multidimensional simulation. As suggested by Franke et al. [26], to avoid prescribing patterns in the dataset, the strain rates were randomly chosen in the range above. This new dataset consisted of $\sim 11,000$ points, corresponding to 30 different strain rates. The simulations were carried out using a in-house code based on the `OpenSMOKE++` framework [27].

The same procedure used for the previous 2D dataset was adopted to find the number of components to retain and the number of clusters for the composition space, which led to 5 PCs and 35 clusters. Also in this case, as for the 2D training, increasing the number of clusters led to a progressive decrease of the number of species and an increase of the accuracy, while as soon as the constraint $n_{obs} > n_{var}$ was violated the accuracy of the simulation started to decrease, confirming the existence of an optimum for k. For each cluster, a reduced mechanism was then generated again as the union of species and reactions corresponding to the individual reduced mechanisms from the sample points constituting them. The same target species were kept (fuel and oxidizer) for the DRGEP, as well as the same tolerance thresholds, to have a proper comparison beteween the

simulations. Table 7 reports the statistics from the DRGEP reduction regarding the number of species and the uniformity of the mechanisms in each cluster.

Table 7: Partitioning of composition space by means of LPCA for the CFDF dataset: number of species and uniformity coefficients for prescribed tolerances ϵ_{DRGEP}, with $k = 35$

ϵ_{DRGEP}	n_{sp}^{mean}	n_{sp}^{max}	λ_{mean}	λ_{max}
0.03	29	45	0.074	0.14
0.02	32	49	0.071	0.14
0.01	37	55	0.063	0.14
0.005	41	59	0.057	0.13

From a first analysis of the reduction results from DRGEP in Table 7, it emerges that the mechanisms generated from 1D simulations were more sensitive to the reduction threshold. Lowering the reduction tolerances from 0.03 to 0.005 led to a reduction of the disuniformity factor λ_{mean} of the 23%, while with the same tolerances λ_{mean} was reduced by 11% for the 2D dataset. At the same time, the maximum number of species and the maximum value of disuniformity coefficient λ were up to $\sim 20\%$ larger than the DRGEP reduction based on the 2D dataset. Given that, in this case it is fair to expect larger speed-up factors due to the lower mean number of species.

Steady-state conditions ($A = 0$) were firstly simulated using the new reduced mechanisms derived from the detailed 1D CFDF simulations. The adaptive simulations showed a moderately lower accuracy compared to the ones based on the 2D training, on equal terms of reduction tolerance ($\epsilon_{DRGEP} = 0.005$), but at the same time a larger speed-up factor due to the lower number of species, as also expected from the reduction statistics. The results obtained with relatively large reduction tolerances (ϵ_{DRGEP} from 0.03 to 0.01) were characterized by very large mean NRMSE (averaged over all the selected species) as shown in Figure 14a, and they seem to be not acceptable especially with respect to the one obtained using the same reduction tolerance on a 2D training dataset. The accuracy of the simulation carried out with $\epsilon_{DRGEP} = 0.005$, instead, seems somehow comparable for the two cases, as shown in Figures 14a and 14b. As regards the error dependence of specific variables with the reduction tolerance ϵ_{DRGEP} in Figure 14c, also in this case, as for the 2D training, target species (fuel and oxidizer), temperature and CO had a linear dependence with the degree of the reduction, while for the radicals a non-linear pattern was observed again because of the lift-off height shift for large values of the tolerance.

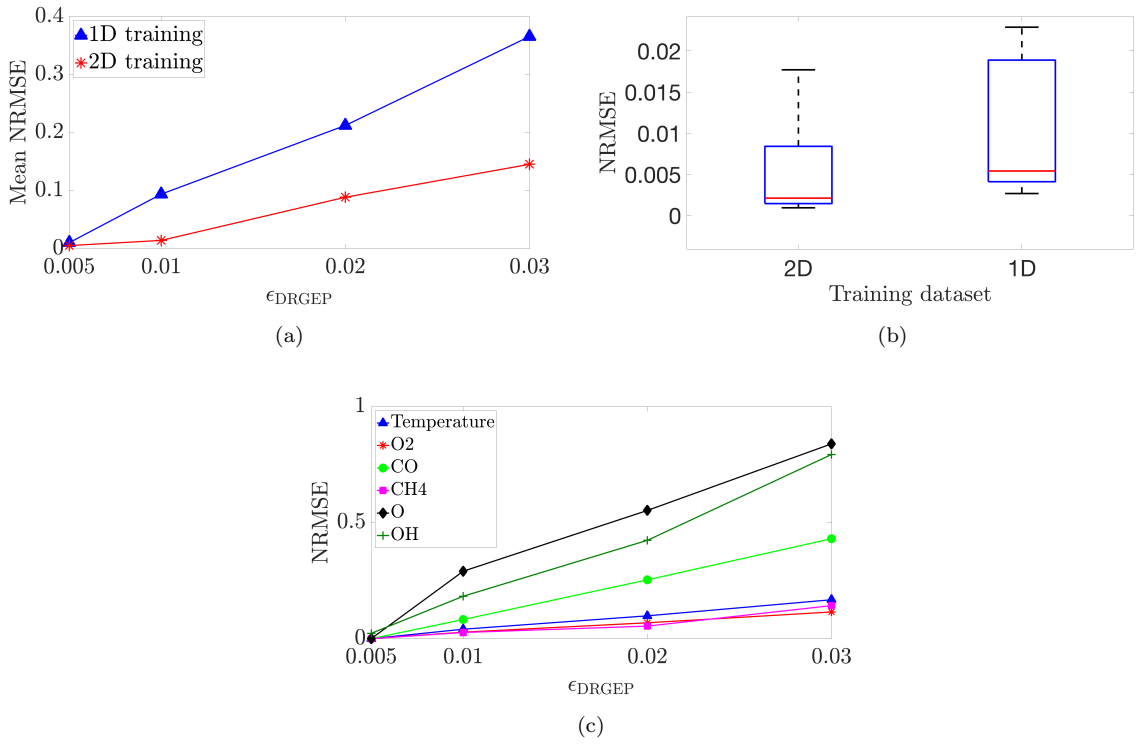


Figure 14: (a). Averaged root mean squared error for several DRGEP tolerances obtained from steady-state simulations ($A = 0$) trained on 1D and 2D datasets; (b). Boxplots of the NRMSE for steady-state simulations ($A = 0$) trained on 1D and 2D datasets with $\epsilon_{DRGEP} = 0.005$; (c). Relation between the error and the degree of the reduction of the mechanisms for temperature and selected species.

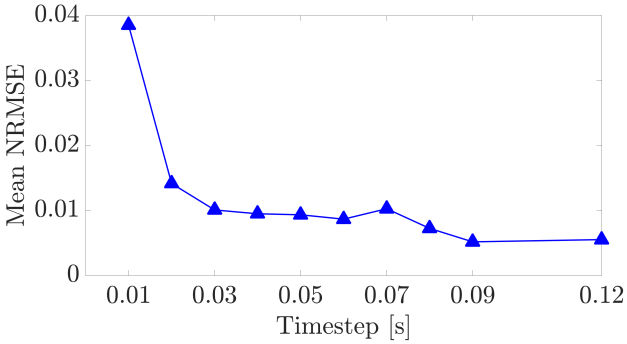
Table 8: Performances of adaptive-chemistry algorithm: average CPU-time per cell (in μs) for chemical step integration ($\bar{\tau}_{chem}$) and relative mean speed-up factor (S_{chem}) for the steady flame ($A = 0$) using the reduced mechanisms obtained with the training of a 1D CFDF dataset, using $k = 35$.

ϵ_{DRGEP}	$\bar{\tau}_{chem}$	S_{chem}
0.03	1.55	9.67
0.02	2.02	7.40
0.01	2.50	6.01
0.005	3.04	4.93
detailed	15.02	-

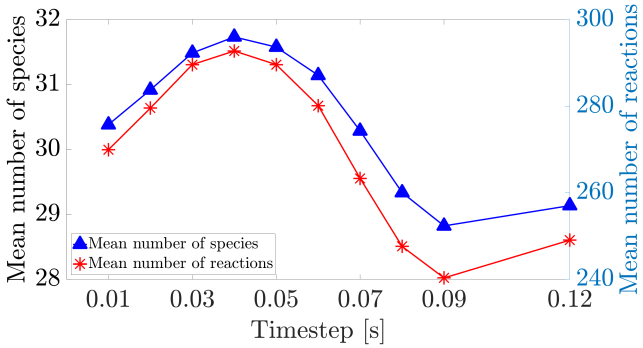
Lastly, the adaptive-chemistry approach was tested for an unsteady case: the flame obtained by means of a perturbation of the velocity profile with frequency $f = 10$ Hz and amplitude $A = 0.25$ was simulated using the reduced mechanisms obtained from the partitioning of the 1D CFDF dataset. From Figure 15a, where the analysis of the mean NRMSE over time is shown, it is possible to see that also in this case the accuracy using mechanisms obtained from a lower dimensional training dataset was large, even if slightly lower than the one obtained using the training of the 2D simulation, as also verified for the simulation in steady conditions. The error curve behaviour has a less oscillatory nature

than the curves examined in the previous paragraphs since, after a maximum at $t = 0.01$ s, it seems to reach a stable, asymptotic value of 1%. In Figure 15b, it is possible to see that in this last case the difference between the minimum and the maximum mean number of species for the considered timesteps is bigger than the one registered in case of 2D training. From the same figure, also the number of reactions for the mechanisms generated with 1D training data appears to be significantly lower than the one reported for the adaptive simulation using mechanisms generated with 2D training data. While the minimum mean number of reactions for the 1D case was equal to $\sim$ 240 and the maximum mean number to $\sim$ 290, for the first simulation using the 2D training data the minimum mean number of reactions was 365 and the maximum $\sim$ 390.

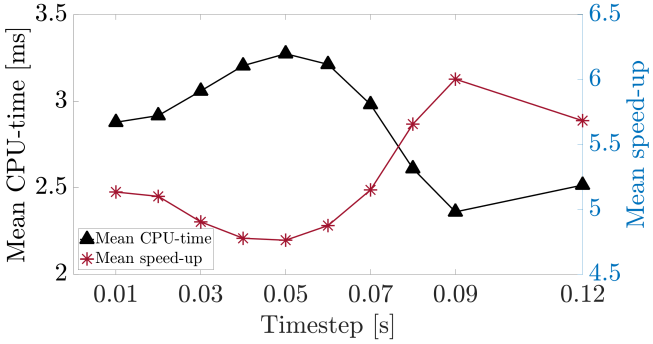
Thus, at this point it is clear that it is possible to perform multidimensional simulations (2D and 3D) of steady and unsteady reacting flows with a training dataset obtained from a lower dimensional simulation, at two conditions. The first, necessary condition, is that the dataset should cover the thermo-chemical conditions met in the application of interest. The second is the appropriate choice of the reduction tolerance. In this case, the 1D CFDF dataset covered a very wide region of the composition space for radicals and for heavier species like benzene, as shown in Figure 16, where the composition spaces visited by the 2D simulation and the 1D CFDF simulations used as training dataset are compared. Anyway, for some species like HO_2 or CH_2O there were still some zones in the composition space spanned by the 2D laminar coflow flame which were not properly covered by the 1D CFDF simulations, as shown in Figure 17. In this case, the reduction becomes the limiting factor, since the calculated DRGEP coefficients are calculated on a less comprehensive set of compositions. This could explain the results, using the reduced mechanisms obtained from 1D simulations, characterized by a slightly lower accuracy if compared to the ones obtained using a training dataset obtained from 2D simulations (parity plots included in the supplemental material).



(a)



(b)



(c)

Figure 15: (a). Averaged normalized root mean squared error; (b). Mean number of species and reactions; (c). Mean CPU-time for chemistry step and relative speed-up factor compared to detailed simulation. All the images refer to several timesteps of the unsteady adaptive simulation obtained with a sinusoidal perturbation in the velocity profile ($f = 10$ Hz, $A = 0.25$), with the training made on the 1D CFDF dataset, $\epsilon_{DRGEP} = 0.005$.

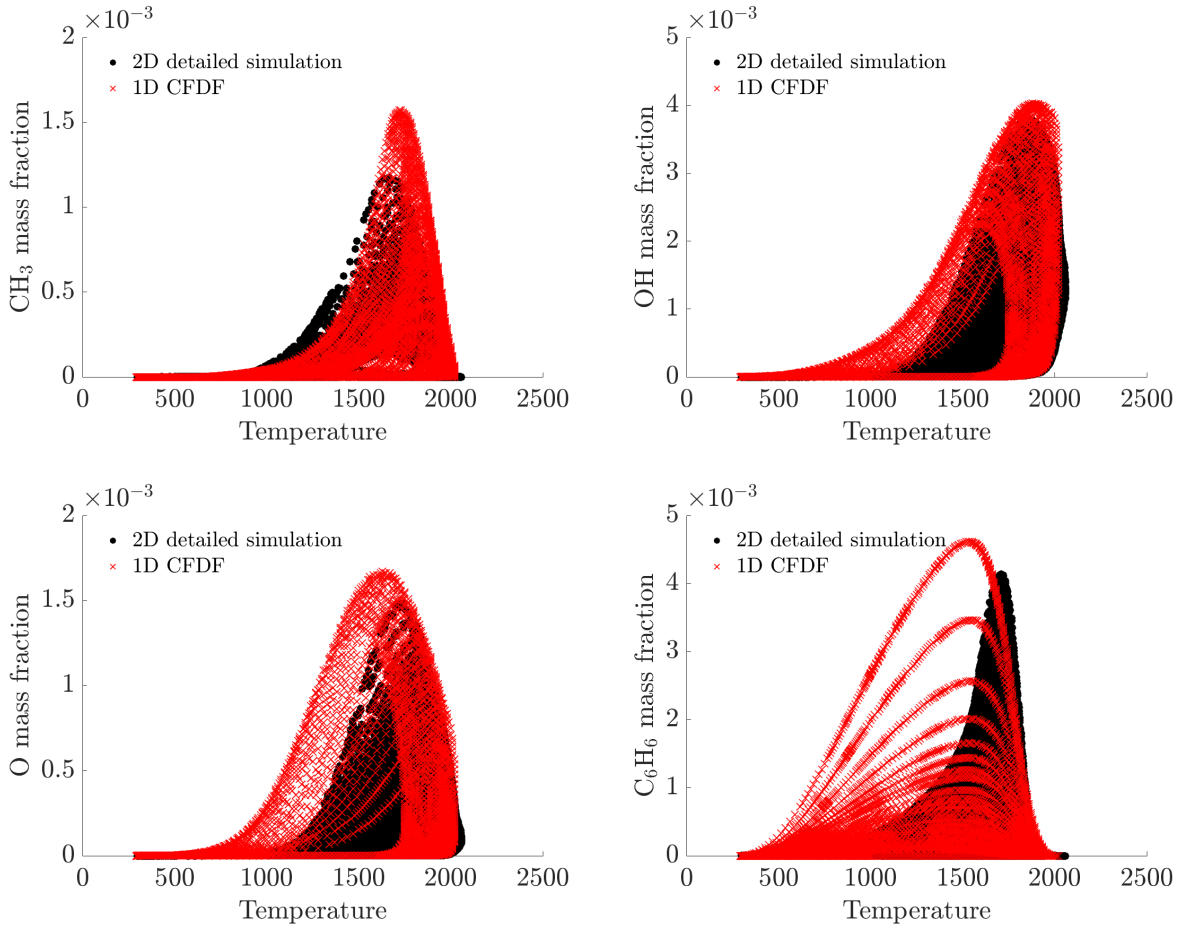


Figure 16: Generated dataset: mass fraction of selected species as a function of temperature for the 2D detailed simulation (black dots) and the 1D couterflow (CFDF) simulation (red cross).

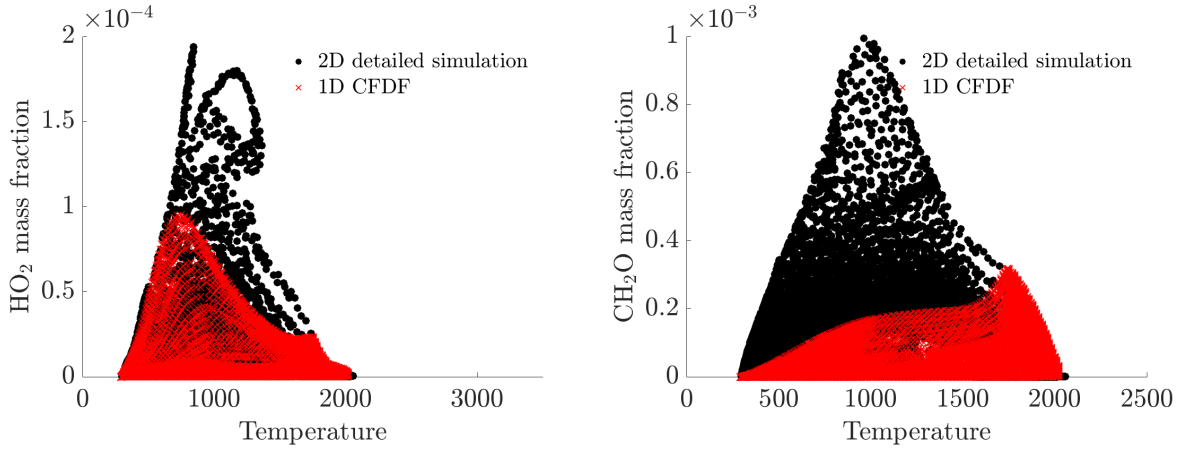


Figure 17: Generated dataset: mass fraction of selected species as a function of temperature for the 2D detailed simulation (black dots) and the 1D couterflow (CFDF) simulation (red cross).

5. Conclusions

The main objective of the SPARC methodology proposed in this work is paving the way for multi-dimensional CFD simulations of laminar flames with complex kinetic mechanisms (hundreds of species), which are currently virtually unfeasible because of the computational cost. The methodology is based on the pre-partitioning of a composition sample space (obtained from previous simulation of the same system or generated via counterflow diffusion flames) in clusters via LPCA. Then, a reduced kinetic mechanism, generated via DRGEP, is associated to each cluster. During the CFD simulation, each cell is properly classified (in order to recognize the cluster to which it belongs) and the corresponding reduced mechanism is adopted. This results in a reduction of CPU time, with minimal loss of accuracy.

The method was successfully applied for the simulation of a 2D laminar coflow diffusion flame burning methane in air. The proposed approach was able to correctly describe the flame, both in steady-state and unsteady conditions. Despite the relatively small size of the adopted mechanisms (84 species), a significant reduction of CPU time was observed: in particular, results showed speed-up factors for the chemical step of ~ 4, thus indicating that chemistry was no longer the bottleneck in the flame under investigation, compared to the time for the transport step. The performances from a speed-up point of view are expected to improve if more complex fuels are adopted: since larger and stiffer kinetic mechanisms are required, a higher level of reduction via DRGEP can be reached, with corresponding larger speed-up factors. This approach could also be exploited to include realistic chemistry in the context of Direct Numerical Simulations (DNS). Alleviating the cost of detailed chemical kinetics in these context, indeed, would allow to include more realistic chemistry within the simulations, thus providing very valuable data for the validation of turbulent combustion models for RANS and LES.

Several improvements and extensions of the proposed technique are possible. In particular, coupling with ISAT (In Situ Adaptive Tabulation) [28] is relatively easy to implement: for each cluster, a separate ISAT table can be created and updated for performing the chemical step. A second extension may consist in applying an *on-the-fly* additional mechanism reduction through a fast reduction technique for clusters showing a high level of non-uniformity, for which the union of mechanisms would lead to large mechanisms.

Moreover, the computational savings can be further (and significantly) increased if the number of transport equations to be solved are also reduced, based on the local conditions. However, this task requires the development of appropriate techniques to ensure that the profiles of the chemical species are continuous throughout the computational domain and can be done using mapping methods based on the combination of PCA, LPCA and non-linear regression techniques.

6. Acknowledgements

The first author acknowledges the support of the Fonds National de la Recherche Scientifique (FRS-FNRS) through a FRIA fellowship.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 714605.

We acknowledge the CINECA award under the ISCRA initiative, for the availability of high performance computing resources and support (ISCRA-B: HP10BGXWCZ).

References

[1] T. Lu and C.K. Law. Toward accommodating realistic fuel chemistry in large-scale computations. *Progress in Energy and Combustion Science*, 35(2):192–215, 2009.

[2] T. Lu and C.K. Law. A directed relation graph method for mechanism reduction. *Proceedings of the Combustion Institute*, 30(1):1333–1341, 2005.

[3] Z. Ren and S.B. Pope. Second-order splitting schemes for a class of reactive systems. *Journal of Computational Physics*, 227(17):8165–8176, 2008.

[4] Y. Liang, S.B. Pope, and P. Pepiot. A pre-partitioned adaptive chemistry methodology for the efficient implementation of combustion chemistry in particle pdf methods. *Combustion and Flame*, 162(9), 2015.

[5] L. Liang, J.G. Stevens, and J.T. Farrell. A dynamic adaptive chemistry scheme for reactive flow computations. *Proceedings of the Combustion Institute*, 32 I:527–534, 2009.

[6] F. Contino, H. Jeanmart, T. Lucchini, and G. D'Errico. Coupling of in situ adaptive tabulation and dynamic adaptive chemistry: An effective method for solving combustion in engine simulations. *Proceedings of the Combustion Institute*, 33(2):3057–3064, 2011.

[7] D.A. Schwer, P. Lu, and W.H. Green Jr. An adaptive chemistry approach to modeling complex kinetics in reacting flows. *Combustion and Flame*, 133(4):451–465, 2003.

[8] I. Banerjee and M.G. Ierapetritou. An adaptive reduction scheme to model reactive flow. *Combustion and Flame*, 144(3):619–633, 2006.

[9] Nandakishore Kambhatla and Todd K Leen. Dimension reduction by local principal component analysis. *Neural computation*, 9(7):1493–1516, 1997.

[10] P. Pepiot-Desjardins and H. Pitsch. An efficient error-propagation-based reduction method for large chemical kinetic mechanisms. *Combustion and Flame*, 154(1-2):67–81, 2008.

[11] Alessandro Parente and James C Sutherland. Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combustion and flame*, 160(2):340–350, 2013.

[12] Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.

[13] Alessandro Parente, JC Sutherland, BB Dally, L Tognotti, and PJ Smith. Investigation of the mild combustion regime via principal component analysis. *Proceedings of the Combustion Institute*, 33(2):3333–3341, 2011.

[14] James C Sutherland and Alessandro Parente. Combustion modeling using principal component analysis. *Proceedings of the Combustion Institute*, 32(1):1563–1570, 2009.

[15] G Aversano, A Bellemans, Z Li, A Coussement, O Gicquel, and A Parente. Application of reduced-order models based on pca & kriging for the development of digital twins of reacting flow applications. *Computers & Chemical Engineering*, 2018.

[16] Benjamin J Isaac, Axel Coussement, Olivier Gicquel, Philip J Smith, and Alessandro Parente. Reduced-order pca models for chemical reacting flows. *Combustion and flame*, 161(11):2785–2800, 2014.

[17] DS Broomhead, R Indik, AC Newell, and DA Rand. Local adaptive galerkin bases for large-dimensional dynamical systems. *Nonlinearity*, 4(2):159, 1991.

[18] Keinosuke Fukunaga and David R Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on Computers*, 100(2):176–183, 1971.

[19] T Hediger, A Passamante, and Mary Eileen Farrell. Characterizing attractors using local intrinsic dimensions calculated by singular-value decomposition and information-theoretic criteria. *Physical Review A*, 41(10):5325, 1990.

[20] Alessandro Parente, James C Sutherland, Leonardo Tognotti, and Philip J Smith. Identification of low-dimensional manifolds in turbulent flames. *Proceedings of the Combustion Institute*, 32(1):1579–1586, 2009.

[21] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[22] A. Cuoci, A. Frassoldati, T. Faravelli, and E. Ranzi. Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method. *Energy and Fuels*, 27(12):7730–7753, 2013.

[23] R.K. Mohammed, M.A. Tanoff, M.D. Smooke, A.M. Schaffer, and M.B. Long. Computational and experimental study of a forced, time-varying, axisymmetric, laminar diffusion flame. *Symposium (International) on Combustion*, 27(1):693–702, 1998.

[24] E. Ranzi, A. Frassoldati, A. Stagni, M. Pelucchi, A. Cuoci, and T. Faravelli. Reduced kinetic schemes of complex reaction systems: fossil and biomass-derived transportation fuels. *International Journal of Chemical Kinetics*, 46:512–542, 2014.

[25] E. Ranzi, A. Frassoldati, R. Grana, A. Cuoci, T. Faravelli, A.P. Kelley, and C.K. Law. Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels. *Progress in Energy and Combustion Science*, 38(4):468–501, 2012.

[26] L.L.C. Franke, A.K. Chatzopoulos, and S. Rigopoulos. Tabulation of combustion chemistry via artificial neural networks (anns): Methodology and application to les-pdf simulation of sydney flame l. *Combustion and Flame*, 185:245–260, 2017.

[27] A. Cuoci, A. Frassoldati, T. Faravelli, and E. Ranzi. Opensmoke++: An object-oriented framework for the numerical modeling of reactive systems with detailed kinetic mechanisms. *Computer Physics Communications*, 192:237–264, 2015.

[28] S.B. Pope. Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation. *Combustion Theory and Modelling*, 1(1):41–63, 1997.