Proceedings of Global Power and Propulsion Society
ISSN-Nr: 2504-4400
GPPS Xi'an21
18th – 20th October, 2021
www.gpps.global

GPPS-TC-2021-0349

Automatic extraction of Chemical Reactor Networks from CFD data via advanced clustering algorithms

Matteo Savarese
Université Libre de Bruxelles
Matteo.savarese@ulb.be
Bruxelles, Belgium

Alberto Cuoci
Politecnico of Milan
Alberto.cuoci@polimi.it
Milan, Italy

Ward De Paepe
Université de Mons
Ward.depaepe@umons.ac.be
Mons, Belgium

Alessandro Parente
Université Libre de Bruxelles
Alessandro.parente@ulb.be
Bruxelles, Belgium

ABSTRACT

The design of cleaner and more sustainable combustion technologies represents nowadays a key task. Reliable numerical models, able to cope with a large variety of configurations, combustion processes and fueling mixtures are needed, especially for future applications in combustion monitoring and control, for thermal and environmental performances, which are of critical importance. In this work, alternative, low-computational cost modelling tools for pollutants and thermal efficiency predictions, represented by Chemical Reactor Networks (CRN), are designed from Computational Fluid Dynamic (CFD) simulations assessing a novel methodology, by exploring new possibilities offered by Machine Learning (ML) algorithms. In particular, unsupervised learning approaches are employed, in order to extract the key features of the system flow-field, adopting advanced clustering algorithms, such as Local Principal Component Analysis (LPCA) and K-Means, thus providing an efficient and automatic identification of similar thermo-chemical state compartments in the computational domain. The identified zones are modelled in a post-processing phase as a network of interconnected chemical reactors, and detailed kinetic mechanisms are employed for low concentration pollutants predictions. The case study, a quasi-industrial, flameless-capable combustion furnace, fed with methane-hydrogen mixtures in different compositions at a nominal power of 15 kW, has been investigated numerically by performing 2D CFD simulations with reduced chemistry and subsequently CRN simulations has been carried out with detailed kinetics, adopting the aforementioned approach. Results are validated upon experimental data, in order to provide a novel methodology for CRN design applications, which can be suited for future GTs applications.

INTRODUCTION

In the context of the energy transition, increasing attention needs to be paid to develop sustainable combustion technologies. A valid alternative for an efficient and clean energy conversion system is represented by micro Gas Turbines combined with non conventional combustion regimes, such as Moderate and Intense Low-oxygen Dilution (MILD) combustion (Cavaliere and de Joannon, 2004), thus providing high thermal efficiency, large fuel flexibility and low pollutants emissions (NO_x, CO and Soot).

MILD combustion, also referred to as Flameless (Wünning and Wünning, 1997) or Colorless Distributed Combustion (Arghode and Gupta, 2010) is achieved by means of preheating reactants above their self-ignition temperature and through a strong entrainment of inert combustion products within the reaction region, thus lowering the oxygen concentration at which combustion takes place, resulting in a suppression of temperature peaks, which in turn leads to the abatement of thermal pollutants, such as NO_x, SO_x and soot (Christo and Dally, 2005). The development of a Colorless Distributed Combustion (CDC) chamber for Gas Turbines application has been shown by (Arghode and Gupta, 2011), while (Zornek et al., 2015) have demonstrated experimentally the applicability of the flameless regime to a Turbec T100 to burn low-calorific, biomass derived fuels with a consistent reduction in NO_x and CO emissions.

This work is licensed under **Attribution 4.0 International (CC BY 4.0)** See:
https://creativecommons.org/licenses/by/4.0/legalcode

In order to assist the design of such advanced technologies, reliable numerical models are required, especially for predicting low concentration pollutants, a crucial aspect in assessing environmental performances of combustion systems. Computational Fluid Dynamics (CFD) has been extensively used for modelling turbulent reacting flows, but predicting pollutants emissions still remains an open research challenge, due to computational cost limitations. In this framework, detailed kinetic mechanisms should be employed, in order to account for all the possible formation pathways of minor chemical species, thus leading to unfeasible computational costs associated with the high dimensionality of the problem, despite the constant increase in computational resources (Lu and Law, 2009). For this reason, alternative tools to model complex reactive systems, such as equivalent Chemical Reactor Networks (CRN), have increased their popularity in the combustion community, since they are capable of handling highly detailed chemical mechanisms, consisting of hundreds of species and thousands of reactions, while maintaining the computational requirements on an affordable level. Stemming from the theory proposed by Levenspiel (Levenspiel, 1997), complex reactive flows may be approximated by a series of interconnected, ideal chemical reactor models, namely Perfectly Stirred Reactors (PSR) and Plug Flow Reactors (PFR), reducing the problem to a 0-D or 1-D configuration, where equations for conservation of mass and species are solved only in few "blocks", thus considerably lowering the number and the complexity of the equations to be solved, with respect to a CFD simulation with detailed kinetics. This modelling technique has been applied to industrial furnaces (Faravelli et al., 2001) and burners (de Toni et al., 2013), showing its versatility towards different configurations. It has also been successfully applied to premixed combustion in gas turbines (Lee et al., 2011; Park et al., 2013). In this framework, several works (Hao, 2014; Lee et al., 2011; Nguyen, 2019; Nguyen et al., 2017; Novosselov et al., 2006) focused on the manual design of small-sized CRN of gas turbine combustors from CFD simulations carried out with simplified kinetics, in order to reduce the associated computational cost. The system domain is then arbitrarily divided into different zones, according to temperature, composition and velocity similarities, by observing the main flow-field features provided by CFD simulations, obtaining CRN models able to predict emissions (NO and CO) and thermal performances over a certain range of operating conditions (i.e. equivalence ratio). However, manually designing an equivalent CRN model is a time-consuming and experience-required task, since an automatic and more systematic methodology to obtain a reduced and simple CRN model from CFD data is still missing, thus particularly relying on the user ability.

Examples of automatically generated CRN from CFD data are available in literature. An integrated CFD-CRN procedure is firstly presented by Falcitelli (Falcitelli et al., 2002) and more recently improved by (Monaghan et al., 2012), (Cuoci et al., 2013) and (Stagni et al., 2014) with a fully coupled CFD-CRN approach for detailed pollutants formation analysis by post-processing CFD data with a kinetic post processor. Even though this approach has obtained remarkable results, its adoption aimed at obtaining also a spatial representation of minor chemical species, and the elevated number of reactors included ($\sim 10^3 - 10^4$) only partially reduce the complexity of a CFD computational grid, and the CPU time associated with the solution of such a large network remains still high (~50 min) (Stagni et al., 2014) .

What makes CRN modelling very appealing, is the possibility to use simple configurations to perform fast input-output kinetics calculations, in order to employ this tool for combustion monitoring and control applications. In this sense, a first example of a small-sized CRN of a jet stirred reactor to predict in real time lean blowout for gas turbines applications is presented by Kaluri (Kaluri et al., 2018) and improved by Gupta (Gupta et al., 2019) by building a CRN-based control system also able to prevent the blowout itself. However, even in this case the CRN has been manually designed by observing the general features of the flow-field from CFD simulations, and a procedure to automatically identify a simple network configuration from CFD data is still missing in literature, according to the authors' knowledge, thus limiting the use of CRN modelling only to experienced designers.

In this framework, Machine Learning (ML) algorithms may represent an effective way to understand and extract flow-field key features from CFD simulations data. ML is receveing increasing attention from the combustion community. Advanced statistical tools, such as Principal Component Analysis (PCA), have been used to identify lower-dimensional manifolds in turbulent reactive flows (Parente et al., 2009; Parente and Sutherland, 2013), moreover, PCA in its local formulation (Local PCA), based on a local reconstruction error minimization, has also been used as an unsupervised clustering approach to partition thermo-chemical data for adaptive chemistry applications, to speed up CFD calculations (D'Alessio et al., 2020b, 2020a). In this context, post-processing CFD data by unsupervised clustering algorithms, such as K-Means or Local PCA (LPCA), can represent an effective solution for the identification of similar thermo-chemical state zones in computational domains, by grouping computational cells that exhibit similar temperature and chemical composition.

The application of clustering algorithms for the extraction of CRN from CFD data is investigated in this work, to establish a novel and automatic methodlogy for the design of equivalent CRN models of simplified configurations. The objective is to develop new solutions for an efficient design of reduced-order and physics-based combustion models, which can be used for fast input-output predictions of thermal and environmental performances, providing a useful modelling tool for future applications in combustion processes optimization, real-time monitoring and control.

METHODOLOGY

In this work, a novel methodology for the automatic extraction of equivalent CRN models from CFD data is presented, representing a first step for future GTs applications. In particular, a numerical CFD campaign has been performed on a quasi-industrial MILD combustion furnace (Ferrarotti et al., 2017), operating at a nominal power of 15 kW and fueled with $CH_4 - H_2$ mixtures in variable proportion, with an equivalence ratio of 0.8. The CFD data have been post-processed by applying clustering algorithms, namely K-Means and Local PCA, to identify different zones in the domain by partitioning computational cells showing similar characteristic in terms of temperature and chemical composition. The identified compartments are then modeled as a network of Perfectly Stirred Reactors (PSR) by calculating their volume and the mass flowrates exchanged between clusters. The mass flowrates exchanged across the identified compartments are meant to represent the connections between the reactors, thus schematically approximating the complex flow-field of the original system. Results are then exported for detailed kinetic calculations using Chemkin Pro®, and the reactor network is solved. A sensitivity analysis with the number of clusters (or reactors) has been performed on a single case, e.g. a fixed fuel composition. The most accurate network has been tested on all the cases available, to observe the level of generalization of this approach, employing different kinetic mechanisms, to assess also the effect of the chosen kinetics. Finally, objective criteria to determine an *a priori* number of reactors, for each available case have been tested, relying on clustering evaluation indexes. Results are then validated upon experimental data available from previous works (Ferrarotti et al., 2018; Ferrarotti, 2020).

Case study

The furnace analysed in this paper is schematically represented in Figure 1. The stainless steel combustion chamber has a cubic internal section of 700mm on each side and it is well insulated with a layer of ceramic fiberboards of 200mm thick (2). The furnace has a global nominal power of 20 kW and it is equipped with a finned heat exchanger (1) to recover heat from the flue gases. Fuel and air are preheated through the recuperative heat exchanger and injected co-axially at the bottom-center of the chamber (3). The internal recirculation of flue gases, necessary to achieve MILD regime, is guaranteed by the high velocity jet. Moreover, 4 finned air-cooling tubes (4) are present, which allow for flexible variation of extracted heat from the system, thus simulating an external load. For experimental measurements, an air-cooled suction pyrometer equipped with a 1.5 mm diameter N-type thermocouple (Nicrosil/ Nisil) is used to measure the in-flame temperature profiles, while the temperature of the furnace T_f (5) and the temperature of the flue gases T_{fg} (6) are measured by two shielded N-type thermocouples with the position showed in Figure 1. Finally, exhaust gas composition is measured with electrochemical sensors with nominal accuracies for CO (±2 ppm), NO (±5 ppm), NO2 (±5 ppm), and O2 (±0.8% of reading) and with a non-dispersive infrared (NDIR) sensor for CO2 (±1% of reading + 0.3%). The experimental campaign showed in this paper is already available from previous works (Ferrarotti et al., 2018, Ferrarotti, 2020).

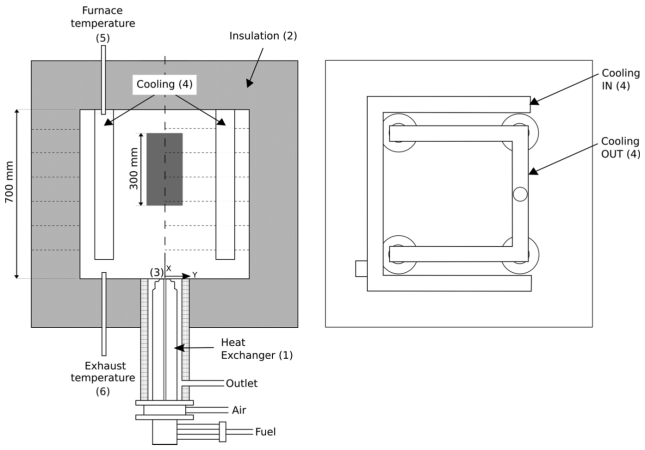

Figure 1: Schematic representation of the MILD combustion furnace: (left) front section (right) view from top

CFD Numerical model

In order to provide the required data for subsequent CRN design, CFD simulations of the experimental cases investigated have been performed. In particular, Reynolds-Averaged Navier-Stokes (RANS) equations are solved on a 2-dimensional grid, since it has been observed that 2-dimensional domains are able to preserve spatial cell connectivity when clustering algorithms are applied.

To model turbulence, the *standard* $k - \epsilon$ model was employed. For chemistry, the Kee chemical mechanism, comprising 17 species and 34 chemical reactions, was chosen, representing a trade-off between chemistry details and

computational cost. Special attention needs to be paid to Turbulence-Chemistry Interactions (TCI) modelling, since it is known that in MILD regime, due to the diffuse and slowered combustion kinetics, induced by the oxygen dilution, chemical and turbulent mixing timescales become comparable (Lewandowski et al., 2020), thus making the hypothesis of infinitely fast kinetics not suitable for this regime. For this reason, the Partially Stirred Reactor (PaSR) (Zhiyi et al., 2017) model was employed, to correctly account for both chemical and mixing timescales. In the PaSR model, the computational cell is divided into a reactive and a non reactive part, and the final species concentration of the cell is determined by the mass exchanged from the two zones, driven by turbulence. In particular, the mass fraction of the reactive zone with respect to the computational cell is calculated as:

$$\kappa = \frac{\tau_C}{\tau_{mix}+\tau_C} \quad (1)$$

where τ_C and τ_{mix} are the chemical and mixing timescales, respectively, that can be estimated in different ways. In this work, the approach used on the same case study, from (Ferrarotti et al., 2018), has been employed, with the static version for τ_{mix} evaluation, with the model constant $C_{mix} = 0.5$ (Ferrarotti et al., 2019; Li et al., 2018).

The mean source term for RANS equation closure is given by a mass transfer between the reactive and the non reactive part of the cell, as follows:

$$\bar{\omega}_i = \kappa \frac{\rho \left(Y_i^* - Y_i^0\right)}{\tau^*} \quad (2)$$

where ρ is the mixture density and τ^* represents the residence time in the reactive zone, which is modelled as an ideal reactor, evolving from Y_i^0 (nass fraction of i-th species in the non-reactive zone) to Y_i^* (mass fraction of the *i*-th species in the reactive zone), following a time-splitting approach:

$$\frac{dY_i^*}{dt} = \frac{\dot{\omega}_i^*}{\rho} \quad (3)$$

Clustering CFD data

Temperature and chemical composition data available from a single CFD simulation, representing the value of state variables in each computational cell, are exported, and we can indicate as $\mathbf{x_j}$ the vector of length n_{var} containing the data in the *i*-th cell, that may also be referred to as the *i*-th observation in the dataset. The data matrix $\mathbf{X}$ will be formed by stacking in rows the available observations, thus having n_{obs} rows (number of observations) and n_{var} columns. The data needs to be pre-processed, by performing centering and/or scaling according to statistics criteria, and, subsequently, a clustering algorithm is applied, by partitioning the observations into a user-defined number of clusters, following an unsupervised learning approach.

Data pre-processing

When dealing with multi-variate dataset, it is a good practice to standardize the data before applying any clustering algorithm. In particular, to the *i*-th variable of each observation the mean value of the *i*-th variable of all the observations is subtracted, in order to center the data (centering). Subsequently, each centered observation needs to be divided by a scaling value (scaling), according to the chosen criteria, in order to normalize raw variables, that usually may show very different values and scales between each other. The process of centering and scaling the *i*-th variable of the *j*-th observation can be indicated as shown in eq. 6:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \bar{x}_{i,j}}{d_i} \quad (6)$$

d_i represents the scaling factor, which can be obtained following different criteria available in statistics (Parente and Sutherland, 2013). In this work, the autoscale criterium has been used, which means that d_i is calculated as the standard deviation of the i-th variable considering all the observations.

Data clustering

Once the data have been pre-processed, a clustering algorithm can be applied, following an unsupervised approach. K-Means is a popular algorithm (Lloyd, 1982) which aims to group the observations in *k* partitions by finding the best μ_j, or cluster *centroids*, according to the following minimization problem:

$$\min_{\mu_j} \sum_{j=1}^{k} \sum_{x_j \in C_j} \left|\left| x_j - \mu_j \right|\right|^2 \quad (7)$$

where C_j denotes the subdomain of data belonging to the j-th cluster, so that the sum of within-clusters Euclidean distances is minimized. The minimization is not solved as a pure optimization problem, but it is indirectly solved following an heuristic algorithm, described below:

1. The *k-means* cluster centroids are initialized and the Euclidean distance between each observation to each centroids is computed
2. Each observation is labelled with the corresponding cluster index that minimize the Euclidean distance
3. The means of all the observations in each cluster are computed, thus updating the cluster centroids. The distance is calculated again between the observations and the updated centroids values, and step (2) is repeated
4. The steps 2 and 3 are repeated until convergence is reached

Local Principal Component Analysis (LPCA), on the other hand, is an unsupervised learning approach that also involves dimensionality reduction and it is based on the minimization of a local reconstruction error:

$$\epsilon_r = \left\| x_i^p - x_i \right\| \quad (8)$$

where x_i^p is the reconstructed observation from the p-truncated dimensional space obtained by performing PCA (Kambhatla and Leen, 1997). The original, centered and scaled, data matrix $\boldsymbol{X} \in \mathbb{R}^{n_{var}}$ is partitioned in k clusters, and in each cluster PCA is performed, thus finding k reduced basis of Local Principal Components (LPCs) $\boldsymbol{A}^j \in \mathbb{R}^p$ with $p < n_{var}$. At this point, for each observation it is possible to iteratively compute k local reconstruction errors, and label the observation with the cluster index for which the reconstruction error is minimum. The algorithm proceeds as follows:

1. Initialization of the centroids
2. Partition each observation according to equation (8)
3. The cluster centroids are updated according to the new partition
4. PCA is performed in each cluster

The steps from 2 to 4 are repeated until a convergence criteria is met: in this case until the global mean reconstruction error is below a fixed threshold and/or the cluster centroids no longer change.

CRN modelling

Once the clustering has been performed, it is possible to visualize the different clusters as the identified zones in the computational domain that exhibit similar thermo-chemical characteristic. A typical output resulting from applying clustering to CFD data is represented in Figure 3.

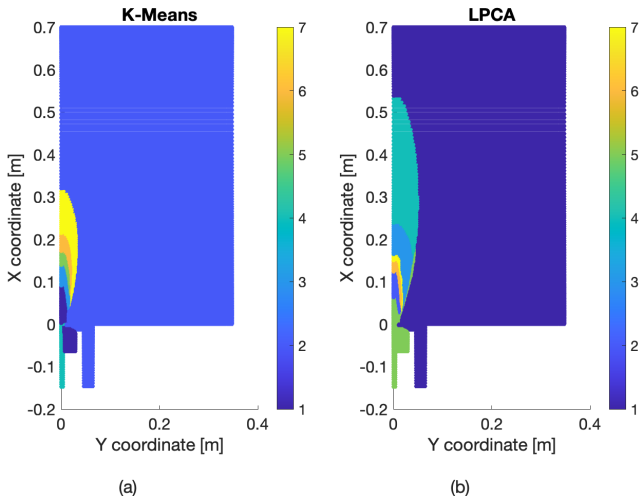

Figure 2: A typical visualization on the CFD domain of a clustering output

The volume of each cluster is calculated by summing the individual volumes of the cells belonging to that cluster. The mass flowrates across neighbors clusters are then retrieved from the CFD simulation. Results are exported in Chemkin Pro®, creating a network of Perfectly Stirred Reactors with the same volume obtained by the clustering and assigning the connections according to the computed mass flowrates. The choice of using only PSRs is justified by the small volume of the in-flame reactors, which is obtained by the clustering. To account for initial incomplete mixing of the reactants, the

cluster representing the injection zone is modelled as a PSR with an infinitesimal volume, thus working as a mixer, since the residence time is too low and ignition cannot take place. The boundary conditions are then specified, namely inlets, outlet and heat losses, which are estimated from the experiments and arbitrarily assigned to the largest post-flame reactor, while assuming the other reactors adiabatic. Once a detailed kinetic mechanism is specified, including all the minor pollutants, the network (mass and energy conservation equations) can be finally solved by the global solver of Chemkin Pro®, which ensure rapid convergence (~10 s).

RESULTS

CFD results

CFD results are shown in Figure 4. Results are in good agreement with experimental data, even though a mismatch is observed in the cases corresponding to Figure 4 (b) and 4 (e). Regarding the case in Figure 4 (b), an explaination can be related to the fact that a slight addiction of hydrogen can greatly affect the position of the reactive zone, and in that specific case can result in a transition from a pure MILD to a partially flame regime, thus being difficult to predict correctly. As for the case in Figure 4 (e), the mismatch is probably due to the model constant, which may overestimates the reactive zone.

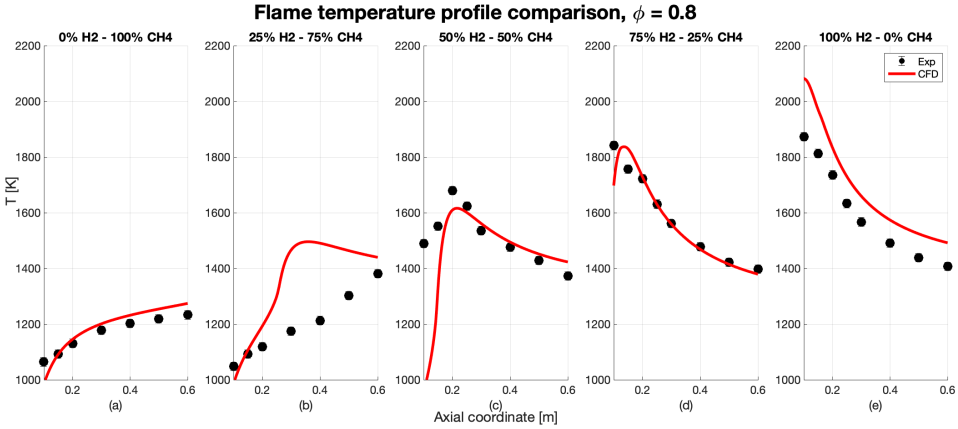

Figure 3: Comparison between the experimental and the CFD flame temperature profile

Chemical Reactor Network results

First, the effect of the clustering algorithm and the main clustering set-up parameter, namely the number of clusters chosen, needs to be assessed both in a qualitative and quantitative manner, by observing the physical shape of the resulting clustering output and by comparing the predictions of temperature and NO emissions of the extracted network with experimental data. Attention is focused on a single experimental case, namely the 75% H_2 – 25% CH_4 mixture, and a sensitivity analysis with the number of reactors and the clustering algorithms employed, namely K-Means and LPCA, is performed in order to assess their impact on the predictions. The GRI 3.0 (Smith et al., 2021) mechanism has been used at this stage. Then, the network which showed better agreements with experimental values, was used to simulate all the experimental cases, to assess the level of generalization of this CRN approach. The effect of the kinetic mechanism employed was evaluated at this stage, by performing simulations with three different chemical mechanisms: the GRI 3.0, GRI 2.11 (C.T. Bowman et al., 2021) and the POLIMI C1-C3 (RANZI et al., 2014). Finally, since an objective criteria to efficiently select a reasonable *a priori* number of cluster is needed, to have a more systematic CRN design process, a network was extracted on every case available, by choosing the number of clusters which optimizes clustering evaluation indexes available in literature (Davies and Bouldin, 1979).

Effect of clustering parameters

The K-Means and the LPCA clustering algorithm were used to cluster the computational domain of the 75% H_2 – 25% CH_4 mixture, by using 5, 8 and 11 clusters, since we want to maintain a very low amout of units, to speed up the calculations. The GRI 3.0 chemical mechanism was employed and the network was solved with Chemkin Pro®. Predictions of outlet temperature, maximum temperature and NO are compared with experimental values. Results are reported in Table 1. We can observe that increasing the number of reactors employed also results in increased accuracy, meaning that the flow-field is better represented by the augmented network structure, since more features of the flow-field itself are included. Regarding the algorithms, Local PCA shows better predictions with respect to K-Means, and it is able to match very closely experimental results in the 11 reactors case. This can be explained by observing the typical clustering output differences in Figure 3, where the Local PCA seems to better capture the flame extension, resulting in a correct prediction of the residence time spent by the mixture in the reactive zone, which is of key importance for pollutants emission characterization. The

maximum temperature appears a very sensitive parameter, which closely depends by the recirculation and the residence time in the ignition reactor, values that are changing when clustering parameters are modified.

Table 1: Results of different networks extracted on the 75% H2 case using K-Means and Local PCA with different number of reactors

K-Means				
Variable	***Exp value***	***5 reactors***	***8 reactors***	***11 reactors***
Outlet temperature [K]	1255 ± 20	1281	1275	1275
Max temperature [K]	1837 ± 20	1753	1787	1944
Outlet NO [ppm]	35.6 ± 5	2.1	10.7	22.1
Local PCA				
Variable	***Exp value***	***5 reactors***	***8 reactors***	***11 reactors***
Outlet temperature [K]	1255 ± 20	1275	1275	1275
Max temperature [K]	1837 ± 20	1524	2041	1981
Outlet NO [ppm]	35.6 ± 5	3.2	51.7	36.3

Network generalization

The most accurate network identified before (LPCA, 11 reactors) was applied to all the experimental cases available, to determine to which extent a CRN generated on a certain case can generalize a wide range of operating conditions. Different chemical mechanisms were employed at this stage, to assess also the effect of chemical kinetics on the obtained results. Indeed, considering the composition variability of the fuel mixtures, the properties of the chemical mechanisms can play a major role.

Results in Figure 5, are showing that the CRN extracted on the 75% H_2 mixture provides good predictions in closeby conditions but it fails to capture the combustion features of mixtures richer in CH_4. This is due to changes in the flow-field occurring at different fuel compositions, which cannot be captured completely by one representative case, highlighting the case-specific nature of the CRN approach.

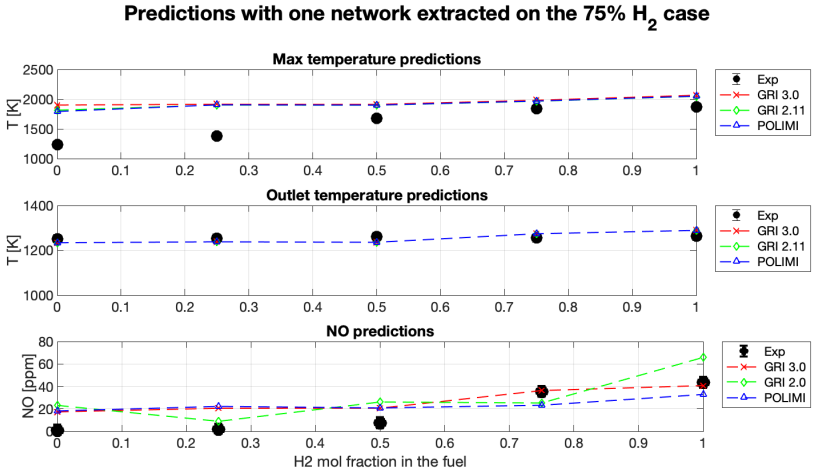

Figure 4: Results from the CRN generated on the 75% H2 case on all the cases, by using different chemical mechanisms

Network extracted optimizing clustering evaluation indexes

To assess the impact of quantitative metric on the *a priori* determination of the number of clusters, a single CRN was extracted for each experimental point, optimising, for each case, the Davies-Bouldin index (DB) (Davies and Bouldin, 1979). The DB index is calculated as the ratio of within-cluster and between-cluster distance in the data space:

$$DB = \frac{1}{k}\sum_{i=1}^{k} \max_{j\neq i}\{D_{i,j}\} \quad (9)$$

where $D_{i,j}$ is the within-to-between cluster distance ratio for the i-th and j-th clusters, or in other terms:

$$D_{i,j} = \frac{\overline{d}_i + \overline{d}_j}{d_{i,j}} \quad (10)$$

In Eq 10, $\overline{d}_i$ and $\overline{d}_j$ are the average distances between each point in the i-th and j-th cluster and the centroid of i-th and j-th cluster, respectively, while $d_{i,j}$ is the Euclidean distance between the i-th and j-th cluster centroids.

For each experimental case, a parametric study varying the number of clusters using LPCA was performed. Then, the number of clusters minimising the DB index was chosen, and a network with that number of clusters was extracted and simulated,. Results are reported in Figure 6.

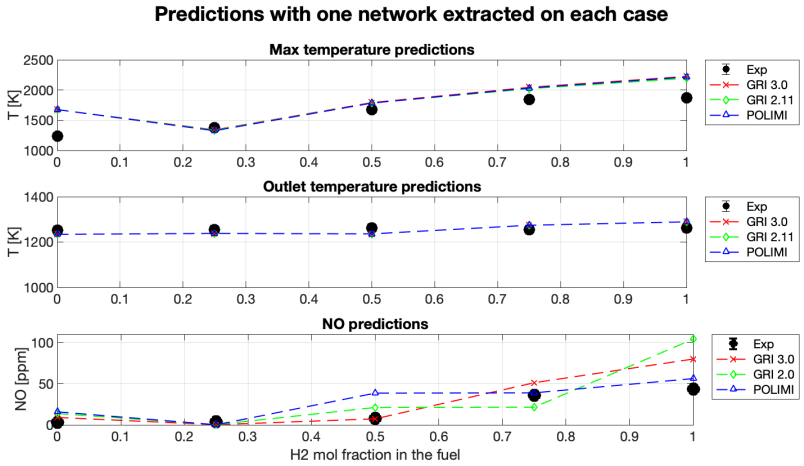

Figure 5: Results of CRN extracted for each single case by minimizing the Davies-Bouldin index

Figure 6 shows that results improved with respect to using a single network to predict the whole case set. The variation of the DB index with the number of clusters along the different cases is reported in Figure 7.

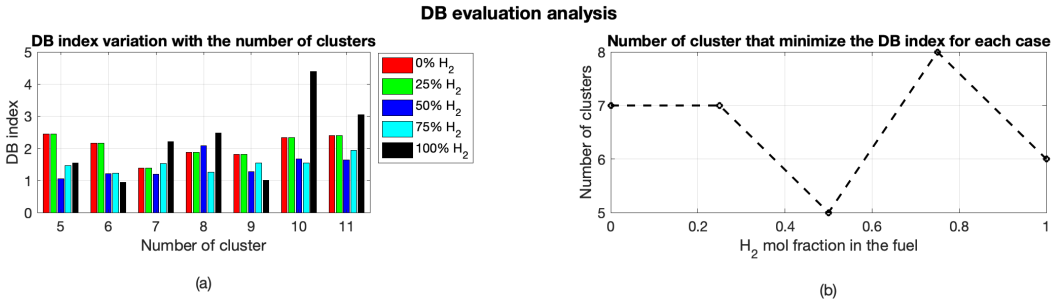

Figure 6: DB index variation with the number of clusters, along the different cases

We can see that no substantial variation of this index appears in varying the number of clusters in the range of fuel composition investigated, except from the case with pure H_2. This means that increasing the number of cluters does not greatly affect the clustering structure and explains the little sensitivity to predictions for the case with 75% H_2. However, the DB index can be very useful to reveal substantial changes in the clustering effectiveness, as with the case with 100% H_2, since going from 9 to 10 clusters results in a great increase of the index value, meaning that the cells are not yet grouped in an afficient way, which can also be non-optimal for CRN applications.

CONCLUSIONS

In this work, a novel methodology is presented for the extraction of Chemical Reactor Networks from CFD data using unsupervised clustering algorithms. The effect of clustering settings and parameters on the CRN output results is discussed, and simulations with CRN and detailed kinetics are performed. A possible *a priori* criterion for the identification of the optimal number of clusters is also illustrated.

Predictions of temperature and low concentration pollutants (NO) obtained from CRN appears in good agreement with experimental values. Special attention needs to be paid to the computational time required to solve the networks, which is very small (~10 s), making the whole CRN design process and modelling very appealing for control and optimisation applications.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Energy Transition Fund of Belgium.

REFERENCES

Arghode, V.K. and Gupta, A.K. (2010). Effect of flow field for colorless distributed combustion (CDC) for gas turbine combustion. *Applied Energy* 87, pp. 1631-1640.

Arghode, V.K. and Gupta, A.K. (2011). Development of high intensity CDC combustor for gas turbine engines. *Applied Energy* 88, pp. 963-973.

Bowman, C.T., Hanson, R.K., Davidson, D.F., Gardiner, W.C., Lissianski, V., Smith, G.P., Golden, D.M., Frenklach, M. and Goldenberg, M. GRI-MECH 2.11, Available at: http://www.me.berkeley.edu/gri_mech/ [Accessed 09/08/2021]

Cavaliere, A. and de Joannon, M., (2004). Mild combustion. *Progress in Energy and Combustion Science*, 30, pp. 329-366

Christo, F.C. and Dally, B.B. (2005). Modeling turbulent reacting jets issuing into a hot and diluted coflow. *Combustion and Flame*, 142, pp. 117-129.

Cuoci, A., Frassoldati, A., Stagni, A., Faravelli, T., Ranzi, E. and Buzzi-Ferraris, G. (2013). Numerical modeling of NOx formation in turbulent flames using a kinetic post-processing technique. *Energy and Fuels*, 27, pp. 1104-1122.

D'Alessio, G., Cuoci, A., Aversano, G., Bracconi, M., Stagni, A. and Parente, A. (2020a). Impact of the partitioning method on multidimensional adaptive-chemistry simulations. *Energies,* 13, 2567; doi:10.3390/en13102567.

D'Alessio, G., Parente, A., Stagni, A. and Cuoci, A. (2020b). Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. *Combustion and Flame*, 211, pp. 68-82.

Davies, D.L. and Bouldin, D.W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1 (2), pp. 224-227..

de Toni, A., Hayashi, T. and Schneider, P. (2013). A reactor network model for predicting NOx emissions in an industrial natural gas burner. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 35, pp. 199-206.

Falcitelli, M., Tognotti, L. and Pasini, S. (2002). An algorithm for extracting chemical reactor network models from CFD simulation of industrial combustion systems. *Combustion Science and Technology*, 174, pp. 27-42.

Faravelli, T., Bua, L., Frassoldati, A., Antifora, A., Tognotti, L. and Ranzi, E. (2001). A new procedure for predicting NO x emissions from furnaces. *Computers and Chemical Engineering,* 25, pp. 613-618.

Ferrarotti, M., Fürst, M., Cresci, E., de Paepe, W. and Parente, A. (2018). Key Modeling Aspects in the Simulation of a Quasi-industrial 20 kW Moderate or Intense Low-oxygen Dilution Combustion Chamber. *Energy and Fuels*, 32, pp. 10228-10241.

Ferrarotti, M., Li, Z. and Parente, A. (2019). On the role of mixing models in the simulation of MILD combustion using finite-rate chemistry combustion models. *Proceedings of the Combustion Institute*, 37, pp. 4531-4538.

Ferrarotti, M., Lupant, D. and Parente, A. (2017). Analysis of a 20 kW flameless furnace fired with natural gas. *Energy Procedia*, 120, pp. 104-111.

Ferrarotti, M. (2020). *Experimental and numerical investigation of fuel flexibility and pollutant emissions in novel combustion technologies using renewable synthetic fuels*. Ph.D. Université Libre de Bruxelles

Gupta, S., Malte, P., Brunton, S.L. and Novosselov, I. (2019). Prevention of lean flame blowout using a predictive chemical reactor network control. *Fuel*, 236, pp. 583-588.

Hao, N.T. (2014). A chemical reactor network for oxides of nitrogen emission prediction in gas turbine combustor. *Journal of Thermal Science*, 23, pp. 279-284.

Kaluri, A., Malte, P. and Novosselov, I. (2018). Real-time prediction of lean blowout using chemical reactor network. *Fuel*, 234, pp. 797-808.

Kambhatla, N. and Leen, T.K., 1997. Dimension Reduction by Local Principal Component Analysis. *Neural Computation*, 9, pp. 1493-1516.

Lee, D., Park, J., Jin, J., Lee, M., 2011. A simulation for prediction of nitrogen oxide emissions in lean premixed combustor. *Journal of Mechanical Science and Technology*, 25, pp. 1871-1878.

Levenspiel, O. (1997) *Chemical reaction engineering*, 3rd ed. John Wiley & Sons.

Lewandowski, M.T., Parente A. and Pozorski, J. (2020). Generalised Eddy Dissipation Concept for MILD combustion regime at low local Reynolds and Damköhler numbers. Part 1: Model framework development. *Fuel*, 278, 117743; https://doi.org/10.1016/j.fuel.2020.117743

Li, Z., Ferrarotti, M., Cuoci, A. and Parente, A. (2018). Finite-rate chemistry modelling of non-conventional combustion regimes using a Partially-Stirred Reactor closure: Combustion model formulation and implementation details. *Applied Energy*, 225, pp. 637-655.

Lloyd, S.P. (1982). Least Squares Quantization. *EEE Transactions on Information Theory*, 28(2), pp. 129-137.

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

Monaghan, R.F.D., Tahir, R., Cuoci, A., Bourque, G., Füri, M., Gordon, R.L., Faravelli, T., Frassoldati, A. and Curran, H.J. (2012). Detailed multi-dimensional study of pollutant formation in a methane diffusion flame. *Energy and Fuels*, 26, pp. 1598-1611.

Nguyen, T.H. (2019). Improved Chemical Reactor Network Application for Predicting the Emission of Nitrogen Oxides in a Lean Premixed Gas Turbine Combustor. *Combustion, Explosion and Shock Waves*, 55, pp. 267-273.

Nguyen, T.H., Kim, Seunghan, Park, J., Jung, S., Kim, Shaun (2017). CFD-CRN validation study for NOx emission prediction in lean premixed gas turbine combustor. *Journal of Mechanical Science and Technology*, 31, pp. 4933-4942.

Novosselov, I. v, Malte, P.C., Yuan, S., Srinivasan, R. and Lee, J.C.Y. (2006). Chemical reactor network application to emissions prediction for industial dle gas turbine. *Proceedings of the ASME Turbo Expo*, 1., 10.1115/GT2006-90282.

Parente, A. and Sutherland, J.C. (2013). Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combustion and Flame*, 160, pp. 340-350.

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

Park, J., Nguyen, T.H., Joung, D., Huh, K.Y. and Lee, M.C. (2013). Prediction of NOx and CO emissions from an industrial lean-premixed gas turbine combustor using a chemical reactor network model. *Energy and Fuels*, 27, pp. 1643-1651.

Ranzi, E., Frassoldati, A., Stagni, A., Pelucchi, M., Cuoci, A. and Faravelli, T. (2014). Reduced Kinetic Schemes of Complex Reaction Systems: Fossil and Biomass-Derived Transportation Fuels. *International Journal of Chemical Kinetics*, 46(9), pp. 512-542.

Smith, G.P, Golden, D.M., Frenklach, M., Moriarty, N.W., Eiteneer, B., Goldenberg, M. C., Bowman, T., Hanson, R.K., Song, S., Gardiner, W.C., Lissianski, V.V. and Qin, W. GRI-MECH 3.0, Available at: http://www.me.berkeley.edu/gri_mech/ [Accessed 9/08/2021]

Stagni, A., Cuoci, A., Frassoldati, A., Faravelli, T. and Ranzi, E. (2014). A fully coupled, parallel approach for the post-processing of CFD data through reactor network analysis. *Computers and Chemical Engineering*, 60, pp. 197-212.

Wünning, J.A. and Wünning, J.G. (1997). Flameless oxidation to reduce thermal no formation. *Progress in Energy and Combustion Science*, 23(1), pp. 81-94

Zhiyi, L., Cuoci, A., Sadiki, A. and Parente A. (2017). Finite-rate chemistry modelling of non conventional combustion regimes. *Energy Procedia*, 142, pp. 1570-1576

Zornek, T., Monz, T. and Aigner, M. (2015). Performance analysis of the micro gas turbine Turbec T100 with a new FLOX-combustion system for low calorific fuels. *Applied Energy*, 159, pp. 276-284.