Available online at www.sciencedirect.com

ScienceDirect

Proceedings of the Combustion Institute 38 (2021) 5373–5381

Proceedings of the Combustion Institute

www.elsevier.com/locate/proci

Digital twin of a combustion furnace operating in flameless conditions: reduced-order model development from CFD simulations

Gianmarco Aversano [a,b], Marco Ferrarotti [a,b], Alessandro Parente [a,b,*]

[a] *Aero-Thermo-Mechanics Departement, Université Libre de Bruxelles, Avenue F.D. Roosevelt 51, CP 165/41, Brussels 1050, Belgium*

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

Received 7 November 2019; accepted 3 June 2020
Available online 30 September 2020

Abstract

The present paper presents the first-of-its-kind digital twin for a furnace operating in flameless combustion conditions. A methodology combining data compression, by means of Proper Orthogonal Decomposition (POD), and interpolation, using Kriging, was developed to design physics-based, reduced-order models (ROMs) for the prediction of combustion data at unexplored operating conditions. Three-dimensional simulations with detailed chemistry were carried out, spanning a wide range of operating conditions in terms of fuel composition (methane-hydrogen mixtures from pure methane to pure hydrogen), equivalence ratio (from 0.7 to 1) and air injector diameter (to adjust the air jet entrainment). Based on the available simulations, a ROM was developed, to predict both spatial fields, local and integral values of thermochemical variables at working conditions not included in the ROM development. Results showed that the developed ROM could reliably predict the temperature and main chemical species distribution in the furnace with an overall error below 10%, proving the effectiveness of the approach for the development of digital twins of combustion systems. A remarkable accuracy was observed for the prediction of specific quantities, including wall temperatures, OH decay length, OH peak value and location and exhaust gas composition, including pollutants, with prediction errors always below 5%, showing the potential of the approach to develop soft sensors.

Keywords: Digital twins; Flameless combustion; Kriging; Proper orthogonal decomposition; Reduced-order modelling

* Corresponding author at: Aero-Thermo-Mechanics Departement, Université Libre de Bruxelles, Avenue F.D. Roosevelt 51, CP 165/41, 1050 Brussels, Belgium.

E-mail addresses: Gianmarco.Aversano@ulb.ac.be (G. Aversano), Marco.Ferrarotti@ulb.ac.be (M. Ferrarotti), Alessandro.Parente@ulb.be (A. Parente).

https://doi.org/10.1016/j.proci.2020.06.045

1. Introduction

Energy supply is one of the greatest societal challenges we are facing. The intermittent nature of renewable sources requires the development of long-term storage solutions as well as the availability of high-density energy sources, for transportation and manufacturing. The storage of renewable excess energy in the form of synthetic fuels is a unique opportunity to integrate renewable sources and combustion systems. The future energy mix will likely include a variety of fuels from power-to-fuel technologies, making the role of combustion technologies even more important than it already is. Novel combustion technologies are indeed required to accommodate the fuel flexibility expected in the future. In this context, Flameless Oxidation (FLOX®) [1] or Moderate and Intense Low-oxygen Dilution (MILD) [2] combustion represents a very attractive solution for its little sensitivity to the charge composition and ability to deliver very high combustion efficiency with very low pollutant emissions. While CFD tools have significantly progressed in recent years, their use in real time is still unrealistic, especially for combustion regimes such as MILD combustion, whose description requires the use of detailed chemical mechanisms and advanced turbulence-chemistry interactions approaches. In this context, the availability of physics-based reduced-order models (ROMs) becomes very attractive, to embed the critical aspects of a detailed simulations into simplified relationships between the inputs and outputs that can be used in real time. The development of virtual models, also referred to as *digital twins*, of industrial systems opens up a number of opportunities, such as the use of data to anticipate the response of a system and brainstorm malfunctioning, and the use of simulations to develop new technologies, i.e. virtual prototyping. A definition of digital twins is "an integrated multi-physics, multi-scale, probabilistic simulation of an as-built system, enabled by digital thread, that uses the best available models, sensor information, and input data to mirror and predict activities/performance over the life of its corresponding physical twin" [3]. Combining CFD simulations with real-time data coming from sensors of a real industrial system to foresee a change in its state is possible only if the prediction of the system's state based on the operating conditions reported by these sensors becomes instantaneous [4]. To do so, a set of training simulations must be generated beforehand, for a wide enough range of possible operating conditions. A physics-based ROM can be then developed by using unsupervised learning to extract the key latent features in the data, for which a response surface is subsequently found by a supervised learning technique. Once the mapping between inputs and outputs is embedded in a ROM, the system state can be predicted for new operating conditions, based on real-time data coming from sensors. In a previous study [5], the Authors showed that the combination of an unsupervised data compression method, i.e. Proper Orthogonal Decomposition (POD), with a supervised interpolation technique, i.e. Kriging, could be effectively used for the reconstruction and prediction of two-dimensional laminar methane flames. In the present work, the methodology was extended and applied to the development of a digital twin of a combustion furnace equipped with a FLOX® burner for the prediction of the full state of the furnace (spatial fields of temperature and main chemical species mass fractions), as well as of important scalar quantities at locations of interest (wall temperature, OH peak value and location, OH decay length, exhaust gas composition including pollutants), within a prescribed accuracy. The design space consisted of a design parameter, the air injector diameter, and two measured inputs, the fuel composition (mixture of H_2/CH_4 in molar basis) and the equivalence ratio. The paper is organized as follows: the methods used for the ROM development are described in Section 2. The description of the case study is reported in Section 3. The sensitivity to the training data of the data compression process and of POD+Kriging are reported in Sections 4.1 and 4.2, respectively. The performances of the developed digital twin are discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Methods

2.1. Proper orthogonal decomposition

Consider a snapshot matrix $\mathbf{Y}$ of size $(m \times n)$, where each row of $\mathbf{Y}$ is a vectorized 2D or 3D spatial field of some variable of interest such as temperature, or a concatenation of more than one field, coming from one CFD simulation. Thus, m is the number of available simulations and n is the number of grid points $\times$ the number of considered variables. In combustion-related problems, $n >> m$. Proper Orthogonal Decomposition (POD) seeks $\mathbf{Z}$ of size $(m \times k)$ and $\mathbf{A}$ of size $(n \times k)$ with $k << n$ (hence the reduction), such that the functional $f(\mathbf{Z}, \mathbf{A}) = \frac{1}{2}||\mathbf{Y} - \mathbf{Z}\mathbf{A}^T||^2$ is minimized, subject to $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This problem can be solved by computing the singular value decomposition (SVD) of the matrix $\mathbf{Y}$, which corresponds to finding the eigenvectors of the matrix $\mathbf{C} = \frac{1}{m-1}\mathbf{Y}^T\mathbf{Y}$. These eigenvectors are the columns of $\mathbf{A}$. A low-rank approximation of $\mathbf{Y}$ is found as follows $\mathbf{Y} \approx \mathbf{Z}\mathbf{A}^T = \mathbf{Y}\mathbf{A}\mathbf{A}^T$, where the columns of $\mathbf{A}$ of size $(n \times k)$ are the POD modes, also referred to as spatial shapes or *eigenflames* [5], and $\mathbf{Z}$ of size $(m \times k)$ is the matrix of POD coefficients. Each column of $\mathbf{Z}$ are the k coefficients for the retained k POD modes so that one

particular simulation, or row of $\mathbf{Y}$, can be expressed as a weighted sum of different data-driven eigenflames: $\mathbf{y}(\mathbf{x}) = \sum_{i=1}^{k} \mathbf{a}_i z_i(\mathbf{x})$.

2.2. Kriging

Kriging is an interpolation method in which every realization $z(\mathbf{x})$ (e.g. the POD coefficients) is expressed as a combination of a trend function and a residual [6]:

$$z(\mathbf{x}) = \mu(\mathbf{x}) + s(\mathbf{x}). \quad (1)$$

The trend function $\mu(\mathbf{x})$ is a low-order polynomial regression and provides a global model in the input space. The residuals $s(\mathbf{x})$ are modeled by a Gaussian process with a kernel or correlation function that depends on a set of hyper-parameters to be evaluated by Maximum Likelihood Estimation (MLE) [7,8].

3. Data-set

The semi-industrial combustion furnace used for this study has a nominal power of 20 kW. It is fired by a burner with an integrated metallic finned heat exchanger to recover heat from the flue gases and to preheat the combustion air. The fuel is injected via a centrally located nozzle (ID 8 mm) and surrounded by a coaxial air jet, whose dimensions can be varied to adjust the air jet entrainment (ID 16-20-25 mm). More details can be found in [9]. To generate the samples required for development of the furnace ROM, CFD simulations were carried out using the commercial software Ansys Fluent 19.1. A constant input power of 20 kW was fixed, while the cooling flow rate was set to reach a furnace outlet temperature of T_{out}=1000 °C. Furthermore, the four sides of the furnace were closed with insulated plates. Therefore, a 45° degrees angular sector of the 3D geometry of the furnace was considered, as a result of the symmetry of the problem. The computational grid was first created with tetrahedrons and then converted into polyhedrons. A grid-independency study based on temperature profiles along the longitudinal direction was also performed and more details can be found in [9]. The selected grid consists of about 216k cells.

The standard $k-\epsilon$ turbulence model was used in combination with the PaSR model [9,10] for turbulence-chemistry interactions. Following [9], a C_{mix} of 0.5 was set for the determination of an appropriate mixing scale in the PaSR approach. A sensitivity study was carried out to select a kinetic scheme, comparing the KEE (17 species and 58 reactions) and GRI-2.11 (31 species and 175 reactions) mechanisms. The detailed comparison is presented in the Supplementary material. Being the difference between the two schemes below 3%, KEE was selected for its lower computational cost. The discrete ordinate (DO) radiation model was used, in combination with the weighted-sum-of-gray-gases (WSGG) model, using the coefficients proposed by Smith et al. [11]. The NO modeling was handled by the post-processing tool of ANSYS Fluent, which assumes that NO species have negligible effects on the overall temperature and major species concentration fields. The tool considers Thermal, Prompt and N_2O pathways, but not the NNH route proposed by Konnov et al. [12], which was implemented by means of a bespoke User Defined Function (UDF). The latter appears to be dominant in presence of hydrogen.

Table 1
Validation cases for the CFD model.

Case	Injector [mm]	ϕ [–]	H_2[mole fraction]
a	16	0.93	0.60
b	16	0.78	0.30
c	16	0.83	0.05

Three input parameters were considered to generate the simulation samples: fuel composition (mixture of methane/hydrogen), equivalence ratio and air injection geometry. A design of experiments (DoE) was established using latin hypercube sampling, varying the input parameters in the range 0–100 % (H_2 mole fraction), 0.7–1 (equivalence ratio ϕ) and 16–20–25 mm (air injector size). A total of 45 simulations were carried out. The variables of interest selected for the generation of the furnace ROM were the temperature, major species (CH_4, H_2, O_2, H_2O, OH), minor species (CO and OH), and pollutants (NO).

Before generating the digital twin based on the CFD simulations, it is key to ensure that the numerical simulations are a good representation of reality. The computational model used in this work was already validated for a number of systems, including the Adelaide Jet in Hot Co-flow [13], fed with an equimolar methane-hydrogen mixture, and the same furnace investigated here, fed with natural gas [9]. Nevertheless, an additional validation was carried out, for some of the cases simulated in the present study, listed in Table 1. The comparison is presented for Cases *a* and *b* in Fig. 1, in terms of measured and computed temperature profiles at several axial positions (z=100, 200, 300, 400, 500, 600 mm). The application of a suction pyrometer [9] implies an averaging effect, typical of this measurement technique. For this reason, the CFD results were averaged in a volume defined by a sphere with a radius of 10 mm (equal to the probe diameter) to allow a fair comparison. CFD predictions show excellent agreement with experimental data, with an averaged error below 6%. Similar results were obtained for case *c* and they are shown in the Supplementary material, underlying the fact that the numerical model is flexible enough to handle a wide range of hydrogen content in the fuel. Pollutant emissions (NO) on dry-basis were also

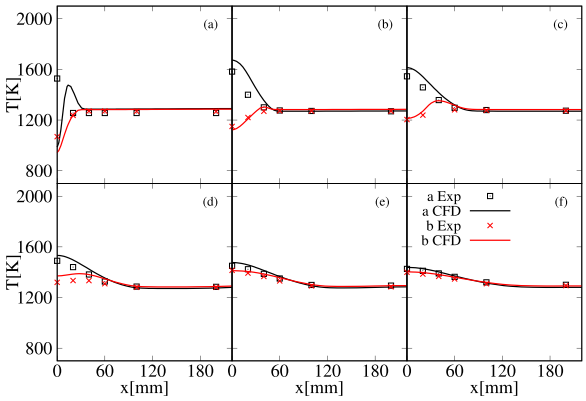


Fig. 1. Validation of the CFD model for Cases *a* and *b* against measured temperature profiles at several axial positions, (a) $z = 100$ mm, (b) $z = 200$ mm, (c) $z = 300$ mm, (d) $z = 400$ mm, (e) $z = 500$ mm and (f) $z = 600$ mm. Averaged experimental uncertainty of 10 K.

Table 2
Experimental vs CFD NO emissions on dry basis. Experimental uncertainty of 5 ppm.

Case	Exp [ppm]	CFD [ppm]
a	13	11
b	2	1
c	1	0.5

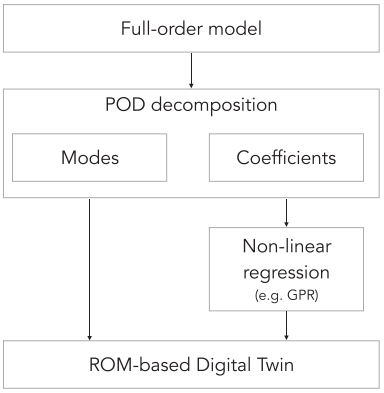


Fig. 2. Reduced-order model generation: identification of invariant and system-dependent information.

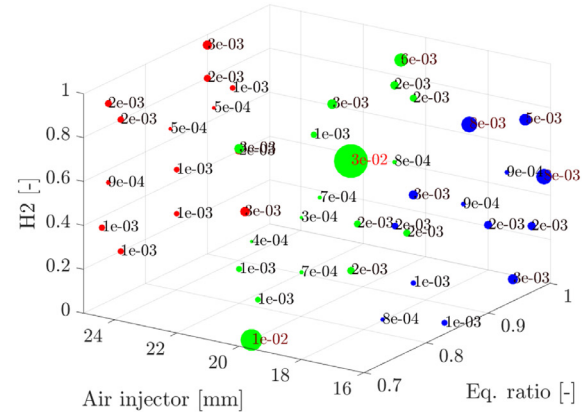
benchmarked against experimental data and results are listed in Table 2. The post-processing approach appears a fair compromise between accuracy and complexity.

4. Reduced order model development

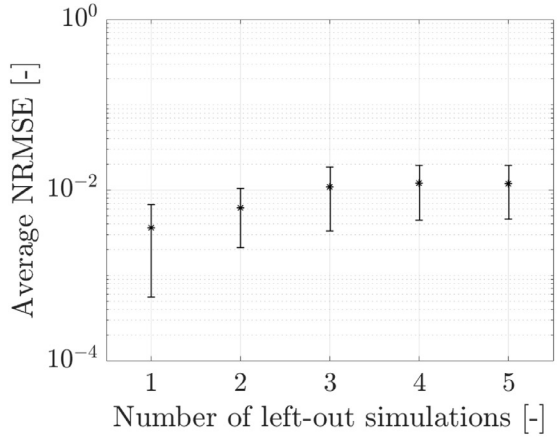
The furnace ROM was developed based on the methodology shown in Fig. 2. The approach, introduced in [5], allows to distinguish between invariant information, the POD modes, and system-dependent ones, the POD coefficients. The POD modes are kept constant, as they represent the intrinsic system physics. The POD coefficients, on the other hand, are used to represent the system variability due to changes in the boundary conditions. This relationship is modelled by means of non-linear regression approaches, Kriging in the present case. The accuracy of the reduced-order model is then dependent on the degree of reduction imposed during the POD decomposition as well as on the training data used to identify the POD modes and coefficients. These aspects are critically discussed in the next sections for the present study. The maximum number of POD modes that could be extracted from the data-set was $m - 1$, where m is the total number of available simulations. In the present study, the data-set related to one particular field (e.g. temperature) consisted of a matrix of size $(m \times l)$, with $m = 45$ and $l = 216,360$, thus a total of 44 POD modes could be identified and used to encode each simulation (a vector of 216,360 real numbers) into a set of 44 coefficients, for which a Kriging response surface was found. The ROM developed in the present work requires the training of a reduced set of scalars (from 216,360 to 44) from the POD decomposition, for which the supervised Kriging method is used. Once built, the evaluation of the ROM is almost instantaneous and can be used in real time, while each of the 45 numerical simulations required 1440 CPU hours on 20 cores.

4.1. Reconstruction of test data

The set of training data to be used for the generation of the reduced-order model out of the 45 available CFD simulations was determined using the sampling strategy described in [5]. This method allows to associate an importance index to each available simulation, based on the influence they



(a)



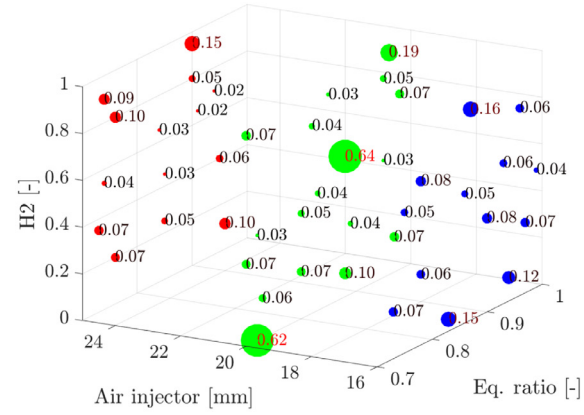
(b)

Fig. 3. (a) Leave-one-out reconstruction errors, visualized in the input parameter space. The reported figures are the average normalized root mean squared errors (NRMSE) across all variables, for the reconstruction of one particular left-out simulation. The sizes of the circles are proportional to the error. Different colours are used for the 3 different values of the air injector size. (b) Average NRMSE for the reconstruction of an increasing number of left-out simulations. Vertical bars represent the standard deviation of the error associated to different combinations of k left out simulations.

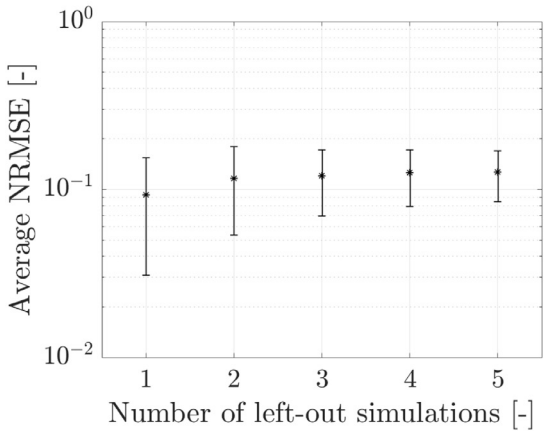
have on the reduced POD basis. To assess the importance of the training data size on the POD basis and understand if enough data had been collected for ROM development, a leave-k-out cross validation analysis was performed, where k was the number of simulations left out from the overall available training data set. Each time, k simulations were left out and the error associated to the reconstruction of left-out simulations from the POD basis was evaluated.

Figure 3(a) shows the average normalized root mean squared errors (NRMSE) across all the available thermo-chemical variables, for the reconstruction of one particular left-out simulation. This corresponds to a leave-one-out (LOO) analysis. In this case, the total number of possible design of experiments (DoE) was equal to the number of available simulations, making it possible to visualise the NMRSE in the design space. Figure 3(a) allows to identify the design points impacting most the POD reconstruction error. It can be observed that very few design points had a considerable impact on the quality of the POD reconstruction.

When $k > 1$, the total number of possible DoE is given by: $m!/[(m-k)!\ k!]$, where m is the total number of available simulations and k is the number of simulations to leave out each time. In the present case, $m = 45$ while k ranges between 1 and 5, thus leading to a very large number of combinations (exceeding 1 million), for $k = 5$. Therefore, only a random subset of all the possible combinations was considered. For a given value of



(a)



(b)

Fig. 4. (a) Leave-one-out prediction errors, visualized in the input parameter space. The reported figures are the average normalized root mean squared errors (NRMSE) across all variables, for the prediction of one particular left-out simulation. The sizes of the circles are proportional to the error. Different colours are used for the 3 different values of the air injector size. (b) Average NRMSE for the prediction of an increasing number of left-out simulations. Vertical bars represent the standard deviation of the error associated to different combinations of k left out simulations.

k, the leave-k-out errors were estimated from random subsets of different sizes. Based on a sensitivity study, the size of the subset was chosen to be 250, as the leave-k-out errors were converging for this value. The analysis was carried out for an increasing value of k, as reported in Fig. 3(b), where the average NRMSE and its standard deviation for the reconstruction of the test data are reported. Two observations can be made. First, the average reconstruction error increases when more simulations are left out of the training data, as expected, and converges at a value of roughly 1%, indicating the ability of the POD basis to reconstruct the test data. Second, the standard deviation of the reconstruction error decreases when k increased, indicating that for high values of k the ROM is more sensitive to the size of the training data than to the location of the training simulations in the input parameter space.

4.2. Prediction of new data

The leave-k-errors for the reconstruction of the left-out data can be used to identify the most relevant simulations for the definition of a reduced basis, as shown in the previous section. However, in the context of predictive ROMs, it is more robust to base the leave-k-out approach on the prediction of the left-out data in order to assess how the developed ROM generalizes to new data [14]. Thus, this section presents the leave-k-out errors relative to the prediction of the left-out data by building a ROM from the included (not left out) simulations. Figure 4(a) shows the NRMSE

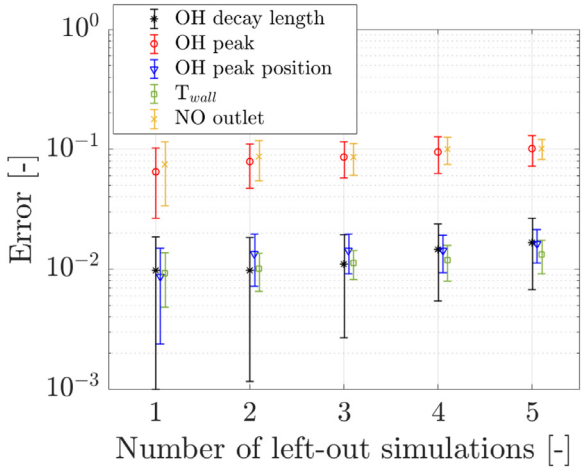


Fig. 5. Leave-k-out relative errors for the prediction of scalar quantities such as wall temperature, OH decay length, value and location of the Y_{OH} peak and NO emissions. Vertical bars represent standard deviations of the error associated to different combinations of k left out simulations.

associated to the prediction of each particular simulation (spatial fields of temperature and main chemical species), when that particular simulation was left out. Although relatively high prediction errors were observed for simulations that had a low influence on the POD basis, some of the highest LOO errors were observed for the prediction of the simulations that had the highest influences on the POD basis as well. Thus, the evaluation of these influences can be taken into consideration as a fast preliminary method to assess the quality of the training data and detect the regions of the input space were more observations are needed. Figure 4(b) shows the average NRMSE for the prediction of an increasing number of left-out simulations, similarly to what was done in Fig. 3. As the LOO prediction errors of Fig. 4(a) indicated the most influential simulations that should always be included in the training set, the leave-k-out errors of Fig. 4(b) were estimated taking this into account. Thus, only the simulations whose influence was $< 15\%$ w.r.t. the most important simulation were taken into consideration as possible test data. Predictably, the prediction errors were greater for higher values of left-out simulations, k. Interestingly, as observed for the reconstruction errors, the standard deviation of the mean prediction NRMSE decreased when k was increased. In the context of stationary systems, it is of major interest to look at quantities that can immediately be compared to sensory data rather than at the full spatial fields, and to predict quantities such as OH-decay and outlet composition. Therefore, leave-k-out errors for the prediction of scalar and integral quantities such as wall temperature, OH decay length and exhaust gas composition are reported in Fig. 5. The OH decay length was estimated as the distance from the inlet (on a vertical axis) at which the OH mass fraction decreased to less than 5% of its maximum value. As such, it can be considered representative of the flame length. The wall temperature was measured at the following axial coordinates z (in mm): 100, 200, 300, 400, 500, 600. Figure 5 shows that the prediction error for the wall temperatures, OH decay length and OH peak location slightly increases when increasing k from $k = 1$ to $k = 5$. Neverthless, the average NRMSE never exceeds 2%, which is remarkable. Higher NRMSE were obtained for the prediciton of the OH peak value, around 10%, but this can be considered acceptable considering the lower concentrations and more localised distribution of OH compared to other scalars. Similarly to Fig. 4, the standard deviations in Fig. 5 decrease for higher values of k. Low standard deviations for the prediction errors are a preferable characteristic of a ROM, to guarantee a lower upper bound for the prediction error. The fact that the developed digital twin can provide access to quantities difficult to measure with physical sensors is a very interesting feature of the approach and opens a number of opportunities for the soft-sensing [15] and control of combustion technologies using models, in this case ROMs.

5. ROM developed from the training data-set determined by leave-k-out analysis

A ROM was developed based on the simulations determined by the leave-k-out analysis of the previous Section. The errors of Fig. 5 were considered relatively low even for $k = 4$, suggesting that the use of a training set of size $m - 4$, i.e., 41, could lead to satisfactory performances as well, especially for the prediction of the scalar and integral

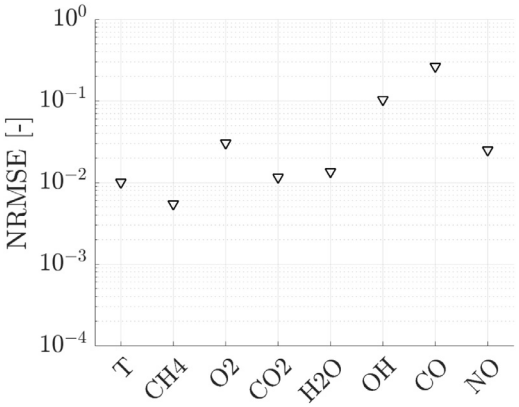


Fig. 6. NRMSE for the prediction of the test data by a ROM based on POD and Kriging.

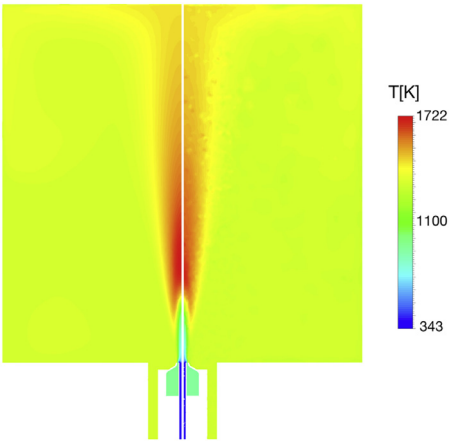


Fig. 7. (*left*) True temperature field from CFD simulation for air injector length of 16 mm, H_2 ratio of 0.60 and equivalence ratio equal to 0.93. (*right*) Predicted temperature field for the same operating conditions.

quantities. Thus, a value of $k = 4$ was chosen and all simulations except the four-dimensional set characterised by the minimum leave-k-out error (simulations 1, 22, 28 and 39 shown in the Supplementary material) were employed as training dataset to find both the reduced POD basis, and train a Kriging model for the prediction of the POD scores. The left-out simulations were used as test data, to assess the ROM's predictive capabilities.

Figure 6 reports the overall NRMSE for all the variables for the prediction of the test data. The fields of temperature and the main chemical species mass fractions and pollutants were predicted with an error below 10%, whereas higher prediction errors were obtained for CO and OH. This was expected, considering that CO and OH display a much more localised distribution with respect to other scalar, thus representing a more challenging target for the ROM.

Figures 7 and 8 compare the true temperature and OH field, respectively, to the ROM predictions, for different unexplored operating conditions. It can be observed how the ROM is able to accurately capture their distribution within the furnace, providing a solution which closely matches the CFD one, with no evident difference. Table 3 reports the errors for the prediction of different scalar quantities such as OH decay length, position and value of the OH peak and exhaust gas composition, for the four left-out simulations. Errors on the wall temperatures, OH decay length and OH mass fraction peak location are remarkably low, below 5% for all cases, w.r.t. the true values for the left-out simulations. The prediction of the OH mass fraction peak location values shows higher prediction errors; nevertheless the error never exceeds 10%, w.r.t. the true values.

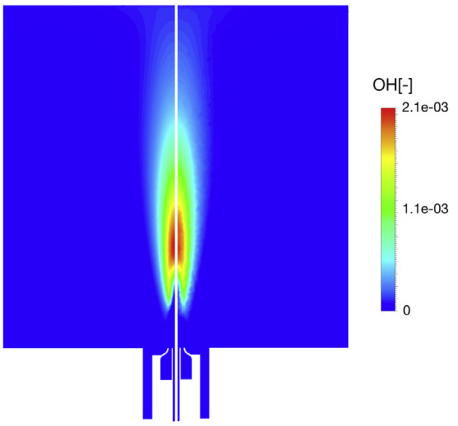


Fig. 8. (*left*) True OH field from CFD simulation for air injector length of 25 mm, H_2 ratio of 0.65 and equivalence ratio equal to 0.91. (*right*) Predicted OH field for the same operating conditions.

Table 3
Digital twin's prediction errors for different scalar quantities of the furnace such as wall temperature, OH decay length, position of the peak of Y_{OH}, value of the peak of Y_{OH}, furnace outlet mass fractions of H_2O, CO_2, CO and NO.

(Error on)	Sim. 1	Sim. 2	Sim. 3	Sim. 4
T_{wall}	0%	3%	1%	0%
OH DECAY LENGTH	9%	0%	1%	0%
POS. OF Y_{OH} PEAK	2%	0%	0%	0%
VALUE OF Y_{OH} PEAK	9%	5%	7%	1%
Y_{H_2O} OUTLET	1%	3%	4%	1%
Y_{CO_2} OUTLET	5%	1%	5%	1%
Y_{CO} OUTLET	3%	0%	0%	1%
Y_{NO} OUTLET	5%	3%	1%	1%

6. Conclusions

In the present work, the first-of-its-kind digital twin for a furnace operating in flameless conditions was developed and validated. A reduced-order model (ROM) based on the combination of Proper Orthogonal Decomposition (POD) and Kriging was developed for the prediction of three-dimensional spatial fields of temperature and chemical species (major, minor and pollutants), as a function of three input parameters, the fuel composition (a mixture of methane and hydrogen from pure methane to pure hydrogen), the equivalence ratio and the air injector diameter. Fourty-five three-dimension CFD simulations were carried out to generate samples for the ROM. Numerical simulations were also validated against available experimental data on the furnace, for different fuel mixture compositions. During the construction of the ROM, POD was used for data compression, thus to represent the original data with a reduced number of features, the POD scores. Kriging was used to find a response surface for these scores at unexplored operating conditions. The influence of each simulation on the reduced basis found by POD was estimated, so to identify the most important simulations to retain as training data for the ROM. The influence of the number of training simulations used for the development of the ROM was also assessed. A leave-*k*-out analysis was carried out to determine how many and which simulations were needed for the training of the ROM, and estimate how the developed ROM would generalize to new data. Results showed that the developed ROM could predict the fields of temperature and CO_2, O_2, H_2O, CH_4, NO mass fractions, at unexplored operating conditions, reliably with an overall prediction error lower than 10%. Higher errors ($< 20\%$) were observed for the prediction of minor species, e.g. CO, and radicals, e.g. OH radicals. In addition, the prediction of scalar quantities at specific locations was characterised by even lower reconstruction errors, below 5%. The latter included wall temperatures, OH decay length, OH peak value and location, as well as exhaust gas composition and temperature, proving the potential of the method for soft sensing and real-time predictions of system change when changing operating conditions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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. The support of the Fondation Wiener-Anspach is acknowledged. The second author wish to thank the Fonds de la Recherche Scientifique FNRS Belgium for financing his research.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.proci.2020.06.045.

References

[1] J. Wünning, J. Wünning, *Prog. Energy Combust. Sci.* 23 (1997) 81–94.
[2] A. Cavaliere, M. de Joannon, *Prog. Energy Combust. Sci.* 30 (2004) 329–366.
[3] E.H. Glaessgen, D.T. Branch, D.S. Stargel, The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles (2019) 1–14.
[4] T.H. Uhlemann, C. Schock, C. Lehmann, S. Freiberger, R. Steinhilper, *Procedia Manuf.* 9 (2017) 113–120, doi:10.1016/j.promfg.2017.04.043.
[5] G. Aversano, A. Bellemans, Z. Li, A. Coussement, O. Gicquel, A. Parente, *Comput. Chem. Eng.* 121 (2019) 422–441, doi:10.1016/j.compchemeng.2018.09.022.
[6] P.G. Constantine, E. Dow, Q. Wang, *SIAM J. Sci. Comput.* 36 (4) (2014) 1500–1524.
[7] S.N. Lophaven, J. Søndergaard, H.B. Nielsen, Kriging Toolbox (2002) 1–28.
[8] M. Seeger, 14, 2004, doi:10.1142/S0129065704001899.
[9] M. Ferrarotti, M. Fürst, E. Cresci, W. de Paepe, A. Parente, *Energy Fuels* 32 (10) (2018) 10228–10241, doi:10.1021/acs.energyfuels.8b01064.
[10] J. Chomiak, *Combustion: A Study in Theory, Fact and Application*, Abacus Press, Philadelphia, PA, 1987.
[11] T.F. Smith, Z.F. Shen, J.N. Friedman, *J. Heat Transf.* 104 (4) (1982) 602, doi:10.1115/1.3245174.
[12] A.A. Konnov, G. Colson, J.D.E. Ruyck, *Combust. Flame* 121 (2000) 548–550.
[13] M. Ferrarotti, Z. Li, A. Parente, *Proc. Combust. Inst.* 37 (4) (2019) 4531–4538.
[14] G.C. Cawley, N.L.C. Talbot, *J. Mach. Learn. Res.* 11 (2010) 2079–2107.
[15] B. Gabrys, S. Strandt, *Comput. Chem. Eng.* (2008).