Results in Engineering 10 (2021) 100223

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.editorialmanager.com/rineng/Default.aspx

Combination of polynomial chaos and Kriging for reduced-order model of reacting flow applications

Gianmarco Aversano [a,b,*], Giuseppe D'Alessio [a,b,c], Axel Coussement [a,b], Francesco Contino [b,d], Alessandro Parente [a,b]

[a] Université Libre de Bruxelles, Aero-Thermo-Mechanics Departement, Avenue F.D. Roosevelt 51, CP 165/41, 1050, Brussels, Belgium
[b] Université Libre de Bruxelles and Vrije Universiteit Brussel, Combustion and Robust Optimization Group (BURN), Brussels, Belgium
[c] CRECK Modeling Lab, Department of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20131, Milano, Italy
[d] Université Catholique de Louvain, Institute of Mechanics, Materials, and Civil Engineering, Louvain-la-Neuve, Belgium

ARTICLE INFO

Keywords:
PCA
Surrogate models
Polynomial chaos
Kriging

ABSTRACT

The combination of Proper Orthogonal Decomposition (POD) with Kriging has been shown to be a reliable choice for the development of Reduced-Order Models (ROMs) for the prediction of combustion data at unexplored operating conditions. In this study, POD is combined with Polynomial Chaos Expansion (PCE), with a combination of PCE and Kriging (PC-Kriging) and with Artificial Neural Networks (ANN) for the development of a ROM that can predict 2D combustion data for unexplored operating conditions. The choice of Non-negative Matrix Factorization (NMF) instead of POD as compression method is also investigated. This method is chosen because it can intrinsically guarantee the non-violation of physical constraints such as positivity of chemical species mass fractions, although POD's data reconstruction errors are lower. The performances of the POD and NMF in combination with the proposed supervised methods are compared, with prediction normalized root mean squared errors (NRMSE) being less than 10% for spatial fields of temperature, CH_4 and O_2 for all approaches.

1. Introduction

High-fidelity expensive computer simulation are necessary for the description of many complex physical systems. This is the case with computational fluid dynamics (CFD) and combustion, where design optimization studies are limited by the computational cost of running a large number of simulations. Due to the non-linearity of CFD and combustion systems, changing operating conditions can lead to drastic changes in the state of the considered system. Thus, complete knowledge about the investigated system's behavior for a full range of operating conditions can only be achieved by running these expensive simulations several times with different inputs, until enough observations of the system's state are obtained.

Surrogate modeling [1] reduces the computational cost of these studies by evaluating only a small subset of the proposed simulations and fitting a computationally cheap model to them. This model can then be quickly evaluated to further guide designs instead of simply running additional CFD simulations, as in Refs. [2,3].

In a previous study [4], the combination of an unsupervised method for data compression such as Principal Component Analysis (PCA) or Proper Orthogonal Decomposition (POD), with a supervised technique such as Kriging for the development of Reduced-Order Models (ROMs) was shown to be a reliable choice for the prediction of combustion data, namely spatial profiles of temperature and chemical species mass fractions, for unseen operating conditions, e.g. inlet velocity and fuel composition of a 2D laminar methane flame. POD could compress the size of the high-dimensional data whose prediction is of interest, leading to a high reduction in the number of surrogate models (SMs) to be trained. In a following study [5], this approach was employed for the construction of a digital twin of a combustion furnace, that could predict the furnace's state based on quantities that could be measured by sensors, such as inlet temperature and equivalence ratio. The present study investigates the use of alternative techniques for either data compression or surrogate modeling, or both.

Because of the reconstruction errors that are involved in low-rank approximations such as the one provided by POD, important physical laws such as positivity of the chemical species mass fractions might be violated. The approach of finding reduced rank non-negative factors to approximate a given non-negative data matrix thus becomes a natural choice. Even though Constrained PCA (CPCA) has shown to be capable to

* Corresponding author. Université Libre de Bruxelles, Aero-Thermo-Mechanics Departement, Avenue F.D. Roosevelt 51, CP 165/41, 1050, Brussels, Belgium
E-mail address: Gianmarco.Aversano1990@gmail.com (G. Aversano).

https://doi.org/10.1016/j.rineng.2021.100223
Received 30 December 2020; Received in revised form 9 March 2021; Accepted 6 April 2021

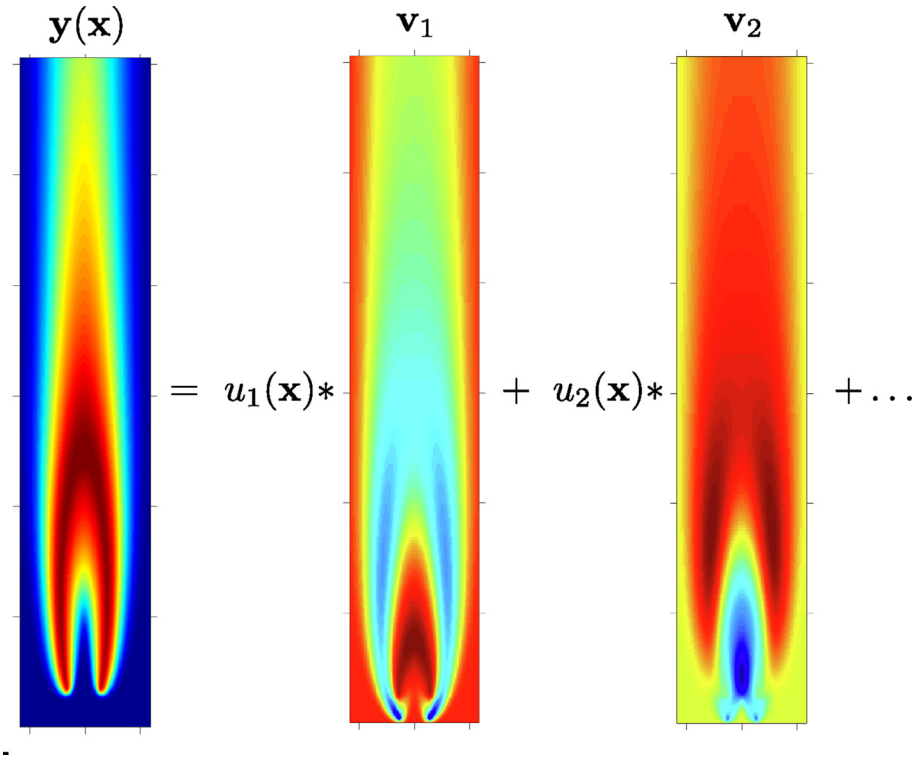

Fig. 1. Illustrative example of low-rank approximation of a spatial field **y** observed for a certain operating condition **x**, represented by a set of coordinates (the

alleviate or solve this problem [4,6], in the present work the choice of Non-negative Matrix Factorization (NMF) is investigated as a data size reduction instead of POD. Thus, the ability of NMF to reconstruct training and test data is assessed, in comparison to POD, as well as the predictive capabilities of a ROM that combines NMF instead of POD with a supervised technique, such as PC-Kriging.

In the present study, also the effects of combining Kriging with Polynomial Chaos Expansion (PCE) for the supervised part of the approach are investigated. As shown in Ref. [7], PCE can act as Kriging trend function, increasing the model's predictive capabilities especially in absence of data, while Kriging helps to interpolate the training data. The combination of the two methods, here referred to as PC-Kriging or PCK, have already produced encouraging results as shown in Refs. [7,8]. However, in the present study the approach is applied to reacting systems, which are multi-physics, multi-scale problems, so to pave the way for the use of this approach to real industrial applications. Besides, no compression method was used before training a SM based on PC-Kriging in Refs. [7,8].

A compression method grants the possibility to encode high-dimensional vectors (such as a spatial field) into a few scalars, as explained in Refs. [3,4,9] in the context of combustion data, thus reducing the number of predictive models to train. However, although the size of the reduced data can be even 10 times smaller than the original data size, the reduced size can still be large if strong non-linearities are inherent to the considered problem, meaning that a large number of predictive models still needs to be trained. Thus, in the present work, the combination of a compression method such as POD or NMF with Artificial Neural Networks (ANNs) [1,10,11] is also investigated as one advantage of ANN is that it makes it possible to train only one net that is able to predict the values of all the targets, namely the POD or NMF scores, for new input parameters. Whereas with the other techniques, it is necessary to train one model per score.

The discussed approaches are tested on 2 designs of experiment (DoE) made out of a total of 64 simulations produced by OpenFOAM, spanning the two input parameters in the range $\frac{24}{89}$ cm/s and $\frac{40}{100}$ % for the inlet velocity and inlet molar fraction of CH_4, respectively. A DoE with 24 training simulations (DoE-A) was firstly used to compare the performances of POD + Kriging, POD + PCE and POD + PC-Kriging. The small number of training simulations was chosen as results from Ref. [7] indicate that the PC-Kriging approach works better for small experimental designs. Then, the approaches are also compared on a bigger experimental design, namely DoE-B, which included 55 training simulations. In both cases, the training samples are randomly chosen. This part of the SM training was not optimized as the main objective of the present study was to compare the methodologies, more than effectively train a ROM. ROMs with low prediction errors are nonetheless constructed in the present work and a leave-*k*-out analysis is also carried out [12].

The paper is organized as follows: the methodology employed for the ROM development is explained briefly in Section 2, as well as the chosen unsupervised and supervised techniques, such as POD, NMF, PCE, Kriging, PC-Kriging and ANN. Section 3 introduces the combustion data-set chosen to validate the developed ROMs and shows the results obtained from data compression performed by means of POD and NMF. Section 4 assesses the predictive capabilities of PCE, Kriging, PC-Kriging and ANN when combined with POD, while Section 5 discusses the performances of an NMF-based ROM using PC-Kriging. Finally, conclusions are drawn in Section 6.

2. Theoretical background

2.1. Methodology

In the present work, a data matrix containing m output of a certain computer model or CFD simulation, indicated by $\mathcal{F}$, will be represented by a matrix $\mathbf{Y}$ of size $(m \times n)$, where n is the size of the output of the considered model. The m combinations of values for the d input parameters to the simulations that produced $\mathbf{Y}$ are collected in the matrix $\mathbf{X}$ of size $(m \times d)$. One particular simulation or output of the considered computer code for one combination of the input parameters $\mathbf{x}$ is indicated

by $\mathbf{y}(\mathbf{x}) = \mathscr{F}(\mathbf{x})$.

In the present work, a low-rank approximation of $\mathbf{Y}$ is constructed, so that each particular simulation output $\mathbf{y}(\mathbf{x})$ can be expressed as a weighted sum of basis functions, as shown in Fig. 1.

By doing so, it is subsequently possible to train SM for unexplored combinations of the input parameter only for the coefficients of the basis functions [4]. Two different methods to perform low-rank approximation of a given data matrix and thus find the basis functions and their coefficients are explained in the next Section. Each method requires the data matrix to be pre-processed [13,14]. In the following Section, it is implied that the data matrix is appropriately centered and scaled for each method.

2.2. Low-rank approximations

Data compression or low-rank approximation is a lower dimensional representation of a higher-dimensional data-set $\mathbf{Y}$ of size $(m \times n)$. This matrix is compressed to or represented by a lower dimensional matrix of size $(m \times k)$, with $k < n$ and k being called the approximation order.

2.2.1. Proper Orthogonal Decomposition

The POD problem can be stated as follows: given a matrix $\mathbf{Y}$ of size $(m \times n)$, find $\mathbf{Z}$ of size $(m \times k)$ and $\mathbf{A}$ of size $(n \times k)$ with $k < n$ such that the functional $f(\mathbf{Z}, \mathbf{A}) = \frac{1}{2}\left|\left|\mathbf{Y} - \mathbf{Z}\mathbf{A}^T\right|\right|^2$ 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 and eigenvalues of the matrix $\mathbf{C} = \frac{1}{m-1}\mathbf{Y}^T\mathbf{Y}$ [15]. The eigenvectors of $\mathbf{C}$ are the POD modes and the associated eigenvalues represent their relevance for the low-rank approximation of $\mathbf{Y}$. The POD modes are thus found all at once, and by ordering them in descending order according to their corresponding eigenvalue and retaining only a subset $k < n$ of them, a low-rank approximation of $\mathbf{Y}$ is possible as follows $\mathbf{Y} \approx \mathbf{Z}\mathbf{A}^T = \mathbf{Y}\mathbf{A}\mathbf{A}^T$. The columns of $\mathbf{A}$ of size $(n \times k)$ are the POD modes and $\mathbf{Z}$ of size $(m \times k)$ is the matrix of POD coefficients. Each row of $\mathbf{Z}$ consists of the k coefficients for the retained k POD modes so that one particular simulation, or row of $\mathbf{Y}$, can be expressed as $\mathbf{y}(\mathbf{x}_j) = \sum_{i=1}^{k} \mathbf{a}_i z_i(\mathbf{x}_j)$. As the solution of the POD problem only leads to the evaluation of $\mathbf{A}$, the matrix $\mathbf{Z}$ can be computed only after $\mathbf{A}$ is known, as follows: $\mathbf{Z} = \mathbf{Y}\mathbf{A}$. This holds for new, unseen data as well: $\mathbf{Z}' = \mathbf{Y}'\mathbf{A}$.

2.2.2. Non-negative matrix factorization

The NMF problem can be stated as follows: given a non-negative matrix $\mathbf{Y}$ of size $(m \times n)$, find non-negative matrix factors $\mathbf{U}$ of size $(m \times k)$ and $\mathbf{V}$ of size $(n \times k)$, with $k < n$, such that the functional $f(\mathbf{U}, \mathbf{V}) = \frac{1}{2}\left|\left|\mathbf{Y} - \mathbf{U}\mathbf{V}^T\right|\right|^2$ is minimized, subject to $u_{ij}, v_{ij} \geq 0 \quad \forall i,j$ [16]. Differently from POD, the non-negative factors or NMF modes are found for a given approximation order k. For a different value of k, the NMF problem needs to be solved again. As for POD, once the NMF problem is solved and thus the NMF modes are found, one particular simulation, or row of $\mathbf{Y}$, can be expressed as $\mathbf{y}(\mathbf{x}_j) = \sum_{i=1}^{k} \mathbf{v}_i u_i(\mathbf{x}_j)$. Differently from POD, both matrices $\mathbf{U}$ and $\mathbf{V}$ are determined by solving the NMF problem. The determination of the NMF scores $\mathbf{U}'$ for new, unseen data $\mathbf{Y}'$ is possible only by solving the least-squares error minimization problem: $\mathbf{U}' = \underset{\mathbf{U}'}{\operatorname{argmin}} \frac{1}{2}\left|\left|\mathbf{Y}' - \mathbf{U}'\mathbf{V}^T\right|\right|^2$, whose solution is taken as $\mathbf{U}' = \mathbf{Y}'(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$.

2.3. Surrogate modeling

After a low-rank approximation of the data matrix is found, by either POD or NMF, the lower-dimensional data, namely the POD or NMF scores, need to be predicted for new values of the input parameters to the computational model, as explained in Section 2.1. To this purpose, the following supervised methods will be used in the present work, so that a response surface for the compressed data is found.

The methods for surrogate modeling presented next require the matrix of the input parameter values $\mathbf{X}$ to be pre-processed, namely centered and scaled, before the surrogate models can be trained, thus this will be implied. Besides, in the following subsections, the surrogate models are described to infer a mapping from a general input space, indicated by $X \in \mathbb{R}^d$, to a general output space, indicated by $Y \in \mathbb{R}$, even though this paper presents a study that makes use of surrogate models to map from the input parameter space to the POD-score space ($\mathbf{x} \in \mathbb{R}^d \rightarrow z \in \mathbb{R}$).

2.3.1. Polynomial chaos expansion

The theoretical background for Polynomial Chaos Expansion (PCE) is reported from Refs. [7,8]. Consider a system whose behavior is represented by a computational model $\mathscr{M}$ which maps the d-dimensional input parameter space to the 1-dimensional output space $\mathscr{M} : \mathbf{x} \in \mathbb{R}^d \rightarrow y \in \mathbb{R}$. In Section 2.1, the computational model was indicated by $\mathscr{F}$ and its output was multi-dimensional. For this reason, in this section the symbol $\mathscr{M}$ will be employed. In the present work, the components of the input vector $\mathbf{x} = x_1, \ldots, x_d$ are assumed independent. The case of dependent input variables can easily be addressed as explained in Ref. [17]. In the present work, the computational model $\mathscr{M}$ is a deterministic mapping from the input to the output space, i.e. repeated evaluations with the same input values lead to the same output value. As explained in Ref. [7], the computational model $\mathscr{M}$ can be approximated by a finite, truncated set of polynomials:

$$y(\mathbf{x}) = \mathscr{M}(\mathbf{x}) \approx \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} a_{\boldsymbol{\alpha}} \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\mathbf{x}), \tag{1}$$

where $a_{\boldsymbol{\alpha}}$ are the expansion coefficients of the multivariate polynomials $\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\mathbf{x})$ and $\boldsymbol{\alpha}$ is the multi-index. Because of the statistical independence of the input variables, the multivariate polynomials are evaluated as the product of uni-variate polynomials $\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\mathbf{x}) = \prod_{i=1}^{d} \boldsymbol{\psi}_{\alpha_i}^{(i)}(x_i)$, where $\boldsymbol{\psi}_{\alpha_i}^{(i)}$ is the polynomial of degree α_i for the i-th variable. The total degree of the (multivariate) polynomials is defined by $|\boldsymbol{\alpha}| = \sum_{i=1}^{d} \alpha_i$. The total number of (multivariate) polynomials depends on the adopted truncation scheme. In the present work, the adopted truncation scheme is the *hyperbolic truncation set* [7]: $\boldsymbol{\alpha} \in \mathbb{N} : \left|\left|\boldsymbol{\alpha}\right|\right|_q \leq N$, where N is the total order of the polynomials and the norm $\left|\left|\cdot\right|\right|_q$ is defined as $\left|\left|\boldsymbol{\alpha}\right|\right|_q = \left(\sum_{i=1}^{d} \alpha_i^q\right)^{1/q}$.

The maximum number of terms in the polynomial basis, attainable for $q = 1$, is given in Eq. (2):

$$p + 1 = \frac{(d+N)!}{d!\,N!}. \tag{2}$$

The expansion coefficients $a_{\boldsymbol{\alpha}}$ can be estimated by a least-square minimization method, as explained in Ref. [7].

2.3.2. Kriging

Kriging is an interpolation method in which every realization $y(\mathbf{x})$ is expressed as a combination of a trend function and a residual [18]:

$$y(\mathbf{x}) = \mu(\mathbf{x}) + s(\mathbf{x}) = \sum_{i=0}^{p} \beta_i f_i(\mathbf{x}) + s(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + s(\mathbf{x}) \tag{3}$$

The trend function $\mu(\mathbf{x})$ is a low-order polynomial regression and provides a global model in the input space. The term $s(\mathbf{x})$ creates a localized deviation weighting the points in the training set that are closer to the target point $\mathbf{x}$. The trend function $\mu(\mathbf{x})$ is expressed as a weighted linear combination of $p+1$ polynomials $\mathbf{f}(\mathbf{x}) = [f_0(\mathbf{x}), \ldots, f_p(\mathbf{x})]^T$ with the weights $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]^T$ determined by generalized least squares (GLS). The subscript p also indicates the degree of the polynomial. The residuals $s(\mathbf{x})$ are modeled by a Gaussian process with a kernel or

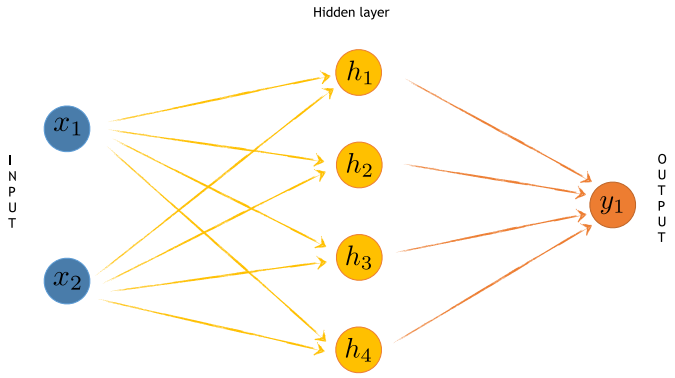

Fig. 2. Illustrative example of an Artificial Neural Network with 2-dimensional input layer, 4-dimensional hidden layer and 1-dimensional output layer. The value of h_i is given by $h_i = \sum_{j=1}^{d} f_l(x_j w_{ji}^{(l)} + b_j^{(l)})$, where d is the size of the input layer, $f_l()$ is the activation function for l-th layer, $w_{ji}^{(l)}$ is the weight from the j-th input to h_i (in layer l) and $b_j^{(l)}$ is the bias for the j-th input (in layer l).

correlation function that depends on a set of hyper-parameters **l** to be evaluated by Maximum Likelihood Estimation (MLE) [19,20].

In the definition of both the trend function and the residual, it is up to the designer to choose the polynomials $\mathbf{f}(\mathbf{x})$ and the correlation model or kernel. In this way, prior knowledge can be added into the problem.

2.3.3. PC-Kriging

As explained in Ref. [7], Kriging is able to interpolate local variations of the output of the computational model. In contrast, polynomial chaos expansions (PCE) are generally used for approximating global behaviors of computational models. The two techniques can be combined if PCE is used as trend function for the Kriging interpolation method. This approach is referred to as PC-Kriging and its formulation is as follows:

$$y(\mathbf{x}) = \mathscr{M}(\mathbf{x}) \approx \sum_{\boldsymbol{\alpha} \in \mathscr{A}} a_{\boldsymbol{\alpha}} \boldsymbol{\psi}_{\boldsymbol{\alpha}}(\mathbf{x}) + s(\mathbf{x}). \tag{4}$$

Building a PC-Kriging meta-model consists of determining the optimal set of polynomials for PCE first and then calibrating the Kriging hyper-parameters **l**.

2.3.4. Artificial Neural Networks

Artificial Neural Networks (ANN) [21,22] are a supervised method that have certain characteristics in common with biological neural networks and have been developed as generalization of mathematical models [23]. An ANN is characterized by its patterns of connections between neurons (architecture), its method on determining suitable values for the weights which its architecture is composed of (training or learning algorithm), its activation functions.

Fig. 2 is an illustrative example of an Artificial Neural Network with 2-dimensional input layer, 4-dimensional hidden layer and 1-dimensional output layer. The size of the output layer can even be greater than 1. In such a case, one ANN can be trained for multiple outputs. For instance, only one net can be trained to predict the values of the POD or NMF scores altogether. In the case of Kriging, PCE or PC-Kriging, this is not possible and the models have to be trained for each score separately.

3. Data-set description and data compression

The configuration of the simulated flame is described in Ref. [24] and in Ref. [25]. The computational domain starts from the exit of the nozzle and extends 122 mm further downstream. The radial direction is expanded to 42.88 mm. A 2D structured axi-symmetric mesh with around 25,600 cells is used and the nozzle radius is resolved with 12 cells. The laminarBuoyantSimpleSMOKE solver is applied. Gravity is turned on.

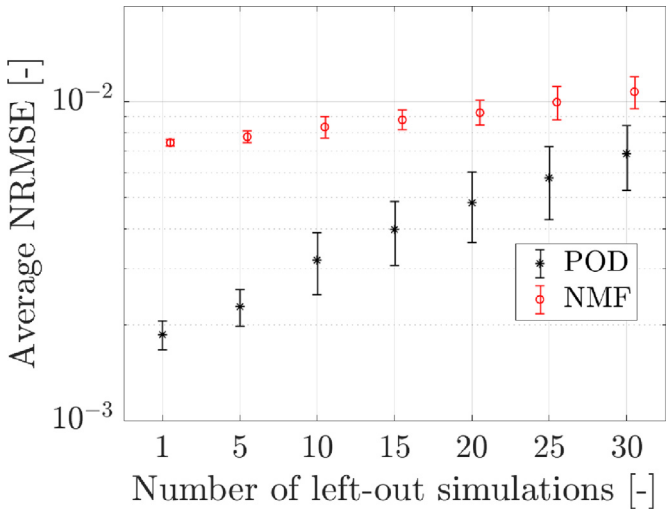

Fig. 3. Average NRMSE for the reconstruction of the test data by POD and NMF (both with a number of 30 modes) for an increasing number of left-out simulations.

The multi-component diffusion model is adopted to consider molecular diffusion. The GRI3.0 mechanism without NO_x (35 species and 219 reactions) is applied. For the velocity boundary condition, the profile provided from Ref. [24] and in Ref. [25] is used. The input parameters are two, namely the inlet velocity and the molar fraction of CH_4 in the inlet stream, which is a mixture of CH_4 and N_2. 64 samples are produced by OpenFOAM, spanning the two input parameters in the range $\frac{24}{55}$ cm/s and $\frac{40}{100}$ % for the inlet velocity and inlet molar fraction of CH_4, respectively. The approaches investigated in the present work will only be performed on the following spatial profiles: CH_4, CO, CO_2, H_2, H_2O, N_2, O_2, OH and temperature.

The ROM was trained using MATLAB on SUSE Linux Enterprise Server 12 SP1 with Intel(R) Xeon(R) CPU E5-2603 v3 @ 1.60 GHz, using a custom implementation and using the ooDACE toolbox [19].

3.1. Results from data compression

One of the advantages of POD is that the POD modes can all be estimated beforehand. Besides, the POD modes come with an eigenvalue associated to them representing the portion of the original data variance that they account for. This makes it possible to solve the POD problem once and then choose the number of modes to retain by looking at the cumulative data variance that they recovered. If NMF is adopted for the compression, instead, a different choice of the approximation order, namely the number of modes or non-negative factors to find, entails a new resolution of the compression problem.

A leave-k-out analysis was carried out in order to assess the influence of the number of training simulations on the POD and NMF solution. Leave-k-out analysis is generally a more robust way to assess how a model will generalize to unseen data. k is the number of simulations left out from the training data set employed to find the POD and NMF reduced basis. For each value of k, different sets of k simulations are left out and the average error (for that value of k) when reconstructing the left-out simulations from the POD and NMF basis is estimated. Fig. 3 reports the normalized root mean squared errors (NRMSE) for the reconstruction of the test data by POD and NMF with 30 modes for an increasing number of left-out simulations. In terms of reconstruction, POD is clearly superior to NMF. The gap between the two methods in average NRMSE for the test data reconstruction decreases for smaller training sizes.

Fig. 4 is a parity plot for the true, observed values for the OH mass fraction and the ones recovered by (a) NMF and (b) POD, for a case where

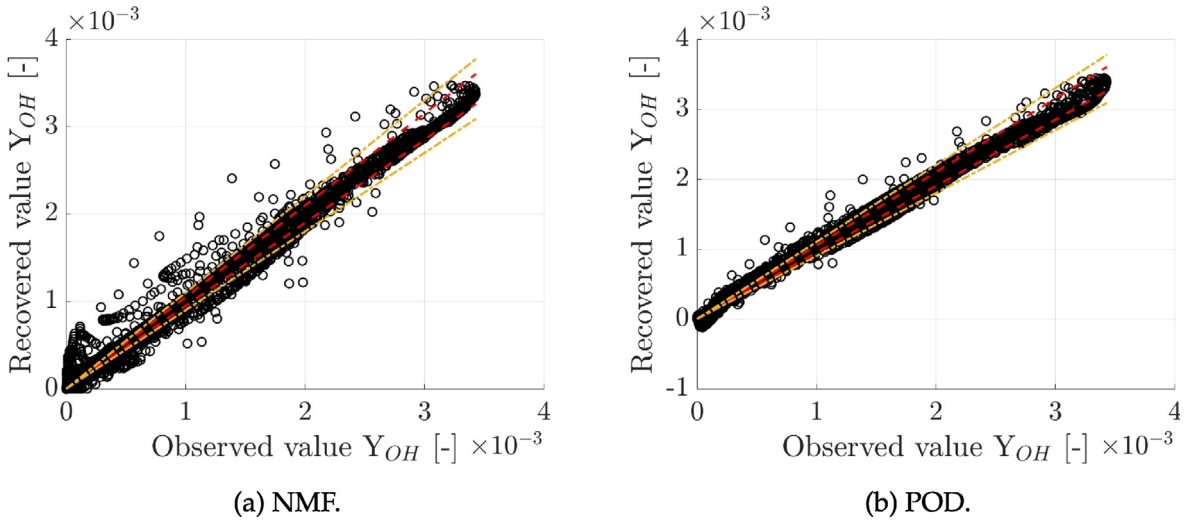

Fig. 4. Reconstruction of OH mass fraction. Although more accurate, POD recovered negative values which are physically impossible. The inner and out dashed lines represent the 5% and 10% error, respectively.

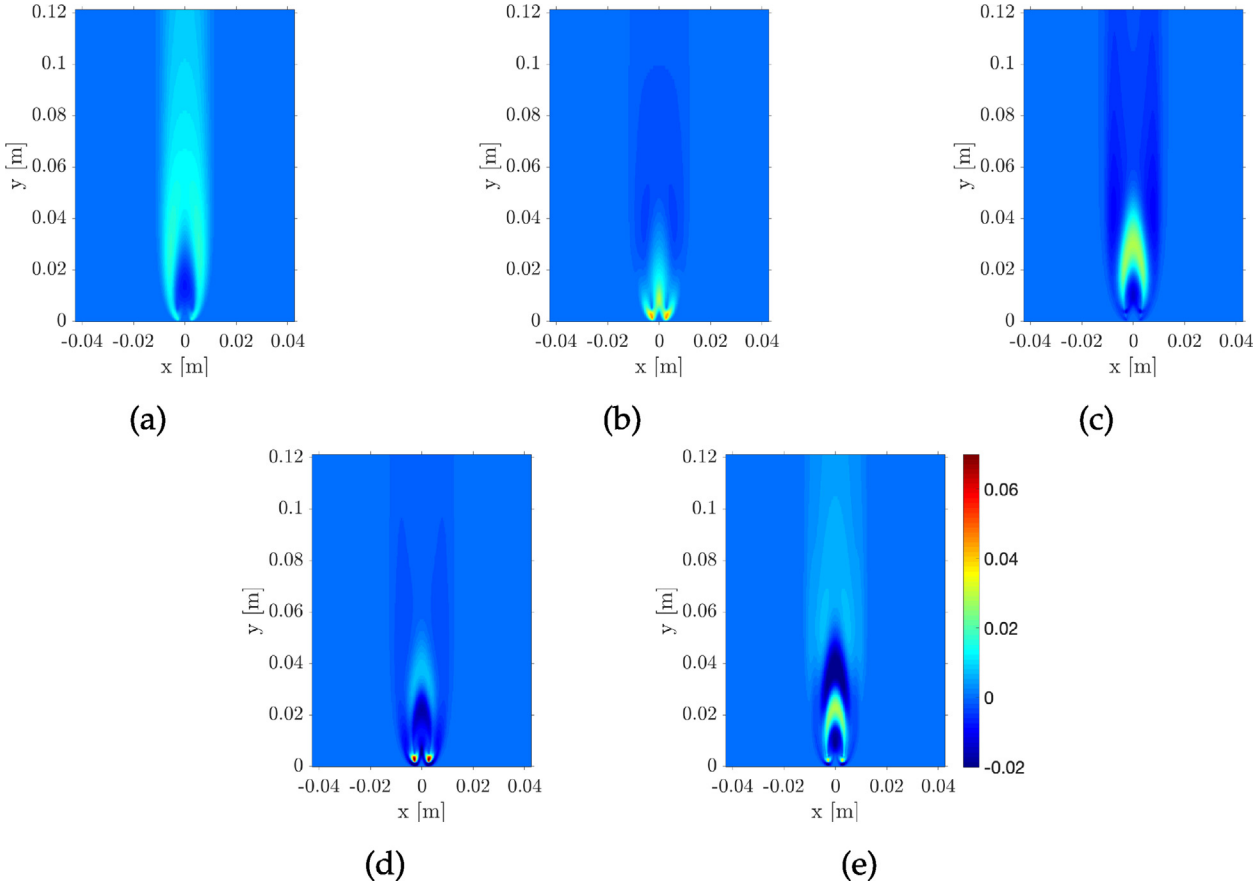

Fig. 5. Data-driven orthogonal basis functions for the temperature field found by POD with a number of 5 retained modes.

10 random simulations are left out. Although NMF's higher data reconstruction errors can be observed here as well, POD recovers negative values of the OH mass fractions, which is physically impossible, while NMF does not.

A physical interpretation of the data-driven basis functions found by POD and NMF can be accomplished by analyzing Figs. 5 and 6, which report the eigenflames found by the two methods when compressing the available simulations to a 5-dimensional reduced manifold. These eigenflames are different. POD's first mode (Fig. 5a) represents the main direction of variation. The second mode (Fig. 5b) is representative of the temperature high gradients due to the cold inlet fuel jet entering the flame zone. The third mode, reported in Fig. 5c, is representative of the reactive zone, where the maximum values of temperature are encountered. NMF's first mode (Fig. 6a) is representative of the temperature field for higher values of inlet velocities, while NMF's second mode is for lower values, with the following modes representing different features to be

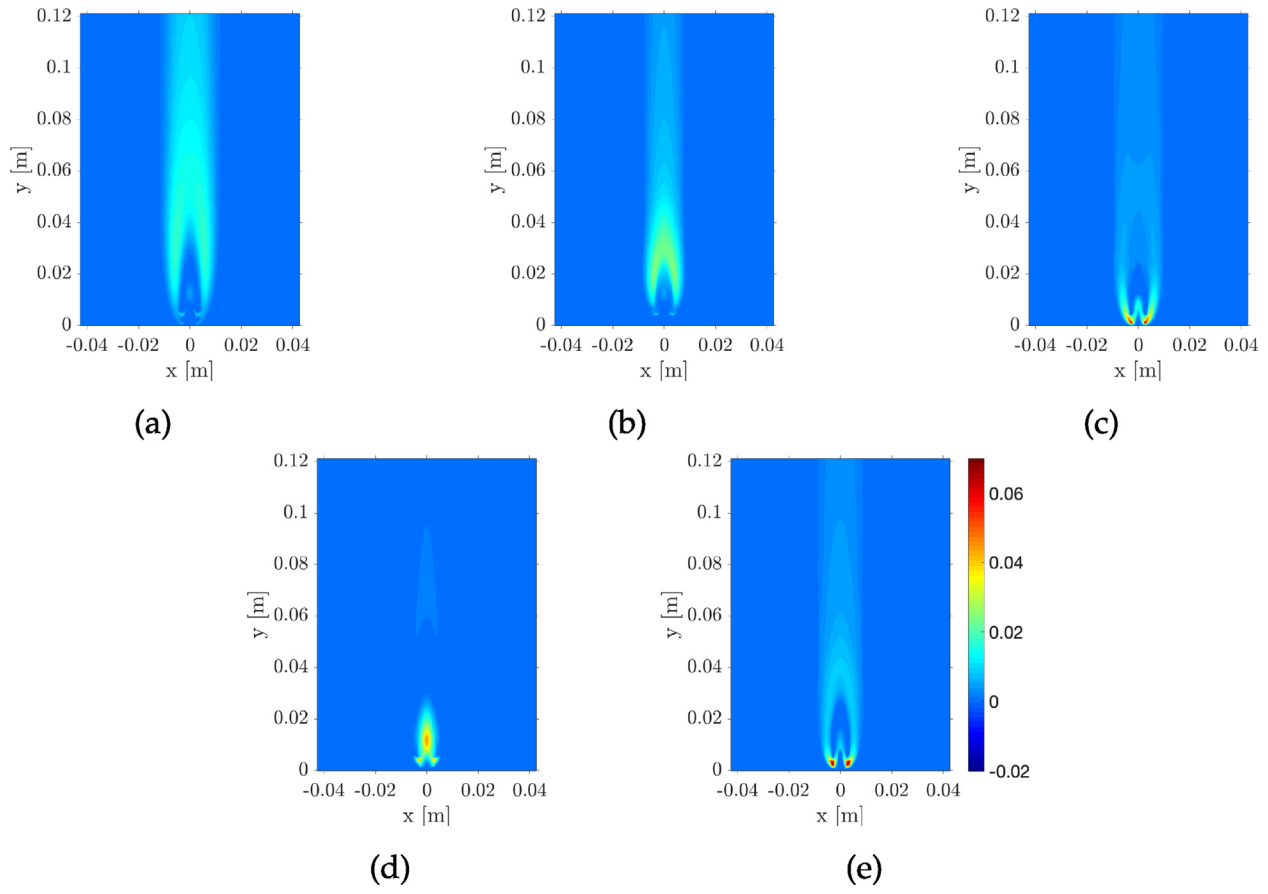

Fig. 6. Data-driven non-negative basis functions for the temperature field found by NMF with a number of 5 retained modes. Same color map as of Fig. 5.

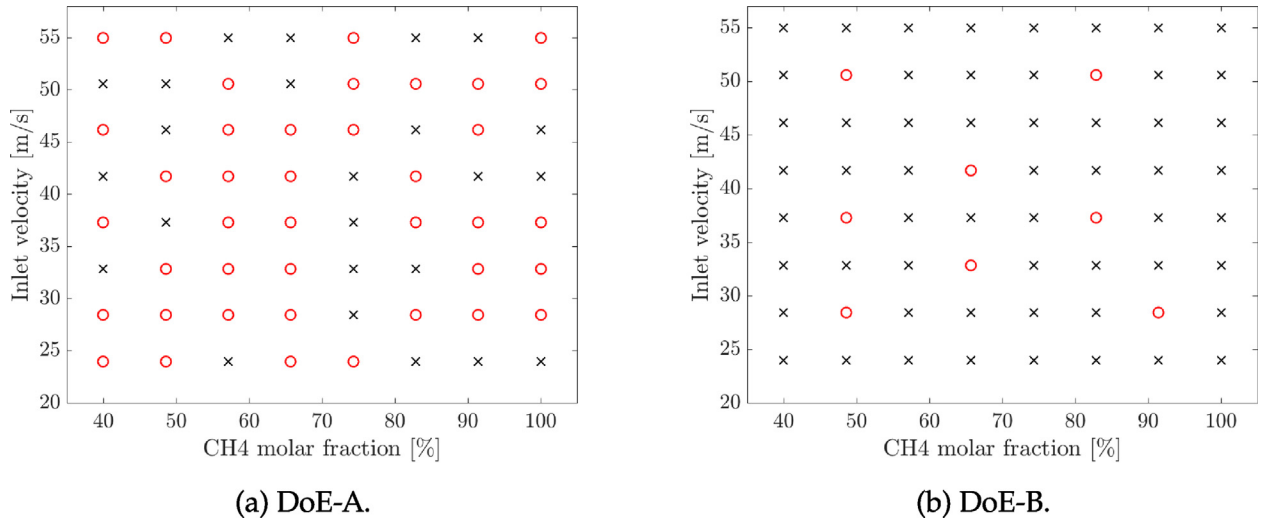

Fig. 7. Random design of experiment for the surrogate model. Black cross: training points; red circle: test points.

added. In both cases, the different flame positions, which depend more on the inlet temperature than on the fuel composition input parameter, play a major role on the shapes of the eigenflames, whereas the maximum values of the different spatial fields, which depend on the fuel composition mainly, are accounted for by the eigenflame coefficients or scores.

In conclusion, despite NMF's additive nature, POD's lower reconstruction errors indicate that this method is a better choice over NMF for the compression of the data of the present study.

3.2. Application of the POD-based ROM

The performance of the ROM developed by combining POD with PCE, Ordinary Kriging (OK) and PC-Kriging are investigated in the present work. Specifically, the effects of changing the values of the PCE parameters such as the total polynomial order N and degree of polynomial interaction q is investigated, as well as the effects of the training data-set's size. The leave-k-out strategy is employed for the analysis, but also two different reference DoEs (randomly chosen) of different sizes are

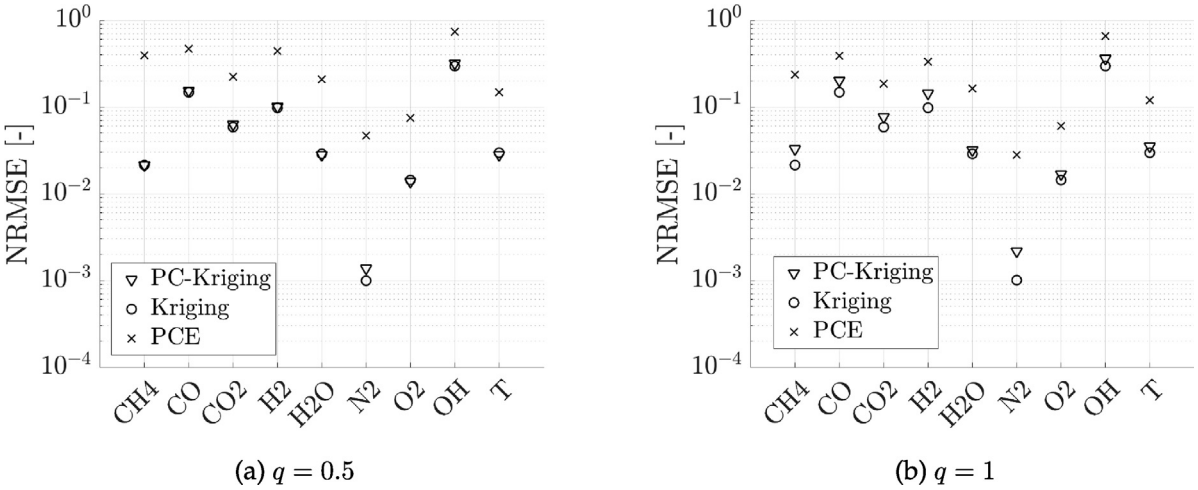

Fig. 8. NRMSE for the prediction of the test data of DoE-A with Matern52 as kernel for the Kriging, with $N = 2$ and with (a) $q = 0.5$ and (b) $q = 1$ for the Polynomial Chaos Expansion.

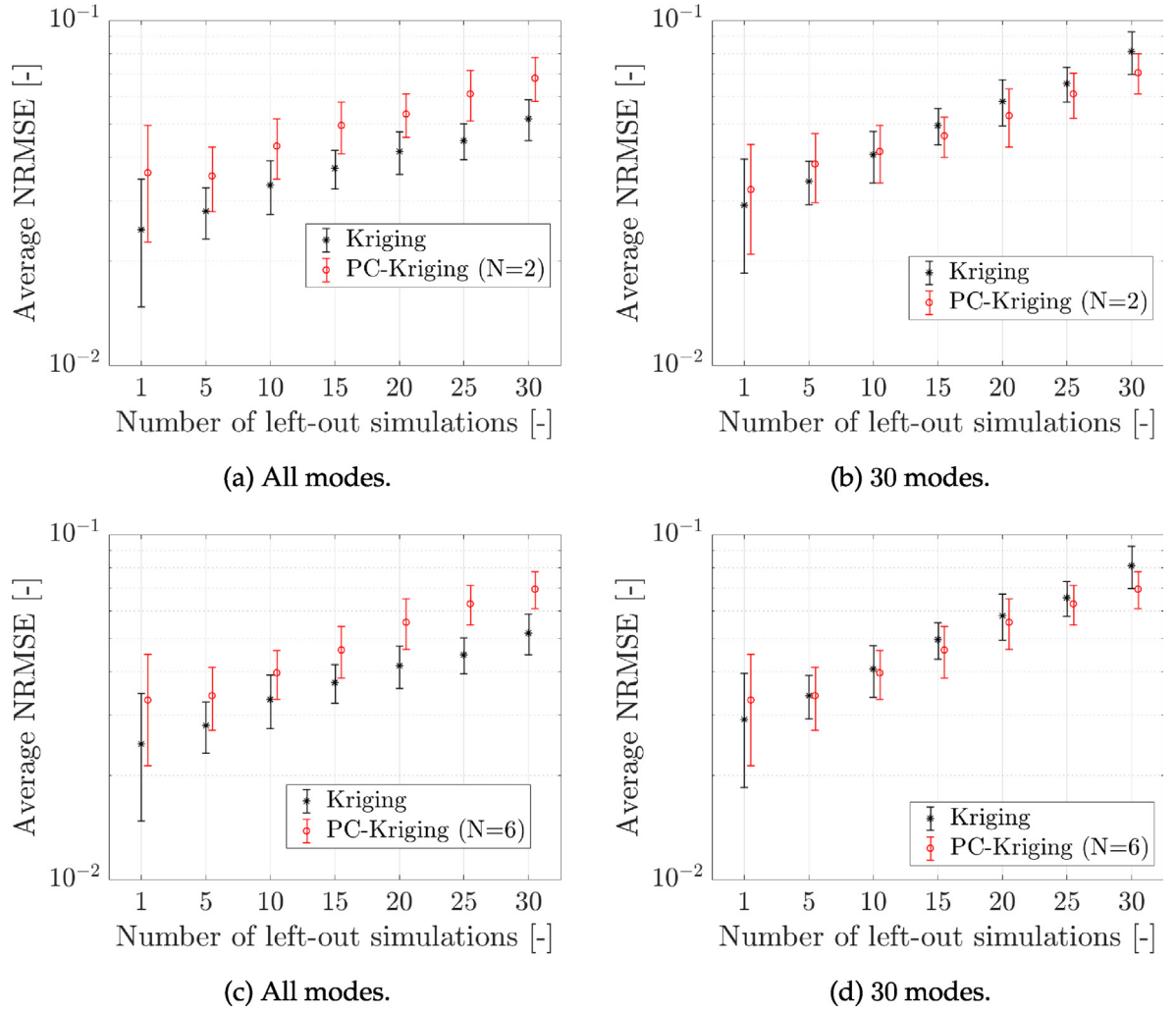

Fig. 9. Average NRMSE among all the considered variables for the prediction of the test data for an increasing number of training simulations by means of POD + Kriging and POD + PC-Kriging with $N = 2$ and $N = 6$ and (a, c) all modes and (b, d) only 30 modes. $q = 0.5$. The height of the bars represents the standard deviation of the NRMSE.

employed when a leave-*k*-out analysis is computationally prohibitive. These DoEs are reported in Fig. 7.

Fig. 8 reports the NRMSE for the prediction of the considered system's variables for DoE-A, for two different values of the parameter q, for $N =$

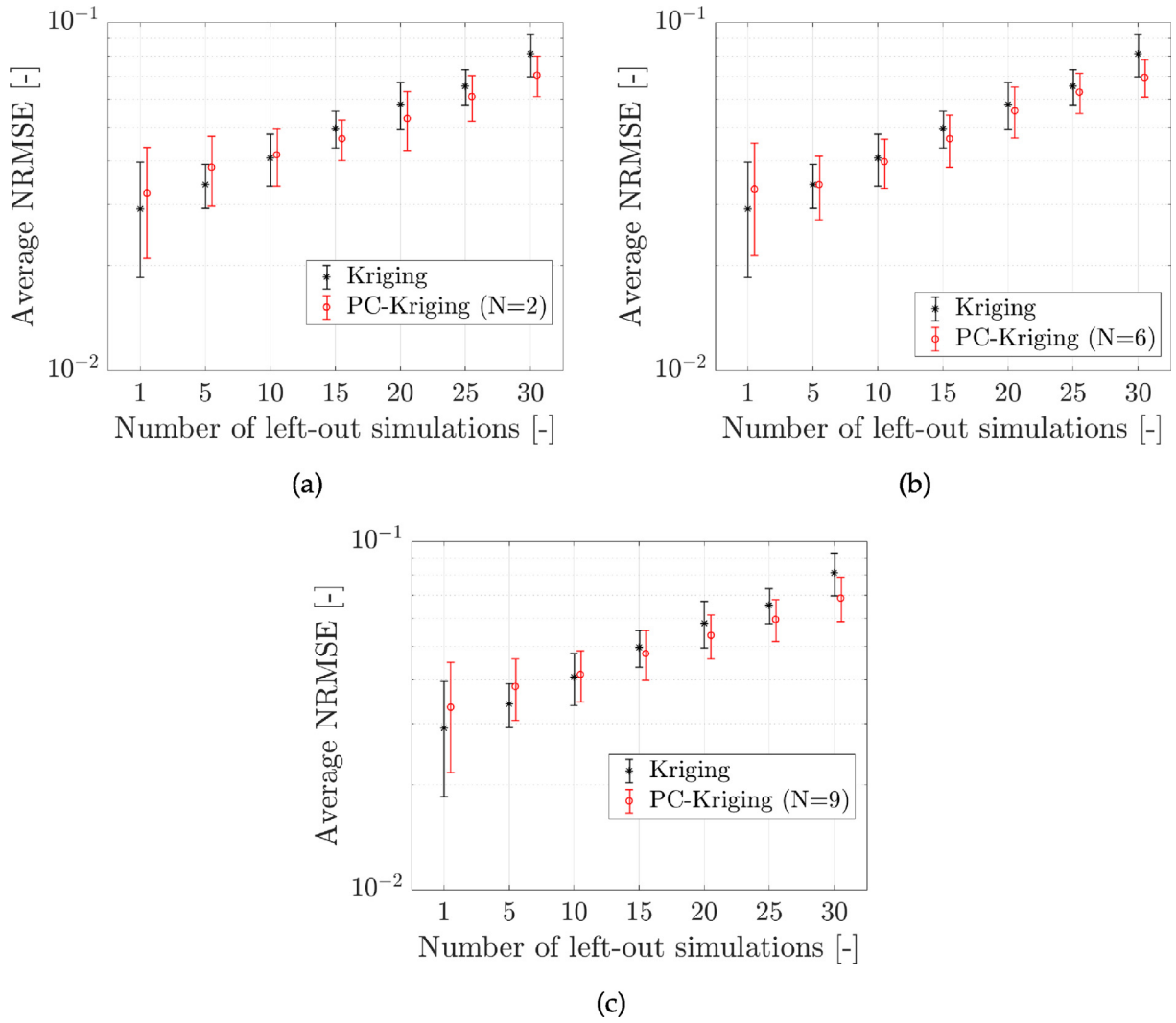

Fig. 10. Average NRMSE among all the considered variables for the prediction of the test data for an increasing number of training simulations by means of POD + Kriging and POD + PC-Kriging with (a) $N = 2$, (b) $N = 6$ and (c) $N = 9$. $q = 0.5$ and 30 modes retained. The height of the bars represents the standard deviation of the NRMSE.

2. In both Fig. 8a and b, the NRMSE for the Kriging model are the same as q is a parameter of PCE. It can be observed that decreasing the value of q led to an improvement for the PC-Kriging model, indicating that the interaction between polynomials of high degrees can be detrimental and meaning that a trade-off in terms of polynomial complexity (optimal values for N and q) should be found when using PC-Kriging.

A leave-k-out analysis is carried out in order to investigate the influence of the number of training simulations on ROM's performances. This analysis is repeated for different values of N, so to determine the sensitivity of the model to this parameter as well. k is the number of simulations left out from the training data set employed to find the POD reduced basis and train the predictive models. Each time, k simulations are left out and the error to predict the left-out simulations is estimated.

Fig. 9 reports the average prediction errors by means of leave-k-out for the POD + Kriging and POD + PC-Kriging models with $N = 2$ and $N = 6$, for a different number of retained POD modes. In particular, all POD modes are retained for the results of Fig. 9a and c, while only 30 POD modes are retained for Fig. 9b and d. POD + Kriging's prediction errors are lower when more modes are retained, while the opposite happened for POD + PC-Kriging. Specifically, when less modes are retained, the POD + PC-Kriging ROM performed better than the POD + Kriging ROM, for smaller training sizes, indicating the PC-Kriging is a better choice than Kriging for coarse experimental designs and low approximation orders. The estimated average prediction errors of the POD + Kriging and POD + PC-Kriging ($N = 2, N = 6$ and $N = 9$) models by means of leave-k-out analysis are reported in Fig. 10, from which it can be observed that POD + PC-Kriging outperformed POD + Kriging for $k > 10$ when $N = 2$, and for $k > 5$ for $N = 6$. The POD + PC-Kriging model's performance declined for $N = 9$, indicating a trade-off for N has to be found. Fig. 11 shows that the NRMSE for POD + PCE with $N = 2$, $N = 6$ and $N = 9$ are approximately 10 times higher with respect to the other models, further confirming that the combination of the two models led to improved performances.

Fig. 12 reports the contours of the true and predicted temperature fields by the 3 ROMs when trained on DoE-A, for one specific combination of the input parameters. Fig. 13 reports the response surfaces to be found for the first 2 POD scores.

3.3. Fixing Kriging length-scales

A POD + PC-Kriging model with small values for the Kriging length-scales is built ($l = 10^{-5}$) so to investigate on the effects of this parameter. This way, the Kriging model could influence the predictions only in the very proximity of the training data and, thus, would allow to see the fine scales of the correlation structure. The NRMSE for the POD + PCE and POD + PC-Kriging models for different values of N and a fixed length-

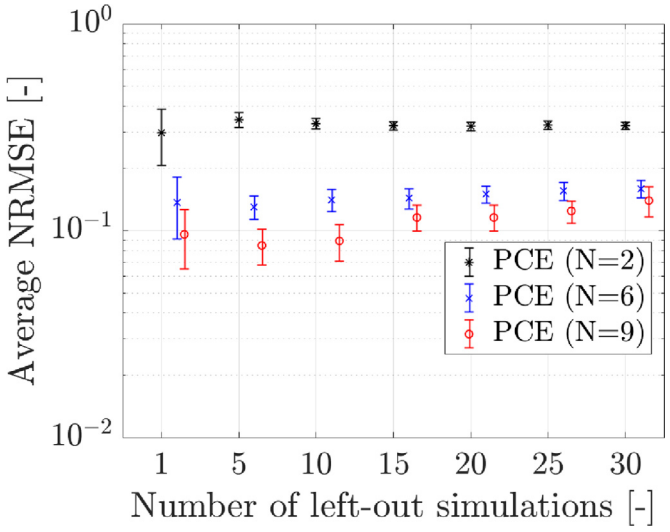

Fig. 11. Average NRMSE among all the considered variables for the prediction of the test data for an increasing number of training simulations by means of POD + PCE with $N = 2$, $N = 6$ and $N = 9$. $q = 0.5$ and 30 modes retained. The height of the bars represents the standard deviation of the NRMSE.

scale $l = 10^{-5}$ are reported in Fig. 14, from which it can be observed that: for both POD + PCE and POD + PC-Kriging, lower prediction errors are obtained when N is increased; the performances of POD + PCE and POD + PC-Kriging for the prediction of the test data coincided, indeed indicating that the predictive capabilities of the PC-Kriging model come from the PCE part when small values for l are set. However, the POD + PC-Kriging model still offers an advantage over the plain POD + PCE model as, thanks to Kriging, it interpolates the training data, differently from PCE which is a regression method. The combination of PCE with Kriging thus offers the possibility of training only the PCE part of the model (estimating the polynomial coefficients) which is computationally cheaper than training a Kriging model, and of manually setting the values of the Kriging kernel length-scales to small values so that its predictions are forced by the Kriging part of the model to be closer to the training values in proximity of the training data and thus to interpolate them.

3.4. Combination of POD and ANN

The use of ANN as supervised technique in combination with POD for the prediction of two random DoEs of different sizes, DoE-A and DoE-B, is also investigated in the present work. These DoEs are reported in Fig. 7. In this case, a leave-k-out analysis is not performed because of its computational cost. As explained in Section 2.3.4, ANN offers the possibility of training one model, namely one neural network, for the prediction of all the POD scores, simultaneously. In the present work, the architecture for the ANN is chosen with one input layer of dimensions $d = 2$ (number of input parameters) and 4 hidden layers of dimensions 8, 32, 128, 256, respectively. The dimension of the output layer is 20 for DoE-A and for DoE-B, given that a total of 20 modes is retained. The activation functions for all the hidden layers are Leaky ReLU with different slopes, specifically the first 2 hidden layers had a slope of 10^{-4} and the last two had a slope of 10^{-3}.

Fig. 15 reports the NRMSE for the prediction of the DoE-A and DoE-B by POD + ANN and by POD + PC-Kriging. Even though only one model for the prediction of the POD scores is necessary to train when using ANN in combination with POD, the resulting prediction errors are higher with respect to the POD + PC-Kriging ROM's. The prediction errors of POD + ANN decreased by $< 50\%$ when increasing the size of the training data-set, whereas the prediction errors of POD + PC-Kriging decreased by a factor of ≈ 10. The training of an ANN consists in a wide range of design choices for its architecture, such as number of hidden layers or type of activation functions, to be cross-validated. Besides, ANNs are usually employed for cases with a high number of observations available, which is not the case for computationally expensive combustion simulations. This makes ANN a more complex choice as supervised method for a ROM, which might not be preferable for the present case, where PC-Kriging is a much simpler method to set up. Because of the mentioned reasons, the use of different ANNs (training one network per score) was not investigated in this study as this method is not expected to perform better than Kriging or PCE for very small data-sets. The advantage of using ANN is to have one surrogate (network) for the simultaneous prediction of all the scores.

In conclusion, the results presented in this Section showed that the POD + Kriging or POD + PC-Kriging ROM performed with lower prediction errors by a factor of ≈ 10 in comparison to POD + PCE. The use of PCE as Kriging trend function could improve the Kriging model's performance for smaller training sizes. In general, the POD + Kriging model

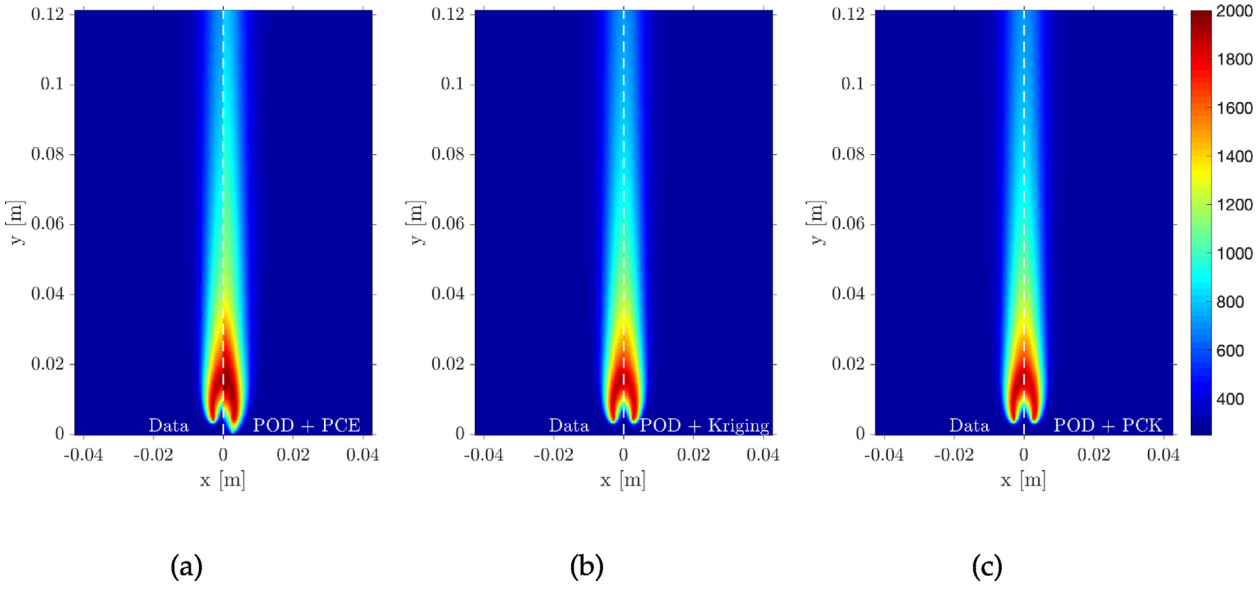

Fig. 12. Left halves: true data. Right halves: (a) prediction of the temperature field by POD + PCE; (b) prediction of the temperature field by POD + Kriging; (c) prediction of the temperature field by POD + PC-Kriging. POD: 20 PCs. Polynomial Chaos: $N = 2$, $q = 0.5$, Legendre. Kriging: Matern52 kernel.

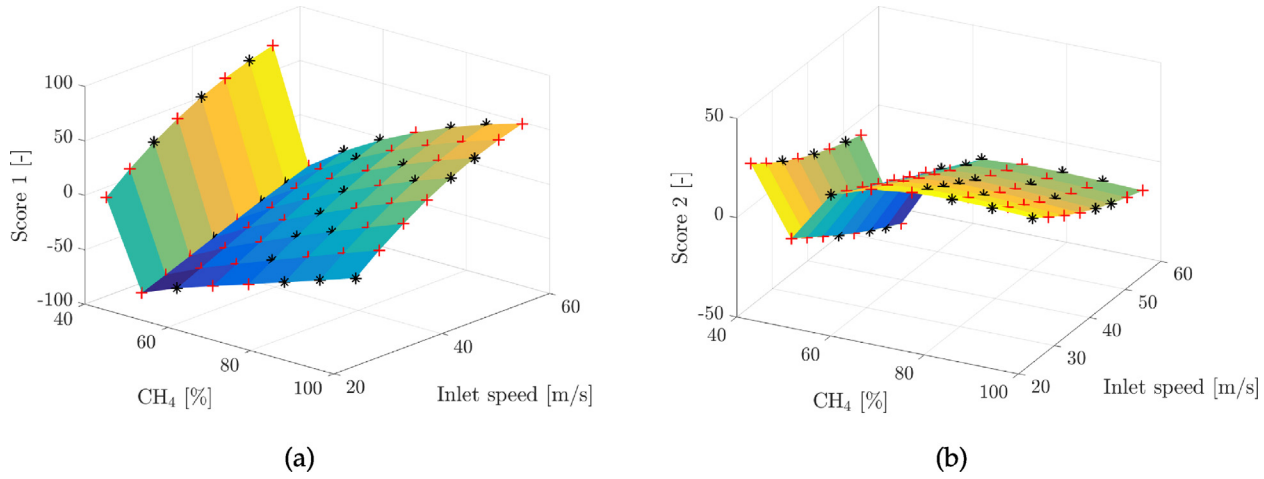

Fig. 13. Response surfaces of the POD score (a) 1 and (b) 2 for the training (black stars) and test (red crosses) data of DoE-A.

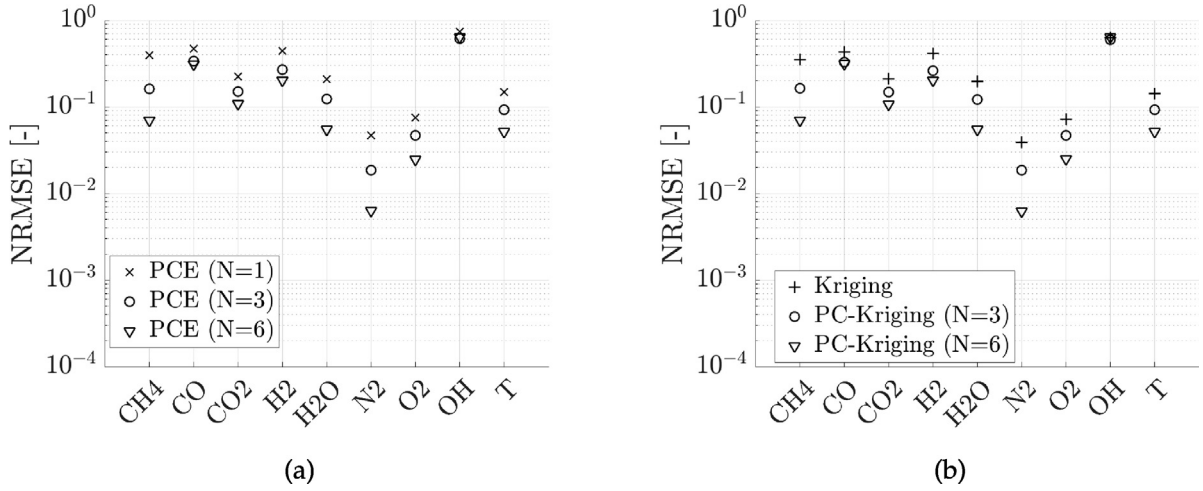

Fig. 14. NRMSE for the prediction of the test data of DoE-A by the (a) PCE and (b) PC-Kriging model with different values for the polynomial order N for a small value of the Kriging length-scale $l = 10^{-5}$.

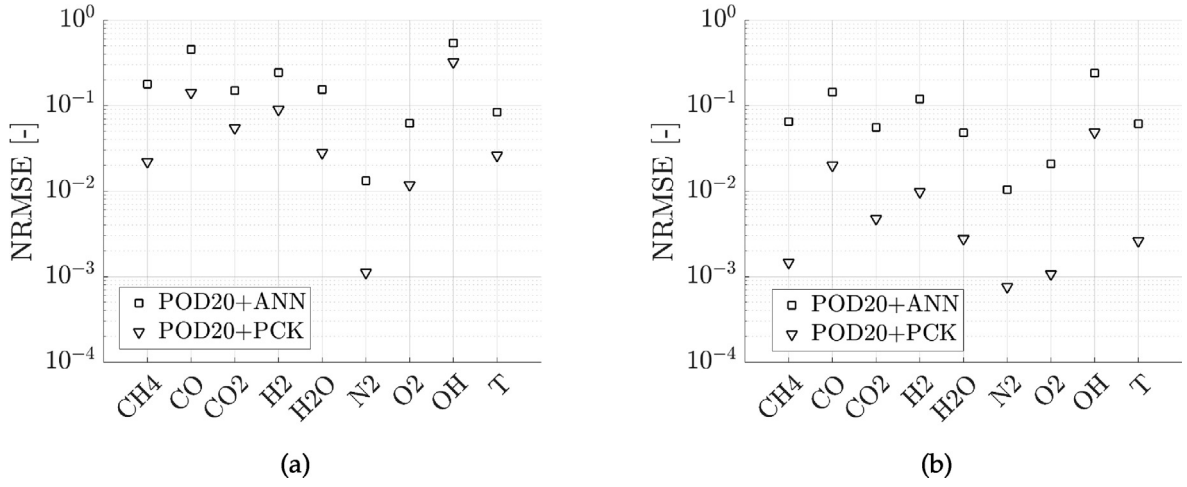

Fig. 15. NRMSE for the prediction of (a) DoE-A and (b) DoE-B by a ROM which combines POD with 20 modes and ANN. The NRMSE by POD + PC-Kriging are also reported for comparison. Kriging: Matern52. Polynomial Chaos: Legendre, $N = 2$, $q = 1.0$.

performed satisfactorily. In fact, the fields of the mass fractions of CH_4, CO_2, H_2O, N_2, O_2 and temperature are predicted with NRMSE $< 1\%$ for DoE-B. The combination of POD with ANN did not lead to satisfactory results as the approach achieved higher prediction errors by a factor of ≈ 10 with respect to POD + PC-Kriging.

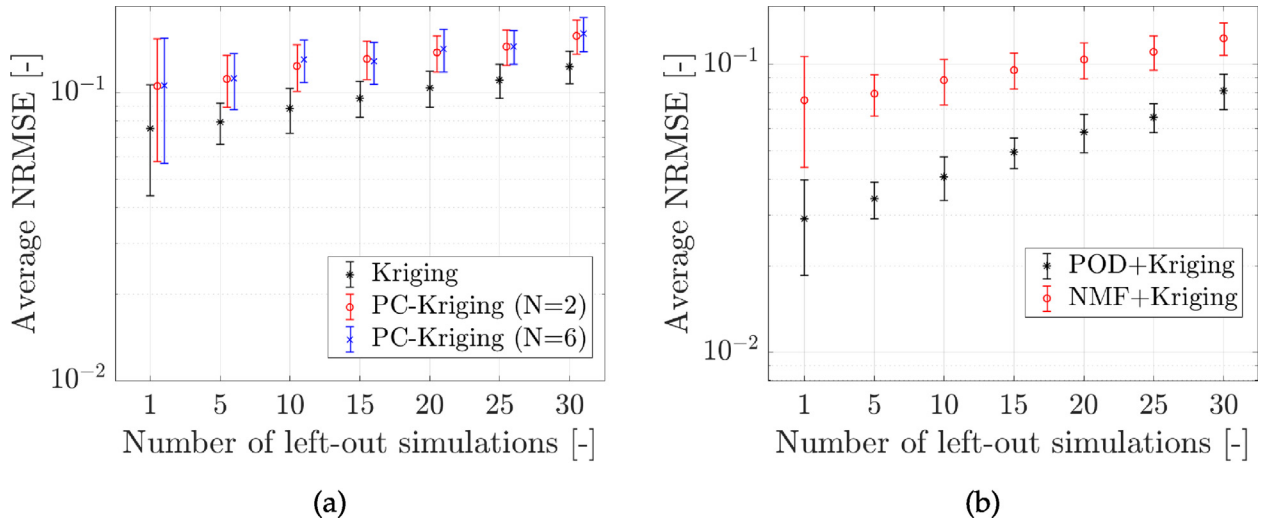

Fig. 16. (a) Average NRMSE among all the considered variables for the prediction of the test data for an increasing number of training simulations by means of NMF + Kriging and NMF + PC-Kriging. The height of the bars represents the standard deviation of the NRMSE. (b) Comparison between POD + Kriging and NMF + Kriging. The POD-based ROM had lower NRMSE for all values for the number of left-out simulations k. 30 modes.

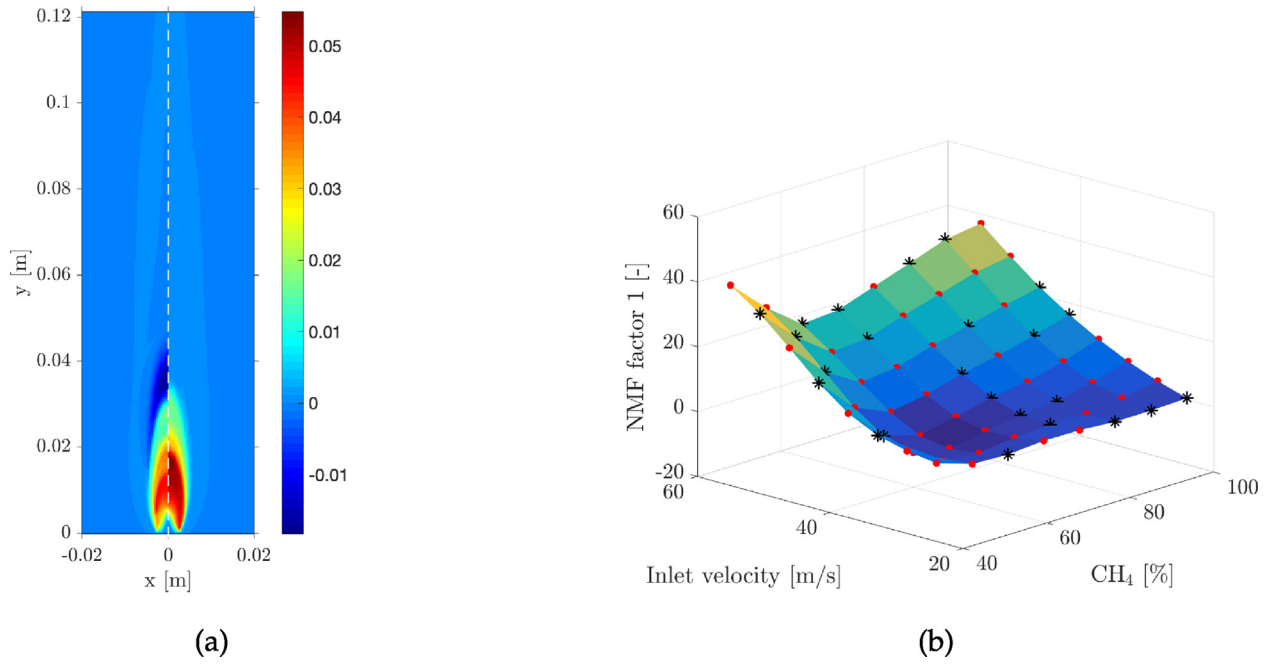

Fig. 17. (a) Predicted CO mass fraction by (left) the NMF-based ROM with no correction and (right) the NMF-based ROM with correction, i.e. after setting the predicted negative-valued NMF scores to zero. (b) Response surface found by PC-Kriging in a NMF-based ROM with 20 modes on DoE-A for one NMF score. PC-Kriging predicted negative values for this score as no positivity constraint is enforced on predictions. Kriging: Matern52. Polynomial Chaos: Legendre, $q = 0.5$, $N = 2$.

4. Application of the NMF-based ROM

The performance of the ROM developed by combining NMF and PC-Kriging is investigated in the present work.

A leave-k-out analysis is carried out in order to investigate the influence of the number of training simulations on performances of a ROM based on NMF as well. k is the number of simulations left out from the training data set employed to find the NMF reduced basis and train the predictive models. Each time, k simulations are left out and the error to reconstruct the left-out simulations from the POD basis is estimated. The estimated average prediction errors of the NMF + Kriging and NMF + PC-Kriging models by means of leave-k-out analysis are reported in Fig. 16a. Fig. 16b compares the performance of the NMF + Kriging model with the POD + Kriging one. The overall higher NRMSE in comparison to a POD-based ROM suggested that POD proved to be a better choice unsupervised method for the proposed approach.

Fig. 17a reports an example of predicted CO mass fraction field by (left) the NMF-based ROM and (right) still the NMF-based ROM after setting the predicted negative-valued scores to zero. As shown in Fig. 17a, because no positivity constraint is enforced on the predicted NMF scores, the predicted CO field by the NMF-based ROM had negative values just like the one predicted by the POD-based ROM, which nullified the advantage that NMF should have provided over POD. A similar issue arose from the combination of Kriging with Constrained POD [6], as shown in Ref. [4]. In fact, as shown in Fig. 17b, which reports the response surface for one NMF score, PC-Kriging predicted negative values for this quantity. However, differently from a POD-based ROM, where no correction on the predicted POD scores can be made to avoid

predicting negative mass fractions, this is possible for NMF. In fact, manually correcting the predicted negative values for all the NMF scores to be equal to 0 led to a less accurate but physically acceptable prediction, as shown in Fig. 17a, which reports the predicted CO mass fraction by the NMF-based ROM when such a correction on the NMF scores is done.

In conclusion, the NMF-based ROMs had higher prediction errors in comparison to the POD-based ROMs. The variables that are predicted with a NRMSE $< 10\%$ for DoE-A are mass fractions of CH_4, CO_2, H_2O, N_2, O_2 and temperature. Although the predicted NMF scores are at times negative in value, leading to predicted negative mass fractions, NMF offered the possibility to correct this by setting the predicted negative values to zero.

5. Conclusions

In the present work, inspired by Refs. [4,8], two data compression techniques such as POD and NMF are combined with a predictive model based on the combination of Polynomial Chaos Expansion (PCE) and Kriging for the development of a Reduced-Order Model (ROM) for the prediction of combustion data, with PCE functioning as Kriging trend. In order to compare the performance of the PC-Kriging interpolation, POD and NMF are also combined with PCE only and Kriging only.

As regards the unsupervised techniques used for data compression for the development of the ROM, the results showed that POD could reconstruct the training data with an NRMSE which is ≈ 10 times lower with respect to NMF. On the other hand, the positivity constraint imposed in the NMF formulation guaranteed that positive physical quantities are not reconstructed with negative values. Both methods generalized to unseen data with similar performances in terms of errors for data reconstruction, with POD having reconstruction errors $\approx 10\%$ lower.

As concerns the supervised part of the ROM, results from a leave-*k*-out analysis showed that the PC-Kriging interpolation performed with lower prediction errors than Kriging for smaller training sizes and for low approximation orders. Both PC-Kriging and Kriging outperformed PCE, when these techniques are combined with POD. PC-Kriging performed better (prediction errors lower by $\approx 10\%$) than Kriging also when in combination with NMF. In general, the POD-based ROM performed with lower prediction errors than the NMF-based ROM, by a factor of ≈ 10 on CH_4 and N_2. The PC-Kriging method performed better than Kriging also when the Kriging length-scales are not optimized and set to low values, so that the Kriging would influence predictions only in the regions close to the training data (and interpolate those data), while PCE would contribute in the regions far from the training data. By doing so, the overall prediction errors are higher than the case where the Kriging length-scales are optimized, although the training of such a model is computationally less demanding as it did not involve the optimization procedure for the evaluation of the length-scales and thus it only involved the evaluation of the PCE coefficients.

The use of ANN as supervised method for a POD-based ROM is also investigated as ANN offers the possibility of training one model for the prediction of all the POD (or NMF) scores, simultaneously. The predictive capabilities of combining POD with ANN are compared to the ones of POD in combination with PC-Kriging. Results showed that, again, POD combined with PC-Kriging had lower prediction errors with respect to the combination of POD and ANN, by a factor of ≈ 10 for nearly all variables involved, as ANN usually requires a high number of observations to perform well, which is not the case in the present work where data from computationally expensive combustion simulations are employed. The training of an ANN also meant that a wide range of design choices for its architecture, such as number of hidden layers or type of activation functions, needed to be explored or cross-validated, which made ANN a more complex choice as supervised method.

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.

Acknowledgements

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. G.D.A. was supported by a FRIA fellowship from the Fonds National de la Recherche Scientique, FRS-FNRS.

References

[1] J. Kabuba, M. Banza, Ion-exchange process for the removal of ni (ii) and co (ii) from wastewater using modified clinoptilolite: modeling by response surface methodology and artificial neural network, Results in Engineering 8 (2020) 100189, https://doi.org/10.1016/j.rineng.2020.100189.
[2] Y. Lang, S.E. Zitney, L.T. Biegler, Optimization of igcc processes with reduced order cfd models, Comput. Chem. Eng. 35 (9) (2011) 1705–1717.
[3] Y.-d. Lang, A. Malacina, L.T. Biegler, S. Munteanu, J.I. Madsen, S.E. Zitney, Reduced order model based on principal component analysis for process simulation and optimization, Energy Fuels 23 (3) (2009) 1695–1706.
[4] G. Aversano, A. Bellemans, Z. Li, A. Coussement, O. Gicquel, A. Parente, Application of reduced-order models based on PCA & Kriging for the development of digital twins of reacting flow applications, Comput. Chem. Eng. 121. doi:10.1016/j.compchemeng.2018.09.022.
[5] G. Aversano, M. Ferrarotti, A. Parente, Digital twin of a combustion furnace operating in flameless conditions: reduced-order model development from CFD simulations, Proc. Combust. Inst.:10.1016/j.proci.2020.06.045.
[6] M. Xiao, P. Breitkopf, R. Filomeno Coelho, C. Knopf-Lenoir, P. Villon, W. Zhang, Constrained Proper Orthogonal Decomposition based on QR-factorization for aerodynamical shape optimization, Appl. Math. Comput. 223 (2013) 254–263, https://doi.org/10.1016/j.amc.2013.07.086.
[7] R. Schöbi, B. Sudret, J. Wiart, HAL Id : hal-01432195.
[8] R. Schöbi, B. Sudret, Combining Polynomial Chaos Expansions and Kriging, 2014, pp. 1–48.
[9] F. Boukouvala, Y. Gao, F. Muzzio, M.G. Ierapetritou, Reduced-order discrete element method modeling, Chem. Eng. Sci. 95 (2013) 12–26.
[10] S. Inazumi, S. Intui, A. Jotisankasa, S. Chaiprakaikeow, K. Kojima, Artificial intelligence system for supporting soil classification, Results in Engineering 8 (2020) 100188, https://doi.org/10.1016/j.rineng.2020.100188.
[11] T. Klotz, R. Pothier, D. Walch, T. Colombo, Prediction of the business jet global 7500 wing deformed shape using fiber bragg gratings and neural network, Results in Engineering 9 (2021) 100190, https://doi.org/10.1016/j.rineng.2020.100190.
[12] G.C. Cawley, N.L.C. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, J. Mach. Learn. Res. 11 (2010) 2079–2107.
[13] A. Parente, J. C. Sutherland, Principal component analysis of turbulent combustion data: data pre-processing and manifold sensitivity, Combust. Flame:10.1016/j.combustflame.2012.09.016.
[14] N. Gillis, Sparse and unique nonnegative matrix factorization through data preprocessing, J. Mach. Learn. Res. 13 (2012) 3349–3386, arXiv:arXiv:1204.2436v1.
[15] C.E. Frouzakis, Y.G. Kevrekidis, J. Lee, K. Boulouchos, A.A. Alonso, Proper orthogonal decomposition of direct numerical simulation data: data reduction and observer, Construction 28 (2000) 75–81.
[16] P. Paatero, U. Tapper, Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, Environmetrics 5 (2) (1994) 111–126, https://doi.org/10.1002/env.3170050203.
[17] R. Lebrun, A. Dutfoy, An innovating analysis of the Nataf transformation from the copula viewpoint, Probabilist. Eng. Mech. 24 (3) (2009) 312–320, https://doi.org/10.1016/j.probengmech.2008.08.001.
[18] P.G. Constantine, E. Dow, Q. Wang, Active subspace methods in theory and practice: applications to kriging surfaces, SIAM J. Sci. Comput. 36 (4) (2014) A1500–A1524, https://doi.org/10.1137/130916138, arXiv:1304.2070.
[19] S.N. Lophaven, J. Søndergaard, H.B. Nielsen, Kriging Toolbox, 2002, pp. 1–28.
[20] M. Seeger, Gaussian Processes for Machine Learning, vol. 14, 2004, https://doi.org/10.1142/S0129065704001899 arXiv:026218253X.
[21] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, An MIT Press book, 2016. https://www.deeplearningbook.org/.
[22] C.M. Bishop, Pattern Recognition and Machine Learning, vol. 53, 2013, https://doi.org/10.1117/1.2819119 arXiv:arXiv:1011.1669v3.

[23] M. Gholamrezaei, K. Ghorbanian, Rotated general regression neural network, IEEE International Conference on Neural Networks - Conference Proceedings 2 (6) (2007) 1959–1964, https://doi.org/10.1109/IJCNN.2007.4371258.

[24] S. Cao, B. Ma, D. Giassi, B.A.V. Bennett, M.B. Long, M.D. Smooke, Effects of pressure and fuel dilution on coflow laminar methane-air diffusion flames: a computational and experimental study, Combust. Theor. Model. 22 (2) (2018) 316–337, https://doi.org/10.1080/13647830.2017.1403051.

[25] S. Cao, B. Ma, B.A. Bennett, D. Giassi, D.P. Stocker, F. Takahashi, M.B. Long, M.D. Smooke, A computational and experimental study of coflow laminar methane/air diffusion flames: effects of fuel dilution, inlet velocity, and gravity, Proc. Combust. Inst. 35 (1) (2015) 897–903, https://doi.org/10.1016/j.proci.2014.05.138.