Application of reduced-order models based on PCA and Kriging for the development of digital twins of reacting flow applications

Gianmarco Aversano, Aurelie Bellemans, Zhiyi Li, Axel Coussement, Olivier Gicquel, Alessandro Parente

Computers and Chemical Engineering 121 (2019) 422–441

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Application of reduced-order models based on PCA & Kriging for the development of digital twins of reacting flow applications

Gianmarco Aversano[a,b,c,*], Aurélie Bellemans[a,b], Zhiyi Li[a,b], Axel Coussement[a,b], Olivier Gicquel[c], Alessandro Parente[a,b]

[a] Université Libre de Bruxelles, École polytechnique de Bruxelles, Aero-Thermo-Mechanics Laboratory, Bruxelles, Belgium
[b] Université Libre de Bruxelles and Vrije Universiteit Brussel, Combustion and Robust Optimization Group (BURN), Brussels, Belgium
[c] Laboratoire EM2C, CNRS, Centrale-Supélec, Université ParisSaclay, 8-10 rue Joliot-Curie, Gif-sur-Yvette 91190 France

ARTICLE INFO

Article history:
Received 8 June 2018
Revised 16 August 2018
Accepted 22 September 2018
Available online 24 October 2018

Keywords:
Principal component analysis
Kriging
Surrogate models

ABSTRACT

Detailed numerical simulations of detailed combustion systems require substantial computational resources, which limit their use for optimization and uncertainty quantification studies. Starting from a limited number of CFD simulations, reduced-order models can be derived using a few detailed function evaluations. In this work, the combination of Principal Component Analysis (PCA) with Kriging is considered to identify accurate low-order models. PCA is used to identify and separate invariants of the system, the PCA modes, from the coefficients that are instead related to the characteristic operating conditions. Kriging is then used to find a response surface for these coefficients. This leads to a surrogate model that allows performing parameter exploration with reduced computational cost. Variations of the classical PCA approach, namely Local and Constrained PCA, are also presented. This methodology is demonstrated on 1D and 2D flames produced by OpenSmoke++ and OpenFoam, respectively, for which accurate surrogate models have been developed.

1. Introduction

In many engineering applications, complex physical systems can only be described by high-fidelity expensive simulations. Due to the non-linearity of these problems, changing the operating conditions, namely the model's input parameters, can drastically change the state of the considered system. Complete knowledge about the investigated system's behavior for a full range of operating conditions can therefore only be achieved by running these expensive simulations several times with different inputs, until enough observations of the system's state are obtained.

In this study, we focus on combustion systems that fall in this category as they are characterized by very complex physical interactions, between chemistry, fluid-dynamics and heat transfer processes. Our objective is to develop advanced Surrogate Models (SMs) that can accurately represent the behavior of complex reacting systems in a wide range of conditions, without the need for expensive Computational Fluid Dynamics (CFD) simulations. This is particularly attractive for the development of digital counterparts of real systems, with application in monitoring, diagnostics and prognostics (Schleich et al., 2017; Uhlemann et al., 2017). To this purpose, we derive techniques from the Machine Learning community.

In our approach a specific computationally-expensive CFD simulation or computer code, referred to as Full-Order Model (FOM) (Bizon and Continillo, 2012; Bizon et al., 2012), is treated as a black box that generates a certain output $\mathbf{y}$ (e.g. the temperature field) given a set of input parameters $\mathbf{x}$ (e.g. the equivalence ratio) and indicated by $\mathcal{F}(\cdot)$:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}). \tag{1}$$

The evaluation of the function $\mathcal{F}(\cdot)$ usually requires many hours of computational time. After enough observations of the FOM's output are available, $\mathbf{y}(\mathbf{x}_i) \quad \forall i = 1, \ldots, M$, a SM can be trained and the output $\mathbf{y}^*$ for a particular set of unexplored inputs $\mathbf{x}^*$ can be predicted without the need to evaluate $\mathcal{F}(\mathbf{x}^*)$. The function $\mathcal{F}(\cdot)$ is therefore approximated by a new function $\mathcal{M}(\cdot)$ whose evaluation is very cheap compared to $\mathcal{F}(\cdot)$:

$$\mathbf{y}^* = \mathcal{F}(\mathbf{x}^*) \approx \mathcal{M}(\mathbf{x}^*). \tag{2}$$

SMs are mathematical models based on available data that try to approximate the underlying *hidden* relationship between in-

* Corresponding author.
E-mail address: gianmarco.aversano@ulb.ac.be (G. Aversano).

https://doi.org/10.1016/j.compchemeng.2018.09.022

put and output. A very simple example of a SM is a linear regression of available data. SMs are useful when this relationship is either not known or comes in the form of a computationally expensive computer code. This is the case of a CFD simulation. SMs are constructed or *trained* from a relatively small set of *training* observations of the models output, which correspond to a set of training locations or points in the model parameter space. Once trained, SMs allow for a fast evaluation of the system's state over a wide range of their input parameters. Therefore, they are very appealing in the context of optimization studies as well as for Uncertainty Quantification (UQ) (Lin, 2017) and global optimization problems (Müller et al., 2013; Regis and Shoemaker, 2005). In Fürst et al. (2018), SMs are used to optimize the performance of chemical kinetics with respect to MILD combustion. In Khuwaileh and Turinsky (2017), SMs are employed in a Bayesian approach to calibrate various neutronics and thermal-hydraulics parameters. SMs are also used in the AlgoRithms for Global Optimization of coNstrAined grey-box compUTational problems (ARGONAUT) framework (Beykal et al., 2018a), which was also utilised in Beykal et al. (2018b) for the optimization of the operation of an oilfield using water-flooding. Ideally, the trained SMs should preserve the physics of the investigated phenomena, and be developed from a limited number of expensive function evaluations, i.e. CFD simulations. Examples of SMs are Radial Basis Functions and Polynomial Chaos Expansion (Crestaux et al., 2009). Examples of SMs used in combustion applications can be found in Lancien et al. (2016).

SMs are generally constructed directly on the analyzed system's output, i.e. directly on the variables of interest like the velocity and temperature fields. For each individual output variable a SM is trained and a response surface is found, indicating the relationship between the variable and the input parameters. If the number of variables of interest is high, many SMs need to be trained. Besides, any correlation between these variables of interest might be lost in the process of training individual SMs. Reducing the number of SMs to train is possible if the original set of variables can be represented by a new set of fewer scalars. This corresponds to the idea that the original variables are actually realization of unknown *latent* variables (Bishop, 2013).

Principal Component Analysis (PCA) (Jolliffe, 2002) offer the potential of preserving the physics of the system while reducing the size of the problem. PCA is a statistical technique used to find a set of orthogonal low-dimensional basis functions, called Principal Components (PCs), to represent an ensemble of high-dimensional data. PCA finds a new, smaller set of uncorrelated variables, often referred to as *PCA scores*, which is representative of the original variables of interest. PCA is also used for data interpretation, usually combined with rotation methods (Bellemans et al., 2018). Once these PCA scores are found, a SM can be built for each one of them. They are indicated as Reduced-Order (Surrogate) Models (ROMs).

SMs usually include interpolation or regression techniques which depend on the choice of some particular design functions. These design functions are defined by a set of so-called hyper-parameters (or also length-scales) whose values affect the SM's predictive abilities. Very often, a good estimation for the value of these hyper-parameters comes via the solution of constrained optimization problems that involve local optima. As shown in Xiao et al. (2010), ROMs are less sensitive to the particular design functions chosen for their construction, which is desirable. ROMs also have a reduced number of variables for which a SM needs to be trained. This means that fewer optimization problems are solved in order to estimate feasible values for the hyper-parameters of the design functions. In addition, in Xiao et al. (2010) it is also shown how ROMs usually scale better than classic SMs for parallel computing. These features are what makes PCA-based ROMs very attractive candidates for the development of physics-preserving SMs.

Combustion problems are well-known for being characterized by a set of strongly inter-dependent variables. In fact, PCA has been employed in Isaac et al. (2014, 2015) to re-parameterize the thermo-chemical state of a reacting system by a small number of progress variables, drastically reducing the number of transport equations to solve, and in the process showing the intrinsic lower-dimensionality of these systems, which will be exploited in the present work. PCA has also been employed in the context of turbulent combustion in Mirgolbabaei et al. (2014), for the a-posteriori validation of a turbulent combustion model based on the solution of transport equations for the principal components (Coussement et al., 2016; Echekki and Mirgolbabaei, 2015) and for on-line process monitoring and fault diagnostics (Yu, 2012).

The objective is to develop advanced SMs, trained on a reduced number of full simulations, able to predict the full system state in unexplored conditions, without running a new simulation. To this end, an approach based on the combination of PCA and Kriging was chosen. PCA is used to extract the invariant (w.r.t. the input parameters) physics-related information of an investigated combustion system and identify the system's coefficients which instead depend on the operating conditions, the PCA scores. Kriging interpolation is then used to find a response surface for these scores. With this strategy it was possible to build a ROM for parameter exploration with reduced computational cost. Kriging was chosen over other regression techniques not only because it provides an estimate of the prediction uncertainty, but also because it allows the user to add prior knowledge on the model by selecting different kernel functions. The use of Kriging for CFD data has also produced encouraging results. In fact, Kriging was employed for the shape optimization of a car engine intake port in Xiao et al. (2012) and for aerodynamical shape optimization problems as shown in Xiao et al. (2014, 2013). However, the application was limited to non-reacting flow problems.

In the present work, the Kriging-PCA approach is extended to combustion applications, to develop a ROM that can faithfully reproduce the temperature and chemical species fields in a reacting flow simulation. The methodology is demonstrated on a methane laminar premixed flame, increasing the complexity of the problem gradually, in terms of number of input parameters (equivalence ratio, inlet temperature and fuel composition), and problem dimensionality (from one to two-dimensional flames). The objective of the present work is to demonstrate the applicability of the proposed methodology for the development of reduced-order models of multi-scale and multi-physics computer models. In this perspective, this work paves the way for the development of digital twins (Haag and Anderl, 2018) of realistic engineering systems. The methodology outlined in the present work shows the advantages of the PCA+Kriging formulation in terms of predictive capabilities and computational efficiency. Indeed, these features are necessary for the development of predictive models of engineering systems, which can be employed for visualization, real-time control, optimization and troubleshooting.

From the application perspective, this paper presents the first application of the Kriging-PCA methodology to combustion problems. From the methodology point of view, this work combines, for the first time, Kriging with a local and a constrained formulation of PCA.

This paper is organized as follows. Section 2 will cover the employed methodology in details, while in Section 3, PCA and its variations, Kriging and a sampling strategy referred to as Adaptive Sampling are shown. Results are presented and discussed in Section 4. In Section 5 conclusions are drawn.

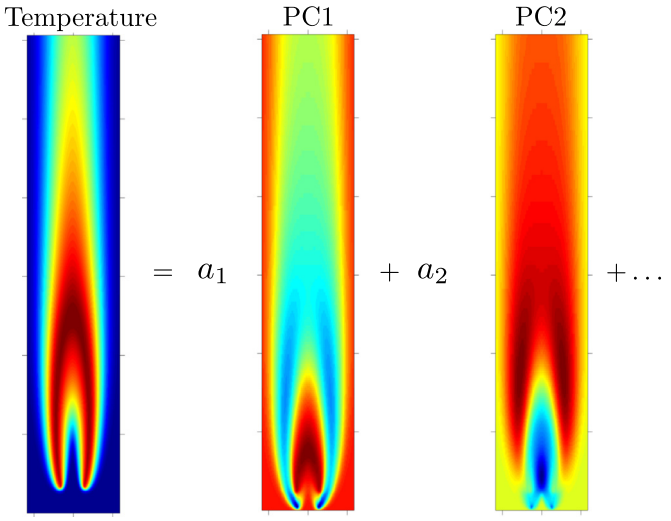

Fig. 1. Illustrative example of PCA applied to combustion data: one particular temperature spatial field is represented by a set of coordinates (the coefficients a_i, called PCA scores) on the basis functions or Principal Components (PCs) found by PCA.

2. Methodology

The methodology used in the present work is sketched in Fig. 1. Consider that a certain high-fidelity simulation model or Full Order Model (FOM) $\mathbf{y}(\mathbf{x}) = F(\mathbf{x}) \in \mathbb{R}^N$ is available, such as a CFD combustion solver. For one value of the input parameter(s) $\mathbf{x}$, the solver returns a vector $\mathbf{y}^{(j)}$ of observations of all the involved physical variables at every grid point:

$$\mathbf{y}^{(j)} = [T(r_1, \mathbf{x}_j), \dots, T(r_L, \mathbf{x}_j), Y_{CH_4}(r_1, \mathbf{x}_j), \dots, Y_{CH_4}(r_L, \mathbf{x}_j), \dots]^T, \tag{3}$$

where L is the total number of grid points, r_i is the ith spatial location and $\mathbf{x}_j$ is the jth point in the input parameter space. This FOM is solved for a limited amount $M < N$ of training points in the input parameter space $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(M)}\} \in \mathcal{D}$, where $\mathcal{D}$ is the region spanned by the training points. Thus, only M simulations are available, one for each of those points: $\mathbf{Y} = \{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(M)}\}$. The full exploration of the region $\mathcal{D}$ is possible only by running the expensive CFD-combustion solver $F(\cdot)$ for every $\mathbf{x} \in \mathcal{D}$. From the data-set $\mathbf{Y}$ of available simulations, PCA is able to extract a set of basis functions $\mathbf{\Phi} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_q\}$, with $q < N$ usually, called PCA modes that are invariant with respect to the input parameters $\mathbf{x}$. A set of coefficients $\mathbf{a}(\mathbf{x}) = \{a_1(\mathbf{x}), a_2(\mathbf{x}), \dots, a_q(\mathbf{x})\}$, called PCA scores and depending on $\mathbf{x}$, is consequently found. An illustrative example is reported in Fig. 2, where a temperature spatial field is represented as a set of coefficients that weight a set of basis functions, i.e. the PCs. These coefficients are less in number than the original number of variables as $q < N$ and can be interpolated in order to acquire knowledge about the system's state for any unexplored point $\mathbf{x}^* \in \mathcal{D}$.

One advantage of this approach is that a much smaller number of variables, namely q PCA scores, are interpolated instead of N. Another advantage is that the N original variables might be correlated. The application of PCA for the detection of latent variables, the PCA scores, preserves this correlation, which might be lost if each original variable is interpolated independently. One additional remark is that considering, for example, $T(r_i, \mathbf{x}_j)$ and $T(r_j, \mathbf{x}_j)$ as two separate variables (rather than using the spatial locations r_i as additional input parameters) also reduces the computational costs, as we shall see.

3. Theory

3.1. PCA

The key idea of Principal Component Analysis (PCA) is to reduce (compress) a large number of interdependent variables (i.e. independent up to the second-order statistical moments) to a smaller number of uncorrelated variables while retaining as much of the original data variance as possible (Bizon et al., 2010; Coussement et al., 2016; Parente and Sutherland, 2013; Parente et al., 2009; Sutherland and Parente, 2009).

For a data-set $\mathbf{Y}(N \times M)$, containing M observations of N original variables, namely temperature and species mass fractions measured at each spatial location of a considered geometrical domain, as described in Section 2, PCA provides an approximation of the original data-set using only $q < N$ linear correlations between the N variables. The quantity q is referred to as approximation order. In general, $q \leq \min(N, M)$. Thus, the vector $\mathbf{y} \in \mathbb{R}^N$ of observed temperature and species mass fractions can be encoded into a lower dimensional vector, $\mathbf{a} \in \mathbb{R}^q$.

Data are usually centered and scaled before PCA is carried out. Here we report six different choices for the scaling of the data:

1. Auto-scaling (STD), each variable is normalized by its standard deviation;
2. RANGE, each variable is normalized by its range;
3. PARETO, each variable is scaled by the square root of its standard deviation;
4. VAST, each variable is scaled by the standard deviation and coefficient of variation;
5. LEVEL, each variable is normalized by the mean of the data;
6. MAX, each variable is scaled by its maximum value.

The centered and scaled data read:

$$\mathbf{Y}_0 = \mathbf{D}^{-1}(\mathbf{Y} - \bar{\mathbf{Y}}), \tag{4}$$

where $\mathbf{D}$ indicates a diagonal matrix of chosen scaling factors, while $\bar{\mathbf{Y}}$ is a matrix containing the mean of each of the N variables, namely $[T(r_1, \mathbf{x}_j), \dots, T(r_L, \mathbf{x}_j), Y_{CH_4}(r_1, \mathbf{x}_j), \dots, Y_{CH_4}(r_L, \mathbf{x}_j), \dots]$ over the M observations.

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_N \end{bmatrix} \quad \bar{\mathbf{Y}} = \begin{bmatrix} \bar{y}_1 & \bar{y}_1 & \bar{y}_1 & \cdots & \bar{y}_1 \\ \bar{y}_2 & \bar{y}_2 & \bar{y}_2 & \cdots & \bar{y}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{y}_N & \bar{y}_N & \bar{y}_N & \cdots & \bar{y}_N \end{bmatrix}. \tag{5}$$

After centering and scaling the data, the covariance matrix $\mathbf{C}$ is evaluated as:

$$\mathbf{C} = \frac{1}{M-1}\mathbf{Y}_0\mathbf{Y}_0^T. \tag{6}$$

This matrix is symmetric and its rank $rank(\mathbf{C}) = rank(\mathbf{Y}) = \min(N, M)$. The set of PCA directions, the Principal Components (PCs) or modes, is found by solving the following set of eigenproblems: $\mathbf{C}\boldsymbol{\phi}_i = \lambda_i \boldsymbol{\phi}_i \quad \forall i = 1, 2, \dots, q$. Each PCA mode $\boldsymbol{\phi}_i$ has an associated eigenvalue, λ_i, which represents the variance of the original data taken into account by that mode (Williams, 2010). The PCA modes can be collected in a $N \times q$ matrix $\mathbf{\Phi} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_q\}$, sorted in descending order of importance.

The number q of PCA modes that are retained is usually much smaller than the dimension N. The PCA modes with the highest eigenvalues are the ones that are kept. Once the PCA modes are found, the data can be encoded in a set of q scalars called PCA scores. The PCA scores corresponding to the realization $\mathbf{y}(\mathbf{x}_j)$ are given by the projection:

$$a_i(\mathbf{x}_j) = \boldsymbol{\phi}_i^T \, \mathbf{y}(\mathbf{x}_j) \tag{7}$$

with $\forall i = 1, \dots, q$ and $\forall j = 1, \dots, M$. The PCA reduction can be ex-

$$\begin{bmatrix} \mathbf{y}^{(1)} = \mathcal{F}(\mathbf{x}^{(1)}) \\ \mathbf{y}^{(2)} = \mathcal{F}(\mathbf{x}^{(2)}) \\ \cdots \\ \mathbf{y}^{(M)} = \mathcal{F}(\mathbf{x}^{(M)}) \end{bmatrix} \rightarrow \text{PCA} \rightarrow \begin{bmatrix} a_1^{(1)}, \ \ldots, \ a_q^{(1)} \\ a_1^{(2)}, \ \ldots, \ a_q^{(2)} \\ \cdots \\ a_1^{(M)}, \ \ldots, \ a_q^{(M)} \\ \boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \ \boldsymbol{\phi}_2, \ \ldots, \ \boldsymbol{\phi}_q) \end{bmatrix} \longrightarrow \begin{array}{c} \text{Kriging response} \\ \text{surface for the} \\ \text{PCA scores} \end{array}$$

Fig. 2. PCA finds the set of Principal Directions $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \ \boldsymbol{\phi}_2, \ \ldots, \ \boldsymbol{\phi}_q)$ and encodes each observation $\mathbf{y}^{(i)} \in \mathbb{R}^N$ into a small set of scalars $a_1^{(i)}, \ \ldots, \ a_q^{(i)}$ for $q < N$. A Kriging response surface is then found for these scalars.

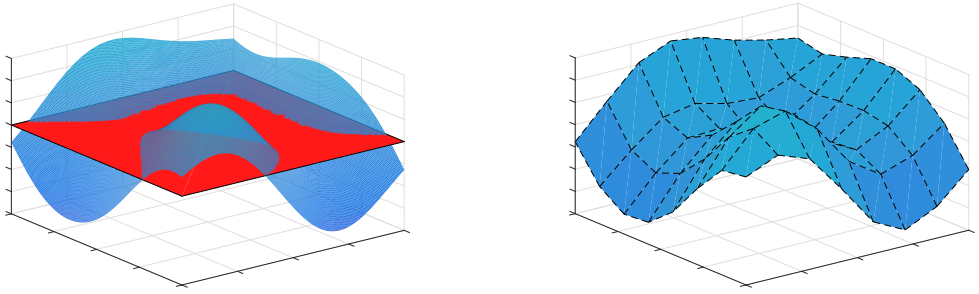

Fig. 3. (*left*) A non-linear hyper-surface is approximated by only one hyper-plane in the data space. (*right*) The same hyper-surface is approximated by a set of local hyper-planes. The application of PCA can lead to better performances if local regions in the data space are detected and PCA is applied locally and independently in each region.

pressed in matrix form as:

$$\mathbf{Y} = \tilde{\mathbf{Y}} + \mathbf{D}\mathbf{Y}_0 \approx \tilde{\mathbf{Y}} + \mathbf{D}\boldsymbol{\Phi}\mathbf{A} = \tilde{\mathbf{Y}}, \tag{8}$$

where $\mathbf{Y}$ is the data matrix as described in Section 2, $\mathbf{A} = \{\mathbf{a}(\mathbf{x}_1), \mathbf{a}(\mathbf{x}_2), \ldots, \mathbf{a}(\mathbf{x}_M)\}$ with $\mathbf{a} \in \mathbb{R}^q$ is the matrix where all the PCA scores for different values of the input parameters are stored and $\tilde{\mathbf{Y}}$ is the reconstruction of the data matrix $\mathbf{Y}$ after the compression achieved by PCA.

3.2. Local PCA

PCA is a linear combination of basis functions. A large number of PCs may be required when applying PCA on highly non-linear systems (Parente et al., 2009; Sutherland and Parente, 2009). Local PCA (LPCA) constructs local models, each pertaining to a different disjoint region of the data space (Kambhatla and Leen, 1997). Within each region, the model complexity is limited, and thus it is possible to construct linear models using PCA (Kambhatla and Leen, 1997; Sahyoun and Djouadi, 2013). Fig. 3 provides a general representation for a set of 3-dimensional observations forming a curved surface. Each axis shows the co-domain for each of the three scalar components that identify the 3-dimensional observations. The figure shows how a local representation of the curved surface can provide a better representation with respect to a single hyper-plane.

The partition in local clusters, where PCA is carried out, is accomplished using a Vector Quantization (VQ) algorithm that minimizes the reconstruction error. The reconstruction error is the squared Euclidean distance from one point or observation in the data-space to the linear manifold that is found by applying PCA in the local region. Mathematically, it can be expressed in a general fashion as:

$$d(\mathbf{z}, \ \mathbf{r}^{(i)}) = \left(\mathbf{z} - \mathbf{r}^{(i)}\right)^T \boldsymbol{\Phi}^{(i)\ T} \boldsymbol{\Phi}^{(i)} \left(\mathbf{z} - \mathbf{r}^{(i)}\right), \tag{9}$$

where $\mathbf{z}$ is the object to be assigned to the cluster $\mathcal{R}^{(i)}$, represented by the reference vector $\mathbf{r}^{(i)}$, defined as the centroid of the ith region: $\mathbf{r}^{(i)} = \mathrm{E}[\mathbf{z} \in \mathcal{R}^{(i)}]$. The cluster i is defined as:

$$\mathcal{R}^{(i)} = \{\mathbf{z} \mid d(\mathbf{z}, \mathbf{r}^{(i)}) \leq d(\mathbf{z}, \mathbf{r}^{(j)}); \ \forall \ j \neq i\}. \tag{10}$$

In this work, the objects to be assigned to different clusters are chosen to be the vectors of observations of one single variable, namely the rows of the data-matrix:

$$\mathbf{Y} = \begin{bmatrix} T(r_1, \mathbf{x}_1) & T(r_1, \mathbf{x}_2) & T(r_1, \mathbf{x}_3) & \cdots & T(r_1, \mathbf{x}_M) \\ \vdots & \vdots & \vdots & & \vdots \\ T(r_L, \mathbf{x}_1) & T(r_L, \mathbf{x}_2) & T(r_L, \mathbf{x}_3) & \cdots & T(r_L, \mathbf{x}_M) \\ Y_{CH_4}(r_1, \mathbf{x}_1) & Y_{CH_4}(r_1, \mathbf{x}_2) & Y_{CH_4}(r_1, \mathbf{x}_3) & \cdots & Y_{CH_4}(r_1, \mathbf{x}_M) \\ \vdots & \vdots & \vdots & & \vdots \\ Y_{CH_4}(r_L, \mathbf{x}_1) & Y_{CH_4}(r_L, \mathbf{x}_2) & Y_{CH_4}(r_L, \mathbf{x}_3) & \cdots & Y_{CH_4}(r_L, \mathbf{x}_M) \\ \vdots & \vdots & \vdots & & \vdots \end{bmatrix}. \tag{11}$$

3.3. Constrained PCA

The truncation of the PCA basis may inevitably involve the violation of important physical laws such as the conservation of mass when the observations $\mathbf{y}(\mathbf{x}_j) \ \forall j = 1, \ldots, M$ are reconstructed from the PCA scores. To avoid that, the PCA scores can be evaluated by solving a constrained minimization problem, where the functional to be minimized is the PCA reconstruction error (Xiao et al., 2010). This approach is usually referred to as Constrained PCA (CPCA). The constraints are the physical laws which are intended not to be violated. This minimization problem can be mathematically expressed as:

$$\begin{aligned} &\textit{minimize}: \quad \mathcal{J}\left(\boldsymbol{\gamma}^{(i)}\right) = \frac{1}{2} \left\| \mathbf{y}^{(i)} - \left(\tilde{\mathbf{y}} + \boldsymbol{\Phi}_k \boldsymbol{\gamma}^{(i)}\right) \right\|^2 \\ &\textit{s.t.}: \quad l_j\left(\tilde{\mathbf{y}} + \boldsymbol{\Phi}_k \boldsymbol{\gamma}^{(i)}\right) = l_j\left(\mathbf{y}^{(i)}\right) = 0 \qquad \forall j = 1, \ldots, N_c \end{aligned} \tag{12}$$

where $\mathbf{y}^{(i)}$ represent the vector $\mathbf{y}(\mathbf{x}_i) \ \forall i = 1, \ldots, M$, also introduced in Sections 2, 3.1 and 3.2, $\boldsymbol{\gamma}$ are the CPCA scores (they have been indicated with a different symbol to differentiate them from the PCA scores), $l_j(\cdot)$ is the function related to the jth constraint and N_c is the number of constraints, which can also be inequality constraints. Minimizing the functional $\mathcal{J}(\cdot)$ when no constraints are enforced leads to the PCA scores $\mathbf{a}$.

It is preferable that the solution of this system be not too computationally expensive. In Xiao et al. (2010), the constrained optimization problem has a straightforward solution due to the linearity of the imposed constraints, which allows for a fast evaluation of the CPCA coefficients. If more complex constraints are imposed,

the solution of the constrained optimization problem for the evaluation of the CPCA scores might involve the reconstruction of the considered physical fields (via Eq. (8)) and the use of more expensive optimization algorithms, making the evaluation of the aforementioned coefficients unfeasible.

3.4. Kriging

Accurate prediction of the PCA scores at unexplored points $\mathbf{x}^* \in \mathcal{D}$ in the input parameter space (e.g. inlet temperature, equivalence ratio, etc.) translate into accurate estimation of the original variables as the mapping from $\mathbf{a}(\mathbf{x}^*)$ to $\mathbf{y}(\mathbf{x}^*)$ is known and explained in Section 3.1. The data-set $\mathbf{A} = \{\mathbf{a}(\mathbf{x}_1),\ \mathbf{a}(\mathbf{x}_2),\ \ldots,\ \mathbf{a}(\mathbf{x}_M)\}$ of PCA scores evaluated at different training points, with $\mathbf{a}(\mathbf{x}) = [a_1(\mathbf{x}), a_2(\mathbf{x}), \ldots, a_q(\mathbf{x})]$, is used to build a response surface in the region $\mathcal{D}$ spanned by $\mathbf{X}$.

Kriging is an interpolation method in which every realization $a(\mathbf{x})$, where a is one PCA score indicated with no subscript for brevity, is expressed as a combination of a trend function and a residual (Constantine et al., 2014):

$$a(\mathbf{x}) = \mu(\mathbf{x}) + z(\mathbf{x}) = \sum_{i=0}^{p} \beta_i f_i(\mathbf{x}) + z(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + z(\mathbf{x}) \tag{13}$$

The trend function $\mu(\mathbf{x})$ is a low-order polynomial regression and provides a global model in the input space. The term $z(\mathbf{x})$ 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 $z(\mathbf{x})$ are modeled by a Gaussian process with a kernel or correlation function that depends on a set of hyper-parameters $\boldsymbol{\theta}$ to be evaluated by Maximum Likelihood Estimation (MLE). Many possible correlation functions are available: linear, quadratic, exponential, Gaussian, Matern 3/2, Matern 5/2, just to name a few (Constantine et al., 2014; Lophaven et al., 2002). A detailed discussion about these functions can be found in Seeger (2004). One of the main differences among these kernels is their smoothness.

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, the designer has the possibility to add prior knowledge into the problem and subsequently let the data speak for themselves by estimating the hyper-parameters $\boldsymbol{\theta}$.

3.5. Adaptive sampling

The sampling strategy employed to explore the region $\mathcal{D}$ of the input space can affect the construction of a PCA+Kriging-based ROM. The construction of high-performing ROMs with a very limited number of samples is possible if an effective sampling strategy is developed. As a PCA-based model strongly depends on its modal basis, a first step towards the improvement of this kind of model consists in improving the basis (Guenot et al., 2013). Given a set of (centered-scaled) observations $\mathbf{Y} = \{\mathbf{y}^{(1)},\ \ldots,\ \mathbf{y}^{(M)}\}$ and its corresponding PCA-based model, we want to choose a new sample point, $\mathbf{x}_{\text{new}}$, that will meet the trade-off between the modal basis improvement and the parametric space exploration. Firstly, the influence of each observation on the modal basis is computed. The influence of the jth observation on the ith mode is defined by:

$$\text{Infl}_{\phi_i}(\mathbf{x}_j) = \frac{1}{|\boldsymbol{\phi}_i^T \boldsymbol{\phi}_i^{-j}|} - 1, \tag{14}$$

where $\boldsymbol{\phi}_i^{-j}$ is the ith basis function evaluated from a data set:

$$\mathbf{Y} = \{\mathbf{y}^{(1)},\ \ldots,\ \mathbf{y}^{(j-1)},\ \mathbf{0},\ \mathbf{y}^{(j+1)},\ \ldots,\ \mathbf{y}^{(M)}\}. \tag{15}$$

The influence of the observation $\mathbf{y}(\mathbf{x}_j)$ on the modal basis is defined by:

$$\text{Infl}_{\text{Basis}}(\mathbf{x}_j) = \sum_{i=1}^{K} s_i \text{Infl}_{\phi_i}(\mathbf{x}_j), \tag{16}$$

where s_i is the singular value of the ith mode. The relative influence of the jth observation on the modal basis is given by:

$$\text{Infl}^{\text{Rel}}_{\text{Basis}}(\mathbf{x}_j) = \frac{\text{Infl}_{\text{Basis}}(\mathbf{x}_j)}{\sum_{l=1}^{M} \text{Infl}_{\text{Basis}}(\mathbf{x}_l)}. \tag{17}$$

After the computation of this equation for each $\mathbf{x}_j$, the parametric space is heavily sampled via a LHS technique and the resulting set of samples is denoted by $\mathbf{Q} = \{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_b\}$. The size b of this set of samples can be chosen as 100 times the parametric space dimension. Then the potential of enrichment $\text{PotBasis}(\boldsymbol{\nu}_i)$ of each candidate sample is computed with respect to the trade-off between the input space exploration and the improvement of the modal basis as:

$$\text{PotBasis}(\boldsymbol{\nu}_i) = d\left(\boldsymbol{\nu}_i, \mathbf{x}_j\right) \text{Infl}^{\text{Rel}}_{\text{Basis}}(\mathbf{x}_j), \tag{18}$$

where $j = \texttt{argmin}_k\ d(\boldsymbol{\nu}_i, \mathbf{x}_k)$ with $d(\cdot,\ \cdot)$ denoting the Euclidean distance. Finally, the new point will be selected to fulfill the following condition:

$$\mathbf{x}_{\text{new}} = \texttt{argmax}_{\boldsymbol{\nu} \in \mathbf{Q}} \text{PotBasis}(\boldsymbol{\nu}). \tag{19}$$

A new sample is chosen as far from the other samples as possible, but at the same time as close to the samples with the highest relative influence as possible. This sampling methodology is named Adaptive Sampling (AS).

4. Results

4.1. 1D flame with one input parameter

The Kriging-PCA approach was tested on a 1D methane/air laminar flame. OpenSMOKE++ (Cuoci et al., 2013a; 2013b) was used to produce a data-set of 21 observations $\{\mathbf{y}^{(1)},\ \mathbf{y}^{(2)},\ \ldots,\ \mathbf{y}^{(M)}\}$ of methane/air flames with GRI 3.0 mechanism for different values of the equivalence ratio ψ, in the range 0.5–2, with a step of 0.05. A subset of size $M = 21$ of the total 31 observations was used as training data-set to build a ROM and referred to as *training* observations or data. The corresponding values of ψ for the training observations are named training points. The 21 training observations were chosen using the Adaptive Sampling technique described in Section 3.5. The remaining 10 observations were used as test data to assess the accuracy of the reduced model. These 10 observations are referred to as *validation* data. The objective was to predict the validation data with acceptable accuracy.

In this paper, the term *reconstruction* is used to indicate the reconstruction of the training data after the compression achieved by PCA, whilst the term *prediction* is used when testing the predictive capabilities of the developed ROM on the validation data.

Each observation $\mathbf{y}^{(i)}$ of the data-set $\mathbf{Y} = \{\mathbf{y}^{(1)},\ \mathbf{y}^{(2)},\ \ldots,\ \mathbf{y}^{(M)}\}$ was a vector of $N = 10,780$ variables: 53 chemical species plus temperature and axial-velocity evaluated at 196 grid points. The size of the matrix $\mathbf{Y}$ was $N \times M = 10,780 \times 21$. Because $N > M$, PCA could find at most $M - 1 = 20$ PCs. Alternatively said, PCA could encode the 10 vectors $\mathbf{y}_m \in \mathbb{R}^N$ contained in the data matrix $\mathbf{Y}$ into a set of at most 20 scalars $\mathbf{a}_m \in \mathbb{R}^{20}$. The data-set was scaled according to the auto-scaling criterion. This criterion was chosen as other criteria tend to prioritize some variables over the others as

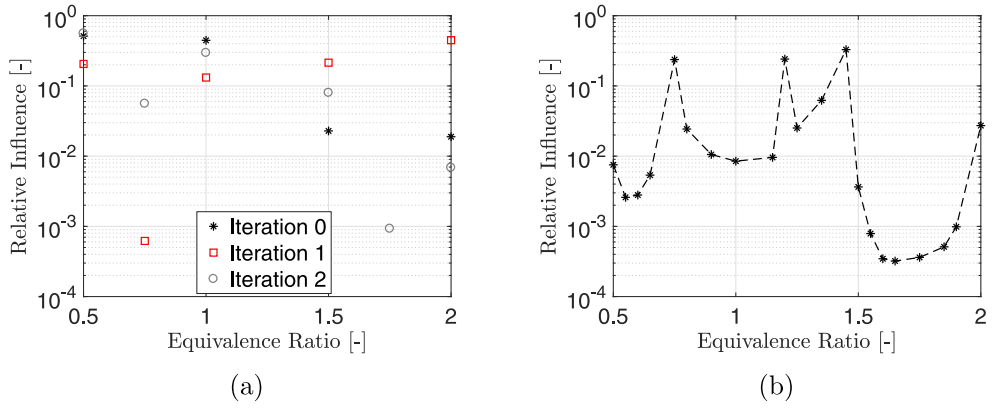

Fig. 4. (a) Starting from 4 initial samples, the relative influence of each sample on the modal basis was evaluated and a new training point was chosen (as explained in Section 3.5). Then, the relative influence of the 5 samples was evaluated again, and a 6th training point was chosen, and so on. (b) Relative influence of each observation on the modal basis evaluated according to Section 3.5 once that a number of 21 total training observation was reached.

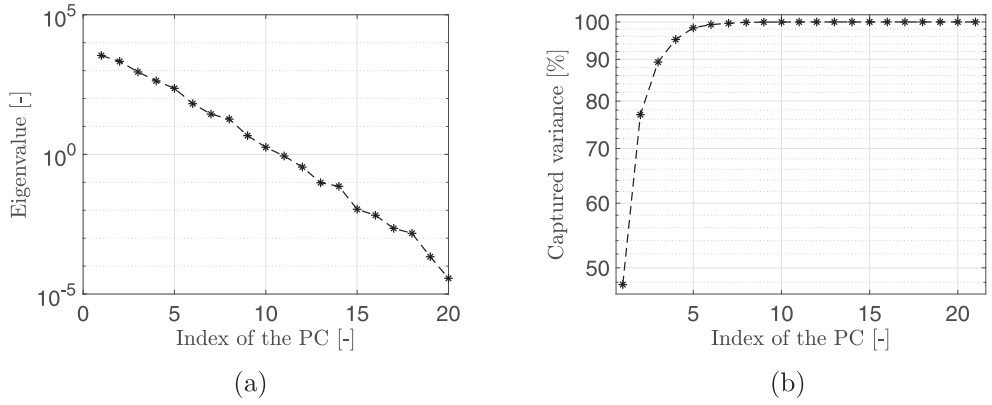

Fig. 5. (a) The spectrum of the eigenvalues associated to each PC provides a criterion for sorting the PCs in descending order as these eigenvalues can be interpreted as the importance of the PC they correspond to. Auto-scaling. (b) The cumulative original data variance that is captured when adding more PCs provides a criterion for the selection of the number of PCs to keep when using global PCA. Auto-scaling. One-parameter case.

shown in Parente and Sutherland (2013). In Fig. 4a, an online evaluation of the relative influence of each *current* training observation on the modal basis (evaluated according to Section 3.5) is reported. Four samples were present at the start (iteration 0), and the 5th training sample was chosen in the region of high values for the equivalence ratio as the already present training observations in that region had a higher relative influence. Once this observation was added to the training data-set, the relative influence of each sample was re-evaluated (iteration 1) and a 6th training observation was chosen. Thus, the relative influence of the new set of samples was re-evaluated (iteration 2). And so on, until the desired number of training samples was reached. In Fig. 4b, the relative influence of each observation on the modal basis is shown, after that 21 training observations have been chosen. Note that the relative influence of each snapshot changes when new observations are introduced.

Fig. 5a shows the eigenvalue spectrum of the PCs found by PCA. Fig. 5b shows the cumulative original data variance that was captured when retaining more PCs. Interestingly, a few number of PCs were enough to recover most of this variance. As the maximum number of PCs that could be found was limited by the number of training observations, the cumulative recovered original data variance, reported in Fig. 5b, could also be used to determine if the number of available training observations was enough for the ROM development. In this case, 5 PCs could recover 98.4% of the original data variance and 10 PCs 99.8%, thus suggesting that a number of 21 training observations could be considered sufficient. In Fig. 6a, the R^2 values for the reconstruction of the training data are reported for different numbers of retained PCs, q. These values, logically, increased with the approximation order, q, because more of the original data variance was accounted for by the reduced basis. Analysis on the R^2 values and, thus, on the reconstruction error is a good estimate about the amount of information lost due to compression, and provides a criterion for the choice of the number of required PCs. Fig. 6b shows the R^2 values for the prediction of the validation data by means of PCA+Kriging. These values are reported for two PCA+Kriging models with a number of 5 PCs retained: with a linear and with a quadratic trend function. Both models used a Matern52 kernel. From these results, we can conclude that the PCA+Kriging model with a linear trend function performed better in terms of predicting capabilities for the present data-set.

In Figs. 7a and b, the effects of the choice for the approximation order q and the choice of different trend functions and kernels for the Kriging interpolation are investigated. A reduction on prediction errors could be observed when increasing the approximation order q. Differences with the data (prediction errors) were still present due to Kriging interpolation and the fact that the data used for validation were not included in the data used to find the PCA modes. For the evaluation of these errors, the validation data-set and the one predicted by the model were scaled (range-scaling) and their difference was evaluated and stored in a new matrix. A mean value for each observation in this matrix was then used as prediction error. The formula for this error is:

$$Err(\mathbf{x}_i) = \frac{1}{N}\sum_{j=1}^{N}\frac{y_j(\mathbf{x}_i) - \tilde{y}_j(\mathbf{x}_i)}{y_j^{max} - y_j^{min}}, \tag{20}$$

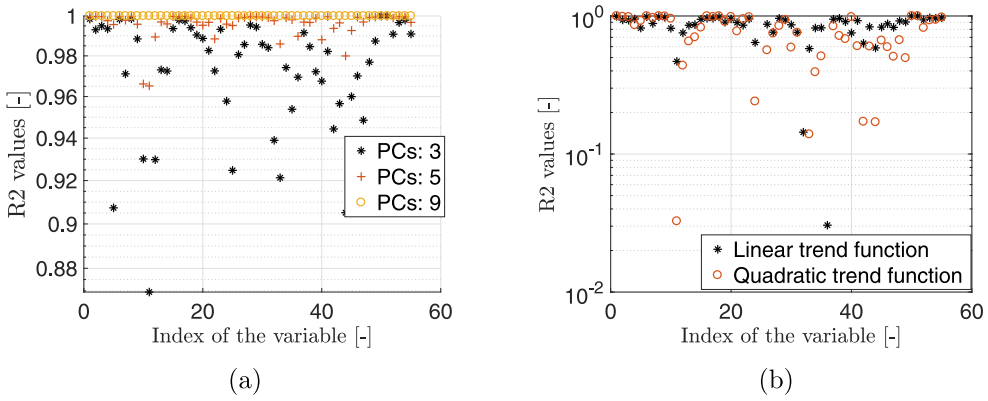

Fig. 6. (a) R-squared values for the reconstruction of the original data by PCA, for different numbers of retained PCs. (b) R-squared values for the prediction of the original data with two different trend functions for Kriging. 5 PCs retained. Matern52 correlation function. One-parameter case.

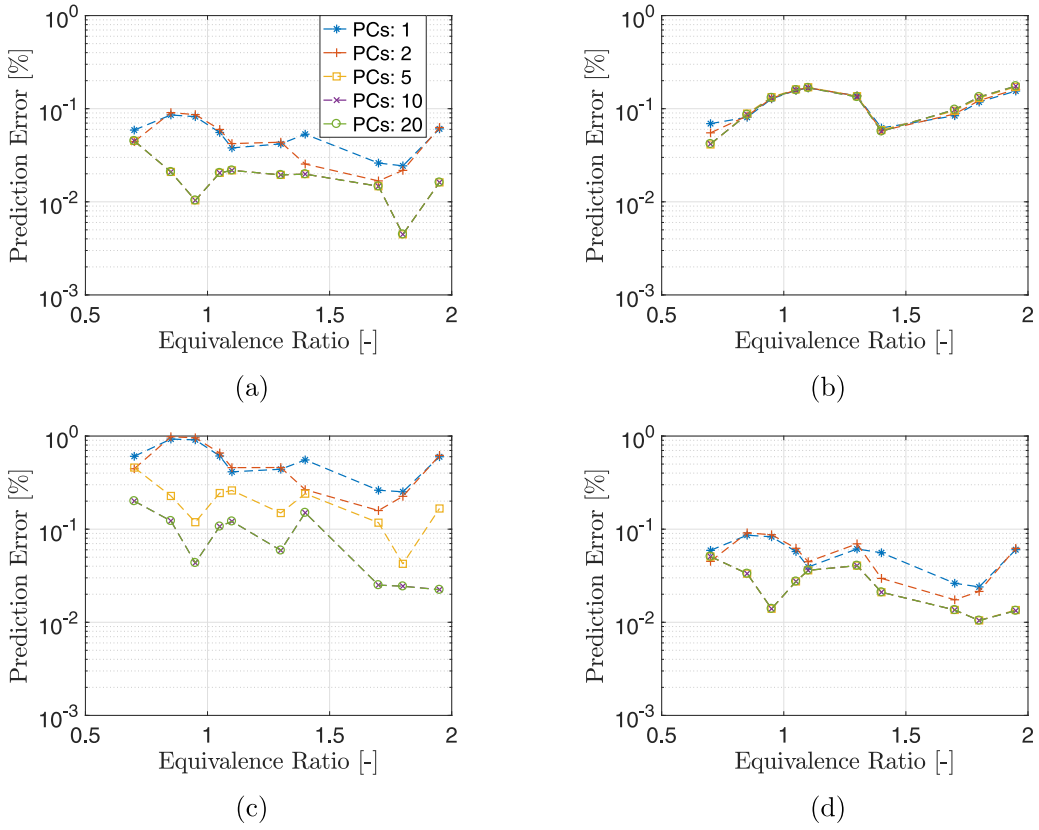

Fig. 7. Prediction errors for different numbers of retained PCs. Global PCA+Kriging. (a) Constant trend function and exponential kernel. (b) Linear trend function and Gaussian kernel. (c) Quadratic trend function and Matern12 kernel. (d) Linear trend function and Matern52 kernel. One-parameter case.

where $\mathbf{x}_i$ is the ith prediction point, N is the number of variables, $y_j(\mathbf{x}_i)$ is the true value of the jth variable for the ith prediction point and $\tilde{y}_j(\mathbf{x}_i)$ is the ROM's prediction for the same variable, for the same prediction point. These errors were below a value of 4% for all observations when $q > 1$. Different choices of trend and correlation function for the Kriging interpolation led to different results. For example, the choice of a Gaussian kernel (see Fig. 7b) for a PCA+Kriging model for this data-set could not reduce the prediction errors further, when increasing q from 2 to 5 or 9. Indeed, although increasing q means accounting for more of the original data variance, it also means that more scalars have to be interpolated.

Next, the effects of performing PCA in separated clusters (local PCA) on the quality of the reduced model was investigated. The effects of the choice of the scaling criterion were also examined. Fig. 8a reports the data reconstruction errors when increasing the number of clusters, for different scaling criteria. Increasing the number of clusters translated into better data reconstructions. The auto-scaling and the Pareto scaling criteria emerged as the ones with the lowest reconstruction errors. The general trend was a decreasing reconstruction error as the number of clusters was increased. Fig. 8b reports the local PCA reconstruction error when the number of retained PCs was increased (auto-scaling), for different number of clusters. It can be observed that the impact of cluster addition on the reconstruction error became less and less significant as the number of clusters increased, indicating the existence of a trade-off, depending on the problem dimensionality and non-linearity. Prediction errors also decreased when adding clus-

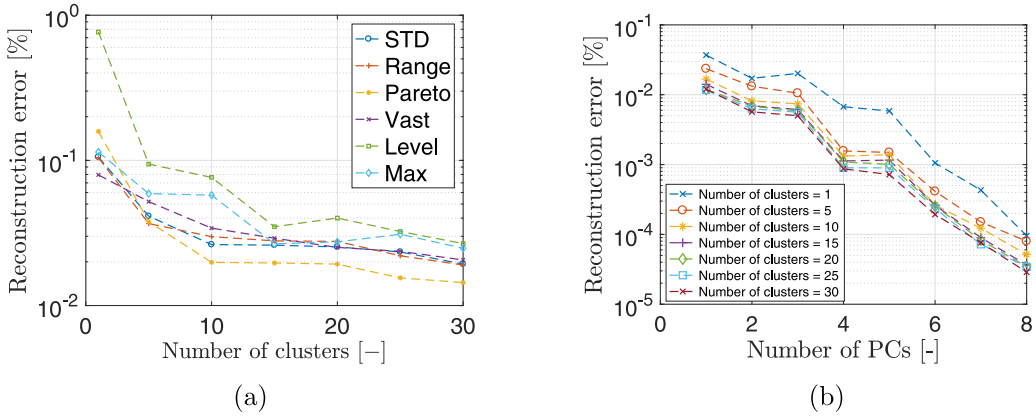

Fig. 8. Local PCA reconstruction errors. (a) Reconstruction errors decreases as the number of clusters used is increased. (b) Reconstruction errors decrease as more PCs are retained (auto-scaling criterion). One-parameter case.

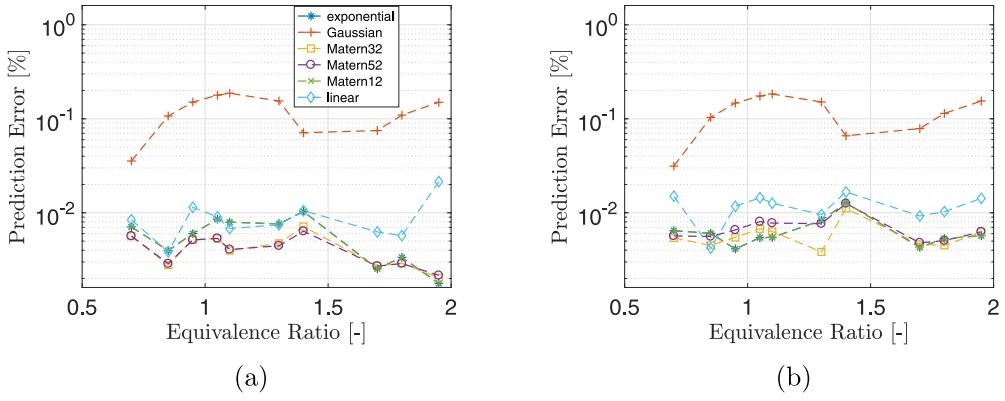

Fig. 9. Local PCA prediction errors for $q = 2$. (a) 5 clusters, auto-scaling, constant trend function. (b) 10 clusters, auto-scaling, constant trend function. One-parameter case.

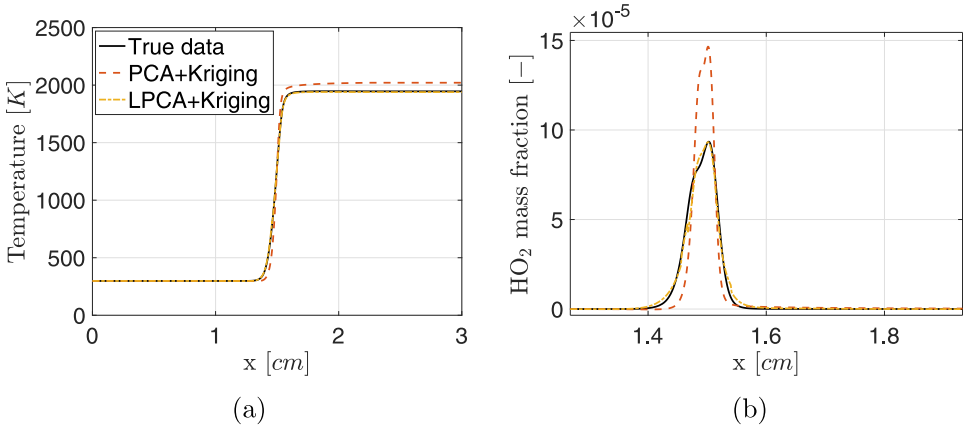

Fig. 10. Prediction of the spatial profile of temperature and HO_2 mass fraction. Comparison between PCA and LPCA (15 clusters) combined with Kriging. One-parameter case.

ters in the local PCA formulation. This can be observed in Figs. 9a and b. Local PCA, in combination with Kriging, provided better performances, in terms of both reconstruction at the training points and predictions. Moreover, the gain in accuracy for the prediction points is a clear proof that also Kriging benefits from the local formulation. Fig. 10 reports the predictions of the spatial profiles of temperature and HO_2 mass fraction by PCA+Kriging and local PCA+Kriging, showing the superior performances of local PCA over PCA.

The clustering strategy explained in Section 3.2 was used to group the rows of the data matrix $\mathbf{Y} \in \mathbb{R}^{N\times M}$ into clusters. The general encoding process for one observation in the data matrix $\mathbf{Y}$, which can be stated as $\mathbf{y} \in \mathbb{R}^N \rightarrow \mathbf{a} \in \mathbb{R}^q$ with $q \ll N$, was possible via the detection of a lower-dimensional manifold spanned by the $M = 21$ observations $\mathbf{y}_m \in \mathbb{R}^N \; \forall \; m = 1, \ldots, M$. The dimension of this lower-dimensional manifold was thus $M - 1 = 20$. A new observation $\mathbf{y}^*$ not already present in the data-set was approximated by a point on the lower-dimensional manifold even if there is no

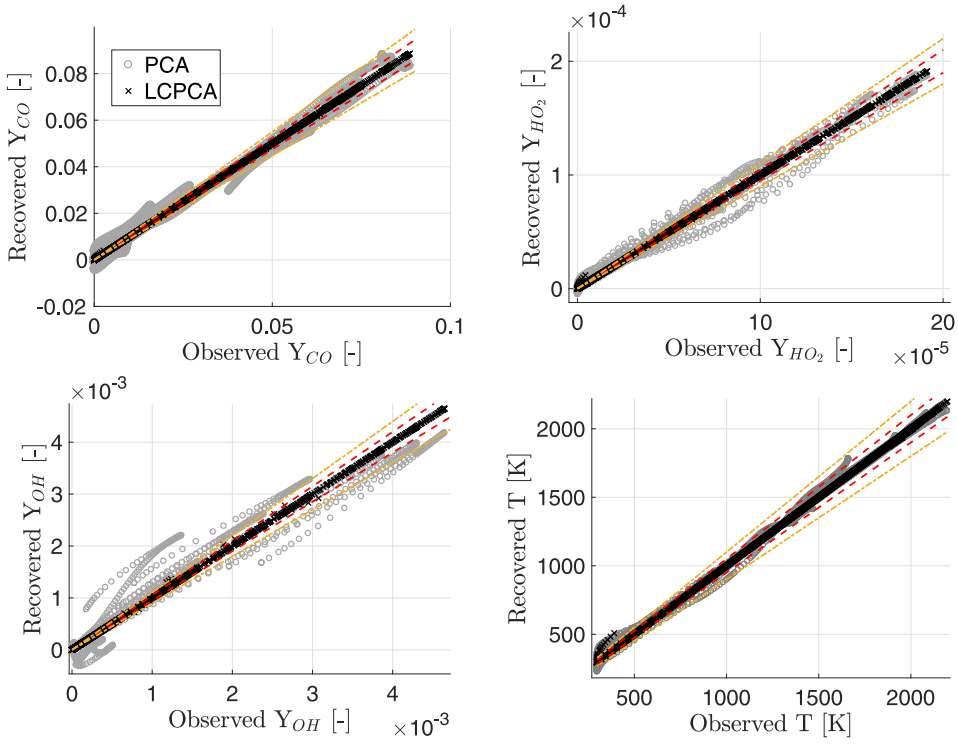

Fig. 11. Parity plots for the reconstruction of the CO, HO_2, OH mass fractions and temperature. PCA: 3 PCs retained. 20 clusters for Local CPCA. Dashed line: 5% error. Dotted line: 10% error. One-parameter case.

guarantee that this new observation lay *close* to this manifold. The clustering process reduced the number of variables for each cluster. This means we were moving closer to a situation where the M samples were enough for the detection of a manifold which could accurately approximate unseen observations.

The PCA-reconstructed mass fraction spatial profile included negative values as shown in the parity plots (Fig. 11). This is a known issue with PCA (Isaac et al., 2015; Parente et al., 2011) since the method itself cannot guarantee that physical constraints are verified, such as the positivity of mass fractions and the conservation of mass ($\sum Y_i = 1$). An increase of the approximation order could indirectly lead to a solution to this issue, but it is here interesting to investigate if the implementation of a constrained version of PCA (CPCA) could guarantee the non-violation of the aforementioned physical laws, even at high compression. As discussed in Section 3.3, the CPCA implementation used in the present work consisted in forcing each rebuilt mass fraction, on every grid point, to be positive.

However, one of the challenges of using CPCA in combination with Kriging is that the chemical species fields recovered from the interpolated CPCA scores might still violate the set of imposed constraints. Even if for each training point a constrained optimization problem is solved for the evaluation of the CPCA scores, no constraint is imposed when these coefficients are interpolated by means of Kriging. Thus, there is no guarantee that the predicted CPCA scores belong to the feasible region delimited by the constraints $l_j(\boldsymbol{\gamma}) = 0$ of (12). CPCA in combination with local PCA guaranteed the satisfaction of the physical constraint, since the reconstruction accuracy improved, as shown in Figs. 11 and 12. Not only the imposed constraint was satisfied by the LCPCA+Kriging model's predictions, but a very satisfactory level of accuracy was also achieved as indicated by the parity plots in Fig. 13, with most of the predicted values being within the 5% or 10% error lines. The R^2 values for the prediction of the validation data, as the number

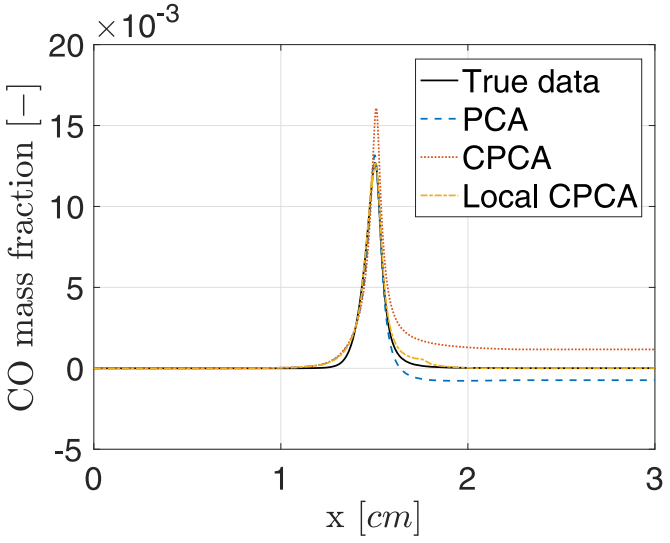

Fig. 12. Reconstruction of the spatial CO mass fraction profile. Comparison between PCA, CPCA and LCPCA (15 clusters). One-parameter case.

of training samples increased, are reported in Fig. 14. These values increased as more observations were employed for the training of the model, as expected.

The results presented in this section showed that different choices for the Kriging kernels, PCA formulation (standard, local and constrained) affect the ROM's predictive capabilities.

4.2. 1D flame with two input parameters

The methodology explained in Section 2 was tested on the same system presented in Section 4.1, using two input parameters.

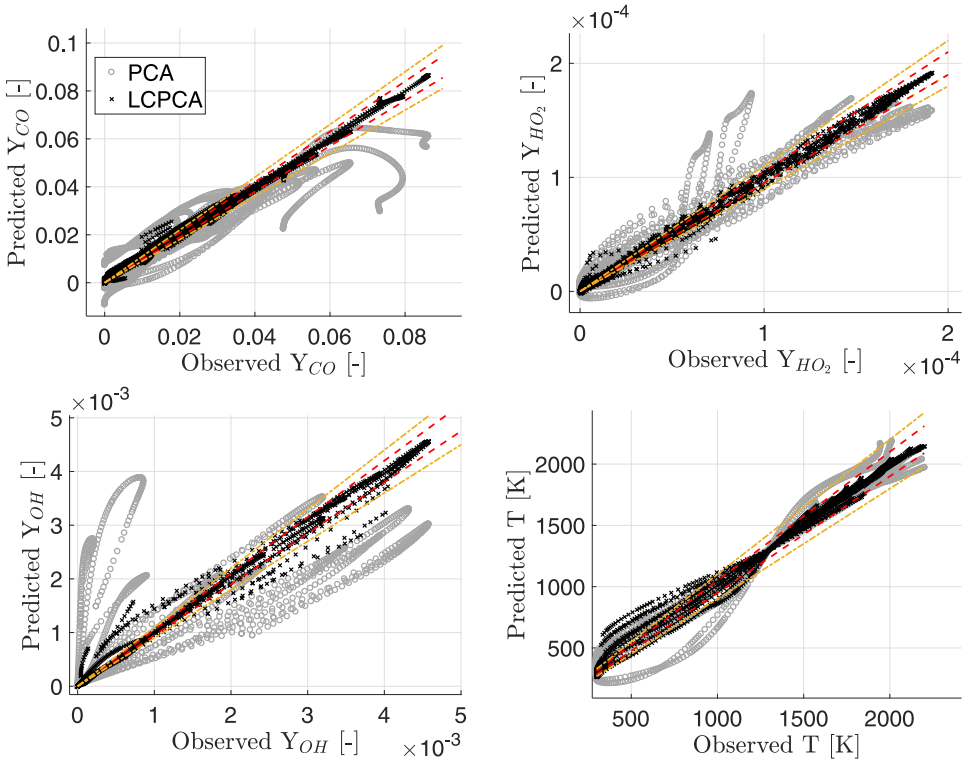

Fig. 13. Parity plots for the predictions of the CO, HO_2, OH mass fractions and temperature. PCA: 3 PCs retained. 50 clusters for local PCA. Kriging: linear trend function, Matern52 kernel. Dashed line: 5% error. Dotted line: 10% error. One-parameter case.

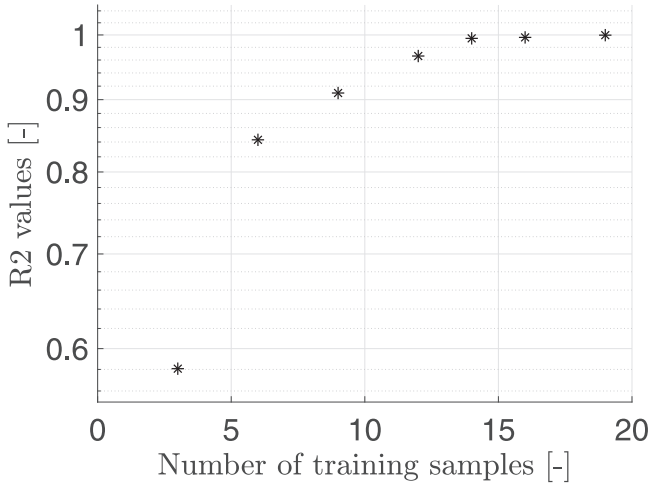

Fig. 14. Global R^2 values for the prediction of the validation data, as the number of training observations increases. Local PCA+Kriging with linear trend function and Matern12 kernel.

A molar fraction of N_2 was introduced in the inlet stream. A total of 147 observations were generated by OpenSMOKE++ in the range 0.5–1.5 for the equivalence ratio and 0.4–1.0 for the molar fraction of N_2. A set of 75 observations was selected with the AS strategy explained in Section 3.5 for the training of the SM. The remaining observations were used as test data. Fig. 15a reports the eigenvalue spectrum of the PCs, for different scaling criteria used. Fig. 15b shows the cumulative original data variance that was captured when more PCs were retained in the reduced basis. This is reported for different scaling criteria applied to the data-set. As for the one-parameter case, most of the total data variance was concentrated on the first few PCs, indicating that PCA was a valid strategy for the reduction of the problem size. Besides, considered that 75 training observations were used, a number of 15 PCs could recover over 99% of the original data variance, indicating that the chosen number of training observations was sufficient, as the prediction errors for this case will indicate.

Fig. 16a shows the prediction errors for 72 validation points, using PCA+Kriging with 20 PCs, a constant trend function and exponential kernel. The prediction errors for a local PCA+Kriging with 50 clusters and direct Kriging (applied directly on the original variables without the use of PCA) are also reported. The prediction errors for the three models are shown in Fig. 16b, replacing the exponential kernel with the Gaussian one. The results confirm that data clustering and the kernel type have a significant impact on prediction errors. Interestingly, direct Kriging provided reconstruction errors higher than the ones obtained combining Kriging with (local) PCA, showing the advantage associated to reducing the number of scalars to regress. In Fig. 17, the parity plots for the predicted values of temperature and CH_4, OH, CO mass fractions are shown. The number of retained PCs was 20, corresponding to a recovered original data variance of 99.78% with the auto-scaling criterion, while the number of clusters was 50. The figure shows that the PCA+Kriging methodology performed efficiently, with clear improvements when the data are clustered. The sum of mass fractions of all chemical species added up to 1 with a maximum error of 4×10^{-3}.

The use of CPCA was also investigated for this case. In Fig. 18a, the rebuilt spatial profiles for CO by means of PCA, CPCA and LCPCA are reported and compared. While the reconstructed CO spatial profile showed negative values when rebuilt by PCA, CPCA provided a non-negative spatial profile. Fig. 18b shows that such a constraint was not violated even after Kriging interpolation,

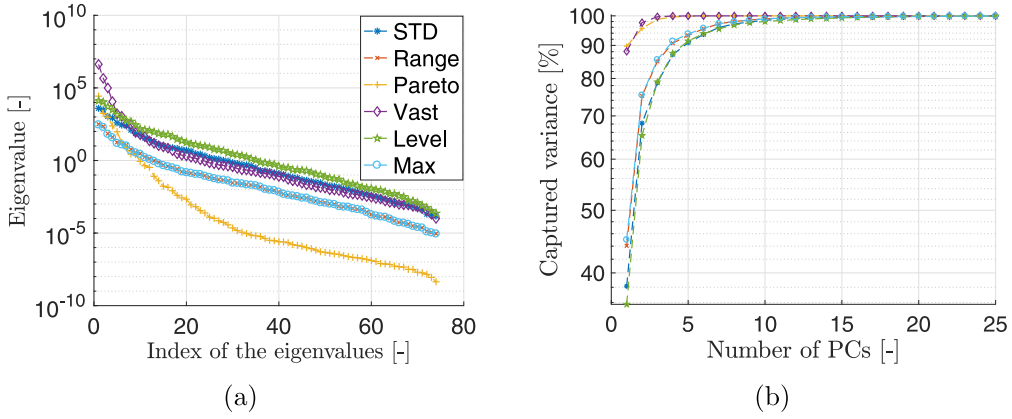

Fig. 15. (a) The spectrum of the eigenvalues associated to each PC provides a criterion for sorting the PCs in descending order as these eigenvalues can be interpreted as the importance of the PC they correspond to. (b) The cumulative original data variance that is captured when adding more PCs provides a criterion for the selection of the number of PCs when using global PCA. Two-parameters case.

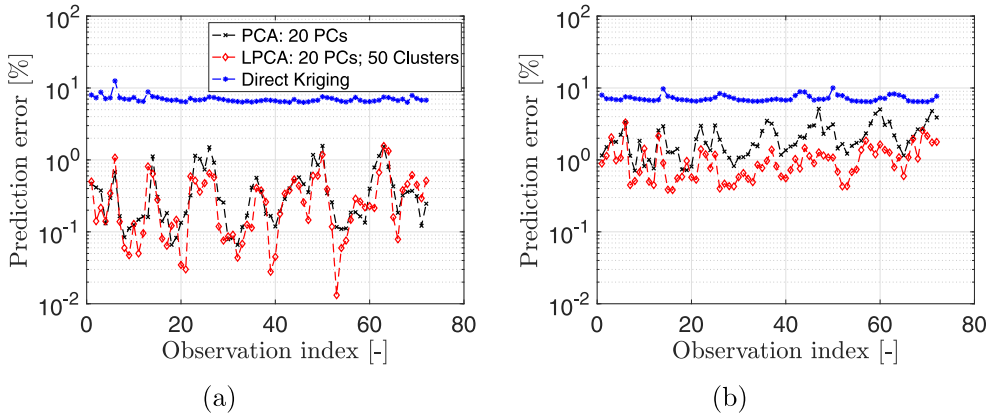

Fig. 16. Prediction errors reported for 3 different SMs. Auto-scaling. Kriging: (a) constant trend function and exponential kernel; (b) constant trend function and Gaussian kernel. Two-parameters case.

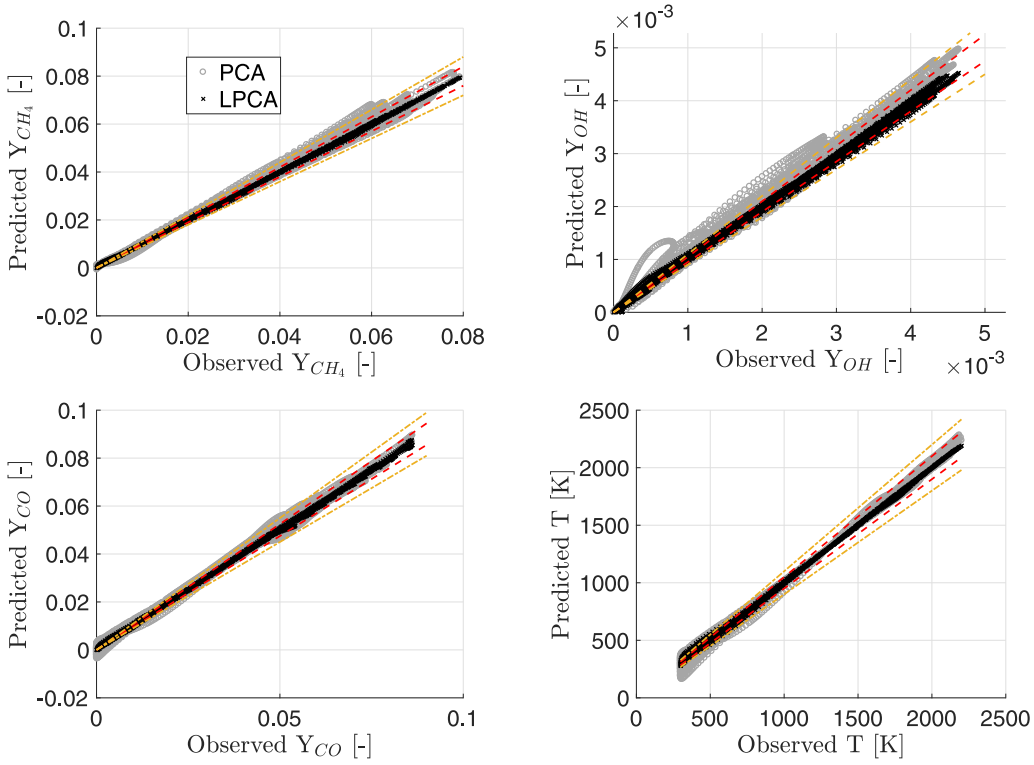

Fig. 17. Parity plots for the prediction of temperature and of the CH_4, OH, CO mass fractions. PCA: 20 PCs retained. Local PCA: 50 clusters. Kriging: linear trend function, Gaussian kernel. Dashed line: 5% error. Dotted line: 10% error. Two-parameter case.

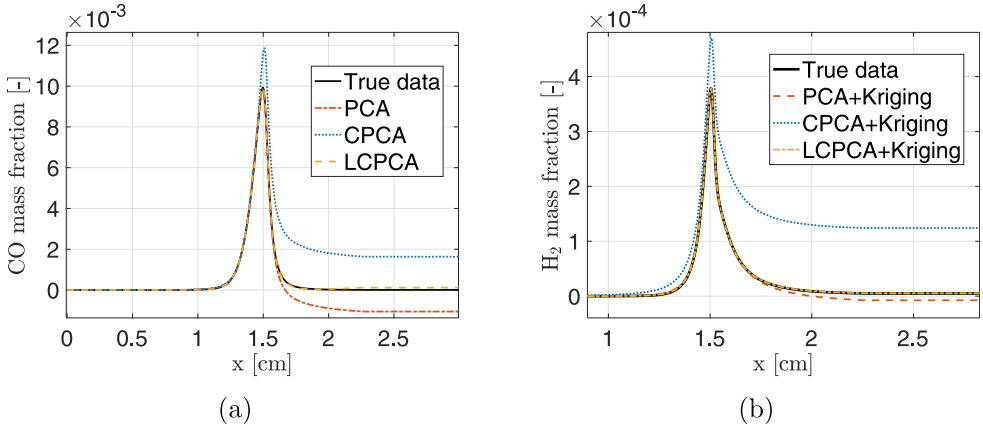

Fig. 18. (a) Reconstruction of the spatial CO mass fraction profile. Comparison between PCA, CPCA and LCPCA (50 clusters). (b) Prediction of the spatial H2 mass fraction profile. Comparison between PCA, CPCA and LCPCA (50 clusters) combined with Kriging. Two-parameter case.

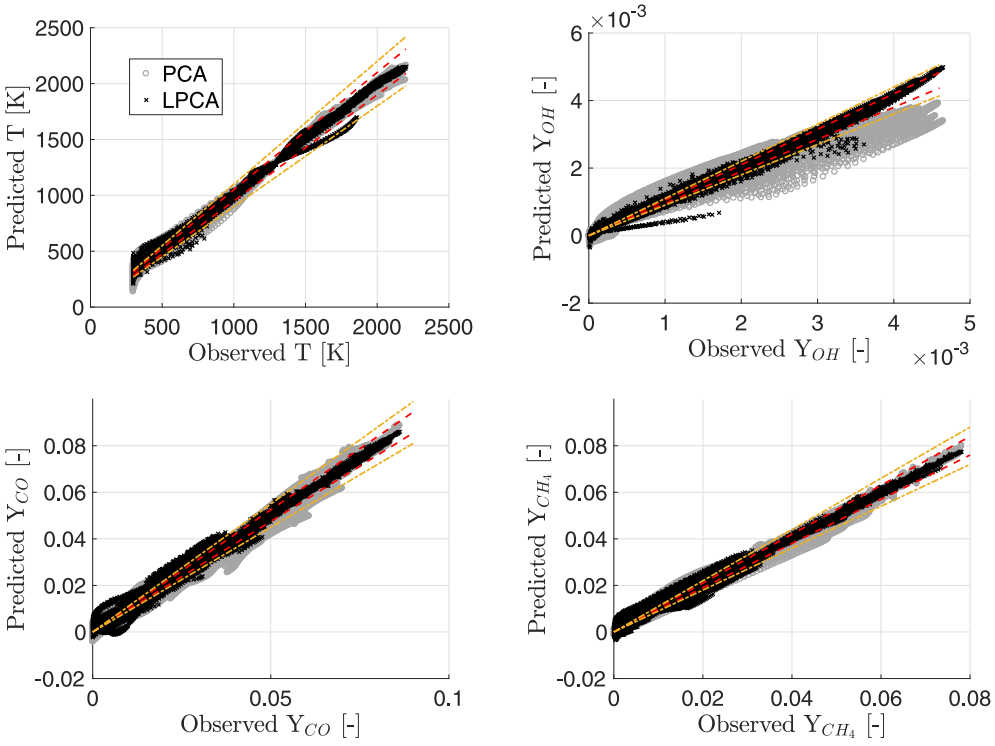

Fig. 19. Parity plots for the predictions of temperature and the OH, CO, CH_4 mass fractions, using 46 training observations. PCA: 30 PCs retained. Local PCA: 90 clusters. Kriging: linear trend function, Gaussian kernel. Dashed line: 5% error. Dotted line: 10% error. Two-parameter case with less training observations.

meaning that the interpolated values for the CPCA scores remained within the feasible region delimited by the constraints indicated in (12). Finally, the spatial profiles provided by the CPCA+Kriging showed significant deviations in the post-reaction zone, while the reconstruction accuracy was significnalty improved using the local PCA formulation.

As explained, most of the data variance was recovered by a small number of PCs, i.e. 15, in comparison to the number of available training samples, i.e. 75. In order to further test the proposed methodology, and to be sure that the observed accuracy was not to due to the high number of samples used for training, another ROM was developed with a smaller number of training observations, i.e. 46. The parity plots for the prediction of temperature and OH, CO, CH_4 mass fractions are shown in Fig. 19. It is clear that even reducing the number of training points, the accuracy of predictions was very satisfactory, indicating that the analysis based on the cumulative data variance recovered by the PCs is an efficient indicator for the amount of information fed to the model by the training samples.

The promising results obtained from the application of the proposed methodology for the one-parameter case, were confirmed with two input parameters. The results confirmed that the combination of PCA, and its variants, with Kriging is a valid choice for the development of reduced-order surrogate models. In the next section, the complexity of the case is increased by adding a third input parameter.

4.3. 1D flame with three input parameters

After validating the predictive capabilities of a ROM developed on the combination of PCA and Kriging on systems with one or two input parameters, one more parameter was added to the system, namely the inlet stream temperature in the range 298–798 K.

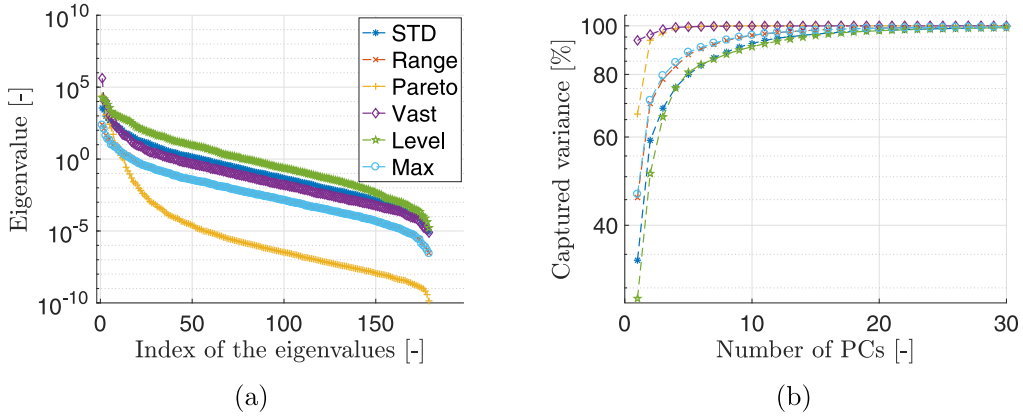

Fig. 20. (a) The spectrum of the eigenvalues associated to each PC provides a criterion for sorting the PCs in descending order as these eigenvalues can be interpreted as the importance of the PC they correspond to. Reported for different scaling criteria. (b) The cumulative original data variance that is captured when adding more PCs provides a criterion for the selection of the number of PCs when using global PCA. Reported for different scaling criteria. Three-parameters case.

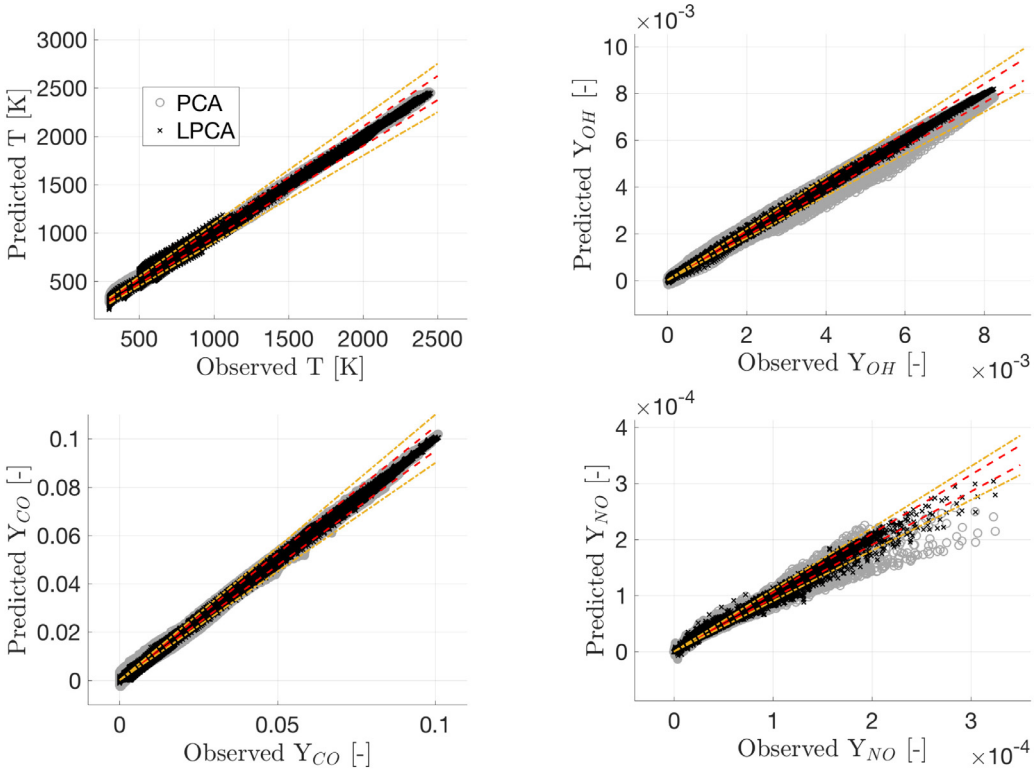

Fig. 21. Parity plots for the prediction of temperature and OH, CO, NO mass fractions. PCA: 48 PCs. LPCA: 20 clusters, 48 PCs. Kriging: quadratic trend function, Matern52 kernel. Dashed line: 5% error. Dotted line: 10% error. Three-parameter case.

Two sets of observations were again generated by OpenSMOKE++. A set of 180 solutions, selected using the AS strategy, was used to train the model. The other set (702 solutions) was used for the validation of the ROM's predictions. Figs. 20a and b show the eigenvalue spectra and original data variance recovery for the different scaling criteria, respectively. Over 99% of the original data variance was captured by 28 PCs. Fig. 21 shows predictions of temperature, CO, OH and NO mass fractions for the PCA+Kriging (48 PCs) and LPCA+Kriging (20 clusters) models, with a quadratic trend function and Matern52 kernel. The application of Local PCA+Kriging notably improved the predictions for some species like OH and NO, whereas good predictions of temperature and CO were already achieved by PCA+Kriging.

The application on the 1D flame allowed to test the potential and limitations of the PCA+Kriging approach. For this reason, it is interesting to test the proposed methodology on a 2D flame with detailed chemistry and transport phenomena.

4.4. 2D flame with two input parameters

The methodology described in Section 2 was applied to a multidimensional flame in more complex configurations in order to test its potential. The configuration of the simulated flame is described in Cao et al. (2013) and in Cao et al. (2015). The computational domain started from the exit of the nozzle and extended 122 mm further downstream. The radial direction was expanded to 42.88 mm. A 2D structured axi-symmetric mesh with around 25,600 cells was used and the nozzle radius was resolved with 12 cells. The laminarBuoyantSimpleSMOKE solver was applied. Gravity was turned on. The multi-component diffusion model was adopted

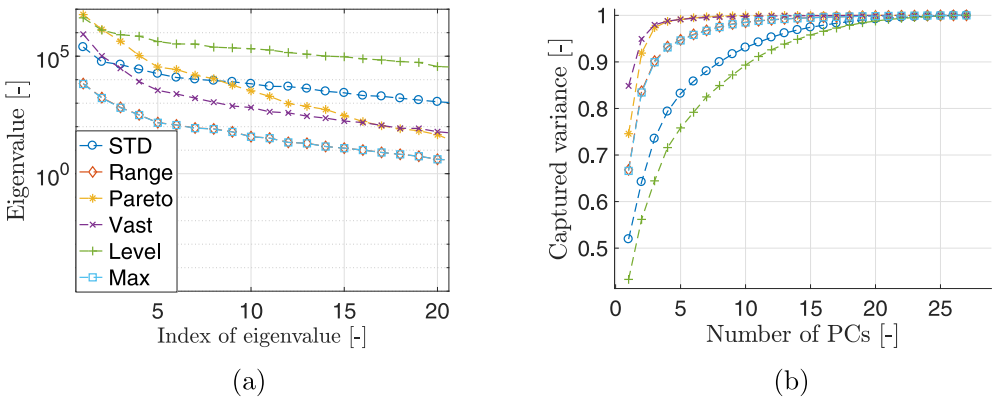

Fig. 22. (a) The spectrum of the eigenvalues associated to each PC provides a criterion for sorting the PCs in descending order as these eigenvalues can be interpreted as the importance of the PC they correspond to. Reported for different scaling criteria. (b) The cumulative original data variance that is captured when adding more PCs provides a criterion for the selection of the number of PCs when using global PCA. Reported for different scaling criteria. 2D flame.

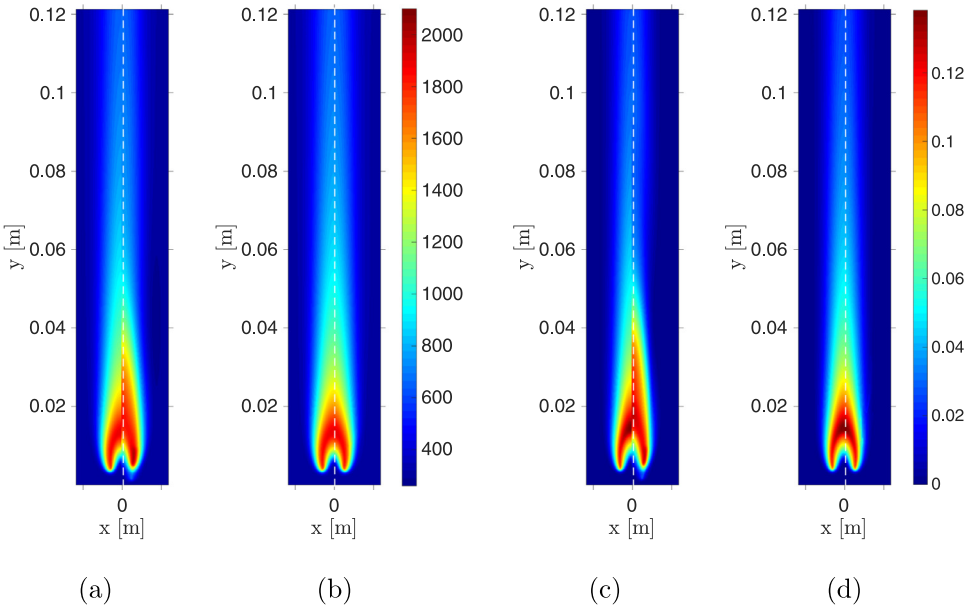

Fig. 23. Left halves: true data. Right halves: (a) reconstruction of the temperature field by PCA; (b) reconstruction of the temperature field by local PCA; (c) reconstruction of the CO_2 mass fraction field by PCA; (d) reconstruction of the CO_2 mass fraction field by local PCA. Inlet velocity: 24 cm/s; inlet CH_4 molar fraction: 40%. 2D flame. PCA: 5 PCs. Local PCA: 100 clusters.

to consider molecular diffusion. The GRI3.0 mechanism without NO_x (35 species and 219 reactions) was chosen. For the velocity boundary condition, the profile provided from Cao et al. (2013) and in Cao et al. (2015) was used. The input parameters were two, namely the inlet velocity and the molar fraction of CH_4 in the inlet stream, which was a mixture of CH_4 and N_2. A total of 30 samples was produced by OpenFOAM, spanning the two input parameters in the range 24–89 cm/s and 40–100% for the inlet velocity and inlet molar fraction of CH_4, respectively. A total of 36 physical variables was considered: 35 chemical species and temperature. 25 observations were chosen to train a PCA+Kriging model using the adaptive strategy described in Section 3.5. Fig. 22a reports the eigenvalue spectrum of a PCA performed on this data-set, for different scaling criteria. Fig. 22b shows the cumulative original data variance that was recovered when more PCs were retained. The recovery of 99% of the original data variance was achieved with 20 PCs for all scaling criteria. This indicated that, despite the high dimensionality of the system, recurrent structures could be found in the data. At the same time, the fact that 20 PCs were needed to recover 99% of the data variance, and that 25 observations were used for the training of the model, also suggested that more training observations were needed, or that a narrower parameter region should have been explored, as we shall see in Section 4.5. Fig. 23 shows the temperature field for an inlet velocity of 24 cm/s and an inlet CH_4 molar fraction of 40%. The original field is shown on the left half of the figures, while PCA and local PCA reconstructions are shown on the right half of the figure. PCA was performed by retaining 5 PCs, while a number of 100 clusters was used for local PCA. The data-set was centered and scaled according to the VAST scaling criterion. The reconstruction performed by PCA could clearly be improved by the local PCA formulation. The mean reconstruction errors were 7% for PCA and 3% for local PCA. The overall performance of the model for the reconstruction of the original data can be analyzed from Fig. 25. These figures report the parity plots for the reconstructed temperature and CO_2, CH_4, CO mass fraction fields. The dashed lines correspond to the 5% error. The dotted lines correspond to the 10% error. It is easily noticeable that the grey points, corresponding to the PCA reconstruction, are more scattered, while the black ones, corresponding to the local PCA reconstruction, lie within the reported 5% or 10% error lines more frequently (Fig. 25).

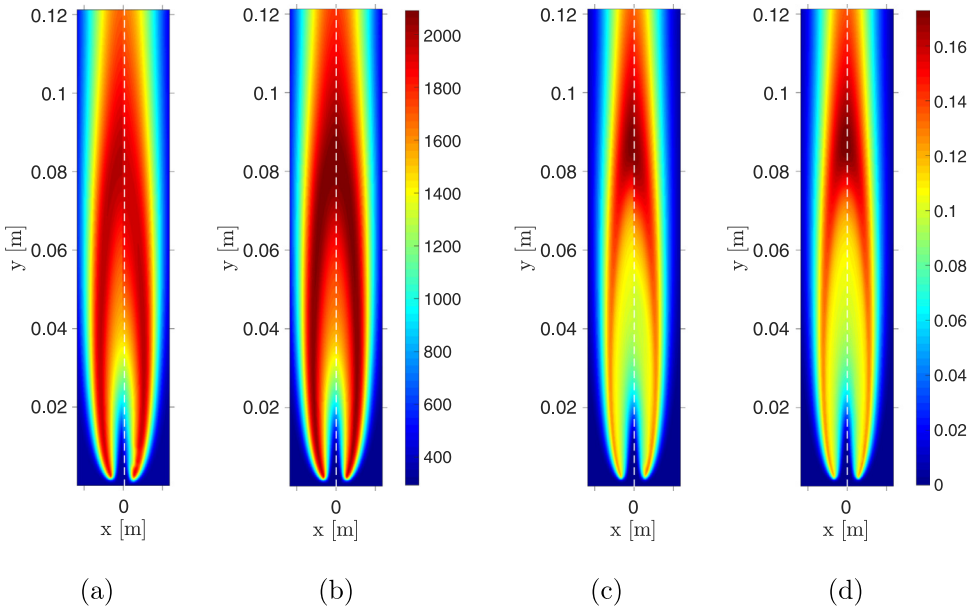

Fig. 24. Left halves: true data. Right halves: (a) reconstruction of the temperature field by PCA; (b) reconstruction of the temperature field by local PCA; (c) reconstruction of the CO_2 mass fraction field by PCA; (d) reconstruction of the CO_2 mass fraction field by local PCA. Inlet velocity: 89 cm/s; inlet CH_4 molar fraction: 85%. 2D flame. PCA: 10 PCs. Local PCA: 100 clusters.

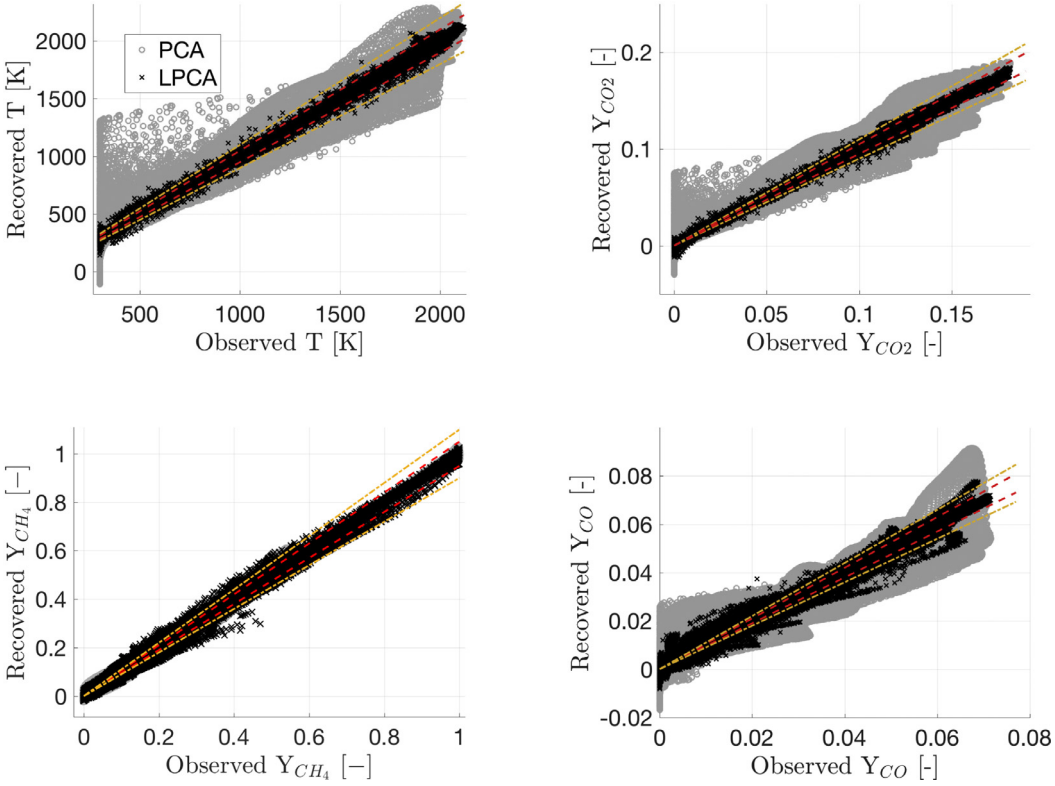

Fig. 25. Parity plot for the reconstruction of temperature and CO_2, CH_4, CO mass fractions (using 25 training samples): comparison between PCA and LPCA (5 PCs, 100 clusters). Dashed line: 5% error. Dotted line: 10% error. 2D flame.

Next, the predictive capabilities of a PCA+Kriging and local PCA+Kriging ROM for the 2D system were investigated. Fig. 26 reports the true temperature field for an inlet velocity of 64 cm/s and an inlet CH_4 molar fraction of 70%, as well as the predictions provided by PCA with 5 PCs, and local PCA with 100 clusters and 5 PCs, both in combination with the same Kriging model (quadratic trend function and exponential kernel). The mean error of the PCA+Kriging model for the temperature field was 8%, and the latter was halved using LPCA+Kriging. Similarly, Fig. 27 reports the CO and OH mass fractions fields for an inlet velocity of 64 cm/s and an inlet CH_4 molar fraction of 70%. The resuls qualitatively confirmed that the local PCA formulation could better reproduce the distribution (shape and extension) of the chemical species contours when compared to PCA. Parity plots for the prediction of temper-

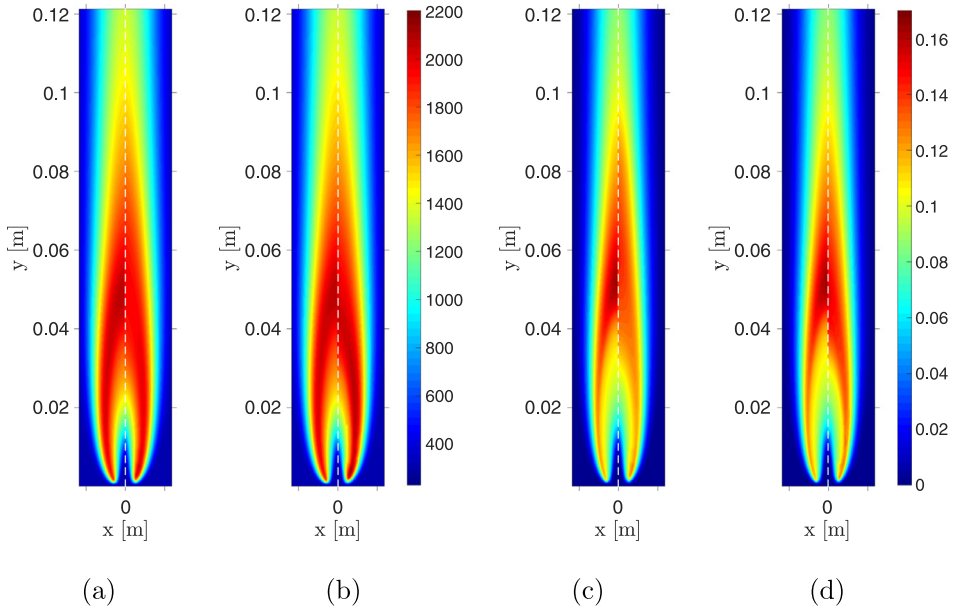

Fig. 26. Left halves: true data. Right halves: (a) prediction of the temperature field by PCA; (b) prediction of the temperature field by local PCA; (c) prediction of the CO_2 mass fraction field by PCA; (d) prediction of the CO_2 mass fraction field by local PCA. Inlet velocity: 64 cm/s; inlet CH_4 molar fraction: 70%. 2D flame. PCA: 5 PCs. Local PCA: 100 clusters. Kriging: quadratic trend function, exponential kernel.

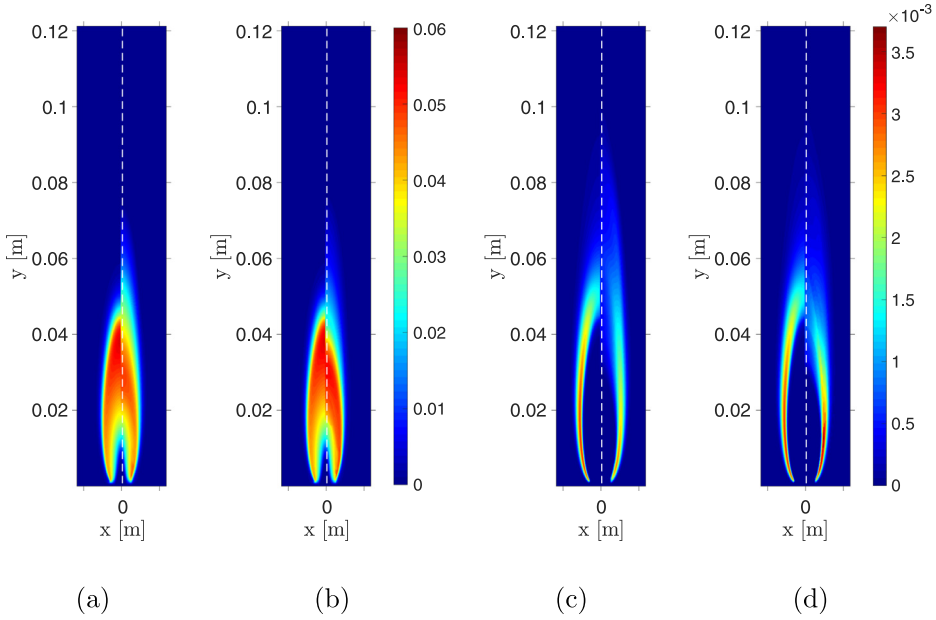

Fig. 27. Left halves: true data. Right halves: (a) prediction of the CO mass fraction field by PCA; (b) prediction of the CO mass fraction field by local PCA; (c) prediction of the OH mass fraction field by PCA; (d) prediction of the OH mass fraction field by local PCA. Inlet velocity: 64 cm/s; inlet CH_4 molar fraction: 70%. PCA: 5 PCs. Local PCA: 100 clusters. Kriging: quadratic trend function, exponential kernel. 2D flame.

ature and CO_2, CH_4, CO mass fractions are reported in Fig. 28. The 5% and 10% error lines are shown again in dashed red and yellow, respectively. The predictions of the local PCA+Kriging model, in black, lie within the 10% error lines more than the PCA+Kriging predictions, in gray, indicating the overall better performance of local PCA+Kriging.

4.5. 2D flame with two input parameters, using a smaller parameter region

The 2D data-set discussed above included observations for a high range of velocities. The parameter region to be explored included strong non-linearities which were difficult to model with *only* 30 training observations, thus limiting the predictive capabilities of the developed ROM. To confirm that, new observations were produced for a narrower range of velocities (24–55 m/s). A total of 23 observations were used for the training of a new ROM based on PCA and Kriging in this region. Predictions were validated for 4 combinations of input parameters: 45% CH_4- 35 m/s, 65% CH_4 - 30 m/s, 75% CH_4 - 40 m/s, 95% CH_4 - 40 m/s. Figs. 29 and 30 report the prediction of the temperature field and of the CO_2, CO, OH mass fraction fields by means of PCA+Kriging and local PCA+Kriging, for an inlet velocity of 40 cm/s and an inlet CH_4

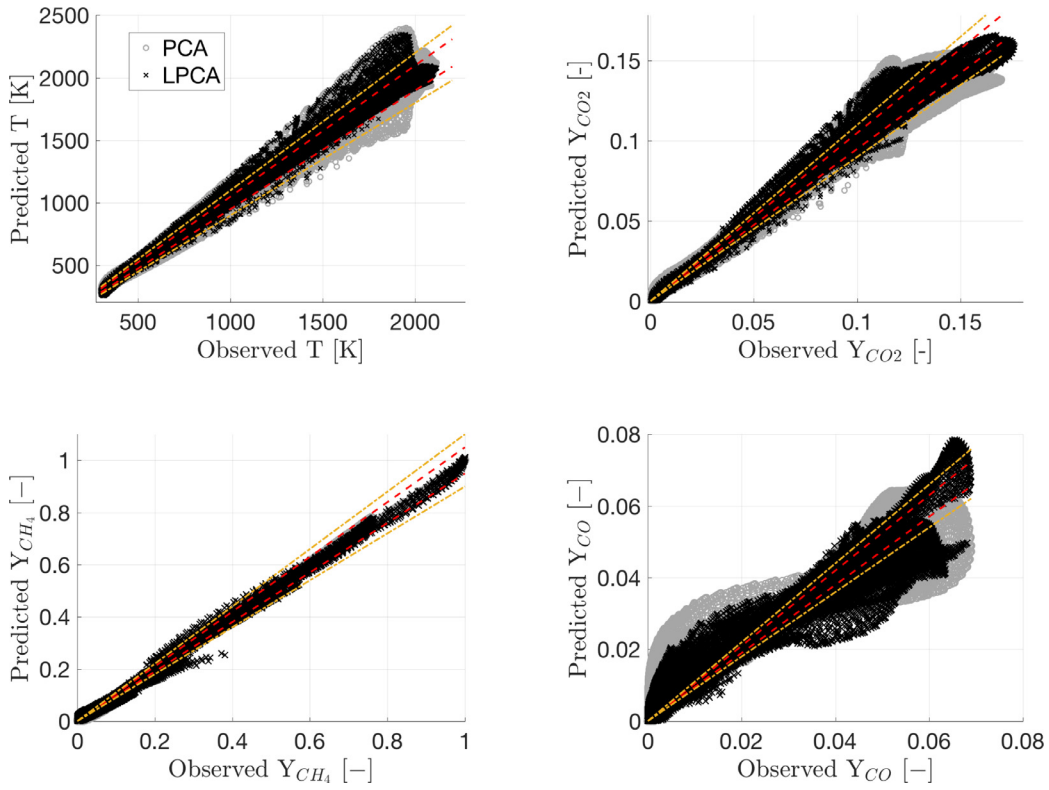

Fig. 28. Parity plot for the predictions of temperature and CO_2, CH_4, CO mass fractions (using 25 training samples): comparison between PCA and LPCA (5 PCs, 100 clusters). Kriging: quadratic trend function, Matern52 kernel. Dashed line: 5% error. Dotted line: 10% error. 2D flame.

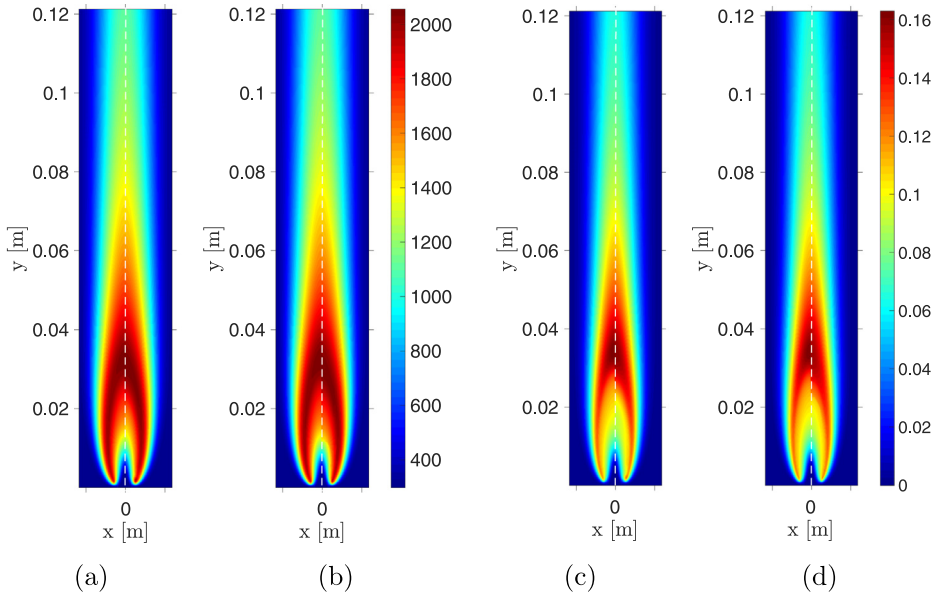

Fig. 29. Left halves: true data. Right halves: (a) prediction of the temperature field by PCA; (b) prediction of the temperature field by local PCA; (c) prediction of the CO_2 mass fraction field by PCA; (d) prediction of the CO_2 mass fraction field by local PCA. Inlet velocity: 40 cm/s; inlet CH_4 molar fraction: 75%. PCA: 5 PCs. Local PCA: 100 clusters. Kriging: quadratic trend function, exponential kernel. 2D flame with a smaller range of velocities.

molar fraction of 75%. PCA was performed with a number of 5 PCs, local PCA with a number of 100 clusters. Kriging was performed with a quadratic trend function and an exponential kernel. Fig. 31 reports the parity plots for the predicted values of temperature and CO_2, CH_4, CO mass fractions. These figures clearly indicate that better predictions were obtained in the region of low inlet velocities in comparison to Fig. 28, where the data spread was higher. In particular, it is observable that most predictions by the local PCA+Kriging model fell within the 5% error region (dashed red). The limitations of the predictive model observed in the previous case were due to the wide range of explored conditions: the inlet velocity was indeed allowed to change roughly by a factor of 4, which led to substantial modifications in the flame topology and structure. Limiting the ratio of velocity to a factor of 2 allowed to develop a ROM whose reliability in the range of conditions was significantly higher.

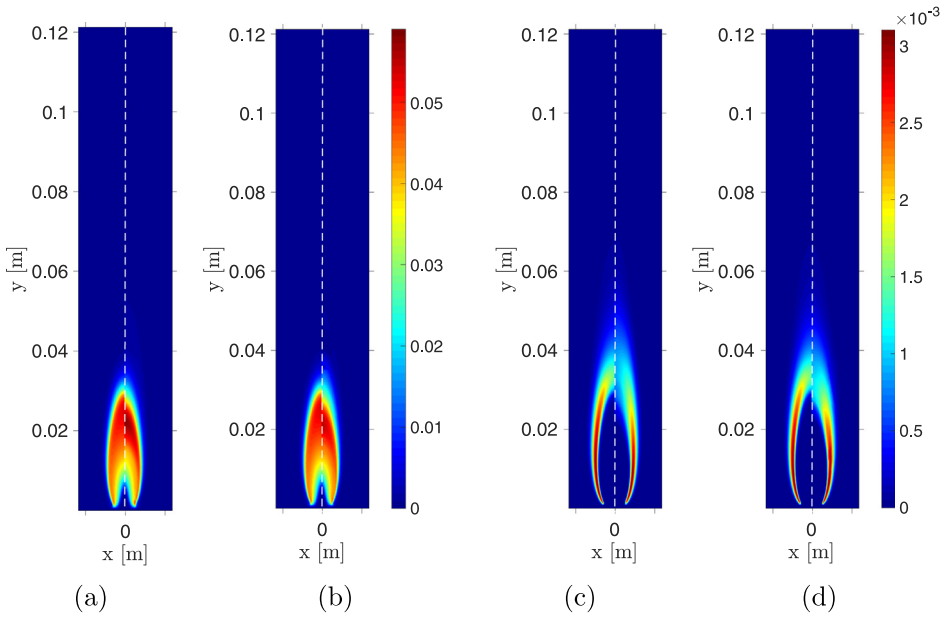

Fig. 30. Left halves: true data. Right halves: (a) prediction of the CO mass fraction field by PCA; (b) prediction of the CO mass fraction field by local PCA; (c) prediction of the OH mass fraction field by PCA; (d) prediction of the OH mass fraction field by local PCA. Inlet velocity: 40 cm/s; inlet CH_4 molar fraction: 75%. PCA: 5 PCs. Local PCA: 100 clusters. Kriging: quadratic trend function, exponential kernel. 2D flame with a smaller range of velocities.

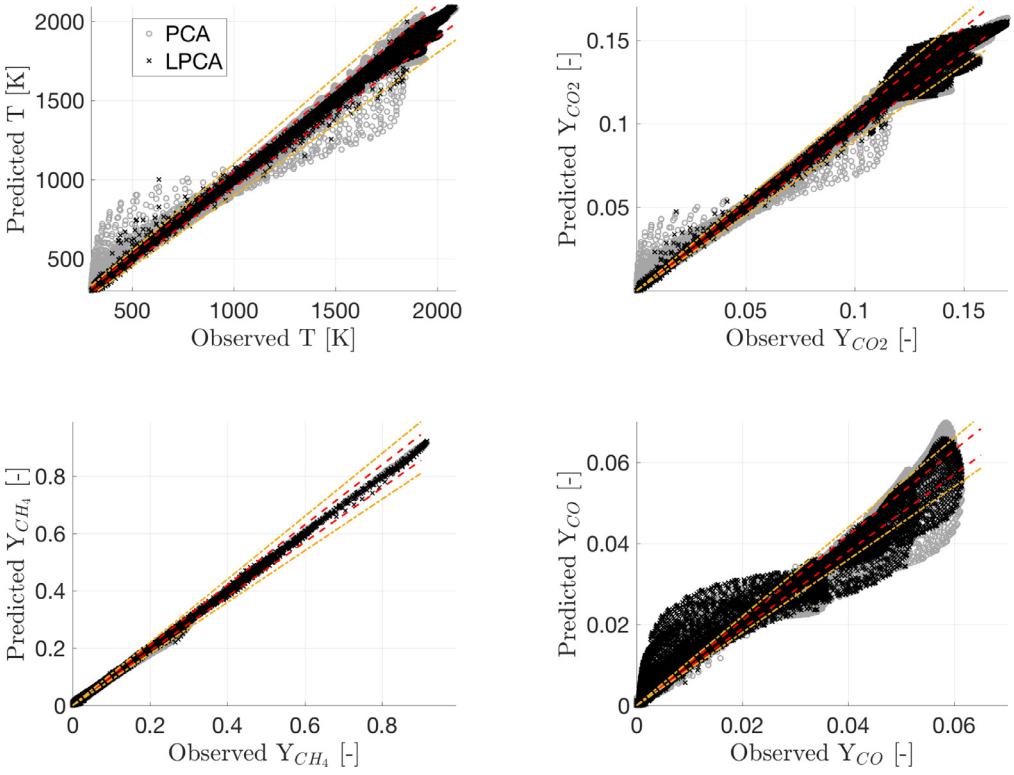

Fig. 31. Parity plot for the predictions of temperature and CO_2, CH_4, CO mass fractions: comparison between PCA and LPCA (5 PCs, 100 clusters). Kriging: quadratic trend function, Matern52 kernel. Dashed line: 5% error. Dotted line: 10% error. 2D flame with a smaller range of velocities.

Fig. 32a reports the cumulative data variance captured by the PCs. In the previous case, the number of training samples was 25 and 20 PCs were needed to capture over 99% of the data variance. In the present case, 15 PCs were sufficient to recover over 99% of data variance for 23 training observations, thus confirming that the parameter region was more efficiently sampled. Fig. 32b reports the global R^2 values for the prediction of the validation data as the number of samples used for the training of a local PCA+Kriging model increased. As expected, these values increased to unity as more training samples were used.

4.6. Performance evaluation

The computational cost of the two-dimensional CFD with two input parameters, described in Section 4.4, was over 30 CPU-hours per simulation. By developing a ROM, as shown throughout this

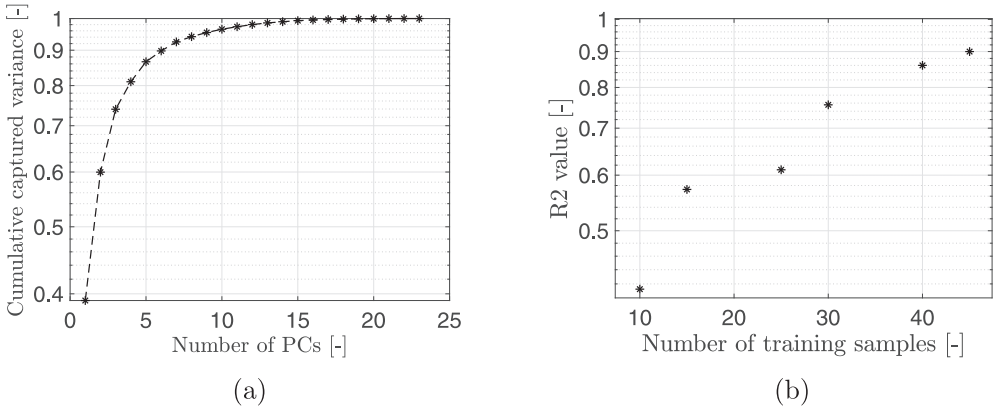

Fig. 32. (a) The cumulative original data variance that is captured when adding more PCs provides a criterion for the selection of the number of PCs when using global PCA. VAST-scaling. 2D flame with a smaller range of velocities. (b) Global R^2 values for the validation data as the number of training observations increases. Local PCA+Kriging with quadratic trend function and Matern52 kernel.

Table 1
Comparison between the computational performances of different Kriging models: a direct Kriging model, a PCA+Kriging model with 10 PCs, and a local PCA+Kriging model with 10 PCs and 100 clusters.

	Kriging	10 PCs	100 clusters
Training time	9.7 hrs	4 s	9 s
Clustering time	-s	-s	53 min
Speed-up	1	8,712	11
#Hyper-parameters	945,360	30	3,000
#Coefficients	0	5,672,160	5,672,160

paper, the outcome of such simulations could be predicted instantaneously. Besides, the costs associated to the training process of a ROM based on PCA and Kriging were also smaller when compared to the computational burden needed for direct Kriging (no PCA compression is performed and one Kriging model is trained per original number of variables).

Table 1 summarizes the training costs associated to different models, namely: a direct Kriging model, a PCA+Kriging model with 10 PCs, and a local PCA+Kriging model with 10 PCs and 100 clusters. As clearly shown in Table 1, the training costs and the number of hyper-parameters associated to a Direct Kriging model were very high when compared to the two ROMs based on PCA or local PCA. On the other hand, after training a (local) PCA+Kriging model, there was the need for additional coefficients to be stored in memory, due to the centering and scaling procedure, and the PCA compression. For a PCA+Kriging model with 10 PCs, 2 vectors of mean values and scaling factors had to be stored as well as 10 PCA modes, for a total of 12 vectors of size 472,680, as the data-set was composed of 36 physical variables evaluated at 13,130 spatial locations (half of the original mesh was analyzed exploiting the problem's symmetry). The same number of coefficients needed to be stored for the local PCA+Kriging model with 10 PCs and 100 clusters.

5. Conclusions

In the present work, a methodology for the development of accurate and robust ROM generation using a combination of PCA and Kriging was presented. The methodology was demonstrated for a 1D laminar flame with an increasing number of input parameters (equivalence ratio, composition of the fuel, inlet temperature), and for a 2D flame with two input parameters (inlet velocity and inlet fuel composition). In all three cases and for the 2D flame as well, both training and test data were available. The training data was employed to generate the ROM. The test data was used to assess the ROM's predictive capabilities.

The results showed that the combination of PCA with Kriging is a valid solution for the development of physics-based SMs. Mainly, we want to stress the following points:

- The developed surrogate models can be used for parameter exploration with low prediction errors: $< 10\%$.
- The local PCA formulation provides an improvement over PCA, as it better deals with the non-linearities and the high dimensionality of the original system.
- CPCA guarantees that the imposed physical constraints are not violated when the data are reconstructed. In this work, the imposed constraint is the positivity of the chemcial specie mass fractions. This is not guaranteed by PCA, while local PCA alleviates this issue by simply improving the accuracy of the data reconstruction.
- Once a ROM is correctly trained, predictions are possible at a reduced computational cost, making parameter exploration much easier, even for very CPU-intensive systems.

The present work represents the first application of the PCA+Kriging methodology to combustion problems. As such, it is intended to be a proof of concept that will pave the way for the application of this methodology to more complex systems. Indeed, as 3D simulations of practical combustion systems usually require significant computational resources, having a low-order model that can reliably and instantaneously predict the outcome of these simulations is precious. Moreover, the promptness of the ROM's predictions is paramount for the development of digital twins for real systems, which can be employed for system control and visualization. A correctly trained ROM also grants the possibility of performing sensitivity analysis of the investigated system w.r.t. its input parameters, and can be employed to solve optimization problems in the context of system design, where the evaluation of the objective function is the computational burden. The training cost of PCA+Kriging ROMs is also lower compared to SMs with no compression. This is very useful when updating a developed ROM in the event of new available data. Despite the relative simplicity of the selected test cases, the present work allowed to investigate the advantages and limitations of the method, as well as its potential for applications to more complex combustion systems. A substantial reduction in the system dimensionality was accomplished via PCA (e.g. from 10,780 to 5 scalars in the one-parameter case), while the use of Kriging allowed to capture the non-linear relation between the reduced set of coefficients and the input parameters, enabling the prediction of non-observed system states.

Future work will extend the proposed methodology to more complex cases, including three-dimensional turbulent reacting flows.

Acknowledgements

This project has received funding from the European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 643134 and from the European Research Council, Starting Grant no 714605.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.compchemeng.2018.09.022.

References

Bellemans, A., Aversano, G., Coussement, A., Parente, A., 2018. Feature extraction from reduced order models based on principal component analysis. Comput. Chem. Eng.. March

Beykal, B., Boukouvala, F., Floudas, C.A., Pistikopoulos, E.N., 2018a. Optimal design of energy systems using constrained grey-box multi-objective optimization. Comput. Chem. Eng. 0, 1–15. doi:10.1016/j.compchemeng.2018.02.017.

Beykal, B., Boukouvala, F., Floudas, C.A., Sorek, N., Zalavadia, H., Gildin, E., 2018b. Global optimization of grey-box computational systems using surrogate functions and application to highly constrained oil-field operations. Comput. Chem. Eng. 114, 99–110. doi:10.1016/j.compchemeng.2018.01.005.

Bishop, C.M., 2013. Pattern Recognition and Machine Learning, 53 doi:10.1117/1.2819119.

Bizon, K., Continillo, G., 2012. Reduced order modelling of chemical reactors with recycle by means of POD-penalty method. Comput. Chem. Eng. 39, 22–32. doi:10.1016/j.compchemeng.2011.10.001.

Bizon, K., Continillo, G., Berezowski, M., Smua-Ostaszewska, J., 2012. Optimal model reduction by empirical spectral methods via sampling of chaotic orbits. Physica D 241 (17), 1441–1449. doi:10.1016/j.physd.2012.05.004.

Bizon, K., Continillo, G., Mancaruso, E., Merola, S.S., Vaglieco, B.M., 2010. POD-based analysis of combustion images in optically accessible engines. Combust. Flame 157 (4), 632–640. doi:10.1016/j.combustflame.2009.12.013.

Cao, S., Bennett, B., Ma, B., Giassi, D., 2013. Effects of fuel dilution and gravity on laminar coflow methane-air diffusion flames: acomputational and experimental investigation, 1–9.

Cao, S., Ma, B., Bennett, B.A., Giassi, D., Stocker, D.P., Takahashi, F., Long, M.B., Smooke, M.D., 2015. 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), 897–903. doi:10.1016/j.proci.2014.05.138.

Constantine, P.G., Dow, E., Wang, Q., 2014. Active subspace methods in theory and practice. SIAM J. Sci. Comput. 36 (4), 1500–1524.

Coussement, A., Isaac, B.J., Gicquel, O., Parente, A., 2016. Assessment of different chemistry reduction methods based on principal component analysis: comparison of the MG-PCA and score-PCA approaches. Combust. Flame 168, 83–97. doi:10.1016/j.combustflame.2016.03.021.

Crestaux, T., Le Maître, O., Martinez, J.M., 2009. Polynomial chaos expansion for sensitivity analysis. Reliab. Eng. Syst. Saf. 94 (7), 1161–1172. doi:10.1016/j.ress.2008.10.008.

Cuoci, A., Frassoldati, A., Faravelli, T., Ranzi, E., 2013a. A computational tool for the detailed kinetic modeling of laminar flames: application to C2H4/CH4 coflow flames. Combust. Flame 160 (5), 870–886. doi:10.1016/j.combustflame.2013.01.011.

Cuoci, A., Frassoldati, A., Faravelli, T., Ranzi, E., 2013b. Numerical modeling of laminar flames with detailed kinetics based on the operator-splitting method. Energy Fuels 27 (12), 7730–7753. doi:10.1021/ef4016334.

Echekki, T., Mirgolbabaei, H., 2015. Principal component transport in turbulent combustion: a posteriori analysis. Combust. Flame 162 (5), 1919–1933. doi:10.1016/j.combustflame.2014.12.011.

Fürst, M., Sabia, P., Lubrano Lavadera, M., Aversano, G., de Joannon, M., Frassoldati, A., Parente, A., 2018. Optimization of chemical kinetics for methane and biomass pyrolysis products in moderate or intense low-Oxygen dilution combustion. Energy Fuels doi:10.1021/acs.energyfuels.8b01022.

Guenot, M., Lepot, I., Sainvitu, C., Goblet, J., Coelho, R.F., 2013. Adaptive sampling strategies for non-intrusive POD-based surrogates. Eng. Comput. 30 (4), 521–547. doi:10.1108/02644401311329352.

Haag, S., Anderl, R., 2018. Digital twin Proof of concept. Manuf. Lett. 10–12. doi:10.1016/j.mfglet.2018.02.006.

Isaac, B.J., Coussement, A., Gicquel, O., Smith, P.J., Parente, A., 2014. Reduced-order PCA models for chemical reacting flows. Combust. Flame 161 (11), 2785–2800. doi:10.1016/j.combustflame.2014.05.011.

Isaac, B.J., Thornock, J.N., Sutherland, J., Smith, P.J., Parente, A., 2015. Advanced regression methods for combustion modelling using principal components. Combust. Flame 162 (6), 2592–2601. doi:10.1016/j.combustflame.2015.03.008.

Jolliffe, I.T., 2002. Principal Component Analysis, Second Edition.

Kambhatla, N., Leen, T., 1997. Dimension reduction by local principal component analysis. Neural Comput. 9 (7), 1493–1516. doi:10.1162/neco.1997.9.7.1493.

Khuwaileh, B.A., Turinsky, P.J., 2017. Surrogate based model calibration for pressurized water reactor physics calculations. Nuclear Eng. Technol. 49 (6), 1219–1225. doi:10.1016/j.net.2017.08.007.

Lancien, T., Dumont, N., Prieur, K., Durox, D., Candel, S., Gicquel, O., Vicquelin, R., 2016. Uncertainty quantification of injected droplet size in mono-dispersed Eulerian simulations.

Lin, G., 2017. On the Bayesian calibration of expensive computer models with input dependent parameters. Spat. Stat. doi:10.1016/j.spasta.2017.08.002.

Lophaven, S. N., Søndergaard, J., Nielsen, H. B., 2002. Kriging toolbox, 1–28.

Mirgolbabaei, H., Echekki, T., Smaoui, N., 2014. A nonlinear principal component analysis approach for turbulent combustion composition space. Int. J. Hydrogen Energy 39 (9), 4622–4633. doi:10.1016/j.ijhydene.2013.12.195.

Müller, J., Shoemaker, C.A., Piché, R., 2013. SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. Comput. Oper. Res. 40 (5), 1383–1400. doi:10.1016/j.cor.2012.08.022.

Parente, A., Sutherland, J.C., 2013. Principal component analysis of turbulent combustion data: data pre-processing and manifold sensitivity. Combust. Flame 160 (2), 340–350. doi:10.1016/j.combustflame.2012.09.016.

Parente, A., Sutherland, J.C., Dally, B.B., Tognotti, L., Smith, P.J., 2011. Investigation of the MILD combustion regime via Principal Component Analysis. Proc. Combust. Inst. 33 (2), 3333–3341. doi:10.1016/j.proci.2010.05.108.

Parente, A., Sutherland, J.C., Tognotti, L., Smith, P.J., 2009. Identification of low-dimensional manifolds in turbulent flames. Proc. Combust. Inst. 32 I (1), 1579–1586. doi:10.1016/j.proci.2008.06.177.

Regis, R.G., Shoemaker, C.A., 2005. Constrained global optimization of expensive black box functions using radial basis functions. J. Global Optim. 31 (1), 153–171. doi:10.1007/s10898-004-0570-0.

Sahyoun, S., Djouadi, S., 2013. Local proper orthogonal decomposition based on space vectors clustering. In: Systems and Control (ICSC), 2013 3rd International Conference on, pp. 665–670. doi:10.1109/ICoSC.2013.6750930.

Schleich, B., Anwer, N., Mathieu, L., Wartzack, S., 2017. Shaping the digital twin for design and production engineering. CIRP Ann. Manuf. Technol. 66 (1), 141–144. doi:10.1016/j.cirp.2017.04.040.

Seeger, M., 2004. Gaussian Processes for Machine Learning., 14 doi:10.1142/S0129065704001899.

Sutherland, J.C., Parente, A., 2009. Combustion modeling using principal component analysis. Proc. Combust. Inst. 32 I (1), 1563–1570. doi:10.1016/j.proci.2008.06.147.

Uhlemann, T.H., Schock, C., Lehmann, C., Freiberger, S., Steinhilper, R., 2017. The digital twin: demonstrating the potential of real time data acquisition in production systems. Procedia Manuf. 9, 113–120. doi:10.1016/j.promfg.2017.04.043.

Williams, L. J., 2010. Principal component analysis 2 (August). 10.1002/wics.101

Xiao, M., Breitkopf, P., Coelho, R.F., Villon, P., Zhang, W., 2014. Proper orthogonal decomposition with high number of linear constraints for aerodynamical shape optimization. Appl. Math. Comput. 247, 1096–1112. doi:10.1016/j.amc.2014.09.068.

Xiao, M., Breitkopf, P., Filomeno Coelho, R., Knopf-Lenoir, C., Sidorkiewicz, M., Villon, P., 2010. Model reduction by CPOD and Kriging. Struct. Multidiscip. Optim. 41 (4), 555–574. doi:10.1007/s00158-009-0434-9.

Xiao, M., Breitkopf, P., Filomeno Coelho, R., Knopf-Lenoir, C., Villon, P., 2012. Enhanced POD projection basis with application to shape optimization of car engine intake port. Struct. Multidiscip. Optim. 46 (1), 129–136. doi:10.1007/s00158-011-0757-1.

Xiao, M., Breitkopf, P., Filomeno Coelho, R., Knopf-Lenoir, C., Villon, P., Zhang, W., 2013. Constrained proper orthogonal decomposition based on QR-factorization for aerodynamical shape optimization. Appl. Math. Comput. 223, 254–263. doi:10.1016/j.amc.2013.07.086.

Yu, J., 2012. Local and global principal component analysis for process monitoring. J. Process Control 22 (7), 1358–1373. doi:10.1016/j.jprocont.2012.06.008.