Reduced order modeling of self-igniting reaction-diffusion system based on POD technique and k-means clustering

Katarzyna Bizon[1], Simone Lombardi[1], Gaetano Continillo[1,2]
[1]Department of Engineering, Università del Sannio Benevento, Italy
[2]Istituto Motori CNR, Naples, Italy

1 Introduction

Although advances in computing power allow nowadays for detailed numerical simulation of a wide variety of systems, there is continuing demand for the development of model order reduction techniques. This is due to the fact that fast but accurate models are necessary in real-time control applications or in the detailed study of the systems characterized by a large number of parameters. Perhaps the most widespread and powerful technique, especially when it comes to the reduction of nonlinear infinite dimensional systems described by evolutionary partial differential equations (PDEs), is proper orthogonal decomposition (POD) [1]. Based on experimental or numerically-derived observations, it provides a set of optimal, in the sense of L^2 norm, empirical basis functions, which can be easily incorporated in a classical subspace projection techniques for model order reduction such as the Galerkin method.

POD has been widely employed for model order reduction in many fields, however there are still a number of issues that need further investigation. Among the others, the policy of the collection of the representative set of experimental or simulation data is considered to be crucial for generating a global basis suitable for the determination of accurate reduced order model (ROM). At present, there exists no unequivocal procedure for optimal snapshot selection. It is well known that a potentially representative ensemble of data can be obtained by combining data from different simulations, conducted for different values of key parameters [2] or characterized by high spatiotemporal complexity [3,4]. However, while the exploration of the parameter space is a well established policy, there is no clear indication in the literature about the influence of the total number of snapshots and their temporal distribution on the ROM performance. Recently, the influence of the number of snapshots both for constant and variable time step sampling was examined in [5]. An alternative approach, based on the mutual information, for the selection of the most uncorrelated snapshots was proposed in [6].

In this work, the method of POD is applied to a problem of a self-ignition of a coal stockpile, described by one-dimensional reaction-diffusion equations [7-9]. Aiming at the improvement of the ROM performance, different sampling strategies for the collection of the observation are studied. Particularly, the performance of a POD model determined from a uniformly time sampled set of representative solutions is compared with an innovative approach based on k-means clustering of solution profiles [10].

Correspondence to: katarzyna.bizon@unisannio.it

2 Mathematical model

The governing equations are those describing a classic reaction-diffusion problem. In case of a heterogeneous reaction in a one-dimensional layer, the model equations, in dimensionless form, are [8]:

$$\frac{\partial Y}{\partial \tau} = Le\frac{\partial^2 Y}{\partial x^2} - \varphi^2 Y \exp\left(\frac{-\gamma}{T}\right)$$

$$\frac{\partial T}{\partial \tau} = \frac{\partial^2 T}{\partial x^2} + \beta\varphi^2 Y \exp\left(\frac{-\gamma}{T}\right)$$

where Y is the concentration of the gas reactant and T is the temperature, assumed to be equal for gas and solid phase. The associated boundary and initial conditions are:

$$T(0,\tau) = T(1,\tau) = 1, \quad Y(0,\tau) = 1, \quad \left.\frac{\partial Y}{\partial x}\right|_{x=1} = 0 \quad \text{for} \quad \tau > 0$$

$$T(x,0) = 1, \quad Y(x,0) = 0 \quad \text{at} \quad \tau > 0, \; x \in [0,1]$$

with x=0 and x=1 corresponding to the top (ambient temperature and concentration at the left boundary) and to the bottom of the pile, respectively. As reported elsewhere [7-9] the system exhibits interesting dynamical behavior. In the present study, the values of the Lewis number, *Le*, the dimensionless heat of reaction, β, and the Thiele modulus, φ, were set, respectively, to the following values [7]:

$$Le = 0.233, \quad \varphi^2 = 70000, \quad \beta = 4.287$$

whereas the dimensionless activation energy, $\gamma = E/(RT_0)$, containing the reference (ambient) temperature, was initially set to γ=12.4. For this set of parameter values, the transient system, after ignition, converges to a period 1 limit cycle (Figure 1).

3 Proper orthogonal decomposition and k-means clustering

For a given set of snapshots $u(\mathbf{x},t_i)$, $i=1,...,M$, the POD basis $\Psi = \{\psi_1,\psi_2,\ldots,\psi_N\}$ can be obtained by solving the eigenvalue problem given by [1]:

$$C\Psi = \lambda C$$

where C is the autocorrelation matrix defined as:

$$C(\mathbf{x},\mathbf{y}) = \left\langle u(\mathbf{x},t_i)u^T(\mathbf{y},t_i)\right\rangle$$

with $\langle\cdot\rangle$ denoting ensemble average over the number of snapshots *M*.

Then, using the POD modes, the solution $u(\mathbf{x},t_i)$ can be expressed, in a truncated form, as:

$$u(\mathbf{x},t_i) = \sum_{k=1}^{K} a_k(t_i)\psi_k(\mathbf{x})$$

where $K \ll N$ is the truncation order, whereas $a_k(t_i), k=1,\ldots,K$ are modal coefficients to be determined solving a system obtained by performing the Galerkin projection of the original system onto POD modes [2-4]. Usually, the snapshots employed for the determination of the POD basis are collected at uniform time-sampling rate. In order to increase the amount of information contained in a given number of POD modes, this classical approach is compared here with a procedure based on the so-called *k-means clustering* [10]. Namely, the solution profiles are first grouped into *M* clusters and then the POD modes are determined from the centroids of the clusters. More precisely, the k-means algorithm groups together subsets of 'nearby' solutions, on the basis of a mutual distance, in this case defined as a squared Euclidean distance.

4 Results and discussion

To test the influence of the number of snapshots and their distribution on the performance of the POD-based ROM, the infinite dimensional model was first reduced to the system of ordinary differential equations employing finite difference approximation of the spatial derivatives with N=201 nodes, constructing what we here call the full order model (FOM). The FOM was then integrated in time for $\tau_f \in [0,15]$, and 10000 snapshots were collected. In the first step, the POD method was applied to a subset made of M=200 uniformly time-sampled solutions. The POD modes computed were then used to determine ROMs of different truncation order. Figure 1 shows a comparison of temperature and concentration evolution at x=0.2 obtained using the FOM and two ROMs constructed employing K=3 and K=10 modes (for each state variable) respectively. As expected, due to the complexity of the solution consisting of a fast transient converging to period-1 stable oscillations, it can be observed that, the performance of the ROM is not satisfactory if too little modes (here K=3, green dots) are employed in the reduction phase. For K=10, the ROM solution (red dots) is so accurate that it is superimposed to the FOM (blue line) solution.

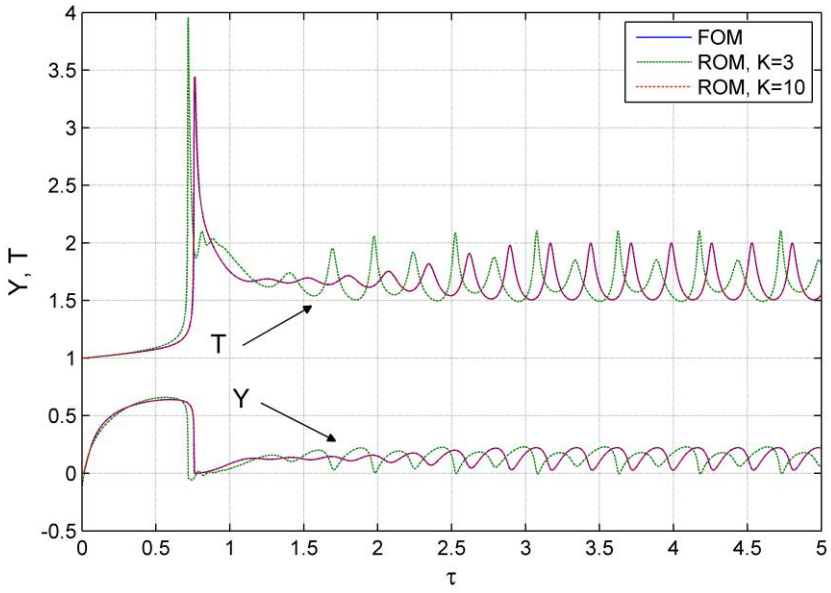

Figure 1. Comparison of temperature and concentration at x=0.2 obtained using FOM and ROM with K=3 and K=10; both ROMs determined from M=200 uniformly distributed in time snapshots.

Successively, in order to investigate the influence of the sampling onto the ROM performance two sampling strategies were utilized and compared. Namely, the POD was performed on:
- snapshots collected at uniform frequency both during transient and steady oscillations; four sets of the total number of snapshots equal to M=25, 50, 100 and 200 were considered for this case;
- centroids of the snapshots grouped into M=200, 100, 50, and 25 clusters.

It is quite evident that, when the dataset is constructed based on the averaging of the clustered solutions, profiles contain more information from the transient, during which solution profiles typically are more distant from each other. Figure 2 shows centroids of the k-means clustered solutions used as the input to the POD procedure for M=25.

Eight data subsets were extracted from the FOM solution data, and POD was performed on each subset. Successively, the ROMs were determined and simulated, initially at reference conditions, i.e. for γ=12.4.

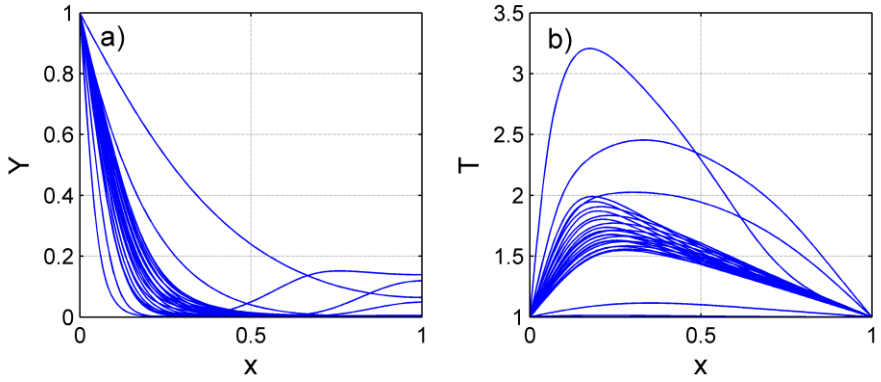

Figure 2. Centroids of the clustered concentration (a) and temperature (b) profiles for M=25.

To evaluate the effect of the two sampling strategies quantitatively, an average least-square truncation error is defined as:

$$\varepsilon_K = \left\langle \frac{1}{N} \left\| u(\mathbf{x},t_i) - \sum_{k=1}^{K} a_k(t_i)\psi_k(\mathbf{x}) \right\|^2 \right\rangle$$

where $\|\cdot\|$ is the L^2 norm while $\langle\cdot\rangle$ denotes an average over M. Figure 3 reports the average least-squares truncation error of ROMs obtained using POD bases determined from subsets containing M=25 and M=200 samples. It can be observed that, for low truncation orders, i.e. $K<8$, the sampling strategy does not influence the accuracy of the ROM solution, whereas for larger truncation orders ($K>8$) the desired accuracy of the ROM solution, i.e. $\varepsilon_K < 10^{-4}$, is achieved. We hence focus on the latter cases for the comparison. With standard POD, as the number of samples M increases, the amount of information included in the POD basis, to some extent, increases and so does the performance of the ROM. On the other hand, with k-means clustering, the best performers are the ROMs constructed by projection of the FOM onto the basis determined with a low number of clusters. This is explained by the fact that clusters, being in low number, are more "distant" from each other and therefore carry the most of information on the system dynamics. Increasing the number of clusters does not increase the amount of information introduced into the data subset, as the clusters simply tend to overlap.

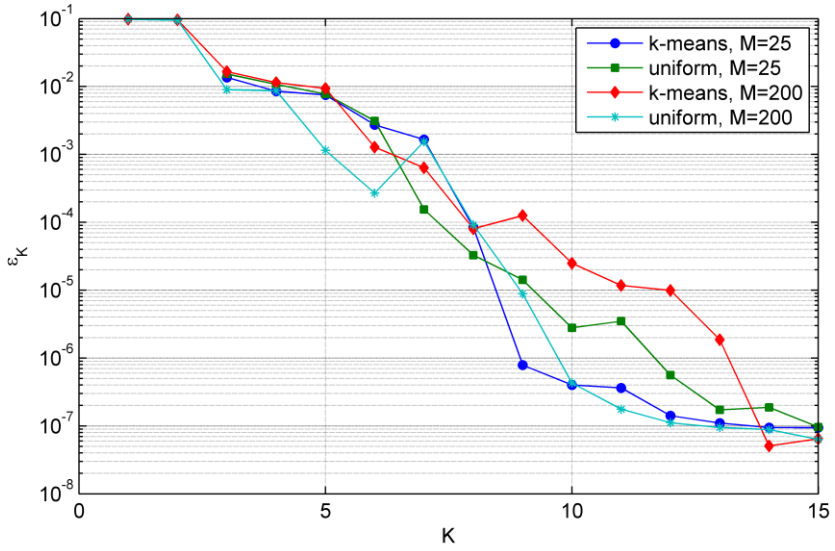

Figure 3. Average least-square truncation error at reference conditions (γ=12.4).

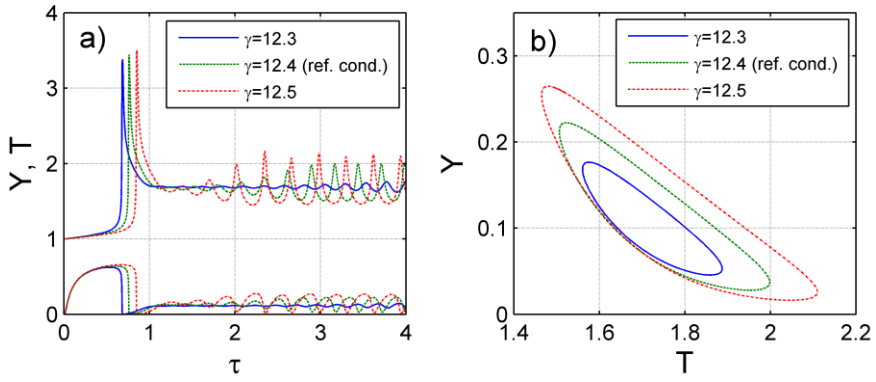

Figure 4. Transient (a) and limit cycles for γ=12.3, γ=12.4 (reference condition) and γ=12.5.

These trends preserve when testing ROMs at off-reference conditions, i.e. at different values of the activation energy γ. Figure 4 shows time series (a) and limit cycles (b) obtained by means of the FOM. For k-means clustering at K=10 or larger, the plots are superimposable to FOM results. To quantify the reconstruction accuracy for various values of K and for different ROM strategies, the average least-square truncation error was evaluated and then plotted in Figure 5. It is interesting to note that the performance of the ROM based on the uniform sampling improves for γ=12.3, different from the reference case, from which the basis was built. This can be explained with the fact that, for γ=12.3, the spatiotemporal complexity of the solution decreases, as it can be seen in Fig. 4a. The opposite is observed for γ=12.5, value that yields a more complex spatiotemporal pattern (the amplitude of the oscillations increases).

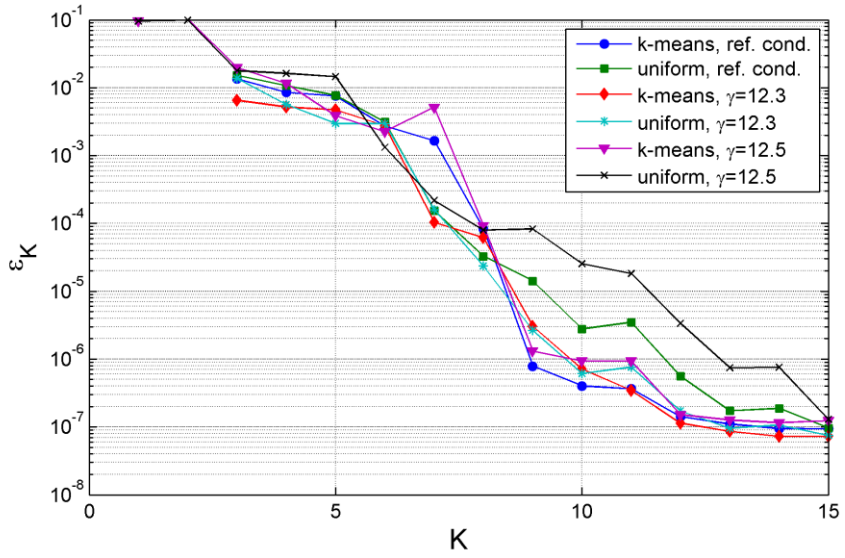

Figure 5. Comparison of average least-square truncation error at reference conditions, γ=12.4, and off-reference conditions, γ=12.3 and γ=12.5.

Another interesting finding is the computational time of both ROMs, normalized here with respect to the CPU time required to solve the FOM. Same number of ODEs, same integration time interval, lower computational effort: the use of adaptive ODE solver demonstrates that the equation system has smoother features. The ROMs determined using k-means clustering tend to be more stable and hence faster then those constructed using uniformly distributed numerical sampling.

Table 1: Computational time normalized with respect to the time required to solve the FOM.

	ref. cond.: γ=12.4		γ=12.3		γ=12.5	
K	k-means	uniform	k-means	uniform	k-means	uniform
3	0.1389	0.2003	0.1704	0.2911	0.1241	0.1897
6	0.2022	0.2329	0.2145	0.2449	0.1974	0.2041
10	**0.3099**	**0.3099**	**0.3279**	**0.3159**	**0.2628**	**0.2852**
15	0.3617	0.3663	0.3590	0.3417	0.3345	0.3334

5 Conclusions

In an attempt of improving the performance of POD-based reduced order models (ROMs) of a reaction-diffusion system, the influence of two different sampling strategies on the solution was investigated. First, POD modes were determined from databases that consisted of various number of samples collected at constant frequency. Then, an innovative approach based on the clustering was introduced and evaluated. It appears that, for low values of the total number of observations, the new approach ensures more stable and more accurate ROMs, with one third of the CPU time with respect to the full order model. The results obtained confirm that a proper choice of the sampling strategy can not only improve the performance of the ROM, but can also reduce significantly the size of the eigenvalue problem to be solved when building the POD modes, and, by designing an appropriate clustering procedure, reduce the amount of data to be computed and stored.

References

[1] Holmes P, Lumley JL, Berkooz G. (1996). Turbulence, coherent structures, dynamical systems and symmetry, Cambridge University Press.

[2] Bizon K, Continillo G, Russo L, Smuła J. (2008). On POD reduced models of tubular reactor with periodic regimes. Comput. Chem. Eng. 32: 105.

[3] Bizon K, Continillo G. (2009). POD/Galerkin reduced order model of tubular reactor with heat recycle by sampling of chaotic orbits. 22nd ICDERS.

[4] Bizon K, Continillo G, Berezowski M, Smuła-Ostaszewska J. (2012). Optimal model reduction by empirical spectral methods via sampling of chaotic orbits. Physica D 241: 1441.

[5] Brenner TA, Fontenot RL, Cizmas PGA, O'Brien TJ, Breault RW. (2012). A reduced-order model for heat transfer in multiphase flow and practical aspects of the proper orthogonal decomposition. Comput. Chem. Eng. 43: 68.

[6] Kostka S, Lynch AC, Huelskamp BC, Kiel BV, Gord JR, Roy S. (2012). Characterization on flame-sheeding behavior behind a bluff-body using proper orthogonal decomposition. Combust. Flame 159: 2872.

[7] Brooks K, Balakotaiah V, Luss D. (1988). Effect of natural convection on spontaneous combustion of coal stockpiles. AIChE Journal 34: 353.

[8] Continillo G, Maffettone PL, Crescitelli S. (1995). On the numerical simulation of the chaotic behaviour of some distributed-parameter systems. In: Biardi G. (Ed.), Chaos and Fractals in Chemical Engineering, World Scientific Publishing Company.

[9] Grepl MA. (2012). Model order reduction of parametrized nonlinear reaction-diffusion systems. Comput. Chem. Eng. 4: 33.

[10] MacQueen J. (1967). Some methods for classification and analysis of multivariate observation. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability 1: 281.