Computers & Fluids 40 (2011) 195–209

Contents lists available at ScienceDirect

Computers & Fluids

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

Towards an adaptive POD/SVD surrogate model for aeronautic design

T. Braconnier [a], M. Ferrier [a], J.-C. Jouhaud [a,*], M. Montagnac [a], P. Sagaut [b]

[a] Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, 42, avenue Gaspard Coriolis, 31057 Toulouse Cedex, France
[b] Institut Jean Le Rond d'Alembert, Université Pierre et Marie Curie, UPMC, UMR CNRS 7190, 4, place Jussieu, 75252 Paris Cedex 5, France

ARTICLE INFO

Article history:
Received 18 January 2010
Received in revised form 1 September 2010
Accepted 2 September 2010
Available online 15 September 2010

Keywords:
Singular Value Decomposition
Proper Orthogonal Decomposition
Aerodynamic design
Error estimation
Sampling
Model reduction

ABSTRACT

A computational methodology is presented to obtain a model reduction of the steady compressible Reynolds-Averaged Navier–Stokes equations with a high-dimensional parameter space. It combines both a reduced-basis method and an adaptive sequential sampling technique. The reduced basis is made of the leading eigenvectors computed by a Singular Value Decomposition of the snapshot span basis. The sampling method uses an a posteriori error estimator combined with a leave-one-out algorithm. This methodology allows a dynamic building of surrogate models for aerodynamic flight data generation and for multidisciplinary design optimization with a small number of full numerical flow simulations. The efficiency of the method is assessed on an analytic test case and a classical aerodynamic transonic flow around a 2D profile.

1. Introduction

The discretization of mathematical models, especially nonlinear partial differential equations (PDE), often leads to high-dimensional algebraic systems in engineering applications. These, such as multidisciplinary optimization [13,14], control design, inverse problem or sensitivity analysis, require a huge number of computations with the full model that gives the true numerical values of interest functions. Even if computers can process quickly large amounts of data, calling the full model for the evaluation of all solutions in a given large parameter space still remains prohibitive in time. Therefore, model reduction is used to obtain an approximation of a solution at a much lower computational cost. The major objective of such methods is to provide an accurate low-cost function estimation over the entire parameter space. Model reduction is currently applied in many different disciplines such as turbulent fluid flows [1–4], control design [5], structural vibration [6,7], numerical imaging and insect gait [8].

As other reduced-basis methods, Proper Orthogonal Decomposition (POD) is a kind of model reduction method. POD was developed by several authors, among which the first was Kosambi [9], and is also known as Principal Component Analysis or Karhunen–Loève Decomposition. POD purpose is to obtain low-dimensional approximate descriptions of high-dimensional phenomena. It has also been applied to repair damaged data and to construct missing data for the characterization of human faces [10]. This idea has been extended to the prediction of aerodynamic flow fields [11,12]. POD requires a set of state solutions, called snapshots, to form the reduced model span. Then, two key points need to be addressed, namely the number of basis vectors to keep and the choice of the snapshots in the parameter space. While the former is a matter of ratio between the modeled to the total information contained in a system, the latter is a sampling issue. As each sample produces the simulation of the full model, the number of samples must be as low as possible and they must capture the physical characteristics with the best possible accuracy.

When no information on the function is available, a priori sampling methods can provide a distribution of points. For a parameter space of large dimension, the uniform sampling method is often too expensive since the dimension is the exponent of the number of points per space direction. The Simple Random sampling method does not tackle properly function variations since it does not spread out the sample points uniformly enough. The Latin Hypercube sampling [40] and orthogonal sampling performing random sampling on space partitions provides a better uniformity. Quasi-Monte Carlo methods such as Hammersley, Halton or Sobol methods [38,39], that use low discrepancy deterministic sequences to generate sample points, are other popular techniques. In the context of expensive large scale system computations, the main drawback of all previous methods is that they do not take into account information on the interest function.

Indeed, if some samples are already known, they can provide useful information on functions. Adaptive sequential sampling methods add points incrementally in the parameter space and

* Corresponding author.
E-mail address: jjouhaud@cerfacs.fr (J.-C. Jouhaud).
URL: http://www.cerfacs.fr (J.-C. Jouhaud).

0045-7930/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compfluid.2010.09.002

the location of a new sample point is chosen according to a criterion which usually depends on the quality estimation of the model at this point. Halton and Sobol methods are sequential methods but they are not adaptive. These methods are the subject of active research in many applications for a long time [41] and, as an example, the greedy methods were recently derived for control problems [42,15].

This article presents an efficient computational methodology combining a reduced-basis method and an adaptive sequential sampling technique based on an a posteriori error estimator to obtain a model reduction of the steady compressible Reynolds-Averaged Navier–Stokes (RANS) equations. It can be applied at least in two major areas in aeronautical industry, namely the generation of flight data and multidisciplinary optimization. The originality of the method consists academic both the dynamic construction of the sampling database based on a Singular Value Decomposition (SVD) of the POD basis and in the fact that the error measure is evaluated through a leave-one-out process that uses all already known computations. This method yields a reliable low-order model with a low computational cost.

This article is organized as follows. The basic physical model used to evaluate sample solutions and the key features of the associated numerical method are presented in Section 2. Then, Section 3 describes the incremental SVD/POD method. Finally, numerical results are presented in Section 4 where the method is illustrated by building the response surface of an analytic test case (Michalewicz's function) and the flight envelope of a 2D transonic profile (RAE 2822).

2. Governing equations and flow solvers

2.1. Physical model

For compressible flows at high Reynolds numbers in aeronautical aerodynamics, a standard approach is to use the RANS equations.

They describe the conservation of mass, momentum and energy of a viscous Newtonian fluid flow. Using Cartesian coordinates, these equations can be expressed in a conservative form as follows:

$$\frac{\partial W}{\partial t} + \nabla.(F_e - F_v) = 0 \tag{1}$$

where the state vector W and the inviscid and viscous fluxes F_e and F_v are expressed as follows:

$$\begin{aligned} W &= [\rho, \rho\vec{U}, \rho E]^T \\ F_e &= [\rho\vec{U}, \rho\vec{U}\otimes\vec{U} + p\bar{\bar{I}}, \vec{U}(\rho E + p)]^T \\ F_v &= [0, \bar{\bar{\tau}}, \bar{\bar{\tau}}.\vec{U} - \vec{q}]^T \end{aligned} \tag{2}$$

where ρ is the density, $\vec{U}$ the velocity vector, p the pressure, E the total energy, $\bar{\bar{\tau}}$ the tensor including shear stresses and Reynolds stresses, and $\vec{q}$ the heat flux given by Fourier's law. The equations are closed with the equation of state for a perfect gas. The laminar viscosity is determined through Sutherland's law. Based on Boussinesq's hypothesis to tackle the turbulent closure problem, the eddy viscosity in Reynolds stresses is computed with the $k-\omega$ turbulence model [16].

2.2. Numerical method

Computations were carried out with the elsA code developed at ONERA and CERFACS [17]. This code solves the 3D compressible RANS equations using a cell-centered finite-volume method on structured grids.

Convective fluxes are discretized with the second order central scheme completed by the model of scalar dissipation introduced by Jameson–Schmidt–Turkel [19]. Diffusive fluxes are also approximated with a second order central scheme. The time integration is driven by the backward Euler implicit scheme. The resulting linear system is solved by the LU-SSOR method [20]. A standard multigrid [18] method combined with local time stepping accelerates the convergence to steady state solutions.

3. Adaptive POD/SVD surrogate model

3.1. Flight domain, parameters and snapshots

In the context of aircraft design, producing a reliable flow database on the full flight domain is a challenge. As a matter of fact, a flight domain may depend on a large number of physical parameters, which leads to huge computational costs when using high-fidelity Computational Fluid Dynamics (CFD) modeling. Then, a good alternative seems to build a surrogate model which offers a good approximation of the whole solution at reduced computational costs [21,22]. Such methods require only a few high-fidelity computations wisely chosen.

In the following, the continuous flight domain (or parameter space) is noted $\mathscr{F}$ and a parameter is noted $\pi = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ where m is the total number of spanned dimensions. Suppose that, for learning purpose, $\mathscr{F}$ is sampled with p parameters according to a sampling strategy. Then, the sample set is noted $\mathscr{F}_d$.

For each parameter $\pi_i \in \mathscr{F}_d$, the computed solution W_{π_i} is referred to as a *snapshot* which corresponds in the present work to a turbulent steady flow solution: the conservative variables, a derived field such as the pressure, or any aerodynamic quantity, for example the C_p distribution. If $\pi_i \in \mathscr{F} \setminus \mathscr{F}_d$, then W_{π_i} is called a *configuration*. In practice, the size n of the snapshot vector is equal to several millions when p is at most some hundreds. Finally, $S \in \mathbb{R}^{n\times p}$ is the matrix built with the snapshots as column vectors and in which snapshots are centered around the snapshot ensemble mean.

$$S = [W_{\pi_1}, \ldots, W_{\pi_p}]\left[I_p - \frac{1}{p}\mathbf{1}_p\right]$$

where $\mathbf{1}_p \in \mathbb{R}^{p\times p}$ is a matrix with 1 as common general term and $I_p \in \mathbb{R}^{p\times p}$ the identity matrix.

3.2. POD/SVD snapshot strategy

The main issue of POD is to find a few basis vectors $\Psi \in \mathbb{R}^{n\times r}$, $\Psi = \{\psi_i\}_{i=1}^r$, $r \leqslant p$ and coefficients β_i that give the best approximation of a configuration W_π in the sense that the smallest error ϵ_r is commited.

$$W_\pi = \sum_{i=1}^{r} \beta_i\psi_i + \epsilon_r \tag{3}$$

The snapshot method [23] consists in expressing the POD basis vectors as a linear combination of snapshots. if p vectors are kept (ie $r = p$), the ith vector is given by Eq. (4).

$$\psi_i = \sum_{j=1}^{p} v_i^j W_{\pi_j} \quad \text{i.e.} \quad \Psi^{(p)} \overset{def}{=} S\left[v_1, \ldots, v_p\right] \tag{4}$$

where v_i^j is the jth component of the ith right-eigenvector $v_i \in \mathbb{R}^p$ of the correlation matrix C given by Eq. (5). The matrix $C \in \mathbb{R}^{p\times p}$ is formed by inner products of every pair of snapshots.

$$C_{ij} = \frac{1}{p}(W_{\pi_i}, W_{\pi_j}) \quad \text{i.e.} \quad C = \frac{1}{p}S^TS \tag{5}$$

Eq. (3) shows that only r vectors are needed to approximate a configuration. Two alternatives are then possible to compute the basis Ψ. The first method is to complete the eigen decomposition of the

correlation matrix C, to sort the eigenvalues $(\lambda_i)_{i=1:p}$ in decreasing order and to find the index r of the strictly positive eigenvalue such that $r = \inf_i \frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{p}\lambda_k} \geqslant \varepsilon$ where ε is an arbitrary error threshold. Then, the POD basis Ψ is made of the r vectors $S\,[v_1,\ldots,v_r]$. The other method, chosen in this study, is to perform the SVD of the snapshot matrix $S = U\Sigma V^T$. The columns of orthogonal matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{p\times p}$ are respectively the left and right singular vectors. The diagonal entries σ_i of $\Sigma \in \mathbb{R}^{n\times p}$ are the singular values of S and the equation $\sigma_i^2 = p\lambda_i$ holds. Therefore, the POD basis is taken as $\Psi = U_r\Sigma_r$ where $U_r \in \mathbb{R}^{n\times r}$ (resp. $\Sigma_r \in \mathbb{R}^{r\times r}$) is the matrix built with the r first columns of U (resp. the $r\times r$ leading upper part of the matrix Σ). The SVD approach [24] has the advantage to be more reliable than the eigenvalue decomposition for computing the smallest spectral information.

An incremental version of the POD basis computation can be derived in order to reduce the memory storage [35]. The underlying idea is to update the POD basis as soon as a new snapshot is available. This way, no snapshot has to be physically stored on disk and the size of SVD remains at most equal to r instead of p. Let $S_k \in \mathbb{R}^{n\times k}$ be the snapshot matrix $[W_{\pi_1},\ldots,W_{\pi_k}]$ and assume the SVD decomposition of S_k has been computed as $S_k = U_k\Sigma_k V_k^T$.

Suppose that the new snapshot $W_{\pi_{k+1}}$ is available and that the POD basis of the matrix $S_{k+1} = [S_k\ W_{\pi_{k+1}}] \in \mathbb{R}^{n\times(k+1)}$ has to be computed. The rank-1 SVD update can be performed as proposed by Brand [33]:

$$M_{k+1} \overset{def}{=} \begin{bmatrix} \Sigma_k & U_k^T W_{\pi_{k+1}} \\ \mathbf{0}^T & \|q\| \end{bmatrix}$$

$$S_{k+1} = [S_k\ W_{\pi_{k+1}}] = \left[U_k \frac{q}{\|q\|}\right] M_{k+1} \begin{bmatrix} V_k^T & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{6}$$

where q is the column vector resulting from the Gram-Schmidt orthogonalization of $W_{\pi_{k+1}}$ against the columns of U_k. The SVD decomposition of S_{k+1} is then obtained by performing the SVD decomposition of the matrix $M_{k+1} \in \mathbb{R}^{(k+1)\times(k+1)}$, $M_{k+1} = A\Sigma_{k+1}B^T$ and then

$$S_{k+1} = \left[U_k \frac{q}{\|q\|}\right] A\Sigma_{k+1}B^T \begin{bmatrix} V_k^T & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} = U_{k+1}\Sigma_{k+1}V_{k+1}^T \tag{7}$$

It should be noticed that this incremental method also requires to recenter the snapshot matrix, which implies similar operations as the rank-1 updating and particularly another SVD decomposition of a square matrix of size $k+1$ [35].

Thus, the most expensive part of the method is the Gram-Schmidt orthogonalization step.

3.3. Reconstruction of solutions

Once the POD basis vectors are computed, any snapshot W_{π_k} in S can be written as:

$$W_{\pi_k} = \sum_{i=1}^{r} v_i^k \psi_i \tag{8}$$

Moreover, a configuration W_π, not included in the snapshot set, can be estimated by:

$$W_\pi \approx \sum_{i=1}^{r} \beta_i \psi_i \tag{9}$$

Finding the coefficients β_i from the components v_i^k is an interpolation problem. It can be considered as a supervised learning problem [25] in which $\mathscr{F}_d$ is referred to as the learning set.

Several techniques are available to perform this reconstruction step, among which cubic splines [26], neural networks [27], radial basis function networks [28] and Kriging [29]. In the Kriging method used in this paper, the function to interpolate is represented by a constant regression model plus localized deviations that match the sample values. Deviations are considered as realizations of a normal distributed random process completely defined by its matrix of correlation and its variance. For Kriging, the correlations between different sample values of the model depend only on the distance between these realizations through a set of parameters that must be adjusted. The method used to adjust these parameters is the maximum of likelihood estimate [30,31].

3.4. Error analysis, splitting and resampling

The originality of this work is emphasized in this section. It relies on a self-adaptive strategy to estimate and increase the accuracy of the POD/SVD surrogate model. This strategy is based on a leave-one-out method to estimate the error induced by the model and a refinement method of the flight domain in order to add snapshots.

3.4.1. Leave-one-out method

The leave-one-out method is a special case of the cross validation algorithm described in [32]. Data are split into two parts: the training set is used to build a surrogate model and the test set is used to validate the latter. In the leave-one-out method, the size of the test set is simply equal to 1.

Consider at some step of the adaptive process that the POD basis $\Psi^{(p)} = \{\psi_i\}_{i=1:p}$ is known and that the SVD decomposition of the snapshot matrix S containing p snapshots of size n is $S = U\Sigma V^T$ where $U \in \mathbb{R}^{n\times p}$ and $V \in \mathbb{R}^{p\times p}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{p\times p}$ is a diagonal matrix. Let $S^{(-k)} \overset{def}{=} [W_{\pi_1},\ldots,W_{\pi_{k-1}},W_{\pi_{k+1}},\ldots,W_{\pi_p}] \in \mathbb{R}^{n\times(p-1)}$ be the matrix that is nothing more than the matrix S where the snapshot k has been removed, and $\Psi^{(-k)} \overset{def}{=} \{\psi_i^{(-k)}\}_{i=1:p-1}$ the POD basis of $S^{(-k)}$.

$\Psi^{(-k)}$ can be efficiently deduced from Ψ using the following SVD downdating algorithm [35]. Let $\widetilde{V} \in \mathbb{R}^{(p-1)\times p}$ be the orthogonal matrix associated to the matrix S to which the kth row is removed. Then,

$$S^{(-k)} = U\Sigma\widetilde{V}^T \tag{10}$$

Substituting $\widetilde{V}$ by its QR factorization [24] into Eq. (10) gives

$$S^{(-k)} = U\Sigma R^T Q^T$$

Performing the SVD decomposition of $\Sigma R^T \overset{def}{=} U'\Sigma' V'^T$, Eq. (10) simplifies to

$$S^{(-k)} = UU'\Sigma' V'^T Q^T \overset{def}{=} U^{(-k)}\Sigma' V^{(-k)^T}$$

which is just the SVD decomposition of $S^{(-k)}$. Thus, the POD basis of $S^{(-k)}$ is given by $\Psi^{(-k)} \overset{def}{=} U^{(-k)}\Sigma'$.

A configuration W_π can then be estimated at π_k in the POD basis $\Psi^{(-k)}$ by $\widetilde{W}_{\pi_k} = \sum_{i=1}^{r}\beta_i\psi_i^{(-k)}$ where β_i are computed using the reconstruction step detailed in Section 3.3. Then, $E_{\pi_k} = \|W_{\pi_k} - \widetilde{W}_{\pi_k}\|_2$ is a good local error estimate for the surrogate model in the neighborhood of π_k. A global error estimate Δ_{F_d} of the learning set F_d can be derived as $\Delta_{F_d} = \frac{1}{p}\sum_{i=1}^{p} E_{\pi_i}^2$ where E_{π_i} is the leave-one-out error estimate at π_i.

3.4.2. Adaptive flight domain refinement

The previous error analysis gives a way to refine the flight domain by detecting regions with large errors. In these regions, the surrogate model can be improved by adding snapshots which will enrich the POD basis. This strategy is called *resampling* and is implemented using a *quad-tree* data structure.

Quad-tree was first introduced in the seventies by Finkel and Bentley [34]. If $d > 1$ is the dimension of the space to be refined, the space is recursively subdivided into 2^d equally sized regions. For each internal node, 2^d children at most (only one child is defined for the nodes at the corner delimiting the space) are defined. The children are taken as the centers of the subdivisions.

For the flight domain, the nodes are the parameters π_i and the quad-tree refinement is performed at the parameter π_k having the largest value of E_{π_k}. Using the quad-tree data structure, new parameters are added near π_k in order to improve the quality of the surrogate model by adding the corresponding new snapshots to the learning set. One or 2^d children are defined for the new

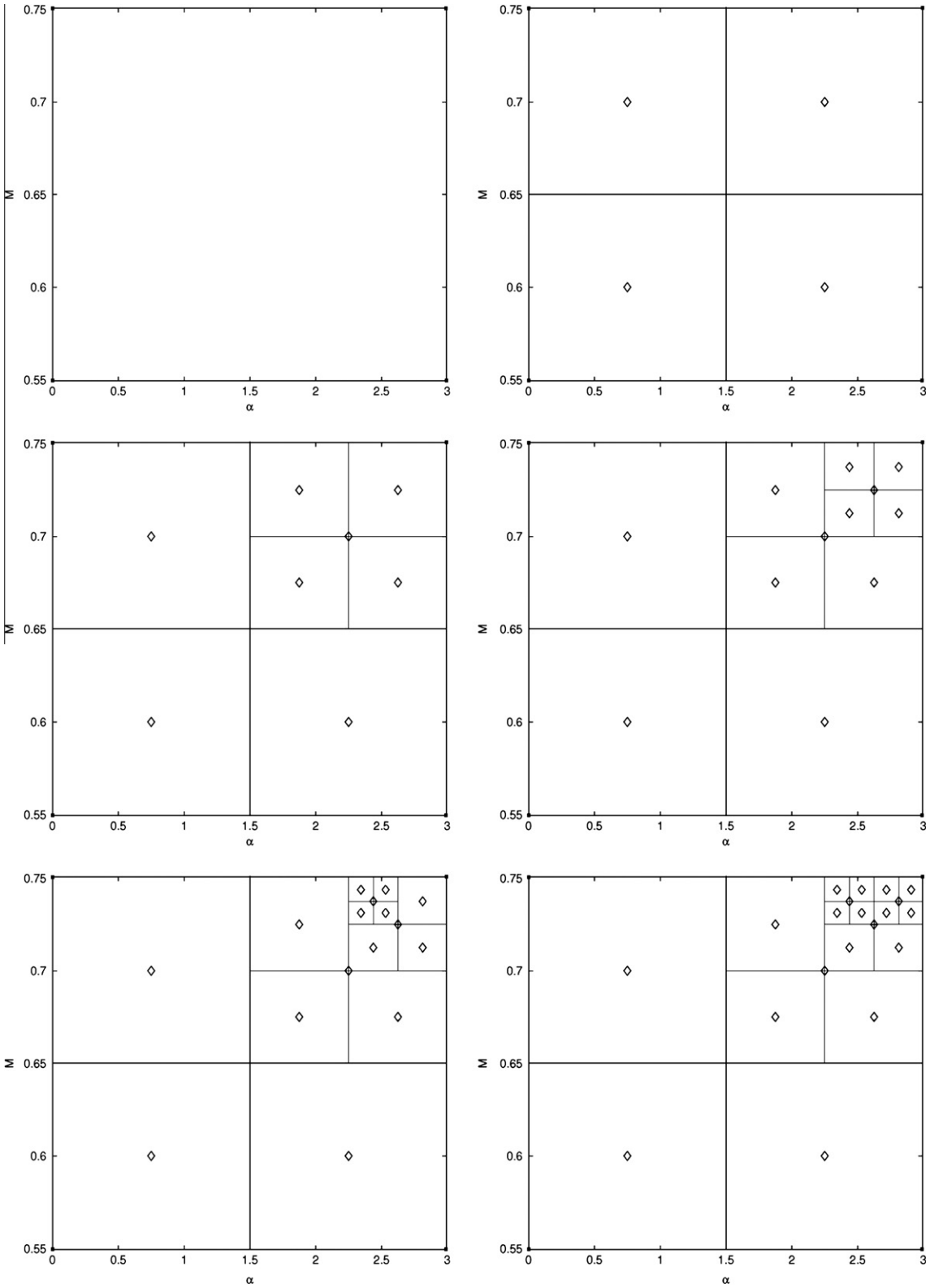

Fig. 1. Adaptive flight domain refinement on the parameter space of the RAE2822 airfoil test case.

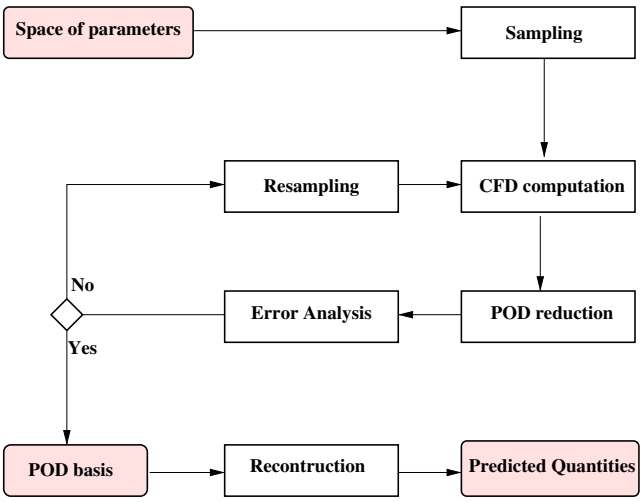

Fig. 2. An adaptive POD/SVD chain.

subdivision of the cell around π_k. The node π_k is then frozen: no more refinement can be performed for it. The process is illustrated on a 2D parameter space in Fig. 1.

Table 1
Michalewicz's function simulations.

Number p of snapshots for	$Er < 13\%$	$Er < 4\%$	$Er < 1\%$
RE	20	28	**36**
UD	25	49	64
HA	22	55	71
SO	34	45	127

Then, new leave-one-out error estimates are computed for the new set of parameters and this process is carried on until the quality indicator Δ_{F_d} is less than a threshold value η given by the user. The resampling process leads to an adaptive POD/SVD Chain which is illustrated in Fig. 2.

3.4.3. Initial sampling

Contrary to optimization algorithms, our aim here is not explicitly to automatically detect the local optimum. In fact, a good estimation of the solution is expected on the whole domain, which implies to detect all the strong variations, not only around the optimum. Nevertheless, when the solution exhibits many different patterns over the whole parameter space, the sampling could miss important features if no snapshots capture these patterns. This problem concerns all sampling techniques.

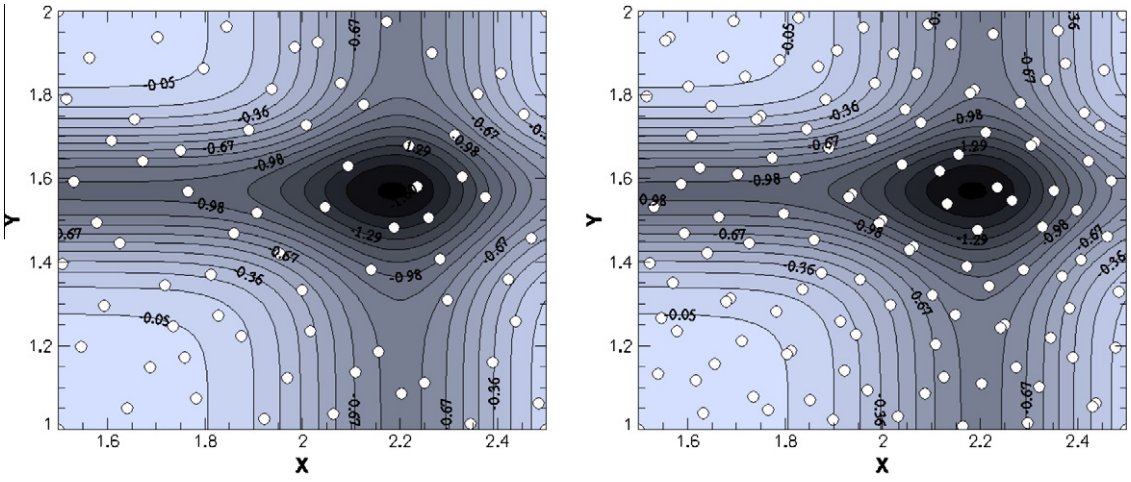

Fig. 3. Michalewicz's function, Halton (left, 71 snapshots) and Sobol (right, 127 snapshots) techniques.

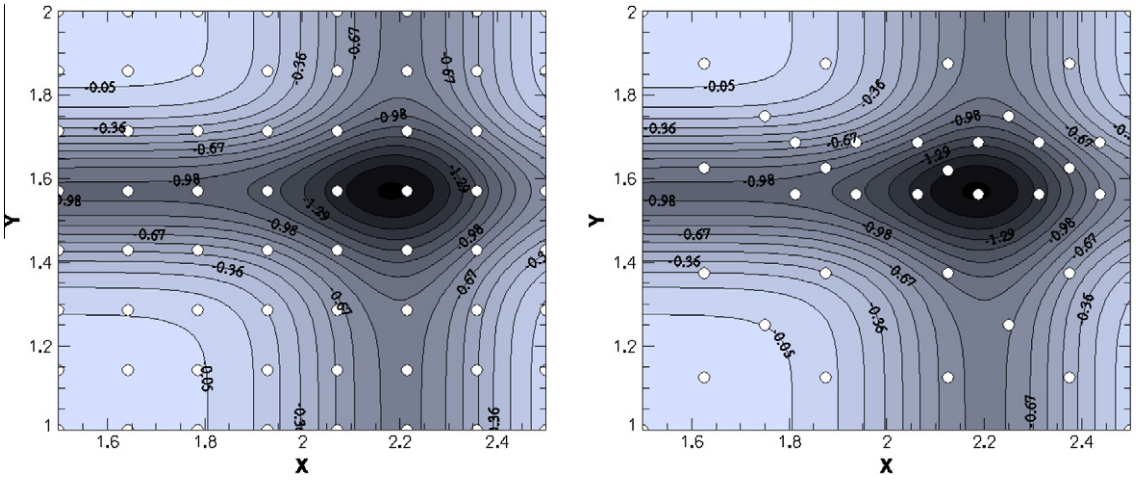

Fig. 4. Michalewicz's function, Uniform Distribution (left, 64 snapshots) and resampling (right, 36 snapshots) techniques.

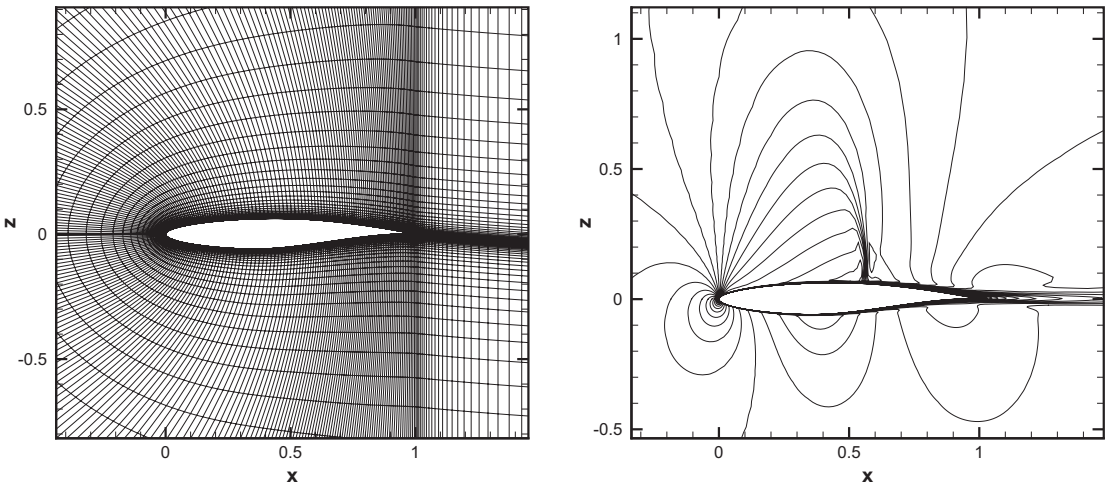

Fig. 5. RAE2822, mesh and Mach number contours at $\pi = (2.79°, 0.73)$.

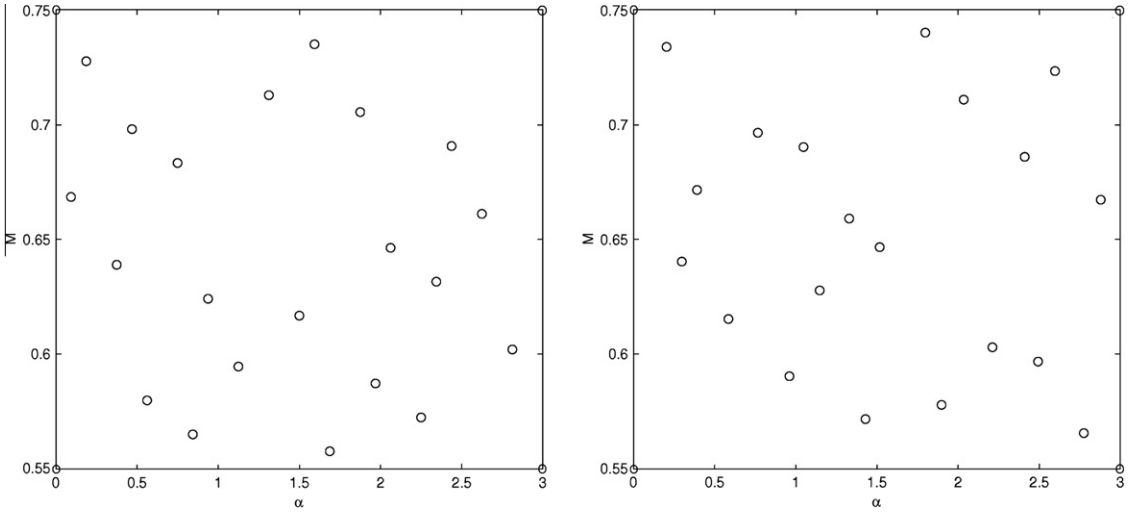

Fig. 6. RAE2822, Halton (left) and Sobol (right) samplings.

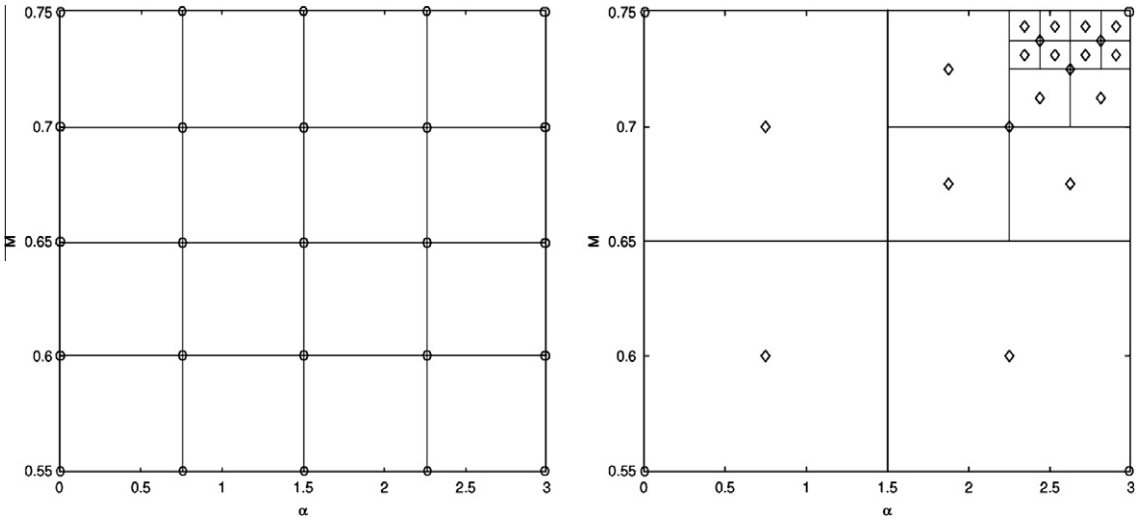

Fig. 7. RAE2822, Uniform Distribution (left) and resampling (right).

In our adaptive POD/SVD Chain, the initial sampling can be realized by *a priori* sampling techniques (Uniform Distribution, Halton or Sobol algorithms) in order to increase the number of initial snapshots whenever the solution is expected to be complex. But for our aerodynamic applications, starting with only the corners of the parameter space seems to be sufficient to enable a relevant enrichment.

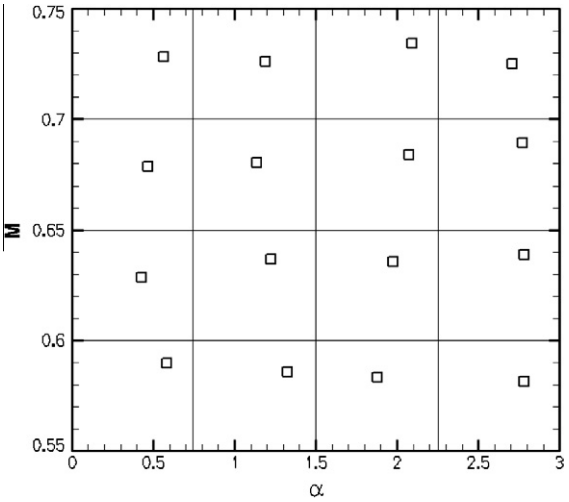

Fig. 8. RAE2822, independent CFD computations (squares).

4. Numerical results

4.1. Basic evaluation on a Michalewicz test function

In order to estimate the effectiveness of our adaptive POD/SVD strategy, computations were firstly performed on a 2D Michalewicz function type [37] which represents a multimodal test function defined as follows:

$$F_{12}(x,y) = -\sin(x)\left[\sin\left(\frac{x^2}{\pi}\right)\right]^{2l} - \sin(x)\left[\sin\left(2\times\frac{y^2}{\pi}\right)\right]^{2l}, \quad 0 \leqslant x,y \leqslant \pi \tag{11}$$

The parameter l defines the "steepness" of the valleys and is fixed to $l = 10$. This test case is studied on the parameter space $[1.5,2.5]\times[1,2]$ where the function has an optimum located at the point $(2.185,1.570)$.

To check the convergence in terms of accuracy, the Resampling algorithm (**RE**) is compared to several strategies: Uniform Distribution

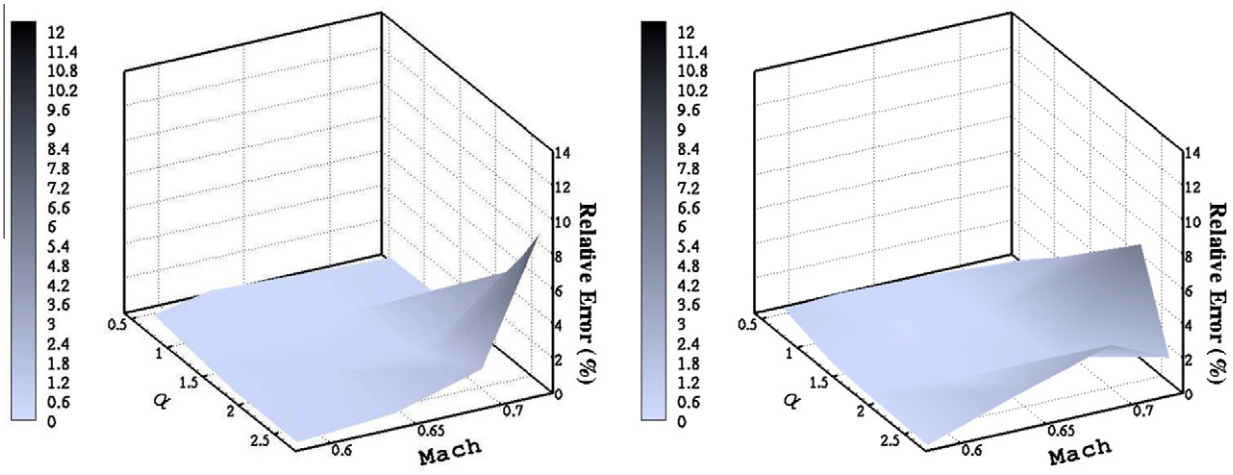

Fig. 9. Relative errors for **HA** (left) and **SO** samplings (right).

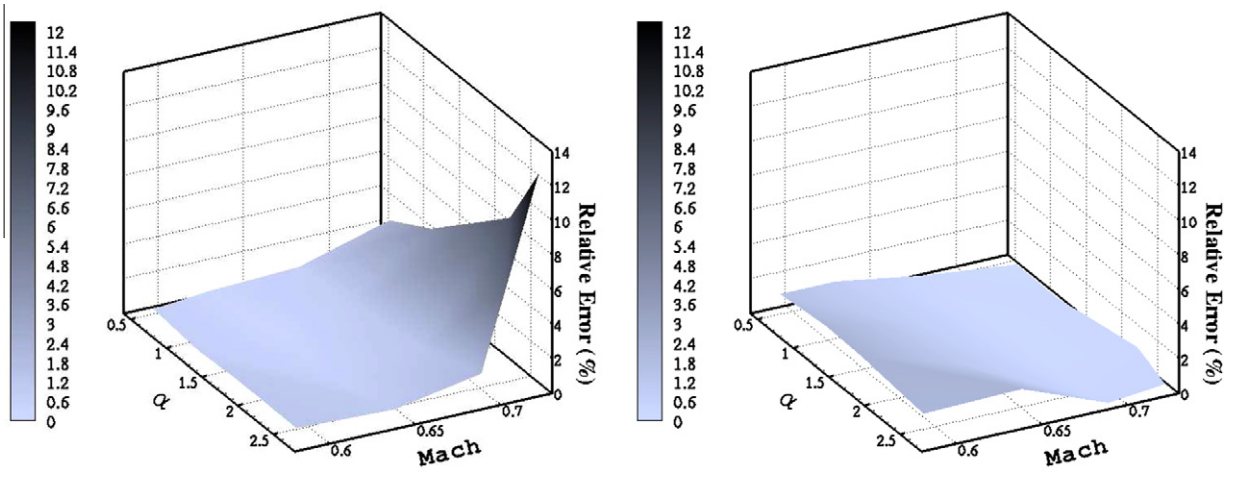

Fig. 10. Relative errors for **UD** (left) and **RE** samplings (right).

(**UD**), Halton algorithm (**HA**) and Sobol algorithm (**SO**), starting from the 4 corners of the parameter space. Figs. 3 and 4 represent the snapshots W_{π_i} ($1 \leqslant i \leqslant p$, where p is the number of snapshots) obtained with these algorithms. For this particular test case: $\pi_i = (x_i, y_i)$ and $W_{\pi_i} = F_{12}(x_i, y_i)$.

UD, HA, and SO algorithms are representative of *a priori* samplings. In case of UD, the sampling is highly deterministic, whereas for HA an SO it is a quasi-random one, i.e. they are actually deterministic but they are of low discrepancy. A HA sampling is formed by reversing the digits in the representation of some sequence of integers in a given base [38]. SO sequences use a base of two to form successively finer uniform partitions of the unit interval, and then reorder the coordinates in each dimension [38,39]. For all these sampling techniques, the index r defined in Section 3.2is equal to 2 ($\varepsilon = 0.98\%$).

Using each POD/SVD surrogate model (UD, HA, SO or RE), 10.201 predictions $\widetilde{W}_{\pi_k}$ have been performed with the parameter π_k lying on the following 101×101 uniform grid: $[x = 1.5\text{–}2.5, \Delta x = 0.01] \times [y = 1.0\text{–}2.0, \Delta y = 0.01]$. These predictions are used to compute the maximum relative error $Er = \max_k \left| \frac{\widetilde{W}_{\pi_k} - W_{\pi_k}}{W_{\pi_k}} \right|$ between the predicted field and the true one.

The comparison between the different algorithms is presented in Table 1 and shows that the RE technique gives the best results. In fact, much less snapshots are needed for the same accuracy.

4.2. RAE2822 airfoil simulations

The adaptive POD/SVD chain has been secondly evaluated considering the turbulent flow around a RAE2822 airfoil with a structured C-type mesh containing 45.056 cells (see Fig. 5). This case has been extensively used for validation of RANS codes applied to transonic airfoil flow since it was retained as an international test case by AGARD [36]. The flight domain is the following: the angle of attack $\alpha \in [0°, 3°]$ and the Mach number $M \in [0.55, 0.75]$. It makes a subsonic-transonic transition occurring, which must be detected and refined by the flight domain sampling procedure.

4.2.1. Analysis of refinement strategy

To estimate the efficiency of the method, we still have performed a comparison between RE, UD, HA and SO algorithms, starting from the 4 corners of the flight domain in the (α, M) plane (see Figs. 6 and 7) and ending with a total of 25 snapshots (CFD computations). For all these algorithms, the first six POD modes ($\varepsilon = 0.98\%$) have been kept. CFD computations have been performed for 16 parameters randomly located inside 16 different regions (see Fig. 8). These parameters are totally different from those used to build the POD/SVD surrogate model. In this way, the predicted values obtained for these parameters have been compared to the exact values.

Figs. 9 and 10 show the relative errors between the CFD computations and the predicted values: $Er_{\pi_k} = \left\| \frac{W_{\pi_k} - W_{\pi}}{W_{\pi}} \right\|_{\infty} \times 100$, where

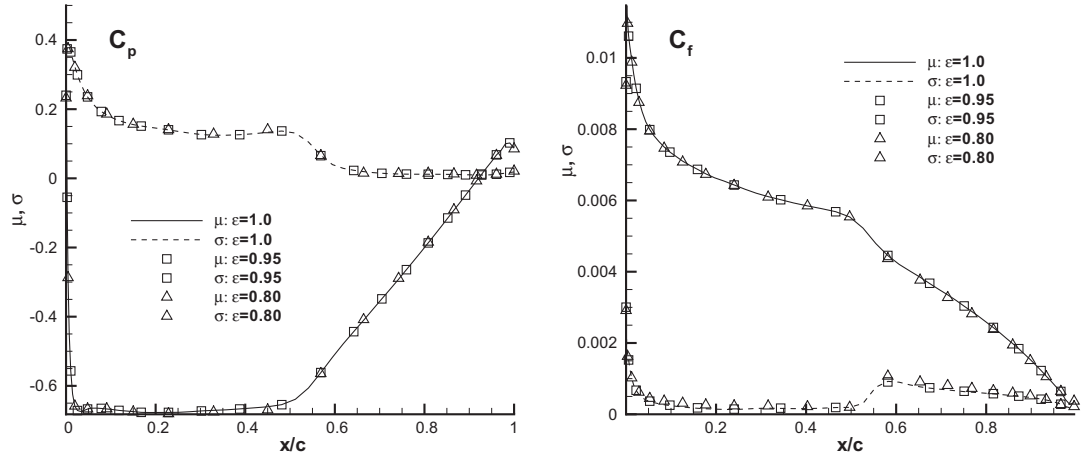

Fig. 11. Mean value (μ) and standard deviation (σ) of C_p and C_f on the airfoil for the uniform sampling case.

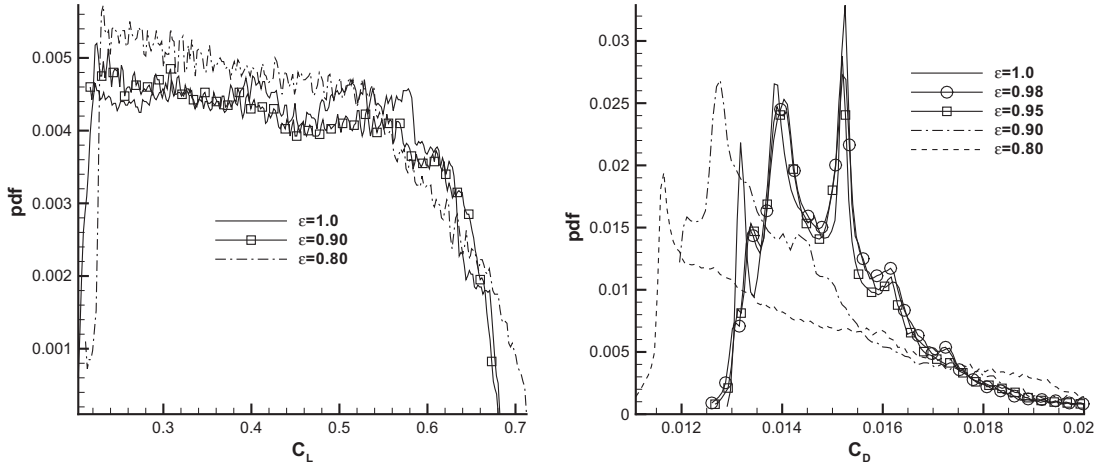

Fig. 12. Probability density function of C_L and C_D for the uniform sampling case.

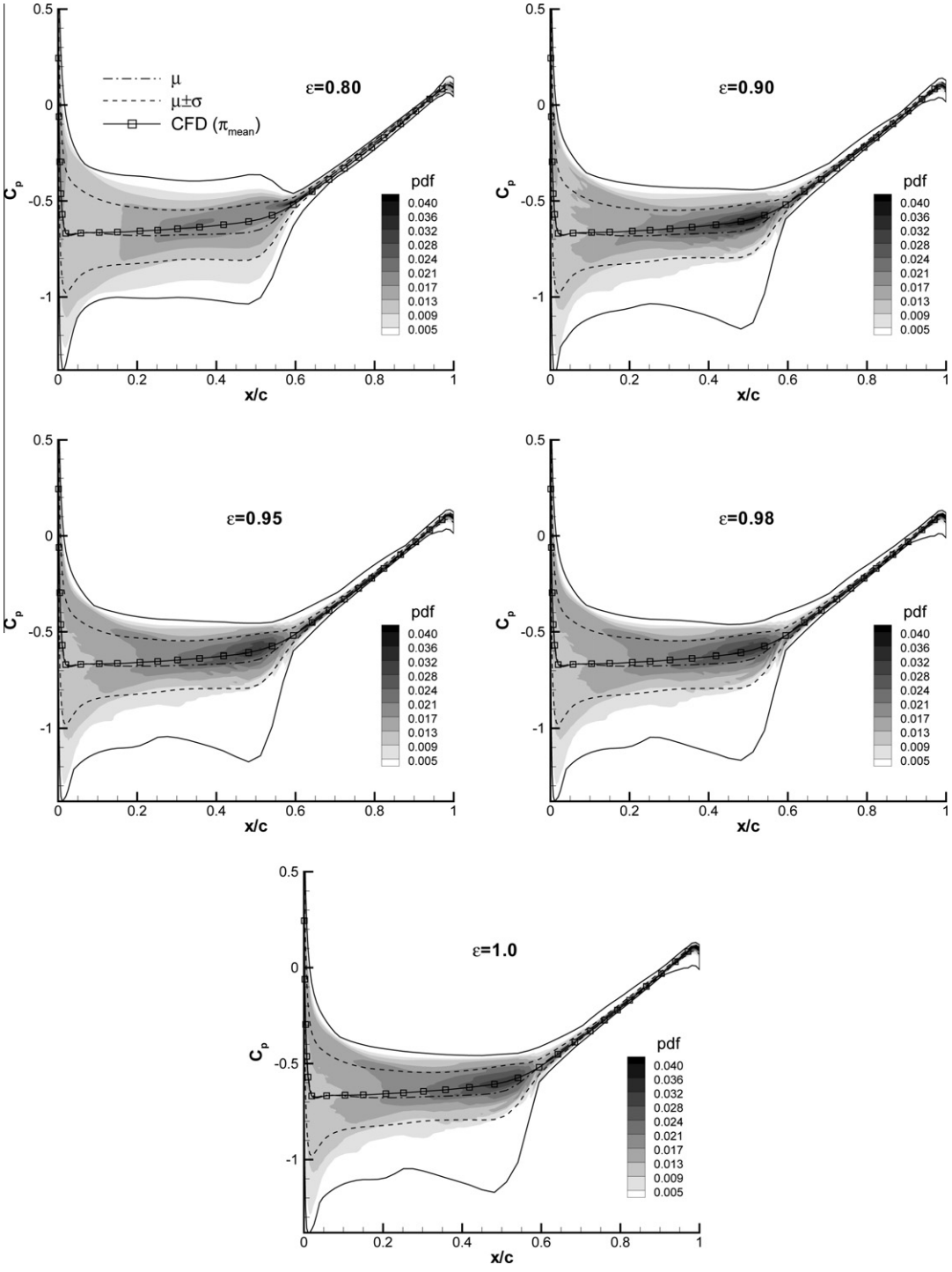

Fig. 13. Probability density function of the extrados C_p distribution for the uniform sampling case.

W_{π_k} are the CFD computations (conservative variables) and W_π the predicted CFD solutions. HA, SO and UD techniques predict well the subsonic area because the errors are small, but the transition zone between subsonic and transonic is not well captured. Only the RE technique indicates a global error smaller than 2% on the entire flight domain. This phenomenon can be explained by the fact that RE automatically adds samples in the top-right corner of the space of parameters where the subsonic-transonic transition occurs (see Fig. 7).

Concerning the CPU costs, the following wall clock time have been obtained on an IBM Nehalem cluster computer: 997.91 s for computing the 25 snapshots with the CFD code, around 2.80 s to compute the POD basis in the case of HA, SO and UD, 7.58 s to reach the final POD basis in the case of RE, and around 0.02 s for reconstructing one desired solution (all algorithms considered). Finally, the RE technique leads to a significant reduction of the global error for only an additional cost of 0.47%.

4.2.2. Statistical analysis of the method

The influence of the error threshold ε on the POD solution is now analyzed in order to find a reasonable range to set it. Two

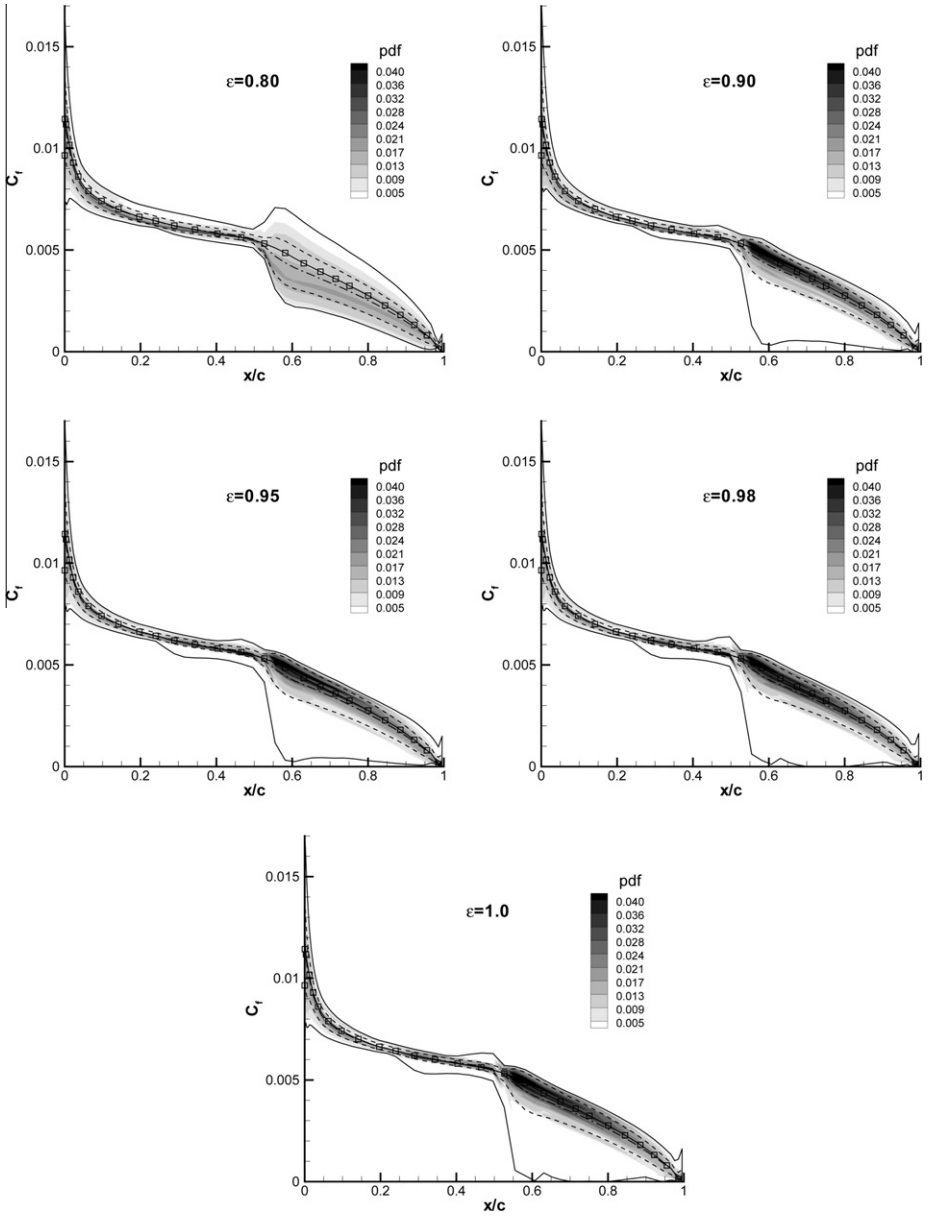

Fig. 14. Probability density function of the extrados C_f distribution for the uniform sampling case.

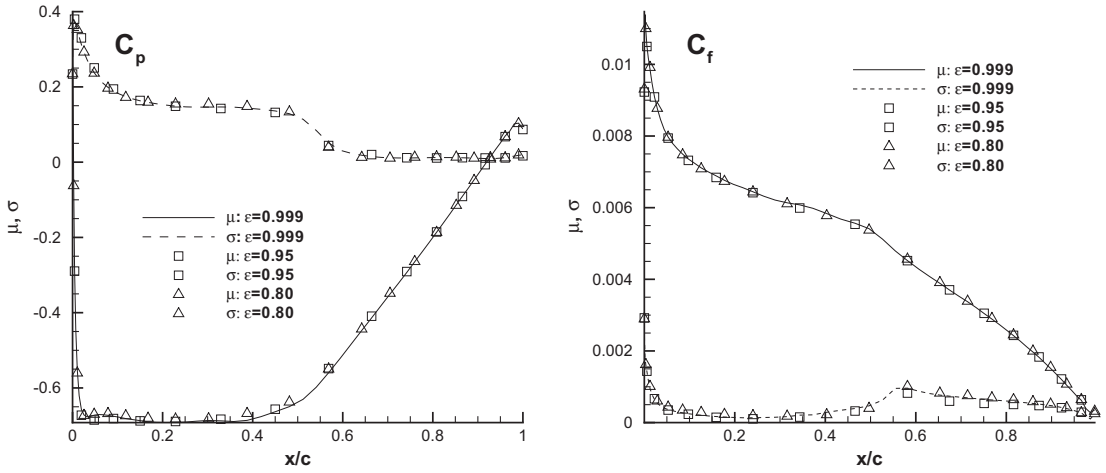

Fig. 15. Mean value (μ) and standard deviation (σ) of C_p and C_f for the resampling strategy.

types of sampling strategies are considered: the uniform sampling and the resampling procedure based on the leave-one-out method.

The POD/SVD procedure enables the reconstruction of the full response surface of the problem. To get a tractable information on the accuracy of the response surface, it is then necessary to focus on quantities of primary interest for aerodynamic purpose.

Therefore, the pressure and skin friction distributions C_p and C_f on the suction side of the airfoil, and the global lift and drag coefficients C_L and C_D are selected to represent snapshots in the POD reduction procedure. Mean values of C_p and C_f (obtained by averaging over the full response surface) and standard deviation distributions computed using uniform sampling are displayed in Fig. 11. It is observed than even with a low-definition POD basis ($\varepsilon = 0.8$), the prediction of global statistical quantities is satisfactory, compared to the case in which all the POD modes are used ($\varepsilon = 1.0$).

This analysis is now refined looking at the probability density functions (PDF) of these quantities (see Fig. 12). Probability density functions give a more accurate view at the variability of the solution than fully averaged quantities such as mean values. The increase of the number of selected POD modes leads to a clear convergence of the PDF of C_L and C_D. Further insight into the variability of the response surface can be gained by plotting the PDF of pressure and friction coefficients along the pressure side of the airfoil. On Figs. 13 and 14, the PDFs are superimposed with: the mean value obtained by averaging the response surface (dash-dot line), the variability envelope obtained by adding/subtracting the standard deviation to the previous mean value (dotted line) and the usual CFD solution obtained running a new computation for the mean value of M and α.

Several conclusions can be drawn looking at these two figures. First, the flight domain mean value is not equal to the solution obtained via a single computation at mean values of the flight domain parameter, showing the need for a full flight domain exploration. Second, the variability of the solution is much larger than the one

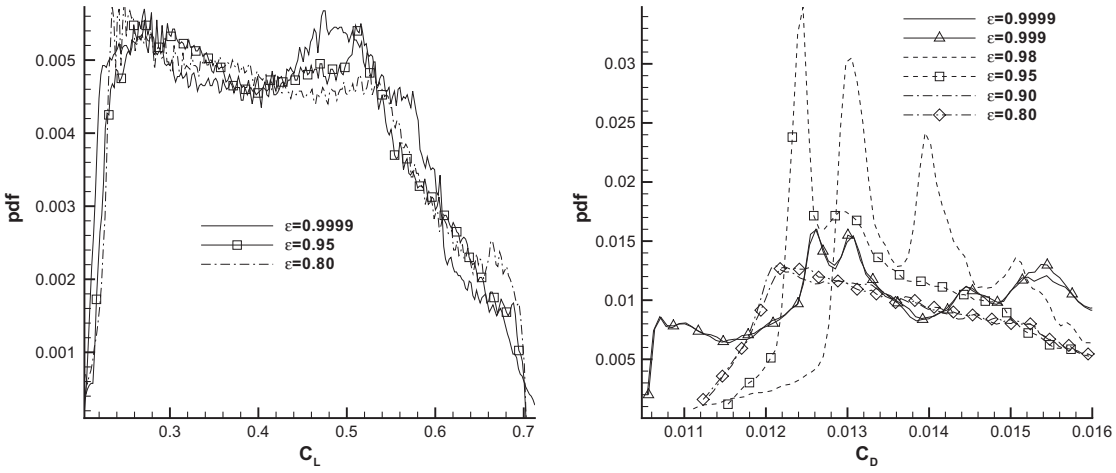

Fig. 16. Probability density function of C_L and C_D for the resampling strategy.

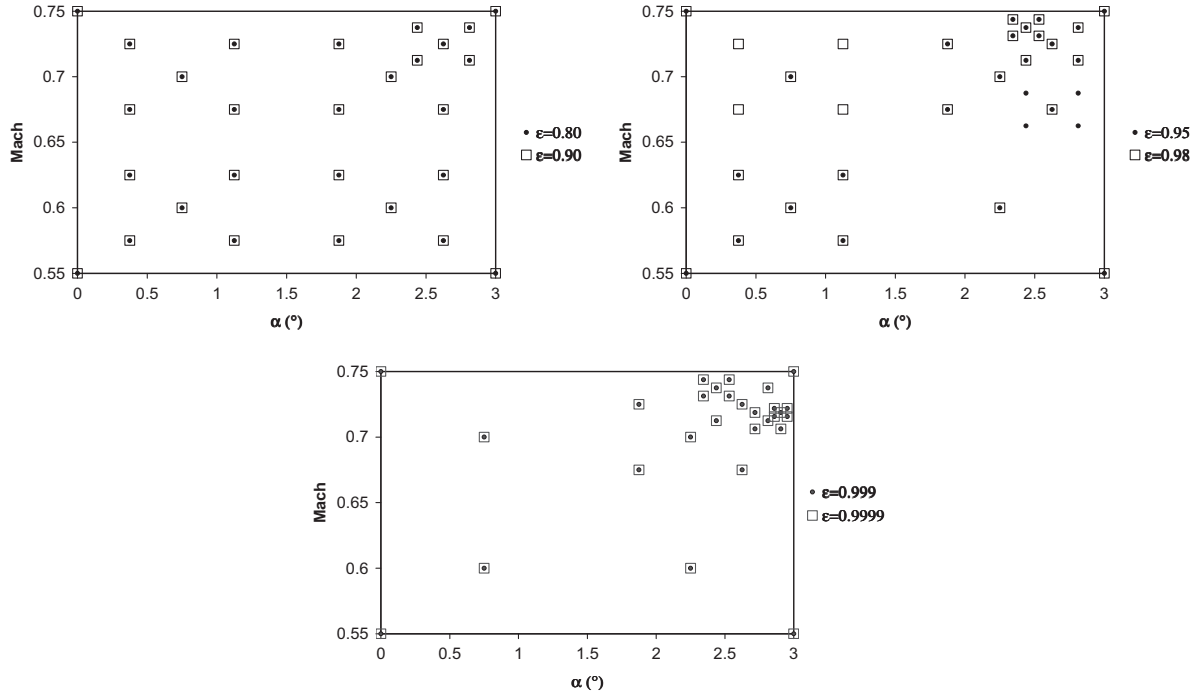

Fig. 17. Effect of ε on the resampling strategy (in each case, the number of snapshots is the same).

predicted using the standard deviation: values with non-negligible probability are observed outside the standard-deviation-based envelope. Thin solid lines correspond to the lower and upper values detected screening the response surface. It is clearly seen that extreme values, which differ from the mean value, are found thanks to the POD/SVD response surface. It is observed that a value of the error threshold too small results in a dramatic smoothing of the PDF, and a significant under-evaluation of the solution variability. A satisfactory description is recovered for $\varepsilon = 0.9$ in the present case.

For the resampling case, Fig. 15 shows the same behavior than for the uniform case. An important result is that very satisfactory prediction of mean values of pressure and skin friction distributions are recovered even with $\varepsilon = 0.8$. But, concerning the PDF of C_L and C_D, Fig. 16 reveals that the PDF can be split into three categories: $\varepsilon = 0.8$–0.9, $\varepsilon = 0.95$–0.98 and $\varepsilon = 0.999$–0.9999. Identical observations can be made in Fig. 18 for C_p and Fig. 19 for C_f. The reason is that increasing ε induces a resampling (see Fig. 17) in the region where the flow dynamics is the most elaborated with high Mach number and high angle of attack. Therefore, the location of the samples within the flight domain is ε-dependent, rendering the convergence of the method much more complex to analyze from a theoretical point of view than in the uniform sampling case.

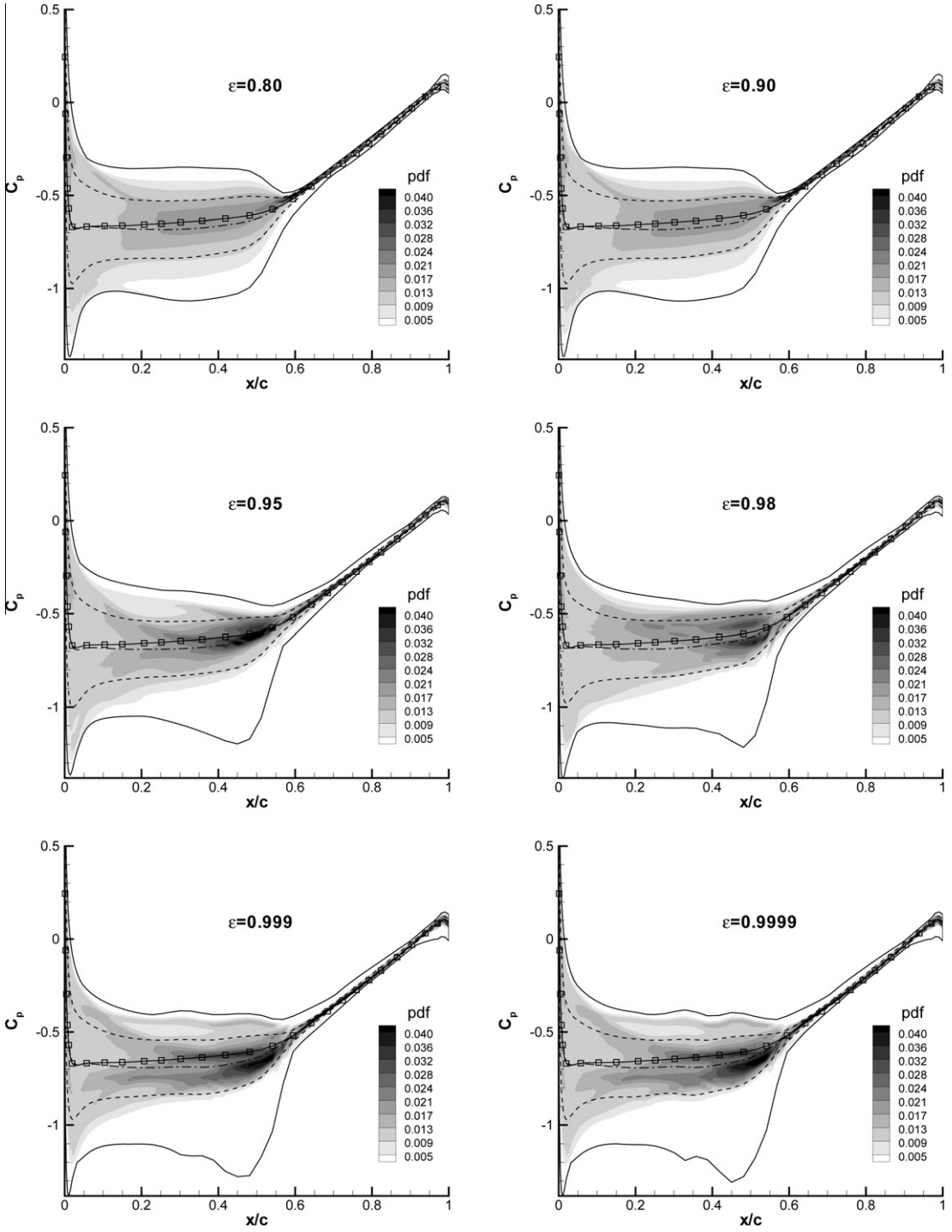

Fig. 18. Probability density function of the extrados C_p distribution for the resampling strategy.

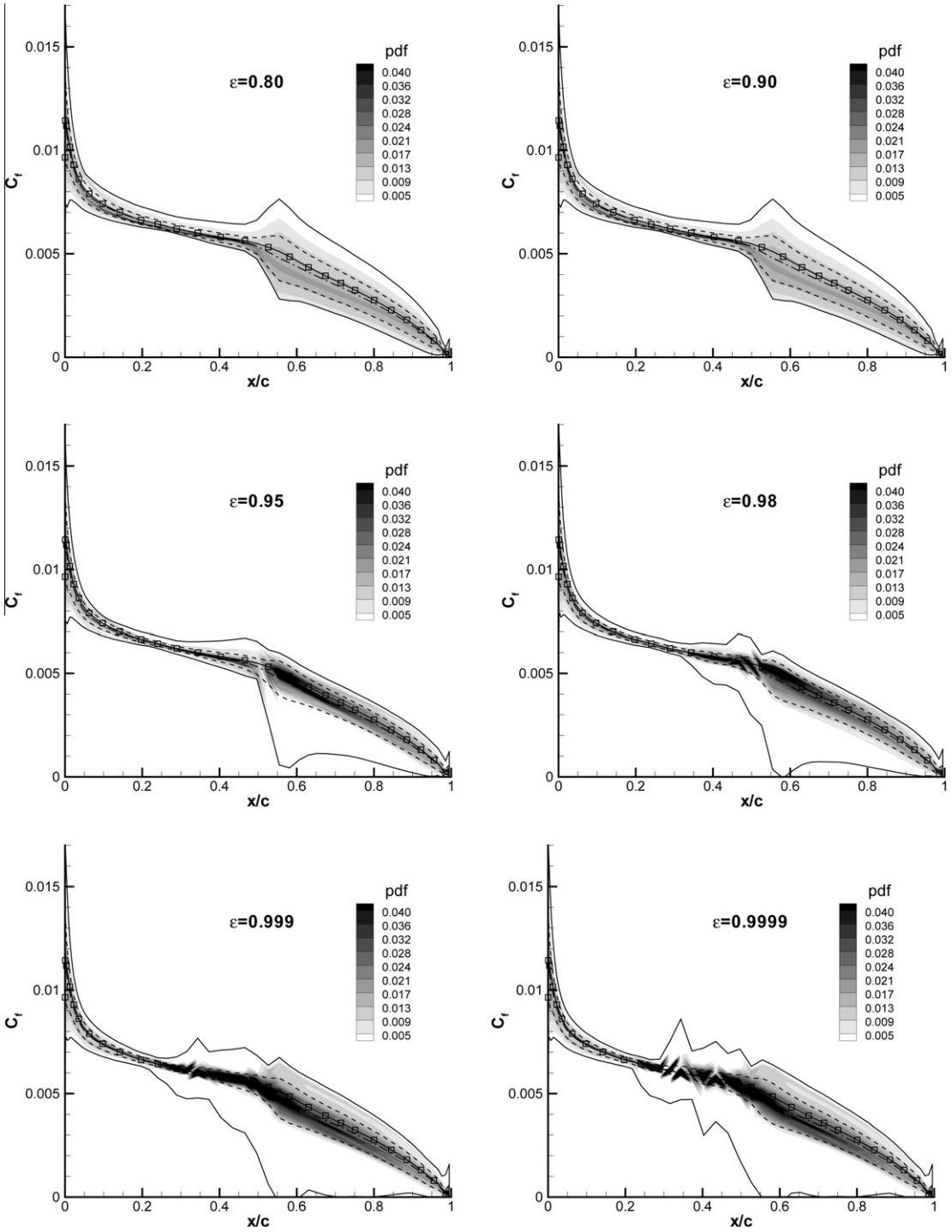

Fig. 19. Probability density function of the extrados C_f distribution for the resampling strategy.

4.2.3. Efficiency of the resampling strategy

To obtain an estimation of the CPU cost/quality ratio, it is interesting to compare the convergence of the PDF's as a function of the number of snapshots involved in the building process of the POD basis. To assess the efficiency of the RE sampling, this latter is compared to a Latin Hypercube (LH) sampling which is representative of *a priori* sampling method (HA, SO, UD, see Fig. 20). A Latin square is a square grid for which only one sample is taken in each row and each column. The LH sampling is a generalization of this concept to a n-dimensional space [40]. The major advantage with such a sampling is that it does not degenerate when n becomes large. For both cases (LH and RE), the objective is to compare the convergence rate as the number of snapshots is increased from $p = 4$ to $p = 44$ with a step of four snapshots. For each p a statistical Convergence Indicator (CI) is calculated as follows:

$$CI = \frac{E_{PDF}(p)}{E_{PDF}(4)} \tag{12}$$

where $E_{PDF}(p)$ is defined by:

$$E_{PDF}(p) = \|PDF(p) - PDF(44)\|_2 \tag{13}$$

According to Eq. (12), CI must take the value 1 for $p = 4$.

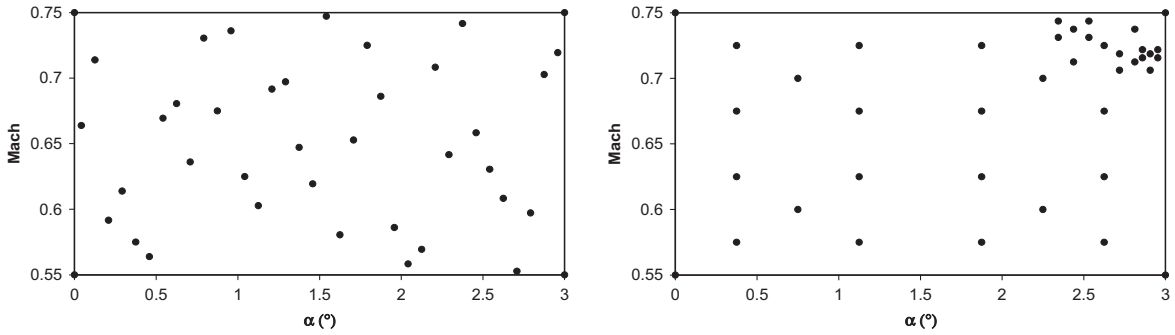

Fig. 20. RAE2822, Latin Hypercube (LH) sampling and adaptive Resampling (RE) with 40 snapshots.

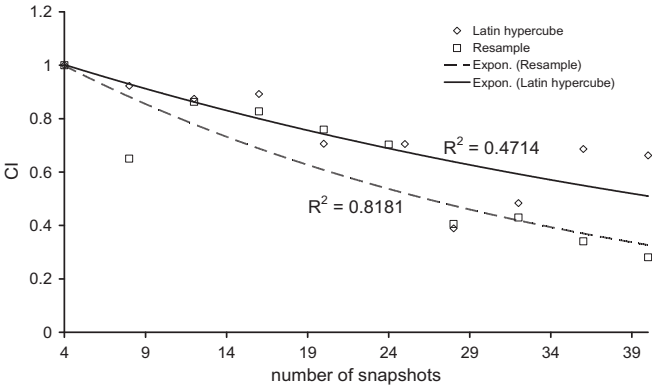

Fig. 21. Evolution of the statistical Convergence Indicator (CI) as a function of the number of snashots.

Fig. 21 presents the evolution of CI for PDFs of C_D. The symbols are the actual values of CI and the lines represent exponential regressions of these values. These regressions are calibrated such that the curve passes by the point of coordinates $(4,1)$. The coefficient of regression R is indicated for both cases.

Firstly, it can be noticed that the data are rather scattered. However, they are more correlated in case of the RE sampling ($R = 0.82$) than in case of the LH one ($R = 0.47$). Secondly, after 24 snapshots the evolution of the data is quite (decreasing) monotonous for the RE case whereas it is very erratic for the LH one. Finally, the regression curve of the RE sampling decreases more rapidly than for the LH sampling which denotes a better efficiency of the RE method to reach a converged state, at least in terms of density of probability. Furthermore, additional costs associated to resampling strategy are less than 1% of global costs.

5. Conclusions

The goal of this paper was to present an efficient way to reconstruct the global aerodynamic fields for a given space of parameters. An automatic POD/SVD resampling strategy have been proposed in order to obtain the best accuracy as possible with the smallest number of snapshots. This strategy is carried on using a two step procedure: first reducing the dimensionality of the flow through an appropriate reduced basis, i.e. the POD basis, then using this reduced data basis to predict the desired quantities and to refine the space of parameters. To achieve this, a relative error quality indicator of the model using the leave-one-out procedure has been developed. It allows a sequential design of computer experiments with a progressive enrichment of the data basis which efficiently reduces the error. The proposed strategy has been tested on an analytic test case and on the 2D turbulent flow around a RAE2822 airfoil.

Compared to other classical a priori sampling such as uniform sampling, Halton and Sobol, the accuracy is controlled on the whole domain so to have an uniform repartition of the error without local extreme values. In terms of data reduction, the important control parameter of the method is ε, the threshold level of energy that is captured by the POD basis. The influence of this parameter is evaluated for two sampling strategies: uniform sampling (UD) and adaptive Resampling (RE). For the studied case, it is observed that a rather low value of ε (0.8) is sufficient to obtain the mean behavior of the aerodynamic forces. Nevertheless, for such a low value, the variability of the response, in terms of PDF, is underevaluated, and higher values of ε should be selected (at least 0.9 for the studied case). Note that, in case of the RE sampling, a great variability of the snapshot repartition in the space of parameters is observed with the variation of ε. Finally, the efficiency of the RE method is assessed by evaluating its statistical speed of convergence, as the number of snapshots is increased, and in terms of PDF repartition. It is shown that this speed of convergence is greater for the RE sampling than for the Latin Hypercube one, which is representative of *a priori* sampling.

References

[1] Lumley JL. Stochastic tools in turbulence. New-York: Academic Press; 1971.
[2] Berkooz G, Holmes P, Lumley JL. The proper orthogonal decomposition in the analysis of turbulent flows. Annu Rev Fluid Mech 1993;25:539–75.
[3] Couplet M, Basdevant C, Sagaut P. Calibrated reduced-order POD-Galerkin system for fluid flow modelling. J Comput Phys 2005;207(1):192–220.
[4] Couplet M, Sagaut P, Basdevant C. Intermodal energy transfers in a proper orthogonal decomposition-Galerkin representation of a turbulent separated flow. J Fluid Mech 2003;491:275–84.
[5] Ito K, Ravindran SS. A reduced-order method for simulation and control of fluid flows. J Comput Phys 1998;143(2):403–25.
[6] Cusumano JP, Sharkady MT, Kimble BW. Experimental measurement of dimensionality and spatial coherence in the dynamics of a flexible-beam impact oscillator. Philos Trans Roy Soc Lond 1994;347:421–38.
[7] Feeny BF, Kappagantu R. On the physical interpretation of proper orthogonal modes in vibrations. J Sound Vib 1998;211:607–16.
[8] Saranli U, Buehler M, Koditschek D. Design, modeling and preliminary control of a compliant hexapod robot. In: Proceeding of IEEE ICRA, San Francisco, CA; 2000.
[9] Kosambi DD. Statistics in function space. J Ind Math Soc 1943;7:76–88.
[10] Everson R, Sirovich L. The Karhunen–Loeve procedure for gappy data. J Opt Soc Am 1995;12:1657–64.
[11] My-Ha D, Lim KM, Khoo BC, Willcox K. Real-time optimization using proper orthogonal decomposition: free surface shape prediction due to underwater bubble dynamics. Comput Fluids 2007;36(3):499–512.
[12] Bui-Thanh T, Damodaran M, Willcox K. Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. AIAA J 2004;42(8):1505–16.
[13] Lieu T, Farhat C. Adaptation of aeroelastic reduced-order models and application to an F-16 configuration. AIAA J 2007;45(6):1244–57.
[14] Amsallem D, Farhat C. Interpolation method for adapting reduced-order models and application to aeroelasticity. AIAA J 2008;46(7):1803–13.

[15] Bui-Thanh T, Willcox K, Ghattas O. Model reduction for large-scale systems with high-dimensional parametric input space. SIAM J Sci Comput 2008;30(6):3270–88.
[16] Wilcox DC. Simulation of transition with a two-equation turbulence model. AIAA J 1994;32(2):247–55.
[17] Cambier L, Veuillot, J-P. Status of the elsA CFD Software for flow simulation and multi-disciplinary applications. In: 46th Aerospace sciences meeting and exhibit, AIAA, Paper 2008-664; 2008.
[18] Jameson A. Steady state solutions of the euler equations for transonic flow by a multigrid method. Advances in scintific comp.. Academic Press; 1982. pp. 37–70.
[19] Jameson A, Schmidt W, Turkel E. Numerical solutions of the euler equations by finite volume methods using Runge-Kutta time-stepping schemes. AIAA Paper 81-1259; 1981.
[20] Yoon S, Jameson A. An LU-SSOR Scheme for the Euler and Navier–Stokes equations. AIAA Paper 87-0600; 1987.
[21] Jouhaud JC, Sagaut P, Labeyrie B. A Kriging approach for CFD/wind tunnel data comparison. ASME J Fluids Eng 2006;128:847–55.
[22] Jouhaud JC, Sagaut P, Montagnac M, Laurenceau J. A surrogate-model based multi-disciplinary shape optimization method with application to a 2D subsonic airfoil. Comput Fluids 2007;36(3):520–9.
[23] Sirovich L. Turbulence and the dynamics of coherent structures. Part 1: coherent structures. Q Appl Math 1987;45(3):561–71.
[24] Golub GH, Van Loan CF. Matrix computations. The Johns Hopkins University Press; 1996.
[25] Rencher AC. Linear models in statistics. Wiley series in probability and statistics; 2000.
[26] Ahlberg H, Nielson E, Walsh J. The theory of splines and their applications. Academic Press; 1967.
[27] Dreyfus G. Neural networks: methodology and applications. Springer-Verlag; 2005.
[28] Jordan MI, Jacobs RA. Hierarchical mixtures of experts and the EM algorithm. Neural Comput 1994;6:181–214.
[29] Krige DG. A statistical approach to some basic mine valuations problems on the witwatersrand. J Chem Metal Min Soc S Afr 1951;52:119–39.
[30] Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. Statistical Science 1989;4(4):409–35.
[31] Laurenceau J, Sagaut P. Building efficient response surfaces of aerodynamic functions with Kriging and cokriging. AIAA J 2008;46(2):498–507.
[32] Geisser S. Predictive inference: an introduction. Chapman and Hall; 1993.
[33] Brand M. Fast low-rank modifications of the thin singular value decomposition. Linear Algebra Appl 2006;415(1):20–30.
[34] Finkel R, Bentley JL. Quad trees: a data structure for retrieval on composite keys. Acta Inform 1974;4(1):1–9.
[35] Melenchon J, Iriondo I, Meler L. Simultaneous and causal appearance learning and tracking. Electron Lett Comput Vis Image Anal 2005;5(3):44–54.
[36] Cook PH, McDonald MA, Firmin MCP. Aerofoil RAE 2822-pressure distributions and boundary layer and wake measurements. AGARD AR 138; 1977.
[37] Michalewicz Z. Genetic algorithms + data structures = evolution programs. Berlin, Heidelberg, New-York: Springer-Verlag; 1992.
[38] Gentle JE. Random number generation and Monte Carlo methods. Statistics and computing. Springer-Verlag; 2003.
[39] Sobol IM, Turchaninov VI, Levitan Yu L, Shukhman BV. Quasi-random sequence generators. Keldysh institute of applied mathematics. Moscow: Russian Academy of Sciences; 1992.
[40] McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 2000;42(1):55–61.
[41] Scheffer T, Wrobel S. Finding the most interesting patterns in a database quickly by using sequential sampling. J Mach Learn Res 2003;3(1):833–62.
[42] Patera AT, Rozza G. Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations. Version 1.0, Copyright MIT 2006–2007, to appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering.