Bayesian inference applied to spatio-temporal reconstruction of flows around a NACA0012 airfoil

Romain Leroux, Ludovic Chatellier, Laurent David

Bayesian inference applied to spatio-temporal reconstruction of flows around a NACA0012 airfoil

Leroux Romain · Ludovic Chatellier · Laurent David

Abstract In this paper, we shall investigate sequential data assimilation techniques to improve the stability of reduced-order models for fluid flows. The reduced-order model used relies on a Galerkin projection of Navier–Stokes equations on proper orthogonal decomposition (POD) basis vectors estimated from snapshots of the flow fields obtained with time-resolved particle image velocimetry (TR-PIV) measurements. The coefficients of the dynamical system are given through a least-squares regression technique applied to the experimental data and lead to a low-order model which is known to diverge, or damp, rapidly in time if left uncontrolled. In this context, a sequential data assimilation method based on a Bayesian approach is proposed. In this formalism, reduced-order models (ROMs) are modeled with discrete time from the hidden Markov processes. Given the whole trajectories of the POD temporal modes, the state of ROM coefficients initially provided by noisy PIV measurements are re-estimated from a Kalman filtering of the sequential data. Results are obtained for the flow around a NACA0012 airfoil at Reynolds numbers of 1000 and 2000 and angles of attack of 10°, 15°, 20° and 30°.

L. Romain (✉)
Université Paul Sabatier, IRIT - Equipe ADRIA,
118 route de Narbonne, 31062 Toulouse Cedex 9, France
e-mail: Romain.Leroux@irit.fr

L. Chatellier · L. David
Institut P', CNRS-Universite de Poitiers-ENSMA,
UPR 3346, Poitiers, France
e-mail: Ludovic.Chatellier@univ-poitiers.fr

L. David
e-mail: Laurent.David@univ-poitiers.fr

L. Chatellier · L. David
Departement Fluides, Thermique, Combustion,
SP2MI-Teleport 2, Boulevard Marie et Pierre Curie,
BP 30179, 86962 Futuroscope Cedex, France

1 Introduction

The flows encountered in environmental or industrial applications are generally non-stationary and coupled with other physical phenomena. The number of degrees of freedom required for a precise description of their dynamics is thus very high. The cost of calculations to model these physical phenomena is prohibitive, particularly those based on systems of partial differential equations, for example, the resolution of the complete Navier–Stokes equations for the numerical simulation of turbulent flows. One possible approach consists in replacing the equations of states by a reduced-order model, thereby shortening the calculation time (Aubry 1991; Deane et al. 1991). The low-order model build this way must allow a faithful approximation of the spatial and temporal evolution of all the physical quantities considered, while inducing a significant reduction in the number of degrees of freedom of the model.

The approach used here is based on Galerkin projection of the physical model on a reduced-dimension basis determined by Proper Orthogonal Decomposition (POD). Thanks to its optimality property, POD allows the definition, via a set of flow solutions originating from a numerical or experimental database, of the best approximation of the flow field database from an energetic perspective, using a limited number of proper functions. A model reduction methodology, called POD-Galerkin, is then used to define a reduced-order flow model by Galerkin projection of Navier–Stokes equations on a POD basis.

The corresponding reduced-order model (ROM) is a system of ordinary differential equations (ODE) of small dimensions; this system governs the evolution of temporal coefficients associated with the POD modes (Galletti et al. 2004; Buffoni et al. 2006). In order to use a reduced-order

model to reconstruct the flow, its accuracy and strength must be guaranteed. Indeed, it has to reproduce as faithfully as possible the complex dynamics contained in the experimental database and at the same time be resistant to variations of the input parameters or to changes in the flow configuration.

Howewer, the POD-Galerkin ROM (or POD ROM) is non-robust and structurally unstable. It determines in general a dynamic system that in many cases may lead to erroneous states after a relative short time of integration (Rempfer 2000; Noack et al. 2003). Unfortunately, these reduced-order models determine in general a dynamical system that in many cases may converge to erroneous states after a relative short time of integration. The origins of the lack of robustness of the POD ROM are evidenced at various levels in Holmes et al. (1996):

- Pressure terms are neglected in the computation of the coefficients of the dynamic system.
- Numerical errors that spread through the model: They may result from the truncation of the basis used to decompose the model variables or to the inaccurate estimation of the time derivatives coefficients of the temporal modes that determine the dynamic system coefficients due to the ill-conditioned least-squares method for the estimation.
- Dissipation loss due to mode truncation.

Methods for the calibration of ODE systems using the POD-Galerkin methodology are therefore necessary in order to guarantee a convergence of these systems own dynamics toward the temporal dynamics of the initial POD modes. The various sources of error mentioned generally combine with each other, which is why the ROM stabilization methods proposed in the literature are based on the resolution of a specific physical phenomenon.

The small number of POD modes selected according to an energy criterion to build the POD ROM implies that the modes with the weakest contribution to the dynamics are neglected. These modes may correspond to small scales and their effects on flow dynamics may be quite significant if they are related for instance to energy dissipation mechanisms. The most commonly used method to control the ROM over time is to include the effect of dissipation of the truncated modes on the ROM through the addition of an artificial viscosity.

Delville et al. (2001), Podvin and Lumley (1998) added a constant viscosity acting in the same way on all the POD ROM. The viscosities finally obtained depend on the nature of the initial data, Aubry et al. (1988) introduced an empirical viscosity, called turbulent viscosity, in order to model the effect of energy exchanges between the selected and truncated POD modes. Cazemier et al. (1998) proposed another definition of this empirical viscosity by adding a conservation constraint of the energy conveyed by the reduced-order model, Karamanos and Karniadakis (2000) used a alternative dissipative model called spectral vanishing viscosity model and Bergmann et al. (2005) consider turbulent viscosities as non-stationary correction coefficients for the constant and linear terms of the ROM which minimize the prediction error of the POD-Galerkin reduced-order model. To make up for the structural instability of the reduced-order model caused by the truncation of the modal basis, the addition to the truncated POD basis of modes that are significant from the viewpoint of the general flow dynamics, but containing little energy, seems to improve the robustness of the reduced-order model (Noack et al. 2003).

These calibration terms are determined *a posteriori* by means of a constrained minimization problem while respecting the formulation of the ODE system resulting from the Galerkin projection. Among the works carried out for this purpose, we can quote Galletti et al. (2005), who used a direct and adjoint model optimization method to determine these coefficients under the nonlinear constraints optimization for the ROM. Couplet et al. (2005) proposed a calibration technique based on the minimization of a linear functional, leading to similar results with a lower numerical cost. Cordier et al. (2010) proposed an improvement of this procedure by using a calibration method based on Tikhonov regularization. Bergmann and Cordier (2008) use trust region methods to control the reduced-order model around a cylinder in laminar flow.

The calibration procedure considered in this paper is based on data assimilation techniques. These techniques serve to reconstruct the state of a dynamic system by combining the information contained in the spatial and temporal evolution equations of the dynamic system under consideration with the physical information contained in the observations of this system over time. These methods are commonly used in fields such as oceanography and meteorology where it is necessary to introduce observation data into the model to take account of spatial and temporal specificities of the physical phenomenon being studied, and thus be able to make an historical analysis of this phenomenon (Cao et al. 2007; Fang et al. 2009; Le Dimet and Talagrand 1986). For that purpose, the use of stochastic methods may be envisaged.

The aim of this article is to investigate sequential data assimilation methods on POD-Galerkin low-order model to reconstruct the flow around a NACA0012 airfoil at Reynolds numbers of 1000 and 2000 and at angles of attack of $10°, 15°, 20°$ and $30°$. For a dynamic system and a series of observations of this system in time, sequential data assimilation is a method which combines the observations with prior knowledge of the current state of the system to obtain updated and improved estimates of the distribution

of true model states or parameters. This work is conducted in a Bayesian statistical framework which includes a forecast model that propagates the systems dynamics forward in time and an observation model that maps the observations to model states and a probability density function, noted pdf, of model and measurements errors. A filtering approach is applied which consists in estimating the system at time k on the basis of observations available up to time k.

The paper is organized as follows. Section 2 briefly recalls the principles of the proper orthogonal decomposition and the standard velocity POD ROM for incompressible flows. The Bayesian approach is introduced, and different calibration methods based on Bayesian filtering of the POD ROM are presented. The linear Kalman filter and the Ensemble Kalman filter are then introduced. The experimental setup used to obtain the flow database is presented in Sect. 3. Numerical results of POD and POD ROM are presented in Sect. 4. Section 5 presents results of the application of Kalman filtering on the linear and quadratic POD ROMs. Finally, Sect. 6 is dedicated to our conclusions.

2 Sequential data assimilation applied to proper orthogonal decomposition

2.1 Reduced-order model based on proper orthogonal decomposition

The proper orthogonal decomposition (POD) is a method used extensively by different authors as a technique to obtain approximate descriptions of large scale or coherent structures in laminar or turbulent flows. This section briefly reviews its main characteristics. The POD method was first introduced in turbulence by Lumley (1967) so as to identify the coherent structures widely known to exist in incompressible turbulent flows. The POD, also known as Karhunen-Loève decomposition, principal component analysis or empirical eigenfunctions method, has been widely used to develop reduced-order flow models so as to approximate the Navier–Stokes equations for the reconstruction of flow fields.

Considering $\Omega \in \mathbb{R}^2$ a physical domain with boundary $\partial\Omega$, we assume that the flow dynamics are governed by the condition of incompressibility and the two-dimensional unsteady Navier–Stokes equations:

$$\begin{cases} \nabla\cdot u = 0 & \text{in} \quad \Omega \times [0,T] \\ \dfrac{\partial u}{\partial t} + \nabla\cdot(u\otimes u) = -\dfrac{1}{\rho}\nabla p + \dfrac{1}{R_e}\Delta u & \text{in} \quad \Omega \times [0,T] \end{cases}$$

where $u(x,t)$ is the velocity vector and $R_e = U_\infty c/\nu$ the Reynold number of the flow, with c the chord length of the NACA0012 profile. The POD consists in looking for the N-orthonormal spatial basis functions $\{\phi_i(x) : \Omega \in \mathbb{R}^2 \longrightarrow \mathbb{R}^2, i = 1,\ldots,N\}$ with temporal coefficients $\{a_i(t) : \mathbb{R}^+ \longrightarrow \mathbb{R}, i = 1,\ldots,N\}$ that is most similar in an average sense to the realizations of the snapshots of the velocity fields $u(x,t_k)$, $k = 1,\ldots,N$. The POD representation is defined for zero mean signals that is why the temporal average is considered here as an additive component that is previously removed from the data. Hence, the expansion of the flow is expressed by separating flow velocity into mean $\bar{u}$ and fluctuating u' parts $u = \bar{u} + \acute{u}$ which leads to the following decomposition:

$$u(x,t) = \bar{u}(x,t) + \sum_{i=1}^{N} a_i(t)\phi_i(x) \tag{1}$$

where $\bar{u}(x,t) = \lim_{T\to\infty}\frac{1}{T}\int u(x,t)\mathrm{d}t$.

The condition of incompressibility is implicity imposed on the basis functions through the data and hence can be dropped. Homogeneous Dirichlet or Neumann boundary conditions are also implicitly satisfied by the basis functions. Generally, the decomposition is taken over the Hilbert space $L^2(\Omega)$. The choice of the inner product becomes a crucial aspect of the decomposition. In the case of incompressible flow, the standard inner product stands:

$$(u,\psi) = \int_\Omega u(x)\psi(x)\mathrm{d}x \tag{2}$$

The associated norm of the scalar product corresponds to kinetic energy of the flow contained in Ω. The POD basis is constructed with a constrained optimization problem (Berkooz et al. 1993) equivalent to the following Fredholm integral eigenvalue problem (Lumley 1967):

$$\int_\Omega R(x,y)\phi(y)\mathrm{d}y = \lambda\phi(x) \tag{3}$$

where $R(x,y)$ is a two-point correlation tensor:

$$R(x,y) = \frac{1}{N}\sum_{i=1}^{N} u(x,t_i)\otimes u(y,t_i) \tag{4}$$

The eigenvalue problem (3) can be formulated as:

$$\sum_{k=1}^{N} C(t_k,t_i)a(t_k) = \lambda a(t_i) \tag{5}$$

where $C = \frac{1}{N}(u(x,t_k),u(x,t_i))$ is a $N\times N$ symmetric and positive definite correlation matrix. It is well known (Lumley 1967) that the solution of the optimization problem (3) is provided by the N eigenfunctions $\{a_1(t),a_2(t),\ldots,a_N(t)\}$ associated with the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ of the matrix C. This method, known as

the snapshot POD method (Sirovitch 1987), is used when the time range is much lower than the velocity field spatial dimension. The spatial basis functions $\phi_i(x)$ can be then calculated from flow realizations $u(x,t_k)$ and the temporal POD modes $a_i(t)$:

$$\phi_i(x) = \frac{1}{N}\sum_{k=0}^{N}\frac{1}{\lambda_i}a_i(t_k)u(x,t_k) \quad \text{for} \quad i = 1,\ldots,N \tag{6}$$

The Galerkin projection of the Navier–Stokes equations onto the subspace spanned by the N vectors of the POD basis formally writes:

$$\left(\frac{\partial u}{\partial t},\phi_i\right) + ((u\cdot\nabla)u,\phi_i) = -\left(\frac{1}{\rho}\nabla p,\phi_i\right) + \left(\frac{1}{R_e}\Delta u,\phi_i\right)$$

The spatial modes are orthonormal, a Galerkin projection of the Navier–Stokes equations, under the hypothesis of incompressible flow, reduces the system of partial differential equations to a system of ordinary differential equations with respect to the temporal coefficients of the decomposition.

Identifying $u(x,t)$ to its modal decomposition (1) leads to a quadratic system of ODE called POD ROM describing the evolution of the temporal coefficients with constant D_i, linear L_{ij} and quadratic C_{ijk} terms:

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}a_i(t) = D_i + \displaystyle\sum_{j=1}^{N} L_{ij}a_j(t) + \sum_{j=1}^{N}\sum_{k=1}^{N} C_{ijk}a_j(t)a_k(t) \\ a_i(0) = (u_{t=0} - \bar{u},\phi_i)_{\Omega} \end{cases} \tag{7}$$

In the following, linear POD ROM refers to the POD ROM defined in (7) without the coefficients C_{ijk} whereas the quadratic POD ROM refers to (7).

The dimension of the system depends on the number of components retained for the analysis. This degree of freedom is usually determined from an energetic criterion called *Relative Information Content* (RIC) (Bergmann and Cordier 2008):

$$\mathrm{RIC}(N_{\mathrm{POD}}) = \sum_{i=1}^{N_{\mathrm{POD}}}\lambda_i \Big/ \sum_{i=1}^{N}\lambda_i \tag{8}$$

The coefficients of the dynamic system are not directly estimated by the Galerkin projection but from the least-squares regression applied to the experimental data using the polynomial identification technique proposed firstly in Perret et al. (2006). As described in Noack et al. (2005), the POD ROM with the quadratic term defined in (7) can efficiently represent incompressible fluid flows. However, this kind of low-order model is known to diverge, or damped, rapidly in time if left uncontrolled. It is thus necessary to stabilize the POD ROM which has led a number of authors to introduce regularization terms in order to stabilize the models (Aubry et al. 1988; Karamanos and Karniadakis 2000; Bergmann et al. 2005). The methods considered in this paper rely on Bayesian methods applied to POD ROM. They include the classical Kalman filter and the Ensemble Kalman filter. These filtering methods of stochastic nature do not introduce any calibration terms and are designed to handle inaccurately specified and truncated dynamic systems.

2.2 Sequential data assimilation

The estimation of the state of a dynamic system using sequential data assimilation can be seen as a probabilistic inference problem, and formulated as the estimation of an *a posteriori* probability distribution of the system state variables at a given time knowing an history of measured data until that time. The state space approach provides a general framework for describing this state estimation and the assimilation estimation can be formulated sequentially by a stochastic filtering problem. This section provides some necessary recalls about the Bayesian formulation of data assimilation and presents two filtering methods, the linear Kalman and Ensemble Kalman filters. More details on sequential data assimilation, Bayesian filtering and Kalman filtering can be found in Le Gland (2009).

2.2.1 State space model

In the following, the state variables at instant k will be represented by a vector x_k of dimension n with initial pdf $p(x_0)$. A sequence of measurements or observations from time 1 to k will be denoted by a set of vectors of dimension n as: $y_{1:k} = \{y_i; i = 1,\ldots,n\}$ where the time between two successive measurements is arbitrarily set to $\Delta k = 1$. We consider the nonlinear discrete-time dynamic system:

$$x_k = f_{k-1}(x_{k-1}) + w_k \tag{9}$$

$$y_k = h_k(x_k) + v_k \tag{10}$$

where the functions f and h represent the system and observation models, respectively. Here, f_k is a nonlinear operator describing the state propagation between two consecutive time steps $k-1$ and k. The true state at time k is assumed to be related to h_k, the observation vector that describes what observation would be measured given the state x_k. The observation y_k is conditionally independent given x_k and the observation is represented by the pdf $p(y_k|x_k)$ which is often named as *likelihood*. In this article, f_k corresponds to the POD ROM and the observations $y_{1:k}$ are the temporal POD modes. We assume that the stochastic processes $(w_k)_{k\geq 0}$ and $(v_k)_{k\geq 1}$ are i.i.d (*independent identically distributed*) additive temporal white Gaussian processes with zero mean and covariance matrices Q_k and R_k respectively.

The state sequence governed by Eq. (9) is a stochastic process which is completely described by its pdf $p(x_k)$

assumed to be Gaussian with covariance matrix P_k. It can be seen as a hidden Markov process according to the conditional pdf $p(x_k|x_{k-1})$, called the *transition distribution* where all the state vectors $x_0, x_1, \ldots, x_k$ are mutually uncorrelated. Under these assumptions, their corresponding pdfs are fully described by their mean and covariance. The presence of errors in both the evolution and the observation steps leads naturally to a probabilistic formulation of the data assimilation problem, where the interest focuses not only on the state x_k but on its probability density function given the observations up to that time $p(x_k|y_{1:k})$ called the *filtering distribution*.

The aim of state estimation in the Bayesian filtering is to estimate the filtering distribution $p(x_k|y_{1:k})$ at each measurement time k. This distribution can be obtained recursively by the Bayesian filtering equations and will constitute the minimum variance estimation between the state variables trajectories at instant k and a history of observations until that time.

2.3 Bayesian filtering

The initialization of the Kalman filter is given by the distribution $p(x_0)$, described by a Gaussian of mean x_0 and covariance P_0. Given the initial *a priori* pdf $p(x_0)$, the transition distribution $p(x_k|x_{k-1})$, and the likelihood $p(y_k|x_k)$, the aim of the filtering is to estimate the optimal state at time k given the observations up to time k. Assuming $p(x_k|y_{1:k-1})$ is known, the filtering distribution $p(x_k|y_{1:k})$ is evaluated in two steps: the correction and update steps.

- The **prediction step** evaluates the predicted filtering distribution $p(x_k|y_{1:k-1})$ from $p(x_{k-1}|y_{1:k-1})$ and the transition distribution $p(x_k|x_{k-1})$:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1} \tag{11}$$

- The **correction step** integrates the new observation y_k through the knowledge of the likelihood $p(y_k|x_k)$. When an observation y_k becomes available at time k, it is possible to obtain this posterior pdf $p(x_k|y_{1:k})$ via the Bayes's rule:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \tag{12}$$

  where $p(y_k|x_k)$ is the observation's likelihood. The normalizing constant $p(y_k|y_{1:k-1})$ is called the *marginal likelihood* and is defined as:

$$p(y_k|y_{1:k-1}) = \int p(y_k|x_k)p(x_k|y_{1:k-1})\mathrm{d}x_k \tag{13}$$

The filtering distribution $p(x_k|y_{1:k})$ of the state x_k given the observations $y_{1:k}$ provides a complete solution of the sequential data assimilation problem, and determines an optimal estimate of the state x_k. However, despite the simple formulation of the solution, it is generally impossible to solve this problem analytically and approximate methods must be used except in the particular case of linear gaussian systems where the corresponding filtering distribution is Gaussian and completely defined from its two firsts moments. The Kalman filter provides an optimal analytic iterative formulation of their expression. For nonlinear systems and non-gaussian distributions, approximations of the optimal Bayesian solution can be derived. The following section recalls the Kalman filter's formulas for linear and nonlinear dynamic system.

2.4 Kalman filtering

2.4.1 Introduction

Kalman filtering, first introduced by Kalman and Bucy (1961), combines observations of the current state of a dynamic system with forecasted results to provide the most accurate estimation, referred to as analysis in the data assimilation community, of the current state of the system. In this section, the basic theory of linear and Ensemble Kalman filters is recalled.

2.4.2 Linear Kalman filter

We will assume that the evolution of state variables is described through a recursive discrete linear time invariant dynamic system model:

$$x_k = Fx_{k-1} + w_k \tag{14}$$

$$y_k = H_k x_k + v_k \tag{15}$$

where F is a $n \times n$ matrix and w_k and v_k are white Gaussian noises with covariance Q_k and R_k. The linearity of the observation operator (10) is also assumed. The initialization of the filter is given by the distribution $p(x_0)$, described by a Gaussian of mean x_0 and covariance P_0. Given the available observations, the Kalman filter computes the filtering distribution $p(x_k|y_{1:k})$ in two steps as described below. The first step is the prediction step where $p(x_k|y_{1:k-1})$ is determined from $p(x_{k-1}|y_{1:k-1})$. The prediction is then corrected with the new available observation y_k. The result of the correction step, also called *analysis*, provides the vector $x_k^a \triangleq \mathbb{E}[x_k|y_{1:k}]$ and the covariance $P_a^k \triangleq \mathbb{E}[(x_k - x_k^a)(x_k - x_k^a)^T|y_{1:k}]$ characterizing the Gaussian filtering distribution at time k. These moments depend on the mean x_k^a and covariance P_k^a. The state space model is considered here as linear and gaussian, hence the transition distribution and the likelihood are also gaussians. So for all $k = 1, \ldots, N$ if $p(x_{k-1}|y_{1:k-1})$ is a gaussian density with mean m_{k-1}^a and covariance P_{k-1}^a we have the following quantities:

$$p(x_k|x_{k-1}) = \mathcal{N}(m_k^f, P_k^f) \tag{16}$$

$$p(y_k|x_k) = \mathcal{N}(m_k^a, P_k^a) \tag{17}$$

with $m_k^f, P_k^f, m_k^a, P_k^a$ defined as follows:

$$m_k^f = F_k m_{k-1}^a \tag{18}$$

$$P_k^f = F_k P_{k-1}^a F_k^T + Q_k \tag{19}$$

$$m_k^a = m_k^f + K_k(y_k - H_k m_k^f) \tag{20}$$

$$P_k^a = (I - K_k H_k) P_k^f \tag{21}$$

where $K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}$ is the so-called Kalman gain (Throughout this paper we use superscripts "f" for *forecasted* and "a" for *analyzed*). A demonstration of these properties can be found in Le Gland (2009). In practice, the process noise and measurement noise covariance matrices might change with each time step or measurement; however, here, we assume they are constant. The covariance matrices in the Kalman filter provide a measure for uncertainty in the predictions and updated state estimate.

In the case of a linear and Gaussian system, the discrete-time Kalman filter is the exact solution of the problem of optimal filtering because it is equivalent to the BLUE (*Best Linear Unbiased Estimator*) (Le Gland 2009) and produces the optimal minimum mean-square error of the underlying system state as described in Kalman and Bucy (1961). However, the computation of covariance matrices is the main issue. Even if the state and measurement equations are linear with additive Gaussian white noise, for sufficiently large-scale problems, such as those which arise in weather forecasting and oceanography, computing and storing the error covariance matrices involved in the Kalman filter are practically impossible and hence, approximations must be made. Besides, for nonlinear systems, the solution exists but it is impossible to determine it analytically.

Several advanced filter algorithms based on the Kalman filters have been developed for data assimilation with large-scale nonlinear models. The Ensemble Kalman filter (EnKF) first proposed by Evensen (1994), including several of its variants, is one of the most used filter algorithms. The Singular Evolutive Extended Kalman (SEEK) filter (Pham et al. 1998) and the Singular Evolutive Interpolated Kalman (SEIK) filter (Pham 2001) provide alternative approaches.

2.4.3 Ensemble Kalman filter

This section presents the ensemble extension of the Kalman filter for systems described by nonlinear dynamics (9) and a linear measurement model (10).

The EnKF is a Monte Carlo, derivative-free, alternative to the extended Kalman filter (Gelb 1974), to approximate the Kalman filter for the linear gaussian state space model. It was developed for the nonlinear gaussian models with great performance compared with the other methods. The Ensemble Kalman filter relies on an ensemble of samples to describe the different probability distributions. The principle of the EnKF is to evaluate the empirical covariance matrix of an ensemble of elements instead of the exact covariance matrix by matrix products. The covariance P_k^f is then replaced by the empirical covariance noted P_k^N and the Kalman gain by the empirical Kalman gain noted K_k^N where N is the number of ensemble elements used in the EnKF.

From initial conditions, the initial pdf is sampled by N members $x_0^{a,i}$ normally distributed. Forecast distribution and filtering distribution are, respectively, approximated through a prediction step and a correction step of the ensemble members. The prediction step consists in propagating the ensembles $x_{k-1}^{a,i}$ through the nonlinear dynamics in order to obtain the forecast ensemble, denoted by $x_k^{f,i}$ (Burgers et al. 1998; Evensen 2003, 2006; Bishop et al. 2001). From the corresponding prediction of these ensembles, the empirical covariance matrix P_k^N is calculated which leads to the computation of the Kalman gain K_k^N of the EnKF.

The EnKF assumes the Gaussian approximation between assimilation times, in order to apply Kalman's formulae. The empirical mean m_k^N of the forecast ensemble is firstly defined by:

$$m_k^N \approx \frac{1}{N}\sum_{i=1}^{N} x_k^{f,i} \tag{22}$$

The empirical ensemble covariance matrix P_k^N is then deduced from the following expression:

$$P_k^N \approx \frac{1}{N-1}\sum_{i=1}^{N}\left[\left(x_k^{f,i} - m_k^N\right)\left(x_k^{f,i} - m_k^N\right)^T\right] \tag{23}$$

Similarly to the Gaussian case presented in the previous section, the empirical Kalman gain K_k^N can be computed from the ensemble variance. We have:

$$K_k^N\left(P_k^N\right) = P_k^N H_k^T \left(H_k P_k^N H_k^T + R_k^N\right)^{-1} \tag{24}$$

In the corrective step, the forecast ensemble members are moved toward the new observation and updated according to the Kalman filter scheme and replaces the covariance matrix by the sample covariance computed from the ensemble:

$$x_k^{a,i} = x_k^{f,i} + K_k^N\left[y_k - \left(H_k x_k^{f,i} + v_k^i\right)\right] \tag{25}$$

The corrected ensemble are then propagated in the state space model at step $k+1$. The algorithm of the EnKF is presented in Fig. 1.

The Ensemble Kalman filter is a gaussian Kalman filter. Errors are statistically represented by an ensemble of points directly propagated by the state equation

Fig. 1 Ensemble Kalman filter algorithm

Initialization of the EnKF : For $i = 1, ..., N$ Generate $x_0^{a,i} \sim \mathcal{N}(0, P_0)$

(i) **Prediction step** : For $i = 1, ..., N$

$$w_k^i \sim \mathcal{N}(0, Q_k)$$
$$x_k^{f,i} = f_k(x_{k-1}^{a,i}) + w_k^i$$
$$m_k^N \approx \frac{1}{N} \sum_{i=1}^N x_k^{f,i}$$
$$P_k^N \approx \frac{1}{N-1} \sum_{i=1}^N [(x_k^{f,i} - m_k^N)(x_k^{f,i} - m_k^N)^T]$$

(ii) **Correction step** : For $i = 1, ..., N$

$$\text{Generate } x_0^{a,i} \sim \mathcal{N}(0, R_k)$$
$$R_k^N = \frac{1}{N-1} \sum_{i=1}^N (v_k^i)(v_k^i)^T$$
$$K_k^N(P_k^N) = P_k^N H_k^T (H_k P_k^N H_k^T + R_k^N)^{-1}$$
$$x_k^{a,i} = x_k^{f,i} + K_k[y_k - (H_k x_k^{f,i} + v_k^i)]$$

without any linearization. The analysis step is the same as the standard Kalman filter. One of the most attractive aspect of the EnKF is that the ensemble size required is usually much smaller than the dimension of model state. So it can be applied to solve high-dimensional problems which cannot be tackled using traditional Kalman filter. The EnKF does not require the linearity of the state equation but the linearity of the observation operator H_k. However, as shown by Evensen (2003), it is possible to use the EnKF for a system in which the observation operator is nonlinear, without using linearization. For the EnKF, adding stochastic perturbations to the observations is essential for each ensemble member. Indeed, as statistic samples become less significant by point coalescence, adding noise to the observations can be interpreted as adding a stochastic term which leads the sample to become more significant (McKean 1969).

Potential loss of rank may occur with the EnKF where random measurement perturbations are used to represent the measurement error covariance matrix. The measurement perturbations introduce sampling errors which can be fully eliminated by a proper sampling of measurement perturbations or avoiding the perturbations as such. The Ensemble Square-Root filter (EnSRF) avoids the measurement's perturbations and thus has a lower analysis error by reducing this additional source of sampling error, without imposing any additional approximations, such as the assumption of uncorrelated measurement errors (Evensen 2004).

2.4.4 Ensemble square-root Kalman filter

This section is devoted to the Ensemble Square-Root Kalman filter (EnSRF) that was introduced by Whitaker and Hamil (2002). The Ensemble Square-Root Kalman filter (EnSRF) modifies the correction equation in such a way as to eliminate the necessity to perturb the observations. The EnSRF algorithm is used to update the ensemble perturbations and is derived starting from the traditional analysis equation for the covariance update in the Kalman filter. Kalman filter of square-root formulation has been used since in meteorology by Bishop et al. (2001) and Whitaker and Hamil (2002), following the initial work of Andrews (1968). The general solution was developed by Andrews (1968) and used in various forms by Bishop et al. (2001), Whitaker and Hamil (2002), Anderson (2001) and Tippett et al. (2003). The Ensemble Square-Root Kalman filter uses the Kalman filter equations for the estimation of the analysis and correction error covariance matrices.

Let X_k be the matrix holding the ensemble members at instant k. The equations of the evolution of the prediction covariance P_k^f and correction error P_k^a of the Kalman filter are:

$$P_k^f = F_k P_{k-1}^a F_k^T + Q_k \tag{26}$$

$$P_k^a = (I - K_k H_k) P_k^f \tag{27}$$

The matrices P_k^f and P_k^a are symmetrical and positive definite, so they can be written as $P_k^f = (X_k^f)(X_k^f)^T$ and $P_k^a = (X_k^a)(X_k^a)^T$. The Ensemble Square-Root Kalman filter replaces the equations of the evolution of the matrices P_k^f and P_k^a by the square root matrices X_k^f ans X_k^a so as to avoid the inverse computation of the matrices P_k^f and P_k^a. From the rewriting of the matrix P_k^a one may obtain:

$$P_k^a = P_k^f - P_k^f H_k^f \left(H_k P_k^f H_k^T + R_k\right)^{-1} H_k P_k^f$$

$$P_k^a = X_k^f \left[I - \left(H_k X_k^f\right)^T \left(H_k X_k^f \left(H_k X_k^f\right)^T + R_k\right)^{-1} H_k X_k^f\right] \left(X_k^f\right)^T$$

$$P_k^a = X_k^f \left[I - V_k^T \left(V_k V_k^T + R_k\right)^{-1} V_k\right] \left(X_k^f\right)^T$$

with $V_k = H_k X_k^f$ and $D_k = V_k V_k^T + R_k$. When observation errors are decorrelated, i.e., the matrix R is diagonal, the observation can be sequentially assimilated one other after the other. In this case, V_k is a column vector and D_k is a scalar. The matrix square root of $I - V_k D_k^{-1} V_k$ can be then calculated by resolving the following equation for the scalar β_k:

$$I - V_k D_k^{-1} V_k = \left(I - \beta_k V_k V_k^T\right)\left(I - \beta_k V_k V_k^T\right)^T \tag{28}$$

which gives the following solution $\beta_k = (D_k + \sqrt{R_k D_k})^{-1}$. The correction of the ensemble is then provided by:

$$X_k^a = X_k^f \left(I - \beta_k P_k^f\right) \tag{29}$$

The linear Kalman filter, the Ensemble Kalman filter and the Ensemble Square-Root Kalman filter are applied on the linear and quadratic POD ROMs so as to improve their abilities to accurately reproduce the velocity fluids around the NACA0012 airfoil from experimental time-resolved data PIV.

3 Description of the experimental setup

The experimental configuration settled for this work consists of the flow around a NACA0012 airfoil of chord $c =$ 60 mm at Reynolds numbers R_e of 1000 and 2000 and at angles of attack of $10°, 15°, 20°$ and $30°$ in a square (160 mm × 160 mm) section water tunnel. Time-resolved 2D-2C PIV measurements were carried out with Nd-YAG laser (Quantel with nominal energy by pulse of 2 × 120 mJ), a pulnix Dual tap Accupixel camera (2,048 × 2,048 px image size), using polyamide seeding particles of 15 µm mean diameter.

The PIV records were analyzed through a cross-correlation technique implemented with a Fast-Fourier-Transform algorithm in a multi-grid process with 3 iterations (1 at 64 × 64 and 2 at 32 × 32) with 75 and 50 % overlapping, respectively, using window shifting with iterative deformations and Gaussian sub-pixel peak localization. The instantaneous vector fields in each snapshot were validated using objective statistical methods (standard deviation and local magnitude difference comparisons) to remove erroneous velocity vectors. The velocity vectors with a signal-to-noise ratio inferior to 1.5 and those with a deviation from the median between 1.5 and 2.5 times the RMS are removed. These erroneous vectors are replaced with the estimated values obtained with the secondary correlation peak. The full series of experiments comprised 2,048 samples of velocity fields obtained at a sampling frequencies $f^{(PIV)} = 6.4$ Hz for the flows at $R_e = 1000$ and $f^{(PIV)} = 12.8$ Hz for the flows at $R_e = 2000$, which correspond, respectively, to a flow velocities of 16.7 and 33.33 mm s^{-1} (Fig. 2).

Figures 3 and 4 illustrate of the spanwise vorticity Ω_z (a dimensioned with U_0) for PIV in the plane $z = 0$ for the case of 15° and 20° at $R_e = 1000$. The purpose here is to work with flows having different dynamics according to a low number of parameters, that is why these two flows were selected as a reference state. The Reynolds numbers based on the chord which are concerned here are low, which makes it possible to suppose that the concerned physical mechanisms remain two-dimensional. Turbulence is absent with such Reynolds numbers so the flow is considered as a swirling flow. Different kinds of vortex shedding of a NACA0012 profile according to the angle of attack and Reynolds number were determined by Huang et al. (2001). Several vortex shedding regimes according to the angle of attack can be distinguished. For each one of these regimes, the identification of the vortex shedding is based on the classification proposed by Huang et al. (2001). For $\alpha = 15°$, the vortex shedding is of type separation vortex, and for $\alpha = 20°$, the vortex shedding is of type leading edge vortex. The sequence $\alpha = 20°$ and $R_e = 1000$ covers 23 vortex shedding cycles with a Strouhal number $S_t = 0.265$ and the $\alpha = 15°$ sequence at $R_e = 1000$ depicting 30 vortex shedding cycles with a Strouhal number $S_t = 0.345$. Sampling rate enables us to obtain approximately 89 snapshots of velocity fields for the case $\alpha = 20°$ at $R_e = 1000$ and 68 snapshots for the case $\alpha = 15°$ at $R_e = 1000$, by vortex shedding period $T_{S_t} = \dfrac{1}{S_t}\dfrac{c}{U_0}$ where S_t is the Strouhal number of the vortex shedding cycles. Figure 3 represents snapshots of the flow fields estimated from PIV for non-dimensional times $t^\star = \dfrac{t}{T_{S_t}}$.

4 Characterization of the linear and quadratic POD ROMs

The snapshot POD method was performed using the sequences of PIV measurements after substraction of their time-averages. The RIC of the N_{POD} first projection modes obtained from the series of 2,048 snapshots of the studied configurations is given in Table 1.

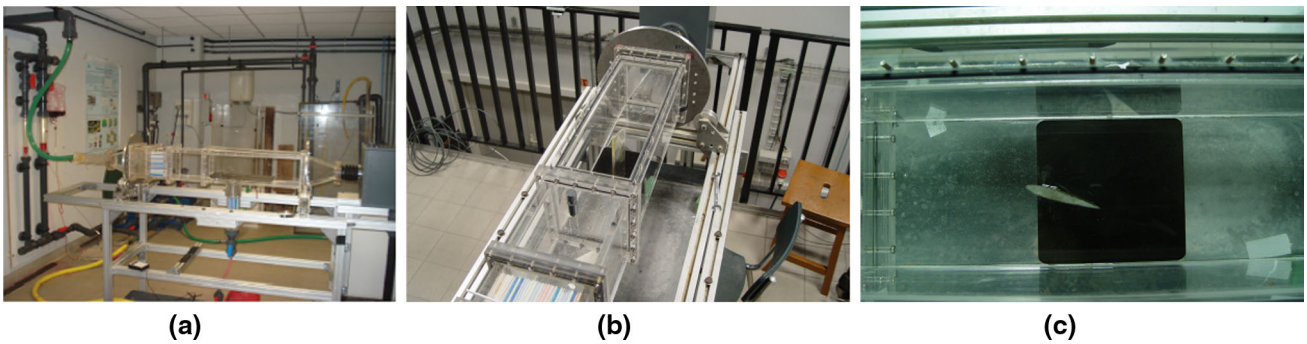

Fig. 2 Experimental setup: **a** square-section, **b** test section and **c** NACA0012 profile at angle of attack of 20°

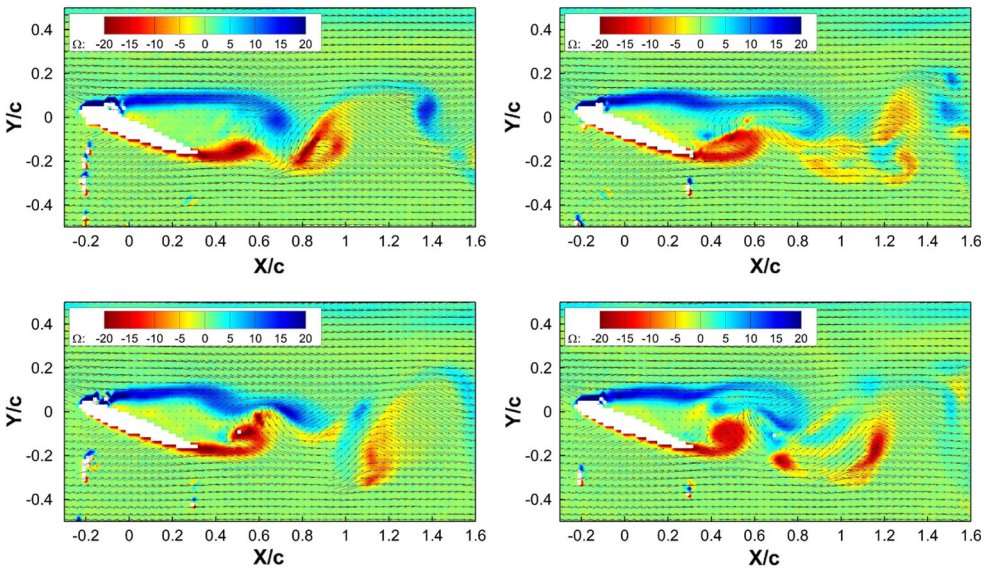

Fig. 3 Snapshots of the vorticity field Ω_z estimated from PIV–$t^\star = 5.75$ and $t^\star = 11.5$, $t^\star = 17.25$ and $t^\star = 23 - \alpha = 20°$ and $R_e = 1000$

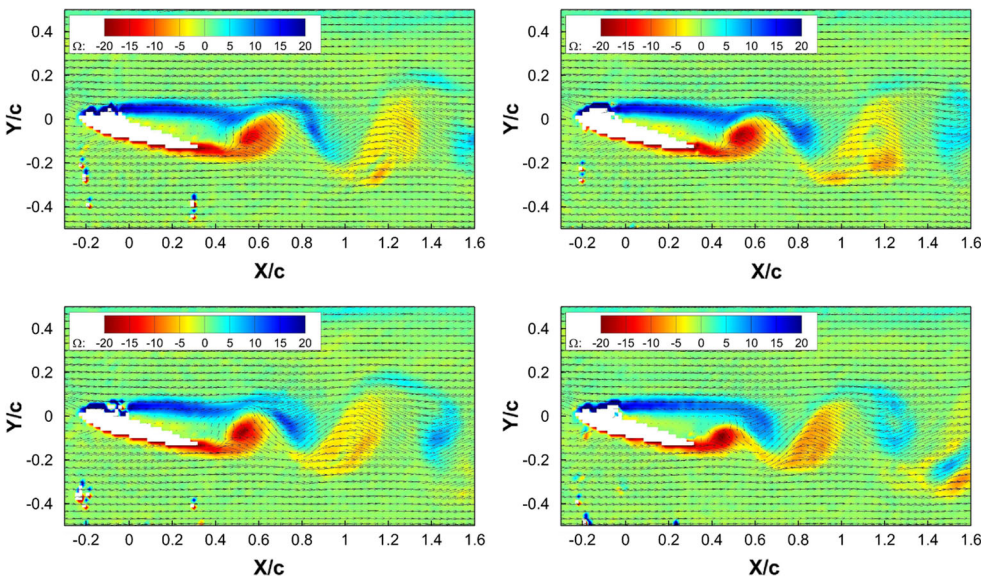

Fig. 4 Snapshots of the vorticity field Ω_z estimated from PIV–$t^\star = 7.48$ and $t^\star = 14.97, t^\star = 22.46$ and $t^\star = 29.94 - \alpha = 15°$ and $R_e = 1000$

For the configurations $\alpha = 20°$ $R_e = 1000$ and $\alpha = 15°$ $R_e = 1000$, the evolution of the RIC of the N_{POD} first POD modes is represented, respectively, in Figs. 5 and 6. For the configuration $\alpha = 15°$ and $R_e = 1000$, the two most energetic modes represent more than 80 % of the total turbulent kinetic energy of the flow, whereas the first 28 modes

concentrate 90 %. These coupled modes correspond to the structure of the von Karman vortices which strongly dominate the flow.

For the studied configurations, the RIC converges slowly to 100 %. Indeed, for most cases, a high number of projection modes are necessary to reconstruct the flow with a satisfactory energetic content. For example, to reconstruct 90 % of the flow, 599 temporal modes are necessary in the case $\alpha = 30°$ and $R_e = 2000$, and 207 in the case $\alpha = 15°$ and $R_e = 2000$. So as to illustrate the velocity field reconstruction from the POD, Figs. 7 and 8 depict the spanwise vorticity Ω_z corresponding to a RIC of 90 % in both cases.

Table 1 RIC in function of the number of projected coefficients N_{POD} for each configuration

	$RIC(N_{POD})$	50 %	60 %	70 %	80 %	90 %
$\alpha = 30° R_e = 2000$	N_{POD}	30	57	114	255	599
$\alpha = 30° R_e = 1000$	N_{POD}	7	12	25	58	181
$\alpha = 20° R_e = 2000$	N_{POD}	14	23	39	69	149
$\alpha = 20° R_e = 1000$	N_{POD}	2	4	10	21	59
$\alpha = 15° R_e = 2000$	N_{POD}	22	35	57	99	207
$\alpha = 15° R_e = 1000$	N_{POD}	*	*	*	2	28
$\alpha = 10° R_e = 2000$	N_{POD}	15	23	37	63	126
$\alpha = 10° R_e = 1000$	N_{POD}	12	25	45	76	139

* Indicate RIC reached with 2 POD modes

Linear and quadratic reduced-order models based on (7) are tested on the flows around the NACA0012 airfoil at Reynolds numbers R_e of 1000 and 2000 and angles of attack of $10°, 15°, 20°$ and $30°$. The snapshot POD method introduced by Sirovich (1987) is first performed using the full sequences of the PIV measurements and used to built the POD basis $\phi_i(x)_{i=1,\ldots,N}$. The 2,048 snapshots uniformly distributed over time are used to compute the discrete form of the temporal correlation matrix C and the POD basis functions are obtained via a projection of the temporal tensor eigenvectors on the whole set of snapshots. Using the most energetic modes and the snapshot mean, the POD-Galerkin model is defined with constant model coefficients, according to Eq. (7). The initial condition of the dynamic system is given by the projection of the first snapshot on the POD basis. The coefficients D_i, L_{ij}, C_{ijk} empirically defining the linear and quadratic POD ROMs are estimated with a polynomial identification. The models are integrated with a constant time

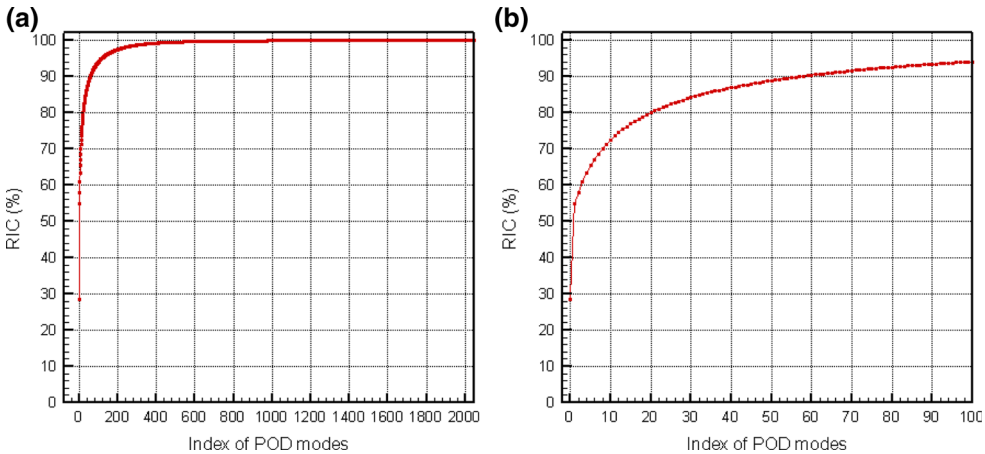

Fig. 5 **a** RIC of the fluctuating modes–$\alpha = 20°$ and $R_e = 1000$ **b** RIC of the fluctuating modes—$\alpha = 20°$ and $R_e = 1000$—zoomed

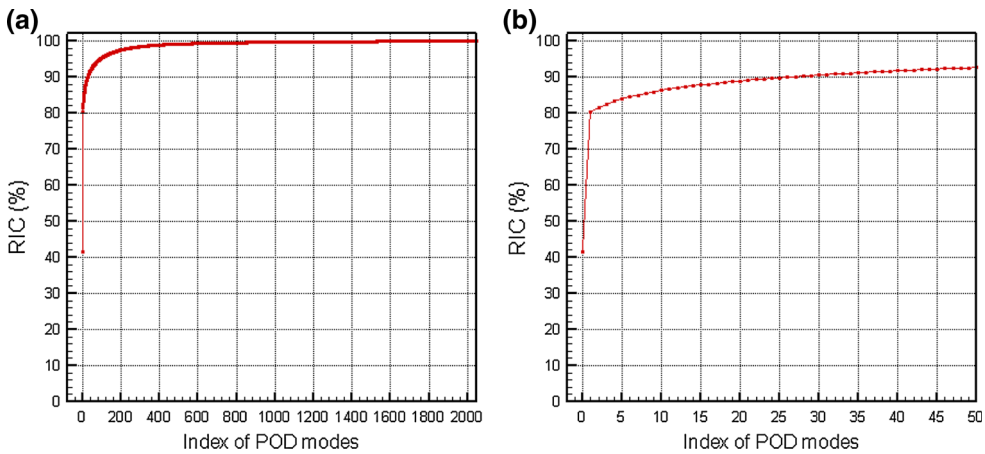

Fig. 6 **a** RIC of the fluctuating modes–$\alpha = 15°$ and $R_e = 1000$, **b** RIC of the fluctuating modes—$\alpha = 15°$ and $R_e = 1000$—zoomed

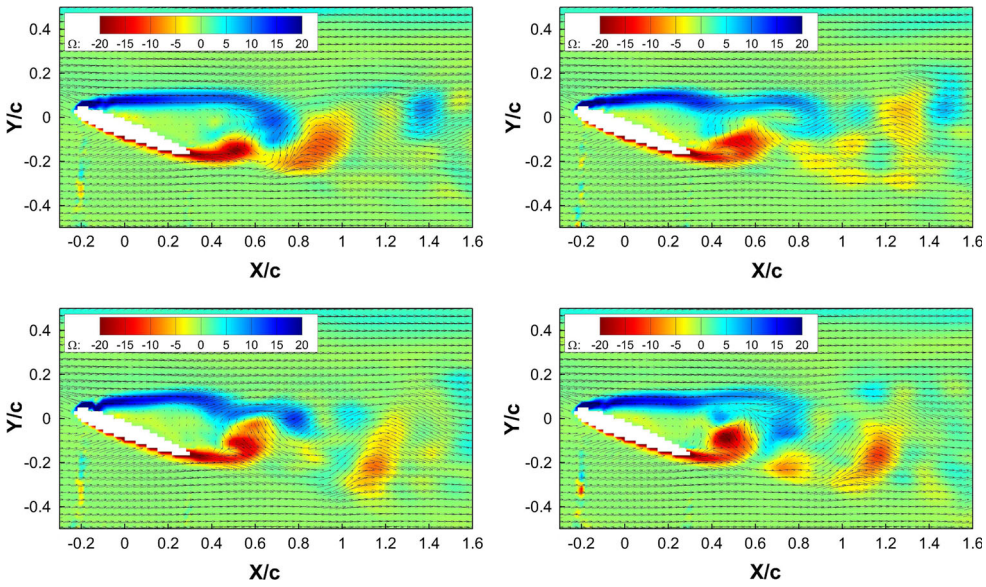

Fig. 7 Snapshots of the vorticity Ω_z field estimated from POD with RIC = 90 %—$t^\star$ = 5.75 and $t^\star$ = 11.5, $t^\star$ = 17.25 and $t^\star$ = 23—$\alpha = 20°$ and $R_e = 1000$

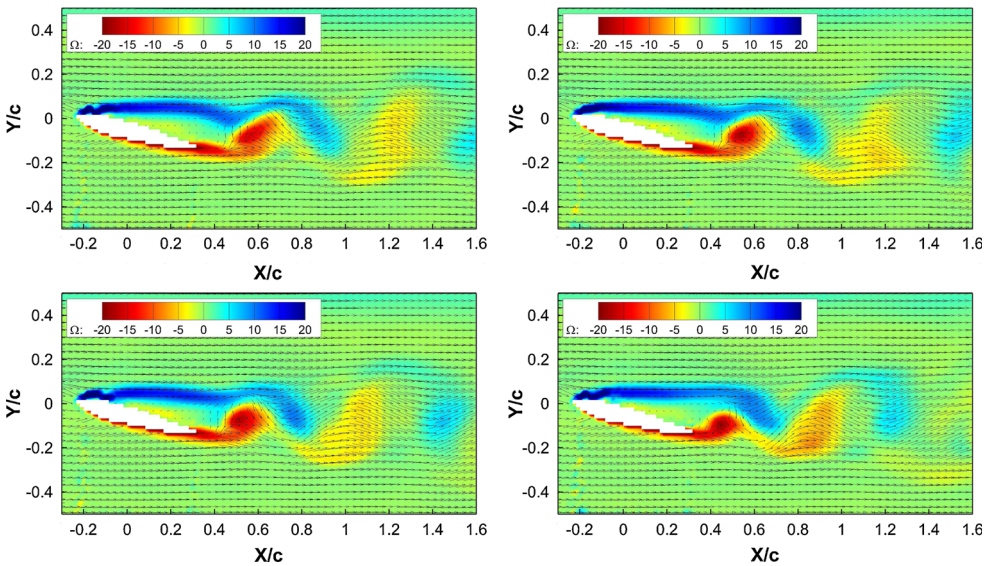

Fig. 8 Snapshots of the vorticity Ω_z field estimated from POD with RIC = 90 %—$t^\star$ = 7.48 and $t^\star$ = 14.97, $t^\star$ = 22.46 and $t^\star$ = 29.94—$\alpha = 15°$ and $R_e = 1000$

step explicit Runge-Kutta method with a timestep $\Delta t = 0.001 \times \Delta t^{(piv)}$, where $\Delta t^{(piv)} = 1/f^{(piv)}$, over the time interval in which the snapshots are taken. Due to the high number of POD modes necessary to obtain a high RIC, only the linear POD ROM will be used at RIC = 90 %. For the quadratic ROM, only the 10 first temporal modes will be used, which correspond to a RIC from 50 to 70 %. The different evolutions of the linear and quadratic POD ROMs are summarized in Table 2. The corresponding behaviors of the POD ROMs are plotted on Figs. 9 and 10 where the original data have also been plotted for comparison purpose. It can be observed that the short-term behaviors of the linear and quadratic POD ROMs are not accurate. In all the configurations tested, errors in amplitude and phase appear quickly at the first time steps and are followed, according to the configurations, by divergence or damping of the amplitude. Both the linear and quadratic models can either diverge or remain stationary, so that none of them can properly reproduce the flow dynamics.

As was already mentioned in Sect. 2, system (7) is clearly unstable when integrated in time. That has led many authors to introduce regularization terms in order to stabilize the model, via constrained optimization methods. We propose, in the following section, to apply linear and nonlinear bayesian inference methods based on sequential data assimilation so as to to stabilize the temporal evolution of the linear and quadratic ROMs without the introduction of corrective terms.

Table 2 Temporal evolution of the linear and quadratic POD ROMs for the different configurations

α°	30	30	30	30
R_e	2000	1000	2000	1000
ROM	Linear	Linear	Quadratic	Quadratic
Evolution	Damping	Damping	Damping	Stationnary
α°	20	20	20	20
R_e	2000	1000	2000	1000
ROM	Linear	Linear	Quadratic	Quadratic
Evolution	Divergence	Divergence	Divergence	Divergence
α°	15	15	15	15
R_e	2000	1000	2000	1000
ROM	Linear	Linear	Quadratic	Quadratic
Evolution	Damping	Damping	Damping	Stationnary
α°	10	10	10	10
R_e	2000	1000	2000	1000
ROM	Linear	Linear	Quadratic	Quadratic
Evolution	Divergence	Damping	Divergence	Stationnary

5 Stabilization of the linear and quadratic POD ROMs by Kalman filtering

This section is devoted to the application of Bayesian inference techniques on the POD ROM so as to improve the estimation of predicted temporal modes with linear and quadratic POD ROMs from discrete sequences of POD temporal modes provided by the snapshot POD method.

Following the notation used in Sect. 2.2 the vector x_k is the n dimensional state space vector to estimated with the function f_k which represents the linear or the quadratic POD ROM. The vector x_k^f contains the predicted POD temporal modes and the vector x_k^a the temporal modes corrected with the Kalman filter. The vector y_k contains the POD temporal modes. The white gaussian noises w_k and v_k of the state and observation equations of the Kalman filter are initialized as:

- $Q = 0.001\ I$ and $R = 0.001\ I$, where I is the identity matrix of size N_{POD} for the linear Kalman filter.
- $w \sim \mathcal{N}(0, 0.01)$ and $v \sim \mathcal{N}(0, 0.01)$ for the EnKF.
- $w \sim \mathcal{N}(0, 0.01)$ for the EnSRF.

In order to initialize the EnKF, the use of $N = 100$ is usually retained, taking into account the fact that moreover it presents one certain margin compared with the $N = 20$ number of ensemble which guarantee the positivity of covariance (Houtekamer et al. 2005). However, to ensure a good convergence of the EnKF, the overall number N used for the EnKF filter will vary according to the number of selected POD coefficients. This is why one choose an overall number of ensemble members N such as if $N_{\mathrm{POD}} < 100$, then $N = 100$ and $100 < N_{\mathrm{POD}} < 200$, then $N = 250$ and $N_{\mathrm{POD}} > 200$, then $N = 500$.

Figures 11 and 13 present the reconstructed temporal modes $a_1(t), a_2(t)$ and $a_5(t), a_{10}(t)$ after the assimilation of the observed POD temporal modes for $\alpha = 20^\circ$ and $R_e = 1000$. Flow reconstruction at different instants and vector fields associated for the case $\alpha = 15^\circ$ and $\alpha = 20^\circ$ to $R_e = 1000$ are presented on the Figs. 18, 19 and 20. These flow reconstruction are obtained by using the EnKF on the linear POD ROM with a RIC of 90 %, with 28 projection modes for the case $\alpha = 15^\circ$ and 59 projection modes for the case $\alpha = 20^\circ$ $R_e = 1000$ (Fig. 12).

Figure 13 shows the effects of Kalman filtering on POD ROM instability. The POD ROM does not diverge much along short time steps, which allows the Kalman filtering to operate efficiently on the POD ROM where the predicted temporal modes are close to their true values. However, the filtering is done using a complete set of observations with low level of noise leading to very low error levels as shown in Fig. 14. Kalman filters used this way are not a straightforward solution for the global problem of the prediction of flow fields using POD ROMS, but they have sufficiently interesting stabilizing properties to be used in this context.

Indeed, as can be observed, the Ensemble Kalman filter enables for the linear reduced-order model to faithfully recover with a good accuracy the trajectories corresponding to the POD modes. The various criteria quantifying the efficiency of the Kalman filters used on the linear and quadratic POD ROMs highlight behaviors which differ according to the configurations studied. The application of the linear Kalman filter to the POD ROM implies an increase in the instantaneous error E in the L^2 norm of the prediction modes according to the number of projection modes in all cases, except the case $\alpha = 30^\circ$ $R_e = 2000$. Indeed, in this case, where 255 projection modes are used, the reconstruction error remains homogenous and in the order of 0.1, which here translates into a divergence of the linear Kalman filter. In other configurations, the maximum reconstruction error committed by the linear Kalman filter is in the order of 8×10^{-3}. Depending on the configurations studied, the reconstruction error of the linear Kalman filter is multiplied by a factor 3, 4 or 6 according to the number of projection modes used (Fig. 14; Table 3). For example, for case $\alpha = 10^\circ$ $R_e = 1000$, the reconstruction error over the first coefficients is 0.001, and 0.006 over the last ones. The efficiency of the linear Kalman filter thus depends on the number of coefficients used on the POD ROM. Despite the increase in the RMSE with the number of coefficients, its values remain low in most configurations

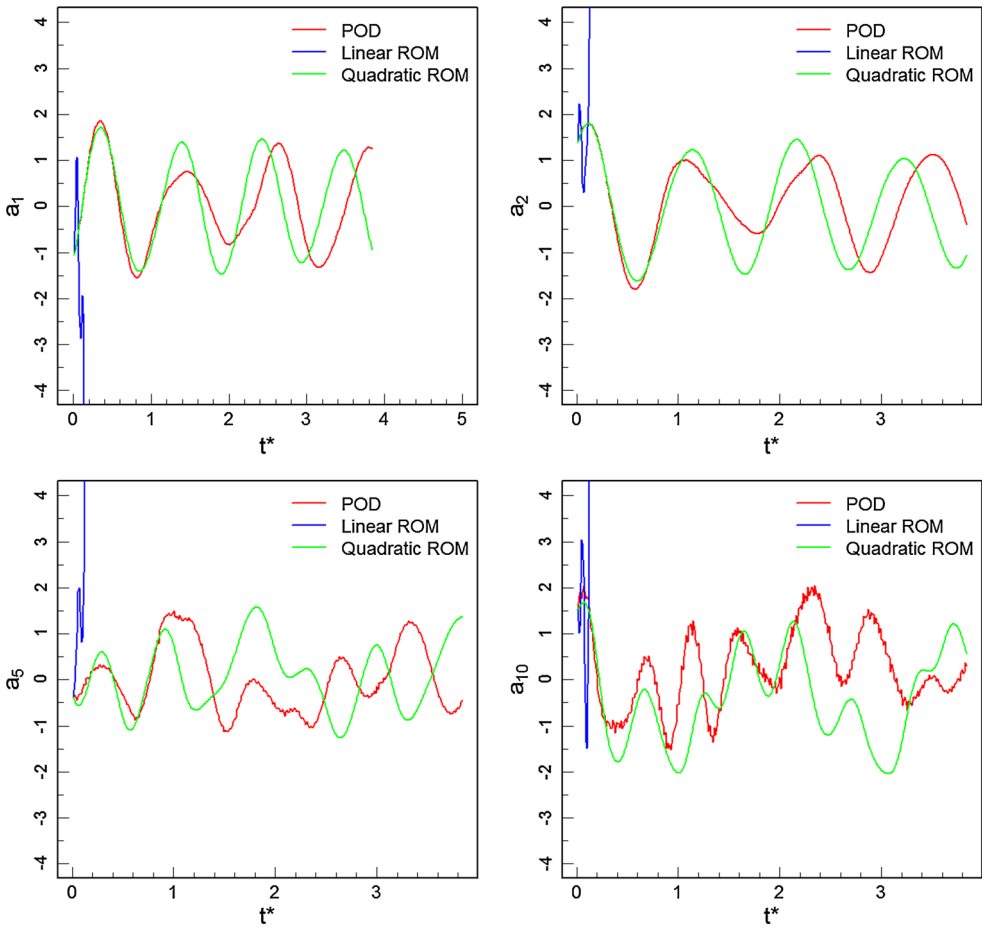

Fig. 9 Predicted coefficients $a_1(t), a_2(t), a_5(t), a_{10}(t)$—$\alpha = 20^\circ$ and $R_e = 1000$

tested provided the number of coefficients is not too high, i.e., lower than 200.

For the EnKF and EnSRF filters applied to the POD-Galerkin linear ROM, the reconstruction error remains homogenous over time. An increase in the reconstruction error with the number of projection modes is observed (Fig. 14) with the increase in the frequency of temporal modes. However, this increase is slight, in the order of 1.5 between the first and last coefficients estimated (Fig. 14). The reconstruction error is small in all configurations tested, and in the order of 10^{-4}. The reconstruction error committed by the EnKF and EnSRF filters is systematically lower than that committed by the linear Kalman filter. For example, for case $\alpha = 20^\circ R_e = 2000$, the maximum reconstruction error of the EnKF filter is 0.00015, that of the EnSRF filter 0.0006 and that of the linear Kalman filter increases from 0.005 to 0.006. The difference in quality between these three Bayesian estimators can also be found in their RMSE (Figs. 15, 16). In each case, the RMSE of the EnKF and EnSRF filters remains in the order of 10^{-4} and 10^{-3} while that of the linear Kalman filter is in the order of 10^{-2} and 10^{-3} (Table 4). The RMSE of the linear Kalman filter is always higher than that of the EnKF and EnSRF filters. Use of the EnKF filter on the quadratic POD ROM results in a reconstruction error in the order of 10^{-4} between the coefficients (Table 5). The RMSE is also low, in the order of 10^{-3}. Unlike the linear case, the small dimension of the state vector does not produce a trend in terms of the evolution of the reconstruction error according to the number of coefficients used (Fig. 17).

The EnKF and EnSRF filters do not diverge over time and the quality of reconstruction of the prediction modes allows an estimation of the prediction modes without a loss of phase or amplitude relation (Figs. 11, 12, 13). The same goes for the linear Kalman filter when it does not diverge. These Bayes estimators allow calibration of the linear POD ROM. Application of the EnKF filter to the quadratic POD ROM also allows efficient calibration of the ROM. The main structures and vector fields of the flow are also reconstructed (Figs. 18, 19, 20).

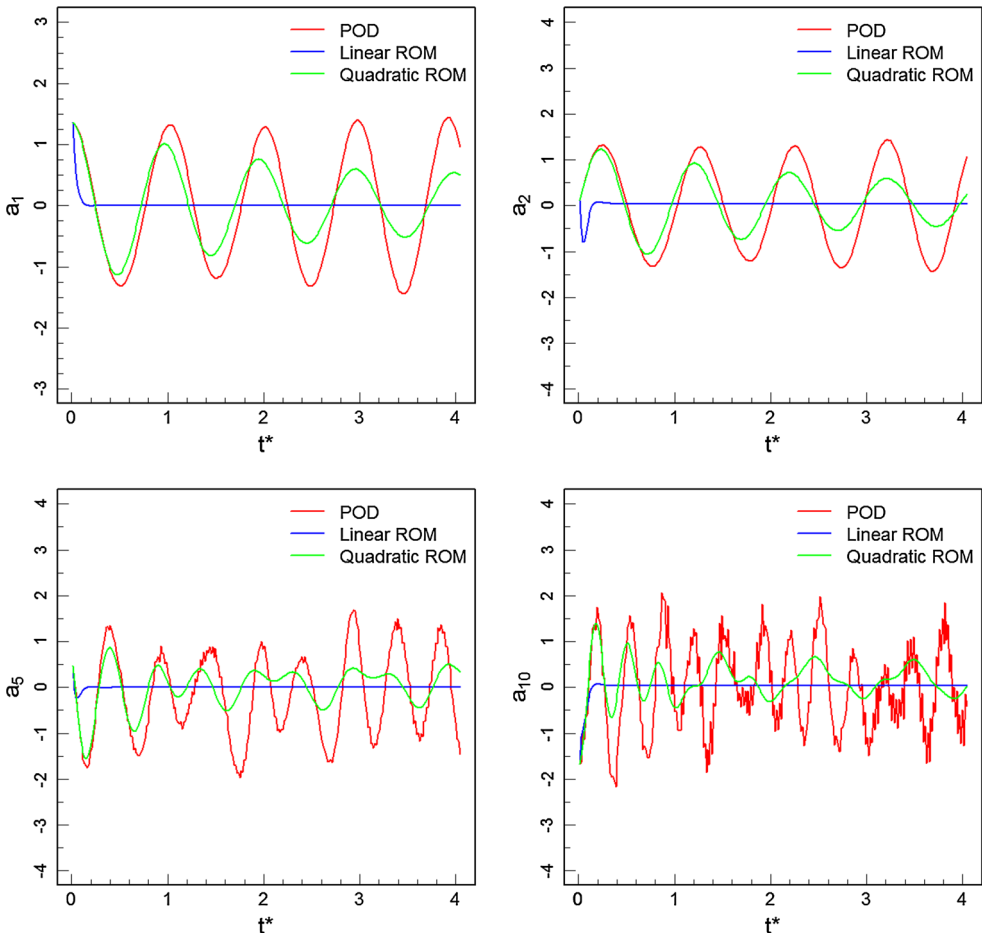

Fig. 10 Predicted coefficients $a_1(t), a_2(t), a_5(t), a_{10}(t)$—$\alpha = 15^\circ$ and $R_e = 1000$

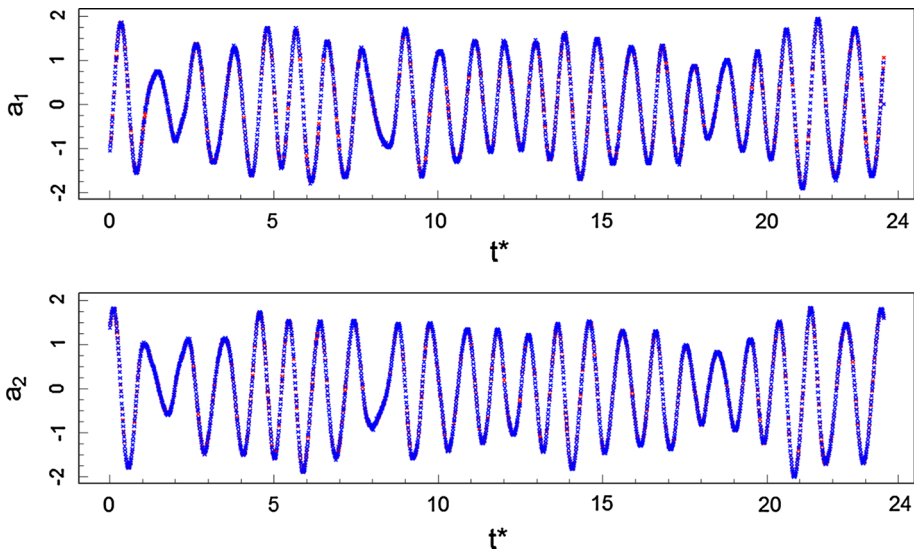

Fig. 11 Reconstructed temporal modes $a_1(t), a_2(t)$—linear ROM—$\alpha = 20^\circ$ and $R_e = 1000$: *red dashed line*, the temporal POD modes; *blue dashed line*, result of the EnKF filtering

6 Conclusion and future work

Bayesian inference methods derived from sequential data assimilation were applied to linear and quadratic POD ROMs in order to reconstruct the flow fields from experimental measurement. The ability of these methods to correct the results of POD ROMs while exploiting avalaible observations has been studied. Different angles of attack and Reynolds numbers were tested in order to apply these methods on the different flow regimes and, consequently, on

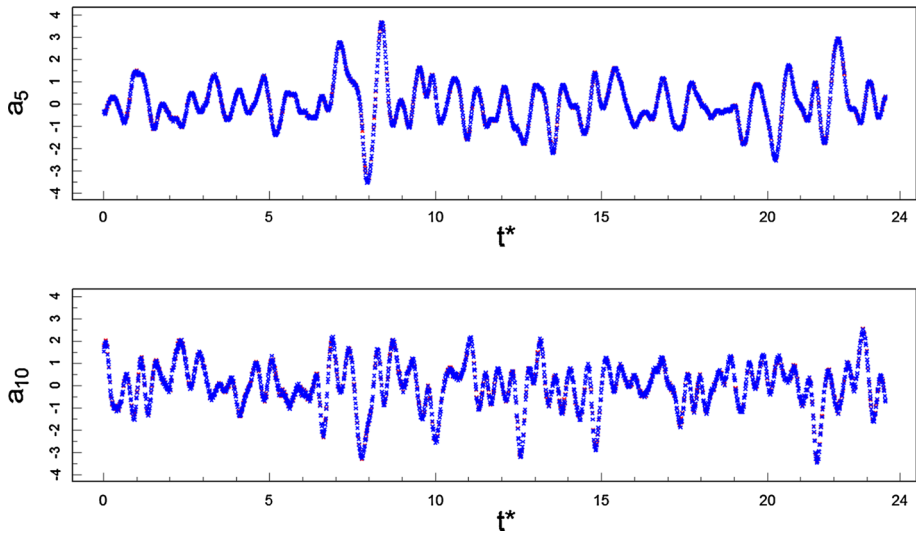

Fig. 12 Reconstructed temporal modes $a_5(t), a_{10}(t)$—linear ROM—$\alpha = 20^\circ$ and $R_e = 1000$: *red dashed line*, the temporal POD modes; *blue dashed line*, result of the EnKF filtering

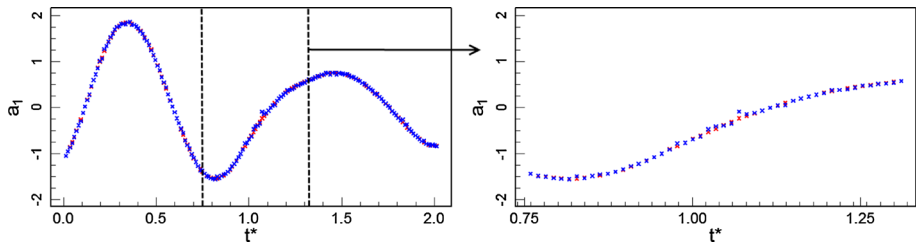

Fig. 13 Reconstructed temporal modes $a_1(t)$—linear ROM—$\alpha = 20^\circ$ and $R_e = 1000$: *red dashed line*, the temporal POD modes; *blue dashed line*, result of the EnKF filtering (zoomed)

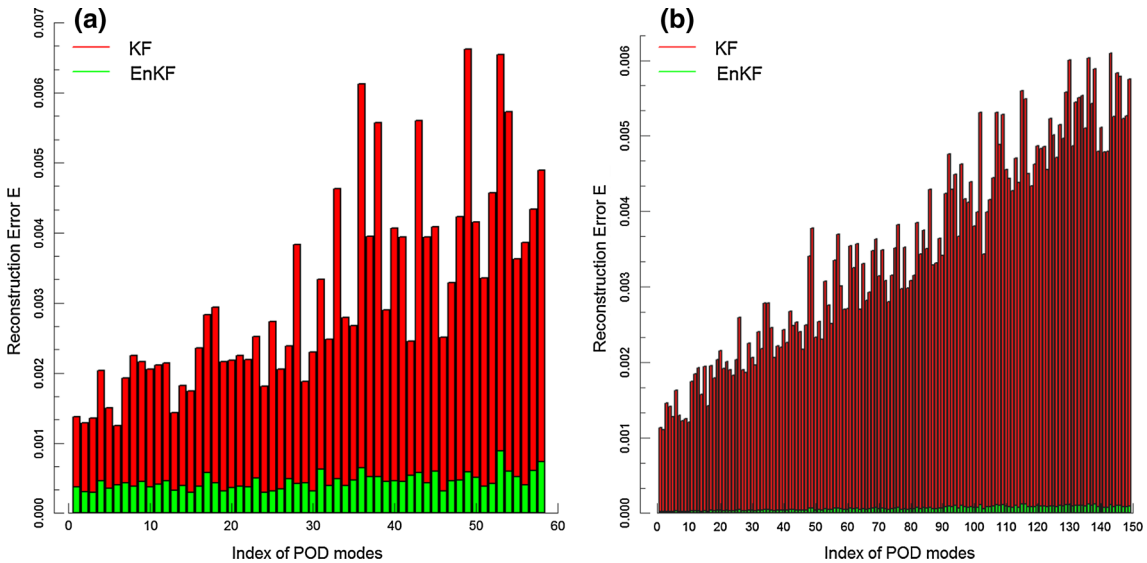

Fig. 14 Reconstruction error E between the reconstructed temporal modes of the linear Kalman, *red bars*, and EnKF filtering, *green bars*: **a** $\alpha = 20^\circ$ and $R_e = 1000$ **b** $\alpha = 20^\circ$ and $R_e = 2000$

different behaviors of the POD ROM. The Bayesian inference methods used here are the following stochastic filtering methods: the linear Kalman filter, the Ensemble Kalman filter and the square-root Ensemble Kalman filter. These Kalman filters enable for the long time PIV sequences to stabilize linear and quadratic reduced-order models that faithfully recover with a certain accuracy the trajectories corresponding to the POD temporal modes.

The Kalman filter sequentially supplies the best least-squares state estimator for a dynamic linear system given a

Table 3 Min to max intervals of error E—linear POD ROM

Lower and upper bound of E			
Re and α	Linear Kalman filter	EnKF	EnSRF
1000 and 10°	[0.0015, 0.0105]	[0.0003, 0.0007]	[0.0004, 0.0009]
1000 and 15°	[0.0075, 0.0626]	[0.00005,0.0007]	[0.0007, 0.0059]
1000 and 20°	[0.0008, 0.0060]	[0.0002, 0.0008]	[0.0001, 0.0008]
1000 and 30°	[0.0008, 0.0086]	[0.0003, 0.0009]	[0.0003, 0.0010]
2000 and 10°	[0.0007, 0.0071]	[0.0002, 0.0004]	[0.000089, 0.0006]
2000 and 15°	[0.0006, 0.0078]	[0.0008, 0.0027]	[0.000081, 0.0007]
2000 and 20°	[0.0010, 0.0059]	[0.000019, 0.00012]	[0.0001, 0.0005]
2000 and 30°	[0.0065, 0.0376]	[0.0002, 0.0004]	[0.0001, 0.0007]

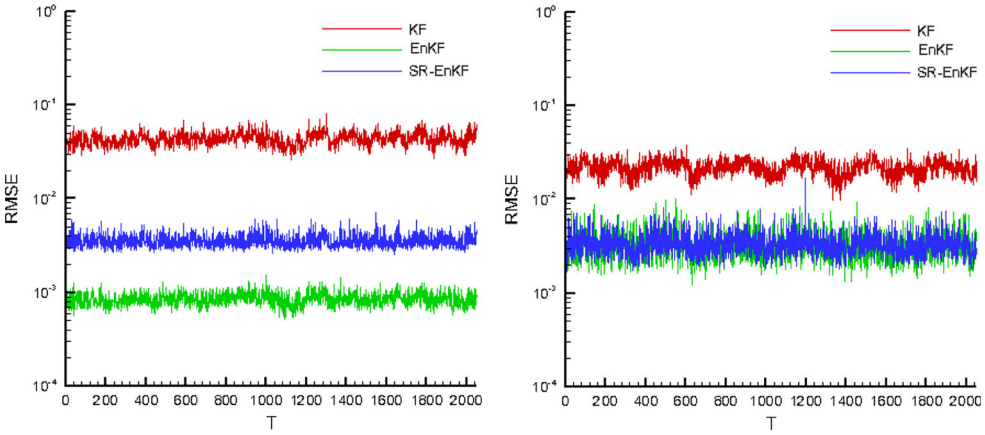

Fig. 15 KF, EnKF and EnSRF RMSE—$\alpha = 20°, R_e = 1000$ and $R_e = 2000$

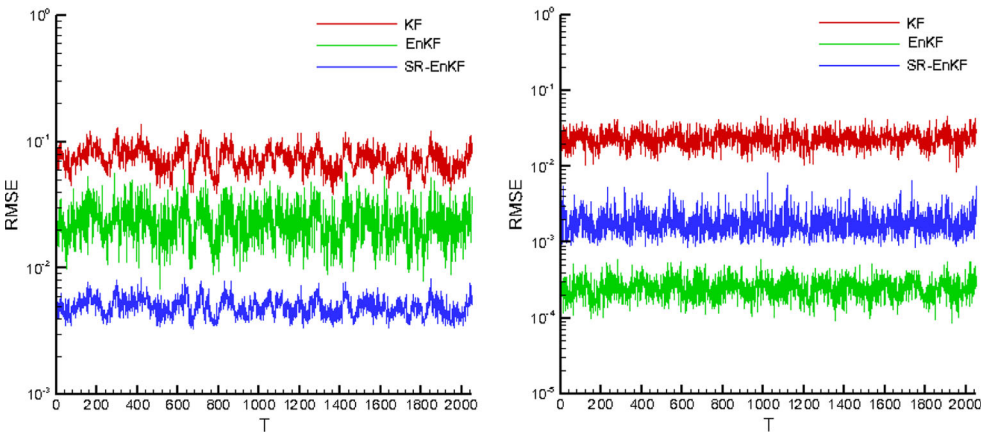

Fig. 16 KF, EnKF and EnSRF RMSE—$\alpha = 15°, R_e = 1000$ and $R_e = 2000$

Table 4 Min to max intervals of RMSE—linear POD ROM

Min to max intervals of RMSE			
Re and α	Kalman filter	EnKF	EnSRF
1000 and 10°	[0.0038, 0.1563]	[0.0036, 0.0138]	[0.0033, 0.0151]
1000 and 15°	[0.0083, 0.0454]	[0.00008, 0.0005]	[0.0008, 0.008]
1000 and 20°	[0.0098, 0.0380]	[0.0012, 0.0110]	[0.0017, 0.0166]
1000 and 30°	[0.0528, 0.1205]	[0.0040, 0.0203]	[0.0050, 0.0194]
2000 and 10°	[0.0231, 0.0817]	[0.0021, 0.0059]	[0.0025, 0.0073]
2000 and 15°	[0.0211, 0.0803]	[0.0048, 0.0473]	[0.0024, 0.0053]
2000 and 20°	[0.0257, 0.0808]	[0.0005, 0.0015]	[0.0025, 0.0071]
2000 and 30°	[0.2972, 0.7433]	[0.0039, 0.0080]	[0.044, 0.0097]

series of available observations. Here, the linear Kalman filter is far less efficient for the same model than the EnKF and EnSRF filters, which result in more homogenous error levels. The EnSRF also gives the lowest reconstruction error and RMSE levels. A difference therefore appears between the linear Kalman filter and the nonlinear Kalman filters based on Monte-Carlo methods. This leads to different behaviors in the growth of instantaneous and global prediction errors in the assimilation interval of the observations. The EnKF and EnSRF filters therefore perform better than the linear Kalman filter. The linear Kalman filter does not offer the same quality of reconstruction, but it does counter the instability of the POD ROM if the number of POD coefficients remains low. The nonlinear filters lead to a significant improvement in the ability of the ROM to show the dynamics of the prediction modes.

Table 5 Min to max intervals of RMSE—Quadratic POD ROM

Min to max intervals of RMSE	
Re and α	EnKF
1000 and 15°	[0.0001, 0.0007]
1000 and 20°	[0.00021, 0.0024]

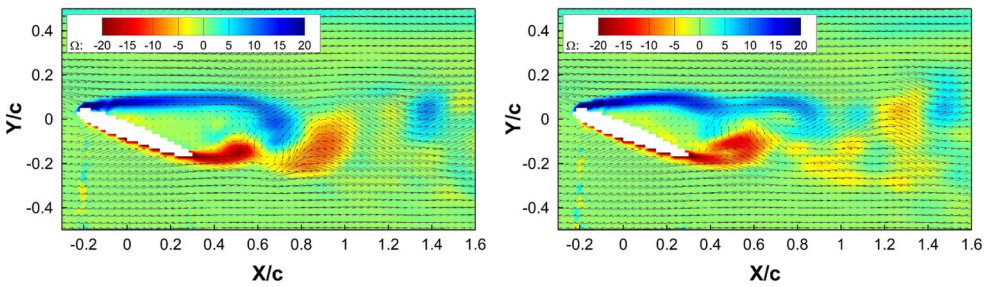

Fig. 17 Snapshots of the vorticity Ω_z field reconstruction with the EnKF applied to the linear POD ROM with a RIC of 90 %—$t^\star = 5.75$ and $t^\star = 11.5$—$\alpha = 20°$ $R_e = 1000$

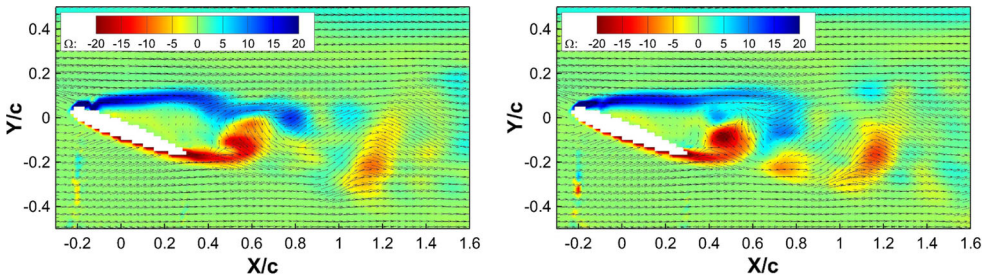

Fig. 18 Snapshots of the vorticity Ω_z field reconstruction with the EnKF applied to the linear POD ROM with a RIC of 90 %—$t^\star = 17.25$ and $t^\star = 23$—$\alpha = 20°$ and $R_e = 1000$

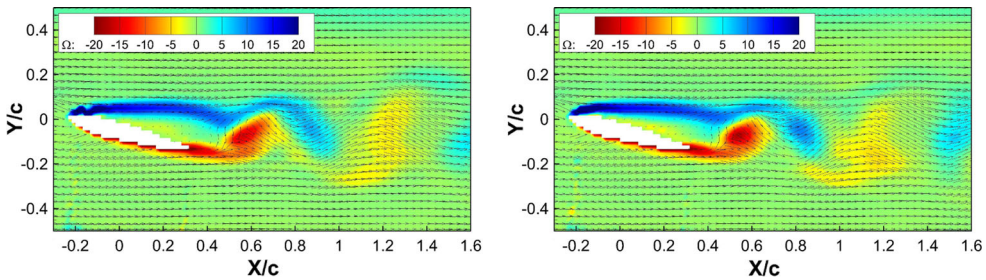

Fig. 19 Snapshots of the vorticity Ω_z field reconstruction with the EnKF applied to the linear POD ROM with a RIC of 90 %—$t^\star = 7.45$ and $t^\star = 14.97$— $\alpha = 15°$ $R_e = 1000$

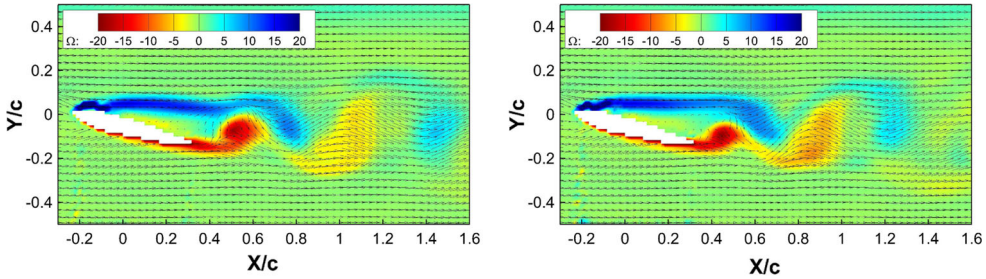

Fig. 20 Snapshots of the vorticity Ω_z field reconstruction with the EnKF applied to the linear POD ROM with a RIC of 90 %—$t^\star = 22.46$ and $t^\star = 29.94$—$\alpha = 15°$ and $R_e = 1000$

Indeed, it is important to carefully choose the N ensemble number used in order to avoid divergence of the filter over time.

The POD ROM, calibrated using Bayesian inference methods, enables an efficient reconstruction of the projection coefficients relative to the reference dynamics for both the linear and quadratic POD ROMs. No amplitude damping, phase shift or amplitude drift are observed. The Kalman filter can stabilize a POD ROM over time without modification of its parameters and without including calibration terms. An analysis of the global relative prediction errors by coefficient and instantaneous error in the L^2 norm confirms a significant improvement in the accuracy of the POD ROM. The low error in each case results in a faithful reconstruction of the coefficients and of the main structures of the flow, as well as the associated vector fields. This validates the application of stochastic filtering as a trustworthy method for calibration of POD ROMs.

The Bayesian inference methods used here allow to taking into account of the noise affecting the experimental measures. An empirical method for adjusting the noise levels of equations of state and of measure was chosen here. Different noise levels and ensemble numbers were tested in order to obtain a sufficiently precise behavior of the ROM, both linear and quadratic. However, it is possible to determine these noise levels using a quantification of the error committed by the PIV, the POD, and the identification of the ROM coefficients.

Acknowledgments This work has supported by the EU-project: Advanced Flow Diagnostics for Aeronautical Research, project no. 265695.

References

Anderson MJ (2001) Permutation tests for univariate or multivariate analysis of variance and regression. Can J Fish Aquat Sci 58(3):626–639

Andrews A (1968) A square root formulation of the kalman covariance equations. AIAA J 6:1165–1166

Aubry N (1991) On the hidden beauty of the proper orthogonal decompositon. Theoret Comput Fluid Dyn 2:339–352

Aubry N, Holmes P, Lumley J, Stone E (1988) The dynamics of coherent structures in the wall region of a turbulent boundary layer. J Fluid Mech 192:125–143

Bergmann M, Cordier L (2008) Optimal control of the cylinder wake in the laminar regime by trust region methods and pod reduced-order models. J Comp Phys 227:7813–7840

Bergmann M, Cordier L, Brancher J (2005) Optimal rotary control of the cylinder wake using pod reduced order model. Phys Fluids 3:1–21

Berkooz G, Holmes P, Lumley J (1993) The proper orthogonal decomposition in the analysis of turbulent flows. Ann Rev Fluid Mech 25:539–575

Bishop C, Etherton B, Majumdar S (2001) Adaptive sampling with the ensemble transform kalman filter. Part I: theoretical aspects. Mon Wea Rev 129(3):420–436

Buffoni M, Camarri S, Iollo A, Salvetti M (2006) Low-dimensional modelling of a confined three dimensional wake flow. J Fluid Mech 569:141–150

Burgers G, Van Leeuwen P, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. Mon Wea Rev 126:1719–1724

Cao Y, Zhu J, Navon I, Luo Z (2007) A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. Int J Numer Methods Fluids 53(10): 1571–1583

Cazemier W, Verstappen R, Veldman A (1998) Proper orthogonal decomposition and low-dimensional models for driven cavity flows. Phys Fluids 10(7):1685–1699

Cordier L, El Majd BA, Favier J (2010) Calibration of POD reduced-order models using Tikhonov regularization. Int J Numer Meth Fluids 63:269–296

Couplet M, Badevant C, Sagaut P (2005) Calibrated reduced-order pod-galerkin systemfor fluid flow modelling. J Comput Phys 207(1):192–220

Deane A, Kevrekidis I, Karniadakis G, Orszag S (1991) Low-dimensional models for complex geometry flows: application to grooved channels and circular cylinders. Phys Fluids A 3(10):2337–2354

Delville J, Ukeiley L, Cordier L, Bonnet J, Glauser M (2001) Examination of large-scale structures in a turbulent plane mixing layer. Part 1. Proper orthogonal decomposition. J Fluid Mech 391:91–122

Evensen E (2004) Sampling strategies and square root analysis schemes for the EnKF. Ocean Dyn 54:539–560

Evensen G (1994) Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99(C5):10143–10162

Evensen G (2003) Ensemble Kalman filter: theoretical formulation and practical implementations. Ocean Dyn 53(4):343–367

Evensen G (2006) Data assimilation the ensemble Kalman filter. Springer, Berlin

Fang F, Pain C, Navon I, Piggot M, GJ G, Goddard A (2009) Reduced order modelling of an adaptative mesh ocean model. Int J Numer Methods Fluids 59(8):827–851

Galletti B, Bruneau C, Zannetti L, Iollo A (2004) Low-order modelling of laminar flow regimes past a confined square cylinder. J Fluid Mech 503:161–170

Galletti B, Bottaro A, Bruneau C, Iollo A (2005) Accurate model reduction of transient flows. RR INRIA 5676:141–148

Gelb A (1974) Applied optimal estimation. MIT Press, Cambridge

Holmes P, Lumley JL, Berkooz G (1996) Turbulence, coherent structures, dynamical systems and symmetry. Cambridge Monographs on Mechanics. Cambridge University Press, New York

Houtekamer P, Mitchell H, Pellerin G, Buehner M, Charron M, L S, B H (2005) Athmospheric data assimilation with an ensemble kalman filter: results with real observations. Mon Wea Rev 133:604–620

Huang R, Wu J, Jeng J, Chen R (2001) Surface flow and vortex shedding of an impulsively started wing. J Fluid Mech 441:265–292

Kalman R, Bucy R (1961) New results in linear filtering and prediction theory. J Basic Eng Trans ASME Ser D 83(3):95–108

Karamanos G, Karniadakis G (2000) A spectral vanishing viscosity method for large eddy simulations. J Comput Phys 163(1):22–50

Le Dimet FX, Talagrand O (1986) Variational algorithm for analysis and assimilation of meteorological observations: theoretical aspects. Tellus 38(A):97–110

Le Gland F (2009) Introduction au filtrage en temps discret—filtre de kalman, filtrage particulaire, modèles de markov cachés. Ecole Nationale Supérieure de Techniques Avancées

Lumley J (1967) The structures of inhomogeneous turbulent flow. In: Yaglom AM, Tatarski VI (eds) Atmospheric turbulence and radioWave propagation. Nauka, Moscow Edition, pp 166–178

McKean H (1969) Propagation of chaos for a class of non-linear parabolic equations. Lectures series in differential equations, vol 2, volume 19 of Van Nostrand Mathematical Studies. Vans Nostrand Reinhold, NewYork, pp 177–194

Noack B, Afanasiev K, Morzynski M, Tadmor G, Thiele F (2003) A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. J Fluid Mech 497:335–363

Noack B, Papas P, Monkewitz P (2005) The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows. J Fluid Mech 523:339–365

Perret L, Collin E, Delville J (2006) Polynomial identification of pod based low-order dynamical system. J Turbul 7(17):1–15

Pham D (2001) Stochastic methods for sequential data assimilation in strongly nonlinear systems. Mon Wea Rev 129:1194–1207

Pham D, Verron J, Gouideau L (1998) Filtres de kalman singuliers évolutifs pour l'assimilation de données en océanographie. CR Acad Sci Paris 326:255–260

Podvin B, Lumley J (1998) A low-dimensional approach for the minimal flow unit. J Fluid Mech 362:121–151

Rempfer D (2000) On low dimensional galerkin models for fluid flows. Theor Comput Fluid Dyn 14(2):75–88

Sirovitch L (1987) Turbulence and the dynamics of coherent structures. Part 1: coherent structures. Quart Appl Math XLV(3): 561–574

Tippett M, Anderson J, Bishop C, Hamill T, Whitaker J (2003) Ensemble square root filters. Mon Wea Rev 31:1485–1490

Whitaker J, Hamil T (2002) Ensemble data assimilation without perturbed observations. Mon Wea Rev 130:913–1924