4D-VARIATIONAL DATA ASSIMILATION FOR POD REDUCED-ORDER MODELS

Gilles Tissot, Laurent Cordier, Nicolas Benard, Bernd R. Noack
PPRIME Institute, CEAT, 43 route de l'aérodrome, 86000 Poitiers, France
Gilles.Tissot@univ-poitiers.fr

ABSTRACT

In flow control, reduced-order models based on Proper Orthogonal Decomposition (POD ROM) are often used as surrogate model for deriving a control law. However, these models are in general too fragile to be used in closed-loop control where the dynamics is strongly modified by the control. Here, a 4D-Variational data assimilation approach (4D-Var) as classically used in meteorology is used to estimate at best the state of the system from inhomogeneous sources of information coming from a model, noisy observations and a background solution. Two complementary strategies (strong constraint 4D-Var and weak constraint 4D-Var) are assessed in the case of a cylinder wake flow with data coming from numerical simulation and PIV experiments.

Keywords : cylinder wake, data assimilation, estimation, POD, reduced-order model

INTRODUCTION

In turbulence, the number of active degrees of freedom is so important that a preliminary step of *model reduction* is compulsory for determining an efficient control strategy. The general objective of model reduction is to extract, from physical insights or mathematical tools, the building blocks which play a dominant role in terms of dynamical modelling. For historical reasons, Proper Orthogonal Decomposition (see Cordier & Bergmann 2008 for an introduction) is the most used reduction approach in the turbulence community. POD is widely employed since it extracts from a sequence of data an orthonormal basis which captures optimally the flow energy. In general, this basis is then used in a Galerkin approach to derive a POD Reduced-Order Model (POD ROM) of the flow by projecting the Navier-Stokes equations onto the POD modes. Unfortunately, this dynamical system is sometimes not sufficiently accurate to predict anything useful in terms of flow control, and identification methods (Cordier *et al.*, 2010) are then used to improve the prediction ability of the POD ROM. In this communication, an alternative procedure is proposed where identification methods are replaced by *data assimilation*.

Data assimilation is a generic methodology which combines heterogeneous observations with the underlying dynamical principles governing the system under observation to estimate at best physical quantities. Starting from a background solution and incoming imperfect informations, an optimal estimation of the true state of the system is determined (see Fig. 1) which takes into account the respective statistical confidences of the different observations. By convention, this true state is often called *analysed solution*. Data assimilation is now routinely used in meteorology and oceanography to combine images coming from satellites, meteorological observations and a dynamical model in order to forecast the weather (Navon, 2009). Since numerical weather prediction is an initial value problem, the goal of assimilation is here to find the best initial condition of the numerical model that minimises the observations errors. In fluid mechanics, data assimilation was more recently introduced for estimating quantities (Papadakis, 2007) and predicting their time evolutions. There are two basic approaches in data assimilation: stochastic estimation that is based on probability considerations (Kalman filtering for instance), and variational data assimilation, which is used in this communication, where the estimation is found as a solution of an optimization problem. When the observations are distributed in time, this approach is referred as four-dimensional variational assimilation or 4D-Var.

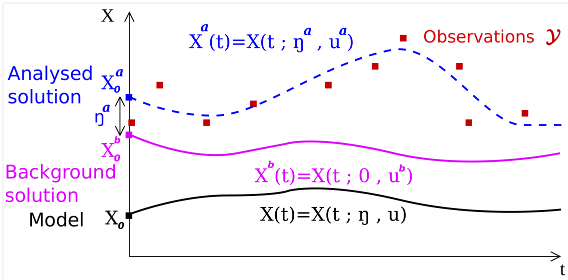

Figure 1. Schematic representation of 4D-Var.

If the dynamical model is assumed to be known perfectly then no extra uncertainty terms are included in the governing equations and the 4D-Var is said with *strong constraint* (see section 1). In that case, the solution of the model is considered to be only dependent on some unknown or imperfectly known parameters: initial condition and coefficients of the model. These values can then be used as control parameters in a minimization problem where the cost functional is built as a sum of an observation error, which measures the difference between the observations and the output of the identified model, and different background errors, which penalize the variation between the background states and the estimated values. In practice, these different terms can be weighted accordingly on the level of statistical confidence that can be evaluated from expert knowledge. Another strategy is to consider a *weak constraint assimilation* approach (see section 2) where the coefficients of the model are assumed to be known in advance, and where an uncertainty function is added to the dynamical model.

The solutions of these constrained optimization problems are computed through an iterative descent algorithm where the gradient of the cost functional with respect to the variation of the control parameters is found by solving an adjoint problem. Computing the solution of an optimality system is known to be computationnaly very expensive since the optimal solution is found iteratively by integrating backward in time the adjoint equation. However, here, the dynamical model corresponds to a POD ROM and then the computation is numerically tractable (D'Adamo *et al.*, 2007).

In this communication, two datasets of increasing dynamical complexity have been used: i) a DNS dataset of a cylinder wake flow at a Reynolds number of 200 (see section 4.1) to validate the algorithm with twin experiments, and ii) an experimental PIV dataset of a cylinder wake flow at a Reynolds number of 40000 (see section 4.2). In the two cases, a POD analysis has first been done and then a POD ROM derived by Galerkin projection of the Navier-Stokes equations onto the POD basis. In the 4D-Var setting, the POD temporal modes served as observations for the assimilation procedure. We will see in section 4 that a qualitative good prediction can be obtained for the two datasets and that relaxing the dynamical constraint can help in the PIV case to follow the observations.

1 STRONG CONSTRAINT 4D-VAR

Let X be the state variable and $\mathbb{M}$ be the non-linear operator relative to the dynamics, the evolution in time of X is given by the following dynamical system:

$$\begin{cases} \dfrac{\partial X}{\partial t}(t)+\mathbb{M}(X(t),u)=0 \\ X(0)=X_0^b+\eta \end{cases} \tag{1}$$

where X_0^b is an *a priori* known initial condition of the system, η an initial condition perturbation and u a set of coefficients of the model.

Strong constraint assimilation consists of finding the solution $X(t)$ of the dynamical system (1) that is closest at the same time to the observations $\mathscr{Y}$ and to regularisation terms called *background* terms. The goal is then to find the control parameters (η,u) which minimize the cost functional $\mathscr{J}$ defined as:

$$\mathscr{J}(u,\eta) = \frac{1}{2}\int_0^T \|\mathscr{Y}(t)-\mathbb{H}(X(t;\eta,u))\|^2_{\mathscr{R}^{-1}}dt + \frac{1}{2}\|\eta\|^2_{\mathscr{B}^{-1}}+\frac{1}{2}\|u-u^b\|^2_{\mathscr{C}^{-1}}. \tag{2}$$

The non-linear operator $\mathbb{H}$, called observation operator, goes from the state space to the observation space. $\mathscr{R}$, $\mathscr{B}$ and $\mathscr{C}$ are covariance tensors of the observation space, state space and control space. They are related respectively to the observations, the state variables' initial conditions and to the control variables. The norms $\|\cdot\|_{\mathscr{R}^{-1}}$, $\|\cdot\|_{\mathscr{B}^{-1}}$ and $\|\cdot\|_{\mathscr{C}^{-1}}$ are induced norms of the inner products $\langle\mathscr{R}\cdot,\cdot\rangle$, $\langle\mathscr{B}\cdot,\cdot\rangle$ and $\langle\mathscr{C}\cdot,\cdot\rangle$. The role of these covariance tensors is to give more or less confidence in the observations and background states. These tensors have a key influence for combining inhomogeneous sources of information in data assimilation. The covariance tensors may be chosen based on extra knowledge of the system. Here, they have been simply defined as diagonal tensor. The optimal control parameters (η^a,u^a) are called *analysed solutions* and the associated analysed dynamics $X^a(t)$ is the best estimation of the system's state, solution of (1) according to the criterion (2). The couple $(0,u^b)$ corresponds to the *background* solution of the problem. Data assimilation is described schematically in Fig. 1.

The minimization of $\mathscr{J}$ is done using the limited storage variant of the BFGS quasi-Newton algorithm (Gilbert & Lemaréchal, 2009). For determining the descent directions, the gradients of the functional with respect to the two control variables η and u need first to be evaluated. The use of finite differences to determine the gradient of the cost functional is in practice unfeasible when the dimension of the state variables is too large. An elegant solution is to write an *adjoint formulation* of the problem. We will see soon that the determination of the gradient of $\mathscr{J}$ with respect to the control variables then corresponds to the forward integration of the dynamical system (1) followed by a backward integration of an adjoint dynamical model to be determined.

In variational data assimilation, the analysed dynamics is found as solution of the constrained minimization problem given by (2) and (1). A classical way for solving this type of constrained optimization problem (Gunzburger, 1997) is by introducing a Lagrangian functional $\mathscr{L}$ defined by

$$\begin{aligned}\mathscr{L}(X(t),u,\eta,\lambda(t),\mu) &= \mathscr{J}(u,\eta)\\ &-\int_0^T\left\langle \frac{\partial X(t)}{\partial t}+\mathbb{M}(X(t),u),\lambda(t)\right\rangle dt\\ &-\langle X(0)-X_0^b-\eta,\mu\rangle\end{aligned} \tag{3}$$

where $\lambda(t)$ and μ are two Lagrange multipliers introduced to enforce the constraints given by (1). When the minimum of $\mathscr{J}$ is reached, $\nabla\mathscr{J}=\nabla\mathscr{L}=0$.

Setting the first variation of $\mathscr{L}$ with respect to X to 0 leads to the adjoint equation

$$\begin{aligned}-\frac{\partial\lambda}{\partial t}(t)+\left(\frac{\partial\mathbb{M}}{\partial X}\right)^*\lambda(t)\\ =\left(\frac{\partial\mathbb{H}}{\partial X}\right)^*\mathscr{R}^{-1}\left(\mathbb{H}(X(t))-\mathscr{Y}(t)\right),\end{aligned} \tag{4}$$

where $\frac{\partial\mathbb{M}}{\partial X}$ and $\frac{\partial\mathbb{H}}{\partial X}$ denote respectively the linear tangent operator[1] of $\mathbb{M}$ and $\mathbb{H}$, and $\left(\frac{\partial\mathbb{M}}{\partial X}\right)^*$ and $\left(\frac{\partial\mathbb{H}}{\partial X}\right)^*$ their adjoint operators[2]. The adjoint equation (4) is defined backward in time with the terminal condition $\lambda(T)=0$.

Setting the first variation of $\mathscr{L}$ with respect to the control parameters u and η to 0 leads to the optimality condi-

[1]The linear tangent of an operator A is the directional derivative operator or Gâteaux derivative of A defined as:

$$\frac{\partial A}{\partial X}(X)\delta X = \lim_{h\to 0}\frac{A(X+h\delta X)-A(X)}{h} \quad \forall\delta X.$$

[2]The adjoint A^* of a linear operator A on a space $\mathscr{D}$ is such that $\forall x,y\in\mathscr{D}, \langle Ax,y\rangle=\langle x,A^*y\rangle$.

tions

$$\begin{cases} \dfrac{\partial \mathscr{J}}{\partial u} = -\displaystyle\int_0^T \left(\dfrac{\partial \mathbb{M}}{\partial u}\right)^* \lambda(t)dt + \mathscr{C}^{-1}(u - u^b) \\ \dfrac{\partial \mathscr{J}}{\partial \eta} = \lambda(0) + \mathscr{B}^{-1}\eta. \end{cases} \tag{5}$$

These optimality conditions can then be evaluated to determine the gradient of $\mathscr{J}$ as soon as the Lagrange multipliers are known, it means as soon as the adjoint equation (4) is integrated backward in time.

2 WEAK CONSTRAINT 4D-VAR

In the strong constraint 4D-Var context (see section 1), the coefficients u of the model were considered to be adjustable parameters that can be optimally determined to reproduce imperfect observations of the dynamical system. Here, this constraint is relaxed and we assume that the coefficients of the dynamical system are directly provided by the data. For taking into account errors of the model, the dynamical system is then defined up to an additive uncertainty function $w(t)$ considered as Gaussian white noise. In the weak constraint 4D-Var approach, the dynamical system becomes:

$$\begin{cases} \dfrac{\partial X(t)}{\partial t} + \mathbb{M}(X(t)) = w(t) \\ X(0) = X_0^b + \eta. \end{cases} \tag{6}$$

In that case, the control parameters of the variational assimilation approach are $w(t)$ and η and the cost functional to be minimized is:

$$\begin{aligned} \mathscr{J}(w(t),\eta) = & \frac{1}{2}\int_0^T \|\mathscr{Y}(t) - \mathbb{H}(X(t))\|^2_{\mathscr{R}^{-1}} dt \\ & + \frac{1}{2}\|\eta\|^2_{\mathscr{B}^{-1}} + \frac{1}{2}\int_0^T \|w(t)\|^2_{\mathscr{W}^{-1}} dt, \end{aligned} \tag{7}$$

where $\mathscr{W}$ is the covariance matrix of $w(t)$. To solve the constrained optimization problem associated to this new formulation, the same procedure as the one described in section 1 is followed. A Lagrangian functional defined as

$$\begin{aligned} & \mathscr{L}(X(t), w(t), \eta, \lambda(t), \mu) = \mathscr{J}(w(t),\eta) \\ & - \int_0^T \left\langle \frac{\partial X(t)}{\partial t} + \mathbb{M}(X(t)) - w(t), \lambda(t) \right\rangle dt \\ & - \langle X(0) - X_0^b - \eta, \mu \rangle, \end{aligned} \tag{8}$$

is first introduced. The first variation of $\mathscr{L}$ with respect to $X(t)$ leads to the same adjoint equation as in the case of the strong constraint 4D-Var *i.e.*

$$\begin{aligned} & -\frac{\partial \lambda}{\partial t}(t) + \left(\frac{\partial \mathbb{M}}{\partial X}\right)^* \lambda(t) \\ & \quad = \left(\frac{\partial \mathbb{H}}{\partial X}\right)^* \mathscr{R}^{-1}\left(\mathbb{H}(X(t)) - \mathscr{Y}(t)\right), \end{aligned} \tag{9}$$

with the terminal condition $\lambda(T) = 0$. Finally, the first variation of $\mathscr{L}$ with respect to the control parameters leads to the following optimality conditions:

$$\begin{cases} \dfrac{\partial \mathscr{J}}{\partial w}(t) = \lambda(t) + \mathscr{W}^{-1}w(t) \\ \dfrac{\partial \mathscr{J}}{\partial \eta} = \lambda(0) + \mathscr{B}^{-1}\eta. \end{cases} \tag{10}$$

In this formulation, the uncertainty function $w(t)$ is part of the dynamical system. Then, since the optimal solution is found on a given time horizon T, it means that the weak constraint 4D-Var approach can not be used to predict the system's state after the end of the assimilation time interval.

3 POD REDUCED-ORDER MODEL

Given the high spatio-temporal complexity of turbulent flows, adopting a model-based approach in flow control is extremely attractive. Indeed, difficult to imagine deriving an efficient control strategy in open-loop and even more, in closed-loop, if no dynamical model is used in the design process. A natural tendency would be to go towards models based on first principles. However, the number of active degrees of freedom in turbulence is so high that it will lead to high fidelity model of very large dimension. A way to cope with this difficulty is to employ surrogate models for developing the control strategy. Proper Orthogonal Decomposition (Cordier & Bergmann, 2008) is often used for this purpose since it extracts, from snapshots, modes that are optimal to capture the energy of the system. Starting from a set of N_s snapshots of velocity fields $\boldsymbol{v}$ taken evenly over a time interval $[0, T_s]$, snapshot POD can be used to determine spatial modes $\boldsymbol{\Phi}_i$ and temporal coefficients $a_i^{\mathrm{P}}(t)$ such that:

$$\boldsymbol{v}(\boldsymbol{x},t) = \boldsymbol{v}_m(\boldsymbol{x}) + \sum_{i=1}^{N_s} a_i^{\mathrm{P}}(t)\boldsymbol{\Phi}_i(\boldsymbol{x}), \tag{11}$$

where $\boldsymbol{v}_m$ corresponds to the average of the snapshots and where $\boldsymbol{x} \in \Omega$, the spatial domain of interest. Truncating the number of modes in (11) to N_{gal}, with $N_{\mathrm{gal}} \ll N_s$, this expansion is substituted into the Galerkin projection of the incompressible Navier-Stokes equations onto the spatial modes $\boldsymbol{\Phi}_i$ to obtain a POD Reduced-Order Model (POD ROM). After some algebraic manipulations (Cordier *et al.*, 2010), the following expression is found for the POD ROM:

$$\frac{da_i^{\mathrm{R}}(t)}{dt} = C_i + \sum_{k=1}^{N_{\mathrm{gal}}} L_{ik} a_k^{\mathrm{R}}(t) + \sum_{k=1}^{N_{\mathrm{gal}}}\sum_{j=1}^{N_{\mathrm{gal}}} Q_{ijk} a_j^{\mathrm{R}}(t) a_k^{\mathrm{R}}(t) \tag{12}$$

where $a_i^{\mathrm{R}}(0) = a_i^{\mathrm{P}}(0)$. The constant, linear and quadratic coefficients, C_i, L_{ik} and Q_{ijk} depend explicitly on $\boldsymbol{\Phi}_i$ and $\boldsymbol{v}_m$ and as such their values can be directly determined. However, it is well known (Cordier *et al.*, 2010) that for different reasons (structural instability of the Galerkin projection, truncation of the POD basis, inaccurate treatment of the boundary and pressure terms) the dynamical system (12) does not represent sufficiently well the correct dynamics. This problem is then perfectly matching the objectives of data assimilation as described in introduction. The temporal coefficients $a_i^{\mathrm{R}}(t)$ can be considered as state variables $X(t)$. The POD ROM (12) can serve as dynamical model

$\mathbb{M}$ and the POD coefficients $a_i^{\mathrm{P}}(t)$ as observations $\mathscr{Y}(t)$. Finally, the coefficients C_i, L_{ik} and Q_{ijk} of (12) can be used as control parameters u in the strong constraint 4D-Var. Concerning the background solutions X_0^b and u^b, they can be determined from the temporal coefficients obtained directly by POD ($a_i^{\mathrm{P}}(t)$) for the initial condition, and from the values determined by Galerkin projection or identification (Cordier *et al.*, 2010) for the coefficients of the POD ROM. The two 4D-Var formulations described in sections 1 and 2 can then be easily applied.

As a final remark, let us comment on the differences between the approach followed in this communication and the reduced-order strategy often employed for 4D-Var data assimilation in meteorology (Robert *et al.*, 2005; Daescu & Navon, 2007). In flow control, a low-order dynamical system is often first developed and then an optimal control approach, or here a variational data assimilation procedure, is then applied. In meteorology, the steps of variational data assimilation and reduced-order modelling are reversed. A full dynamical model is first considered for the data assimilation and since the analysed solution is searched iteratively by integrating forward/backward in time the direct/adjoint systems, the control space is then restricted to a low-dimensional space spanned by the first POD eigenfunctions. In the two strategies, the computational costs are highly decreased. However, the influence of the chosen strategy on the determination of the analysed solution is still not well determined.

4 RESULTS

In this section, the variational data assimilation approach will be tested in a simple flow configuration corresponding to the cylinder wake. Due to its simple geometry and its representative behaviour of separated flows (Zdravkovich, 1997), the viscous flow past a circular cylinder has been extensively used in the past decade as a test bed to develop control strategies (Bergmann *et al.*, 2005). Here, the first objective is to evaluate the ability of 4D-Var to improve the description of the dynamics within the time horizon where the observations are known. By definition (see introduction), this interpolatory requirement should be offered by 4D-Var. A second objective is to assess the predictive behaviour of the dynamical system obtained as solution of 4D-Var and to measure the influence of strong and weak constraint hypothesis on the analysed dynamics.

In a first time (see section 4.1), the 4D-Var approach will be applied on numerical data to test the method and evaluate the role of some numerical parameters. In particular, the assimilation procedure will be exercised with a convenient methodology, termed as twin experiments. In a second time (see section 4.2), an experimental dataset based on PIV data will be used to analyse this time the influence of the dynamical complexity of the observed dynamics on the analysed solution obtained by 4D-Var.

4.1 Numerical dataset

4D-Var The 4D-Var approach is here applied to a two-dimensional incompressible cylinder wake flow at $Re = 200$. The database was computed using a finite-element code (DNS code Icare, IMFT/university of Toulouse, see Favier 2007 for details) and contains $N_s = 200$ two-dimensional snapshots of the flow velocity, taken over a period $T_s = 12$ *i.e.* over more than two periods of vortex shedding ($T_{\mathrm{vs}} = 5$). Since 4D-Var is applied on POD-ROM with observations corresponding to the POD temporal coefficients a_i^{P}, snapshot POD is first applied on the previous data. The first six POD modes are here sufficient to represent 99.9% of the flow energy (see Cordier *et al.* 2010 for more details on the procedure and on the POD results) indicating that $N_{\mathrm{gal}} = 6$ should be sufficient for the order of the POD ROM (12).

The strong constraint 4D-Var, as described in section 1, is now applied. The background of the initial condition is given by $a_i^{\mathrm{P}}(t=0)$ and the background of the coefficients C_i, L_{ik} and Q_{ijk} of (12) are equal to zero. These background values are also used to initialize the coefficients of the dynamical system (12) at the beginning of the iterative procedure. Lastly, the covariant matrices are chosen as $\mathscr{R}^{-1} = \mathbb{I}$ and $\mathscr{B}^{-1} = \mathscr{C}^{-1} = \sigma^2\mathbb{I}$ where $\sigma = 10^{-3}$ and $\mathbb{I}$ denotes the identity matrix.

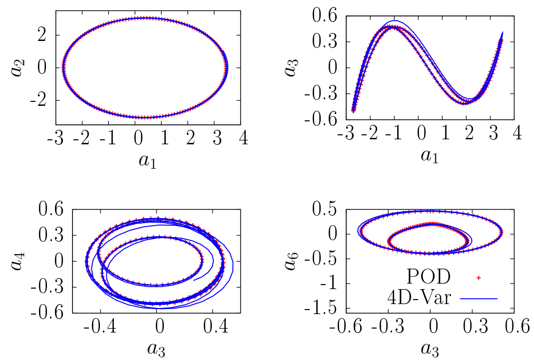

Figure 2. Results of 4D-Var for the DNS dataset (perfect observations).

Figure 2 represents the results of strong constraint 4D-Var for the DNS dataset when perfect observations are used. The 4D-Var approach has been first applied over the 200 time steps contained in the database (assimilation window). Then, the analysed dynamical model has been integrated in time over 400 time steps (forecast window) to conclude on the predictive character of the model. Very good agreements are obtained between the observations and the analysed dynamics. Moreover, the analysed model can predict correctly the dynamics over twice the assimilation period.

Twin experiments Twin experiments (Titaud *et al.*, 2010) is a procedure used to test 4D-Var performances. It consists first to get a "true state" of the system. This true state can be known analytically or, as we will do, can come from a 4D-Var procedure with perfect observations. In a second step, modified observations are artificially generated by under-sampling and noising the perfect observations. At this point, 4D-Var can then be performed and the analysed solution be compared with the true state. This procedure shows the ability of the model to represent the state under observation and to predict its future. The analysed dynamics obtained previously with perfect observations (see Fig. 2) is now considered as true state for the twin experiments. The previous observations $a_i^{\mathrm{P}}(t)$ are modified by adding a Gaussian noise defined as $\mathscr{X}/\sigma_i$ where $\mathscr{X} \simeq \mathscr{N}(0, \sigma_t^2)$ with $\sigma_t = 0.2$ and $\sigma_i = 2[\frac{i+1}{2}]$ where $[x]$ returns the nearest integer to x. These noisy states are taken as observations for the twin experiments. The strong constraint 4D-Var approach is applied with the same numerical parameters as in the previous case.

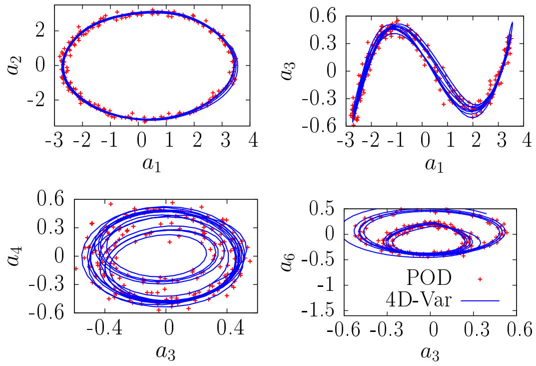

Figure 3. Results of 4D-Var twin experiments for the DNS dataset (noised observations).

Figure 3 represents the results of the twin experiments for the DNS dataset. Despite the noisy observations used in the 4D-Var approach, the analysed dynamics is in good agreement with the expected true state. Moreover, the forecast dynamics corresponds to the correct attractor of the system. The interest of the twin experiments framework is to fairly compare different assimilation procedures based on the values of the error $e(t)$ defined as

$$e(t) = \sqrt{\sum_{i=1}^{N_{\text{gal}}} \left(a_i(t) - a_i^{\text{True}}(t)\right)^2}, \tag{13}$$

where $a_i(t)$ corresponds to the i^{th} temporal coefficient of the POD expansion. Figure 4 represents the time evolution of (13) for the twin experiments. The minimum level of error that can be obtained corresponds to the case where perfect observations are used (Fig. 2). These reference values are represented in Fig. 4 for comparison with the case where the observations are noised. In addition, the time error for the analysed dynamics is also given in Fig. 4. As it can be expected from the 4D-Var approach, the value of error is systematically lower than for the noised observations over the assimilation time window. Moreover, this error does not increase in the forecast window meaning that the predictability of the analysed dynamical system is satisfactory.

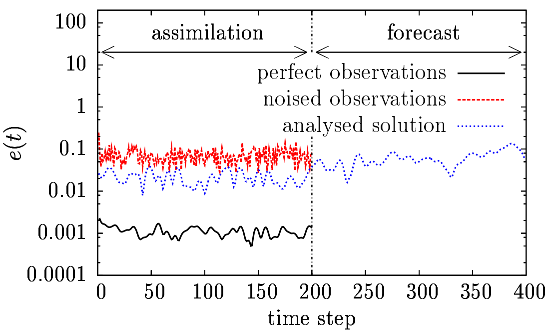

Figure 4. Time evolutions of errors. Comparison between the case of perfect observations and the case of twin experiments.

4.2 Experimental dataset

In section 4.1, the strong constraint 4D-Var approach has been applied successfully for a laminar cylinder wake obtained by numerical simulations. Here, data assimilation will be applied on data obtained by 2D-2C PIV measurements for a turbulent cylinder wake (Benard *et al.*, 2010) corresponding to a sub-critical flow regime ($Re_D = DU_\infty/\nu = 40000$ where $D = 40\,$mm is the cylinder diameter and $U_\infty = 15.6\ m.s^{-1}$ is the free-stream velocity). The database contains $N_s = 5130$ snapshots taken at a sampling frequency $f_s = 1\,$k Hz over approximatively 400 periods of vortex shedding. First, a snapshot POD has been performed on this dataset. The first 16 POD modes capture 31% of the flow energy. It was found to be sufficient for describing the wake flow.

Strong constraint 4D-Var A strong constraint 4D-Var approach is now applied to the first 128 time steps contained in the database. Compared to the case of section 4.1, the dynamics is much more complex and the PIV data correspond to a 2D description of a pure 3D physical phenomenon. The observations are then noised and incomplete, a typical situation where data assimilation should help to reconstruct optimally the flow states. As previously, a POD ROM derived this time with $N_{\text{gal}} = 16$ served as dynamical model for the assimilation.

Similarly to the case of the numerical database, the background of the initial condition is chosen as $a_i^{\text{P}}(t = 0)$. For the background of the coefficients of the model (12), the situation is different since the configuration is more complex. In the previous case, the coefficients were initialized to zero. Here, a preliminary identification is performed following the method described in Cordier *et al.* (2010). The background coefficients C_i^b, L_{ik}^b and Q_{ijk}^b are searched as to minimize a quadratic cost functional $\mathscr{I}$ given by

$$\mathscr{I}(C_i, L_{ik}, Q_{ijk}) = \tag{14}$$
$$\frac{1}{T_{\text{cal}}}\int_0^{T_{\text{cal}}} \left\| \frac{d\boldsymbol{a}^{\text{P}}}{dt}(t) - \mathbb{M}_{\text{gal}}(\boldsymbol{a}^{\text{P}}(t), C_i, L_{ik}, Q_{ijk}) \right\|^2 dt,$$

where $\mathbb{M}_{\text{gal}}$ corresponds to the right hand side of (12) and where $T_{\text{cal}} = 102.4\,$ms is the length of the identification window. Finally, the same values as in section 4.1 are chosen for the covariant matrices.

The results of strong constraint 4D-Var are represented in Fig. 5. For the first POD modes, the analysed dynamics is well reproduced and is smoother than the observations, especially for the modes 1 and 2 corresponding to the Von Kármán vortex shedding. For higher order modes, the estimation obtained by 4D-Var has a smaller amplitude than the observations. This difference of quality between the estimation of the large and fine scales can be explained by: i) POD truncation that neglects in the POD ROM the effect of the fine scales on the large scales of turbulence, ii) 2D observations used for describing a 3D phenomenon, and iii) signal-to-noise ratio that is lower for the fine scales. The objective of the next section is to see if the 4D-Var estimation can be improved by using a weak constraint 4D-Var approach.

Weak constraint 4D-Var Here, the dynamical constraint given by the POD ROM (12) is relaxed and the modelling error is considered to be represented globally with an additive Gaussian noise $w(t)$. The same background solutions $(0, C_i^b, L_{ik}^b, Q_{ijk}^b)$ as the ones used previously in the strong constraint 4D-Var approach is now employed for performing a weak constraint 4D-Var. In addition, the values of the covariance matrices are also the same as in the strong constraint 4D-Var. The analysed solution is shown in Fig. 5 for comparison with the solution obtained by strong 4D-Var.

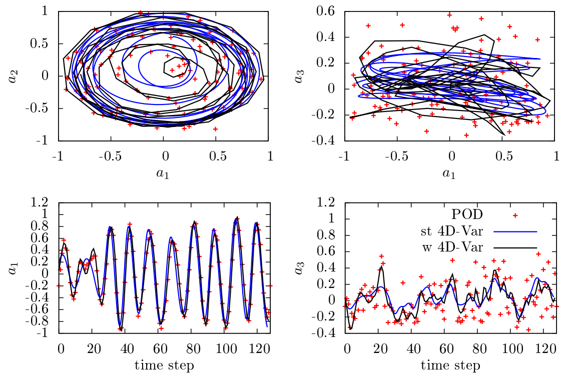

Figure 5. Results of 4D-Var for the PIV dataset. Comparison between strong (st) and weak constraint (w) 4D-Var. Axes associated to a temporal coefficient are multiplied by a factor 1000.

The estimations obtained by weak constraint 4D-Var approach are qualitatively slightly better than those obtained by strong constraint 4D-Var. However, we have to keep in mind that the additive noise $w(t)$ is defined only over the assimilation window preventing the use of the analysed model for forecasting the flow dynamics.

CONCLUSION

4D-Var data assimilation is a well established method in meteorology and oceanography. Essentially, it combines different inhomogeneous sources of information (data, dynamical model) to estimate optimally the true state of the system and to potentially predict its future dynamics. In this paper, this approach has been applied in a fluid mechanics context to a POD ROM of a cylinder wake flow. The control parameters were respectively the initial conditions and the coefficients of the POD ROM for the strong constraint 4D-Var and the initial conditions and an additive noise in the weak constraint 4D-Var.

The strong constraint 4D-Var procedure has first been tested on a numerical dataset corresponding to a low value of Reynolds number. The original dynamics was well reproduced in the case of the perfect observations and also in twin experiments. Moreover, a good predictability of the analysed dynamical model was found. For the experimental database and the strong constraint 4D-Var, only the dynamics of the higher POD modes is well reproduced. The dynamical constraint was then relaxed by considering a weak constraint 4D-Var. With this approach, the estimation was qualitatively improved over the assimilation window but no prediction is possible.

An important result of this study was to demonstrate numerically that the modification of the initial condition plays a crucial role for capturing the correct dynamics with a POD ROM. Finally, the success of the 4D-Var approach is deeply linked to the accurate estimation of the different covariance matrices used in the procedure. This point should be studied more carefully in the future.

ACKNOWLEDGMENT

Nicolas BENARD (Electro-Fluido-Dynamic group of the Pprime institute) is warmly acknowledged for providing us with the PIV experimental database of the cylinder wake flow. This work has received support from the National Agency for Research on reference ANR-08-BLAN-0115-01.

References

Benard, N., Debien, A., David, L. & Moreau, E. 2010 Analyse par PIV rapide du sillage d'un cylindre manipulé par actionneurs plasmas. In *Congrès Francophone de Techniques Laser*. Vandoeuvre-lès-Nancy 14-17 septembre.

Bergmann, M., Cordier, L. & Brancher, J.P. 2005 Optimal rotary control of the cylinder wake using Proper Orthogonal Decomposition Reduced-Order Model. *Physics of Fluids* **17**, 097101.

Cordier, L., Abou El Majd, B. & Favier, J. 2010 Calibration of POD Reduced-Order Models using Tikhonov regularization. *International Journal for Numerical Methods in Fluids* **63** (2).

Cordier, L. & Bergmann, M. 2008 Proper Orthogonal Decomposition: an overview. In *Lecture series 2002-04, 2003-03 and 2008-01 on post-processing of experimental and numerical data*. Von Kármán Institute for Fluid Dynamics.

D'Adamo, J., Papadakis, N., Mémin, E. & Artana, G. 2007 Variational assimilation of POD low-order dynamical systems. *Journal of Turbulence* **8** (9), 1–22.

Daescu, D.N. & Navon, I.M. 2007 Efficiency of a POD-based reduced second-order adjoint model in 4D-Var data assimilation. *International Journal for Numerical Methods in Fluids* **53** (6), 985–1004.

Favier, J. 2007 Contrôle d'écoulements : approche expérimentale et modélisation de dimension réduite. PhD thesis, Institut National Polytechnique de Toulouse.

Gilbert, J.C. & Lemaréchal, C. 2009 The module M1QN3 – Version 3.3. *INRIA Rocquencourt & Rhône-Alpes* .

Gunzburger, M. D. 1997 Introduction into mathematical aspects of flow control and optimization. In *Lecture series 1997-05 on inverse design and optimization methods*. Von Kármán Institute for Fluid Dynamics.

Navon, I. M. 2009 Data assimilation for Numerical Weather Prediction: a review. In *Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications*. Springer Book.

Papadakis, N. 2007 Assimilation de données images : application au suivi de courbes et de champs de vecteurs. PhD thesis, Université de Rennes I.

Robert, C., Durbiano, S., Blayo, E., Verron, J., Blum, J. & Le Dimet, F.-X. 2005 A reduced-order strategy for 4D-Var data assimilation. *Journal of Marine Systems* **57** (1-2), 70–82.

Titaud, O., Vidard, A., Souopgui, I. & Le Dimet, F.-X. 2010 Assimilation of image sequences in numerical models. *Tellus A* **62** (1), 30–47.

Zdravkovich, M.M. 1997 *Flow Around Circular Cylinders, vol. 1. Fundamentals*. Oxford University Press.