DATA ASSIMILATION APPLIED TO NUMERICAL 2D-CYLINDER WAKE FLOW

N. Kumar*, F. Kerhervé, L. Cordier

Institut Pprime, CNRS – Université de Poitiers – ISAE-ENSMA, UPR 3346
11 Boulevard Marie et Pierre Curie, BP 30179, F86962 Futuroscope Chasseneuil Cedex, France
*Corresponding author: nishant.kumar@cnrs.pprime.fr

Abstract

The design of model-based flow controllers requires the knowledge of a dynamical model of the flow. However, real-time and robust estimation of the flow state remains a challenging task when only limited spatial and temporal discrete measurements are available. In this study, the objective is to draw upon the methodologies implemented classically in meteorology to develop dynamic observers for flow control applications. A well established data assimilation method based on Kalman filter is considered. This approach is here extended to both estimate model states and flow parameters. The strategy is numerically demonstrated on a POD Reduced-Order Model of a 2D-cylinder wake flow at low Reynolds number.

Keywords : data assimilation, Dual Ensemble Kalman Filter, POD Reduced-Order Model, cylinder wake flow

1 Introduction

Despite decades of intensive research in shape optimization, aerodynamic mechanisms such as separation and mixing still represent an important source of energy expenditure in transport vehicles. The manufacturers have consequently developed, over the years, a range of strategies to improve the aerodynamic performance of their vehicles. One of these strategies consists in using active devices that can be actuated on and off based on the change of the inflow conditions. To drive such actuators, command laws and more globally valid controllers are required. To design these controllers, different control strategies can be employed, one of which being the model-based approach. This one requires the knowledge of a dynamical model of the flow that one wants to control in real-time. In reality, the identification of such a model is challenging because it is generally obtained from limited spatial and time discrete measurements.

This problem of state estimation is well-known in the field of meteorology where data is collected at several locations, with spatial and temporal scales varying in orders of magnitude, to estimate the evolution of the weather (Talagrand and Courtier, 1987). Classical methodologies involve the use of data assimilation (DA) techniques where data is assimilated, when available, with a dynamical model to correct the state estimate obtained by the model, and eventually, to correct the model itself for further, more accurate predictions. In the specific context of aerodynamic-related problems, the complexity resides essentially in the large number of degrees of freedom of the flow dynamics, in the broad range of scales to examine and in non-linearities driving the flow dynamics. To design a physical model that

can be managed in real-time applications, the Navier-Stokes equations are traditionally projected onto an appropriate reduced basis leading to a reduced-order model of the flow (Rowley and Dawson, 2017). Proper Orthogonal Decomposition (POD, Lumley 1967) is typically used to design the orthonormal basis which captures the flow features which are believed essential to represent the dynamics. 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 for application in flow control, and identification methods are then used to improve the prediction ability of the POD ROM (Cordier et al., 2010). An alternative is to use the ROM in conjunction with DA techniques to design dynamic observers able to predict the flow state from limited information (Papadakis, 2007).

In this paper, the stochastic data assimilation method known as Dual Ensemble Kalman Filter (Dual EnKF) is used on a numerical dataset of a $2D$-cylinder wake flow considered at low Reynolds number ($Re = 100$). In the DA setting, the POD temporal modes serve as observations for the assimilation procedure which is used to test the predictability of the DA method. In § 2, an overview of snapshot POD (§ 2.1), POD ROM (§ 2.2), and Dual EnKF (§ 2.3) is first given. POD Reduced-Order model is then applied to the cylinder wake flow in § 3. The use of data assimilation to predict the temporal POD coefficients for the numerical flow is discussed in § 4. Section 5 offers concluding remarks.

2 Model reduction and ensemble Kalman filter

2.1 Snapshot POD

Proper Orthogonal Decomposition is a reduced-order model approach which was introduced to the fluid dynamics community as a mathematical tool to extract coherent structures from turbulent flow fields (Lumley, 1967). It provides an objective algorithm to decompose a set of data into a minimal number of basis functions or *modes* to capture as much as possible the original dynamics. It is a data-driven method and therefore requires only the flow field data for modal decomposition and does not necessarily require the knowledge of the dynamical equations. POD extracts the modes based on optimizing the mean square of the field variable under observation (Cordier and Tissot, 2013).

Let $\boldsymbol{\chi}$ be the spatial coordinates and t represents the scalar time, we assume that the unsteady component of the vector field $\boldsymbol{u}(\boldsymbol{\chi}, t)$ can be decomposed as

$$\boldsymbol{u}(\boldsymbol{\chi}, t) - \overline{\boldsymbol{u}}(\boldsymbol{\chi}) = \sum_i a_i^{\mathrm{POD}}(t)\boldsymbol{\phi}_i^{\mathrm{POD}}(\boldsymbol{\chi}), \tag{1}$$

where $\overline{\boldsymbol{u}}$ is the temporal mean of $\boldsymbol{u}$. In (1), $\boldsymbol{\phi}_i^{\mathrm{POD}}(\boldsymbol{\chi})$ and $a_i^{\mathrm{POD}}(t)$ represent the spatial modes and the time-dependent projection coefficients, respectively. In the POD framework, we seek the optimal set of basis functions for a given flow field data.

The POD method requires an ensemble of snapshots of any scalar (*e.g.* pressure, temperature) or vectorial (*e.g.* velocity, vorticity) fields defined at discrete spatial points $\boldsymbol{\chi}_i$ ($i = 1, \cdots, N_\chi$) and discrete time instants t_k ($k = 1, \cdots, N_t$). From this data set, we prepare snapshots of the flow field stacked in terms of a collection of column vectors $\boldsymbol{u}'(\boldsymbol{\chi}_i, t_k)$. We then consider a collection of finite-dimensional

data vectors that represent the flow field fluctuation

$$\boldsymbol{u}'(\boldsymbol{\chi}_i, t_k) = \boldsymbol{u}(\boldsymbol{\chi}_i, t_k) - \overline{\boldsymbol{u}}(\boldsymbol{\chi}_i) \in \mathbb{R}^{N_s}, \quad i = 1, 2, \ldots, N_\chi, \quad k = 1, 2, \ldots, N_t \tag{2}$$

where N_s is the number of the spatial degrees of freedom of the data and N_t is the number of snapshots. For a fluid flow data, N_s is equal to the number of grid points (N_χ) times the number of components (N_c) to be considered in the data. The snapshot vector can be stacked in matrix form as

$$\boldsymbol{X} = \left[\boldsymbol{u}'(\boldsymbol{\chi}_i, t_1)\ \boldsymbol{u}'(\boldsymbol{\chi}_i, t_2)\ \ldots \boldsymbol{u}'(\boldsymbol{\chi}_i, t_{N_t})\right] \in \mathbb{R}^{N_s \times N_t}. \tag{3}$$

To obtain the POD modes, we employ the *method of snapshots*. The objective is to find the optimal basis vectors that can best represent the flow data in an average sense. In other words, we seek the vectors $\boldsymbol{\phi}_i^{\mathrm{POD}}(\boldsymbol{\chi})$ in (1) that can represent $\boldsymbol{u}'(\boldsymbol{\chi}_i, t_k)$ in an optimal manner and with the least number of modes. The solution to this problem can be determined by finding the eigenvectors $\boldsymbol{\psi}_i$ and eigenvalues λ_i from

$$\boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\psi}_i = \lambda_i \boldsymbol{\psi}_i, \quad \boldsymbol{\psi}_i \in \mathbb{R}^{N_t} \tag{4}$$

where $\boldsymbol{X}^T\boldsymbol{X}$ is the temporal correlation matrix[1] of size $N_t \times N_t$.

In most numerical cases, spatial size of the data is very large compared to the temporal size, $N_s \gg N_t$. The method of snapshots thus uses a smaller correlation matrix than the spatial correlation matrix $\boldsymbol{X}\boldsymbol{X}^T$ of size $N_s \times N_s$ used in classical *spatial POD method*. It has been shown in literature that the temporal correlation matrix will yield the same dominant spatial modes while giving rise to a much smaller and computationally more tractable eigenvalue problem (Rowley and Dawson, 2017). The number of snapshots N_t in this case should be chosen such that the important fluctuations in the flow field are well resolved in time.

After obtaining the eigenvectors $\boldsymbol{\psi}_i$ from the smaller eigenvalue problem (4), the (spatial) POD modes can be recovered as

$$\boldsymbol{\phi}_i = \frac{1}{\sqrt{\lambda_i}} \boldsymbol{X} \boldsymbol{\psi}_i \in \mathbb{R}^{N_s}, \quad i = 1, 2, \ldots, N_t, \tag{5}$$

which can be also written as

$$\boldsymbol{\Phi} = \boldsymbol{X} \boldsymbol{\Psi} \boldsymbol{\Lambda}^{-1/2}, \tag{6}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2\ \ldots\ \boldsymbol{\phi}_{N_t}] \in \mathbb{R}^{N_s \times N_t}$ and $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1\ \boldsymbol{\psi}_2\ \ldots\ \boldsymbol{\psi}_{N_t}] \in \mathbb{R}^{N_t \times N_t}$. Note that these matrices are orthonormal, *i.e.* $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Phi}^T = \boldsymbol{I}$ and $\boldsymbol{\Psi}^T\boldsymbol{\Psi} = \boldsymbol{\Psi}\boldsymbol{\Psi}^T = \boldsymbol{I}$. The temporal coefficients are determined accordingly as

$$a_i(t) = \left\langle \boldsymbol{u}'(\boldsymbol{\chi}, t), \boldsymbol{\phi}_i(\boldsymbol{\chi}) \right\rangle, \quad i = 1, 2, \ldots, N_t, \tag{7}$$

where $\langle \cdot \rangle$ represents the spatial inner product defined as

$$\left\langle \boldsymbol{v}^{\mathrm{I}}(\boldsymbol{\chi}, t), \boldsymbol{v}^{\mathrm{II}}(\boldsymbol{\chi}, t) \right\rangle = \int_\Omega \boldsymbol{v}^{\mathrm{I}}(\boldsymbol{\chi}, t) \cdot \boldsymbol{v}^{\mathrm{II}}(\boldsymbol{\chi}, t)\, \mathrm{d}\boldsymbol{\chi} \tag{8}$$

for two given vector fields $\boldsymbol{v}^{\mathrm{I}}$ and $\boldsymbol{v}^{\mathrm{II}}$ defined in a spatial domain Ω.

[1] For simplicity, we have considered the field data to be placed on a uniform grid. In general, the cell volume needs to be included in the formulation to represent the inner product. The covariance matrix should therefore be written as $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$ where $\boldsymbol{W}$ holds the spatial weights.

2.2 POD Reduced-Order Model (POD ROM)

A POD ROM may be derived by the standard Galerkin projection of the Navier-Stokes equations onto the spatial POD modes $\boldsymbol{\phi}_i$. After some algebraic manipulation (Cordier et al., 2010), the following expression is obtained for the POD ROM

$$\dot{a}_i^{\mathrm{ROM}}(t) = \frac{da_i^{\mathrm{ROM}}(t)}{dt} = C_i + \sum_{j=1}^{N_{\mathrm{Gal}}} L_{ij} a_j^{\mathrm{ROM}}(t) + \sum_{j=1}^{N_{\mathrm{Gal}}} \sum_{k=j}^{N_{\mathrm{Gal}}} Q_{ijk} a_j^{\mathrm{ROM}}(t) a_k^{\mathrm{ROM}}(t), \tag{9}$$

where the superscript ROM is used to separately identify the temporal coefficients obtained from the POD ROM as compared to the ones ($a_i^{\mathrm{POD}}(t)$) which are directly obtained from snapshot POD. The parameters C_i, L_{ij} and Q_{ijk} are the constant, linear and quadratic coefficients which depend exclusively on the spatial POD modes $\boldsymbol{\phi}_i$ and mean field $\overline{\boldsymbol{u}}$. As such, their values can be directly determined (see Cordier et al., 2010). For the initial conditions, we have

$$a_i^{\mathrm{ROM}}(0) = a_i^{\mathrm{POD}}(0) = \langle \boldsymbol{u}(\boldsymbol{\chi}, 0) - \overline{\boldsymbol{u}}(\boldsymbol{\chi}), \boldsymbol{\phi}_i(\boldsymbol{\chi}) \rangle \,. \tag{10}$$

The number of model parameters C_i, L_{ij} and Q_{ijk} is given as $N_b = N_0 + N_1 + N_2$, where

- $N_0 = 1$ for the constant terms C_i,
- $N_1 = N_{\mathrm{Gal}}$ for the linear terms L_{ij}, and
- $N_2 = N_{\mathrm{Gal}}(N_{\mathrm{Gal}} + 1)/2$ for the quadratic terms Q_{ijk}.

From the knowledge of N_t samples of the quantities $\dot{a}_i^{\mathrm{ROM}}(t_k)$ and $a_i^{\mathrm{ROM}}(t_k)$ at discrete times t_k, the dynamical parameters can be determined by a least mean-square estimation procedure. This problem leads to the minimisation of an error function $\varepsilon = \|\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{B}\|_2$ (Perret et al., 2006), with

- $\boldsymbol{A} \in \mathbb{R}^{N_t \times N_b}$ as the matrix of the terms 1, $a_j^{\mathrm{ROM}}(t_k)$ and $a_j^{\mathrm{ROM}}(t_k) a_k^{\mathrm{ROM}}(t_k)$,
- $\boldsymbol{\theta} \in \mathbb{R}^{N_b \times 1}$ as the vector of unknown parameters, and
- $\boldsymbol{B} \in \mathbb{R}^{N_t \times 1}$ containing the N_t values of $\dot{a}_i^{\mathrm{ROM}}(t_k)$.

The estimation problem is solved by forming a linear system $(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{\theta} = \boldsymbol{A}^T\boldsymbol{B}$ and using SVD of the matrix $\boldsymbol{A}$ to find the pseudo-inverse of $\boldsymbol{A}^T\boldsymbol{A}$ and solve for the parameters in $\boldsymbol{\theta}$.

However, it is well known that for various reasons (structural instability of the Galerkin projection, truncation of the POD basis, inaccurate treatment of the boundary and pressure terms), the dynamical system does not represent sufficiently well the correct dynamics. This problem is now tackled by data assimilation.

2.3 Dual Ensemble Kalman Filter (Dual EnKF)

As the POD ROM determined by Galerkin projection is not sufficient in general to replicate the correct dynamics, we seek data assimilation methods to correctly identify the model parameters. In the following, we briefly present the specific DA method known as Dual Ensemble Kalman Filter which has been used in the current study and then discuss its applicability to the POD ROM.

Data assimilation is a generic methodology which combines heterogeneous observations with the underlying dynamical principles governing the system under consideration to estimate at best the states and/or physical quantities parameterizing the dynamics. Starting from a background solution and incoming imperfect information, an optimal estimation of the unknowns of the system is determined which takes into account the respective statistical confidences of the different observations.

We consider a non-linear, time-discrete dynamic system given as

$$\boldsymbol{x}_{k+1}^t = \mathcal{M}_{k,k+1}(\boldsymbol{x}_k^t, \boldsymbol{\theta}_k) + \boldsymbol{\eta}_k, \tag{11}$$

where the superscript t stands for *true*. In (11), $k = 0, \ldots, N-1$ is the time index and N is the total number of time steps. In this system, the propagation of the true state $\boldsymbol{x}_k^t \in \mathbb{R}^{N_s}$ (where N_s is the dimension of the system of the state after spatial discretisation) from time t_k to t_{k+1} is performed by a non-linear function $\mathcal{M}_{k,k+1}$. The propagator depends on the vector of model parameters $\boldsymbol{\theta}_k$ at time t_k. The model error $\boldsymbol{\eta}_k$ is considered to be Gaussian distributed with zero mean and covariance matrix $\boldsymbol{Q}_k$ *i.e.*

$$\boldsymbol{\eta}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k). \tag{12}$$

The observation vector $\boldsymbol{y}_{k+1}^o \in \mathbb{R}^{N_o}$ (where N_o is the dimension of the observations) at time t_k is given by

$$\boldsymbol{y}_{k+1}^o = \mathcal{H}_{k+1}(\boldsymbol{x}_{k+1}^t) + \boldsymbol{\epsilon}_{k+1}, \tag{13}$$

where the observation function $\mathcal{H}_{k+1}$ is also non-linear. The observation noise vector $\boldsymbol{\epsilon}_{k+1}$ is constructed by taking all the sources of observation error into consideration. It is approximated as a Gaussian distribution with zero mean and covariance $\boldsymbol{R}_{k+1}$ *i.e.*

$$\boldsymbol{\epsilon}_{k+1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{k+1}). \tag{14}$$

If the model parameters are known to be true, a DA method is specifically used for the state estimation. On the other hand, in practice, we frequently encounter cases where the parameters are unknown or imprecise.

One of the approaches to solve this dual estimation problem is to use an iterative method where the filtering is alternatively applied to estimate the state and parameters in the course of each iteration. More precisely, the model parameters are obtained from the corrected state at the previous time state and the new state is evaluated from the current analysed parameters. This method is formally known as Dual EnKF (Moradkhani et al., 2005).

The mechanism of evolution of the Kalman filter is based on a prediction-correction scheme. First, an ensemble of N_e forecasted parameters is created using a forced random walk of the parameters set up according to a Gaussian law $\boldsymbol{\xi}_k$ with zero mean and covariance matrix $\boldsymbol{C}_k$ *i.e.*

$$\boldsymbol{\xi}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C}_k). \tag{15}$$

During the prediction step, the ensemble of forecasted parameters is constructed as

$$\boldsymbol{\theta}_{k+1}^{f,(n)} = \boldsymbol{\theta}_k^{a,(n)} + \boldsymbol{\xi}_k^{(n)}, \tag{16}$$

where the superscript (n) represents the n-th member of the ensemble of parameters. In (16), the su-

perscripts f and a denote the forecasted and assimilated values of the parameters, respectively. The predicted ensemble of parameters are then updated using the available observations. For this, the predicted state at the instant t_{k+1} is built starting from the predicted parameters at the instant t_{k+1} and the corrected state at the instant t_k *i.e.*

$$\boldsymbol{x}_{k+1}^{f,(n)} = \mathcal{M}_{k,k+1}(\boldsymbol{x}_k^{a,(n)}, \boldsymbol{\theta}_{k+1}^{f,(n)}). \tag{17}$$

The predicted ensemble of observations is then obtained as

$$\boldsymbol{y}_{k+1}^{f,(n)} = \mathcal{H}_{k+1}(\boldsymbol{x}_{k+1}^{f,(n)}). \tag{18}$$

Next, the random observations are built to be centered around the actual observation $\boldsymbol{y}_{k+1}^{o}$ following

$$\boldsymbol{y}_{k+1}^{o,(n)} = \boldsymbol{y}_{k+1}^{o} + \boldsymbol{\epsilon}_{k+1}^{o,(n)}, \tag{19}$$

where the covariance matrix of the observation error at the instant t_{k+1} is given as

$$\boldsymbol{R}_{k+1}^{e} = \frac{1}{N_e - 1} \sum_{n=1}^{N_e} (\boldsymbol{\epsilon}_{k+1}^{o,(n)})(\boldsymbol{\epsilon}_{k+1}^{o,(n)})^T. \tag{20}$$

As the number of Monte Carlo elements tends to infinity, the empirical estimator $\boldsymbol{R}_{k+1}^{e}$ tends to the covariance matrix of the observation error $\boldsymbol{R}_{k+1}$.

Given the random observations from (19) and the empirical estimator (20), the update of the predicted ensemble of parameters is realised by the Kalman analysis, given as

$$\boldsymbol{\theta}_{k+1}^{a,(n)} = \boldsymbol{\theta}_{k+1}^{f,(n)} + \boldsymbol{K}_{k+1}^{\theta,e}(\boldsymbol{y}_{k+1}^{o,(n)} - \boldsymbol{y}_{k+1}^{f,(n)}). \tag{21}$$

The Kalman gain allowing the correction of the parameter trajectories is expressed using the ensemble predictions as

$$\boldsymbol{K}_{k+1}^{\theta,e} = \boldsymbol{P}_{\theta y,k+1}^{f,e}(\boldsymbol{P}_{yy,k+1}^{f,e} + \boldsymbol{R}_{k+1}^{e})^{-1}. \tag{22}$$

In this expression, the terms $\boldsymbol{P}_{\theta y,k+1}^{f,e}$ and $\boldsymbol{P}_{yy,k+1}^{f,e}$ designate at the instant t_{k+1} the unbiased empirical estimators of the cross-correlation matrix between the parameters and the observations, and the correlations between the observations, respectively.

Once the predicted parameters have been updated, the second filter is implemented to estimate the state of the system at the instant t_{k+1} by assuming the parameters as known. In other words, the parameters just corrected by the first filter are trusted. The following prediction-correction mechanism is identical to the classical ensemble Kalman filter. In the first step, the predicted ensemble of states at the instant t_{k+1} is built starting from the ensemble of assimilated states at the instant t_k and the corrected ensemble of parameters at the instant t_{k+1} *i.e.*

$$\boldsymbol{x}_{k+1}^{f,(n)} = \mathcal{M}_{k,k+1}(\boldsymbol{x}_k^{a,(n)}, \boldsymbol{\theta}_{k+1}^{a,(n)}) + \boldsymbol{\eta}_k^{(n)}. \tag{23}$$

For the second time, an ensemble of observations is calculated as

$$\boldsymbol{y}_{k+1}^{f,(n)} = \mathcal{H}_{k+1}(\boldsymbol{x}_{k+1}^{f,(n)}). \tag{24}$$

At this stage, the observations are again considered as random variables and the covariance matrix of the observation error involved in calculating the Kalman gain is also empirically estimated. Next, the update of the predicted ensemble of the state is realised by Kalman analysis, given as

$$\boldsymbol{x}_{k+1}^{a,(n)} = \boldsymbol{x}_{k+1}^{f,(n)} + \boldsymbol{K}_{k+1}^{e}(\boldsymbol{y}_{k+1}^{o,(n)} - \boldsymbol{y}_{k+1}^{f,(n)}). \tag{25}$$

The use of a Kalman gain permits the correction of the state trajectories as

$$\boldsymbol{K}_{k+1}^{e} = \boldsymbol{P}_{xy,k+1}^{f,e}(\boldsymbol{P}_{yy,k+1}^{f,e} + \boldsymbol{R}_{k+1}^{e})^{-1}. \tag{26}$$

In this expression, $\boldsymbol{P}_{xy,k+1}^{f,e}$ and $\boldsymbol{P}_{yy,k+1}^{f,e}$ designate at the instant t_{k+1} the unbiased empirical estimators of the cross-correlation matrix between the state and the observations, and the correlation matrix between the observations, respectively. The prediction-correction step is repeated by using the analysed state at t_{k+1} from (25) and new parameter forecast at t_{k+2} from (16) to build the new state forecast in (17). The phase of prediction-correction of the system state can precede the operation on the parameters without consequence on the quality of the estimation.

Now, we can return to the POD ROM and present the use of the Dual EnKF method to correctly identify the model parameters that will help to replicate the correct dynamics. In the DA problem, the temporal coefficients $a_i^{\mathrm{ROM}}(t_k)$ can be considered to form the state vector $\boldsymbol{x}_k$. The POD ROM in (10) serves as the dynamical model $\mathcal{M}_{k,k+1}$ and the POD coefficients $a_i^{\mathrm{POD}}(t_k)$ form the observations vector $\boldsymbol{y}_k^o$. Finally, the coefficients C_i, L_{ij} and Q_{ijk} form the parameters $\boldsymbol{\theta}$. The initial background conditions $\boldsymbol{x}_0^b$ and $\boldsymbol{\theta}^b$ are determined by using the temporal coefficients directly obtained from POD at the initial condition, and from the parameter values identified in § 2.2 for the POD ROM.

3 POD ROM of a 2D-cylinder wake flow

In order to test the data assimilation framework in a simple numerical configuration, we choose in this paper to estimate the dynamics of a 2D-cylinder wake flow. The snapshots necessary to determine the POD modes (see § 2.1) are obtained from a numerical simulation performed using the finite-element solver *FreeFem++* (Hecht, 2012). A POD reduced-order model is then built using the procedure described in § 2.2. This POD ROM serves as the dynamical model used for the application of the Dual EnKF method presented in § 2.3.

The computational domain along with the dimensions and boundary conditions is shown in Fig. 1. All the dimensions have been parameterised with the diameter of the cylinder D, here set to one. For the purpose of mesh size control, we divide the domain into sub-domains and use the automatic mesh generation offered by *FreeFem++*'s `buildmesh` command (Hecht, 2012). The mesh has $N_{\mathrm{vert}} = 10263$ vertices and $N_{\mathrm{tri}} = 20360$ triangles.

To solve the space discretised Navier-Stokes equations, an optimised Newton method is used. This approach is a variant of the classical Newton method for which the non-linear term is discretised semi-implicitly. The finite element space for velocity and pressure variables are discretised using $P2$ and $P1$ elements, respectively. After convergence of the iterative procedure, the pressure field is such that the resulting velocity is divergence free.

The Reynolds number of the numerical simulation based on the cylinder diameter D is set to $Re = 100$.

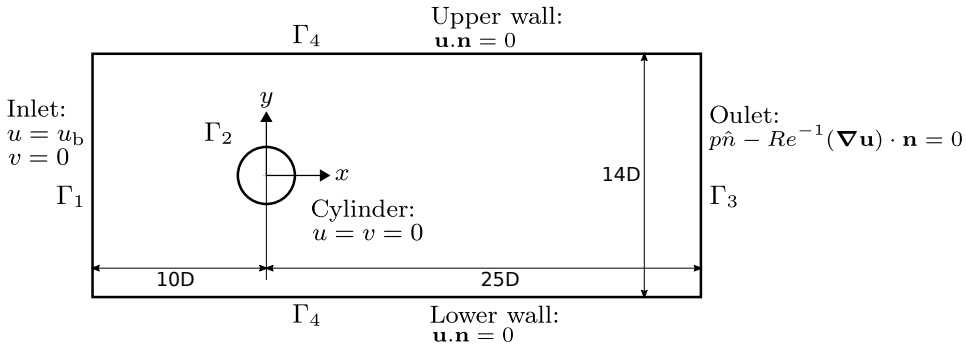


Figure 1: Computational domain and boundary conditions for the 2D-cylinder wake flow.

The snapshots are taken in a time range of $[150, 250]$ s after all initial disturbances have been canceled. The simulation generates $N_t = 1000$ snapshots at a sampling frequency of $f_s = 10$ Hz. The snapshots contain the information of the two components of the fluctuating velocity vector $\boldsymbol{u}' = [u', v']^T$. The degree of freedom (number of discretised state variables) for the problem is $N_s = 81772$. The time evolution of the drag and lift coefficients are shown in Fig. 2.

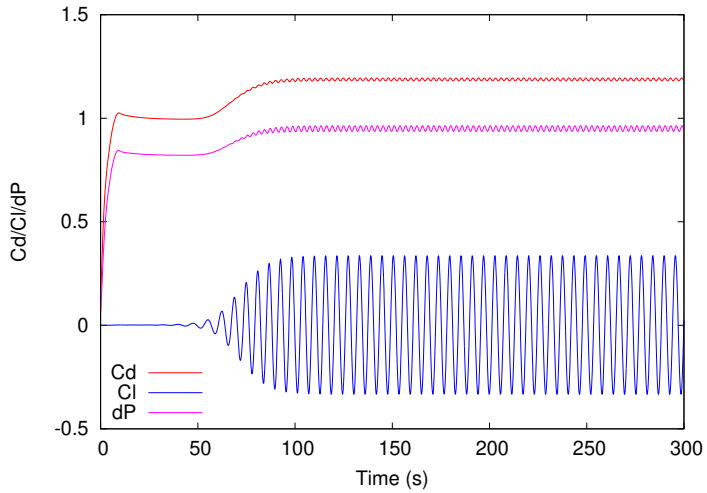


Figure 2: Time evolution of drag coefficient (C_d), lift coefficient (C_l), and pressure drop across the cylinder's leading and trailing surfaces (dP) for the $2D$-cylinder flow at $Re = 100$. The periodic vortex shedding is observed from $t \simeq 100$ s.

The POD modes are then calculated for the fluctuating velocity fields obtained from the snapshot database. The energy content of the modes is plotted in Fig. 3. We observe that the first four modes in total represent 99% of the energy. Therefore, for the further evaluation of the ROM, we reduce the degrees of freedom of the problem to $N_{\text{Gal}} = 4$, where $N_{\text{Gal}} \ll N_t$.

The four most dominant temporal and spatial modes are plotted in Fig. 4 and Fig. 5, respectively. The temporal coefficients a_i^{POD} show a sinusoidal response with the amplitude for the first pair (a_1^{POD} and a_2^{POD}), greater by an order of magnitude when compared with the next pair of modes (a_3^{POD} and a_4^{POD}). The frequency of the dynamics of the first pair of modes is half of that of the second pair. The spatial mode pairs are shifted spatially as a result of the convective nature of the flow. The first spatial mode pair (ϕ_1^{POD} and ϕ_2^{POD}) jointly depicts the dynamical vortex shedding and their downstream convection. The next mode pair (ϕ_3^{POD} and ϕ_4^{POD}) corresponds to smaller scale structures attributed to the manifestation of the separated shear layers along the sides of the cylinder and their longitudinal expansion further

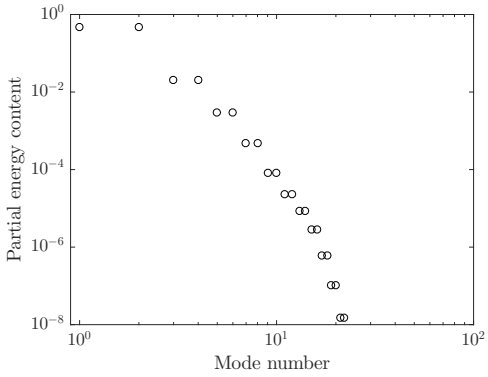
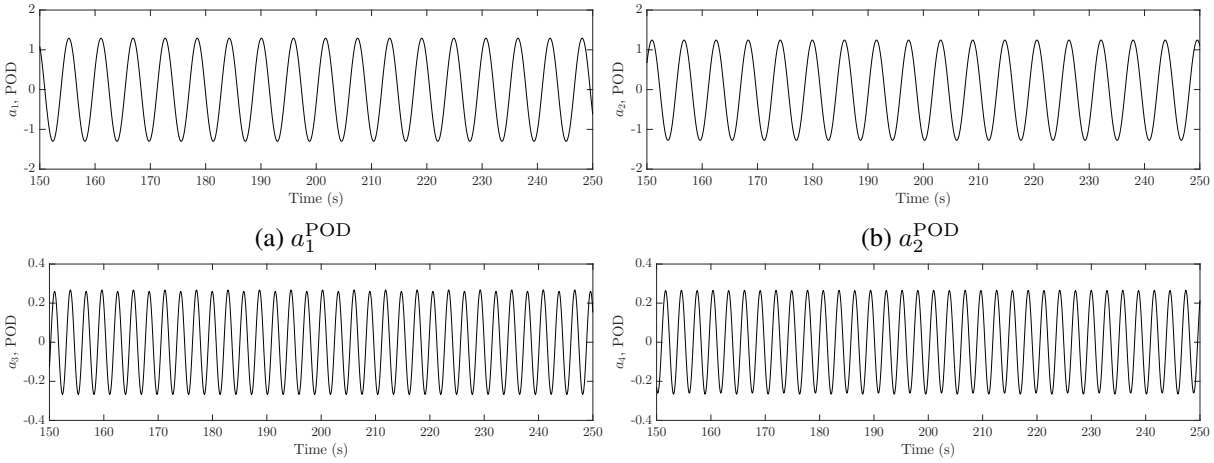


Figure 3: Energy content of the POD modes obtained for the fluctuating velocity fields of the $2D$-cylinder wake flow at $Re = 100$. The first four modes are the most dominant and account for 99% of the total energy.

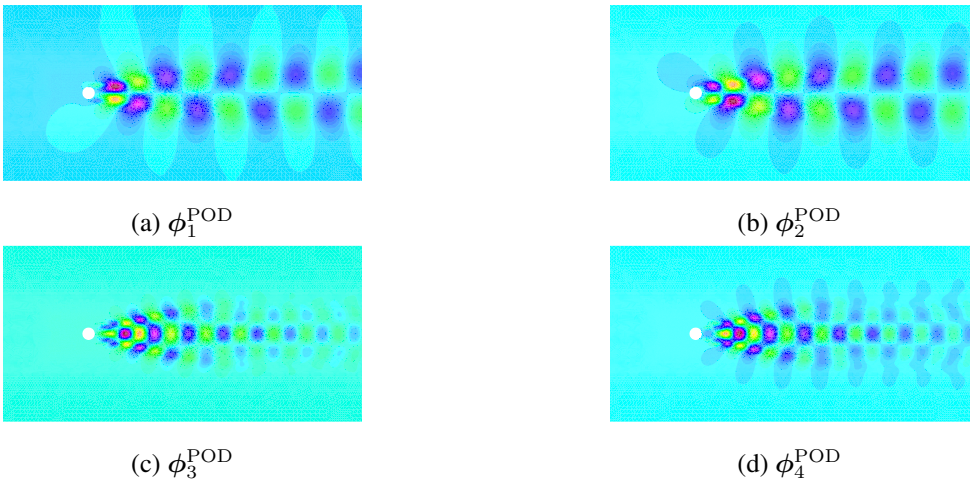
downstream.



(a) a_1^{POD} (b) a_2^{POD}

(c) a_3^{POD} (d) a_4^{POD}

Figure 4: Time evolution of the temporal POD coefficients a_i^{POD} corresponding to the four most energetic modes for the $2D$-cylinder wake flow at $Re = 100$.



(a) ϕ_1^{POD} (b) ϕ_2^{POD}

(c) ϕ_3^{POD} (d) ϕ_4^{POD}

Figure 5: Spatial POD modes ϕ_i^{POD} corresponding to the four most energetic modes for the $2D$-cylinder wake flow at $Re = 100$.

Finally, knowing the time evolution of the temporal POD modes, we can apply the identification method described in § 2.2. For the DA application (see § 4), the parameters of the ROM is identified using the first 50 s of the time range $[150, 200]$, the remaining time range $(200, 250]$ s being used for verification of the procedure. As we retained $N_{\mathrm{Gal}} = 4$ modes in the current case, we have $N_b = 15$ unknown parameters in the model for each mode. Moreover, as we sample the time span $[150, 200]$ s at a sampling frequency of 10 Hz, we have $N_t = 501$ for the identification. The parameters of the POD ROM obtained from the subsequent $L2$-minimisation are given in Table 1.

Table 1: Initial conditions and identified coefficients of the POD ROM for the $2D$-cylinder wake flow at $Re = 100$ ($i = 1, 2, 3, 4$).

$a_1^{\mathrm{ROM}}(0)$	$a_2^{\mathrm{ROM}}(0)$	$a_3^{\mathrm{ROM}}(0)$	$a_4^{\mathrm{ROM}}(0)$
1.08222	0.66946	-0.12609	-0.23122

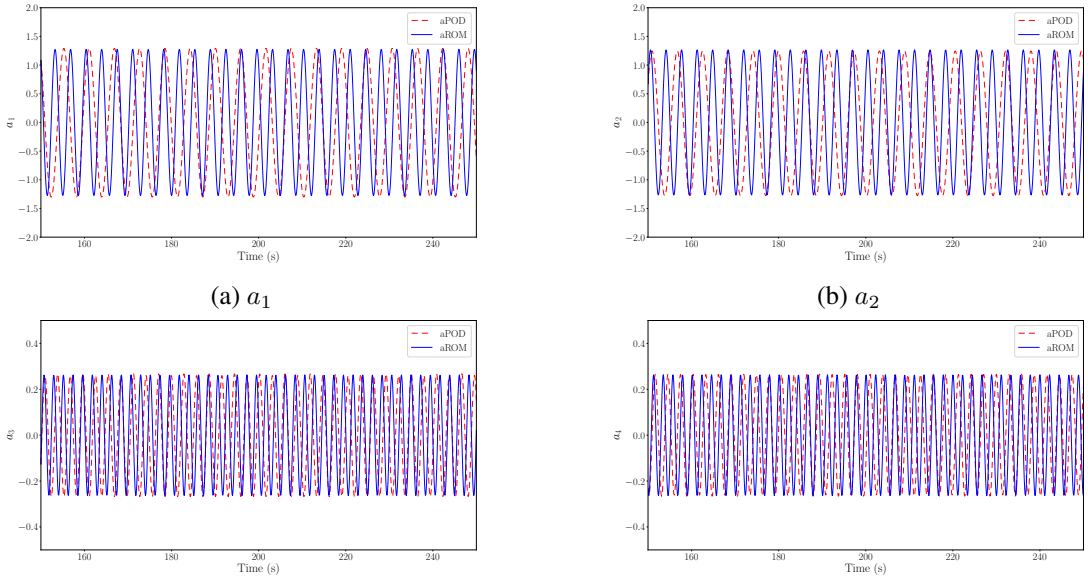
	$\dot{a}_1^{\mathrm{ROM}}$	$\dot{a}_2^{\mathrm{ROM}}$	$\dot{a}_3^{\mathrm{ROM}}$	$\dot{a}_4^{\mathrm{ROM}}$
C_i	0.362064	-0.098294	-0.000607	0.035313
L_{i1}	0.092654	1.749100	-0.083613	-0.224252
L_{i2}	-1.774020	0.063207	0.210161	-0.057925
L_{i3}	-0.283876	-0.197742	0.313415	2.837220
L_{i4}	0.246881	-0.256934	-2.829170	0.294689
Q_{i11}	-6.12E-07	0.001337	0.025321	-0.037705
Q_{i12}	-0.001339	-0.023584	0.080276	0.058309
Q_{i13}	-0.025362	-0.027676	0.014383	0.096379
Q_{i14}	0.037990	-0.022375	-0.093679	0.002203
Q_{i22}	0.023581	-1.03E-06	-0.031571	0.043928
Q_{i23}	-0.052789	0.031436	0.011520	-0.023409
Q_{i24}	-0.035259	-0.043554	0.032056	0.016921
Q_{i33}	-0.014382	-0.011518	4.85E-06	-0.000360
Q_{i34}	-0.002704	-0.008651	0.000320	-0.012959
Q_{i44}	-0.002201	-0.016919	0.013066	-7.09E-05

4 Results of Dual EnKF

As already mentioned in § 2.2, the POD ROM is in general not sufficiently accurate to describe the original dynamics followed by the POD modes. This is illustrated in Fig. 6 where the ROM built using the parameter values in Table 1 is employed to determine the time evolution of the POD coefficients. We observe that the true dynamics are not exactly captured by the identified ROM. This was the main motivation to use a DA method like the Dual EnKF to correct the values of the ROM parameters.

Before proceeding with the DA, we emphasize the fact that most of the physical systems have only a few non-linear terms in the dynamics, making the right hand side of the equations (9) *sparse* in a high-dimensional nonlinear function space. Hereafter, we exploit this sparsity in the non-linear model to balance the model complexity with accuracy. To introduce sparsity in our solution, we follow the procedure presented in Brunton et al. (2016) and introduce the concept of *sparsification knob*. The role of this quantity is to set the maximum value allowed for any parameter. By first running the data assimilation with no sparsification, a value of 0.5 has been found empirically for the sparsification knob. In this section, the covariances are set to diagonal matrices equal to $\boldsymbol{Q} = 10^{-4}\boldsymbol{I}$, $\boldsymbol{C} = 10^{-4}\boldsymbol{I}$ and $\boldsymbol{R} = 10^{-5}\boldsymbol{I}$. The parameters obtained from the Dual EnKF DA are given in Table 2. As anticipated, the parameter space is sparse as only four significant parameters are retained to build the ROM.

The estimation and forecast results of the Dual EnKF assimilation are shown in Fig. 7. The data is



(a) a_1 (b) a_2

(c) a_3 (d) a_4

Figure 6: True a_i^{POD} (dashed) and estimated a_i^{ROM} (solid) temporal POD modes ($i = 1, 2, 3, 4$) obtained with the identified POD ROM ($N_{\mathrm{Gal}} = 4$) for the $2D$-cylinder wake flow at $Re = 100$.

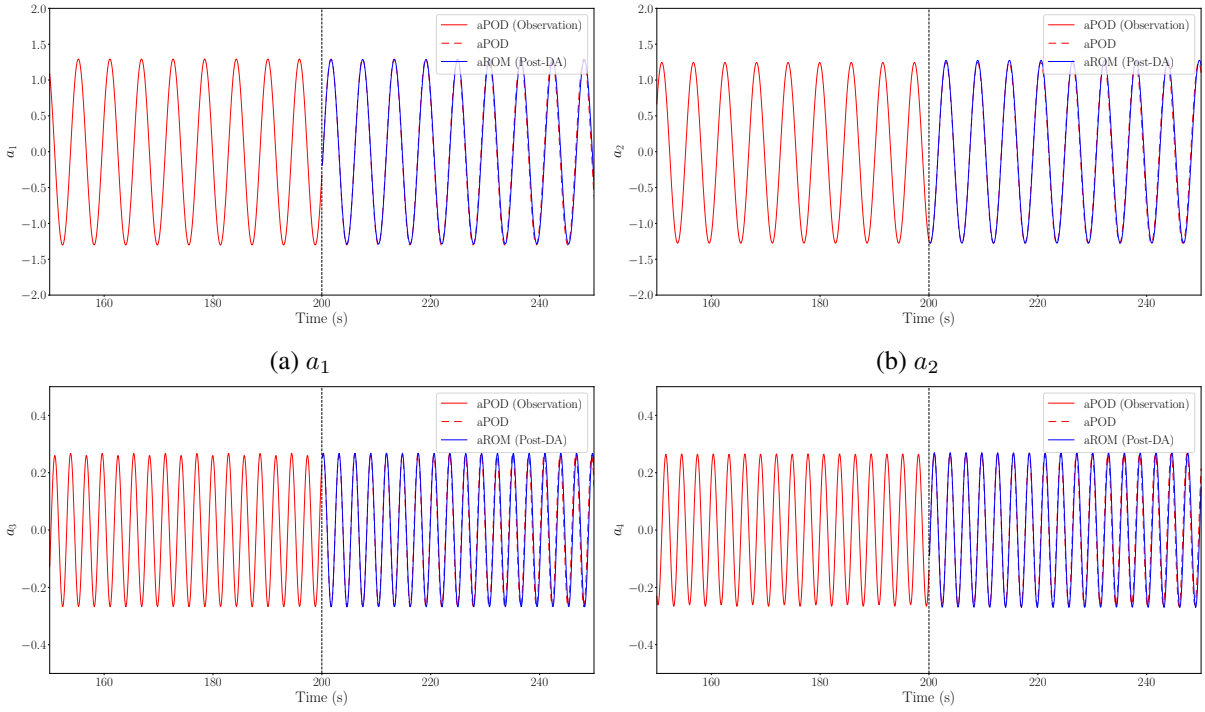
Table 2: Coefficients of the POD-ROM issued from Dual EnKF for the $2D$-cylinder wake flow at $Re = 100$ ($i = 1, 2, 3, 4$).

	$\dot{a}_1^{\mathrm{ROM}}$	$\dot{a}_2^{\mathrm{ROM}}$	$\dot{a}_3^{\mathrm{ROM}}$	$\dot{a}_4^{\mathrm{ROM}}$
C_i	0	0	0	0
L_{i1}	0	1.09517	0	0
L_{i2}	-1.06468	0	0	0
L_{i3}	0	0	0	2.16846
L_{i4}	0	0	-2.14971	0
Q_{i11}	0	0	0	0
Q_{i12}	0	0	0	0
Q_{i13}	0	0	0	0
Q_{i14}	0	0	0	0
Q_{i22}	0	0	0	0
Q_{i23}	0	0	0	0
Q_{i24}	0	0	0	0
Q_{i33}	0	0	0	0
Q_{i34}	0	0	0	0
Q_{i44}	0	0	0	0

assimilated for the time window $[150, 200]$ s and the parameters of the ROM are estimated. The ROM with the assimilated parameters is then used to forecast the evolution of the temporal coefficients in the time window $t = (200, 250]$ s. The results are compared with the reference POD coefficients. A good agreement between the forecast and the reference values is observed as a result of the corrected parameters of the POD ROM.

5 Conclusion

The Dual EnKF data assimilation is a well-established method in meteorology. It combines different inhomogeneous sources of information (data, dynamical model) for the dual estimation of the true state



(a) a_1 (b) a_2

(c) a_3 (d) a_4

Figure 7: Time evolution of the four most energetic temporal POD coefficients for the $2D$-cylinder wake flow at $Re = 100$. Comparison between the values obtained by POD and the ones obtained by the POD ROM with the parameters determined by Dual EnKF. The true values a_i^{POD} are used for $t \leq 200$ s as observation in the DA. The dotted line at $t = 200$ s indicates the end of the DA window and beginning of the forecast window.

and model parameters. This allows the prediction of the dynamical evolution with time. In this paper, this approach has been applied in a fluid mechanics context to a POD ROM of a $2D$-cylinder wake flow at low Reynolds number ($Re = 100$). We observe that the initial POD ROM obtained from identification is not sufficient to predict the long-term dynamics. Therefore, we apply the Dual EnKF algorithm to correct the model parameters and subsequently apply sparsification owing to the non-complex nature of the dynamics. We observe that the original dynamics was well reproduced and a good predictability of the analysed dynamical model was found. In the future, this DA method needs to be expanded with the inclusion of parameters related to control and applied to experimental flow cases. It is a work in progress.

References

S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *PNAS*, 113(15):3932–3937, 2016.

L. Cordier and G. Tissot. Model Reduction, POD and Data Assimilation. In *Lecture series 2014-01 on Advanced Post-Processing of Experimental and Numerical Data*, pages 1–89. Von Kármán Institute for Fluid Dynamics, 2013. ISBN 978-2-87516-061-4.

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

F. Hecht. New development in FreeFem++. *Journal of Numerical Mathematics*, 20(3-4):251–265, 2012.

J. L. Lumley. Atmospheric Turbulence and Wave Propagation. the structure of inhomogeneous turbulence. pages 166–178, 1967.

H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2):135 – 147, 2005.

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

L. Perret, E. Collin, and J. Delville. Polynomial identification of POD based low-order dynamical system. *Journal of Turbulence*, (7):N17, 2006.

C. W. Rowley and S. T. M. Dawson. Model Reduction for Flow Analysis and Control. *Annu. Rev. Fluid Mech.*, 49(1):387–417, 2017.

O. Talagrand and P. Courtier. Variational Assimilation of Meteorological Observations with the Adjoint Vorticity Equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478): 1311–1328, 1987.