Numerical Simulations with Data Assimilation Using an Adaptive POD Procedure

Gabriel Dimitriu[1], Narcisa Apreutesei[2] and Răzvan Ştefănescu[1]

[1] "Gr. T. Popa" University of Medicine and Pharmacy,
Department of Mathematics and Informatics,
dimitriu.gabriel@gmail.com, rastefanescu@yahoo.co.uk,
700115 Iaşi, Romania

[2] "Gh. Asachi" Technical University, Department of Mathematics,
napreut@gmail.com, 700506 Iaşi, Romania

Abstract. In this study the proper orthogonal decomposition (POD) methodology to model reduction is applied to construct a reduced-order control space for simple advection-diffusion equations. Several 4D-Var data assimilation experiments associated with these models are carried out in the reduced control space. Emphasis is laid on the performance evaluation of an adaptive POD procedure, with respect to the solution obtained with the classical 4D-Var (full model), and POD 4D-Var data assimilation. Despite some perturbation factors characterizing the model dynamics, the adaptive POD scheme presents better numerical robustness compared to the other methods, and provides accurate results.

1 Introduction

Data assimilation represents a methodology to combine the results of a large-scale numerical model with the measurements available to obtain an optimal reconstruction of the state of the system. The four-dimensional variational data assimilation (4D-Var) method has been a very successful technique used in operational numerical weather prediction at many weather forecast centers ([6], [8], [9], [10]).

Proper orthogonal decomposition technique has been used to obtain low dimensional dynamical models of many applications in engineering and science ([1], [2], [3], [4], [5], [7]). Basically, the idea starts with an ensemple of data, called *snapshots*, collected from an experiment or a numerical procedure of a physical system. The POD technique is then used to produce a set of basis functions which spans the snapshots collection. The goal of the approach is to represent the ensemble of data in terms of an optimal coordinate system. That is, the snapshots can be generated by a smallest possible set of basis functions.

The drawback of the POD 4D-Var consists of the fact that the optimal solution can only be sought within the space spanned by the POD basis of background fields. When observations lay outside of the POD space, the POD 4D-Var solution may fail to fit observations sufficiently. The above limitation of the POD 4D-Var can be improved by implementing an adaptive POD 4D-Var scheme. In this study an adaptive proper orthogonal decomposition (A POD) procedure is applied to set up a reduced-order control space for the one-dimensional diffusion-advection equations.

The outline of this paper is as follows. The numerical model under study is introduced in Section 2. Section 3 is dedicated to reviewing the 4D-Var assimilation procedure, together with POD 4D-Var method and its adaptive variant. Section 4 contains numerical results from data assimilation experiments using 4D-Var, POD 4D-Var and adaptive POD 4D-Var methods. The paper ends with some conclusions.

2 Numerical Model

Our model under study is a one-dimensional diffusion-advection equation defined by the following partial differential equation:

$$\frac{\partial \mathbf{c}}{\partial t} = a\frac{\partial^2 \mathbf{c}}{\partial x^2} + b\frac{\partial \mathbf{c}}{\partial x} + f(x,t)\,, \quad (x,t) \in \Omega \times (0,T), \quad \mathbf{c}(x,0) = x(x-1)\,. \tag{1}$$

Here, the spatial domain $\Omega = [0,1]$, and the coefficients a and b are positive constants. The function f is chosen to be $f(x,t) = (-x^2 + (1+2b)x + 2a - b)e^t$. Then, the exact (analytical) solution of (1) is given by $\mathbf{c}_{exact}(x,t) = x(1-x)e^t$. Details about the numerical implementation of a data assimilation algorithm for a similar model of (1) is presented in [10].

3 Adaptive POD 4D-Var Assimilation Procedure

For a successful implementation of the POD 4D-Var in data assimilation problems, it is of most importance to construct an accurate POD reduced model. In what follows, we briefly present a description of this procedure (see [2], [3], [7]).

For a temporal-spatial flow $\mathbf{c}(x,t)$, we denoted by $\mathbf{c}_1, \ldots, \mathbf{c}_n$ a set adequately chosen in a time interval $[0,T]$, that is $\mathbf{c}_i = \mathbf{c}(x,t_i)$. Defining the mean $\overline{\mathbf{c}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}_i$, we expand $\mathbf{c}(x,t)$ as

$$\mathbf{c}^{\mathrm{POD}}(x,t) = \overline{\mathbf{c}}(x) + \sum_{i=1}^{M} \beta_i(t)\Phi_i(x)\,, \tag{2}$$

where $\Phi_i(x)$ – the ith element of POD basis –, and M are appropriately chosen to capture the dynamics of the flow as follows:

1. Calculate the mean $\overline{\mathbf{c}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}_i$;
2. Set up the correlation matrix $K = [k_{ij}]$, where $k_{ij} = \int_{\Omega}(\mathbf{c}_i - \overline{\mathbf{c}})(\mathbf{c}_j - \overline{\mathbf{c}})\,dx$;
3. Compute the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and the corresponding orthogonal eigenvectors $v_1, v_2, \ldots, v_n$ of K;
4. Set $\Phi_i := \sum_{j=1}^{n} v_j^i(\mathbf{c}_i - \overline{\mathbf{c}})$.

Now, we introduce a relative information content to select a low-dimensional basis of size $M \ll n$, by neglecting modes corresponding to the small eigenvalues. Thus, we define the index

$$I(k) = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$

and choose M, such that $M = \operatorname{argmin}\{I(m) : I(m) \geq \gamma\}$, where $0 \leq \gamma \leq 1$ is the percentage of total information captured by the reduced space $X^M = \operatorname{span}\{\Phi_1, \Phi_2, \ldots, \Phi_M\}$. The tolerance parameter γ must be chosen to be near the unity in order to capture most of the energy of the snapshot basis. The reduced order model is then obtained by expanding the solution as in (2).

Generally, an atmospheric or oceanic model is usually governed by the following dynamic system

$$\frac{d\mathbf{c}}{dt} = F(\mathbf{c}, t)\,, \qquad \mathbf{c}(x,0) = \mathbf{c}^0(x)\,. \tag{3}$$

To obtain a reduced model of (3), we first solve (3) for a set of snapshots and follow above procedures, then use a Galerkin projection of the model equations onto the space X^M spanned by the POD basis elements (replacing $\mathbf{c}$ in (3) by the expansion (2), then multiplying Φ_i and integrating over spatial domain Ω):

$$\frac{d\beta_i}{dt} = < F\left(\overline{\mathbf{c}} + \sum_{i=1}^{M} \beta_i \Phi_i,\, t\right), \Phi_i >\,, \qquad \beta_i(0) = < \mathbf{c}(x,0) - \overline{\mathbf{c}}(x),\, \Phi_i(x) > \,. \tag{4}$$

Equation (4) defines a reduced model of (3). In the following, we will analyze applying this model reduction to 4D-Var formulation. In this context, the forward model and the adjoint model for computing the cost function and its gradient represent the reduced model and its corresponding adjoint, respectively.

At the assimilation time interval $[0, T]$, a prior estimate or 'background estimate', $\mathbf{c}_b$ of the initial state $\mathbf{c}_0$ is assumed to be known and the initial random errors $(\mathbf{c}_0 - \mathbf{c}_b)$ are assumed to be Gaussian with covariance matrix $\mathbf{B}$.

The aim of the data asimilation is to minimize the square error between the model predictions and the observed system states, weighted by the inverse of the covariance matrices, over the assimilation interval. The initial state $\mathbf{c}_0$ is treated as the required control variable in the optimization process. Thus, the objective function associated with the data assimilation for (3) is expressed by

$$J(\mathbf{c}_0) = (\mathbf{c}_0 - \mathbf{c}_b)^T \mathbf{B}^{-1} (\mathbf{c}_0 - \mathbf{c}_b) + (\mathbf{Hc} - \mathbf{y}^{\mathrm{o}})^T \mathbf{R}^{-1} (\mathbf{Hc} - \mathbf{y}^{\mathrm{o}})\,. \tag{5}$$

Here, $\mathbf{H}$ is an observation operator, and $\mathbf{R}$ is the observation error covariance matrix.

In POD 4D-Var, we look for an optimal solution of (5) to minimize the cost function $J(\mathbf{c}_0^M) = J(\beta_1(0), \ldots, \beta_M(0))$ given by

$$J(\mathbf{c}_0^M) = (\mathbf{c}_0^{\mathrm{POD}} - \mathbf{c}_b) \mathbf{B}^{-1} (\mathbf{c}_0^{\mathrm{POD}} - \mathbf{c}_b) + (\mathbf{Hc}^{\mathrm{POD}} - \mathbf{y}^{\mathrm{o}}) \mathbf{R}^{-1} (\mathbf{Hc}^{\mathrm{POD}} - \mathbf{y}^{\mathrm{o}})\,, \tag{6}$$

where $\mathbf{c}_0^{\mathrm{POD}}$ is the control vector.

In (6), $\mathbf{c}_0^{\mathrm{POD}}(x) = \mathbf{c}_0^{\mathrm{POD}}(x,0)$ and $\mathbf{c}^{\mathrm{POD}}(x) = \mathbf{c}^{\mathrm{POD}}(x,t)$ are expressed by

$$\mathbf{c}_0^{\mathrm{POD}}(x) = \overline{\mathbf{c}}(x) + \sum_{i=1}^{M} \beta_i(0)\Phi_i(x), \qquad \mathbf{c}^{\mathrm{POD}}(x) = \overline{\mathbf{c}}(x) + \sum_{i=1}^{M} \beta_i(t)\Phi_i(x)\,.$$

Therefore, in POD 4D-Var the control variables are $\beta_1(0), \ldots, \beta_M(0)$. As explained later, the dimension of the POD reduced space could be much smaller than that the original space. As a consequence, the forward model is the reduced model (4) which can be very efficiently solved. The adjoint model of (4) is then used to calculate the gradient of the cost function (6) and that will significantly reduce both the computational cost and the programming effort.

The POD model in POD 4D-Var assimilation is established by construction of a set of snapshots, which is taken from the background trajectory, or integrate original model (3) with background initial conditions. The `A-POD` algorithm used in our numerical experiments is presented below:

`A-POD` Algorithm:

`Step 1`: Set $k = 1$, the iteration level for POD procedure, and the initial guess controls $\mathbf{c}_0^k$.

`Step 2`: Set up the snapshots ensemble from the solution of the full forward model, with the controls $\mathbf{c}_0^k$.

`Step 3`: Compute the POD bases (the number of POD bases is chosen to capture a prescribed energy level, γ, mentioned above).

`Step 4`: Project the controls $\mathbf{c}_0^k$ on the reduced space $\beta_{k,iter}$ ($iter = 1$).

`Step 5`: Optimize the initial controls $\beta_{k,iter}$ (here, $iter$ denotes the iteration of the optimization process, completely carried out on the reduced space).

`Step 6`: (i) Check the value of the cost function (5). If $|J_{iter}| < \varepsilon$ (where ε is the tolerance for the optimization), then `STOP`;

(ii) If $|J_{iter}| > \varepsilon$ and $|J_{iter} - J_{iter-1}| > 10^{-3}$, then set $iter = iter + 1$ and go to `Step 5`;

(iii) If $|J_{iter}| > \varepsilon$ and $|J_{iter} - J_{iter-1}| < 10^{-3}$, then update the POD bases:

(a) Find the new controls $\mathbf{c}_0^{k+1}$ by projecting the optimization controls $\beta_{k,iter}$ onto the original domain.

(b) Set $k = k + 1$ and go to `Step 2`.

4 Numerical results

This section is devoted to the numerical results of several data assimilation experiments carried out to examine the performance of A POD 4D-Var procedure, by comparing it with the full 4D-Var and POD 4D-Var data assimilation. All the performed tests have used as 'true' (exact) solution $\mathbf{c}_0$ of the assimilation problem, that one computed from the analytical solution $\mathbf{c}_{exact}$ of (1) given in Section 2.

We set in our approach $T = 1$, and used 31 discretization points in space, and 69 points in time interval [0, 1]. By means of a perturbed initial condition

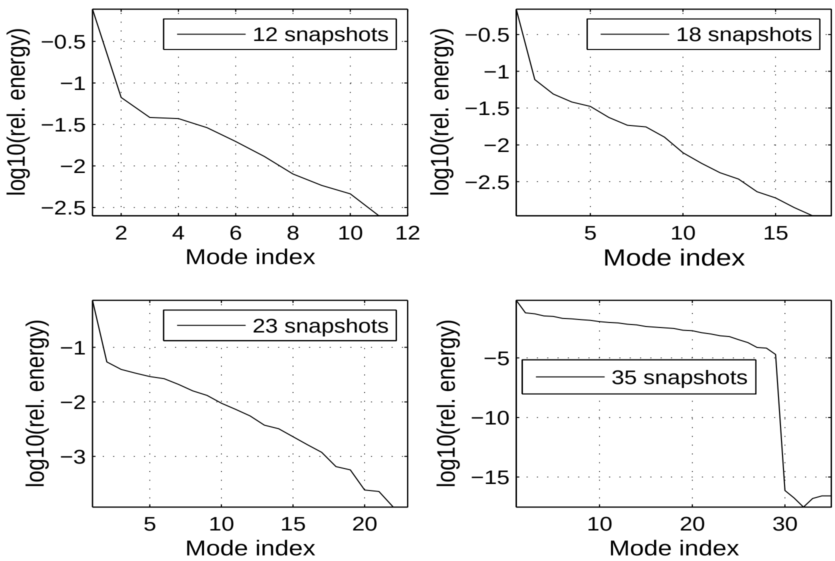

Fig. 1. Relative energy spectrum for POD modes, corresponding to different sets of snapshots for the system state.

we generated the observed state $\mathbf{y}^o$. We assumed that the correlation matrices in (5) and (6) are diagonal matrices, given by $\mathbf{B} = \sigma_b^2 I$ and $\mathbf{R} = \sigma_o^2 I$, with I denoting the identity matrix of appropriate order. We set $\sigma_b^2 = 0.05^2$ and $\sigma_o^2 = 0.1^2$, representing the variances for the background and observational errors, respectively.

The numerical tests have been performed with the following values of the diffusion and advection parameters: $a = 0.0001$ and $b = 1$. The background value $\mathbf{c}_b$ and the observation $\mathbf{y}^o$ were obtained by perturbing the exact solution of the state equation.

Figure 1 illustrates the decay of the eigenvalues, in case when the spatial correlation matrix, K, was calculated using 12, 18, 23 and 35 snapshots, respectively. The dimension of the POD reduced model depends on the number of basis functions. We found that only few basis functions are required to capture a high percentage from the dynamic of the system. For the presented cases, we can remark in Table 1 that more than 90% from the full variability characterizing the model dynamics can be captured with 4, 16 and 18 POD modes, respectively. Thus, choosing $M = 16$ modes and $n = 36$ snapshots, the captured energy represents 98.53% of the total energy, while when $M = 18$, the captured energy percentage is 99.16%.

We performed several numerical tests by implementing the three distinct procedures for data assimilation problem (1), (5): the standard (full) 4D-var, the

Table 1. The values of the index $I(k)$ for different numbers of snapshots (n), and POD modes (M).

Number of snapshots (n)	Number of modes (M)	Index $I(M)$
12	4	91.74%
18	16	99.89%
36	16	98.53%
36	18	99.16%

POD 4D-Var, and the adaptive POD (A POD) scheme. In all cases, the performance of POD scheme with respect to A POD was evaluated by computing RMS errors of the assimilation solution (see Table 2). The numerical solution of the optimal control problem was obtained using *fminunc* – the Matlab unconstrained minimization routine. Its algorithm is based on the BFGS quasi-Newton method with a mixed quadratic and cubic line search procedure. The decreasing values of the cost function J as well as of the its gradient $\frac{\partial J}{\partial \mathbf{c}_0}$ normalized to their values at the first iteration is presented in Figure 2.

Table 2. The values of the RMS errors (RMSE) calculated for POD and A POD procedures, corresponding to different numbers of snapshots and POD modes.

Number of snapshots (n)	Number of modes (M)	RMSE for POD	RMSE for A POD
12	4	0.1193	0.0874
12	12	0.073	0.0532
18	16	0.0684	0.0476
36	16	0.064	0.042
23	22	0.0444	0.0403
36	22	0.0427	0.0351

The comparative assimilation results associated with our experiments for certain selections of M and n are depicted in Figures 3, 4 and 5. Obviously, one can conclude that better results are obtained when one is using more POD modes and more snapshots.

We also remark that the variance of the observational errors chosen twice bigger than the variance of the background errors caused a relative high instability for the assimilation solution (even for the best case, where the reduced model was constructed with $n = 36$ snapshots and $M = 22$ modes, see Figure 5, bottom plot). A certain contribution to this oscillatory behaviour of the solution could also be attributed to the fact the advection strongly dominates diffusion in our model.

5 Conclusions

In this study, we applied to a simple diffusion-advection model, both a standard 4D-Var assimilation scheme and a reduced order approach to 4D-Var assimilation using an adaptive POD procedure. The numerical results from several POD models (constructed with different numbers of snapshots and POD modes) are compared with those of the original model. Our numerical tests showed that variability of the original model could be captured reasonably well by applying an adaptive POD scheme to a low dimensional system set up with 36 snapshots and 22 leading POD basis functions.

Several issues, including more general models (containing distributed diffusion and advection parameters, and also nonlinear source terms, directly depending on state variable, **c**) are under investigation.

Acknowledgement. The paper is supported by the project ID 342/2009, CNCSIS, Romania.

References

1. Antoulas, A., Sorensen, S., Gallivan, K.A.: Model Reduction of Large-Scale Dynamical Systems. *Lecture Notes in Computer Science*, 3038 (2004), 740–747.
2. Cao, Y., Zhu, J., Luo, Z., Navon, I.M.: Reduced Order Modeling of the Upper Tropical Pacific Ocean Model Using Proper Orthogonal Decomposition. Accepted for publication in *Computer & Mathematics with Applications* (2006).
3. Cao, Y.H., Zhu, J., Navon, I.M.: A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *Int. J. Numer. Meth. Fluids* 53(10) (2007), 1571–1583.
4. Dimitriu, G.: Using Singular Value Decomposition in Conjunction with Data Assimilation Procedures. LNCS 4310 (2007), 435–442.
5. Fang, F., Pain, C.C., Navon, I.M., Piggot, M.D., Gorman, G.J., Allison, P.A., Goddard, A.J.H.: Reduced-order modelling of an adaptive mesh ocean model. *Int. J. Numer. Meth. Fluids* 59 (2009), 827–851.
6. Ghil, M., Malanotte-Rizzoli, P.: In *Data Assimilation in Meteorology and Oceanography. Advances in Geophysics*. R. Dunowska, B. Saltzman (eds.), 33 (1991), 141–266.
7. Luo, Z., Zhu, J., Wang, R., Navon, I.M.: Proper Orthogonal Decomposition Approach and Error Estimation of Mixed Finite Element Methods for the Tropical Pacific Ocean Reduced Gravity Model. Submitted to *Computer Methods in Applied Mechanics and Engineering.*
8. Qui, C.J., Zhang, L., Shao, A.M.: An Expicit four-dimensional variational data assimilation method. *Sci. China Ser. D-Earth Sci.*, 50(8) (2007), 1232–1240.
9. Thépaut, J.N., Hoffman R.N., Courtier, P.: Interactions of dynamics and observations in a 4D variational assimilation. *Mon. Weather Rev.*, 121 (1993), 3393–3414.
10. Zlatev, Z., Brandth, J., Havasi, À.: Implementation issues related to variational data assimilation: some preliminary results and conclusions, *Working Group on Matrix Computations and Statistics*, Copenhagen, Denmark, April 1-3, (2005).

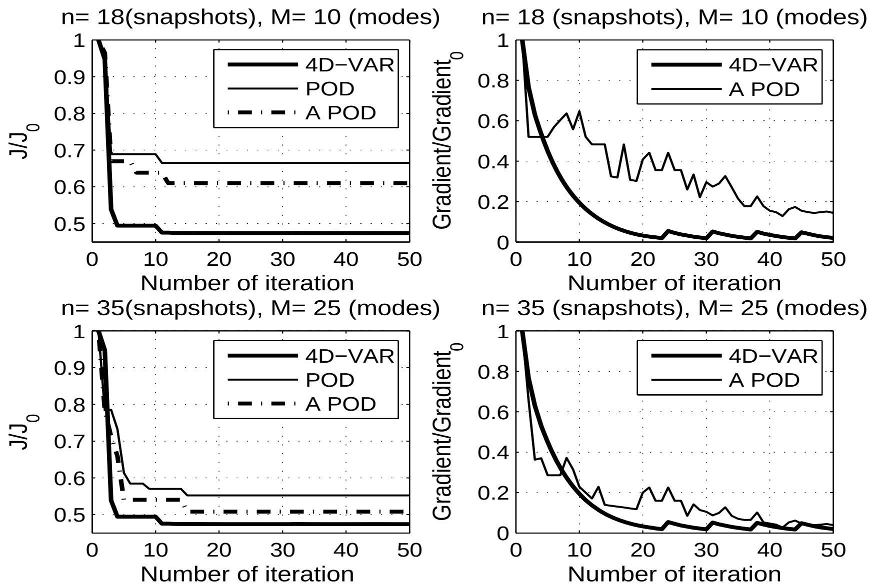

Fig. 2. The decreasing of the cost function J and of the gradient $\frac{\partial J}{\partial \mathbf{c}_0}$, for different values of snapshots and POD modes, compared to the 4D-Var method with full order model and reduced order model using POD 4D-Var and adaptive POD 4D-Var techniques.

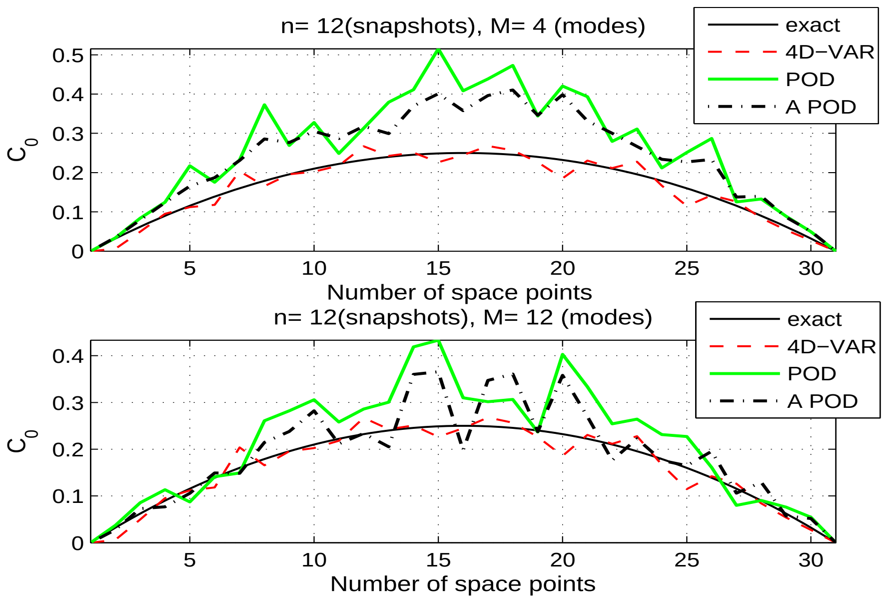

Fig. 3. Comparative results on the assimilation of $\mathbf{c}_0$, maintaining the same number of snapshots ($n = 12$) and increasing the number of POD modes, from $M = 4$ to $M = 12$.

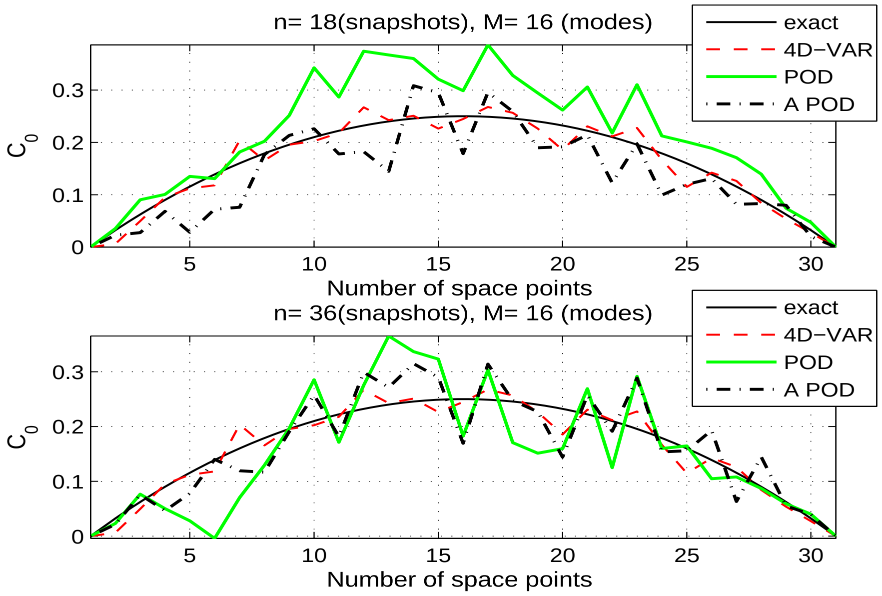

Fig. 4. Comparative results on the assimilation of $\mathbf{c}_0$, maintaining the same number of POD modes ($M = 16$) and increasing the number of snapshots, from $n = 18$ to $n = 36$.

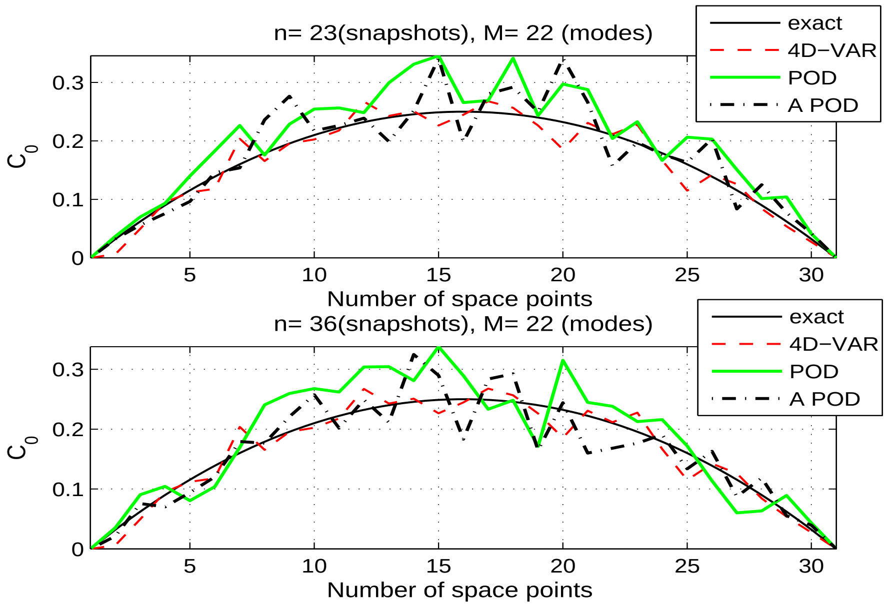

Fig. 5. Comparative results on the assimilation of $\mathbf{c}_0$, maintaining the same number of POD modes ($M = 22$) and increasing the number of snapshots, from $n = 23$ to $n = 36$.