A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition

Yanhua Cao,
Institute of atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
Email: caoyh@mail.iap.ac.cn

Jiang Zhu,
Institute of atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
Email: jzhu@mail.iap.ac.cn

I. Michael Navon,
School of Computational Science and Department of Mathematics, Florida State University,
Tallahassee, FL 32306-4120, USA.
Email: navon@csit.fsu.edu

Zhendong Luo,
College of Fundamental Sciences, North China University of Technology, Beijing, 100041, China
Email: luozhendong@sohu.com
A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition

Abstract

A major hurdle in use of four dimensional variational data assimilation for realistic general circulation models is the dimension of the control space and the high computational cost in computing the cost function and its gradient that require integration model and its adjoint model). In this paper, we propose a 4DVAR approach based on proper orthogonal decomposition (POD). POD is an efficient way to carry out reduced order modeling by identifying the few most energetic modes in a sequence of snapshots from a time-dependent system, and providing a means of obtaining a low-dimensional description of the system's dynamics. The POD based 4DVAR not only reduces the dimension of control space, but also reduces the size of dynamical model, both in dramatic ways. The novelty of our approach also consists in the inclusion of adaptability, applied when in the process of iterative control the new control variables depart significantly from the ones on which the POD model was based upon. In addition, these approaches also allow to conveniently constructing the adjoint model.

The proposed POD based 4DVAR methods are tested and demonstrated using a reduced gravity wave ocean model in Pacific domain in the context of identical twin data assimilation experiments. The results show that POD-based 4DVAR methods have the potential to approximate the performance of full order 4DVAR with a much smaller computational cost (less than 1/100 computer time of the full order 4DVAR). This study also shows that further research efforts in this direction are worth pursuing and may lead ultimately to a practical implementation in both operational NWP and ocean forecasts.

Key words: proper orthogonal decomposition, variational data assimilation,
ocean-modeling.
1. Introduction

Four dimensional variational data assimilation (4DVAR) is a powerful tool to obtain dynamically consistent atmospheric and oceanic flows that optimally fit observations. Since its introduction (see LeDimet and Talagrand, 1986), 4DVAR has been applied to numerical weather prediction (NWP) (e.g., Courtier et al. 1994), ocean general circulation estimation (e.g., Stammer; Awaji et al. 2003) and atmosphere-ocean-land coupled modeling (Awaji et al. 2003). However, a major hurdle in use of 4D-Var for realistic general circulation models is the dimension of the control space, generally equal to the size of the model state variable and typically of order $10^7 – 10^8$. Current ways to obtain feasible implementations of 4D-Var consist mainly of the incremental method (Courtier et al. 1994) which is the method adopted at all operational centers implementing 4D-Var. Additionally check-pointing (Griewank 2000, Griewank and Walter, 2000) and parallelization are also used. The incremental method proposed by Courtier et al. (1994) consists in generating a succession of quadratic problems which can be solved in inner loop using a coarse resolution corrected by full model runs in few outer loops –however the method is characterized by the fact that the dimension of the control space remains very large in realistic applications (see Li et al. (2000), Gauthier (2003), Tremolet (2004)). Memory storage requirements impose a severe limitation on the size of assimilation studies, even on the largest computers. Checkpointing strategies (Restrepo et al. 1998, Griewank and Walter, 2000) have been developed to address the explosive growth in both on-line computer memory and remote storage requirements for computing the gradient by the forward/adjoint technique that characterizes large-scale assimilation studies. It was shown that the tradeoff between the storage requirements and the computational time might be optimized such that the storage and computational time grow only logarithmically (Griewank 1992).

Parallelization using message-passing interface (MPI) is currently used to implement 4D-Var (ECMWF, NCEP, and WRF). In order to reduce the computational cost of
4D-Var data assimilation we can consider carrying out the minimization of the cost functional in a space whose dimension is much smaller than that of the original one. A way to drastically decrease the dimension of the control space without significantly compromising the quality of the final solution but sizably decreasing the cost in memory and CPU time of 4D-Var motivates us to choose to project the control variable on a basis of characteristic vectors capturing most of the energy and the main directions of variability of the model, i.e. SVD, EOF, Lyapunov or bred vectors. One would then attempt to control the vector of initial conditions in the reduced space model.

Up to now, most efforts of model reduction have centered on Kalman and extended Kalman filter data assimilation techniques (Todling et al. 1994, 1998; Pham et al. 1998; Cane et al. 1996; Dee 1990; Evensen 1992; Fukumori 1995; Fukumori and Malanotte-Rizzoli 1995; Hoang et al. 1997; Verlaan and Heemink 1997; Hoteit and Pham 2003). In particular, Cane et al. (1996) employed a reduced order method in which the state space is reduced through the projection onto a linear subspace spanned by a small set of basis functions, using an empirical orthogonal function (EOF) analysis. This filter is referred to as the reduced order extended Kalman (ROEK) filter.

Some initial efforts aiming at the reduction of the dimension of the control variable - referred to as reduced order strategy for 4D-Var ocean data assimilation were put forward initially by Blayo et al. (1998), Durbiano (2001) and more recently by Hoteit et al. (2004) and Robert et al. (2005). They used a low dimension space based on the first few EOF’s or empirical orthogonal functions, which can be computed from a sampling of the model trajectory. Hoteit et al. (2004) used the reduced order model for part of the 4-D VAR assimilation then switched to the full model in a manner done earlier by Peterson (1989).

The proper orthogonal decomposition (POD) is an efficient way to reduced order
modeling by identifying the few most energetic modes in a time-dependent system, thus providing a means of obtaining a low-dimensional description of the system’s dynamics. It was successfully used in a variety of fields including signal analysis and pattern recognition (see Fukunaga 1990), fluid dynamics and coherent structures (see Aubry et al. 1988; Holmes et al. 1996; Ma and Karniadakis 2002; Bansch 1991; Willcox, K. et al., 2002) and more recently in control theory (see Afanasiev, et al. 2001; Arian, et al. 2000; Kepler et al. 2000; Ly and Tran 2002; Ly and Tran 2002) and inverse problems (see Banks et al. 2000). Moreover, Atwell et al. (2000) had successfully utilized POD to compute reduced-order controllers. For a comprehensive description of POD theory and state of the art research, see Gunzburger (2003) and Gunzburger et al. (2004).

In this paper we apply POD to 4DVAR our first aim being to explore the feasibility of significant reduction in the computational cost of 4DVAR. Our basic approach will build on the POD-based adaptive control of Hinze and Kunisch (2000) and Arian, Fahl and Sachs (2002). The novelty of our approach resides also in the inclusion of adaptivity, applied when in the process of iterative control the new initial condition departs significantly from the one on which the POD model was based upon. The paper is arranged as follows. A brief review of POD is given in section 2. A 4DVAR formulation based on POD and an adaptive POD 4DVAR are proposed in section 3. The numerical model used in this study is a reduced gravity ocean model and its POD model is described in section 4. The accuracy of the POD model is also examined in section 4. Section 5 contains results from identical twin data assimilation experiments using 4DVAR, POD 4DVAR and adaptive POD 4DVAR, respectively. Finally, Section 6 provides main conclusions and discussions of some related issues of this study.

2. POD

POD has been shown an efficient way to reduced order modeling by identifying the
few most energetic modes in a time-dependent system, thus providing a means of obtaining a low-dimensional description of the system’s dynamics. For a complex temporal-spatial flow $U(t, x)$, we denoted by $U^1, ..., U^n$ an ensemble adequately chosen in a time interval $[0, T_x]$, that is $U^i = U(t_i, x)$. Define the mean:

$$\bar{U} = \frac{1}{n} \sum_{i=1}^{n} U^i$$  \hfill (2.1)

We expand $U(t, x)$ as

$$U^{POD}(t, x) = \bar{U}(x) + \sum_{i=1}^{M} c_i(t)\Phi_i(x)$$  \hfill (2.2)

where the POD basis vector $\Phi_i(x)$ and $M$ are judiciously chosen to capture the dynamics of the flow as follows. First, define the spatial correlation matrix $K (n \times n)$ with entries

$$K_{ij} = \int_{\Omega} (U^i - \bar{U})^T (U^j - \bar{U}) d\Omega, \quad 1 \leq i, j \leq n$$  \hfill (2.3)

Next the eigenvalue $K\nu = \lambda \nu$  

$$\hfill (2.4)$$

is solved to obtain the eigenvalues $\lambda_1, ..., \lambda_n$ and the orthonormal eigenvectors $\nu_1, ..., \nu_n$ (if rank (K) $< n$ only the eigenvectors associated to the nonzero eigenvalues are computed). The POD basis vectors are obtained by defining

$$\Phi_i = \sum_{k=1}^{n} (\nu_i)_k (U^k - \bar{U}), \quad i = 1, ..., M$$  \hfill (2.5)

which are then normalized $\Phi_i = \Phi_i/\|\Phi_i\|$ to obtain an orthonormal basis.

One can define a relative information content to choose a low-dimensional basis of size $M (<<n)$ by neglecting modes corresponding to the small eigenvalues. We define

$$I(m) = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$  \hfill (2.6)

and choose $M$ such that

$$M = \arg \min \{I(m) : I(m) \geq \gamma\}$$
where $0 \leq \gamma \leq 1$ is the percentage of total information captured by the reduced space $D^M = \text{span}\{\Phi_1, \ldots, \Phi_M\}$. The tolerance $\gamma$ 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 (2.2).

For an atmospheric or oceanic flow $U(t, x)$, it is usually governed by a dynamic model

$$\frac{dU}{dt} = F(U, t).$$

$$U(0, x) = U_0(x).$$

To obtain a reduced model of (2.7), we can first solve (2.7) for an ensemble of snapshots and follow above procedures, then use a Galerkin projection of the model equations onto the space spanned by the POD basis elements (replacing $U$ in (2.7) by (2.2), then multiplying $\Phi_i$ and integrating over spatial domain):

$$\frac{dc_i}{dt} = \left\langle F(\hat{U} + \sum_{i=1}^{M} c_i \Phi_i, t), \Phi_i \right\rangle.
\quad (2.8)$$

$$c_i(t = 0) = c_i(0)$$

Equation (2.8) defines a reduced model of (2.7). In the following sections we will discuss applying this model reduction to 4DVAR in which the forward model and the adjoint model for computing the cost function and its gradient is the reduced model and its corresponding adjoint.

3. POD-4DVAR

3.1 POD-4DVAR

At the analysis time $[0, T_N]$, strong constraint 4DVAR looks for an optimal solution of (2.7) to minimize a cost function

$$J(U_0) = (U_0 - U_b)^T B^{-1} (U_0 - U_b) + (HU - y^o)^T O^{-1} (HU - y^o).$$

(3.1)

In POD 4DVAR, we look for an optimal solution of (2.7) to minimize the cost function
\[ J(c_1(0), \ldots, c_M(0)) = (U_0^{\text{POD}} - U_b)^T B^{-1} (U_0^{\text{POD}} - U_b) + (H U_0^{\text{POD}} - y^\alpha)^T O^{-1} (H U_0^{\text{POD}} - y^\alpha) \] (3.2)

where \( U_0^{\text{POD}} \) is the control vector, \( H \) is an observation operator, \( B \) is the background error covariance matrix and \( O \) is the observation error covariance matrix.

In (3.2),

\[ U_0^{\text{POD}}(x) = U_0^{\text{POD}}(x,0) = \overline{U}(x) + \sum_{i=1}^{M} c_i(0) \Phi_i(x), \]

\[ U^{\text{POD}} = U^{\text{POD}}(x,t) = \overline{U}(x) + \sum_{i=1}^{M} c_i(t) \Phi_i(x). \]

In POD 4DVAR, the control variables are \( c_1(0), \ldots, c_M(0) \). As shown later, the dimension of the POD reduced space could be much smaller than that the original space. In addition, the forward model is the reduced model (2.8) which can be very efficiently solved. The adjoint model of (2.8) is used to calculate the gradient of the cost function (3.2) and that will greatly reduce both the computational cost and coding effort.

To establish POD model in POD 4DVAR, we need first to obtain an ensemble of snapshots, which is taken from the background trajectory, or integrate original model (2.7) with background initial conditions.

### 3.3 Adaptive POD-4DVAR

Since the POD model is based on the solution of the original model for a specified initial condition, it might be a poor model when the new initial condition is significantly different from the one on which the POD model is based upon. Therefore, we propose an adaptive POD 4DVAR procedure as follows:

(i) Establish POD model using background initial conditions and then perform optimization iterations to approximate the optimal solution of the cost function (3.2);

(ii) If after a number of iterations, the cost function cannot be reduced
significantly as measured by a preset criterion, we generate a new set of snapshots by integrating the original model using the newest initial conditions;

(iii) Establish a new POD model using the new set of snapshots and continue optimization iteration;

(iv) Check if the optimality conditions are reached, if yes, stop; otherwise, go to step (ii).

4. Model and POD reduced model

4.1 Model

The numerical model used here is a reduced-gravity model. The equations (Seager et al. 1988) for the depth-averaged currents are

$$\frac{\partial u}{\partial t} - f v = -g \frac{\partial h}{\partial x} + \frac{\tau_x}{\rho_o H} + A \nabla^2 u - \alpha u$$

$$\frac{\partial v}{\partial t} + f u = -g \frac{\partial h}{\partial y} + \frac{\tau_y}{\rho_o H} + A \nabla^2 v - \alpha v$$

(4.1)

$$\frac{\partial h}{\partial t} + H \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0,$$

where \((u,v)\) are the horizontal velocity components of the depth-averaged currents; \(h\) the total layer thickness; \(f\) the Coriolis force; \(H\) the mean depth of the layer; \(\rho_o\) the density of water; and \(A\) the horizontal eddy viscosity coefficient and \(\alpha\) is the friction coefficient. The wind stress is calculated by the aerodynamic bulk formula

$$\left( \tau_x, \tau_y \right) = \rho_a c_D \sqrt{u_{wind}^2 + v_{wind}^2} \left( u_{wind}, v_{wind} \right),$$

where \(\rho_a\) is the density of the air; \(c_D\) the wind stress drag coefficient; \(U\) the wind speed vector; and \((u_{wind}, v_{wind})\) the components of the wind velocity.

In this study, we applied the model to the tropic Pacific Ocean domain (29°S-29°N, 120°E-70°W). This chosen model domain allows all possible equatorially trapped waves, which can be excited for example by the applied wind forcing (Moore and Philander 1978). The model is discretized on the Arakawa C-grid, and all the model
boundaries are closed. The no-normal flow and no-slip conditions are applied at these solid boundaries. The time integration uses a leapfrog scheme, with a forward scheme applied every 10th time step to eliminate the computational mode. We choose the spatial interval for the dynamical model to be $\Delta x = \Delta y = 0.5^\circ$ and the time step to be $\Delta t = 100\,\text{s}$. This temporal-spatial resolution will allow to resolve all possible waves and to make the model integration numerically stable. The model is driven by the Florida State University (FSU) climatology monthly mean winds (Stricherz et al. 1992). The data are projected into each time step by a linear interpolation and into each grid point by a bilinear interpolation. The values of numerical parameters used in the model integration are listed in Table 1. It takes about 20 years for the model to reach a periodic constant seasonal cycle; at that time, the main seasonal variability of dynamical fields has been successfully captured. The currents and the upper layer thickness of the 21st year are saved for POD reduced model and data assimilation experiments as described below.

4.2 Construction of POD reduced model
For successful POD 4DVAR, it is crucial to construct an accurate POD reduced model. In this section, we demonstrate in detail the construction of the above reduced gravity model (referred as full model thereafter) and check its accuracy of approximation to the full model.

The procedure for computing the POD reduced order spaces $X_h^{POD}, X_u^{POD}, X_v^{POD}$ consists of the following steps.

(i) Obtain the snapshots. At first, full model was integrated for 20 years. During the 21st year these equations are solved at $n$ time steps (then snapshots) 

$$\{h_1(\bar{x}), h_2(\bar{x}), \ldots, h_n(\bar{x}); u_1(\bar{x}), u_2(\bar{x}), \ldots, u_n(\bar{x}); v_1(\bar{x}), v_2(\bar{x}), \ldots, v_n(\bar{x})\}$$

at an increment of $360/n$ day for $\bar{x} \in \Omega$ (here $\Omega$ denotes the two-dimensional rectangular domain). These snapshots are discrete data over $\Omega$.

(ii) Compute the covariant matrix $D_h, D_u, D_v$. The matrix elements of $D_h, D_u, D_v$ are
given as \( D_h = A_h^T A_h, D_u = A_u^T A_u, D_v = A_v^T A_v \). Here the space-time transposed technique is used.

(iii) Solve the eigenvalue problem \( D_h V_h = \lambda_h V_h; D_u V_u = \lambda_u V_u; D_v V_v = \lambda_v V_v \). Since \( D_h, D_u, D_v \) are all nonnegative Hermitian matrices, they all have a complete set of orthogonal eigenvectors with the corresponding eigenvalues arranged in ascending order as \( \lambda_{h1} \geq \lambda_{h2} \geq \cdots \geq \lambda_{hn} \geq 0; \lambda_{u1} \geq \lambda_{u2} \geq \cdots \geq \lambda_{un} \geq 0; \lambda_{v1} \geq \lambda_{v2} \geq \cdots \geq \lambda_{vn} \geq 0 \) respectively.

(iv) Compute the POD basis vector. The POD basis elements \( \Phi_{hi}(\bar{x}); \Phi_{ui}(\bar{x}); \Phi_{vi}(\bar{x}) \) such that

\[
X_h^{POD} = \text{span}\{ \Phi_{hi}(\bar{x}), \Phi_{h2}(\bar{x}), \ldots, \Phi_{hn}(\bar{x}) \}
\]

\[
X_u^{POD} = \text{span}\{ \Phi_{ui}(\bar{x}), \Phi_{u2}(\bar{x}), \ldots, \Phi_{un}(\bar{x}) \}
\]

\[
X_v^{POD} = \text{span}\{ \Phi_{vi}(\bar{x}), \Phi_{v2}(\bar{x}), \ldots, \Phi_{vn}(\bar{x}) \}
\]

are defined as

\[
\Phi_{hi} = \sum_{i=1}^{n} a_{hi}^k c_{hi}; \Phi_{ui} = \sum_{i=1}^{n} a_{ui}^k c_{ui}; \Phi_{vi} = \sum_{i=1}^{n} a_{vi}^k c_{vi},
\]

where \( 1 \leq k \leq n \) and \( a_{hi}^k, a_{ui}^k, a_{vi}^k \) are the elements of the eigenvectors \( A_h V_h^k / \sqrt{\lambda_{hi}}; A_u V_u^k / \sqrt{\lambda_{ui}}; A_v V_v^k / \sqrt{\lambda_{vi}} \) corresponding to the eigenvalue \( \lambda_{hi}, \lambda_{ui}, \lambda_{vi} \) respectively.

(v) Construct the POD reduced model. Using above basis functions, for different model variables \( u, v, h \), we can approximate the full model solution using different modes respectively, as the following:

\[
u(\bar{x}, t) = \overline{v}(\bar{x}) + \sum_{i=1}^{n_v} \beta_i^v(t) \Phi_{vi}(\bar{x})
\]

\[
u(\bar{x}, t) = \overline{v}(\bar{x}) + \sum_{i=1}^{n_v} \beta_i^v(t) \Phi_{vi}(\bar{x}) \quad (4.2)
\]

\[
h(\bar{x}, t) = \overline{h}(\bar{x}) + \sum_{i=1}^{n_h} \beta_i^h(t) \Phi_{hi}(\bar{x})
\]

with coefficients \( \beta_i^u(i = 1, \ldots, n_u); \beta_i^v(i = 1, \ldots, n_v); \beta_i^h(i = 1, \ldots, n_h) \) to be determined.
Substituting (4.2) into (4.1) and multiplying

\[ \Phi_{ui}(i = 1, \ldots, n_u); \Phi_{vi}(i = 1, \ldots, n_v); \Phi_{hi}(i = 1, \ldots, n_h) \]

on both sides, then integrating over whole model domain respectively. Since the basis functions are orthonormal, the reduced system of ODEs is as follows

\[
\frac{\partial \beta^u_j(t)}{\partial t} = f_1(t, \beta^u_i(t), \ldots, \beta^u_{n_u}(t), \beta^v_i(t), \ldots, \beta^v_{n_v}(t), \beta^h_i(t), \ldots, \beta^h_{n_h}(t)), \quad j = 1, \ldots, n_u,
\]

\[
\frac{\partial \beta^v_j(t)}{\partial t} = f_2(t, \beta^v_i(t), \ldots, \beta^v_{n_u}(t), \beta^v_i(t), \ldots, \beta^v_{n_v}(t), \beta^h_i(t), \ldots, \beta^h_{n_h}(t)), \quad j = 1, \ldots, n_v, \tag{4.3}
\]

\[
\frac{\partial \beta^h_j(t)}{\partial t} = f_3(t, \beta^h_i(t), \ldots, \beta^h_{n_u}(t), \beta^h_i(t), \ldots, \beta^h_{n_v}(t), \beta^h_i(t), \ldots, \beta^h_{n_h}(t)), \quad j = 1, \ldots, n_h.
\]

The initial conditions of (4.3) are

\[
\beta^u_j(0) = (u(\bar{x},0) - \bar{u}(\bar{x}), \Phi_{ui}(\bar{x})), \quad j = 1, \ldots, n_u,
\]

\[
\beta^v_j(0) = (v(\bar{x},0) - \bar{v}(\bar{x}), \Phi_{vi}(\bar{x})), \quad j = 1, \ldots, n_v,
\]

\[
\beta^h_j(0) = (h(\bar{x},0) - \bar{h}(\bar{x}), \Phi_{hi}(\bar{x})), \quad j = 1, \ldots, n_h.
\]

Using the Euler-backwards differencing scheme to solve the above ODE (4.3) problems, the approximated solutions can be obtained.

### 4.3 Accuracy of POD reduced model

The accuracy of POD reduced model had been discussed in detail (see Cao et al. 2005). Here we only display the results in a succinct manner. We consider approximate one-year results of the full model by the POD reduced model. First, we found that no more than 30 snapshots are required to obtain good approximation from POD reduced model. The approximation is very accurate both in terms of root mean square error (RMSE) and in terms of correlations. The overall RMSE is about 1m and correlations are about 0.99 (see Table 2 and Table 3 more details).

The dimension of the POD reduced model depends on the number of basis functions.
We found that only few basis functions (POD modes) are required to capture a high percentage of variability. Figure 1 shows the captured energy by different numbers of POD modes. 99% of variability can be captured by 11 POD modes. The dimension of the POD reduced model is 33 if 11 modes are used. That constitutes a significant reduction considering that the dimension of the full model is $10^4$.

5. POD 4DVAR experiments

5.1 Assimilation experiments

In this section, we present identical twin data assimilation experiments to examine the performances of POD 4DVAR and adaptive POD 4DVAR by comparing them with 4DVAR. The “true” seasonal cycle of tropic Pacific is generated by forcing the model using FSU climatology monthly wind fields as described in the previous section. From the twelve-month’s truth, we generate a set of observations of $h$ that have uncorrelated Gaussian observational errors of zero mean and 0.06 m of variances. Observations are sampled at the one by one degree resolution and a 10-day temporal resolution. This observation network and error characteristics imitate the Topex/POSEIDON/JASON-1 satellite sea surface height observations.

The control variables in these experiments are the initial conditions only. The cost function consists of observation term and the background terms. The observation error covariance matrix is a diagonal one with $0.06^2$ as diagonal elements. The background field is taken from the true state, but on the 100th Day. The background covariance matrix is assumed diagonal and the variances are determined, based on truth-minus-background.

The first experiment is the standard 4DVAR. The dimension of optimization problem exceeds $10^4$. In 4DVAR experiment, we make a preconditioning by the inverse of square root of the background error covariance matrix. The second experiment is POD 4DVAR. The POD model is constructed in the way described in section 4, but the snapshots are taken from the background model results. The number of the
snapshots is 60. Six, nine and ten POD basis functions for $h$ and $u, \nu$, respectively, are used respectively that is sufficient to capture 99% of variability of the snapshots. The dimension of the optimization problem is 25. The third experiment is the adaptive POD 4DVAR that can update the POD model during the optimization procedures. In the adaptive POD 4DVAR experiments, the optimization comprises several outer iterations. In each outer iteration, the POD model is updated from a new set of snapshots that are taken from the full model results based on the result of the previous outer iteration. We stop the present outer iteration and switch to a new outer iteration following the criterion that the gradient should decrease by at least three orders of magnitude from the initial gradient value in the outer iteration minimization.

The numerical solution of the optimal control problem is obtained using the M1QN3 large-scale unconstrained minimization routine, which is based on a limited memory quasi-Newton method. In our case the used memory is five.

5.2 Results
Here we present the numerical results for the three experiments. The assimilation window is one year. The true state is taken from the 21st year of 21-year simulation. The background field for initial condition in the three experiments is taken from the true state on the 100th day. The number of snapshots used in POD-4DVAR and adaptive POD-4DVAR is 60 and the energy captured is more than 99%.

Figure 2 shows the history of the cost function and its gradient during the 4DVAR experiment. The cost function was reduced by more than three orders of magnitude. The gradient of the cost function is also sufficiently reduced that indicates that 4DVAR can successfully approximate the minimum after 85 iterations.

Figure 3 shows the history of the minimization of the cost function and its gradient in the POD 4DVAR experiment. The reduction of the cost function is less than that obtained in the 4DVAR experiment. The gradient is reduced by more than 3 orders in
magnitude. The POD 4DVAR has the limitation that the optimal solution can only be sought within the space spanned by POD basis of background fields. When observations lay outside of the POD space, the POD 4DVAR solution may fail to fit observations sufficiently. This limitation can be improved by adaptively updating POD bases during the optimization. Figure 4 shows the history of the minimization of the cost function and its gradient in the adaptive POD 4DVAR experiment. The cost function is reduced much more than in POD 4DVAR experiment, closer to that in 4DVAR experiments. The final value of the cost function obtained is about 1/20 of that of the first guess.

Figure 5 shows RMSE and correlation coefficients of the outcomes of the three experiments comparing to the true state. All the three experiments have smaller errors than the background. The 4DVAR yields upper layer thickness results that turn out to have the smallest errors. The adaptive POD 4DVAR result exhibits smaller errors than those of POD-4DVAR in term of upper layer thickness. The similar results hold in the term of correlation coefficients.

Figure 6 shows error comparison between the true state and results obtained from 4DVAR and adaptive POD-4DVAR at the initial time. Both 4DVAR and adaptive POD 4DVAR improved initial field significantly. 4DVAR result is basically true field plus some white noise originating from noise in observations. The adaptive POD 4DVAR result does not fit observations as close as that in 4DVAR. Patch structure in figure 6 (c) indicates that some small scale features in observations cannot resolved by POD bases.

5.3 Comparison of computational costs

The computational cost of POD 4DVAR and adaptive POD-4DVAR is much cheaper than that of the full 4DVAR. Since the POD model and its adjoint model are much smaller in size than their full order counterparts, the integrations of POD model and its adjoint model are extremely fast and require less than 1/100 computer time of the
full order models.

6. Conclusions

In this paper, we proposed a reduced order approach to 4DVAR using POD. The approach not only reduces the dimension of the control space, but also reduces the size of the dynamical model, both in dramatic ways. This approach also entails a convenient way of constructing the adjoint model. Further, an adaptive POD 4DVAR is also proposed. To test the POD approach to 4DVAR, a reduced-gravity tropical Pacific model is used to perform identical twin experiments in which conventional 4DVAR, POD 4DVAR and adaptive POD 4DVAR are tested and compared to each other. The main conclusions drawn from this study are:

- The POD model can accurately approximate the full order model with much smaller size;
- The POD 4DVAR has the limitation that the optimal solution can only be sought within the space spanned by POD basis of background fields. When observations lay outside of the POD space, the POD 4DVAR solution may fail to fit observations sufficiently;
- The above limitation of POD 4DVAR can be improved by implementing adaptive POD 4DVAR, with few additional computational time requirements;
- The adaptive POD 4DVAR is capable of delivering comparable results as full order 4DVAR with much less computational cost.

As an initial effort to dramatically reduce computational cost of 4DVAR, the testing assimilation experiments in this study are not as realistic as those in realistic applications. We use a simple model and model simulated data. The control variables are initial condition only. However the results are very promising and show that further research efforts in this direction are worth pursuing and may lead ultimately to a practical implementation of POD-4DVAR in operational NWP and ocean forecasts. In future study, real data and more realistic ocean general circulation models should
be tested. For ocean models, the atmospheric forcing fields should also be included in control variables.

Acknowledgement

This study is supported by Natural Science Foundation of China (40437017, 40225015). Prof. I. M. Navon acknowledges the support from the NSF grant number ATM-9731472.
References


Ma X. and G. Karniadakis. 2002. A low-dimensional model for simulating


<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g'$</td>
<td>$3.7 \times 10^{-2}$</td>
<td>Reduced gravity</td>
</tr>
<tr>
<td>$C_D$</td>
<td>$1.5 \times 10^{-3}$</td>
<td>Wind stress drag coefficient</td>
</tr>
<tr>
<td>$H$</td>
<td>150 m</td>
<td>Mean depth of upper layer</td>
</tr>
<tr>
<td>$\rho_a$</td>
<td>1.2 kg m$^{-3}$</td>
<td>Density of air</td>
</tr>
<tr>
<td>$\rho_0$</td>
<td>1025 kg m$^{-3}$</td>
<td>Density of seawater</td>
</tr>
<tr>
<td>$A$</td>
<td>750 m$^2$ sec$^{-1}$</td>
<td>Coefficient of horizontal viscosity</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$2.5 \times 10^{-5}$</td>
<td>Coefficient of bottom friction</td>
</tr>
</tbody>
</table>

Table 1 Model parameters.
Table 2 RMSE as to 5 snapshots, 20 snapshots and 30 snapshots for different captured energy; (a) upper layer thickness $h$, (b) the zonal current velocity $u$.

(a) RMSE of $h$ (m)

<table>
<thead>
<tr>
<th>RMSE of $h$</th>
<th>95% energy</th>
<th>99% energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 snapshots</td>
<td>1.31539011</td>
<td>0.88490134</td>
</tr>
<tr>
<td>20 snapshots</td>
<td>1.29849041</td>
<td>0.88701826</td>
</tr>
<tr>
<td>30 snapshots</td>
<td>1.27734923</td>
<td>1.07926083</td>
</tr>
</tbody>
</table>

(b) RMSE of $u$ (m/s)

<table>
<thead>
<tr>
<th>RMSE of $u$</th>
<th>95% energy</th>
<th>99% energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 snapshots</td>
<td>0.00761431</td>
<td>0.00669807</td>
</tr>
<tr>
<td>20 snapshots</td>
<td>0.00680718</td>
<td>0.00542305</td>
</tr>
<tr>
<td>30 snapshots</td>
<td>0.00711650</td>
<td>0.00504097</td>
</tr>
</tbody>
</table>
(a) Correlation of $h$

<table>
<thead>
<tr>
<th>Correlation of $h$</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 snapshots</td>
<td>95.8</td>
<td>98.6</td>
<td>99.3</td>
<td>97.9</td>
<td>99.4</td>
<td>96.1</td>
<td>98.8</td>
<td>98.6</td>
<td>99.6</td>
<td>99.0</td>
<td>99.9</td>
<td></td>
</tr>
<tr>
<td>20 snapshots</td>
<td>98.7</td>
<td>99.3</td>
<td>99.6</td>
<td>99.3</td>
<td>98.9</td>
<td>98.0</td>
<td>98.7</td>
<td>99.3</td>
<td>99.6</td>
<td>99.4</td>
<td>99.5</td>
<td></td>
</tr>
<tr>
<td>30 snapshots</td>
<td>98.7</td>
<td>99.4</td>
<td>99.6</td>
<td>99.4</td>
<td>99.2</td>
<td>98.2</td>
<td>98.5</td>
<td>99.3</td>
<td>99.6</td>
<td>99.6</td>
<td>99.5</td>
<td>99.3</td>
</tr>
</tbody>
</table>

(b) Correlation of $u$

<table>
<thead>
<tr>
<th>Correlation of $u$</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 snapshots</td>
<td>84.2</td>
<td>96.2</td>
<td>98.5</td>
<td>92.1</td>
<td>98.8</td>
<td>92.2</td>
<td>97.6</td>
<td>97.0</td>
<td>93.4</td>
<td>97.5</td>
<td>93.6</td>
<td>99.2</td>
</tr>
<tr>
<td>20 snapshots</td>
<td>97.3</td>
<td>98.6</td>
<td>99.0</td>
<td>98.4</td>
<td>96.9</td>
<td>97.9</td>
<td>99.0</td>
<td>98.9</td>
<td>99.0</td>
<td>98.4</td>
<td>98.4</td>
<td>98.9</td>
</tr>
<tr>
<td>30 snapshots</td>
<td>97.1</td>
<td>98.4</td>
<td>98.5</td>
<td>98.8</td>
<td>96.7</td>
<td>97.3</td>
<td>99.0</td>
<td>99.2</td>
<td>98.2</td>
<td>98.7</td>
<td>97.8</td>
<td>98.4</td>
</tr>
</tbody>
</table>

Table 3 Correlation as to 5 snapshots, 20 snapshots and 30 snapshots for (a) upper layer thickness $h$, (b) zonal current velocity $u$, energy captured 99%.
Figure 1 The POD modes capture energy: (a) 5 snapshots, (b) 20 snapshots, (c) 30 snapshots; black line: the upper layer thickness $h$ (m), red line: zonal current velocity $u$ (m/s), and blue line: meridional current velocity $v$ (m/s).
Figure 2 Evolution of the cost function and gradient in 4DVAR experiment. (a) cost function; (b) gradient as a function of the number of minimization iterations.
Figure 3 Evolution of the cost function and gradient in POD-4DVAR experiment. (a) cost function; (b) gradient as a function of the number of minimization iterations.
Figure 4 Evolution of the cost function and gradient in adaptive POD-4DVAR experiment. (a) cost function; (b) gradient as a function of the number of minimization iterations.
Figure 5 RMSE and correlation of the results compared to the true state for upper layer thickness.
Figure 6 Errors between the true state and the numerical approximations for upper layer thickness $h$ (m) in the initial time. (a) the error between the true state and background state, (b) the error between the true state and 4DVAR, (c) the error between the true state and adaptive POD-4DVAR.