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,
Department of Mathematics, Capital Normal University, Beijing, 100037, China
Email: luozhendong@sohu.com

Abstract

Four dimensional variational data assimilation (4DVAR) is a powerful tool for data assimilation in meteorology and oceanography. However, a major hurdle in use of 4DVAR 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 $10^7 - 10^8$) and the high computational cost in computing the cost function and its gradient that require integration model and its adjoint model). Current ways to obtain feasible implementations of 4D-Var consist mainly of the incremental method that 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.

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 4DVAR methods converge faster 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, reduced order, 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 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 Durbiano et al. (2002) 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) 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 section3. 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. Discussions of some related issues are presented in section 6. Finally, Section 7 provides main conclusions of this study.

2. POD

For a complex temporal-spatial flow $U(t,x)$, we denoted by $U^1,...,U^n$ an ensemble adequately chosen in a time interval $[0,T_N]$, that is $U^i = U(t_i,x)$。Define the

mean：

$$\overline{U} = \frac{1}{n}\sum_{i=1}^{n} U^i \qquad (2.1)$$

We expand $U(t,x)$ as

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

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 $\boldsymbol{K}$ (nxn) with entries

$$K_{i,j} = \int_{\Omega} (U^i - \overline{U})^T (U^j - \overline{U}) d\Omega, \qquad 1 \le i, j \le n \qquad (2.3)$$

Next the eigenvalue

$$Kv = \lambda v \qquad (2.4)$$

is solved to obtain the eigenvalues $\lambda_1,...,\lambda_n$ and the orthonormal eigenvectors $v_1,...,v_n$ (if rank (K) < n only the eigenvectors associated to the nonzero eigenvalues are computed). The POD basis vectors are obtained by defining

$$\Phi_i = \sum_{k=1}^{n} (v_i)_k (U^k - \overline{U}), \qquad i = 1,...,M \qquad (2.5)$$

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

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

and choose M such that

$$M = \arg\min\{I(m) : I(m) \ge \gamma\}$$

where $0 \le \gamma \le 1$ is the percentage of total information captured by the reduced space $D^M = span\{\Phi_1,...,\Phi_M\}$. The tolerance 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 (2.2).

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

$$\begin{aligned} &\frac{dU}{dt} = F(U,t) \\ &U(0,x) = U_0(x) \end{aligned} . \qquad (2.7)$$

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 Φ_i and integrating over spatial domain) :

$$\begin{aligned} &\frac{dc_i}{dt} = \left\langle F(\bar{y} + \sum_{i=1}^{M} c_i \Phi_i, t), \Phi_i \right\rangle \\ &c_i(t=0) = c_i(0) \end{aligned} . \qquad (2.8)$$

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

$$J(U_0) = (U_0 - U_b)^T B^{-1} (U_0 - U_b) + (HU - y^o)^T O^{-1} (HU - y^o) . \qquad (3.1)$$

In POD 4DVAR, we look for an optimal solution of (2.7) to minimize the cost function

$$J(c_1(0), \cdots, c_M(0)) = (U_0^{POD} - U_b)^T B^{-1} (U_0^{POD} - U_b) + (HU^{POD} - y^o)^T O^{-1} (HU^{POD} - y^o)$$

(3.2)

where U_0^{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),

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

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

In POD 4DVAR, the control variables are $c_1(0), \cdots, c_M(0)$. 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 computation 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, 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

$$\frac{\partial u}{\partial t} - fv = -g' \frac{\partial h}{\partial x} + \frac{\tau^x}{\rho_0 H} + A\nabla^2 u - \alpha u$$

$$\frac{\partial v}{\partial t} + fu = -g' \frac{\partial h}{\partial y} + \frac{\tau^y}{\rho_0 H} + A\nabla^2 v - \alpha v$$

(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; ρ_o the density of water; and A the horizontal eddy viscosity coefficient and α is the friction coefficient. The wind stress is calculated by the aerodynamic bulk formula

$$(\tau^x, \tau^y) = \rho_a c_D \sqrt{u_{\text{wind}}{}^2 + v_{\text{wind}}{}^2}\,(u_{\text{wind}}, v_{\text{wind}}),$$

where ρ_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^{o}S-29^{o}N, 120^{o}E-70^{o}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 $\Delta x = \Delta y = 0.5^0$ and the time step to be $\Delta t = 100\,\mathrm{s}$. 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)

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

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

(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 $\Phi_{hi}(\vec{x}); \Phi_{ui}(\vec{x}); \Phi_{vi}(\vec{x})$ such that

$$X_h^{POD} = \text{span}\{\Phi_{h1}(\vec{x}), \Phi_{h2}(\vec{x}), \cdots, \Phi_{hn}(\vec{x})\}$$

$$X_u^{POD} = \text{span}\{\Phi_{u1}(\vec{x}), \Phi_{u2}(\vec{x}), \cdots, \Phi_{un}(\vec{x})\}$$

$$X_v^{POD} = \text{span}\{\Phi_{v1}(\vec{x}), \Phi_{v2}(\vec{x}), \cdots, \Phi_{vn}(\vec{x})\}$$

are defined as

$$\Phi_{hk} = \sum_{i=1}^{n} a_{hi}^k c_{hi}; \Phi_{uk} = \sum_{i=1}^{n} a_{ui}^k c_{ui}; \Phi_{vk} = \sum_{i=1}^{n} a_{vi}^k c_{vi},$$

where $1 \leq k \leq n$ and $a_{hi}^k, a_{ui}^k, a_{vi}^k$ are the elements of the eigenvalues $A_h V_h^k / \sqrt{\lambda_{hk}}, A_u V_u^k / \sqrt{\lambda_{uk}}, A_v V_v^k / \sqrt{\lambda_{vk}}$ corresponding to the eigenvalue $\lambda_h^k, \lambda_u^k, \lambda_v^k$ respectively.

(v) Construct the POD reduced model. Using above basis functions, we can approximate the full model solution using the form:

$$\begin{aligned} u(\vec{x},t) &= \bar{u}(\vec{x}) + \sum_{i=1}^{nu} \beta_i^u(t)\Phi_{ui}(\vec{x}) \\ v(\vec{x},t) &= \bar{v}(\vec{x}) + \sum_{i=1}^{nv} \beta_i^v(t)\Phi_{vi}(\vec{x}) \\ h(\vec{x},t) &= \bar{h}(\vec{x}) + \sum_{i=1}^{nh} \beta_i^h(t)\Phi_{hi}(\vec{x}) \end{aligned}, \quad (4.2)$$

with coefficients $\beta_i^u(i=1,\cdots,nu); \beta_i^v(i=1,\cdots,nv); \beta_i^h(i=1,\cdots,nh)$ to be determined. Substituting (4.2) into (4.1) and multiplying

$$\Phi_{ui}(i=1,\cdots,nu);\ \Phi_{vi}(i=1,\cdots,nv);\ \Phi_{hi}(i=1,\cdots,nh)$$

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

$$\frac{\partial \beta_j^u(t)}{\partial t} = f_1(t, \beta_1^u(t), \cdots \beta_{nu}^u(t), \beta_1^v(t), \cdots \beta_{nv}^v(t), \beta_1^h(t), \cdots, \beta_{nh}^h(t)), \quad j = 1, \cdots, n_u,$$

$$\frac{\partial \beta_j^v(t)}{\partial t} = f_2(t, \beta_1^u(t), \cdots \beta_{nu}^u(t), \beta_1^v(t), \cdots \beta_{nv}^v(t), \beta_1^h(t), \cdots, \beta_{nh}^h(t)), \; j = 1, \cdots, n_v, \qquad (4.3)$$

$$\frac{\partial \beta_j^h(t)}{\partial t} = f_3(t, \beta_1^u(t), \cdots \beta_{nu}^u(t), \beta_1^v(t), \cdots \beta_{nv}^v(t), \beta_1^h(t), \cdots, \beta_{nh}^h(t)), \qquad j = 1, \cdots, n_h.$$

The initial conditions of (4.3) are

$$\beta_i^u(0) = (u(\vec{x},0) - \overline{u}(\vec{x}), \Phi_{ui}(\vec{x})), \qquad j = 1, \cdots, n_u,$$

$$\beta_i^v(0) = (v(\vec{x},0) - \overline{v}(\vec{x}), \Phi_{vi}(\vec{x})), \qquad j = 1, \cdots, n_v,$$

$$\beta_i^h(0) = (h(\vec{x},0) - \overline{h}(\vec{x}), \Phi_{hi}(\vec{x})), \qquad j = 1, \cdots, n_h.$$

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 is significant 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 initial conditions only. The cost function consists of observation terms 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 different background fields are chosen for different assimilation window in the following sections. The background covariance matrix is assumed diagonal and the variances are determined, based on truth-minus-background.

We divide these experiments into two groups. In the first group, the length of assimilation window is one month, that is, the observations of the first month are assimilated. In the second group experiments assimilation window is one year, the corresponding observations of the twelve months are assimilated.

In 4DVAR experiments, we carry out a preconditioning by inverse of square root of the background error covariance matrix. In POD 4DVAR experiments, the POD

model is constructed in the way described in section 4, but the snapshots are taken from the background model results within the assimilation window. The number of the snapshots is 30 for one month and 60 for one year, respectively. The POD basis functions are used to capture at least 99% of variability of the snapshots. 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.

5.2 Results from one-month experiments

In this section we present the numerical results for the first group. The assimilation window is one month. The background field is taken from the true state on the 10^{th} day. The number of snapshots used in POD-4DVAR and adaptive POD-4DVAR is 30 and the energy captured is at least 99.99%.

Figure 2 shows the history of the cost function and its gradient during the standard 4DVAR experiment. The reduction of the cost function is more than 2 orders in magnitude. The gradient is reduced by more than 3 orders in magnitude.

Figure 3 shows the history of the minimization of the cost function and its gradient during the POD 4DVAR experiment. The reduction of the cost function is more than that obtained from standard 4DVAR experiment. The gradient is reduced by more than 4 orders in magnitude.

Figure 4a shows the error between the true state and the background state, Figure 4b presents the error between the true state and 4DVAR and Figure 4c displays the error between the true state and POD-4DVAR at the initial time for the upper layer thickness. From these figures, one can see clearly that the results from 4DVAR improve on background and the figure shows that the results from POD-4DVAR are better compared to those obtained in 4DVAR, which has the smallest error.

5.3 Results from one-year experiments

Here we present the numerical results for the second group. The assimilation window is one year. As discussed in previous section, the background field is taken from the true state on the 100^{th} day. The number of snapshots used in POD-4DVAR and adaptive POD-4DVAR is 60 and the energy captured is more than 99%.

Figure 5 shows the history of the cost function and its gradient during the 4DVAR experiment. The cost function is reduced during the first several iterations, but then cannot be decreased any more. The gradient of the cost function is also not sufficiently reduced. There are several possible reasons responsible for the failure of 4DVAR experiment. First one may suspect inaccurate coding of the adjoint model. However we performed strict gradient check and found the adjoint model is accurate up to round-off errors. The second reason may be due to bad conditioning of the problem since the assimilation window is too long. We did some other 4DVAR experiments with different assimilation windows and found that when the length of the window is less than 3 months, 4DVAR can successfully reach the minimum as displayed in section 5.2 for one month. Therefore we conclude that the failure of 4DVAR experiment is due to bad conditioning caused by too long assimilation window. It has been known in NWP that an assimilation window longer than 12 hours could causes failure of the 4DVAR assimilation (Pires et al 1996 , Lawless et al 2005, Li , Navon et al 1993). Overcoming the problem and increasing the length of the assimilation window, thus enabling more data to be assimilated, is an open

problem remaining to be solved.

Figure 6 shows the history of the minimization of the cost function and its gradient during the POD 4DVAR experiment. The reduction of the cost function is more than that obtained in the 4DVAR experiment. The gradient is reduced by more than 3 orders in magnitude.

Figure 7 shows the history of the minimization of the cost function and its gradient during the adaptive POD 4DVAR experiment. The cost function is reduced much more than in either 4DVAR or POD 4DVAR experiments. The final value of the cost function obtained is about 1/20 of that of the first guess.

Figure 8 shows RMSEs of the three experiments comparing to the true state. The POD 4DVAR results have smaller errors than those of 4DVAR in term of upper layer thickness. The adaptive POD 4DVAR yields upper layer thickness results that turn out to have the smallest errors. For the zonal current field, the adaptive POD 4DVAR has also the smallest errors, while 4DVAR and POD 4DVAR have larger errors than the background.

Figure 9 shows a comparison between the true state and numerical results obtained from POD-4DVAR and adaptive POD-4DVAR at the initial time. It shows that the results from adaptive POD-4DVAR are better compared to those obtained with POD-4DVAR. The results appear to be sufficiently close to the true state for the long assimilation window.

5.4 Comparison of computational cost

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 only less than 1/100 computer time of

the full order models. It also preconditions the minimization process. Moreover, if one is using the adaptive POD technique approximately it turns out to be the best one amongst the three experiments in terms of both computational time and the approximation to the true state of the results. However, one should keep in mind that this adaptive POD method may not applicable for any case.

6. Conclusions and discussions

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 potential to improve performance of 4DVAR with much smaller computation and memory requirements;
- 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;
- For a long assimilation window, 4DVAR may fail due to bad conditioning or due to occurrence of multiple-minima while the POD 4DVAR and adaptive POD 4DVAR can provide a potential tool to overcome this problem to some extent;
- This study shows 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.

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

Afanasiev, K. and Hinze, M. 2001. Adaptive control of a wake flow using proper orthogonal decomposition. *Lect. Notes Pure Appl. Math.*, **216,** 317-332.

Arian, E., Fahl, M. and Sachs, E.W. 2000. Trust-region proper orthogonal decomposition for flow control. *NASA/CR-. 2000-210124, ICASE Report*, no. 2000-25.

Atwell, JA., Borggaard, JT. and King, BB. 2001. Reduced order controllers for Burgers' equation with a nonlinear observer. *Int. J. Appl. Math. Comput. Sci.*, **11 (6)**, 1311-1330.

Aubry, N., Holmes, P., Lumley, JL. and Stone, E. 1988. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.*, **192**, 115-173.

Awaji T., Masuda S. , Shuhei Masuda, Yoichi Ishikawa, Nozomi Sugiura,Takahiro Toyoda and Tomohiro Nakamura 2003.State estimation of the North Pacific Ocean by a Four-Dimensional variational data assimilation experiment. *Journal of Oceanography* **59 (6).** 931-943.

Banks, H.T., Joyner, M.L., Winchesky, B. and Winfree, W.P. 2000. Nondestructive evaluation using a reduced-order computational methodology. *Inverse Problems*, **16**, 1-17.

Bänsch, E. 1991. An adaptive finite-element-strategy for the three-dimensional time-dependent Navier-Stokes Equations. *J. Comp. Math.*, **36**, 3-28.

Blayo, E. , J. Blum and J. Verron, 1998: Assimilation Variationnelle de Donnees en Ocenaographie et reduction de la dimension de l'espace de controle: *In Equations aux Derivees Partielles et Applications*, 199-218.

Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C. and Busalacchi, A.J., 1996: Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.*, **101**, 22599-22617.

Cao Yanhua, Jiang Zhu, Zhengdong Luo and I.M Navon,. 2005. Reduced order modeling of the upper tropical Pacific ocean model using proper orthgonal decomposition. *Computers & mathematics with Applications.* In press.

Courtier, P, J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1388.

Dee, D.P., 1990. Simplification of the Kalman filter for meteorological data assimilation. *Q.J.R. Meteorol. Soc.*, **117,** 365-384.

Dommenget D and Stammer D. Nov 2004**.** Assessing ENSO simulations and predictions using adjoint ocean state estimation. *JOURNAL OF CLIMATE* **17 (22)**: 4301-4315.

Durbiano, S. ,2001: Vecteurs caracteristiques de modeles oceaniques pour la reduction d'ordre en assimilation de donnees. Ph.D. Thesis , Universite Joseph Fourier, Laboratoire de Modelisation et calcul. Grenoble, France, 214pp

Durbiano, S. ,E. Blayo, J. Verron, J. Blum and F-X Le Dimet. 2002: A reduced order strategy for 4D-Var ocean data assimilation. Submitted to *Q.J.R. Meteorol. Soc*.

Evensen, G., 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res*., **97 (C11)**, 17905- 17924.

Fukumori, I., Malanotte-Rizzoli P., 1995: An approximate Kalman filter for ocean data assimilation - an example with an idealized gulf-stream model. *Journal of Geophysical Research-Oceans* **100 (C4)**: 6777-6793.

Fukumori, I., 1995: Assimilation of TOPEX sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific Ocean. *Journal of Geophysical Research-Oceans*, **Vol 100 (C12)**, 25027-25039 .

Fukunaga, K. 1990. Introduction to Statistical Recognition. Academic Press, New York.

Gauthier, P. 2003. Operational implementation of variational data assimilation. In Data Assimilation for the Earth System. NATO Science Series. IV. *Earth and Environmental Sciences*, **Vol. 26**, p.167-176.

Griewank, A., 1992: Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation. *Optimization Methods and Software*, **1**, 35-54.

Griewank, A. 1993: Some bounds on the complexity of gradients, Jacobians, and Hessians. *Complexity in Nonlinear Optimization,* P. M. Pardalos, Ed., World Scientific, 128–161.

Griewank, A and Walther , A., 2000: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation . *ACM Transactions on Mathematical Software*,. **26**, 19 -- 45 .

Griewank, A. 2000: Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. *Frontiers in Applied Mathematics*, **Vol. 19, SIAM**, 369 pp.

Gunzburger, Max D., 2003. Perspectives in flow control and optimization Society for Industrial and Applied Mathematics, Philadelphia, 261 pp.

Gunzburger, Max D., 2004: Reduced-order Modeling, Data Compression, and the Design of Experiments. Second DOE Workshop on Multiscale Mathematics. July 20-22, Broomfield, Colorado.

Holmes, P., Lumley, J.L. and Berkooz, G. 1996. Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge Monographs on Mechanics, Cambridge University Press.

Hoang, S., R. Baraille, O. Talagrand, X. Carton and P. De Mey,1997: Adaptive filtering: application to satellite data assimilation in oceanography. *Dynamics of Atmospheres and Oceans*. **Vol. 27** , No 1-4, 257-281.

Hoteit, I., Kohl, A. and D. Stammer, 2004: Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. ECCO Report Series No 27. Submitted to *Ocean Modeling*, 26pp.

Ibrahim Hoteit and Dinh-Tuan Pham. 2003: Evolution of the Reduced State Space and Data assimilation Schemes Based on the Kalman Filter. *Journal of the Meteorological Society of Japan*, **Vol. 81**, No. 1, 21-39.

Kepler, G. M., H. T. Tran, and H. T. Banks. 2000. Reduced order model compensator control of species transport in a CVD reactor. *Optimal Control Application & Methods,* **21**:143 – 160.

Lawless, A.S. Gratton, and Nichols, N.K., 2005. An investigation of incremental 4D-Var using non-tangent linear models, *Quart. J. Roy. Meteor. Soc.,* **131,** 459-476.

Le Dimet F.X., Talagrand O. 1986. "Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects". *Tellus*, ***38A*:** 97-110**.**

Li, Y., Navon, I. M., Courtier, P. and Gauthier , P. 1993.Variational data assimilation with a semi-Lagrangian semi-implicit global shallow water equation model and its adjoint, *Monthly Weather Review*, **121, No. 6**, 1759-1769 .

Li, Z., Navon, I.M., and Yanqiu Zhu, Y. 2000. Performance of 4D-Var with different strategies for the use of adjoint physics with the FSU global spectral model. *Monthly Weather Review* **128 (3)**, 668-688.

Ly, H. V. and Tran H. T. 2002. Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor .Quarterly of Applied Mathematics 60 (4), 631-656.

Ma X. and G. Karniadakis. 2002. A low-dimensional model for simulating three-dimensional cylinder flow, *J. Fluid Mech.*, **458.** 181-190.

Peterson, J.S., 1989: The reduced basis method for incompressible viscous flow calculations. *SIAM J. Sci. Stat. Comput.*, **Vol 10, No 4,** 777-786.

Pham D.T., Verron J. and Roubaix M.C., 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems.* **16 (3-4)**, 323-340.

Pires, C., Vautard, R. and Talagrand, O., 1996. On Extending the limits of Variational Assimilation in Nonlinear Chaotic Systems. *Tellus*, **48A**, 96-121.

Ravindran S.S. 2000. A reduced-order approach for optimal control of fluids using proper orthogonal decomposition .*International Journal for Numerical Methods in Fluids* **34 (5)**, 425-448.

Restrepo, J., Leaf, G. and Griewank, A. 1998. Circumventing storage limitations in variational data assimilation studies. *SIAM J. Sci. Comput.* **19**, 1586-1605.

Robert C, Durbiano S, Blayo E, Verron J, Blum J and Le Dimet F.X. 2005. A

reduced-order strategy for 4D-Var data assimilation**.** *Journal of Marine Systems* **57 (1-2)**: 70-82.

Todling, R., Cohn, S.E. 1994, Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev*. **122**, 2530—2557.

Todling, R., Cohn, S.E. and Sivakumaran, N.S., 1998
Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Mon. Wea. Rev.* **126**, 2274--2286

Tremolet, Y. 2004. Diagnostics of linear and incremental approximations in 4D-Var. *Q.J.R. Meteorol. Soc.***130,(601)**, 2233-2251.

Verlaan, M. and Heemink, A.W., 1997: Tidal flow forecasting using reduced rank square root filters. *Stoch. Hydrol. Hydraul.* **11 (5)**, 349-368.

Parameter	Value	Remarks
g'	3.7×10^{-2}	Reduced gravity
C_D	1.5×10^{-3}	Wind stress drag coefficient
H	150 m	Mean depth of upper layer
ρ_a	1.2 kg m^{-3}	Density of air
ρ_0	1025 kg m^{-3}	Density of seawater
A	750 m^2 sec^{-1}	Coefficient of horizontal viscosity
α	2.5×10^{-5}	Coefficient of bottom friction

Table 1 Model parameters.

RMSE of h (m)

RMSE of h	95% energy	99% energy
5 snapshots	1.31539011	0.88490134
20 snapshots	1.29849041	0.88701826
30 snapshots	1.27734923	1.07926083

RMSE of u (m/s)

RMSE of u	95% energy	99% energy
5 snapshots	0.00761431	0.00669807
20 snapshots	0.00680718	0.00542305
30 snapshots	0.00711650	0.00504097

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.

Correlation of h

Correlation of h	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
5 snapshots	95.8	98.6	99.3	97.9	99.4	96.1	98.8	98.6	98.8	99.6	99.0	99.9
20 snapshots	98.7	99.3	99.6	99.3	98.9	98.0	98.7	99.3	99.6	99.6	99.4	99.5
30 snapshots	98.7	99.4	99.6	99.4	99.2	98.2	98.5	99.3	99.6	99.6	99.5	99.3

Correlation of u

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

Table 3 Correlation as to 5 snapshots, 20 snapshots and 30 snapshots for upper layer thickness h and zonal current velocity u, energy captured 99%.

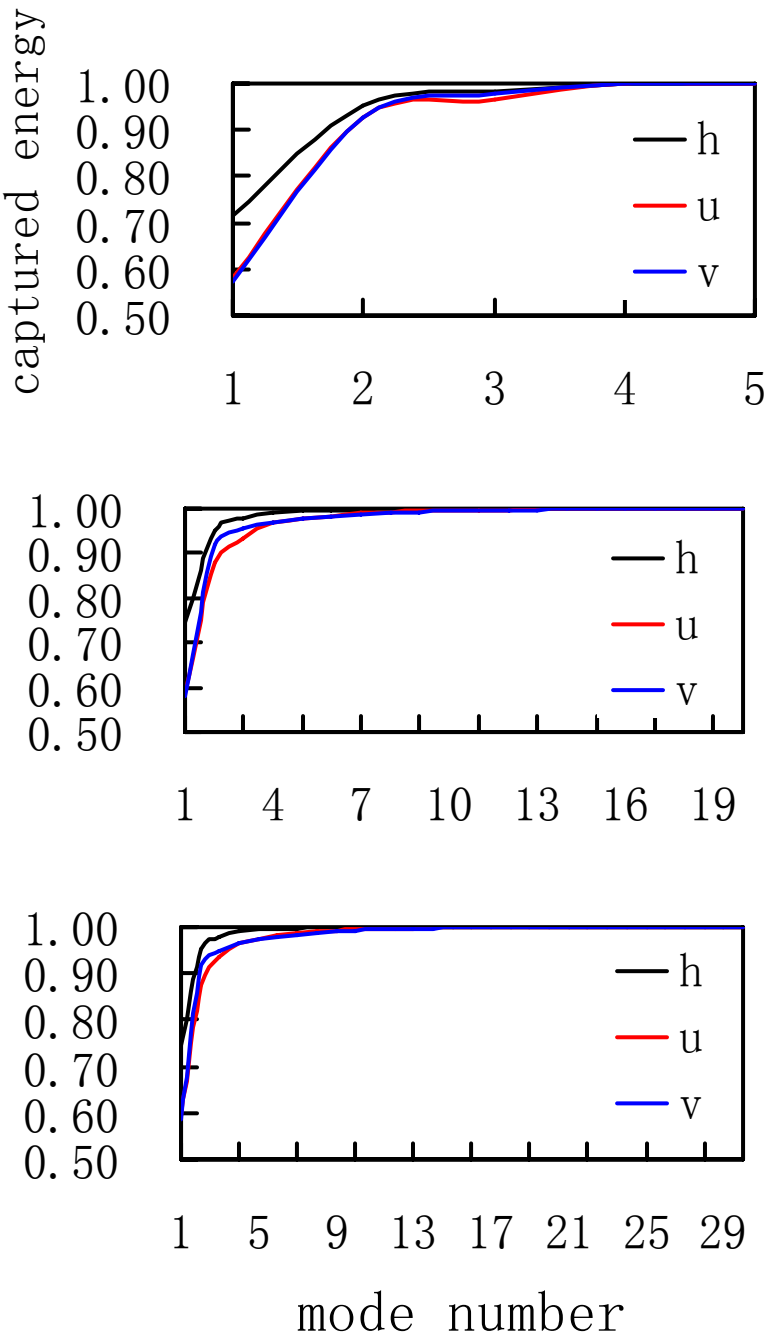

Figure 1 The POD modes capture energy: (a) 5 snapshots, (b) 20 snapshots, (c) 30 snapshots; black line: the upper layer thickness h (m), red line: zonal current velocity u (m/s), and blue line: meridional current velocity v (m/s).

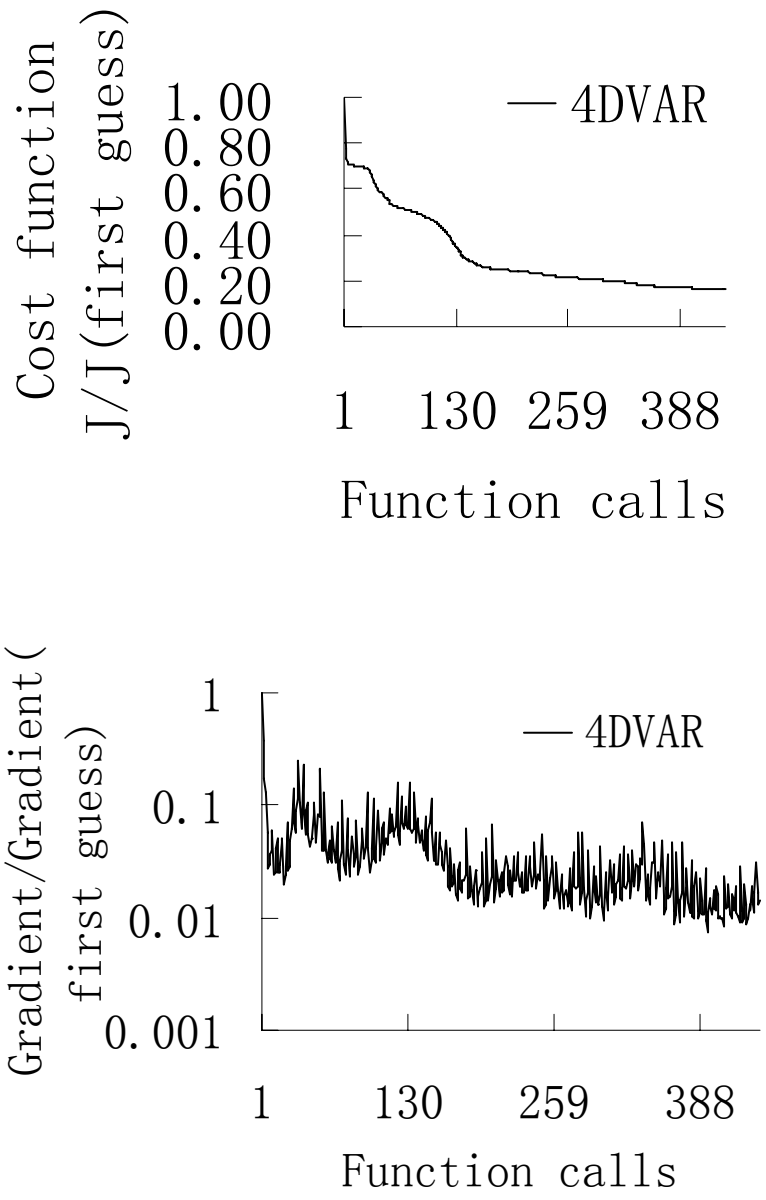

Figure 2 Evolution of the cost function and gradient in standard 4DVAR, (a) cost function (b) gradient as a function of the number of minimization iterations.

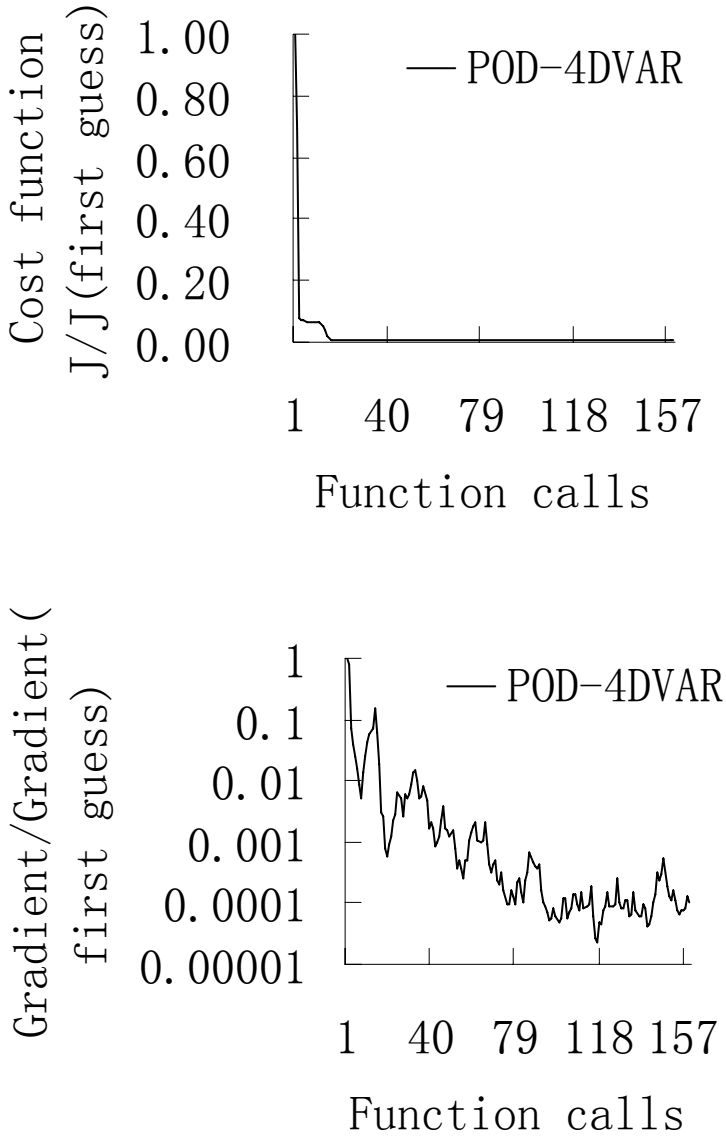

Figure 3 Evolution of the cost function and gradient in POD-4DVAR, (a) cost function (b) gradient as a function of the number of minimization iterations.

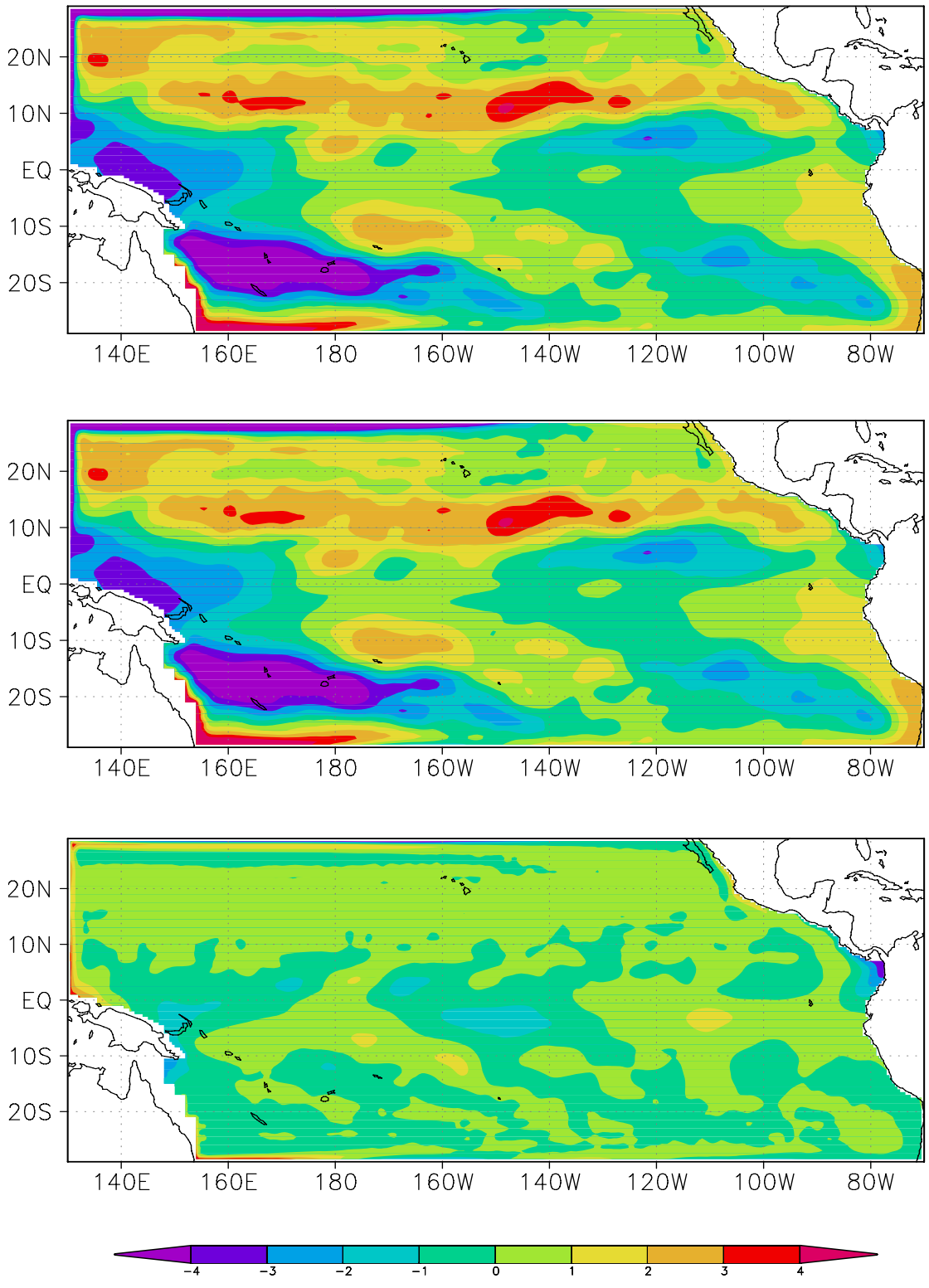

Figure 4 The error about upper layer thickness h (m) in the initial time between (a) the background state and the true state, (b) the 4DVAR and the true state, (c) the POD-4DVAR and the true state.

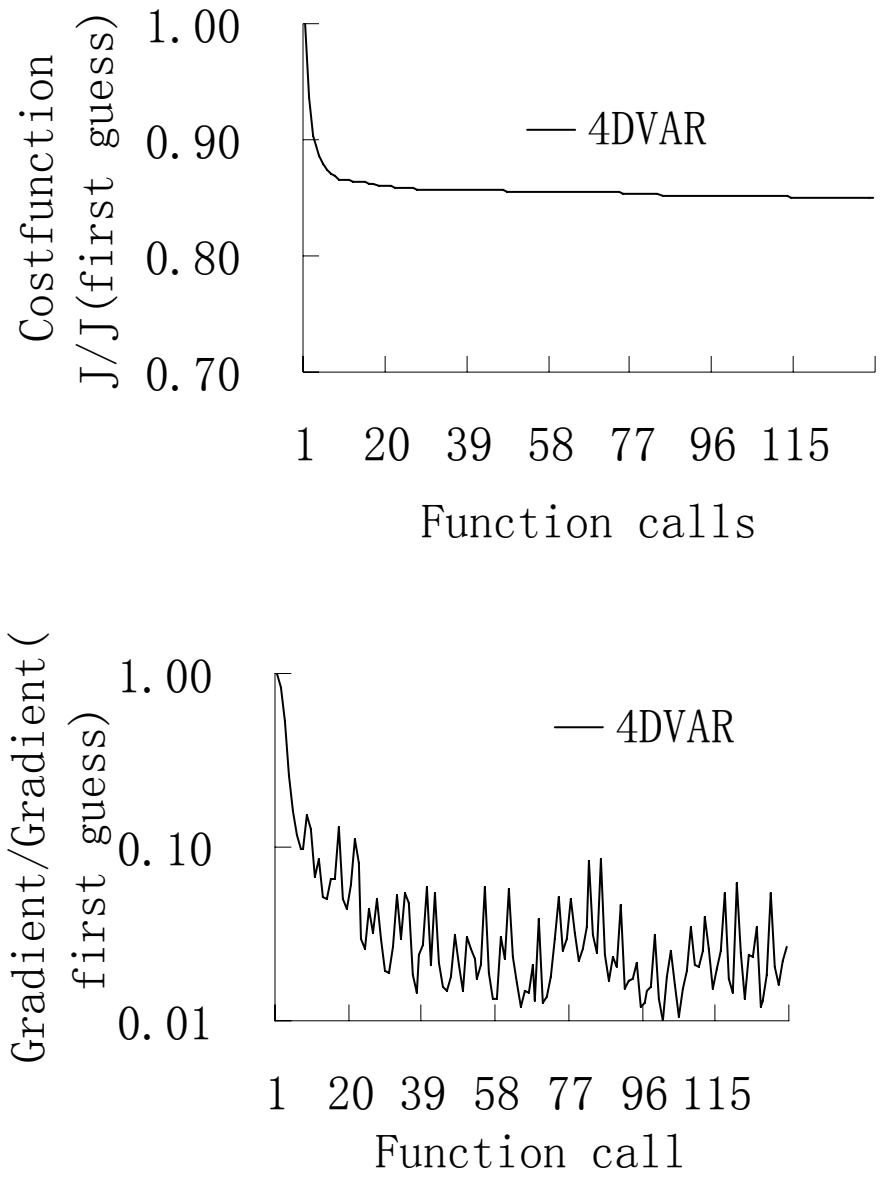

Figure 5 Evolution of the cost function and gradient in standard 4DVAR, (a) cost function (b) gradient as a function of the number of minimization iterations

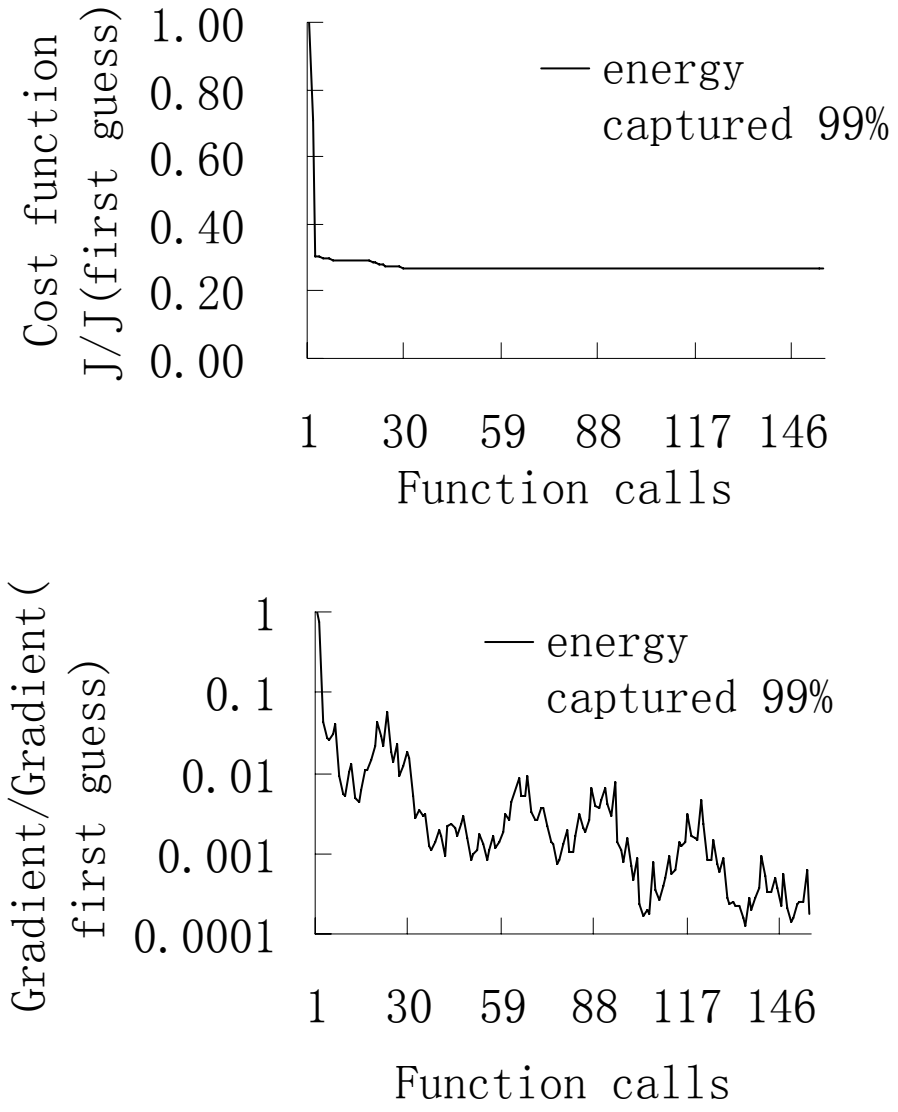

Figure 6 Evolution of the cost function and gradient in POD-4DVAR, energy captured 99% (a) cost function (b) gradient as a function of the number of minimization iterations.

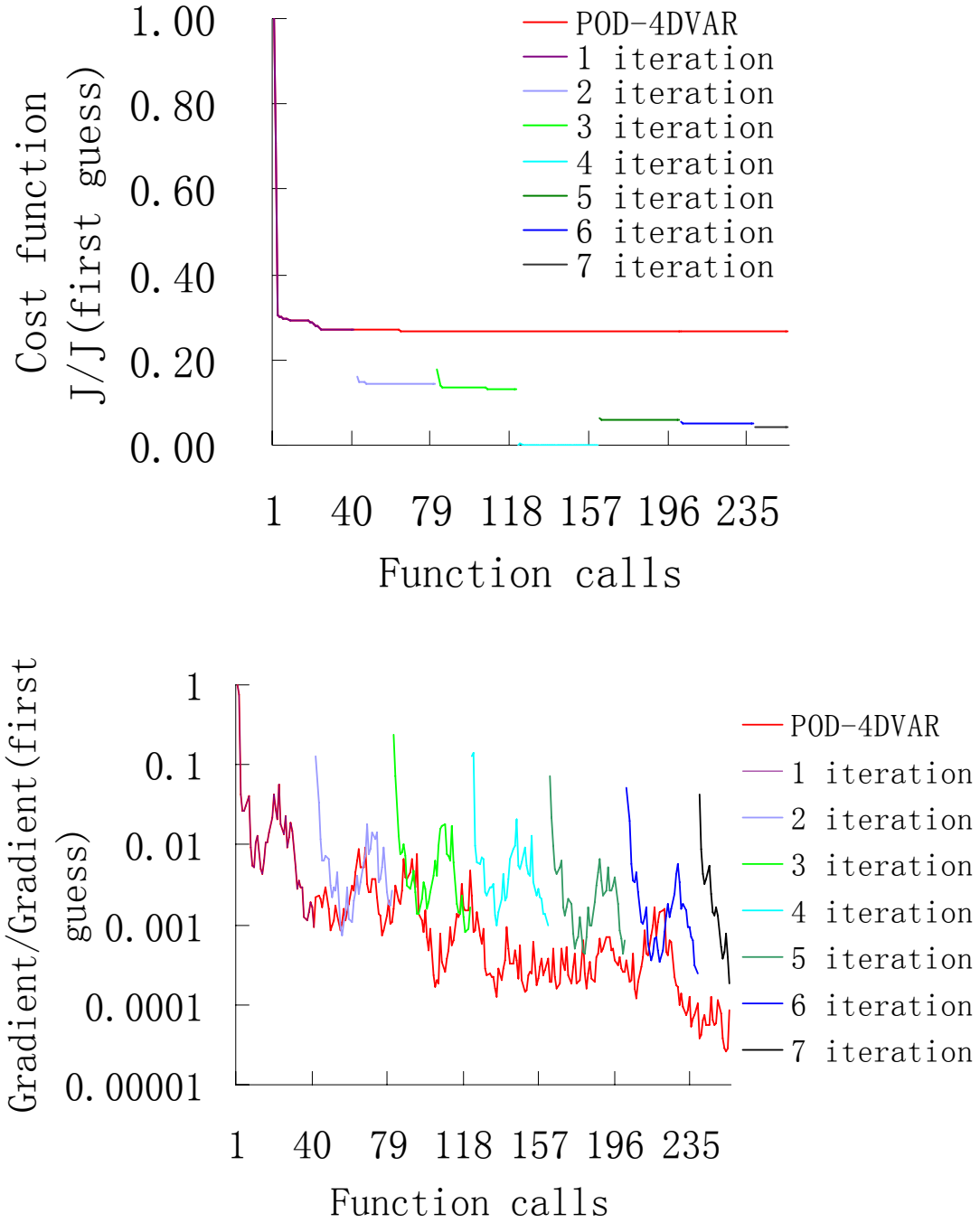

Figure 7 Evolution of the cost function and gradient in adaptive POD-4DVAR: (a) cost function (b) gradient as a function of the number of minimization iterations

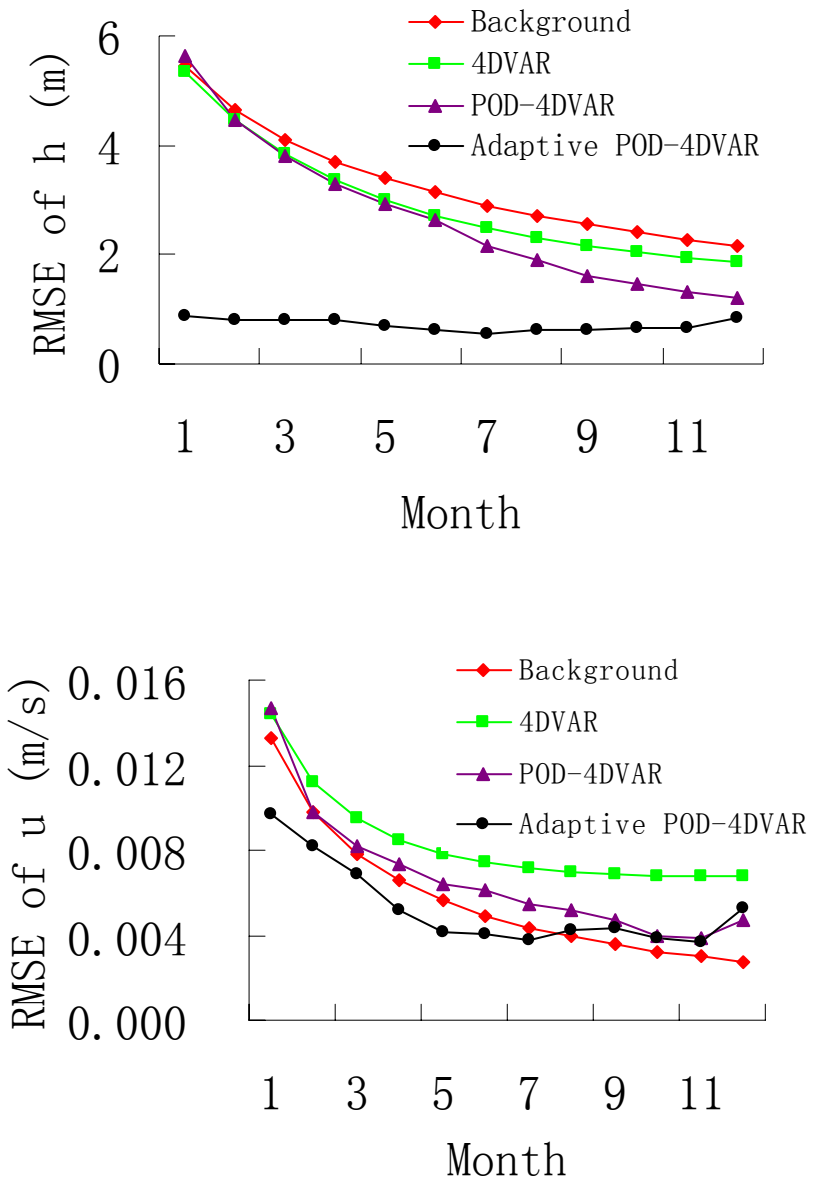

Figure 8 RMSEs of the results comparing to the true state.

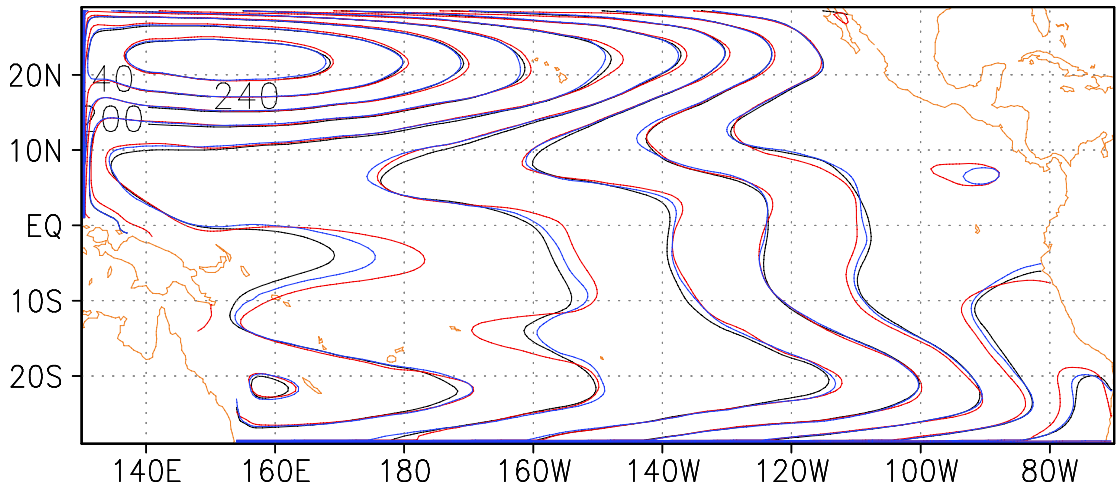

Figure 9 In case of 60 snapshots, energy captured 99%, the true state and the reduced order approximation for upper layer thickness h (m) in the initial time. black isoline: true state, red isoline: POD-4DVAR approximation, blue isoline: adaptive POD-4DVAR approximation.