Reduced Order Modeling of the Upper Tropical Pacific Ocean Model Using Proper Orthogonal Decomposition

Yanhua Cao[a,b] Jiang Zhu[b] Zhendong Luo[a] and Navon, I.M.[c]

[a] Department of Mathematics, Capital Normal University, Beijing, 100037, China

[b] Institute of atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

[c] School of Computational Science and Department of Mathematics, Florida State University, Tallahassee, FL 32306-4120, USA.

Abstract

The proper orthogonal decomposition (POD) is shown to be an efficient model reduction technique for simulating physical processes governed by partial differential equations. In this paper, we make an initial effort to investigate problems related to POD reduced modeling of a large-scale upper ocean circulation in the tropic Pacific domain. We construct different POD models with different choices of snapshots and different number of POD basis functions. The results from these different POD models are compared with that of the original model. The main findings are: (1) the large-scale seasonal variability of the tropic Pacific obtained by the original model is well captured by a low dimensional system of order of 22, which is constructed using 20 snapshots and 7 leading POD basis functions. (2) the RMS errors for the upper ocean layer thickness of the POD model of order of 22 are less than 1m that is less than 1% of the average thickness and the correlations between the upper ocean layer thickness with that from the POD model is around 0.99. (3) Retaining modes that capture 99% energy is necessary in order to construct POD models yielding a high accuracy.

Key words POD, reduced order model, PDE, Galerkin methods, ODE.

AMS subject classifications 76N10, 65K10, 49J20, 35C10.

1. Introduction

The proper orthogonal decomposition (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. Since it was originally introduced by Karhunen in 1946 (see [1]) and Loeve in 1945 (see [2]), the method has been extensively used in research in recent years and successfully applied to a variety of fields. One of these important applications was the application to spatially organized motions in fluid flows, such as cylinder flows (see [3]). POD was also used for identification of coherent structures, signal analysis and pattern recognition (see [4, 5, 6]). Many researchers have also applied the POD technique to optimal control problems. For instance, this method has been used for Burger's equation (see [7, 8, 9]), the Ginzburg-Landau equation and the Bénard convection (see [10]), and in other fluid control problems [11, 12, 13, 14, 15, 16, 17]). More recently POD has also been used in inverse problems (see [18]). In addition, the method has also been used for industrial applications such as supersonic jet modeling (see [19]), thermal processing of foods (see [20, 21]), and study of the dynamic wind pressures acting on buildings ([22]), to name but a few. For a comprehensive description of POD theory and state of the art POD research, see [23, 24].

Compared with above efforts, little attention was paid to application of POD to large-scale geofluid dynamics such as atmospheric or oceanic systems. In general these dynamic systems are quite complex and their discrete models are hard to solve due to their large dimensions (typical 10^6-10^8). Uzunoglu et al (see [25]) applied POD to adaptively reduce an ensemble for numerical weather forecasting. Another obvious application of POD in weather forecasting and operational oceanography is the four-dimensional variational (4DVAR) data assimilation problem. 4D-VAR looks for an optimal solution of an atmospheric or oceanic general circulation model that fits

observations over a certain period (analysis interval) best. 4D-VAR is an optimal control problem. 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 (see [26]), check-pointing (see [27]) and parallelization. However, each of these three methods have their typical defects. The incremental method is characterized by the fact that the dimension of the control space remains very large in realistic applications (see [28,29,30]). Memory storage requirements impose a severe limitation on the size of assimilation studies, even on the largest computers. Checkpointing strategies (see [31]) have been developed to address the explosive growth in both on-line computer memory and remote storage requirements of computing the gradient by the forward/adjoint technique, which characterizes large-scale assimilation studies. POD provides a potential candidate technique that can dramatically reduce computation and memory burdens of 4D-VAR. Cao et al (see [32]) made an initial effort to explore the feasibility of application of POD to 4D-VAR.

Prior to applying POD to various atmospheric and oceanic problems, it is essential to study problems related to construction of POD reduced models: how to choose the number of POD snapshots; how to decide the modes used in such system and how the different modes of basis functions used to reconstruct the solution is affecting the resulting simulation results. These problems have not been studied as of now for large-scale atmospheric or oceanic models. In this paper, we will study these problems with an upper ocean system in the tropical Pacific domain.

The paper is arranged as follows. The upper tropical Pacific Ocean model is described in §2. The POD technique is briefly presented in §3. The issues on the implementation and numerical calculations with POD used in the context of simulating the upper layer thickness and the current in this ocean model are finally discussed in §4.

2. Model of upper tropic Pacific

2.1 Description of the physical model

The numerical model used in this paper is Cane 's reduced-gravity model with a constant-depth surface layer (Cane 1979; Seager et al. 1988), which is studying the ocean dynamics in tropical regions.

The model is a reduced-gravity, linear transport model, consisting of two layers above the thermocline with the same constant density (Figure 1). It is assumed that below the thermocline, the ocean is of a higher density, which is sufficiently deep so that its velocity vanishes and there is no density difference across the base of the surface layer, that is, we regard the surface layer as part of the upper layer. The equations for the depth-averaged currents are

$$\frac{\partial u}{\partial t} - fv = -g' \frac{\partial h}{\partial x} + \frac{\tau^x}{\rho_0 H} + A\nabla^2 u - \alpha u \qquad (2.1a)$$

$$\frac{\partial v}{\partial t} + fu = -g' \frac{\partial h}{\partial y} + \frac{\tau^y}{\rho_0 H} + A\nabla^2 v - \alpha v \qquad (2.1b)$$

$$\frac{\partial h}{\partial t} + H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0, \qquad (2.1c)$$

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_0 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

$$(\tau^x, \tau^y) = \rho_a c_D |\mathrm{U}|_{wind} (U_{wind}, V_{eind}),$$

where ρ_a is the density of the air; c_D the wind stress drag coefficient; U_{wind} the wind speed vector; and (U_{wind}, V_{eind}) the components of the wind velocity.

2.2 Numerical scheme

The dynamical model equations (2.1a)-(2.1c) are governed by wave dynamics. In addition, the chosen model domain ranges from 29°S – 29°N, 120°E – 70°W. This chosen model domain allows all possible equatorially trapped waves, to be excited by the applied wind forcing (Moore and Philander 1978). We choose the spatial interval for the dynamical model to be $\Delta x = \Delta y = 0.5^\circ$ and the time step to be $\Delta t = 100$ s. This temporal-spatial resolution will allow to resolve all possible waves and to render the model integration numerically stable. The model (2.1a)-(2.1c) is driven by the FSU (Florida State University) climatological monthly mean winds (Stricherz et al. 1992). By a linear interpolation, the data are projected onto each time step and into each grid point. In Table 1, the values of the numerical parameters used in the model integration are listed. It takes about 20 years for the model to reach a periodic constant seasonal cycle; at that time, it has successfully captured the main seasonal variability of dynamical fields. The currents and the upper layer thickness of the 21-st year are saved for the process.

The model is discretized on the Arakawa C-grid, and all the model boundaries are closed. At these solid boundaries, we apply the no-normal flow and no-slip conditions. The time integration uses a leapfrog scheme, with a forward scheme every 10th time step to eliminate the computational mode. Every integration day a mass-compensation is carried out.

3. Proper Orthogonal Decomposition

We denote by $U_i(\vec{x}),\ \ i = 1,2,\ldots,n$ the set of n observations (also called snapshots) of some physical process taken at position $\vec{x}$. In this section, we consider the discrete Karhunen-Loève expansion to find an optimal representation of the ensemble of snapshots.

In general, each sample of snapshots $U_i(\vec{x})$ which is defined on a set of m node $\vec{x}$ stands for a m dimensional vector $\vec{u}_i$ as follows:

$$\vec{u}_i = \begin{bmatrix} \vec{u}_{i1} & \vec{u}_{i2} & \cdots & \vec{u}_{im} \end{bmatrix}^T \tag{3.1}$$

where $\vec{u}_{ij}$ represent j component of the vector $\vec{u}_i$. Define the mean vector:

$$\overline{u}_k = \frac{1}{n}\sum_{i=1}^{n} \vec{u}_{ik}, k = 1, \cdots, m. \tag{3.2}$$

We can form a new ensemble by subtracting from the mean as follows:

$$\vec{v}_{ik} = v_i(\vec{x}_k) = \vec{u}_{ik} - \overline{u}_k, k = 1, \cdots, m. \tag{3.3}$$

To find an optimal compressed description of the sequence of data (3.3), one description of the process is a series expansion in terms of a set of basis functions. Intuitively, the basis functions should represent the members of the ensemble in some sense. Such a coordinate system is provided by the Karhunen-Loève expansion. Actually here the basis functions Φ are admixtures of the snapshots given by:

$$\Phi(\vec{x}_k) = \sum_{i=1}^{n} a_{ik} v_i(\vec{x}_k), k = 1, m. \tag{3.4}$$

Here, the coefficients a_{ik} are to be determined so that Φ given by (3.4) will most resemble the ensemble (3.3). More specifically, one looks for a function Φ to maximize

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}(v_{ik}\Phi(\vec{x}_k))^2, \tag{3.5}$$

subject to

$$\sum_{k=1}^{m}(\Phi(\vec{x}_k))^2 = 1.$$

Let matrix A denote the new ensemble:

$$A = \begin{pmatrix} \vec{v}_{11} & \vec{v}_{21} & \cdots & \vec{v}_{n1}) \\ \vec{v}_{12} & \vec{v}_{22} & \cdots & \vec{v}_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ \vec{v}_{1m} & \vec{v}_{2m} & \cdots & \vec{v}_{nm} \end{pmatrix}_{m \times n}$$

Here, the discrete covariance matrix of the ensemble $\vec{u}$ is

$$C\lambda_k = AA^T y_k = \lambda_k y_k. \tag{3.6}$$

Thus, with the POD mode computed, one must solve an $m \times m$ eigenvalue problem. For a discretization of an ocean problem, the dimension m often exceeds 10^4, so it is often not feasible to the direct solution of this eigenvalue problem. The $m \times m$ eigenvalue problem can be transformed into an $n \times n$ eigenvalue problem [Sirovich, 1987]. The $n \times n$ eigenvalue problem can be solved with the method of snapshots,

$$Dw_k = A^T A w_k = \lambda_k w_k, \quad w_k \in R^n, \tag{3.7}$$

where D is a symmetric and nonnegative matrix, λ_k are the eigenvalues. We can choose the eigenvectors w_k to be orthonormal, and give the POD modes by $\phi_k = A w_k / \sqrt{\lambda_k}$. In matrix form, $\Phi = AW$, where $\Phi = [\phi_1, \ldots, \phi_n]$, $W = [w_1, \ldots, w_n]$. It is shown that the cost functional

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m} (v_{ik} \Phi(\vec{x})_k)^2 = \lambda,$$

which is maximized when the coefficients a_i's of (3.4) are the elements of the eigenvector corresponding to the largest eigenvalue of D.

The $n \times n$ eigenvalue problem (3.7) is more efficient than the $m \times m$ eigenvalue problem (3.6) when the number of snapshots n is smaller than m.

4. POD reduced model

In this section, the POD method is applied to the above upper tropical Pacific Ocean model. This method can provide a systematic way of creating a reduced basis space with the state of the system at n different time instances and m different space locations. As in general reduced order basis methods, one can derive the states from full order numerical computations and n should be sufficiently large so that the snapshots u_i may contain all the salient features of the dynamics being considered. Therefore, through a nonlinear Galerkin procedure the POD basis functions Φ_i with the original dynamics offer the possibility of achieving a high fidelity model (albeit) with a possible large dimension *n*.

To achieve model reduction, we first choose $k << n$ then carry out a nonlinear Galerkin procedure with the set of elements $\{\Phi_1, \Phi_2, ..., \Phi_k\}$. How to choose the values of n and k constitutes a crucial question. Since the associated POD eigenvalues are ordered $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, 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

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

and choose M such that

$$M = \arg\min\{I(k) : I(k) \geq \gamma\}$$

where $0 \leq \gamma \leq 1$ 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 in the vicinity of the unity in order to capture most of the energy of the snapshot basis. Here for our case, if the POD is constructed for n=5 and a reduced order model with k=3 it yields a ratio of about 0.98; and if n=20 or n=30 with k=7 it yields a ratio of above 0.99 for the percentage of kinetic energy retained.

We are now returning to the upper tropical Pacific Ocean model of §2 to apply the POD technique. Therefore, we solve equations (2.1a-2.1c) for the steady state solutions of upper layer thickness and velocity field after 20 years time integration. The 21st year results are depicted graphically in Figure 2 .

4.1 Construction of POD Basis Vectors

We compute the POD reduced order spaces $X_h^{POD}, X_u^{POD}, X_v^{POD}$ using the following algorithmic steps.

(i) Obtain the snapshot. First, integrate equations (2.1a)-(2.1c) 20 years. During the 21st year solve these equations at n (n=5, 20, 30) 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 covariance matrix. D_h, D_u, D_v. The matrix elements of D_h, D_u, D_v are given as $D_h = A_h^T A_h, D_u = A_u^T A_u, D_v = A_v^T A_v$ which is depicted in §3. 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 matrix, 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.

4.2 Reconstruction of Solutions through POD basis Vectors

Since the scales in model variables h, u, v are not uniform, different modes can be chosen to reconstruct the solutions.

In this section, we will take into account the problem of approximation of the

infinite-dimensional equations (2.1a)-(2.1c) by a sequence of finite-dimensional problems with combination of Galerkin approximations and POD basis elements.

First, different modes of the basis functions will be used to reconstruct model variables, which assume the following forms

$$\begin{aligned} u(\vec{x},t) &= \bar{u}(\vec{x}) + \sum_{i=1}^{n_u} \beta_i^u(t)\Phi_{ui}(\vec{x}) \\ v(\vec{x},t) &= \bar{v}(\vec{x}) + \sum_{i=1}^{n_v} \beta_i^v(t)\Phi_{vi}(\vec{x}) \\ h(\vec{x},t) &= \bar{h}(\vec{x}) + \sum_{i=1}^{n_h} \beta_i^h(t)\Phi_{hi}(\vec{x}) \end{aligned} .$$

Once the coefficients $\beta_i^u(i=1,\cdots,n_u);\beta_i^v(i=1,\cdots,n_v);\beta_i^h(i=1,\cdots,n_h)$ have been obtained, then we substitute $u(\vec{x},t),v(\vec{x},t),h(\vec{x},t)$ into equations (2.1a)-(2.1c) and multiply $\Phi_{ui}(i=1,\cdots,nu);\ \Phi_{vi}(i=1,\cdots,nv);\ \Phi_{hi}(i=1,\cdots,nh)$ and finally integrate respectively in terms of $\vec{x}$. Since the basis functions are orthonormal, the system of ODE is as follows

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

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

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

along with the initial condition

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

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

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

By solving the above ODE problems using a difference scheme, one obtains the reconstructed solutions.

4.3 Numerical results

In this section, we report results of numerical computations related to the approaches presented in the previous paragraphs.

Here, if n=5, the first four POD modes (Figure 3), capture nearly 100% of the characteristics of the five observations. While for n=20 or n=30, the first seven POD modes capture about 99% energy. It can be clearly seen that for the upper layer thickness h, the same modes may capture the most energy, next is u and the least is v. Thus, different POD modes may be used to reconstruct h, u, v fields respectively.

To quantify the performance of the reduced basis method, we use two metrics namely the root mean square error (RMSE) and correlation of the difference between the full order and the reduced order simulation. This is obtained by first taking the twelve-month's full order results and the corresponding twelve-month's reduced order results and computing the error, for example, for variable u this yields

$$RMSE_{um} = \sqrt{\frac{1}{M}\sum_{i=1}^{M} | \hat{u}_m(z_i) - u_m(z_i) |^2} ,$$

where M is the number of node, the index m denotes the month, $\hat{u}_m$ is the full order approximation and u_m denotes the reduced order approximation. The average RMS error is defined as:

$$RMSE_u = \frac{1}{12}\sum_{m=1}^{12} RMSE_{um} = \frac{1}{12}\sum_{m=1}^{12} \sqrt{\frac{1}{M}\sum_{i=1}^{M} | \hat{u}_m(z_i) - u_m(z_i) |^2} .$$

the correlation as:

$$CORRELATION_{um} = \frac{\sum_{i=1}^{M} (\hat{u}_m(z_i) - \bar{\hat{u}}(z_i))(u_m(z_i) - \bar{u}(z_i))}{\sqrt{\sum_{i=1}^{M} (\hat{u}_m(z_i) - \bar{\hat{u}}(z_i))^2 . \sum_{i=1}^{M} (u_m(z_i) - \bar{u}(z_i))^2}} ,$$

where $\bar{\hat{u}}$ and $\bar{u}$ are the average of full order approximation and reduced order approximation respectively. Similarly compute the RMSE and the correlation for other model variables v and h. Table 2 presents the average RMSE in reduced order approximations using different modes as to n=5, n=20, and n=30 snapshots. Note that from these simulations, on one hand, as the span of the reduced basis space increases, the RMSE decreases as long as the same number snapshots is used. On the

other hand, for different number of snapshots but for the same energy percentage captured, the RMSE decrease stops at 30 snapshots. The correlation for twelve months is displayed in Table 3. Clearly, when increasing the POD mode, the correlation increases also for the same snapshots. This increase stops at 30 snapshots (Table 3) and the reported best approximation obtained with 30 snapshots produced a correlation at the same as level as the approximation 20 snapshots.

However, one must also note that a simple linear independence is not a sufficient criterion for choosing the POD mode. It only provides one with some reference. The comparison between the full order and the reduced order is displayed in Figure 4a and Figure 4b for a retained energy percentage of 95% and 99% respectively about upper layer thickness h . From these figures, we can see there is a little improvement between either 20 snapshots or 30 snapshots compared to 5 snapshots, but there is almost no difference between 20 snapshots and 30 snapshots. The contrast between the full order approximation and numerical results obtained using energy captured at 95% for 20 snapshots about currents is displayed in Figure 5a and 5b. It shows that the reduced order approximation may be sufficiently close to the full order approximation. Other experiments have also been carried out, with either more or less snapshots taken and for different percentages of energy captured. From the computational cost and memory storage aspects, 20 snapshots and the energy captured at 99% level yielded the best results.

5. Conclusions

We studied problems related to POD reduced modeling of a large-scale upper ocean circulation in the tropic Pacific domain. The large-scale seasonal variation of the upper tropic Pacific is first simulated using a reduced gravity model with spatial resolution of $\Delta x = \Delta y = 0.5^\circ$ and a time step of $\Delta t = 100$ s. Then we constructed different POD models with different choices of snapshots and different number of POD basis functions. The results from these different POD models are compared with that of the

original model. The main conclusions are

- The large-scale seasonal variability of the tropic Pacific obtained by the original model can be captured well by a low dimensional system of order of 22, that is constructed by 20 snapshots and 7 leading POD basis functions.
- By analysis of RMS errors and correlations, we found that the modes that capture 99% energy are necessary to construct POD models.
- RMS errors for the upper ocean layer thickness of the POD model of order of 22 is less than 1m that is less than 1% of the average thickness. The correlations of the upper ocean layer thickness from the POD model is around 0.99.
- Compared with the upper ocean layer thickness, the velocity fields from the POD model are less accurate, especially the meridional component. This remains a problem to be further explored in forthcoming research.

Our preliminary investigations on the use of POD for the upper ocean circulation simulation yield encouraging results and show that POD can be a powerful tool for various applications such as four-dimensional variational data assimilation. These results will be described in a follow-up paper.

References

[1] Karhunen, K., 'Zur Spektraltheorie Stochastischer Prozesse', *Ann. Acad. Sci. Fennicae*, 37, (1946)**.**

[2] M. Loeve, Functions aléatoires de second ordre, Compte Rend. Acad. Sci., Paris, 220, (1945).

[3] X. Ma and G. Karniadaks, A low-dimensional model for simulating three-dimensional cylinder flow, J. Fluid Mech., 181-190, 458(2002).

[4] Holmes, P., Lumley, J.L., Berkooz, G.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge Monographs on Mechanics, Cambridge University Press (1996).

[5] C. Lopez and E. Garcia-Herandez, Low-dimensional dynamical system model

for observed coherent structures in ocean satellite data, Physica A, 233-250, 328(2003).

[6] L. Sirovich, Turbulence and the dynamics of coherent structures, parts I-III. Q. Appl. Math. XLV(3), 561-590.

[7] Atwell, J.A., Borggaard, J.T., King, B.B.: Reduced order controllers for Burgers' equation with a nonlinear observer. Int. J. Appl. Math. Comput. Sci., 11, 1311{1330 (2001).

[8] D.H. Chambers, R.J. Adrian, P. Moin, D.S. Stewart, and H.J. Sung, Karhunen-Loeve expansion of Burger' model of turbulence, Phys. Fluids, 2573-2582, 31(1988).

[9] K. Kunisch and S. Volkwein, Control of the Burgers' equation by a reduced order approach using proper orthogonal decomposition. J. Optimization Theory Appl., 102 (2) 345-371 (1999).

[10] L. Sirovich, Chaotic dynamics of coherent structures, Physica D, 126-145, 37(1989).

[11] Afanasiev, K., Hinze, M.: Adaptive control of a wake flow using proper orthogonal decomposition. Lect. Notes Pure Appl. Math., 216, 317{332 (2001).

[12] Afanasiev K., Hinze, M.: Entwicklung von Feedback Controllern für abgelöste Strömungen Abschlussbericht des Teilprojekts C3 des Sonderforschungsbereichs 557 'Beeinflussung turbulenter Scherströmungen', Fachbereich Mathematik, Technische Universität Berlin (2001)

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

[14] A.K. Bangia, P.F. Batcho, I.G. Kevrekidis and G.E. Karniadakis, Unsteady two-dimensional flows in complex geometries: comparative bifurcation studies with global eigenfunction expansions. SIAM J. Sci. Comput., 18(3) 775-805 (1997).

[15] G. Berkooz, P. Holmes and J. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows. Ann. Rev. Fluid Mech, 25 777-786 (1993).

[16] Ly HV, Tran HT, Proper orthogonal decomposition for flow calculations and

optimal control in a horizontal CVD reactor .Quarterly of Applied Mathematics 60 (4), 631-656 (2002).

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

[18] Banks, H.T., Joyner, M.L., Winchesky, B., Winfree, W.P.: Nondestructive evaluation using a reduced-order computational methodology. Inverse Problems,

16, 1{17 (2000).

[19]E.Caraballo, M. Saminny, J. Scott, S. Narayan, and J. Debonis, Application of proper orthogonal decomposition to a supersonic axisymmetric jet, AIAA J., 866-877, 41(2003).

[20] Ravindran S.S., Adaptive reduced-order controllers for a thermal flow system using proper orthogonal decomposition SIAM Journal on Scientific Computing 23 (6), 1924-1942 (2002).

[21]E. Balsa-Canto, A. Alonso, and J. Banga, Novel, efficient and reliable method for thermal process design and optimization. Part I: theory. Part II: applications, J. Food Process. Pres., 227-247, 52 (2002).

[22]H. Kikuchi, Y. Tamura, H. Ueda, and K. Hibi, Dynamic wind pressures acting on a tall building model-proper orthogonal decomposition, J. Wind. Eng. Ind. Aerod., 631-646, 71(1997).

[23]Gunzburger, Max D. , Perspectives in flow control and optimization Society for Industrial and Applied Mathematics, Philadelphia, 261, (2003).

[24] Gunzburger, Max D. , Reduced-order Modeling, data Compression, and the design of experiments. Second DOE Workshop on Multiscale Mathematics. July 20-22, Broomfield, Colorado,(2004).

[25] Uzunoglu B., Navon, I.M and Zupanski, M, Adaptive ensemble size reduction, Tellus (2005), in press.

[26] Courtier, P., Th´epaut, J.-N. and Hollingsworth, A., A strategy for operational implementation of 4D-Var, using an incremental approach, Quarterly Journal of the Royal Meteorological Society, 120,1367-1388, (1994).

[27] Griewank, A., evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation *Frontiers in Applied Mathematics*, SIAM. (2000).

[28] Li, Z., Navon, I.M., and Yanqiu Zhu, Y., 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, (2000).

[29] Gauthier, P., 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, (2003).

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

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

[32] Cao, Y., Zhu, J., Navon, I. M, and Luo, Z.D, A reduced order approach to four dimensional variational data assimilation using proper orthogonal decomposition, manuscript to be submitted.

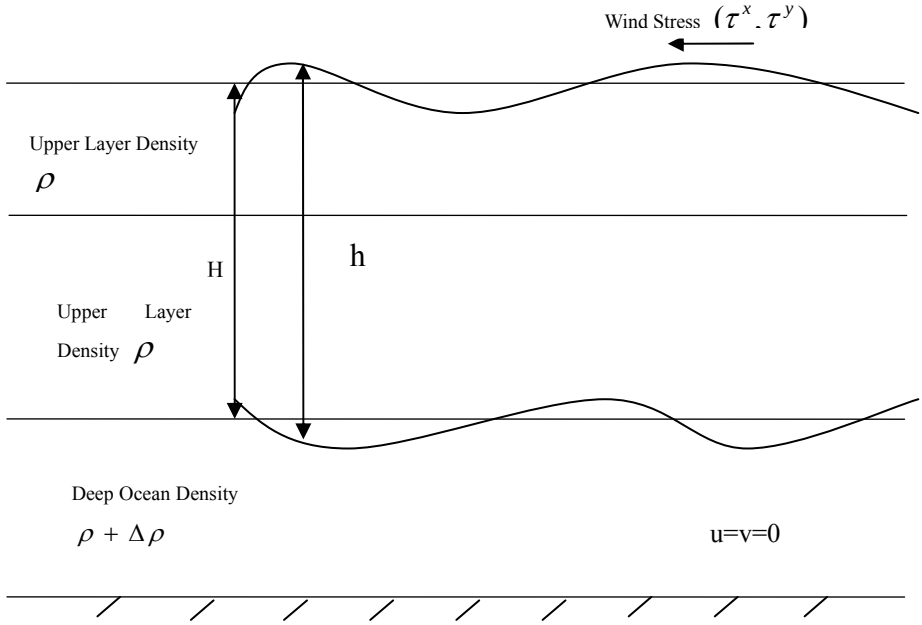


Figure 1 The vertical structure of the reduced-gravity model

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 The values of the model parameters used in the model integration

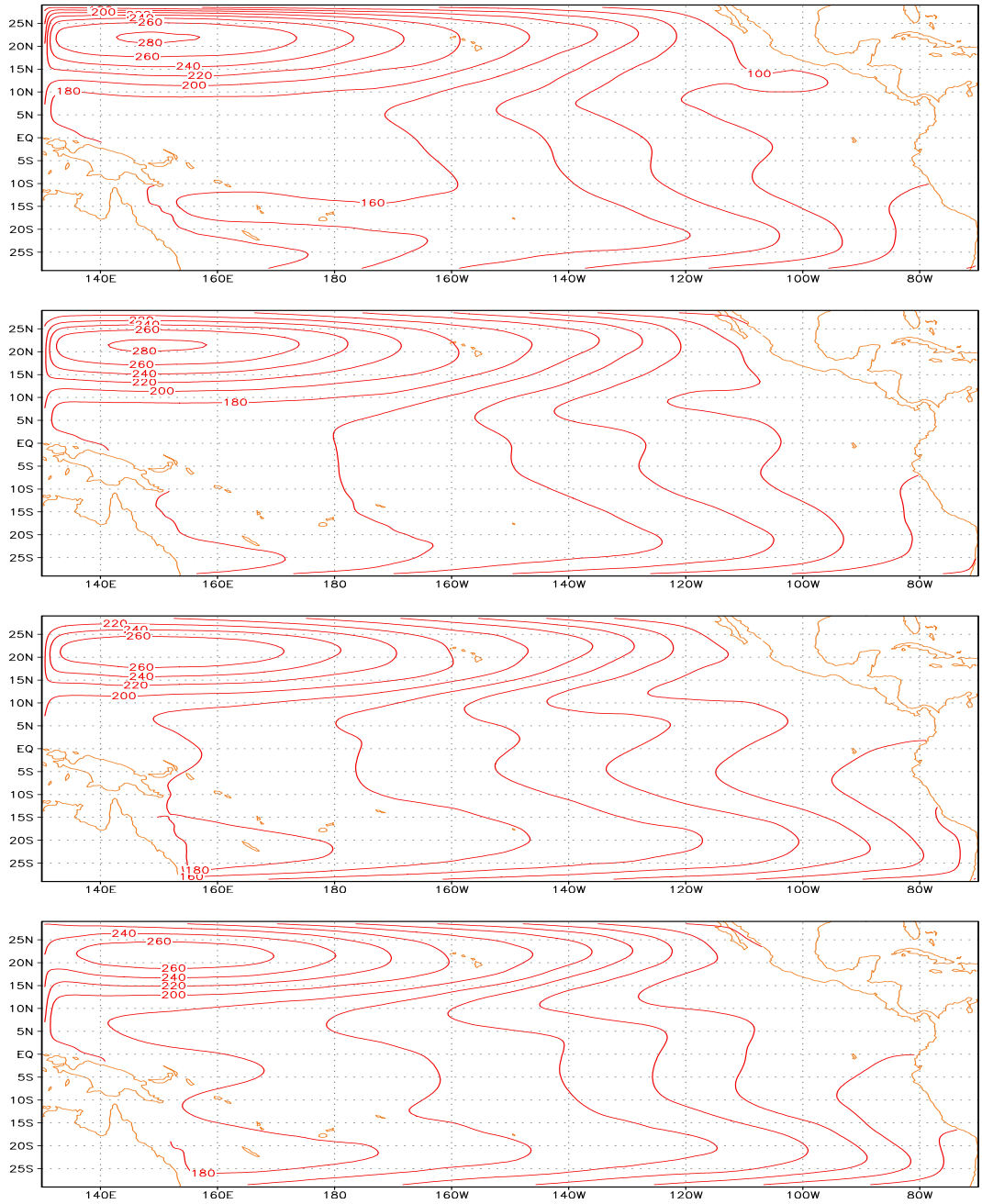
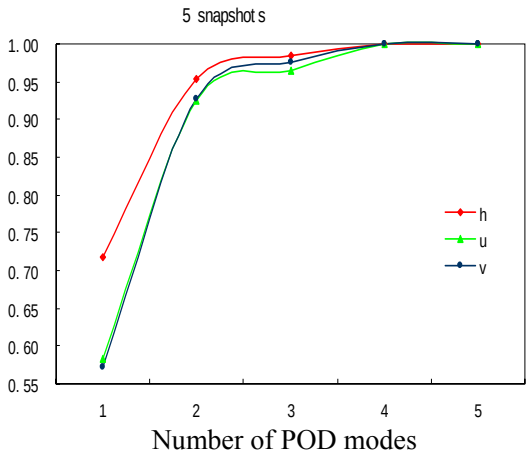
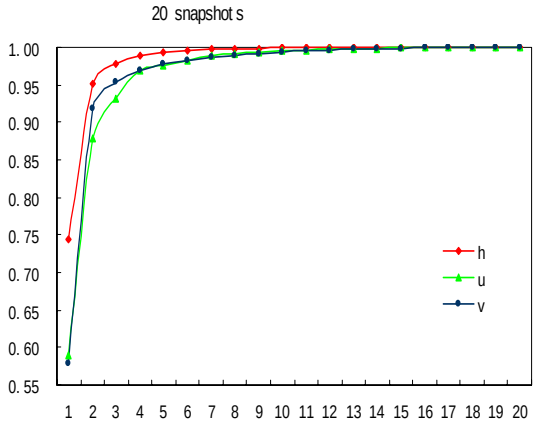


Figure 2 Upper layer thickness of full order approximation (a) February, (b) May, (c) August, and (d) November.





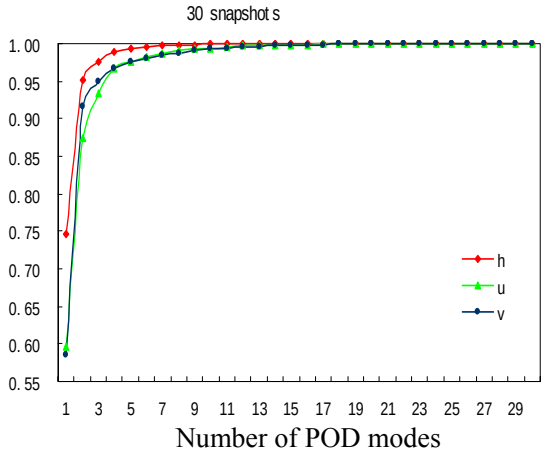


Figure 3 Captured energy vs. number of POD: (a) 5 snapshots, (b) 20 snapshots, (c) 30 snapshots; rhombus line: upper layer thickness h, triangle line: zonal current velocity u, and star line: meridional current velocity v.

RMSE of h	70% energy	95% energy	99% energy
5 snapshots	3.17096138	1.31539011	0.88490134
20 snapshots	2.99194503	1.29849041	0.88701826
30 snapshots	2.97900558	1.27734923	1.07926083

RMSE of u	58% energy	95% energy	99% energy
5 snapshots	0.01243962	0.00761431	0.00669807
20 snapshots	0.01354840	0.00680718	0.00542305
30 snapshots	0.01358298	0.00711650	0.00504097

RMSE of v	54% energy	95% energy	99% energy
5 snapshots	0.01182489	0.00403928	0.00422783
20 snapshots	0.01149406	0.00387623	0.00504759
30 snapshots	0.01146040	0.00474720	0.00536092

Table 2 RMSE as to 5 snapshots, 20 snapshots, and 30 snapshots respectively for different percentages of captured energy; (a) upper layer thickness h (unit: m), (b) the zonal current velocity u (unit: m/s), and (c) the meridional current velocity v (unit: m/s).

Table 3 Correlation coefficients between POD model and the original model

a. POD modes for a 95% captured energy.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Correlation of h												
5 snapshots	0.95	0.97	0.99	0.96	0.93	0.93	0.94	0.97	0.97	0.99	0.98	0.98
20 snapshots	0.96	0.97	0.99	0.95	0.91	0.94	0.95	0.97	0.97	0.99	0.98	0.97
30 snapshots	0.96	0.97	0.99	0.95	0.92	0.95	0.95	0.97	0.97	0.99	0.98	0.96
Correlation of u												
5 snapshots	0.79	0.94	0.98	0.85	0.92	0.89	0.95	0.97	0.95	0.94	0.90	0.99
20 snapshots	0.90	0.98	0.98	0.95	0.94	0.91	0.97	0.99	0.97	0.95	0.96	0.98
30 snapshots	0.95	0.98	0.98	0.92	0.91	0.93	0.97	0.95	0.94	0.98	0.93	0.96
Correlation of v												
5 snapshots	0.93	0.94	0.95	0.95	0.95	0.88	0.87	0.98	0.90	0.97	0.92	0.96
20 snapshots	0.94	0.99	0.89	0.66	0.91	0.90	0.91	0.99	0.94	0.98	0.99	0.91
30 snapshots	0.81	0.92	0.89	0.61	0.83	0.91	0.84	0.94	0.93	0.97	0.84	0.88

b. POD modes for a 99% captured energy

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Correlation of h												
5 snapshots	0.96	0.99	0.99	0.98	0.99	0.96	0.99	0.99	0.99	0.99	0.99	0.99
20 snapshots	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99
30 snapshots	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99
Correlation of u												
5 snapshots	0.84	0.96	0.99	0.92	0.99	0.92	0.98	0.97	0.93	0.98	0.94	0.99
20 snapshots	0.97	0.99	0.99	0.98	0.967	0.98	0.99	0.99	0.99	0.98	0.98	0.99
30 snapshots	0.97	0.98	0.99	0.99	0.97	0.97	0.99	0.99	0.98	0.99	0.98	0.98
Correlation of v												
5 snapshots	0.86	0.95	0.98.7	0.93	0.98	0.91	0.93	0.97	0.92	0.96	0.92	0.97
20 snapshots	0.85	0.88	0.91	0.82	0.86	0.79	0.88	0.95	0.98	0.95	0.91	0.93
30 snapshots	0.81	0.87	0.91	0.95	0.91	0.85	0.88	0.97	0.97	0.95	0.96	0.97

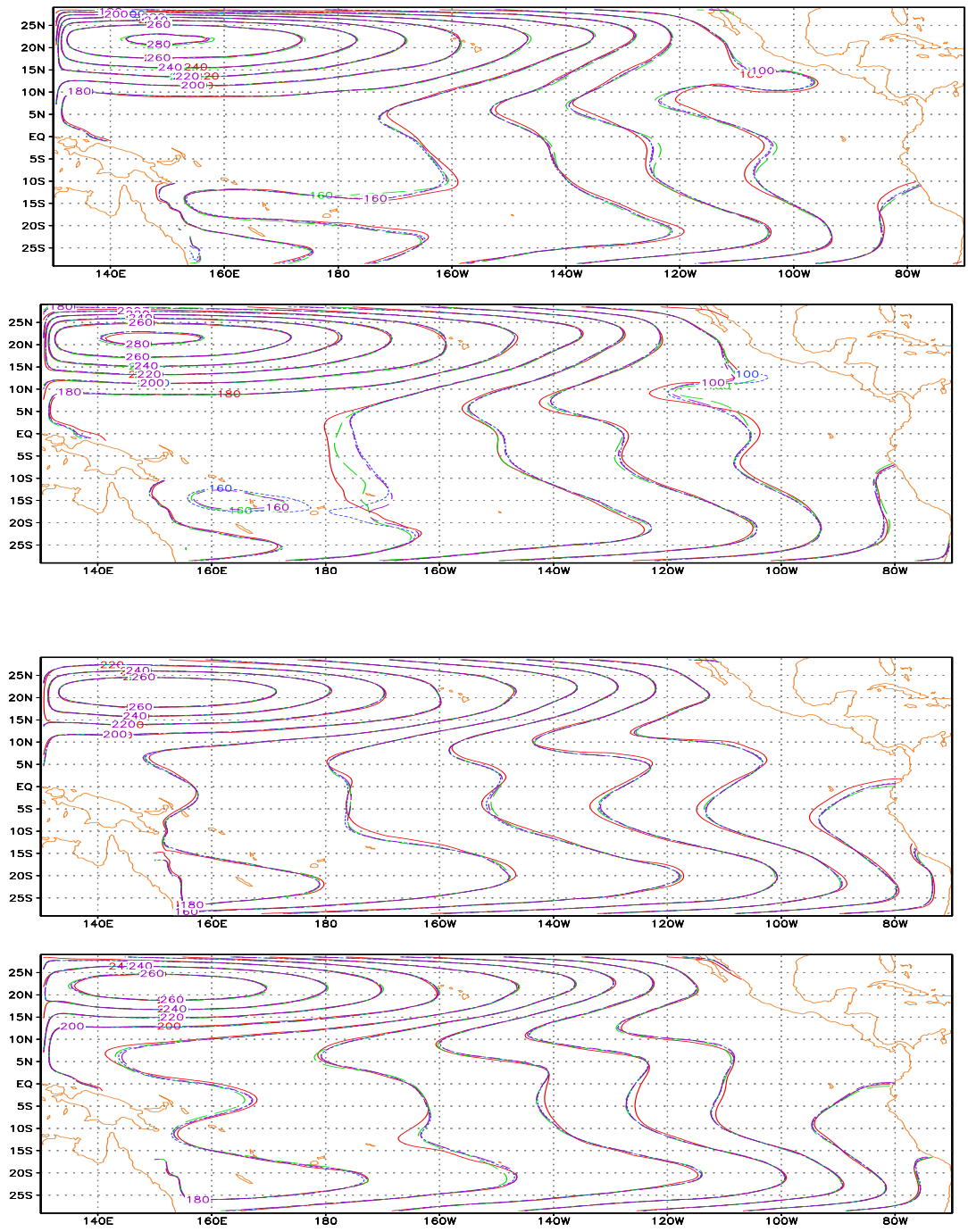


Figure 4. The case of 5 snapshots, 20 snapshots, 30 snapshots, for energy captured at 95% level, comparing the full model approximation and the reduced order approximation for upper layer thickness: (a) February, (b) May, (c) August, and (d) November. The red isolines represent full order approximation, the green long discontinuous isoline: 5 snapshots, the dark red short discontinuous isoline: 20 snapshots, the purple long and short discontinuous isoline: 30 snapshots, respectively.

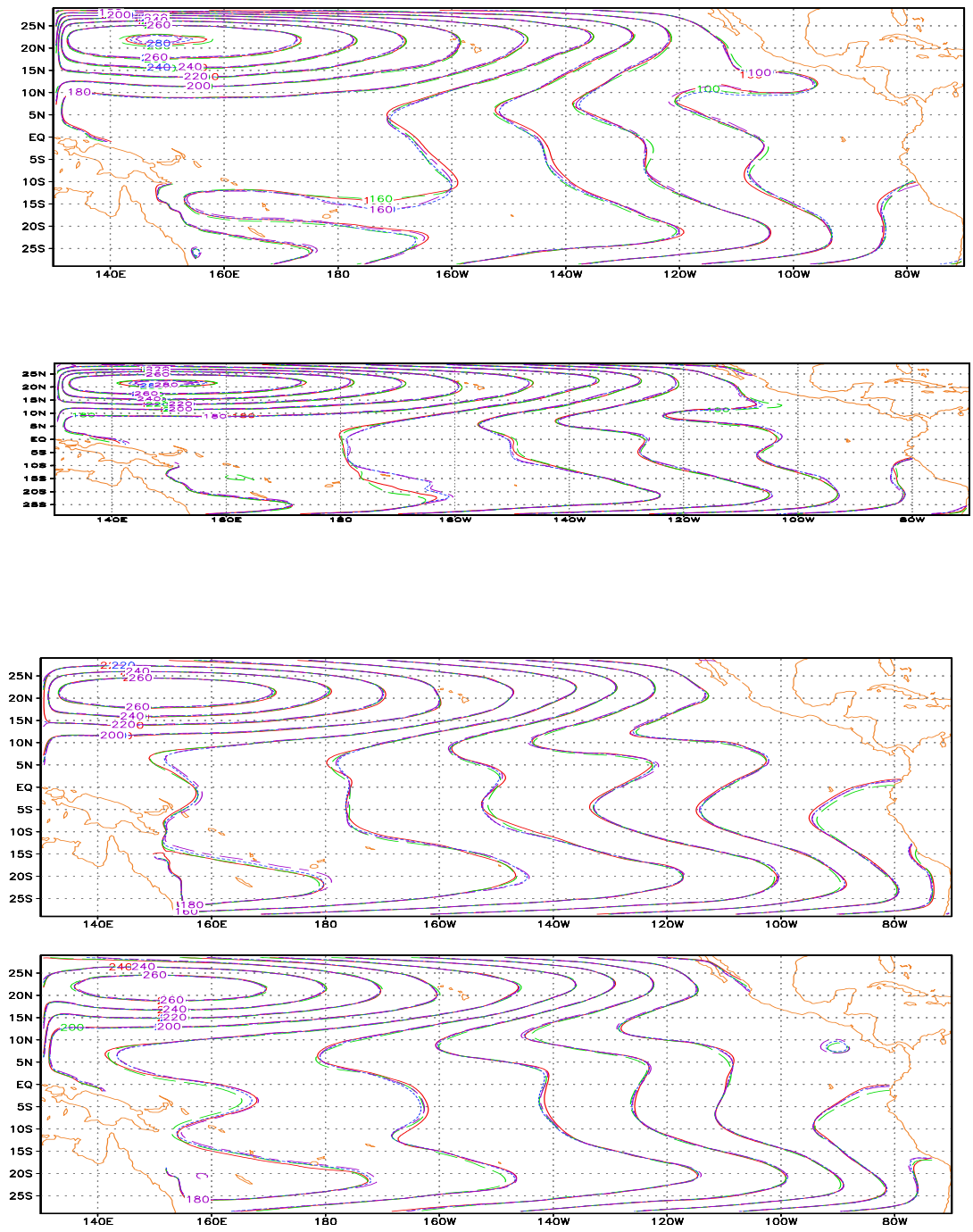


Figure 5. The case of 5 snapshots, 20 snapshots, 30 snapshots, for energy captured at 99% level, comparing the full model approximation and the reduced order approximations for upper layer thickness: (a) February, (b) May, (c) August, and (d) November. The red isolines represents full order approximation, the green long discontinuous isoline: 5 snapshots, the dark red short discontinuous isoline: 20 snapshots, the purple long and short discontinuous line: 30 snapshots, respectively.