Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER Computers and Mathematics with Applications 52 (2006) 1373–1386
An International Journal
computers & mathematics with applications
www.elsevier.com/locate/camwa

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

YANHUA CAO
Department of Mathematics, Capital Normal University
Beijing 100037, China
and
Institute of Atmospheric Physics, Chinese Academy of Sciences
Beijing, 100029, China

JIANG ZHU
Institute of Atmospheric Physics, Chinese Academy of Sciences
Beijing, 100029, China

ZHENDONG LUO
College of Fundamental Sciences North China University of Technology
Beijing, 100041, China

I. M. NAVON
School of Computational Science and Department of Mathematics
Florida State University, Tallahassee, FL 32306-4120, U.S.A.

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 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 22 are less than 1m that is less than 1% of the average thickness and the correlation 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.

Keywords—POD, Reduced order model, PDE, Galerkin methods, ODE.

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 Loève in 1945 (see [2]),

0898-1221/06/$ - see front matter Typeset by AMS-TEX
doi:10.1016/j.camwa.2006.11.012

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 [4–6]). Many researchers have also applied the POD technique to optimal control problems. For instance, this method has been used for Burgers' equation (see [7–9]), the Ginzburg-Landau equation and the Bénard convection (see [10]), and in other fluid control problems [11–17]). 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 [28–30]). 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 are 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 Section 2. The POD technique is briefly presented in Section 3. The issues of 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 Section 4.

2. MODEL OF UPPER TROPIC PACIFIC

2.1. Description of the Physical Model

The numerical model used in this paper is a 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,

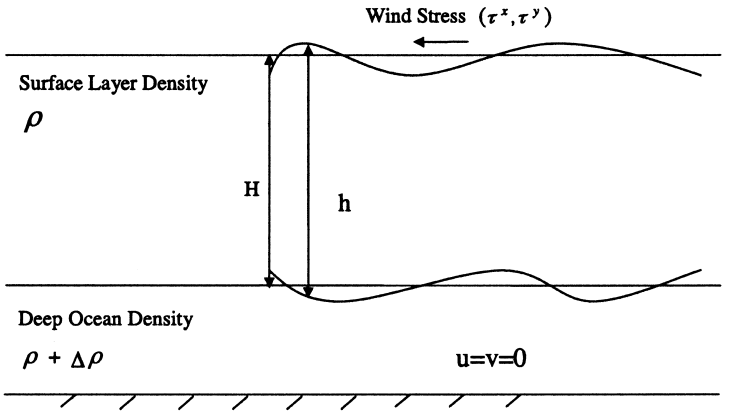

Figure 1. The vertical structure of the reduced-gravity model.

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

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

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

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

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 $(\tau^x, \tau^y) = \rho_a C_D |\sqrt{U_{\text{wind}}^2 + V_{\text{wind}}^2}(U_{\text{wind}}, V_{\text{wind}})$, where ρ_a is the density of the air; C_D the wind stress drag coefficient; and $(U_{\text{wind}}, V_{\text{wind}})$ 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 $\Delta x = \Delta y = 0.5°$ and the time step to be $\Delta t = 100\,\text{s}$. This temporal-spatial resolution will allow resolving 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$^{\text{st}}$ 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 10$^{\text{th}}$ time step to eliminate the computational mode. Every integration day a mass-compensation is carried out.

Table 1. The values of the model parameters used in the 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	$1.2\,\mathrm{kg \cdot m^{-3}}$	Density of air
ρ_0	$1025\,\mathrm{kg \cdot m^{-3}}$	Density of seawater
A	$750\,\mathrm{m^2 \cdot sec{-}1}$	Coefficient of horizontal viscosity
α	2.5×10^{-5}	Coefficient of bottom friction

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:

$$\vec{u}_i = [\vec{u}_{i1}, \vec{u}_{i2}, \ldots, \vec{u}_{im}]^{\top} \tag{3.1}$$

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

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

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

$$\vec{v}_{ik} = \vec{v}_i(\vec{x}_k) = \vec{u}_{ik} - \bar{u}_k, \qquad k = 1, \ldots, m. \tag{3.3}$$

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

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

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

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

subject to

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

Let matrix A denote the new ensemble

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

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

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

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 get 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,

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

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 $\phi_k = Aw_k/\sqrt{\lambda_k}$. In matrix form, $\Phi = AW$, where $\Phi = [\phi_1, \phi_2, \ldots, \phi_n]$, $W = [w_1, w_2, \ldots, w_n]$. It is shown that the cost functional

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\left(\vec{v}_{ik}\Phi\left(\vec{x}\right)\right)^2 = \lambda,$$

which is maximized when the coefficients a_is of (3.4) are the element 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 station. 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 $\vec{u}_i$ may contain all 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 $k \ll n$ then carry out a nonlinear Galerkin procedure with the set of elements $\{\Phi_1, \Phi_2, \ldots, \Phi_k\}$. How to choose the values of n and k is 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 $M \ll n$ 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

$$M = \operatorname{argmin}\{I(k) : I(k) \geq \gamma\},$$

where $0 \leq \gamma \leq 1$ is the percentage of total information captured by the reduced space $D^M = \operatorname{span}\{\Phi_1, \Phi_2, \ldots, \Phi_M\}$.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 $k = 3$ yields a ratio of about 0.99; 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 Section 2 to apply the POD technique. Therefore, solve equations (2.1a)–(2.1c) for the steady state solutions of upper layer thickness and velocity field after 20 years time integration. The 21^{st} year results are depicted graphically in Figure 2.

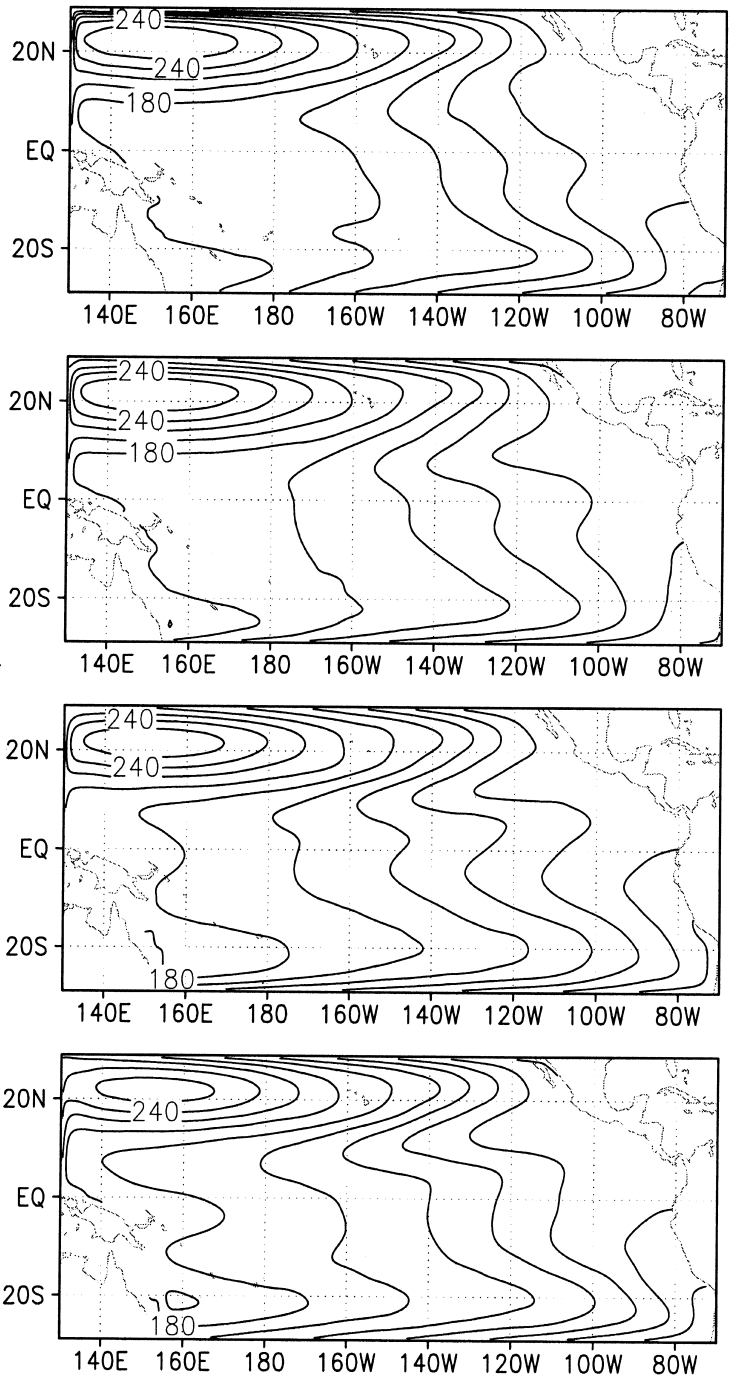

Figure 2. Upper layer thickness of full order approximation in February, May, August, and November.

4.1. Construction of POD Basis Vectors

We compute the POD reduced order spaces $X_h^{\mathrm{POD}}, X_u^{\mathrm{POD}}, X_v^{\mathrm{POD}}$ as the following 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) $\{h_1(\vec{x}), h_2(\vec{x}), \ldots, h_n(\vec{x}); u_1(\vec{x}), u_2(\vec{x}), \ldots, u_n(\vec{x}); v_1(\vec{x}), v_2(\vec{x}), \ldots, v_n(\vec{x})\}$ 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 $D_h = A_h^\top A_h, D_u = A_u^\top A_u, D_v = A_v^\top A_v$ which is depicted in Section 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 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

$$
\begin{aligned}
X_h^{\mathrm{POD}} &= \operatorname{span}\{\Phi_{h1}(\vec{x}), \Phi_{h2}(\vec{x}), \ldots, \Phi_{hn}(\vec{x})\}, \\
X_u^{\mathrm{POD}} &= \operatorname{span}\{\Phi_{u1}(\vec{x}), \Phi_{u2}(\vec{x}), \ldots, \Phi_{un}(\vec{x})\}, \\
X_v^{\mathrm{POD}} &= \operatorname{span}\{\Phi_{v1}(\vec{x}), \Phi_{v2}(\vec{x}), \ldots, \Phi_{vn}(\vec{x})\},
\end{aligned}
$$

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, respectively.

4.2. Reconstruction of Solutions through POD Basis Vectors

Since the scales in model variables h, u, v are not uniform, thus 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

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

Once the coefficients β_i^h $(i = 1, \ldots, n_h)$; β_i^u $(i = 1, \ldots, n_u)$; β_i^v $(i = 1, \ldots, n_v)$ have been obtained, then substituting $h(\vec{x}, t), u(\vec{x}, t), v(\vec{x}, t)$ into equations (2.1a)–(2.1c) and multiplying by ϕ_{hi} $(i = 1, \ldots, n_h)$; ϕ_{ui} $(i = 1, \ldots, n_u)$; ϕ_{vi} $(i = 1, \ldots, n_v)$ and integrating respectively in terms of $\vec{x}$. Since the basis functions are orthonormal, the system of ODE is as follows

$$
\begin{aligned}
\frac{\partial \beta_j^h(t)}{\partial t} &= f_1\left(t, \beta_1^h(t), \ldots, \beta_{n_h}^h(t); \beta_1^u(t), \ldots, \beta_{n_u}^u(t); \beta_1^v(t), \ldots, \beta_{n_v}^v(t)\right), \qquad j = 1, \ldots, n_h, \\
\frac{\partial \beta_j^u(t)}{\partial t} &= f_2\left(t, \beta_1^h(t), \ldots, \beta_{n_h}^h(t); \beta_1^u(t), \ldots, \beta_{n_u}^u(t); \beta_1^v(t), \ldots, \beta_{n_v}^v(t)\right), \qquad j = 1, \ldots, n_u, \\
\frac{\partial \beta_j^v(t)}{\partial t} &= f_3\left(t, \beta_1^h(t), \ldots, \beta_{n_h}^h(t); \beta_1^u(t), \ldots, \beta_{n_u}^u(t); \beta_1^v(t), \ldots, \beta_{n_v}^v(t)\right), \qquad j = 1, \ldots, n_v,
\end{aligned}
$$

along with the initial condition

$$\beta_i^h(0) = \left(h\left(\vec{x},0\right) - \bar{h}\left(\vec{x}\right), \phi_{hi}\left(\vec{x}\right)\right), \qquad i = 1,\ldots,n_h,$$
$$\beta_i^u(0) = \left(u\left(\vec{x},0\right) - \bar{u}\left(\vec{x}\right), \phi_{ui}\left(\vec{x}\right)\right), \qquad i = 1,\ldots,n_u,$$
$$\beta_i^v(0) = \left(v\left(\vec{x},0\right) - \bar{v}\left(\vec{x}\right), \phi_{vi}\left(\vec{x}\right)\right), \qquad i = 1,\ldots,n_v.$$

Solving the above ODE problems using a finite difference scheme, one can obtain the reconstructed solutions.

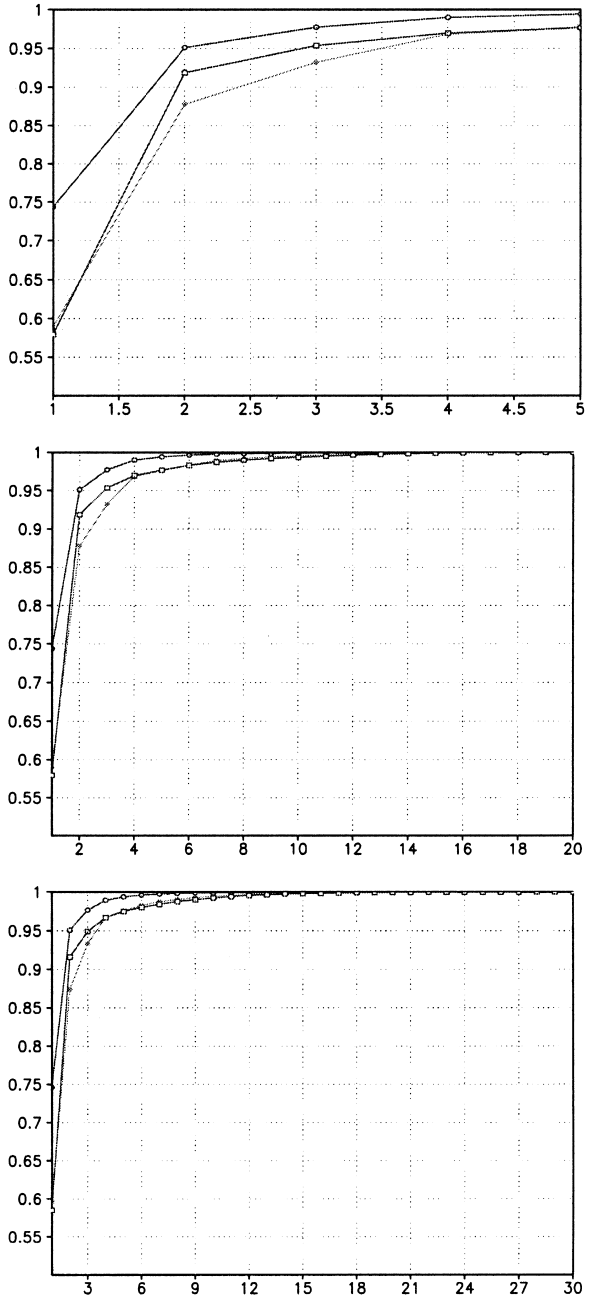

Figure 3. The POD modes capture of energy in case of five snapshots, 20 snapshots, and 30 snapshots; rhombus line: upper layer thickness h (m), triangle line: zonal current velocity u (m/s), and star line: meridional current velocity v (m/s).

Table 2. RMSE of five snapshots, 20 snapshots, and 30 snapshots, respectively, for different percentages of captured energy.

(a). Upper layer thickness h (unit: m).

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

(b). The zonal current velocity u (unit: m/s).

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

(c). The meridional current velocity v (unit: m/s).

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

4.3. Numerical Results

In this section, we report 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 also could be seen clearly, 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, and v, 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 twelve-month's full order results and the corresponding twelve-month's reduced order results and computing the error, for example, for variable u it yields

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

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

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

the correlation as:

$$\mathrm{CORRELATION}_{um} = \frac{\sum_{i=1}^{M} \left(\hat{u}_m(z_i) - \bar{\hat{u}}_m(z_i)\right)\left(u_m(z_i) - \bar{u}_m(z_i)\right)}{\sqrt{\sum_{i=1}^{M} \left(\hat{u}_m(z_i) - \bar{\hat{u}}_m(z_i)\right)^2 \sum_{i=1}^{M} \left(u_m(z_i) - \bar{u}_m(z_i)\right)^2}},$$

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 h and v.

Table 3. Correlation of five snapshots, 20 snapshots, and 30 snapshots for upper layer thickness h (unit: m), zonal current velocity u (unit: m/s) and meridional current velocity v (unit: m/s), (a) energy captured 95%, and (b) energy captured 99

(a). POD modes for a 95% captured energy.

Correlation of h	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
5 snapshots	0.948	0.969	0.986	0.960	0.934	0.931	0.940	0.965	0.968	0.986	0.980	0.978
20 snapshots	0.959	0.970	0.985	0.950	0.913	0.942	0.950	0.968	0.971	0.988	0.982	0.965
30 snapshots	0.958	0.970	0.985	0.954	0.924	0.948	0.950	0.968	0.971	0.989	0.981	0.961

Correlation of u	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
5 snapshots	0.793	0.941	0.984	0.846	0.923	0.894	0.953	0.966	0.950	0.936	0.902	0.997
20 snapshots	0.901	0.980	0.980	0.951	0.944	0.912	0.968	0.986	0.974	0.954	0.962	0.979
30 snapshots	0.949	0.982	0.982	0.919	0.912	0.930	0.965	0.952	0.936	0.977	0.931	0.960

Correlation of v	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
5 snapshots	0.931	0.942	0.946	0.952	0.952	0.879	0.866	0.976	0.897	0.968	0.916	0.962
20 snapshots	0.940	0.990	0.890	0.951	0.944	0.899	0.907	0.985	0.943	0.975	0.988	0.910
30 snapshots	0.813	0.918	0.888	0.613	0.830	0.906	0.838	0.935	0.927	0.967	0.835	0.879

(b) POD modes for a 99% captured energy.

Correlation of h	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
5 snapshots	0.958	0.986	0.993	0.979	0.994	0.961	0.988	0.986	0.988	0.996	0.990	0.999
20 snapshots	0.987	0.993	0.996	0.993	0.989	0.980	0.987	0.993	0.996	0.996	0.994	0.995
30 snapshots	0.987	0.994	0.996	0.994	0.992	0.982	0.985	0.993	0.996	0.996	0.995	0.993

Correlation of u	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
5 snapshots	0.842	0.962	0.985	0.921	0.988	0.922	0.976	0.970	0.934	0.975	0.936	0.992
20 snapshots	0.973	0.986	0.990	0.984	0.969	0.979	0.990	0.989	0.990	0.984	0.984	0.989
30 snapshots	0.971	0.984	0.985	0.988	0.967	0.973	0.990	0.992	0.982	0.987	0.978	0.984

Correlation of v	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
5 snapshots	0.858	0.953	0.987	0.926	0.977	0.905	0.934	0.972	0.917	0.963	0.922	0.972
20 snapshots	0.845	0.876	0.905	0.823	0.857	0.792	0.882	0.952	0.975	0.952	0.908	0.930
30 snapshots	0.806	0.871	0.912	0.954	0.909	0.847	0.881	0.973	0.965	0.950	0.956	0.969

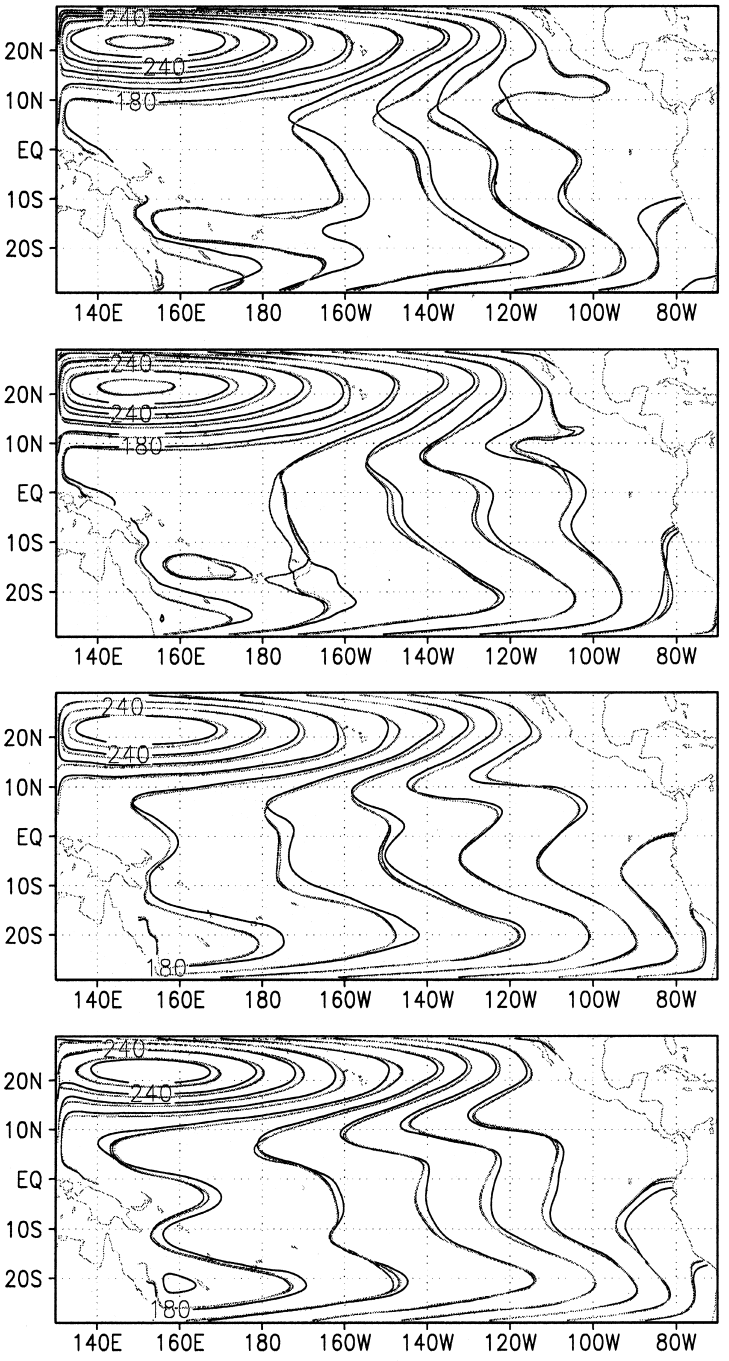

Figure 4. Upper layer thickness in February, May, August and November in case of five snapshots, 20 snapshots, 30 snapshots, energy capture 95%, the full model approximation and the reduced order approximation. Black isoline: full order approximation, red isoline: five snapshots, green isoline: 20 snapshots, blue isoline: 30 snapshots.

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, with the span of the reduced basis space increasing, the RMSE decreases for the same number of snapshots it retained. On the other hand, for different number of snapshots but for the same

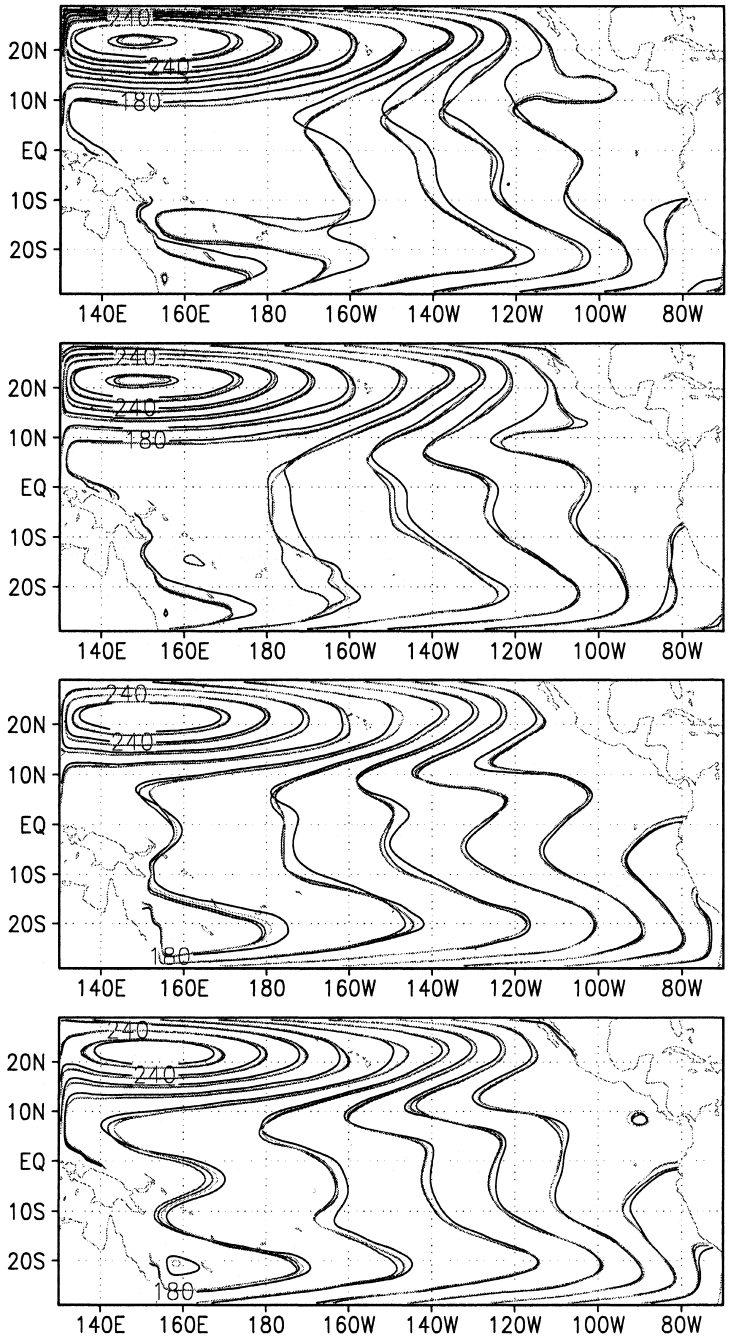

Figure 5. Upper layer thickness in February, May, August, and November in case of five snapshots, 20 snapshots, 30 snapshots, energy capture 99%, the full model approximation and the reduced order approximation. Black isoline: full order approximation, red isoline: five snapshots, green isoline: 20 snapshots, blue isoline: 30 snapshots.

energy 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. The increment stops at 30 snapshots (Table 3) where the best approximation obtained with 20 snapshots produced a correlation of the same level as the approximation 20 snapshots.

However one must also note that a simple linear independence is not a sufficient criterion when 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 4 for a retained energy of 95% respectively about upper layer thickness. From these figures, we can see there is little improvement between either 20 snapshots or 30 snapshots compared to five snapshots, but there is almost no difference between 20 snapshots and 30 snapshots. The contrast between the full order approximation and numerical results using energy captured 99% for different snapshots about upper layer thickness is displayed in Figure 5 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 taken snapshots and for different percentages of energy captured. From the computational cost and memory storage aspects, 20 snapshots and the energy captured 99% 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 $\Delta x = \Delta y = 0.5$ and a time step of $\Delta t = 100\,\mathrm{s}$. 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 as follows.

- 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 seven 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. K. Karhunen, Zur Spektraltheorie Stochastischer Prozesse, *Ann. Acad. Sci. Fennicae* **37**, (1946).
2. M. Loève, Fonctions aleatoires 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.* **458**, 181–190, (2002.
4. P. Holmes, J.L. Lumley and G. Berkooz, 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* **328**, 233–250, (2003.
6. L. Sirovich, Turbulence and the dynamics of coherent structures, Parts I–III., *Q. Appl. Math.* **XLV** (3), 561–590.
7. J.A. Atwell, J.T. Borggaard and B.B. King, 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-Loève expansion of Burgers' model of turbulence, *Phys. Fluids* **31**, 2573–2582, (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* **37**, 126–145, (1989).
11. K. Afanasiev and M. Hinze, Adaptive control of a wake flow using proper orthogonal decomposition, In *Lect. Notes Pure Appl. Math., Volume 216*, pp. 317–332, (2001).
12. K. Afanasiev and M. Hinze, Entwicklung von Feedback-Controllern zur Beeinflussung abgelöster Strömungen. Abschlussbericht des Teilprojekts C3 des Sonderforschungsbereichs TP A4, SFB 557, TU Berlin, (2000).
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. H.V. Ly and H.T. Tran, Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor, *Quarterly of Applied Mathematics* **60** (4), 631–656, (2002).
17. S.S. Ravindran, 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. H.T. Banks, M.L. Joyner, B. Winchesky and W.P. Winfree, 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.* **41**, 866–877, (2003).
20. S.S. Ravindran, 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.* **52**, 227–247, (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.* **71**, 631–646, (1997).
23. M.D. Gunzburger, *Perspectives in Flow Control and Optimization*, pp. 261, Society for Industrial and Applied Mathematics, Philadelphia, PA, (2003).
24. M.D. Gunzburger, Reduced-order Modeling, data compression, and the design of experiments, Second DOE Workshop on Multiscale Mathematics, July 20–22, Broomfield, Colorado, (2004).
25. B. Uzunoglu, S.J. Fletcher, I.M. Navon and M. Zupanski, Adaptive ensemble size reduction, *Quarterly Journal of the Royal Meteorological Society*, (submitted) (2005).
26. P. Courtier, J.-N. Thépaut and A. Hollingsworth, A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society* **120**, 1367–1388, (1994).
27. A. Griewank, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation Frontiers in Applied Mathematics*, SIAM, (2000).
28. Z. Li, I.M. Navon and Y.Y. Zhu, 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. P. Gauthier, Operational implementation of variational data assimilation. In Data Assimilation for the Earth System, *NATO Science Series IV. Earth and Environmental Sciences* **26**, 167– 176, (2003).
30. Y. Tremolet, Diagnostics of linear and incremental approximations in 4D-Var, *Q. J. R. Meteorol. Soc.* **130** (601), 2233–2251, (2004).
31. J. Restrepo, G. Leaf and A. Griewank, Circumventing storage limitations in variational data assimilation studies, *SIAM J. Sci. Comput.* **19**, 1586–1605, (1998).
32. Y. Cao, J. Zhu, I.M. Navon and Z. Luo, A reduced order approach to four dimensional variational data assimilation using proper orthogonal decomposition, *Journal for Numerical Methods in Fluids* (to appear); Published online DOI:10.1002/fld.1365.