

# AN EQUATION- FREE REDUCED ORDER MODELING APPROACH TO TROPIC PACIFIC SIMULATION

Ruiwen Wang<sup>a,b</sup> Jiang Zhu<sup>b</sup> Zhendong Luo<sup>c</sup> and Navon, I.M.<sup>d</sup>

<sup>a</sup> Department of Mathematics, Capital Normal University, Beijing, 100037, China

<sup>b</sup> Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

<sup>c</sup> School of Science, Beijing Jiaotong University, Beijing, 100044, China

<sup>d</sup> School of Computational Science and Department of Mathematics,  
Florida State University, Tallahassee, FL 32306-4120, USA.

## Abstract

The “equation-free” (EF) method is based on a procedure, which is often used in complex, multi-scale problems, given an initial macroscopic variable, by lifting, evolve in Direct Numerical Simulations (DNS), restriction, and projective integration. In this paper, we apply the Equation-free method to the reduced modeling of a large-scale upper ocean circulation in the tropic Pacific domain. We carry out a series of experiments to discuss its convergence and error, and we also discuss some factors that affect the results, such as the number of snapshots and basis functions based on proper orthogonal decomposition (POD) mode, large-scale time step, short time step and the number of iterations during the course of the DNS running in a large-scale time step. Compared with POD method, we do not need availability of explicit equation or right-hand-side (RHS), and we may obtain better results for horizontal velocity components of depth-averaged currents than those obtained in POD case. The results from the equation-free based on different POD modes are compared with results of POD and DNS. We also illustrate the convergence and error of the EF method. The main findings are: we reduce the computation work enormously in contrast to DNS, for instance, the computational cost of the equation-free POD-assisted method is less than 10% of that for DNS. The method can capture the main variability by a low dimensional system based on equation-free POD mode. We also attained a RMS error for the upper ocean layer thickness that is less than 1m i.e. less than 1% of the average thickness and the correlations between the upper layer thickness are around 0.99 based on equation-free POD method, while the convergence and the error are illustrated through numerical experiments.

**Key words:** multi-scale problem, equation-free, PDE.

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

Corresponding author. Tel.: +1 850 644 6560; fax: +1 850 644 0098.

E-mail address: [navon@csit.fsu.edu](mailto:navon@csit.fsu.edu) (I.M. Navon).

## 1 Introduction

The Proper orthogonal decomposition (POD) is a methodology that constitutes a very efficient way to perform reduced modeling by identifying the few most energetic modes in a sequence of snapshots from a time-dependent system, and subsequently providing a means of obtaining a low-dimensional description of the system's dynamics [1]. The method of snapshots which is first proposed in [2] for flow system is a very effective and easy to carry out approach for obtaining POD basis sets. The proper orthogonal decomposition (POD) was originally introduced by Karhunen in 1946 (see [3]) and Loève in 1945 (see [4]), and the method has been extensively used in research in recent years and successfully applied to a variety of fields, such as in conjunction with experimental e.g., [5,6,7,8]) and with numerical studies (e.g., [9,10,11,12]) in thermal convection, shear layers, cavity flows and external flows, to mention just a few. In recent years, our understanding of the tropical ocean has increased. There is a vast and growing literature on the design of the ocean model based on partial differential equations (PDE) for physical systems. Such modes are often hard to solve because of the high order system that describe the state. Another obvious application of POD in weather forecasting and operational oceanography is the four-dimensional variational (4DVAR) data assimilation problem. However, a major difficulty 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 [13]), check-pointing (see [14]) and parallelization. However, each of these three methods has their obvious shortcomings. The dimension of the control space remains very large in realistic applications for the incremental method (see [15, 16]). The size of assimilation studies is imposed a limitation for memory storage requirements, even on the largest computers. Check-pointing strategies (see [17]) 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 technique that can dramatically reduce computation and memory burdens of 4D-VAR. Cao et al (see [18]) made an initial effort to explore the feasibility of application of POD to 4D-VAR. If someone was to apply the POD method to an ocean model, the feasibility and efficiency of the POD technique in ocean calculations are already demonstrated (e.g., [19]). In the aforementioned works, we can obtain low-dimensional system dynamics systems directly from the Galerkin projection of the governing equations on the POD modes. However, it is well known that the reduced systems resulting from truncated Galerkin projections may result in spurious asymptotic states (e.g., [20]), and it is difficult to obtain the explicit form of the right-hand-side (RHS) which consists of POD coefficients of the evolution equation. Thus, the Equation-free POD-assisted method is first used to resolve the question above in incompressible flows (see [21]). It is well known that the tropic Pacific Ocean model is more complicated than the molecular dynamics model, and it is very difficult to obtain accurate information due to the lack of direct measurements and insufficient knowledge of air-sea exchange processes. Since Ocean forecasting is very important to human activity, the topic we are investigating is more complicated and significant than the two-dimensional flow past a circular cylinder. Similarly, the Equation-free

method also can be used to resolve problems related to weather and land numerical simulation. In this paper, we apply the equation-free POD-assisted method to the simulation of the upper tropical Pacific Ocean model based on POD model. In this context, the role of the “detailed, microscopic” model is played by the numerical simulation (DNS) simulator, and the “unavailable coarse-grained” equations are the evolution equations about the coefficients of the projection of the full DNS solution on master POD modes. We often obtain these equations by performing a Galerkin projection of the original equations on the POD basis; to truncate this Galerkin projection would give an approximation of these equations. The equation-free method tries to resolve this model without deriving it in closed form. We use full DNS simulator to estimate the right-hand-side of this Galerkin model, and the equation-free acceleration technique we will illustrate is projective integration.

In this paper, we make use of an Equation-free POD-assisted method, which use “equation-free” [22, 23, 24] projective integration [25, 26, 27]. The equation-free procedure is devised for the efficient computational study of complex, multi-scale problems. The basic process is as follows (two level)[21] (see Figure 1): (a) to devise and implement short-time numerical experiments with “the best available “ microscopic model, and subsequently (b) to estimate quantities (derivative) required in numerical computation of the (unavailable) macroscopic equations for the coarse-grained system behavior [28,29] by using the numerical results of such microscopic computations. Thus, we can estimate the closures that are required to obtain explicit macroscopic equations on demand; we can perform numerical analysis tasks by running the microscopic simulation directly. This framework has been applied to many types of problems, such as bifurcation analysis of complex systems and homogenization of random media [22, 23, 25, 28, 30, and 31].

The paper is organized as follows: In section 2 the upper tropical Pacific Ocean model is described. The POD technique and its mathematical properties and equation-free POD model are presented in section 3. In section 4, the numerical calculations using equation-free POD in the context of simulating the upper layer thickness and currents in this ocean model, and a comparison with DNS and POD are discussed. In section 5 we discuss the convergence and error of the method. We summarize the results in section 6.

## 2 Reduced-gravity Model of Upper Tropic Pacific

### 2.1 Description of the physical model

A reduced-gravity model with a constant-depth surface layer is used in this paper (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 2). 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 not 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

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

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

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

where  $(u, v)$  are the horizontal velocity components of the depth-averaged currents;  $h$  the total layer thickness;  $f$  the Coriolis force;  $\hat{H}$  the mean depth of the layer;  $\rho_0$  the density of water; and  $A$  the horizontal eddy viscosity coefficient and  $\alpha$  is the friction coefficient. The wind stress is calculated by the aerodynamic bulk formula

$$(\tau^x, \tau^y) = \rho_a c_D \sqrt{U_{wind}^2 + V_{wind}^2} (U_{wind}, V_{wind}),$$

Where  $\rho_a$  is the density of the air,  $c_D$  the wind stress drag coefficient; and  $(U_{wind}, V_{wind})$  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  $29^\circ \text{S} \sim 29^\circ \text{N}$ ,  $120^\circ \text{E} \sim 70^\circ \text{W}$ . 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 well allows 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-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. Computational Formulation of Equation-free Method

### A Simple Introduction to Proper Orthogonal Decomposition

In order to illuminate the idea of the proper orthogonal decomposition method, we will introduce the POD method in continuous case. Because the main idea for both continuous and discrete cases is same, we carry out the numerical experiments in the discrete case.

Let  $U_i(\vec{x}) (i = 1, 2, \dots, n)$  denote the set of  $n$  observations (also called snapshots) of some physical process taken at position  $\vec{x} = (x, y)$ . The average of the ensemble of snapshots is given by

$$\bar{U} = \langle U \rangle = \frac{1}{n} \sum_{i=1}^n U_i(\vec{x}). \quad (3.1)$$

We form new ensemble by focusing on deviations from the mean as follows:

$$V_i = U_i - \bar{U}. \quad (3.2)$$

We wish to find an optimal compressed description of the sequence of data (3.2). One description of the process is a series expansion in terms of a set of basis functions. Intuitively, the basis functions should in some sense be representative of the members of the ensemble. Such a coordinate system, which possesses several optimality properties (to be discussed in the sequel), is provided by the Karhunen-Loève expansion (see [35]), where the basis functions  $\Phi$  are, in fact, admixtures of the snapshots and are given by

$$\Phi = \sum_{i=1}^n \eta_i V_i(\vec{x}), \quad (3.3)$$

where the coefficients  $\eta_i$  are to be determined such that  $\Phi$  given by (3.3) will most resemble the ensemble  $\{V_i(\vec{x})\}_{i=1}^n$ . More specifically, POD seeks a function  $\Phi$  such that

$$\frac{1}{n} \sum_{i=1}^n |(V_i, \Phi)|^2, \quad (3.4)$$

subject to

$$(\Phi, \Phi) = \|\Phi\|^2 = 1 \quad (3.5)$$

is minimized, where  $(\cdot, \cdot)$  and  $\|\cdot\|$  denote the usual  $L^2$ -inner product and  $L^2$ -norm, respectively.

It follows that (see, [36]) the basis functions are the eigenfunctions of the integral equation

$$\int_{\Omega} C(\vec{x}, \vec{x}') \Phi(\vec{x}') d\vec{x}' = \lambda \Phi(\vec{x}), \quad (3.6)$$

where the kernel is given by

$$C(\vec{x}, \vec{x}') = \frac{1}{n} \sum_{i=1}^n V_i(\vec{x}) V_i(\vec{x}'). \quad (3.7)$$

Substituting (3.3) into (3.6) yields the following eigenvalue problem:

$$\sum_{i=1}^n L_{i,j} \eta_j = \lambda \eta_i, \quad (3.8)$$

where  $L_{ij} = \frac{1}{n}(V_i, V_j)$ ,  $L = (L_{ij})_{n \times n}$  is a symmetric and nonnegative matrix. Thus, we see that our problem amounts to solving for the eigenvectors of a  $n \times n$  matrix where  $n$  is the size of the ensemble of snapshots. A straightforward calculation (see also [36]) shows that the cost functional

$$\frac{1}{n} \sum_{i=1}^n |(V_i, \Phi)|^2 = (\lambda \Phi, \Phi) = \lambda, \quad (3.9)$$

which is maximized when the coefficients  $\eta_i$  ( $i = 1, 2, \dots, n$ ) of (3.8) are the elements of the eigenvector corresponding to the largest eigenvalue of  $L$ .

The non-negative definiteness of correlation (3.9) assures that  $\lambda \geq 0$  and we order the eigenvalues by  $\lambda_i \geq \lambda_{i+1}$ . After determining the POD basis  $\Phi_k$ ,  $k = 1, 2, \dots, n$ , the POD expansion coefficients can be determined by

$$a_k(t) = (V(t, X), \Phi_k(X)).$$

By denoting  $\mathbf{a}(t) = \{a_k(t)\}$  (see [21]), we define a *restriction operator*  $\Upsilon$  such that

$$\mathbf{a}(t) = \Upsilon V(t, X) \equiv \{(V(t, X), \Phi_k(X)), t \in A, \forall k\}, \quad (3.10)$$

and a *lifting operator*  $\Psi$  such that

$$V(t, X) = \Psi \mathbf{a}(t) \equiv \sum_k a_k(t) \Phi_k(X), t \in R^+ \quad (3.11)$$

Clearly, both  $\Upsilon$  and  $\Psi$  are linear operators, and  $\Upsilon \Psi \mathbf{a} = \mathbf{I} \mathbf{a} \equiv \mathbf{a}$ , where  $\mathbf{I}$  is unit operator. The evolution of the POD coefficients  $\mathbf{a}(t)$  is governed by a process

$$\frac{d\mathbf{a}}{dt}(t) = \mathbf{Y}(t; \mathbf{a}(t)), \quad (3.12)$$

Where the explicit form of right-hand-side (RHS) terms are unknown. In the normal truncated POD-Galerkin procedure, the RHS terms are derived from the equation (2.1a)-(2.1c) by a Galerkin approach, resulting in a (possibly large) set of coupled ODEs.

In this section, in order to perform the numerical experiments, we consider the discrete Karhunen-Loève expansion to find an optimal representation of the ensemble of snapshots (e.g., [1], [10], [19], and [37]).

### 3.1 The Equation-free POD Model

Equation-free method is usually used to resolve different scale problems between the macroscopic level and microscopic level. The whole method is often made of two levels, that is, “inner” simulator and “outer” simulator. The “inner” simulator is micro direct numerical simulation, and the “outer” simulator consists of many types of continuous mathematic methods at macro level, such as finite difference, finite element, finite volume element, optimization. How to link between macro scale and micro scale is the key problem. Here, we often have some steps as described below.

Full DNS is used to *estimate* the right hand side (RHS) of these Galerkin ODEs on demand in our approach, and we can accelerate their numerical integration, without approximating them in closed form. In this paper, the equation-free implementation of a numerical task is illustrated: numerical integration, which is called “projective integration” in an equation-free context. We will provide a detailed description for every step of the method.

Equation-free single step projective integration consists of the following main components, from  $T = T^n$  to  $T = T^{n+1}$ ,

(a)  $m_f \geq 1$  steps of micro level (short time step) computation (at time step  $\Delta t$  where

$$\Delta T_f = m_f \Delta t \text{ and } T_c^n = T^n + \Delta T_f).$$

(b) **Restriction** to macro variables and **estimation** of the time-derivatives of their evolution;

(c) One step of macro level projective integration with step size  $\Delta T_c$ ;

(d) **Lifting** from the projected values of the macro variables to *consistent* micro level initial conditions.

The large (coarse) step  $\Delta T_c \geq \Delta t$  will be usually chosen to be  $\Delta T_c = m_c \Delta t$  where  $m_c \geq 1$ .

The global time step is  $\Delta T = T^{n+1} - T^n = \Delta T_f + \Delta T_c = (m_f + m_c) \Delta t$ . Figure 2 presents a graphical illustration of the notation.

In our case, the “inner” simulator is the fully resolved DNS discretization for tropic Pacific model. The “outer” (coarse) model is the unavailable in closed form Galerkin sets of ODEs that are made up of coefficients of the evolution equations based on the first few low-POD modes.

Specially, the process of POD-assisted projective integration is made up of the following steps (see Figure 2): Given  $\mathbf{a}^n \equiv \mathbf{a}(T^n)$ ,

1. **Lifting**: at  $T = T^n$ , obtain  $V(T^n, X) = \Psi \mathbf{a}(T^n)$ .

2. **Micro scale computation**: Resolve the equations (2.1a)-(2.1c) by DNS for a short period of time, for  $T^n \leq T \leq T_c^n = T^n + m_f \Delta t$ . We can compute this by appropriate difference scheme with a small time step  $\Delta t$ .

3. **Restriction**: compute the POD coefficients  $\mathbf{a}(T) = \Upsilon V(T, X)$  for  $T^n \leq T \leq T_c^n$ , and estimate the time derivative  $\frac{d\mathbf{a}}{dt}$  at  $T = T_c^n$ .

4. **Projective integration**: Integrate equation (3.12) to  $T^{n+1}$  by standard ODE techniques to obtain  $\mathbf{a}(T^{n+1})$ . The time step here is  $\Delta T_c \equiv m_c \Delta t = T^{n+1} - T_c^n$ , where  $m_c \geq m_f \geq 1$ .

5. Return to step 1 until the final integration time is reached.

**Remark:** As discussed above, we can obtain the right-hand-side  $\mathbf{Y}(t; \mathbf{a}(t))$  of equations (3.12) from the equations (2.1a)-(2.1c) by a Galerkin projection. This procedure may result in rather intricate forms and often suffers long-term dynamics, (see [20, 32]). This “equation-free” approach can avoid these difficulties by using “just enough” full DNS simulation due to its not

requiring the explicit form of the RHS of (3.12). After having already performed a short DNS run, the procedure is as follows:

Obtain the POD variable time series  $\mathbf{a}(T) = \mathbf{Y}^T V(T, X)$  for  $T^n \leq T \leq T_c^n = T^n + m_f \Delta t$ .

Approximate the RHS of (3.12) at  $T = T_c^n$ , here through  $\mathbf{Y}(T_c^n) = \frac{d\mathbf{a}}{dt}(T_c^n + O(\Delta t^\eta))$ . So we only

estimate the time derivative  $\frac{d\mathbf{a}}{dt}(T_c^n)$ , we can do it by  $\frac{\mathbf{a}^{k+1,N} - \mathbf{a}^{k,N}}{\Delta t}$ , here

$T_c^n = T(N \Delta T + T_k)$ ,  $T_k = k \Delta t$ . In our case,  $\eta = 1$ .

Once the RHS of (3.12) is estimated numerically, one can integrate it through Euler difference scheme.

$$\mathbf{a}^{N+1} = \mathbf{a}^{k,N} + (\Delta T - T_k) \frac{\mathbf{a}^{k+1,N} - \mathbf{a}^{k,N}}{\Delta t}. \quad (3.13)$$

Therefore, after the steps above, we obtain what we want, that is,  $\mathbf{a}^{N+1}$ , repeatedly.

#### 4 Numerical Results

In this section, we present numerical results of the equation-free POD model for the upper tropical Pacific Ocean model. Here we select the three sets of different snapshots for discussing the effect of the number of snapshots and basis functions, that is  $n=20$ ,  $n=36$ , and  $n=60$ . For all of them, the first seven POD modes can capture about 99% energy. It can also be clearly seen, that for the upper layer thickness  $h$ , the same modes capture the most energy, followed next by  $u$  and last by  $v$ . Thus, different POD modes may be used to reconstruct  $h, u, v$  respectively. To quantify the performance of the reduced basis method, we use two metrics, namely the root mean square error (RMSE) and the correlation of the difference between the full order and the reduced order simulation. We can obtain it by first taking twelve-month's full order results and the corresponding twelve-month's reduced order results. We made a comparison between POD, equation-free POD and full DNS simulation by RMSE. The formulation of computing the error is as follows:

$$RMSE_{hm} = \sqrt{\frac{1}{M} \sum_{i=1}^M |\tilde{h}_m(z_i) - h_m(z_i)|^2},$$

where  $M$  is the number of nodal points, the index  $m$  denotes the month,  $\tilde{h}_m$  is the full order approximation and  $h_m$  is the reduced order approximation. The average RMS error is defined as:

$$RMSE_h = \frac{1}{12} \sum_{m=1}^{12} E_{hm} = \frac{1}{12} \sum_{m=1}^{12} \sqrt{\frac{1}{M} \sum_{i=1}^M |\tilde{h}_m(z_i) - h_m(z_i)|^2}.$$

And the correlation is defined as:

$$CORRELATION_{hm} = \frac{\sum_{i=1}^M (\tilde{h}_m(z_i) - \bar{\tilde{h}}(z_i))(h_m(z_i) - \bar{h}(z_i))}{\sqrt{\sum_{i=1}^M (\tilde{h}_m(z_i) - \bar{\tilde{h}}(z_i))^2 \cdot \sum_{i=1}^M (h_m(z_i) - \bar{h}(z_i))^2}},$$

where  $\bar{\tilde{h}}$  and  $\bar{h}$  is the average of full order approximation and reduced order approximation respectively. The RMS error and the correlation for the other model variables  $v$  and  $u$  are computed similarly. Table 2 presents the average RMSE in reduced order approximations using different modes namely  $n = 20$ ,  $n = 36$ , and  $n = 60$  snapshots. Note that from these simulations, with the span of the reduced basis space increasing, the RMSE decreases as long as the same number snapshots are used. Compared with POD, equation-free POD results in a bit of improvement, especially for  $v$  by capturing 60% energy. We also find that the RMSE continues to decrease with more energy being captured in equation-free POD than in POD. The correlation for twelve months is displayed in Table 3. Clearly, when increasing the POD mode, the correlation also increases for the same number of snapshots. We find that the results from equation-free POD are almost the same as those of POD. (Figure 3)-(Figure 5) present some comparisons made (among DNS, POD and equation-free POD ) under the different scenarios.

## 5 Convergence and Error

In the problem we investigate, we also need to pay attention to the issues of convergence and error. Before we consider the convergence, we first prove the consistency of the EF method. Let equations (2.1a)-(2.1c) be rewritten in vector form. The equation is as follows:

$$\frac{\partial \mathbf{P}}{\partial t}(t, \mathbf{x}) = \mathbf{f}(t; \mathbf{P}(t, \mathbf{x})) \quad , \quad (5.1)$$

where  $\mathbf{P}$  is a vector, which consists of  $(h, u, v)$ , and  $\mathbf{f}$  is also a vector function. Let  $\Delta T$ ,  $\Delta t$  denote the large-scale time step and short time step respectively,  $\rho$  is the maximum mesh interval, and we apply a numerical scheme that is consistent and stable. Let  $\bar{H}(t, \mathbf{x})$  be the numerical solution of  $\mathbf{p}(t, \mathbf{x})$ ,  $H(t, \mathbf{x})$  is the numerical solution of the EF method. A selected scheme is as follows,

$$\bar{H}(T^{N+1}, \mathbf{x}) = \bar{H}(T^N, \mathbf{x}) + \Delta T \psi(T^N; \Delta T, \rho, \dots). \quad (5.2)$$

Since the scheme is consistent and stable, we have following requirement at a given moment  $T = T^N$ ,

$$\lim_{\Delta T, \rho \rightarrow 0} \psi(T^N, \Delta T, \rho, \dots) = \mathbf{f}(T^N; \mathbf{P}^N), \quad (5.3)$$

$$\lim_{\Delta T, \rho \rightarrow 0} \|\bar{H}^N - \mathbf{P}\| = 0. \quad (5.4)$$

Where,  $\|\cdot\|$  is  $L^2$ -norm.

From (3.10), (3.11), (3.13) we obtained

$$\begin{aligned}
H^{N+1} &\equiv H(T^{N+1}) = \Psi \mathbf{a}^N \equiv \Psi \mathbf{a}(T^{N+1}) \\
&= \Psi \left[ \mathbf{a}^{k,N} + \Delta T_c \frac{\mathbf{a}^{k+1,N} - \mathbf{a}^{k,N}}{\Delta t} + o(\Delta T_c^2) \right] \\
&= \Psi \Upsilon \left[ H^{k,N} + \Delta T_c \frac{H^{k+1,N} - H^{k,N}}{\Delta t} + o(\Delta T_c^2) \right] \\
&= H^N + \Delta t \sum_{k=1}^{m_f} \psi((N\Delta T + k\Delta t); H) + \Delta T_c \frac{\partial H}{\partial t}(T_c^N) + o(\Delta t^2) + o(\Delta T_c^2) \\
&= H^N + \Delta T \phi(T^N; H, \dots). \tag{5.5}
\end{aligned}$$

where

$$\phi(T^N; H, \dots) = \frac{1}{\Delta T} \left[ \Delta t \sum_{k=1}^{m_f} \psi((N\Delta T + k\Delta t); H) + \Delta T_c \frac{\partial \mathbf{p}}{\partial t}(T_c^N) + o(\Delta t^2) + o(\Delta T_c^2) \right].$$

In the case of the limit of  $\Delta T \rightarrow 0, T_c^N \rightarrow T^N$ , and  $\Delta t, \Delta T_c \rightarrow 0$ . By virtue of property (5.3), we have

$$\begin{aligned}
\lim_{\Delta T, \rho \rightarrow 0} \phi(T^N; \mathbf{p}, \dots) &= \lim_{\Delta T, \rho \rightarrow 0} \frac{1}{\Delta T} \left[ \Delta t \sum_{k=1}^{m_f} \psi((N\Delta T + k\Delta t); \mathbf{p}) \right. \\
&\quad \left. + \Delta T_c \frac{\partial \mathbf{p}}{\partial t}(T_c^N) + o(\Delta t^2) + o(\Delta T_c^2) \right] \\
&= \mathbf{f}(T^N; \mathbf{P}^N).
\end{aligned}$$

Having finished illustrating the consistency from a theoretic angle, the convergence of the EF method may be obtained at a given moment  $T = T^N$ , through (5.3),(5.5)

$$\begin{aligned}
H^{N+1} &= H^N + \Delta T \phi(T^N; H, \dots), \\
H^N &= H^{N-1} + \Delta T \phi(T^{N-1}; H, \dots), \\
&\dots \\
H^1 &= H^0 + \Delta T \phi(T^0; H, \dots),
\end{aligned}$$

So, we obtain

$$H^{N+1} = \Delta T \sum_{m=0}^N \phi(T^m; H, \dots).$$

Similarly,  $\bar{H}^{N+1} = \Delta T \sum_{m=0}^N \psi(T^m; \bar{H}, \dots)$ , using (5.4),

$$\begin{aligned}
\lim_{\Delta T, \rho \rightarrow 0} \|H^N(\Delta T, \rho) - \mathbf{p}(t, \mathbf{x})\| &= \lim_{\Delta T, \rho \rightarrow 0} \|H^N - \bar{H}^N + \bar{H}^N - \mathbf{P}\| \\
&\leq \lim_{\Delta T, \rho \rightarrow 0} (\|H^N - \bar{H}^N\| + \|\bar{H}^N - \mathbf{P}\|) \\
&= \lim_{\Delta T, \rho \rightarrow 0} \|H^N - \bar{H}^N\| + 0 \\
&= \lim_{\Delta T, \rho \rightarrow 0} \|\Delta T \sum_{m=0}^N (\phi(T^m; \bar{H}, \dots) - \psi(T^m; H, \dots))\| \\
&= 0.
\end{aligned}$$

Where  $\|\cdot\|$  is  $L^2$ -norm,  $H(t, \mathbf{x})$  is the EF method solution,  $\bar{H}(t, \mathbf{x})$  is full DNS solution, and  $\mathbf{p}(t, \mathbf{x})$  is the true solution of original problem. The convergence of the method is thus proved.

The relative error is used to highlight the convergence and error in numerical experiments.

The relative error can be expressed by  $\frac{\|H - \bar{H}\|}{\|\bar{H}\|}$ , where  $H, \bar{H}$  are the EF method solution and

full DNS solution, respectively, and  $\|\cdot\|$  denotes the Euclidian-norm. The results of analysis of the numerical experiments are following.

We discuss the factors that affect the results of convergence and error, including time step and the number of iterations. In our case, due to the different magnitude, we choose the relative error, and we think that it can reflect the convergence truly. We select the case of  $n=36$ , and we first discuss the difference between large-scale time step and short time step results. One factor is number of iterations during the course of the DNS running in a large-scale time step, and we also select several sets of experiments. We give the error fold line figure when we select the number of iterations to be 2, 10, 20, 40, 100, respectively, see figure 6. We find that the results are much improved when we increase the number of iterations, and the error is much smaller. Similarly, we decided to choose the number of iterations 20 for the sake of saving computational time. The other factor is the time step, which in the case of short time step is  $\Delta t=100$  seconds, We select the large-scale time step size as  $\Delta T=6$  hours, 1 day, 5 days and 10 days respectively. We can illustrate that the method is convergent by providing a numerical example, see figure 7. In addition to the RMS error, we compute the relative error between reduced solution and full solution for  $h$ , and we attained a relative error of below  $10^{-3}$  of the magnitude of the average height  $h$ , that is about 150 meters. The results indicate that the smaller the large time step is, the smaller the error for  $h, v$ , with a reverse effect for  $u$ . Generally speaking we conclude that the result is better when a large-scale time step of  $\Delta T=1$  day is used, in particular when we consider computation time savings. The relative error is presented to show the convergence of the method, see figure 8. Here, we also consider another issue related to scheme of time derivative estimation. In our case, we use the Forward Euler Projection to discretize the time derivative. The polynomial extrapolation approximation is another way to estimate the time derivative. In our case, we find that the Forward Euler Projection yields the same result as obtained with polynomial extrapolation when we use 1-th order polynomial. However, the result is unbelievably improved when we

choose the 2-th order polynomial, and this result may be related to our specific problem.

## 6 Summary

We applied the equation-free POD method to reduced modeling of a large-scale upper ocean circulation in the tropic Pacific domain. Three sets of snapshots are chosen to analyze the difference between POD and equation-free POD, and we found that some results of the equation-free POD were the same as those of POD, while other were slightly better than those of POD. By numerical experimentation, we analyzed the factors that affected the results, and some improvements were obtained in resolving the POD related problems. Generally speaking, the number of snapshots was found to have little influence on results beyond a certain threshold number. Certainly, increasing the number of snapshots is good for us. Here, we often choose the number of basis functions that capture 99% energy. We found that it was not always the case that the larger the number of basis functions we choose, the better the results obtained. For a given number of snapshots and basis functions, we also discussed the effect induced by a large-scale time step and by the number of iterations. The results indicated that the ratio between large time step and short time step should be confined to a certain range, here, the result is acceptable when the large time is equal to 1 day.

We also think it appropriate to take the number of iterations to be equal to 20 for saving computation time, though a larger number of iterations is beneficial for our problem. The convergence and error between approximate solution and full solution were also discussed, and we provided an illustration of convergence and error by calculating relative error and RMS error.

From the results, we can find that the magnitude of the error is below  $10^{-3}$ . However the most obvious advantage of using the equation-free POD is that we can solve problems for which the corresponding equations are unavailable or without closed form. Although the results of using equation-free POD are not obviously better (they are the same as POD in cases) than those of POD, we are still satisfied with them for our applying the new method to the tropic Pacific Ocean simulation due to its enormous decrease in computation work compared with DNS. For instance, the computational cost of the equation-free POD-assisted method is less than 10% that required by DNS. In our case, we think it important that the *lifting* step-given the accurate initial values of macro variables (here, low-POD coefficients) be implemented. Projective integration is one of the traditional continuum numerical procedures implemented in an equation-free frame. Certainly, we must further investigate the issues that emerge in the process of using equation-free POD method, for instance, the computational cost is not less than that of POD.

## References

- [1] G. Berkooz, P. Holmes and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Ann. Rev. Fluid Mech.*, 25:539-575, 1993.
- [2] L. Sirovich. Turbulence and the dynamics of coherent structures, Parts i, ii and iii. *Quart. Appl. Math.*, XLV:561-590, 1987.

- [3] Karhunen, K., Zur Spektraltheorie Stochastischer Prozesse, Ann. Acad. Sci. Fennicae, 37 (1946).
- [4] M. Loeve, Fonctions aleatoires de second ordre, Comptes Rend. Acad. Sci., Paris, 220(1945).
- [5] R.E. Arndt, D.F. Long and M. N. Glauser. The proper orthogonal decomposition of pressure fluctuations surrounding a turbulent jet. J. Fluid Mech., 340:1-33, 1997.
- [6] J.H. Citriniti and W.K. George. Reconstruction of the global velocity field in the axisymmetric mixing layer utilizing the proper orthogonal decomposition. J. Fluid Mech., 418:137-166, 2000.
- [7] J. Delville, L. Ukeiley, L. Cordier, J.P. Bonnet, and M. Glauser. Examination of large-scale structures in a turbulent plane mixing layer. Part 1. Proper orthogonal decomposition. J. Fluid Mech., 391:91-122, 1999.
- [8] A. Glezer, Z. Kadioglu, and A.J. Pearlstein. Development of an extended proper orthogonal decomposition and its application to a time periodically forced plane mixing layer. Phys. Fluids, 1(8):1363, 1989.
- [9] N. Aubry, P. Holmes, J. L. Stone and J.L. Lumley. The dynamics of coherent structures in the wall region of a turbulent boundary layer. J. Fluid. Mech., 340:1-33, 1997.
- [10] W. Cazemier, R. W. Verstappen and A.E. Veldman. Proper orthogonal decomposition and low-dimensional models for driven cavity flows. Phys. Fluids, 10(7):1685-1699, 1998.
- [11] A.E. Deane, I.G. Kevrekidis, G. E. Karniadakis, and S.A. Orszag. Low dimensional models for complex geometry flows: Application to grooved channels and circular cylinders. Phys. Fluids A, 3(10):2337-2354, 1991.
- [12] A. Liakopoulos, P.A. Blythe and H. Gunes. A reduced dynamical model of convective flows in tall laterally heated cavities. Proc. R. Soc. Lond. A, 453:663, 1997.
- [13] Courtier, P., Thépaut, 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).
- [14] Griewank, A., Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation Frontiers in Applied Mathematics, SIAM. (2000).
- [15] 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).
- [16] 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).
- [17] Restrepo, J., Leaf, G., Griewank, A., Circumventing storage limitations in variational data assimilation studies. SIAM J. Sci. Comput. 19, 1586-1605, (1998).
- [18] 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 submitted to Int. J. Num. Methods in Fluids.
- [19] Yanhua Cao, Jiang Zhu, Zhendong Luo, Navon, I. M. Reduced order modeling of the upper

tropical Pacific Ocean model using proper orthogonal decomposition, (accepted). *Computers and Mathematics with Applications*.

[20] C.Foias, M.S.Jolly, I.G. Kevrekidis and E.S. Titi. Dissipativity of the numerical schemes. *Nonlinearity*, 4:591-613,1991.

[21] S. Sirisup, D. Xiu , GE Karniadakis and IG Kevrekidis, Equation/Galerkin-free POD-assisted Computation of Incompressible Flows, *J. Comput. Phys*, 207, 568-587 (2005).

[22] C. Theodoropoulos, Y. Qian and I.G. Kevrekidis. Coarse stability and bifurcation analysis using time-steppers: a reaction-diffusion example. *Proc. Natl. Acad. Sci.*,97,2000.

[23] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, and C. Theodoropoulos. Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Comm. Math. Sci.*,1(4):715-762,2003.

[24] I.G. Kevrekidis, C.W. Gear and G. Hummer. Equation-free: the computer-assisted analysis of complex, multiscale systems. *A. I. Ch. E. Journal*, 50(7):1346-354,2004.

[25] C.W. Gear and I.G. Kevrekidis. Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum. *SIAM J. Sci. Comput.*, 24:1091-1106, 2003.

[26] C.W. Gear and I. G. Kevrekidis. Telescopic projective methods for parabolic differential equations. *J. Comput. Phys.*, 187(1):95-109, 2003.

[27] R. Rico-Martinez, C.W. Gear, and I.G. Kevrekidis. Coarse projective kMC integration: forward/reverse initial and boundary value problems. *J. Comput. Phys.*, 196:474-489, 2004.

[28] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, and C. Theodoropoulos. Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Comm. Math. Sci.*, 1(4):715-762, 2003.

[29] I.G. Kevrekidis, C.W. Gear, and G. Hummer. Equation-free: the computer-assisted analysis of complex, multiscale systems. *A. I. Ch. E Journal*, 50(7):1346-1354, 2004.

[30] A. Makeev, D. Maroudas, and I.G. Kevrekidis. Coarse stability and bifurcation analysis using stochastic simulators: kinetic Monte Carlo examples. *J. Chem. Phys. Chem. Phys.*, 116:10083-10091,2002.

[31] A. Makeev, D. Maroudas, A. Panagiotopoulos, and I.G. Kevrekidis. Coarse bifurcation analysis of kinetic Monte Carlo simulations: a lattice-gas model with lateral interactions. *J. Chem. Phys.*, 117:8229{8240, 2002.

[32] S. Sirisup and G. E. Karniadakis. A spectral viscosity method for correcting the long-term behavior of POD models. *J. Comput. Phys.*, 194(1):92{116, 2004.

[33] C. W. Gear, IGK and C. Theodoropoulos, 'Coarse' integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods *Comp. Chem. Engng.* 26 pp.941-963 (2002). 3.

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

[35] Fukunaga K. *Introduction to Statistical Recognition*, Academic Press, 1990.

[36] Ly H V and Tran H T. Proper orthogonal decomposition for flow calculations and optimal

control in a horizontal CVD reactor, Center for Research in Scientific Computation, North Carolina State University, technical report CRSC-TR98-12 (1998).

[37] Sirovich L. Turbulence and the dynamics of coherent structures, parts I-III. *Q. Appl. Math.* XLV (3), 1987: pp.561-590.

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

Table 1 The values of the model parameters used in the model

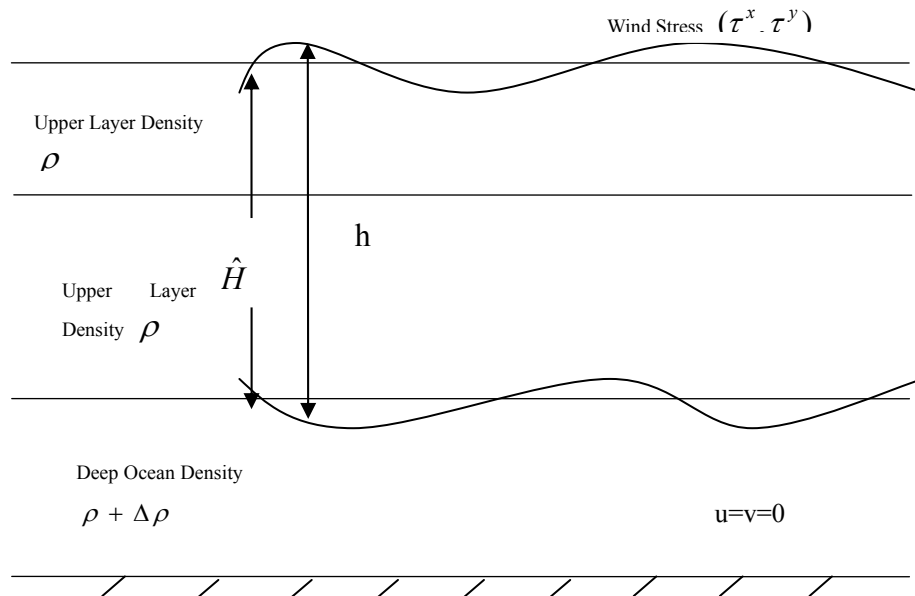


Figure 1 The vertical structure of the reduced-gravity model

POD

RMSE of h	60%	95%	99%
20 snapshots	2.76744771	1.25615931	0.60962808
36 snapshots	2.76414776	1.24391854	0.59819734
60 snapshots	2.68031812	1.24963772	0.62677878

Equation-free POD

RMSE of h	60%	95%	99%
20snapshots	2.41852546	1.12094295	0.51566243
36snapshots	2.41472030	1.11343133	0.48440239
60snapshots	2.41544986	1.11462390	0.48794851

POD

RMSE of u	60%	95%	99%
20snapshots	0.01915840	0.00635201	0.00546203
36snapshots	0.01916046	0.00411278	0.00509165
60snapshots	0.01926533	0.00524721	0.00656431

Equation-free POD

RMSE of u	60%	95%	99%
20snapshots	0.01127066	0.00545016	0.00348107
36snapshots	0.01115794	0.00487991	0.00250289
60snapshots	0.01116655	0.00493398	0.00261936

POD

RMSE of v	60%	95%	99%
20snapshots	0.01203851	0.00397984	0.00547067
36snapshots	0.01194210	0.00426739	0.00512042
60snapshots	0.01091984	0.00378621	0.00410236

Equation-free POD

RMSE of v	60%	95%	99%
20snapshots	0.00531464	0.00207548	0.00154492

36snapshots	0.00529146	0.00176191	0.00101180
60snapshots	0.00530750	0.00181613	0.00111099

Table 2 RMSE of  $h, u, v$ , and a comparison between POD and equation-free POD for different percentages of captured energy, here, snapshots is  $n=20, 36, 60$

POD

h (99%)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
20 snapshots	0.992	0.991	0.995	0.992	0.992	0.983	0.985	0.991	0.995	0.994	0.992	0.991
36 snapshots	0.992	0.993	0.995	0.993	0.992	0.984	0.985	0.990	0.994	0.994	0.993	0.990
60 snapshots	0.992	0.996	0.994	0.992	0.993	0.983	0.986	0.991	0.992	0.994	0.993	0.990

Equation-free POD

h (99%)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
20 snapshots	0.992	0.996	0.996	0.993	0.993	0.986	0.986	0.991	0.995	0.994	0.993	0.991
36 snapshots	0.994	0.997	0.996	0.994	0.994	0.985	0.987	0.992	0.995	0.995	0.994	0.991
60 snapshots	0.994	0.996	0.996	0.994	0.994	0.985	0.987	0.992	0.995	0.995	0.994	0.990

POD

u (99%)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
20 snapshots	0.776	0.883	0.950	0.890	0.872	0.897	0.910	0.945	0.940	0.920	0.920	0.861
36 snapshots	0.791	0.889	0.949	0.900	0.885	0.885	0.916	0.949	0.934	0.922	0.925	0.858
60 snapshots	0.794	0.898	0.950	0.898	0.878	0.884	0.919	0.951	0.937	0.928	0.925	0.852

Equation-free POD

u (99%)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
20 snapshots	0.974	0.989	0.993	0.952	0.975	0.989	0.990	0.992	0.996	0.951	0.956	0.992
36 snapshots	0.990	0.995	0.992	0.990	0.990	0.989	0.996	0.995	0.993	0.987	0.973	0.990
60 snapshots	0.989	0.994	0.992	0.988	0.989	0.989	0.995	0.995	0.993	0.986	0.970	0.989

Table 3 Correlation as to 20 snapshots, 36 snapshots, and 60 snapshots for upper layer thickness  $h$  (unit: m), zonal current velocity  $v$  (unit: m/s) a comparison between POD and equation-free POD in energy percentage of 99%

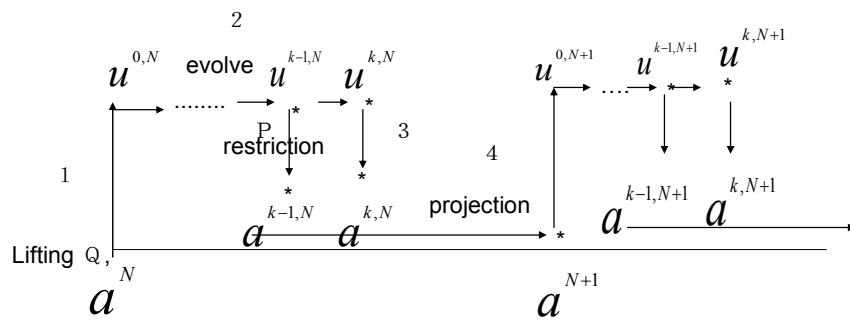
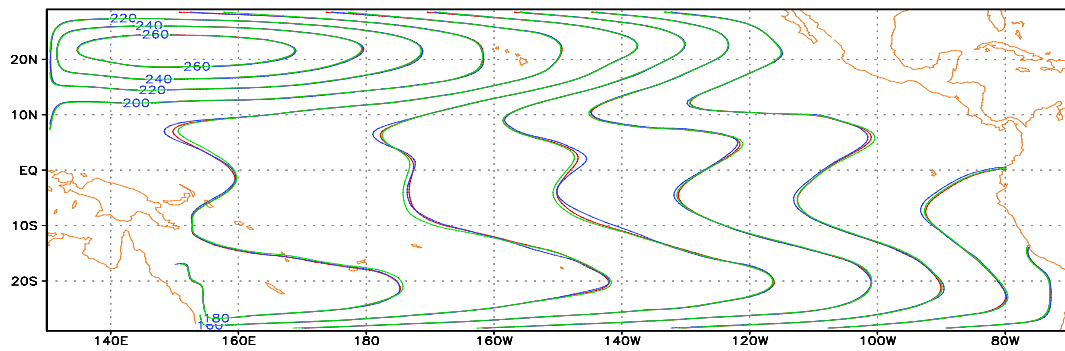
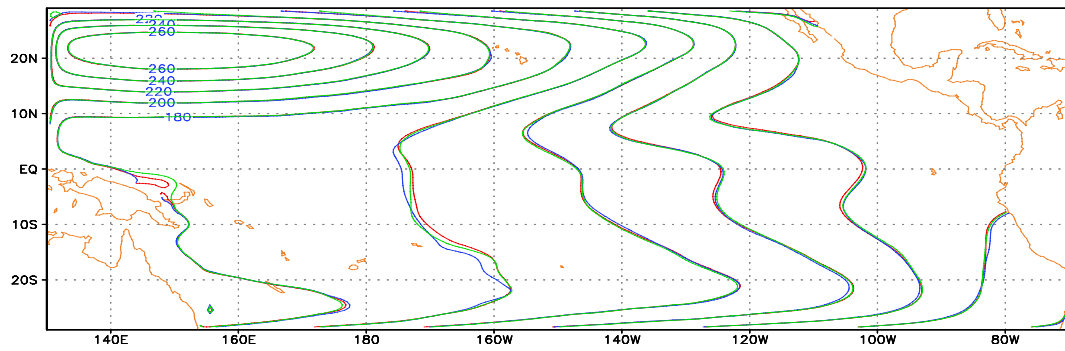
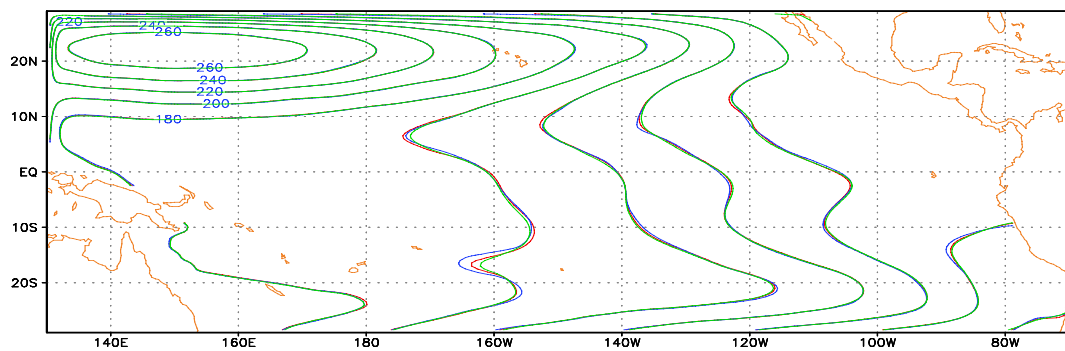


Figure 2 Sketch of POD-assisted projective integration



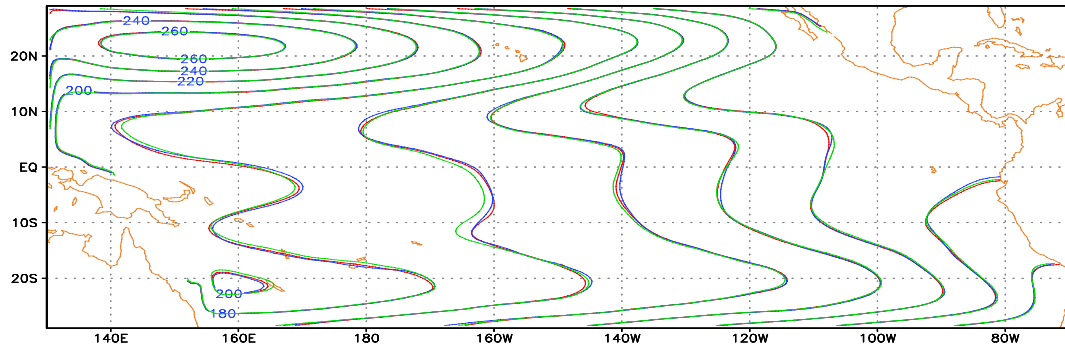
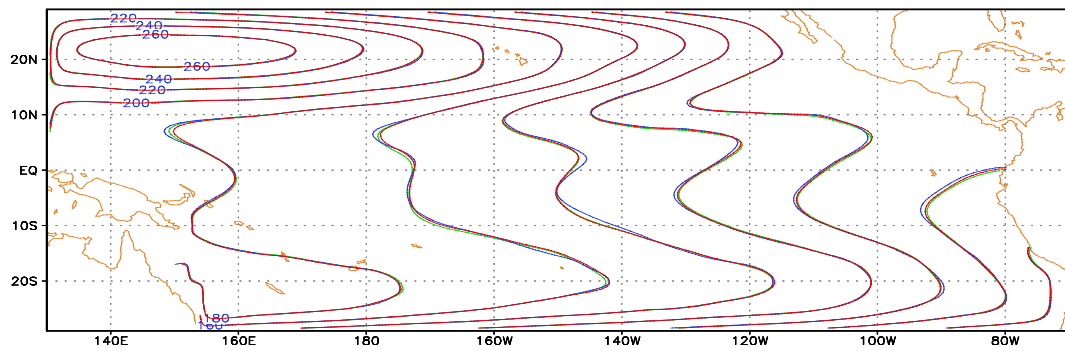
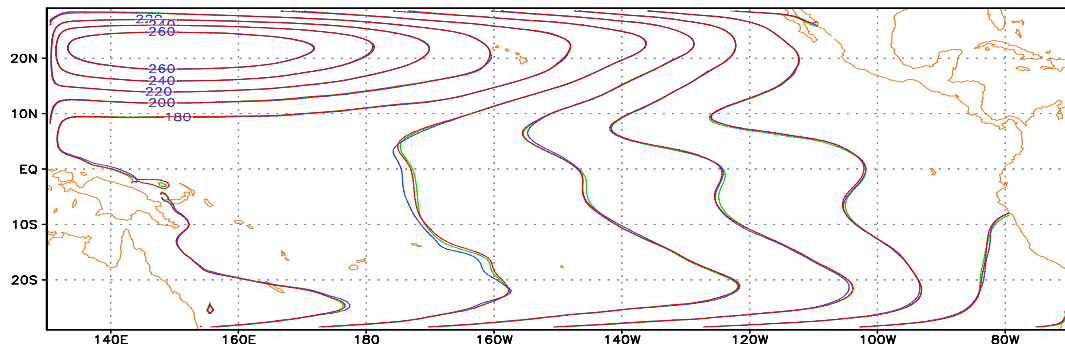
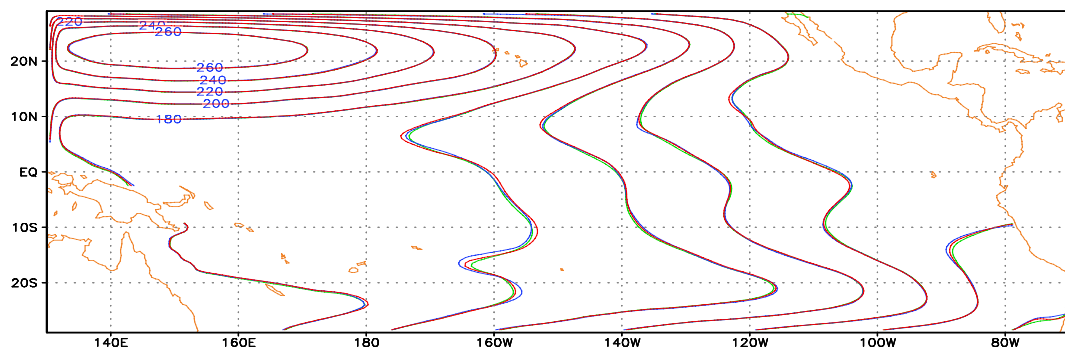


Figure 3 Upper layer thickness in February, May, August and November in the case of 20 snapshots. The number of POD basis is determined by capturing 99% of energy produced by the full model approximation. Blue: full order approximation; red: equation-free POD; green : POD



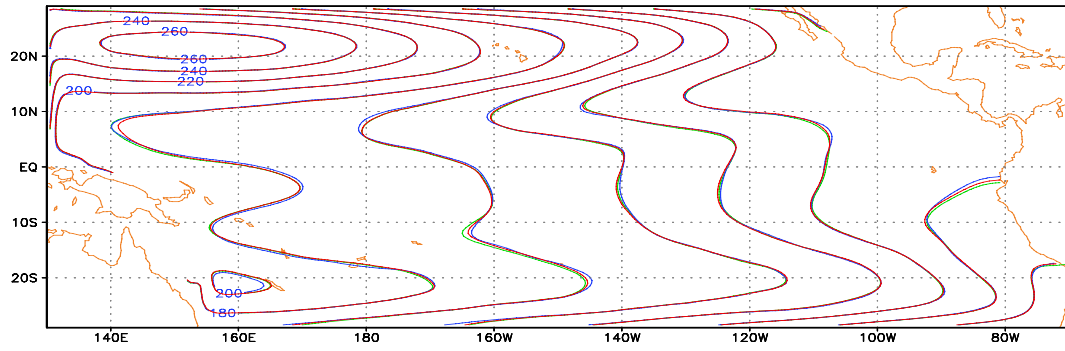
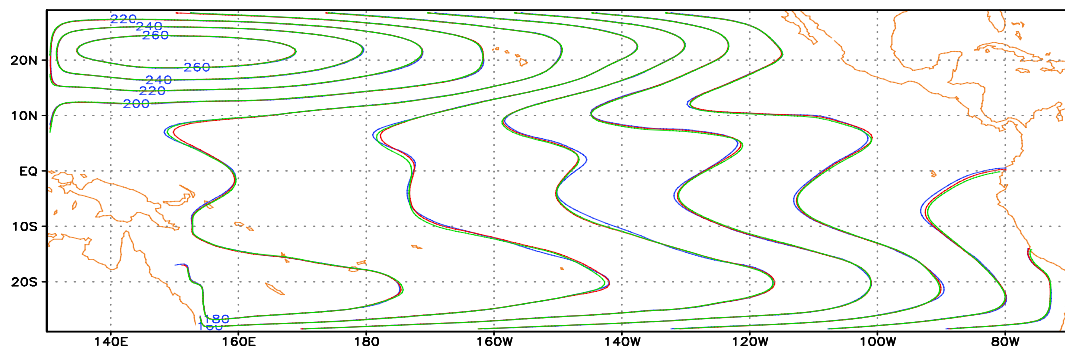
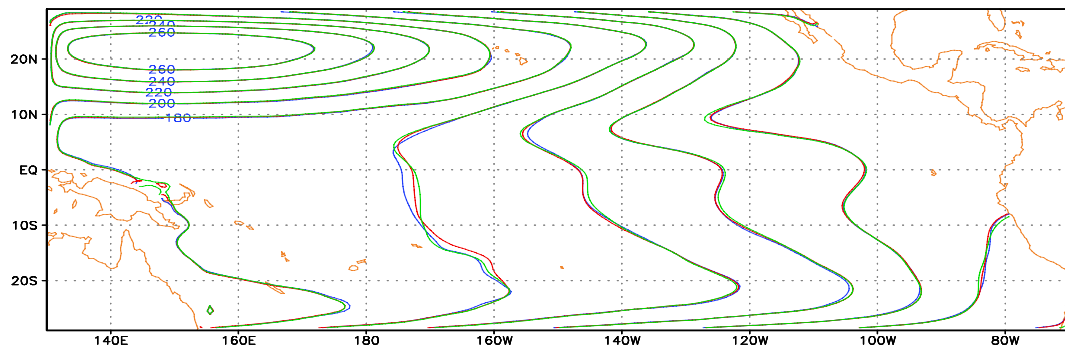
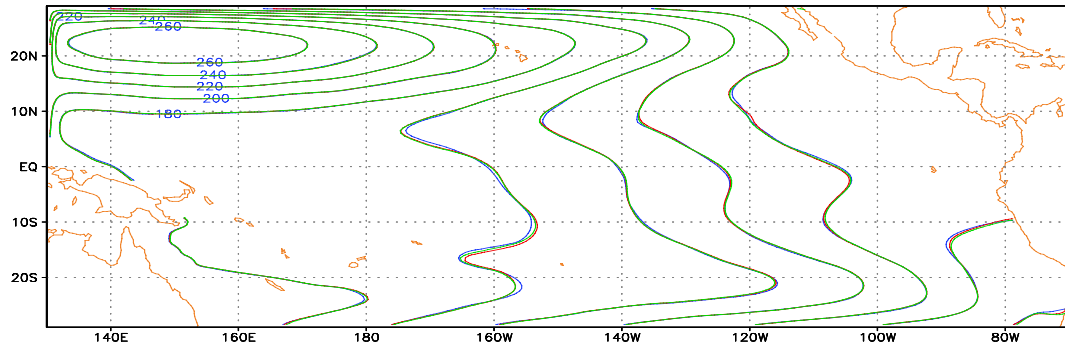


Figure 4 The same as figure 3, but for the case of 36 snapshots.



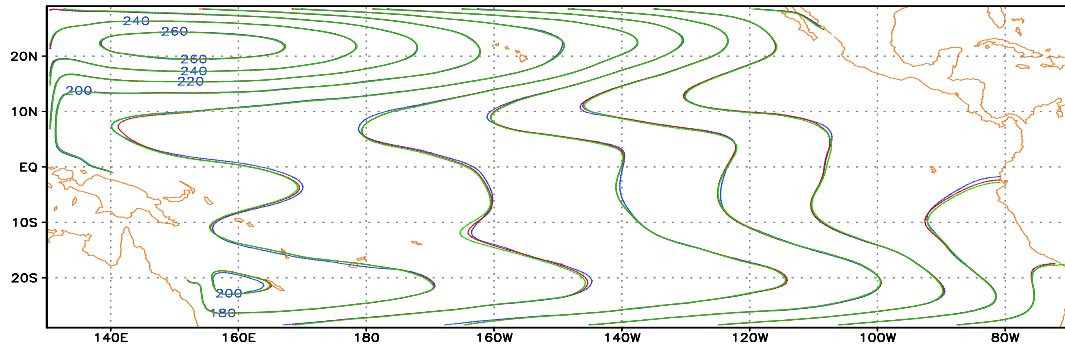
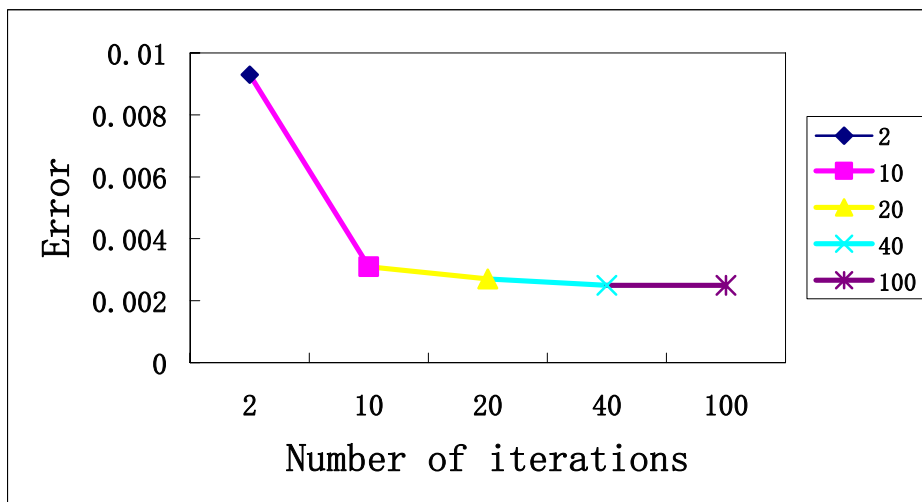
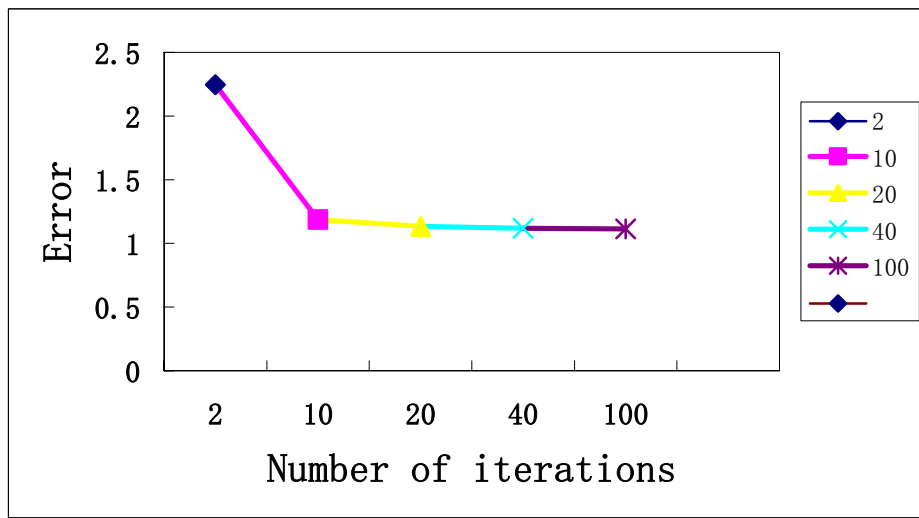


Figure 5 The same as figure 3, but for the case of 60snapshots.



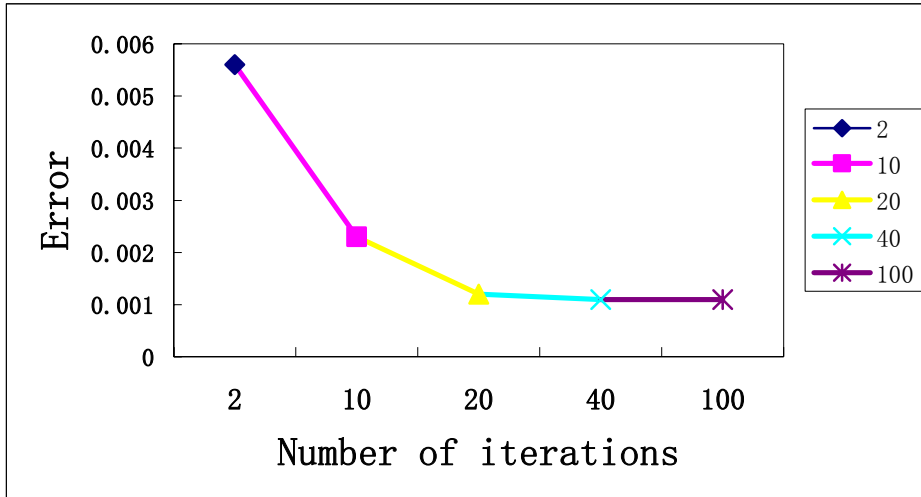
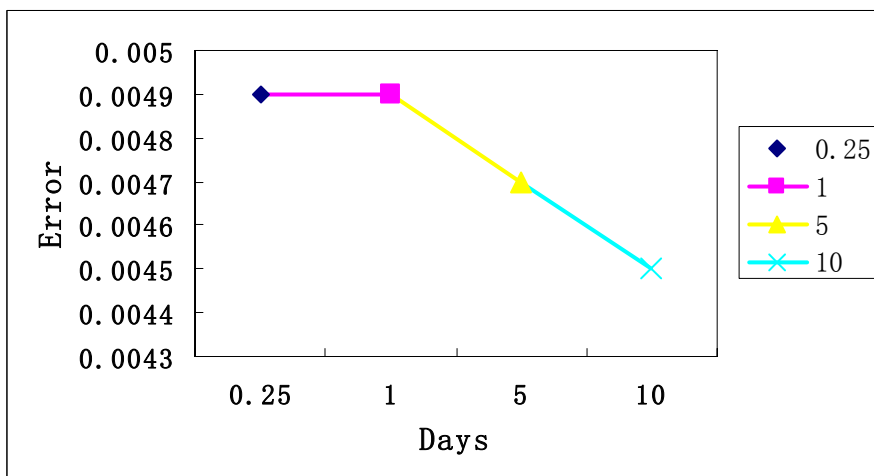
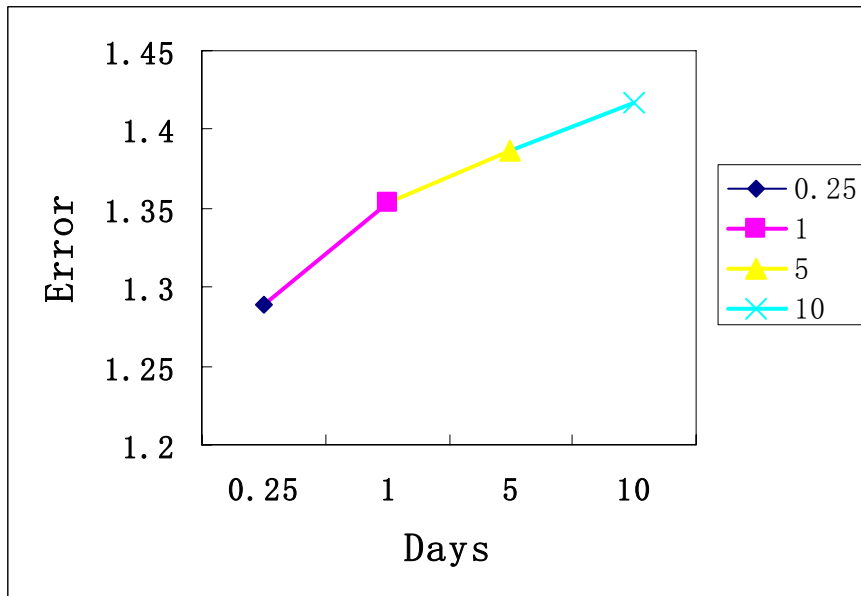


Figure 6 Root mean square error of the equation-free POD as a function of the number of iterations.



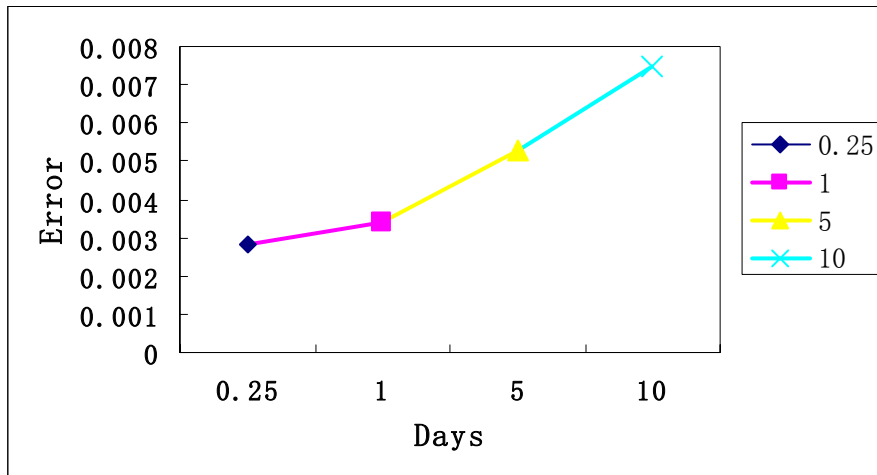


Figure 7 Root mean square error of the equation-free POD as a function of the large-scale time step.

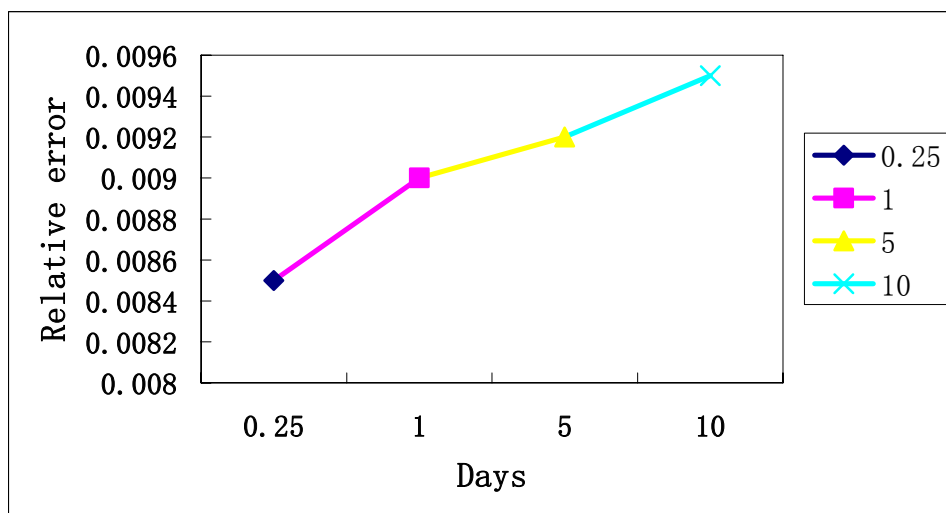
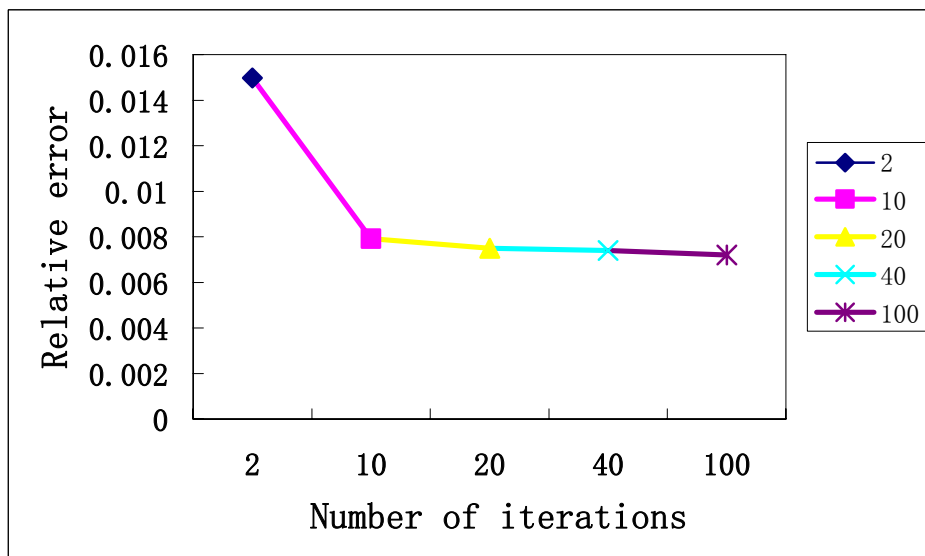


Figure 8 Relative error of the equation-free POD as a function of the number of iterations and large-scale time step, respectively.