AN EQUATION- FREE, REDUCED-ORDER MODELING APPROACH TO TROPICAL PACIFIC SIMULATION

RUIWEN WANG[†]

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

School of Science, Beijing Jiaotong University, Beijing, 100044, China

I.M. NAVON

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

The "equation-free" (EF) method is often used in complex, multi-scale problems. In such cases it is necessary to know the closed form of the required evolution equations about macroscopic variables within some applied fields. Conceptually such equations exist, however, they are not available in closed form. The EF method can bypass this difficulty. This method can obtain macroscopic information by implementing models at a microscopic level. Given an initial macroscopic variable, through lifting we can obtain the associated microscopic variable, which may be evolved using Direct Numerical Simulations (DNS) and by restriction, we can obtain the necessary macroscopic information and the projective integration to obtain the desired quantities. In this paper we apply the EF POD-assisted method to the reduced modeling of a large-scale upper ocean circulation in the tropical Pacific domain. The computation cost is reduced dramatically. Compared with the POD method, the method provided more accurate results and it did not require the availability of any explicit equations or the right-hand-side (RHS) of the evolution equation.

[†] E-mail: wangrw@mail.iap.ac.cn

Contract/grant sponsor: Natural Science Foundation of China; contract/grant numbers: 40437017, 40225015.

Contract/grant sponsor: NSF; contract/grant number: ATM-9731472

1. Introduction

Proper Orthogonal Decomposition (POD) methodology is one that provides a very efficient way to perform reduced modeling via the identification of the most energetic modes in a sequence of snapshots from a time-dependent system and subsequently providing a means for obtaining a low-dimensional description of that system's dynamics (see [1]). The method of snapshots that was first proposed in Ref. [2] for a flow system was a very effective and easy to carry out approach for obtaining POD basis sets. The POD method was originally introduced by Karhunen in 1946 (see [3]) and Loeve in 1945 (see [4]), and the method has been extensively used in research within recent years and has been successfully applied to a variety of fields, both in conjunction with experimental efforts (see [5, 6, 7, 8]) and other numerical studies (see [9, 10, 11, 12]), including thermal convection, shear layers, cavity flows and external flows.

In recent years, our understanding of the tropical ocean has increased dramatically. There is a vast and growing amount of literature based on the design of ocean models focused on the discretization of partial differential equations for physical systems. Such models are often hard to solve due to the high order systems that are needed to describe the state. So, the reduced-order modeling is derived by using the POD method to approximate the full modeling (see [9]). In order to improve the prediction of ocean and atmosphere, the data assimilation is a kind of effective method, 4D-VAR is a kind of data assimilation methods. However, a basal difficulty in the use of 4DVAR for realistic general circulation models is that we need to run the full model repeatedly, and the computation cost is very expensive. How to reduce computation cost and computation time is a very valuable work in solving the actual problem. EF method can be employed to solve this difficulty effectively. Another major difficulty in the 4D-VAR is the dimension of the control space, which is generally equal to the size of the model-state variable and is typically on the order of 10^7-10^8. Current ways to obtain feasible implementations of 4DVAR consist mainly of the incremental method (see [14]), check-pointing (see [15]) and parallelization. However, each of these three methods has their obvious shortcomings. POD provides a potential technique that can dramatically reduce the computation and memory burdens associated with 4DVAR (see [19]). In the aforementioned works, we can obtain a low-dimensional dynamics system directly from the Galerkin projection of the governing equations on the POD modes. However, it is well known that the reduced systems resulting from the truncated Galerkin projection may result in spurious asymptotic states (see

[20]) and may be difficult to obtain the explicit form of the right-hand-side (RHS) of the equation, which consists of POD coefficients of the evolution equations.

The equation free (EF) POD-assisted method was first used to resolve the above questions within incompressible flows (see [21]). The basic process is as follows (see Fig. 1): First one must devise and implement short-time numerical experiments with "the best available microscopic model, and the subsequently estimate quantities (derivative) that are required for numerical computation of the unavailable macroscopic equations for the coarse-grained system behavior (see[28,29]) by using the numerical results of aforementioned microscopic computations. Since ocean forecasting is very important for human activity, the topic we are investigating is more complicated and significant than a two-dimensional flow past a circular cylinder. In this paper, we apply the EF POD-assisted method to the simulation of the upper tropical Pacific Ocean model based on the POD model. Thus, we can estimate the closures that are required to obtain explicit macroscopic equations on demand and perform numerical analysis by running the microscopic simulations directly. This framework has been applied to solve many types of problems, such as bifurcation analysis of complex systems and homogenization of random media (see [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, its mathematical properties and EF POD model are presented in section 3. In section 4 we analyse the convergence and accuracy of the method. In section 5, the numerical calculations using EF POD method in the context of simulating the upper layer thickness and currents in this ocean model, as well as comparisons with the full model and POD model are discussed. We summarize the results we obtained 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 was used in this paper, which was used in the study of ocean dynamics in tropical regions (Cane 1979; Seager et al..1988). The model is a reduced-gravity, linear transport model, consisting of two layers above the thermocline with the same constant density (Fig. 1).

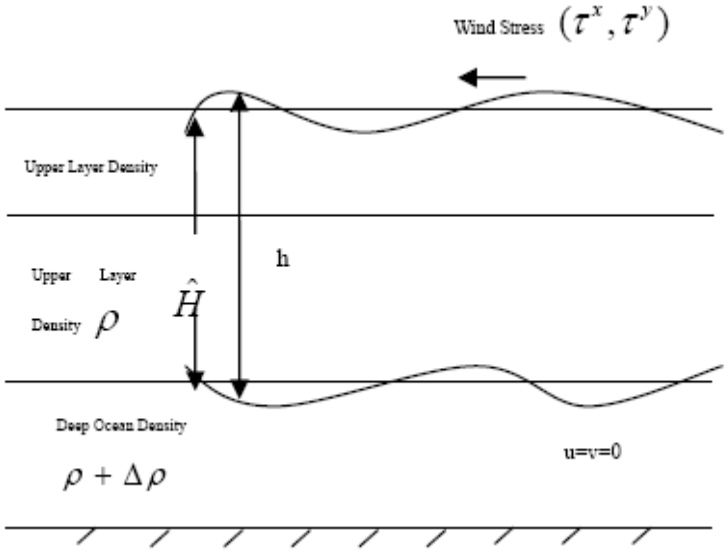


Figure 1. The vertical structure of the reduced-gravity model.

It is assumed that below the thermocline, the ocean is higher in 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:

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

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

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

where (u, v) are the horizontal velocity components of the depth-averaged currents, h is the total layer thickness, f the Coriolis force, $\hat{H}$ the mean depth of the layer, ρ_0 the density of water, A the horizontal eddy viscosity

coefficient and α is the coefficient of friction. The wind stress was calculated via the aerodynamic bulk formula,

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

Where ρ_a is the density of air, c_D the wind-stress drag coefficient and (U_{wind}, V_{wind}) are the components of the wind velocity.

2.2. *Numerical Scheme*

The dynamical model equations [eq. (2.1a)-(2.1c)] are governed by wave dynamics. In addition, the model domain ranges were chosen to cover the region from 29°S ~ 29°N, 120°E ~ 70°W. This chosen model domain allowed all possible equatorially trapped waves to be excited by the applied wind forcing (Moore and Philander 1978). We chose 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 allowed all possible waves to be resolved and rendered the model integration numerically stable. The model was driven by the FSU (Florida State University) climatological monthly-mean winds (Stricherz *et al.*.1992). Through a linear interpolation, the data were projected onto each time step and then into each grid point. In Table 1, the values of the numerical parameters that were used in the model integration are listed.

Table 1. Values of the model parameters used for the full-model calculations.

Parameter	Value	Remarks
g'	3.7×10^{-2}	Reduced gravity
C_D	1.5×10^{-3}	Wind stress drag coefficient
$\hat{H}$	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

It takes about 20 computational years for the model to reach a periodic constant seasonal cycle. At that time, the computation has successfully captured the main seasonal variability of the dynamical fields. The currents and the upper layer thickness of the 21st year were saved for the process.

The model was discretized on the Arakawa C-grid, and all model boundaries were closed. At these solid boundaries, we applied the no-normal flow and no-slip conditions. The time integration used a leap-frog scheme, with a forward scheme every 10th time step used to eliminate the computational mode. On every integration day, a mass-compensation was carried out.

3. Computational Formulation of EF Method

A Simple Introduction to POD

In order to express the premise of the POD method, we will introduce the POD method in a continuous case. Since the basis for both the continuous and discrete cases are the same, we will describe the continuous case, despite having carried out the numerical experiments using the discrete case.

Let $U_i(\vec{x})$ ($i = 1,2,\ldots,N$) denote a 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,

$$\overline{U} = < U > = \frac{1}{N}\sum_{i=1}^{N} U_i(\vec{x}). \qquad (3.1)$$

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

$$V_i = U_i - \overline{U}\ . \qquad (3.2)$$

We wish to find an optimal compressed description of the sequence of data using eq. (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 optimal properties, is provided by the Karhunen-Loève

expansion (see [35]), where the basis functions Φ were in fact admixtures of the snapshots and are given by

$$\Phi = \sum_{i=1}^{N} \beta_{ik} V_i(\vec{x}), \qquad (3.3)$$

where the coefficients β_{ik} are to be determined such that the value of Φ given by eq. (3.3) will most closely resemble the ensemble $\{V_i(\vec{x})\}_{i=1}^{N}$. More specifically, the POD seeks a function Φ such that

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

that is subject to

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

which is maximized, 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,

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

where the kernel is given by

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

Thus, we see that our problem from eq. (3.4) amounts to solving for the eigenvectors of a $N \times N$ matrix (see also [36]), where N is the size of the ensemble of the snapshots as follows:

$$\frac{1}{N}\sum_{i=1}^{N} |(V_i, \Phi)|^2 = (\lambda\Phi, \Phi) = \lambda, \qquad (3.8)$$

substituting eq. (3.3) into eq. (3.6), yields the following eigenvalue problem:

$$\sum_{i=1}^{N} \left(\frac{1}{N}\int_{\Omega} V_i(\vec{x}')V_j(\vec{x}')d\vec{x}'\right)\eta_j = \lambda\eta_i, i = 1, 2, \cdots, N. \qquad (3.9)$$

Namely,

$$\sum_{i=1}^{N} L_{i,j}\eta_j = \lambda\eta_i, \qquad (3.10)$$

where, $L_{ij} = \frac{1}{N}(V_i, V_j)$, $L = (L_{ij})_{N\times N}$ .We then may rewrite Eq. (3.9) in matrix form,

$$\mathbf{C}\mathbf{w} = \lambda\mathbf{w}, \qquad (3.11)$$

where $\mathbf{C}$ is a symmetric and non-negative matrix,

$\mathbf{C}_{ik} = \frac{1}{N}\int_{\Omega} V_i(\vec{x}')V_k(\vec{x}')d\vec{x}', \mathbf{w} = [\eta_1, \eta_2, \cdots, \eta_N]', \eta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{iN}](i = 1, 2, \cdots, N)$

, are the elements of the eigenvector corresponding to the eigenvalue of λ_i .

By using eq. (3.3), we can obtain the POD basis, $\Phi_i, i = 1, 2, \cdots N$. To achieve model reduction, one can define a relative information content to choose a low-dimensional basis of size $M(<< N)$, by neglecting modes corresponding to small eigenvalues. We therefore define

$$I(k) = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{n}\lambda_i}, \qquad (3.12)$$

and choose M such that

$$M = \arg\min\{I(k) : I(k) \geq \sigma\}, \qquad (3.13)$$

where $0 \leq \sigma \leq 1$ is the percentage of total information captured by the reduced space $B^M = \text{span}\{\Phi_1, \Phi_2, \cdots, \Phi_M\}$. The tolerance σ must be chosen to be in the vicinity of unity, in order to capture most of the energy of the snapshot basis. The reduced model can then be obtained by the following expansion:

$$U^{POD}(t, \vec{x}) = \overline{U}(\vec{x}) + \sum_{i=1}^{M}\alpha_i(t)\Phi_i(\vec{x}). \qquad (3.14)$$

Assuming that the initial condition is unknown, for t =0,

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

Therefore, we obtained reduced control variables $\alpha_i(0)(i = 1,2,\cdots,M)$. We can reconstruct the reduced order spaces and find the solution in such spaces. We can then use the Galerkin approximation to project the full model to the reduced order spaces spanned by POD basis functions, and then obtain the system of ODEs about coefficients.

Supposing that the equations of the full order model have the following vector form:

$$\frac{\partial U}{\partial t} - K(U) = 0. \qquad (3.15)$$

the POD-reduced model equation can be denoted as,

$$\frac{\partial U^{POD}}{\partial t} - K(U^{POD}) = 0 \ . \qquad (3.16)$$

Substituting eq. (3.14) into eq. (3.16) yields the following ODEs about coefficients:

$$\frac{d\alpha_i}{dt} - (K(\overline{U} + \sum_{j=1}^{M} \alpha_j(t)\Phi_j), \Phi_i)_\Omega = 0, i = 1,2,\cdots,M \ . \qquad (3.17)$$

where $\alpha_i(0)(i = 1,2,\cdots,M)$ are given.

So we can obtain the POD approximation of the full model by eq. (3.14) and eq. (3.17)

The vector form of eq. (3.17) is,

$$\frac{d\mathbf{a}}{dt} = \mathbf{Y}(t;\mathbf{a}(t)). \qquad (3.18)$$

It is very difficult or unavailable to know $\mathbf{Y}(t;\mathbf{a}(t))$ in closed form, and to do so entails a large difficulty in the application of the POD method. So, the EF POD-assisted method combines the EF method with the POD method to overcome the difficulty above.

3.1. *The EF POD Model*

The POD procedure extracts an empirical orthogonal basis using a modal decomposition from an ensemble of signals. So one can represent $V(t,\vec{x})$ in the form,

$$V(t,\vec{x}) = \sum_i \alpha_i(t)\Phi_i \qquad (3.19)$$

where $\{\Phi_i(\vec{x})\}_{i=0}^{\infty}$ is the basis extracted from the above eigenvalue problem (see eq.(3.3)-(3.11)) and $\{\alpha_i(t)\}_{i=0}^{\infty}$ is the eigenvector corresponding to the eigenvalue. Since the basis functions are orthogonal, we can represent $\{\alpha_i(t)\}_{i=0}^{\infty}$ in the form,

$$\alpha_i(t) = (V(t,\vec{x}),\Phi_i(\vec{x})) \qquad (3.20)$$

Here, we let $\mathbf{a}(t) = \{\alpha_i(t)\}$ ([21]) and we define a restriction operator Υ, subject to,

$$\mathbf{a}(t) = \Upsilon V(t,\vec{x}) \equiv \{(V(t,\vec{x}),\Phi_i(\vec{x})\}, t \in W, \forall i\}, \qquad (3.21)$$

and a lifting operator Ψ, subject to,

$$V(t,\vec{x}) = \Psi\mathbf{a}(t) \equiv \sum_i \alpha_i(t)\Phi_i(\vec{x}), t \in R^+, \qquad (3.22)$$

where W is a time interval, Υ and Ψ are linear operators, and $\Upsilon\ \Psi\ \mathbf{a} = I\ \mathbf{a} \equiv \mathbf{a}$, I is the unit operator.

The whole method is often made of two levels, that is, an "inner" and "outer" simulator. The "inner" simulator is the microscopic, direct numerical simulation and the "outer" consists of many types of continuous mathematical methods at the macro scale level, such as finite difference, finite element and finite volume element optimization. The manner in which to link between macro and micro scales is the key problem.

Here, the "inner" simulator is the fully-resolved, full-model numerical simulation for the tropical Pacific model. The "outer" (coarse) model is the unavailable, in closed form Galerkin sets of ODEs, which consist of coefficients of the evolution equations based on the first few low-POD modes. $V(t,\vec{x})$ is regarded as the microscopic variable and the coefficient $\mathbf{a}(t)$ is regarded as macroscopic variable, based on the POD reduced spaces. The process of the EF POD-assisted method was made up of the following steps (see Fig. 2): Given $\mathbf{a}^n \equiv \mathbf{a}(T^n)$,

1. Lifting: at $T = T^n$, get $V(T^n, \vec{x}) = \Psi \mathbf{a}(T^n)$
2. Microscopic variable computation: resolve the full model using eq. (2.1a)-(2.1c) and obtain $V(T, \vec{x})$ for a time interval $T^n \le T \le T_*^n = T^n + k\Delta t$, where, Δt is the time step of the DNS of the full model, with k being an integer.
3. Restriction: compute the coefficients $\mathbf{a}(T) = \Upsilon V(T, \vec{x})$ for $T^n \le T \le T_*^n = T^n + k\Delta t$, and estimate the time derivative at $T = T_*^n$.
4. Projective integration: integrate eq. (3.18) to T^{n+1} by the standard ODE techniques to obtain $\mathbf{a}(T^{n+1})$.
5. Return to Step 1 until the final integration time is attained.

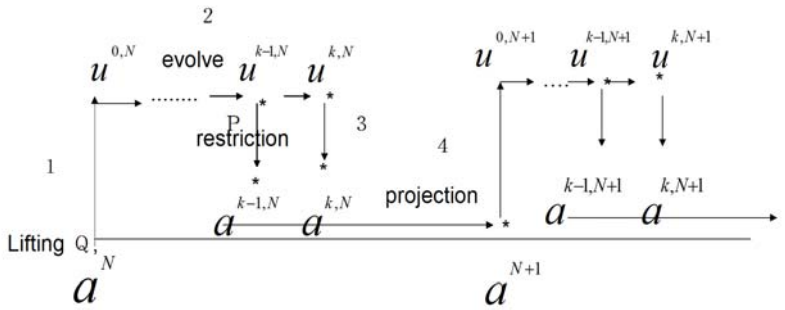


Figure 2. Sketch of the EF POD-assisted method, a^N is the time coefficient of the outer model, $u^{k,N}$ is the variable for the inner model.

Remark: As discussed above, we can obtained the right-hand-side, $\mathbf{Y}(t;a(t))$ of eq. (3.18) from the eq. (2.1a)-(2.1c) by a Galerkin projection. This procedure may result in rather intricate forms and often suffers long-term dynamics, because not only we have to compute complicated ODEs but also POD model may diverge and will approach, after a long-time integration, another erroneous state (see [20, 32]). This "EF" approach can avoid these difficulties by using "just enough" full DNS simulations without requiring the explicit form of the RHS of eq. (3.18).
Having obtained the POD time coefficients for $T^n \le T \le T_*^n = T^n + k\Delta t$ in Step 3, we can obtain the time coefficients $\mathbf{a}$ (T^{n+1}) by using the following formula:

$$\mathbf{a}^{n+1} = \mathbf{a}(T_*^n) + \sum_{l=1}^{I_2} \frac{(\Delta T - T_k)}{l!} \frac{\partial^{(l-1)}}{\partial t^{l-1}} \mathbf{Y}(T_*^n) + O((\Delta T - T_k)^{I_2+1}), \quad (3.23)$$

where I_2 is an integer,

$$\mathbf{Y}(T_*^n) = \sum_{j=0}^{k} \sigma_j \mathbf{a}(T_j^{'}) = \frac{d\mathbf{a}}{dt}(T_*^n) + O(\Delta t^{I_1}),$$
$$\Delta T = T^{n+1} - T^n, T_k = k\Delta t, \quad (3.24)$$
$$1 \le j \le k, T_j^{'} = T_*^n - j\Delta t.$$

where I_1 is the order of the approximation, $\{\sigma_j\}_{j=0}^{k}$ is a set of consistent coefficients, such that $\sum \sigma_j \mathbf{a}(T_j^{'}) = \frac{d\mathbf{a}}{dt}(T_*^n) + O(\Delta t^{I_1})$, and the higher order derivatives of $\mathbf{Y}(t)$ in eq. (3.23) are approximated by way of eq. (3.24). We call eq. (3.23) a projective integration scheme.

4. Convergence and Accuracy

First of all, we consider the convergence of the EF POD-assisted method.

We denote the finite-term POD expansion as,

$$V_K(t, \vec{x}) = \sum_{i=1}^{K} \alpha_i(t) \Phi_i \ , \quad (4.1)$$

and correspondingly, the truncated restriction and truncated lifting operators as Υ_K and Ψ_K, which are subject to,

$$V_K = I_K V \equiv \Psi_K \Upsilon_K V \quad (4.2)$$

We represent eq. (2.1a)-(2.1c) in vector form as follows:

$$\frac{\partial \mathbf{U}}{\partial t}(t, \vec{x}) = \mathbf{f}(t; \mathbf{U}(t, \vec{x})), \quad (4.3)$$

here $\mathbf{U}(t, \vec{x})$ is a vector, consisting 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, ρ is the maximum mesh interval and we applied a numerical scheme that was both consistent and stable. Let $\mathbf{U}$ be the full-order

approximation and $\overline{\overline{\mathbf{U}}}$ the reduced-order approximation resulting from the EF-POD method. A selected scheme of DNS is as follows:

$$\overline{\mathbf{U}}(T^{k+1},\vec{x}) = \overline{\mathbf{U}}(T^{k},\vec{x}) + \Delta t\psi(T^{k};\Delta t,\rho,\cdots) \tag{4.4}$$

Since the scheme was consistent and stable, we had the following requirements, at a given moment $T = T^k$,

$$\lim_{\Delta t,\rho\to 0}\psi(T^{k},\Delta t,\rho,\cdots) = \mathbf{f}(T^{k};\mathbf{U}^{k}) \tag{4.5}$$

$$\lim_{\Delta t,\rho\to 0}\| \overline{\mathbf{U}}^{k} - \mathbf{U}\| = 0\,. \tag{4.6}$$

by virtue of eq. (3.23),

$$\begin{aligned}
&\overline{\overline{\mathbf{U}}}^{N+1} \equiv \overline{\overline{\mathbf{U}}}(T^{N+1}) = \Psi_K\mathbf{a}(T^{N+1})\\
&= \Psi_K[\mathbf{a}(T_*^{N}) + \sum_{l=1}^{I_2}\frac{(\Delta T - T_k)^l}{l!}\frac{\partial^{(l-1)}}{\partial t^{l-1}}\mathbf{Y}(T_*^{n}) + O((\Delta T - T_k)^{I_2+1})]\\
&= \Psi_K[\mathbf{a}(T_*^{N}) + \sum_{l=1}^{I_2}\frac{(\Delta T - T_k)^l}{l!}\sum\nolimits_j \sigma_j\mathbf{a}(T_j) + O((\Delta T - T_k)^{I_2+1})]\\
&= \Psi_K\Upsilon_K[\mathbf{U}(T_*^{N}) + \sum_{l=1}^{I_2}\frac{(\Delta T - T_k)^l}{l!}\sum\nolimits_j \sigma_j\mathbf{U}(T_j) + O((\Delta T - T_k)^{I_2+1})]\\
&= I_K[\mathbf{U}(T^{N}) + \Delta t\sum\nolimits_{m=1}^{k}\psi((N\Delta T + m\Delta t);\mathbf{U}) + \sum_{l=1}^{I_2}\frac{(\Delta T - T_k)^l}{l!}\frac{\partial\mathbf{U}}{\partial t}(T_*^{N})\\
&\quad + O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1})]\\
&= \overline{\overline{\mathbf{U}}}^{N} + \Delta T\phi(T^{N};\mathbf{U},\cdots)
\end{aligned} \tag{4.7}$$

where

$$\begin{aligned}
\phi(T^{N};\mathbf{U},\cdots) = \frac{1}{\Delta T}I_K[&\Delta t\sum\nolimits_{m=1}^{k}\psi((N\Delta T + m\Delta t);\mathbf{U}) + \sum_{l=1}^{I_2}\frac{(\Delta T - T_k)^l}{l!}\frac{\partial\mathbf{U}}{\partial t}(T_*^{N})\\
&+ O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1})]
\end{aligned} \tag{4.8}$$

Since $\Delta T \to 0, T_*^N \to T^N, \Delta t, \Delta T_* = \Delta T - T_k \to 0$, by virtue of eq. (4.5),

$$\begin{aligned}
\lim_{\Delta T,\rho\to 0}\phi(T^{N};\mathbf{U},\cdots) = \lim_{\Delta T,\rho\to 0}\frac{1}{\Delta T}I_K[&\Delta t l\mathbf{f}(T^{N},\mathrm{U}^{N}) + \sum_{l=1}^{I_2}\frac{(\Delta T - T_k)^l}{l!}\frac{\partial\mathbf{U}}{\partial t}(T_*^{N})\\
&+ O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1})]\\
= I_K[&\mathbf{f}(T^{N};\mathrm{U}^{N})]
\end{aligned} \tag{4.9}$$

Let $\mathbf{U}^{POD} = \Psi_K \Upsilon_K \mathbf{U} = I_K \mathbf{U}$ be the numerical solution of the POD method,

$$\| \mathbf{U} - \mathbf{U}^{POD} \| \to Q^{-z}, \text{ as } Q \to \infty,$$

where the real number $z > 0$ denotes the convergence rate. By virtue of eq. (4.9),

$$\lim_{\Delta T, \rho \to 0} \phi(T^N; \mathbf{U}, \cdots) = \mathbf{f}(T^N; \mathbf{U}^N)$$

So we have finished illustrating the consistency theoretically.

We then consider the truncation error θ, let $\mathbf{U}(T^{N+1}) \equiv \mathbf{U}^{N+1}$,

$$\begin{aligned}
\theta &= \| \mathbf{U}^{N+1} - \overline{\overline{\mathbf{U}}}^{N+1} \| \\
&= \| \mathbf{U}^{N+1} - \mathbf{U}^{N+1^{POD}} + \mathbf{U}^{N+1^{POD}} - \overline{\overline{\mathbf{U}}}^{N+1} \| \\
&\le \| \mathbf{U}^{N+1} - \mathbf{U}^{N+1^{POD}} \| + \| \mathbf{U}^{N+1^{POD}} - \overline{\overline{\mathbf{U}}}^{N+1} \| \\
&= \theta_1 + \theta_2
\end{aligned}$$

Clearly, we only need to estimate θ_2,

$$\begin{aligned}
&\overline{\overline{\mathbf{U}}}^{N+1} = \mathbf{U}^{N^{POD}} + \Delta T \phi(T^N; \mathbf{U}, \cdots) \\
&= I_K[\mathbf{U}(T^N) + \Delta t \sum\nolimits_{m=1}^{k} (\psi((N\Delta T + m\Delta t); \mathbf{U}) + \beta) + \sum_{l=1}^{I_2} \frac{(\Delta T - T_k)^l}{l!} \frac{\partial \mathbf{U}}{\partial t}(T_*^N) \\
&\quad + O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1})] \\
&= I_K[\mathbf{U}(T_*^N) + \Delta t l \beta + \sum_{l=1}^{I_2} \frac{(\Delta T - T_k)^l}{l!} \frac{\partial \mathbf{U}}{\partial t}(T_*^N) + O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1})] \\
&= \mathbf{U}^{N+1^{POD}} + I_K[\Delta t l \beta + O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1})] \\
&= \mathbf{U}^{N+1^{POD}} + \lambda_k
\end{aligned}$$

We get

$$\| \theta_2 = \| \mathbf{U}^{N+1^{POD}} - \overline{\overline{\mathbf{U}}}^{N+1} \| = \| \lambda_k \| = \Delta t \beta + O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1})$$

where $\beta = O(\Delta t^{l_1}, \rho^{l_2})$ is the scheme truncation error of the full model.

So the final truncation error is,

$$\theta = \theta_1 + \theta_2 = O(Q^{-z}) + O(\Delta t^{I_1}) + O((\Delta T - T_k)^{I_2+1}) + O(\Delta t^{l_1}, \rho^{l_2}) .$$

5. Numerical results and Error analysis

In this section, we present numerical results of the EF POD-assisted method for the upper tropical Pacific Ocean model. The number of snapshots was selected to be 12, 36, and 60, respectively. To quantify the performance of the reduced-basis method, we used two metrics, namely the root-mean-square error (RMSE) and the correlation of the difference between the full-order and the reduced-order simulations.

This was obtained by first taking twelve-month's of full-order results along with the corresponding twelve-month's reduced-order results and computing the error. For example, for the variable h,

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

where N is the number of the node, the index m denotes the month, $\tilde{h}_m$ is the full-order approximation and h_m is the reduced-order approximation. The average RMSE is defined as,

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

and the correlation is defined as,

$$CORRELATION_{hm} = \frac{\sum_{i=1}^{N} (\tilde{h}_m(z_i) - \overline{\tilde{h}}(z_i))(h_m(z_i) - \overline{h}(z_i))}{\sqrt{\sum_{i=1}^{N} (\tilde{h}_m(z_i) - \overline{\tilde{h}}(z_i))^2 \sum_{i=1}^{N} (h_m(z_i) - \overline{h}(z_i))^2}}$$

First of all, we considered snapshots in the case of 36, and others were similar. Let the large-scale time step be taken as ΔT =5 hours, while the short time step as Δt =100 seconds. We let I_2 =1 in the projective integration scheme of Eq. (3.23), and we researched the effect on RMSE when k changes in $T_k = k\Delta t$, where k is the number of iterations in the full model during one large time step. From Table 2, we can find that the RMSE decreases when T_k increases.

Table 2. The RMSE of h, u, v for different numbers of iterations, k ($T_k = k\Delta t$) in the full model using a large time step ΔT .

T_k	h	u	v
$\frac{\Delta T}{18}$	0.41024467	0.00306552	0.00173173
$\frac{\Delta T}{9}$	0.40551540	0.00302744	0.00160071
$\frac{\Delta T}{6}$	0.40461191	0.00301344	0.00153364

To save computation cost, we select $T_k = \frac{\Delta T}{18}$ (k =10), considering varied large time steps, ΔT . From Table 3, we find that it is appropriate for the large time step to be chosen as either 5 or 10 hours. That is, a better result is obtained when the ratio between large and short time steps is either 180:1 or 360:1.

Table 3. The RMSE of h, u, v for different large time steps.

ΔT	h	u	v
5 hours	0.41024467	0.00306552	0.00173173
10 hours	0.40536597	0.00302397	0.00157455
1 day	0.61680775	0.00434816	0.00220632

Let ΔT =5 hours, $T_k = \frac{\Delta T}{18}$, and we research I_2 =1, 2, 3 respectively, using the projective integration of eq. (3.23). The RMSEs are shown in Table 4.

Table 4. The RMSE for different I_2 using the projective integration scheme.

	I_2 =1	I_2 =2	I_2 =3
h	0.40551540	0.40523764	0.40523800
u	0.00302744	0.00302595	0.00302595
v	0.00160071	0.00159335	0.00159336

We also researched different snapshots for 12, 36, and 60 (see Table 5).

Table 5. The RMSE of h, u, v for calculations using 12, 36, 60 snapshots, respectively.

	h	u	v
12	0.42748689	0.00264911	0.00128677
36	0.41024467	0.00306552	0.00173173
60	0.41343936	0.00322827	0.00189658

We compared the EF POD with the POD method, and found that the EF POD resulted in smaller RMSE (see Table 6).

Table 6. Comparison between the full POD and EF-POD methods using the RMSE of h, u, v as the comparison metric.

	h	u	v
POD	0.89352083	0.00811650	0.00669807
EF-POD	0.40551540	0.00302744	0.00160071

The better correlations of the EF POD-assisted method were also obtained when compared with that of the POD in different month (see Table 7).

Table 7. Correlation comparison between the full POD and EF-POD methods for (a) upper layer thickness h (unit: m), (b) zonal current velocity u (unit: m/s) and (c) meridional current velocity v (unit: m/s).

(a)

	1	2	3	4	5	6	7	8	9	10	11	12
POD	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
EF-POD	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9

(b)

	1	2	3	4	5	6	7	8	9	10	11	12
POD	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8
EF-POD	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9

(c)

	1	2	3	4	5	6	7	8	9	10	11	12
POD	0.8	0.9	0.8	0.6	0.8	0.9	.08	0.9	0.9	0.9	0.8	0.8
EF-POD	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9

The main difference between the EF POD and the POD methods is in the manner in which the time coefficient is obtained; here we also compared the time coefficient of both methods with that of the empirical orthogonal function (EOF). Since the first mode captures the majority of the energy, we only analyzed the first mode time coefficient about h, u, v.

From fig. 3-5, we can find the time coefficients based on the EF POD were very similar to that of EOF, with the EF POD method providing a more convenient manner of extracting them.

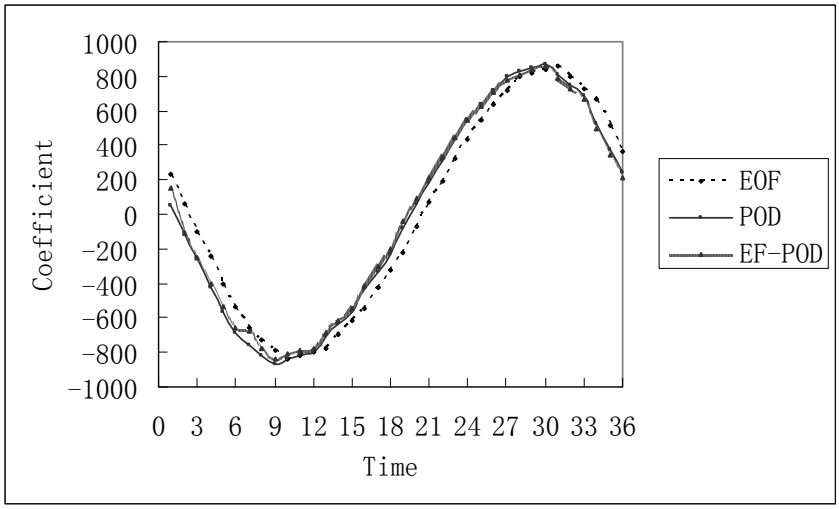


Figure 3. The first-mode time coefficient of h over one year, which are from EOF, POD and EF-POD respectively.

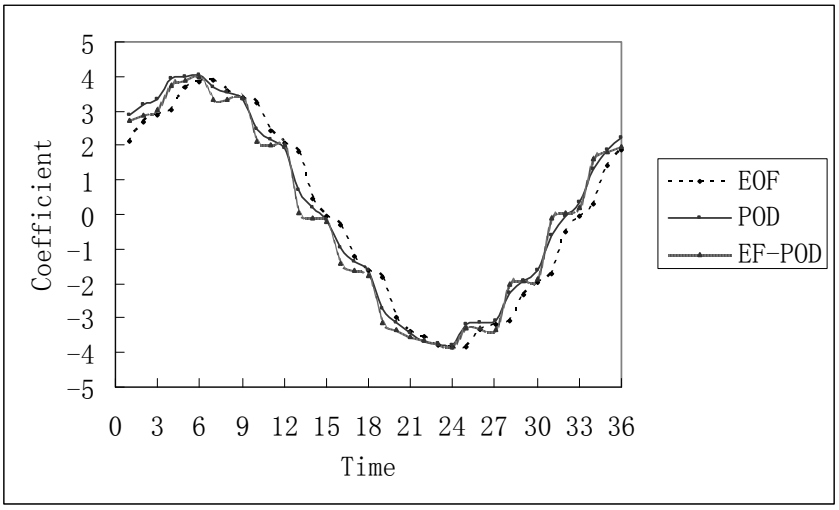


Figure 4. The same to the figure 3, the result is from the zonal velocity u.

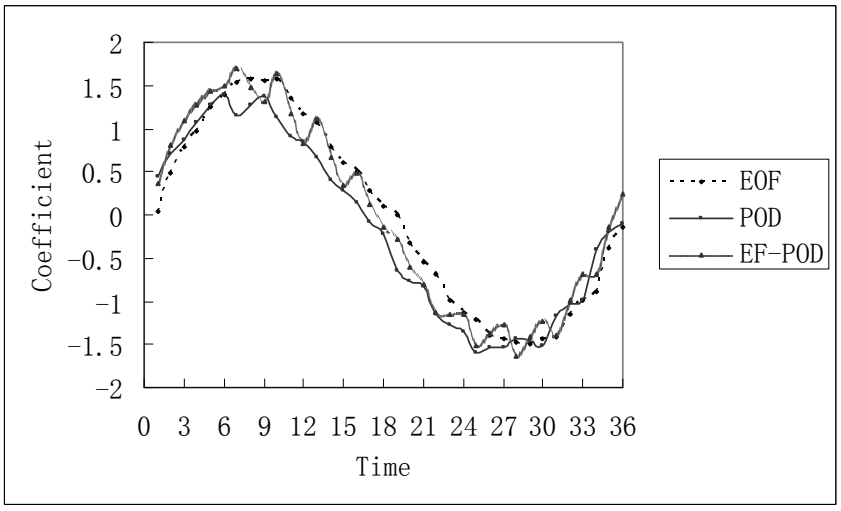
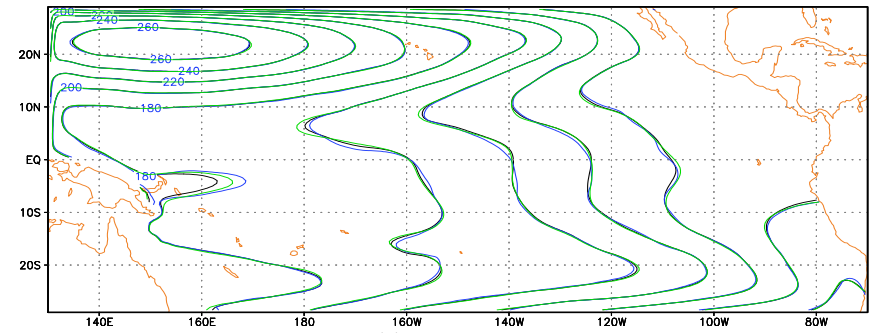
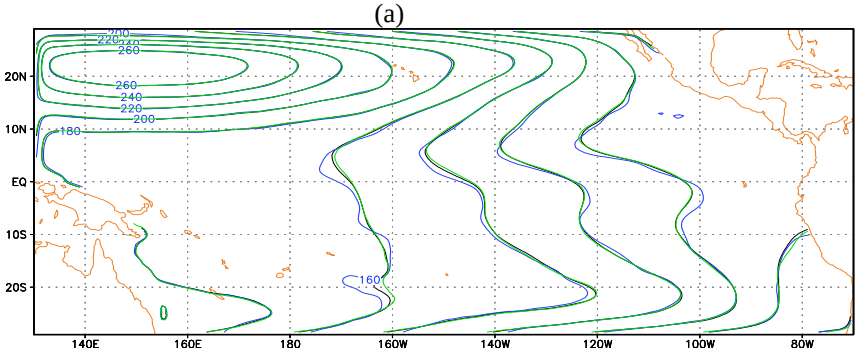


Figure 5. The same to the figure 3, the result is from the meridional velocity v.

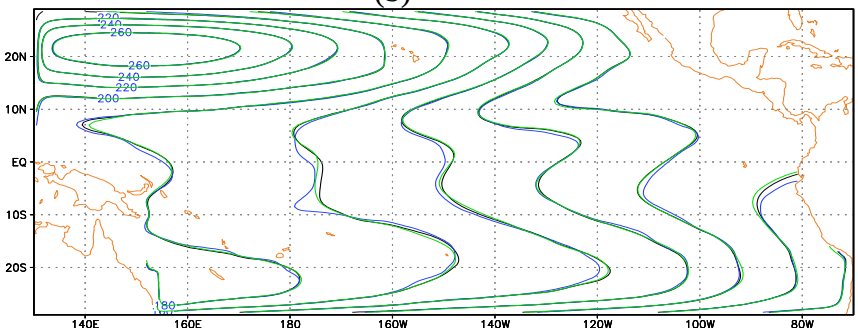
A comparison between the full and reduced orders is displayed in Fig. 6 about the upper layer thickness and in Fig. 7 and Fig. 8 about the streamline figure of velocity.


20N
10N
EQ
10S
20S
140E
160E
180
160W
140W
120W
100W
80W

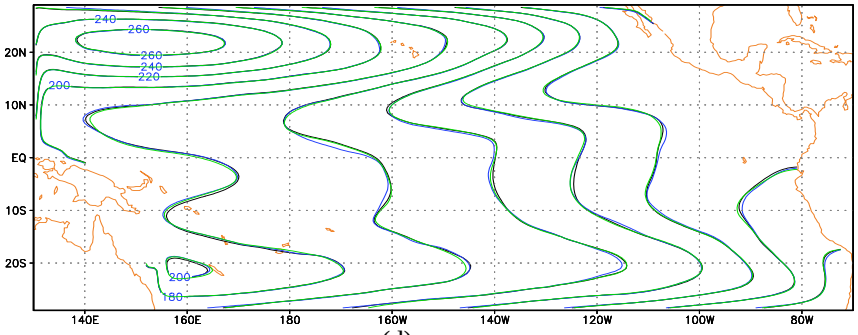
(a)


20N
10N
EQ
10S
20S
140E
160E
180
160W
140W
120W
100W
80W

(b)

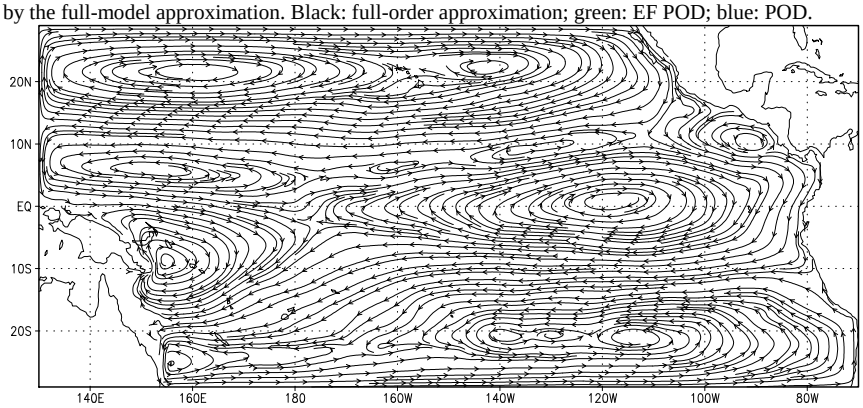

20N
10N
EQ
10S
20S
140E
160E
180
160W
140W
120W
100W
80W

(c)

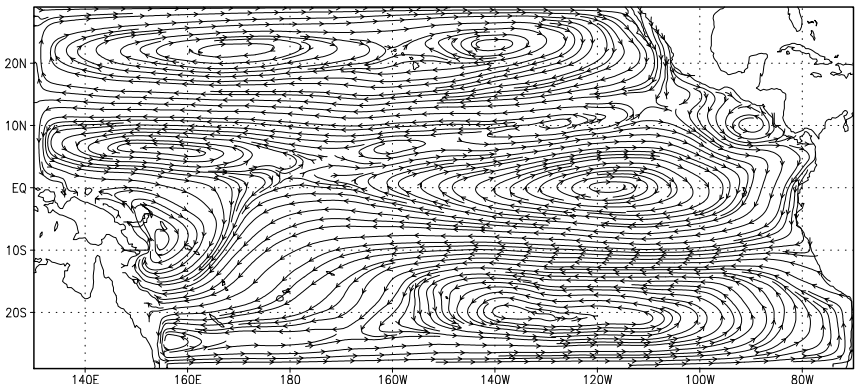


(d)

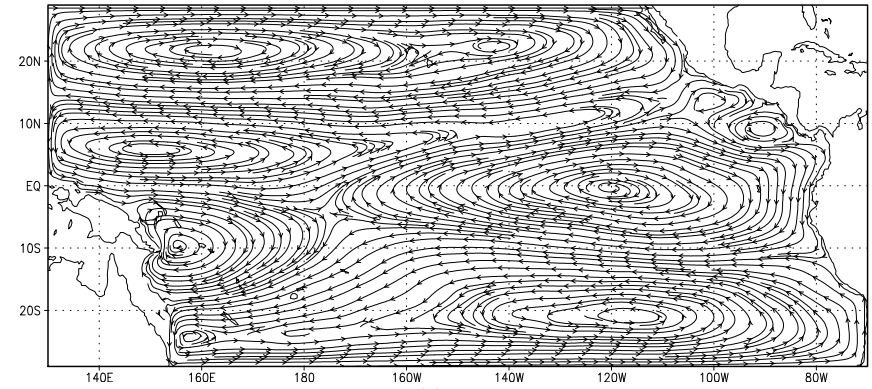
Figure 6. Upper layer thickness h in (a) January, (b) April, (c) July and (d) November for the case of 36 snapshots. The number of the POD basis was determined by capturing 99% of energy produced by the full-model approximation. Black: full-order approximation; green: EF POD; blue: POD.



(a)

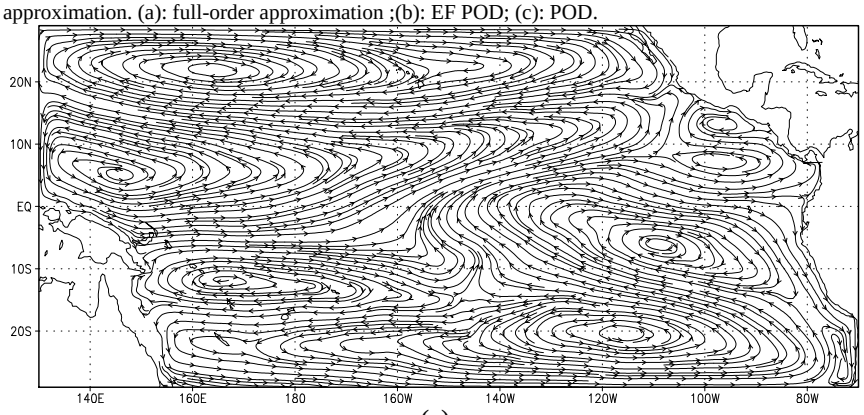


(b)

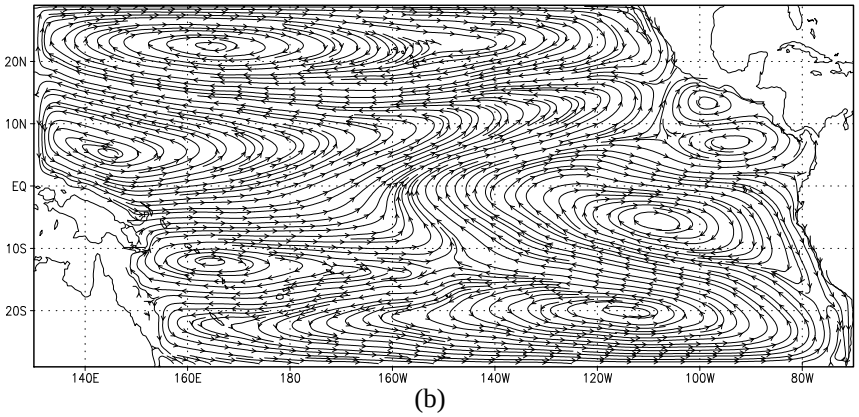


(c)

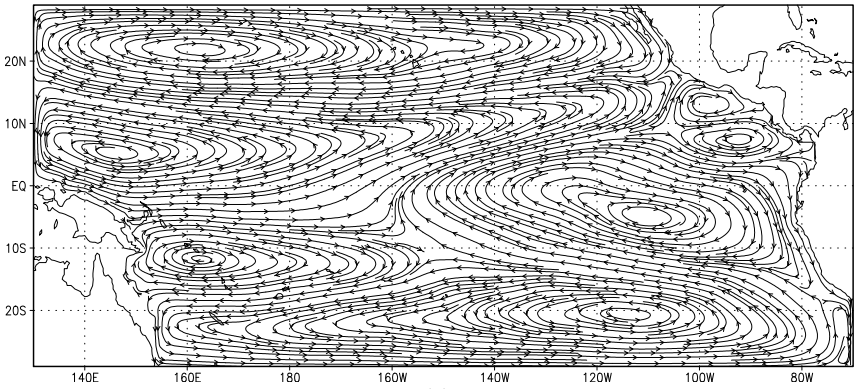
Figure 7. Streamline figure of the velocity during June for the case of 36 snapshots. The number of the POD basis was determined by capturing 99% of the energy produced by the full-model approximation. (a): full-order approximation ;(b): EF POD; (c): POD.



(a)



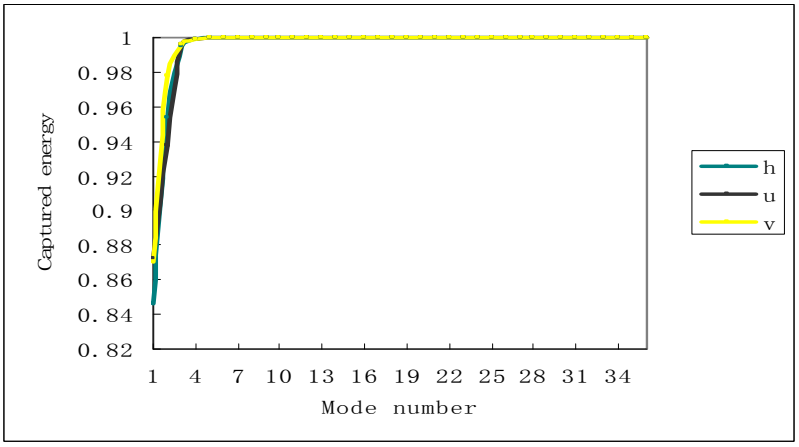
(b)



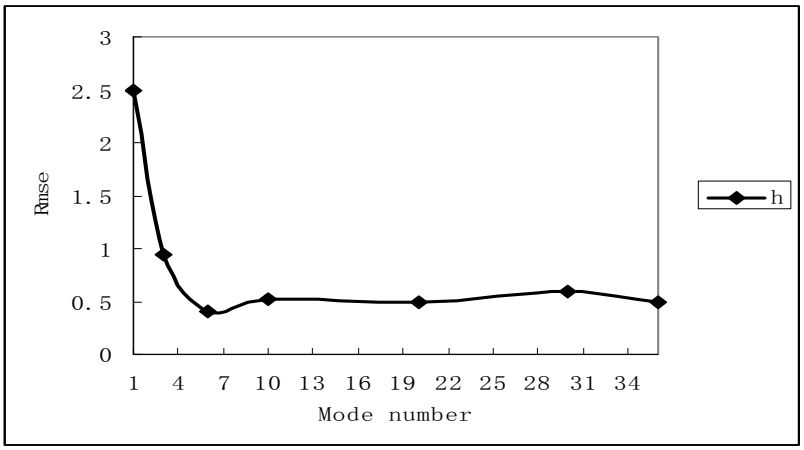
(c)

Figure 8. The same to the figure 7, the results are from the December.

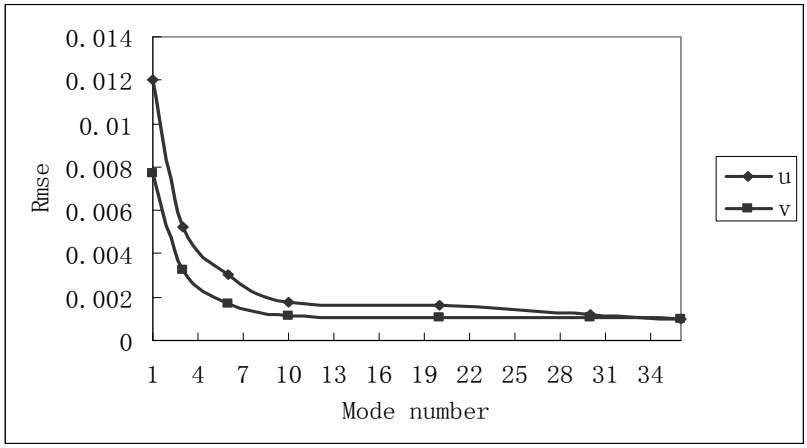
The modes that capture 99% energy are necessary to construct POD models, especially, if a satisfying result was obtained when the number of POD modes ($M = 6$ for h) is employed (Fig. 9).



(a)



(b)



(c)

Figure 9. (a) The POD-modes capture energy, (b) the relationship between the mode number and the RMSE for the upper layer thickness, h and (c) the relationship between the mode number and the RMSE for the velocity u,v.

Finally, to find the difference for the different number of snapshots, we show the upper layer thickness in June and December for the cases of 12, 36 and 60 snapshots in Fig. 10. It is evident that the difference is very small.

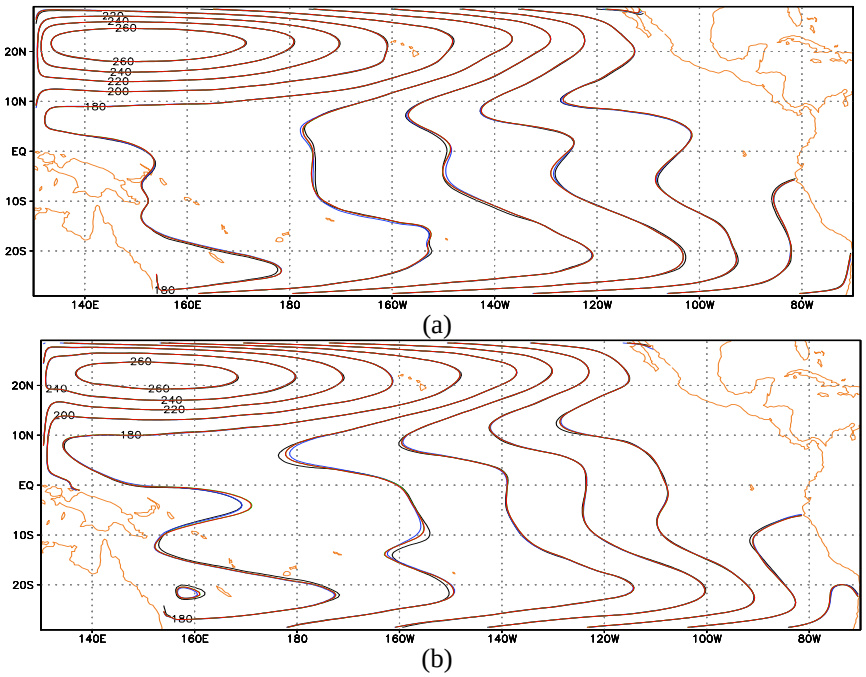


Figure 10. Upper layer thickness during (a) June and (b) December for the case of 12, 36 and 60 snapshots, using both the full-model approximation and the reduced-order approximation. Black : full-order approximation, red : 12 snapshots, green : 36 snapshots, blue : 60 snapshots

6. Summary

We applied the EF POD method to the reduced modeling of a large-scale upper ocean circulation in the tropical Pacific domain. We discussed the convergence and accuracy, and we analyzed the factors that affect the results in the numerical experiment. To quantify the performance of the EF POD method, we used two metrics, namely the RMSE and the correlation of the difference between the full-order and the reduced-order simulations. The main findings are as follows: when the ratio between the large and short time steps was approximately 180:1 or 360:1, we found that better results were collected via EF POD than with POD alone. When we select the ratio between these two time steps to be about 180:1, the computation cost was reduced dramatically. We found that the modes that capture 99% of the energy were required to construct the POD models, and the number of POD modes ($M = 6$ for h) was the best choice in this model. The RMSE for the upper ocean layer thickness was about 0.3% of the average thickness and the correlations between the upper layer thickness with that from

the EF POD model was around 0.99. The size of the RMSE of the velocity between the full model and the reduced model was about 10^{-3}, and the correlation was about 0.9. The results obtained above were all better than those that were obtained using the full POD method. Another advantage for EF POD method was its ability to allow one to bypass the derivation of the RHS of the macroscopic evolution equation.

References

1. G. Berkooz, P. Holmes and J. L. Lumley, *Ann. Rev. Fluid Mech*. **25**, 539(1993).
2. L. Sirovich, *Quart. Appl. Math*. **XLV**, 56(1987).
3. Karhunen, K., *Ann. Acad. Sci.* 37(1946).
4. M. Loeve, *Compte Rend. Acad. Sci*. 220(1945).
5. R.E. Arndt, D.F. Long and M. N. Glauser, *J. Fluid Mech*. **340**, 1(1997).
6. J.H. Citriniti and W.K. George, *J. Fluid Mech*. **418**, 13(2000).
7. J. Delville, L. Ukeiley, L. Cordier, J.P. Bonnet, and M. Glauser, *J. Fluid Mech*. **391**, 91(1999).
8. A.Glezer, Z.Kadioglu, and A.J.Pearlstein, *Phys. Fluids*. **1(8)**, 1363(1989).
9. Yanhua Cao, Jiang Zhu, Zhendong Luo, Navon, I. M., *Computers and Mathematics with Applications*. **Vol 52, Issues 8-9**, 1373(2006).
10. N. Aubry, P. Holmes J. L. Stone and J.L. Lumley, *J. Fluid. Mech.* **340**, 1(1997).
11. W.Cazemier, R. W. Verstappen and A.E.Veldman, *Phys. Fluids*. **10(7)**, 1685(1998).
12. A.E.Deane, I.G. Kevrekidis, G .E. Karniadakis, and S.A.Orszag, *Phys. Fluids.* **A, 3(10)**, 2337(1991).
13. A.Liakopoulos, P.A.Blythe and H.Gunes, *Proc. R. Soc. Lond*. A, 453(1997).
14. Courtier, P., Th'epaut, J.-N. and Hollingsworth, A, *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367 (1994).
15. Griewank, A., *SIAM*. (2000).
16. Li, Z., Navon, I.M., and Yanqiu Zhu, Y., *Monthly Weather Review*. **128 (3)**, 668 (2000).
17. Gauthier, P., *Earth and Environmental Sciences.* **Vol. 26**,167 (2003).
18. Restrepo, J., Leaf, G., Griewank, A., *SIAM J. Sci. Comput*. **19**, 1586(1998).
19. Cao, Y., Zhu, J., Navon, I. M, and Luo, Z.D, *Int. J. Num. Methods in Fluids.* **Vol 53, Issue 10**, 1571(2007).
20. C.Foias, M.S.Jolly, I.G. Kevrekidis and E.S. Titi., *Nonlinearity.* **4**, 591(1991).

21. S. Sirisup, D. Xiu, GE Karniadakis and IG Kevrekidis, *J. Comput. Phys.* **207**, 568(2005).
22. C. Theodoropoulos, Y. Qian and I.G. Kevrekidis, *Proc. Natl. Acad. Sci*. **97**, 2000.
23. I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, and C. Theodoropoulos, *Comm. Math. Sci*. **1(4)**, 715(2003).
24. I.G. Kevrekidis, C.W. Gear and G. Hummer, *A. I. Ch. E. Journal*. **50(7)**, 1346(2004).
25. C.W. Gear and I.G. Kevrekidis, *SIAM J. Sci. Comput*. **24**, 1091(2003).
26. C.W. Gear and I. G. Kevrekidis, *J. Comput. Phys*. **187(1)**, 95(2003).
27. R. Rico-Martinez, C.W. Gear, and I.G. Kevrekidis, *J. Comput. Phys*. **196**, 474(2004).
28. I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, and C. Theodoropoulos, Comm. Math. Sci. **1(4)**, 715(2003).
29. I.G. Kevrekidis, C.W. Gear, and G. Hummer, *A. I. Ch. E Journal*. **50(7)**, 1346(2004).
30. A. Makeev, D. Maroudas, and I.G. Kevrekidis, J. Chem. Phys. Chem. Phys. **116**, 10083(2002).
31. A. Makeev, D. Maroudas, A. Panagiotopoulos, and I.G. Kevrekidis, *J. Chem. Phys*. **117**,8229(2002).
32. S. Sirisup and G. E. Karniadakis, *J. Comput. Phys*. **194(1)**, 92(2004).
33. C. W. Gear, IGK and C. Theodoropoulos, Comp. Chem. Engng. **26**, 941(2002).
34. Tremolet, Y., *Q. J. R. Meteorol. Soc*. **130**, (601), 2233(2004).
35. Fukunaga K. Introduction to Statistical Recognition, Academic Press, 1990.
36. Ly H V and Tran H T, Technical Report CRSC-TR98-12., 1998.
37. Sirovich L, *Q. Appl. Math*. **XLV (3)**, 561(1987).
38. Cane, M. A, *J. Mar. Res*.37, 233(1979).
39. Seager, R., S. E. Zebiak, and M. A. Cane, *J. Geophys. Res*. 93, 1265(1988).
40. Moore, D. W., and S. G. H. Philander, *The Sea*, **Vol. 6**, *Goldberg et al., Eds., Wiley*, 319(1978).
41. Stricherz, J. N., J.J. O'Brien, and D. M. Legler, A mesoscale Air-Sea Interaction Group Technique Report. 216pp, 1992.
42. C. Foias, M.S. Jolly, I.G. Kevrekidis, E.S. Titi, *Nonlinearity*. **4**, 591(1991).