Reduced order ocean model using proper orthogonal decomposition

D.A. Salas-de-León and M.A. Monreal-Gómez

Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México,

E. van-de-Ven and S. Weiland

Electrical Engineering Department, Technische Universiteit Eindhoven,
Den Dolech 2, 5612 AZ Eindhoven, Netherlands.

D. Salas-Monreal

Centro de Ecología y Pesquerías, Universidad Veracruzana,
Calle Hidalgo No. 617, Col. Río Jamapa, CP 94290, Boca del Río, Veracruz, México.

Recibido el 2 de febrero de 2009; aceptado el 15 de abril de 2009

The proper orthogonal decomposition (POD) is shown to be an efficient model reduction technique for simulating physical processes governed by partial differential equations. In this paper, a POD reduced model of a barotropic ocean circulation for coastal region domains was made. The POD basis functions and the results from this POD model were constructed and compared with that of the original model. The main findings were: 1) the variability of the barotropic circulation obtained by the original model is well captured by a low dimensional system of order of 22, which is constructed using 15 snapshots and 7 leading POD basis functions; 2) the RMS errors for the POD model is of order $10^{-4}$ and the correlations between the original results with that from the POD model of more than 0.99; 3) the CPU model time solution is reduced is five times less than the original one; and 4) it is necessary to retain modes that capture more than 99% of the energy is necessary in order to construct POD models yielding a high accuracy.

Keywords: POD; reduced order model; PDE; Galerkin methods; EDP.

La descomposición ortogonal propia (POD) es una técnica eficiente para la reducción de los modelos que describen procesos físicos gobernados por ecuaciones diferenciales parciales. En este trabajo, se hace una reducción de un modelo barotrópico de circulación costera del océano mediante POD. Se construyen las funciones bases y los resultados de la aplicación de la reducción mediante POD se comparan con la solución original del modelo barotrópico. Los principales resultados son: 1) Se reproduce la variabilidad de la circulación barotrópica obtenida con el modelo original con un sistema de baja dimensión de orden 22, el cual se construye usando 15 aproximaciones y 7 funciones básicas de POD. 2) El error (RMS) del modelo mediante POD es del orden de $10^{-4}$ y la correlación entre los resultados del modelo original y los obtenidos mediante la aproximación POD es más de 0.99. 3) El tiempo (CPU) de obtención de la solución mediante POD es cinco veces menor que con el esquema original de solución y 4) Es necesario construir modelos POD que retengan más del 99% de la energía del sistema para que estos sean aceptables.

Descriptors: POD; reducción del orden de modelos; PDE; métodos de Galerkin; EDP.

PACS: 90; 95.75.-z; 95.75.Pq; 92.10.-c; 92.10.Sx

1. Introduction

The proper orthogonal decomposition (POD) is an efficient way to carry out reduced order modelling by identifying the few most energetic modes in a sequence of snapshots from a time-dependent system, and providing a means of obtaining a low-dimensional description of the system’s dynamics. Since its original introduction by Loeve in 1945 [1] and Karhunen in 1946 [2], the method has been extensively used in research in recent years and successfully applied to a variety of fields. One of these important applications was the application to spatially organized motions in fluid flows, such as cylinder flows [3]. POD was also used for identification of coherent structures, signal analysis and pattern recognition [4,5]. Many researchers have also applied the POD technique to optimal control problems. For instance, this method has been used for Burger’s equation [6-8], the Ginzburg-Landau equation and the Bénard convection [9], and in other fluid control problems [10-15]. More recently POD has also been used in inverse problems [16]. In addition, the method has also been used for industrial applications such as supersonic jet modelling [17], thermal processing of foods [15,18], and studies of the dynamic wind pressures acting on buildings [19], to name but a few.

For a comprehensive description of POD theory and state of the art POD research, see Gunzburger [20]. Compared with the above efforts, little attention was paid to application of POD to geophysical fluid dynamics such as atmospheric or oceanic systems. In general these dynamic systems are quite complex and their discrete models are hard to solve due to their large dimensions (typically $10^6$- $10^8$).

In this study, we make a POD reduced modelling of a barotropic ocean circulation model for coastal region domains. We construct the POD basis functions and the results from this POD model are compared with that of the original barotropic model.
2. Description of the barotropic ocean model

The numerical model used in this paper is the Salas-de-León and Monreal-Gómez’s barotropic model [21] with variable-depth and free surface layer, which is studying the tidal currents, wind forcing currents, and free surface dynamics in coastal regions.

The model is a non-linear transport model, consisting of one layer above the maximum depth with the same constant density in the layer (Fig. 1). The equations for the depth-averaged transport are [21]:

\[
\begin{align*}
\frac{\partial \vec{V}}{\partial t} + \nabla \left( h^{-1} \vec{V} \nabla \vec{V} \right) + \vec{k} \times \vec{f} \vec{V} &= -h \frac{\nabla p}{\rho} + \vec{\tau}_b - \vec{\tau}_s + \nu_H \nabla^2 \vec{V} \\
\frac{\partial h}{\partial t} + \nabla \cdot \vec{V} &= 0 \\
\frac{\nabla p}{\rho} &= g \nabla \eta
\end{align*}
\]  

(1)

where
- \( \vec{V} \) Horizontal transport components of the depth-averaged currents 
- \( \vec{V} = U \hat{i} + V \hat{j} ; U = uh, V = vh \)
- \( t \) Time
- \( \vec{\tau}_b \) Bottom friction force
- \( \vec{\tau}_s \) Wind stress force
- \( p \) Pressure
- \( h \) Total layer thickness
- \( H \) Average depth
- \( f \) Coriolis parameter \( (f = 2\omega \sin \phi) \), \( \omega \) angular velocity of the earth and \( \phi \) latitude

Figure 1. Geometry of the model.

Figure 2. The study region, left panel is the bathymetry (depths) of the region in meters.

\begin{align*}
\eta & \quad \text{Free surface anomaly} \ (\eta = h - H) \\
\rho & \quad \text{Density of water} \\
\nu_H & \quad \text{Horizontal eddy viscosity}
\end{align*}
where \( \rho \) is the density of the air, \( C_W \) the wind stress drag coefficient that depends on the wind velocities [22], \( \vec{v} \) the water velocity vector, and \((u_s, v_s)\) the components of the wind velocity.

First step:

\[
V^{n+\frac{1}{2}} = V^n - \frac{\Delta t}{2} \left[ \frac{\partial}{\partial x} \left( \frac{UV}{h} \right) \right]^n - \frac{\Delta t}{2} \left[ \frac{\partial}{\partial y} \left( \frac{V^2}{h} \right) \right]^n - \frac{\Delta t g}{2} \left[ h \frac{\partial h}{\partial y} \right]^n - (f_o + \beta y) \frac{\Delta t}{2} U^n + \frac{v_H \Delta t}{2} \left( \frac{\partial^2 V}{\partial x^2} \right)^n
\]

\[
+ \frac{v_H \Delta t}{2} \left( \frac{\partial^2 V}{\partial y^2} \right)^n + \frac{\Delta t}{2} \left( \frac{\tau_y}{\rho} \right)^n - \frac{D \Delta t}{2} \left( \frac{\vec{V}}{h^2} \right)^{n+\frac{1}{2}}
\]

\[
U^{n+\frac{1}{2}} = U^n - \frac{\Delta t}{2} \left[ \frac{\partial}{\partial x} \left( \frac{U V}{h} \right) \right]^n - \frac{\Delta t}{2} \left[ \frac{\partial}{\partial y} \left( \frac{U^2}{h} \right) \right]^n - \frac{\Delta t g}{2} \left[ h^n \frac{\partial h}{\partial y} \right]^{n+\frac{1}{2}}
\]

\[
+ (f_o + \beta y) \frac{\Delta t}{2} V^{n+\frac{1}{2}} + \frac{v_H \Delta t}{2} \left( \frac{\partial^2 U}{\partial x^2} \right)^n + \frac{v_H \Delta t}{2} \left( \frac{\partial^2 U}{\partial y^2} \right)^n + \frac{\Delta t}{2} \left( \frac{\tau_x}{p} \right)^n - \frac{D \Delta t}{2} \left( \frac{\vec{V}}{h^2} \right)^n U^{n+\frac{1}{2}}
\]

\[
\eta^{n+\frac{1}{2}} = \eta^n - \frac{\Delta t}{2} \left( \frac{\partial U}{\partial x} \right)^{n+\frac{1}{2}} - \frac{\Delta t}{2} \left( \frac{\partial V}{\partial x} \right)^n
\]

The bottom friction coefficients \((|\tau_b|)\) are approximated by [23]:

\[
|\tau_b| = \frac{C_b |\vec{v}| (u_b, v_b)}{\rho h^2}
\]  

where \( \rho \) is the water density, \( C_b \) the bottom friction coefficient that depends on the water velocities that is of the order of \(10^{-3}\) [23], \( \vec{v} \) the water velocity vector, and \((u_b, v_b)\) the components of the water velocity vector.

3. The study region

The barotropic model was successfully applied to a coastal lagoon in the Mexican Caribbean in order to depict the current pattern induced by tides and winds [24]. The coastal lagoon is located in the Sian Ka’an biosphere reserve in the Mexican Caribbean (Fig. 2), and has a north-south length of approximately 88 km, and a maximum width of 3 km. The maximum depth is 6 m, with an average depth of 1.5 m.

The system has two mouths or connections with the adjacent sea, one to the south (Boca Grande) and the other to the northwest (Boca Paila). Water exchanges between the open sea and the lagoon are produced at both mouths, and are forced by tides and the wind stress.

3.1. Numerical scheme

The dynamical model Eq. (1) are governed by long wave dynamics such as tides, via ocean co-oscillations. In addition, the chosen model allows high frequencies waves to be excited by the applied wind forcing [21].

System (1) was semi-implicit finite differences approximate in order to solve it numerically in a modified staggered stencil of the Arakawa-C scheme [21] (Fig. 3).

The resulting approximation is:
Second step:

\[
U^{n+1}_T = U^{n+\frac{1}{2}} - \frac{\Delta t}{2} \left[ \frac{\partial}{\partial x} \left( \frac{U^2}{h} \right) \right]^{n+\frac{1}{2}} - \frac{\Delta t}{2} \left[ \frac{\partial}{\partial y} \left( \frac{U V}{h} \right) \right]^{n+\frac{1}{2}} - \frac{\Delta t g}{2} \left[ h \left( \frac{\partial \eta}{\partial y} \right) \right]^{n+\frac{1}{2}} + (f_o + \beta y) \frac{\Delta t}{2} V^{n+\frac{1}{2}}
\]

\[
+ \frac{v_H \Delta t}{2} \left( \frac{\partial^2 U}{\partial x^2} \right)^{n+\frac{1}{2}} + \frac{v_H \Delta t}{2} \left( \frac{\partial^2 U}{\partial y^2} \right)^{n+\frac{1}{2}} + \frac{\Delta t}{2} \left( \frac{\tau_x}{\rho} \right)^n - D \frac{\Delta t}{2} \left( \frac{V}{h^2} \right)^{n+\frac{1}{2}} V^{n+1}
\]

\[
V^{n+1} = V^{n+\frac{1}{2}} - \frac{\Delta t}{2} \frac{\partial}{\partial x} \left( \frac{U V}{h} \right)^{n+\frac{1}{2}} - \frac{\Delta t}{2} \frac{\partial}{\partial y} \left( \frac{V^2}{h} \right)^{n+\frac{1}{2}} V^{n+1} - \frac{\Delta t g}{2} \left[ h^{n+\frac{1}{2}} \left( \frac{\partial \eta}{\partial y} \right)^{n+1} \right] - (f_o + \beta y) \frac{\Delta t}{2} U^{n+1}
\]

\[
+ \frac{v_H \Delta t}{2} \left( \frac{\partial^2 V}{\partial x^2} \right)^{n+\frac{1}{2}} + \frac{v_H \Delta t}{2} \left( \frac{\partial^2 V}{\partial y^2} \right)^{n+1} + \frac{\Delta t}{2} \left( \frac{\tau_y}{\rho} \right)^{n+\frac{1}{2}} - \frac{\Delta t}{2} \frac{D}{2} \left( \frac{V}{h^2} \right)^{n+\frac{1}{2}} V^{n+1}
\]

\[
\eta^{n+1} = \eta^{n+\frac{1}{2}} - \frac{\Delta t}{2} \frac{\partial}{\partial x} \left( \frac{U}{h} \right)^{n+\frac{1}{2}} - \frac{\Delta t}{2} \frac{\partial}{\partial x} \left( \frac{V}{h} \right)^{n+1}
\]

where

\[
\beta = \frac{\partial f}{\partial y}
\]

Conditions at the solid boundaries are no-normal flow and no-slip conditions, and at the open boundaries the amplitude and phase of the M2 tidal signal and in the free surface the wind stress. The time integration uses a leapfrog scheme.

The spatial interval for the dynamical model was chosen to be 20 m and the time step to be 745.2 s, which is \(\frac{1}{60}\) of the M2 tidal period (12.42 h). This temporal-spatial resolution will make it possible to resolve the M2 tidal wave characteristic and make the model integration numerically stable. It takes about 5 tidal cycles for the model to reach a periodic constant cycle at that time. The model was calibrated using current velocities measured in situ with an acoustic Doppler current profiler (ADCP). Numerically tidally-driven currents during flood and ebb tides are shown in Fig. 4. Velocities reach their highest values near the openings and along the channels. Results of the model agree well with observed currents (more that 0.85 correlation) [24].

4. Proper orthogonal decomposition

4.1. Fundamentals of the proper orthogonal decomposition

For simplicity the proper orthogonal decomposition in the context of scalar fields was introduced: A complex-valued functions defined on an interval \(\Omega\) of the real line. The interval might be the width of the flow, or the computational domain. We restrict ourselves to the space of functions that are square integral, or, in physical terms, fields with finite kinetic energy on this interval so we need an inner product given by [25]:

\[
(f, g) = \int_{\Omega} f(x)g^*(x) \, dx
\]

and a norm:

\[
\|f\| = (f, f)^{\frac{1}{2}}
\]

That is, we find the member that maximizes the normalized inner product with the field \(v\), which is most nearly parallel in function space. This is a classical problem in the calculus of variations where a necessary condition is that \(\phi\) be an eigenfunction of the two-point correlation tensor given by [25]:

\[
\int \langle u(x) u(x') \rangle \phi(x') \, dx' = \lambda \phi(x)
\]

The integral is from 1 to infinity. Almost every member, in a measure sense, of the ensemble may be reproduced by a modal decomposition in the eigenfunctions [10]:

\[
u(x) = \sum_k a_k \phi_k(x)
\]

Equation (10) is the proper orthogonal decomposition.

4.2. Approximation based on n-th order truncation

Spectral Decomposition is based on Fourier expansion [2]:

\[
U(p, t) = \sum_{j=1}^{\infty} a_{uj}(t) \varphi_{uj}(p)
\]

and the approximation based on n-th order truncation is:
4.3. Snapshot creation

To find an optimal compressed description first we proceed to a series of expansion in terms of a set of basis functions. Intuitively, the basis functions should represent the members of the ensemble in some sense. Such a coordinate system is provided by the Karhunen-Loève expansion. Actually here the basis functions $\Phi$ is a mixture of the snapshots so we take snapshots at appropriate points in time:

$$T (\text{snap}_n) = \begin{pmatrix}
U(p_1, t_1) & U(p_1, t_2) & \cdots & U(p_1, t_f) \\
U(p_2, t_1) & U(p_2, t_2) & \cdots & U(p_2, t_f) \\
\vdots & \vdots & \ddots & \vdots \\
U(p_n, t_1) & U(p_n, t_2) & \cdots & U(p_n, t_m)
\end{pmatrix}$$

We denote by $U(p_n, t_m)$, the set of observations (also called snapshots) of some physical process taken at appropriate points in time at positions $i = 1, 2, k$. In this section, we consider the discrete Karhunen-Loève expansion to find an optimal representation of the ensemble of snapshots. In general, each sample of snapshots $U(p_n, t_m)$ which is defined on a set of $n \times m$ node, where $U(p_n, t_m)$ represent components of a vector.

4.4. Missing point estimation

The method to calculate time-variant matrices faster is based on pre-known spatial information in the orthogonal bases [26]. Actually here the basis function $\Phi$ is a mixture of the snapshots. Thus, with the POD mode computed, one must solve an $m \times m$ eigenvalue problem. For a discretization of an ocean region, the dimension often exceeds, so it is often not feasible for the direct solution of this eigenvalue problem. The eigenvalue problem can be transformed into an $m = 10^4$, $m \times m$ that is an $n \times n$ eigenvalue problem [9]. The $n \times n$ eigenvalue problem can be solved with the method of snapshots.

At this moment we must calculate error for every point. We select the $k$ out of $n$ points with the greatest error as [26]:

$$e(X_0) = \| \Phi^T \Phi - I \|$$

5. Results and discussion

In this section, we report the results of the numerical computations related to the approaches presented in the previous paragraphs. The POD method is applied to the above tidal and wind stress model for a coastal tropical lagoon in the Mexican Caribbean. This method can provide a systematic way of creating a reduced basis space with the state of the system at different time instances and different space locations. As in general reduced order basis methods, one can derive the states from full order numerical computations and should be sufficiently large so that the snapshots may contain all the salient features of the dynamics being considered. Therefore, through a nonlinear Galerkin procedure the POD basis functions with the original dynamics offer the possibility of achieving a high fidelity model (albeit) with a possible large dimension.
To achieve model reduction, we carry out a nonlinear Galerkin procedure with the set of elements. How to choose the values of the nonlinear Galerkin transformation is a crucial question. The associated POD eigenvalues should define a relative information content to choose a low-dimensional basis by neglecting modes corresponding to the small eigenvalues in order to capture most of the energy of the snapshot basis. Here for our case, if the POD is constructed for 5 and a reduced order model with 3 it yields a ratio of about 0.98; and if is constructed with 15 it yields a ratio of above 0.99 for the percentage of kinetic energy retained (Fig. 5 and 6).

We are now returning to the barotropic tidal and wind stress model for a coastal lagoon in the Mexican Caribbean to apply the POD technique. Therefore, we solve Eqs. (4) - (6) after 5 tidal cycles of the $M_2$ harmonic. Results using classical model are depicted graphically in Figure 4. The results of the model using POD are graphically almost the same and will not be shown.

To quantify the performance of the reduced basis method, we use two metrics namely the root mean square error (RMSE) and correlation of the difference between the full order and the reduced order simulation. This is obtained by first taking the five tidal cycles full order results and the corresponding five tidal cycles reduced order results and computing the error, for example, for the variable $u$ and $v$ components of the velocity vector $(\vec{v})$; the errors are shown in Figs. 6 and 7. Here, if $n = 10$ basis function, the first four PODs modes (Fig. 6), capture nearly 100%, while for $n = 15$ basis function, the first seven POD (Fig. 6) capture nearly 100% with an error ranging from $10^{-4}$ to $10^{-7}$. Modes capture about 99% of the energy. Thus, different POD modes may be used to reconstruct fields respectively. For different numbers of snapshots but for the same energy percentage captured, the RMSE decrease stops at 15 snapshots. The correlation taking the five tidal cycles full order results and the corresponding five tidal cycles reduced order clearly, when increasing the POD mode, the correlation increases also for the same snapshots. This increase stops at 5 snapshots and the reported best approximation obtained with 15 snapshots produced a correlation at the same level as the approximation 20 snapshots.

However, one must also note that a simple linear independence is not a sufficient criterion for choosing the POD mode. It only provides one with some reference. The error between the full order and the reduced order is displayed in Fig. 7 for a retained energy percentage of 99%. There is a little improvement between either 10 snapshots or 15 snapshots and 5 snapshots, but there is almost no difference between 15 snapshots and 30 snapshots.

Order approximation may be sufficiently close to the full order approximation. Other experiments have also been carried out, with either more or fewer snapshots taken and for different percentages of energy captured, not shown here. From the computational cost and memory storage aspects, 15 snapshots and the energy captured at 99% level yielded the best results.

![Figure 5](image1.png)  
**Figure 5.** Orthogonal base and order evolution of the approximations.

![Figure 6](image2.png)  
**Figure 6.** Computed error with $n = 15$ basis functions. a) Absolute value of the currents, continuous line $u$, and dotted line $v$ components of the velocity vector $\vec{v}$, and b) absolute error of the current compared with the classical barotropic ocean circulation numerical model results.

![Figure 7](image3.png)  
**Figure 7.** Computed absolute error in $|\vec{v}|$ with $n = 15$ base functions and $K = 60$ at time 140 for all positions in the numerical spatial grid.
6. Conclusions

We studied problems related to POD reduced modelling of a coastal ocean circulation model in the Mexican Caribbean area. The large-scale variability of the wind stress and $M_2$ tidal component is first simulated using a barotropic vertically integrated numerical model with spatial resolution of $\Delta x = \Delta y = 20\,\text{m}$ and a time step of $\Delta t = 745.2\,\text{s}$. Then we constructed different POD models with different choices of snapshots and different numbers of POD basis functions. The results from these different POD models are compared with those of the original model. The main conclusions are: 1) the large-scale variability of the wind stress and $M_2$ tidal component obtained by the original model can be captured well by a low-dimensional system of order 22 that is constructed by 15 snapshots and 7 leading POD basis functions; 2) by analysis of RMS errors and correlations, we found that the modes that capture 99% of the energy are necessary to construct POD models; 3) RMS errors for the velocity components of the POD model of order 22 is less than $10^{-4}$ order compared with the original model that is less than 1% of $(u,v)$ in the original model; correlations of the original model from the POD model are around 0.99; and 4) compared with the original model, the velocity fields from the POD model are less accurate than the free surface oscillation results (not shown because the agreement was more than 99% between the original model and the POD). This remains a problem to be further explored in forthcoming research. Our preliminary investigations on the use of POD tidal and wind stress ocean circulation simulation yield encouraging results and show that POD can be a powerful tool for various applications such as four-dimensional variational data assimilation. These results will be described in a follow-up paper.