A POD Reduced order 4D-VAR adaptive mesh ocean modelling approach

F. Fang [a,*], C.C. Pain [a], I.M. Navon [b], M.D. Piggott [a],
G.J. Gorman [a], P. Allison [a], A.J.H. Goddard [a]

[a] *Applied Modelling and Computation Group,*
Department of Earth Science and Engineering,
Imperial College London,
Prince Consort Road, London, SW7 2BP, UK.
URL: `http://amcg.ese.imperial.ac.uk`

[b] *Department of Mathematics*
Florida State University
Tallahassee FL, 32306-4120. USA.

Abstract

A novel Proper Orthogonal Decomposition (POD) inverse model developed for a mesh adaptive ocean model, the Imperial College Ocean Model(ICOM) is herein presented. The new POD model is validated by application to a gyre flow test case where it invertis for initial conditions. (**** Say something quantitative about how well the model performed here, one short sentence will do). Error estimation (including the error in POD reduced modelling and the interpolation error in adaptive meshes) was used to show that quality of reduced adaptive mesh models was (**** say something quantitative about the results of this estimation).

New criteria are applied for POD adaptivity in this approach. When the optimization loses control (i.e., when in the process of optimisation of the new control variables depart significantly from the ones on which the POD reduced model is based upon) the set of POD snapshots are refreshed. when the optimization loses its control, i.e., when in the process of optimisation the new control variables depart significantly from the ones on which the POD reduced model is based upon. Previously developed numerical approaches (Fang et al. (2007)) are employed to accurately represent the geostrophic balance and improve the efficiency of the POD simulation;

The feasibility of the proposed methodology as well as and it's ability to assessing the error in POD results are demonstrated in a range of numerical experiments including:-

keywords: inverse, adjoint, POD; reduced-order modelling; ocean model; finite element; unstructured adaptive mesh

1 Introduction

Given current concerns on the nature of climate change it is clearly crucially important to improve the predictive capabilities of ocean models. Data assimilation techniques are an important means of meeting this aim. By assimilating observations (such as in-situ measurements and remote sensing) into models, unknown inputs such as initial and boundary conditions, bottom friction coefficients, turbulent viscosity, and wind stress etc. can be optimised [1, 2, 3, 4, 5, 6, 7, 8]

A variety of approaches have in the past been used to facilitate assimilation and include statistical interpolation methods, nudging data assimilation, variational (adjoint) methods and the Kalamn filter. In particular, four-dimensional variational (adjoint) method has proved an effi-

---
* Corresponding author
  Email address: `f.fang@imperial.ac.uk` (F. Fang).

cient method for assimilation observed data into simulations [9, 10]. It is capable of producing a best estimate model solution by fitting a numerical simulation to observational data over both space and time. The technique is also able to estimate the error sources caused by uncertainties (boundary conditions, initial condition and parameters) in the model . The solution is derived by minimising a cost function that contains the misfits between the data and dynamical model, as well as the covariances specifying spatial and temporal correlations of errors. The 4D-Var data assimilation has been used widely in both atmospheric and oceanographic models over the past two decades [11, 12, 9, 13, 14, 15, 16, 17, 7, 8, 14, 18, 4, 6, 19, 20]. However, the major difficulty in the implementation of 4D-Var data assimilation is the large dimensionality of the control space (for a discrete realistic model, the size of the control variables is typically in the range $10^6 - 10^8$), and hence demands high memory and computational costs.

Computational cost can however, be minimised by decreasing the dimensions of the control space so as to ensure that the minimisation of the cost function (or error co-variances) is carried out within a low-dimensional space. This can be achieved through an incremental approach (Courtier et al., 1994 [21]) whereby a succession of quadratic problems is generated over increasing time periods. The successive quadratic minimisation problems can then be solved using a coarse resolution in the in-loop of the minimisation. The dimension of the minimisation problem can then be decreased by one or two orders of magnitude. Although such an incremental approach is currently used in operational atmospheric models (Rabier et.al. 2000 [22]), the dimensions of the control space remain quite large in realistic applications.

The use of Empirical Orthogonal Functions (EOF) analysis has been advocated as an approach that can lead to reduced order ocean modelling (Blayo et al.,1998; Durbiano, 2001 [23]). The implementation of this method additionally results in a drastic reduction of the dimension of the control space and thus the minimisation process [24, 25]. Reduced order 4D-Var can also be used to precondition 4D-Var, and reduce computational cost [26]. It has been further proposed (Vermeulen and Heemink,2006 [27]) that efficiencies can be enhanced if the adjoint can be directly implemented in a subspace of the reduced model (determined by the leading EOFs) and then used to approximate the gradient of the criterion. The minimisation process can thus be solved completely in reduced space with negligible computational costs.

Proper Orthogonal Decomposition (POD) methodologies, in combination with the Galerkin projection procedure have also been shown to provide an efficient means of generating reduced order models ([28]; [29], [30]). This technique essentially identifies the most energetic modes in a time-dependent system thus providing a means of obtaining a low-dimensional description of the system's dynamics. POD has been widely and successfully applied to numerous fields, including signal analysis and pattern recognition ([31]), fluid dynamics and coherent structures (Lumley,1967; [32]; [28]; [33]) and image reconstruction ([34]). To improve the accuracy of reduced models, a goal-oriented approach has been used to optimise the POD bases ([35], [36]). The dual-weighted POD approach provides an 'enriched' set of basis functions combining information from both model dynamics and the data assimilation system. The practical utility of this approach has been extended to include ocean and climate modelling and the solution

of inverse problems ([24], [25], [37] and [29]). The POD-based 4D VAR not only reduces the dimension of control space, but also reduces the size of dynamical model, both in dramatic ways ([37, 38]).

Our motivation is to produce a novel POD reduced order 4D-Var for an adaptive ocean model. A POD-based reduced forward model has been developed for ICOM [39] that can simultaneously resolve both small and large scale ocean flows whilst smoothly varying mesh resolution and conforming to complex coastlines and bathymetry. In this work, a further step has been made to introduce the POD approach into an adaptive mesh refinement adjoint model. Using the POD and Galerkin projection approaches, the reduced forward model is derived in a subspace (details in [39]). Once the forward reduced model is available, the reduced order adjoint model can directly obtained from the POD reduced forward model in the subspace, instead of the original forward model. The minimisation procedure is then carried out in the reduced space.

When adaptive meshes are employed in both the forward and adjoint models, the mesh resolution requirements for each model may be spatially and temporally different, as the meshes are adapted according to the flow features of each model. This poses additional challenges for the implementation of an inverse POD-based reduced adaptive model. Such challenges include (a) snapshots can be of different length at different time levels; (b) the POD base of the forward model can differ from the POD base of the adjoint model. To overcome these difficulties, a standard reference fixed mesh is adopted for both the forward and adjoint reduced models. The solutions for both are interpolated from their own mesh onto the same reference fixed mesh at each time level. This allows the same number of base modes for both the reduced forward and adjoint models. Particular attention is paid to the accuracy of the interpolation used this purpose.

In this work, an adaptive POD procedure is employed to improve the reduced model by updating the POD basis. The original reduced basis for inverse problems is calculated by a set of snapshots based on the results from the full forward model with the specified control variables. The optimised control variables via optimisation techniques could be significantly different from the original ones. An updated reduced basis is therefore needed when the optimisation procedure loses its control based on an error criterion defined in a cost function (or Trust-Region POD approach). Ravindran (2002, 2006) [40, 41] proposes an adaptive procedure that successively updates the reduced-order model to be used via a sequential SQP (sequential quadratic programming) constrained optimisation algorithm. Cao et al. (2007) introduced an adaptive POD approach into POD inverse (adjoint) models. The updated POD basis is recalculated using a refreshed set of snapshots based on the latest results obtained from the full forward model driven via the optimised control variables.

The remainder of this paper is structured as follows: In the second section the Proper Orthogonal Decomposition (POD) is described. In the third section the discrete reduced forward and adjoint models are then derived. The mesh adaptivity technique is discussed in detail in section four. In section five the above reduced model is applied to and illustrated by some relevant cases.

Summary and conclusions are drawn in the final section.

2   POD reduced model

The Proper Orthogonal Decomposition (POD) is the most efficient choice among linear decompositions in the sense that it can capture the greatest possible kinetic energy. A POD reduced model for a 3D flow model is described in this section. A 3D flow dynamical model is generally written as:

$$\frac{\partial \mathbf{u}}{\partial t} = f(\mathbf{u}, t, \mathbf{x}), \tag{1}$$

where, $\mathbf{u}$ is a vector containing all variables to be solved (e.g., velocities u, v, w, pressure p and temperature T etc.), t is time and $\mathbf{x} = (x, y, z)^T$ represents the Cartesian coordinate position.

2.1   Proper Orthogonal Decomposition

The model variables $\mathbf{u}$ are sampled at defined checkpoints during the simulation period $[t_1, \ldots, t_K]$, also referred to as snapshots $U = (U_1, \ldots, U_K)^T$ (K being the number of snapshots). The snapshots can be obtained either from a mathematical (numerical) model of the phenomenon or from experiments/observations. The sampled values of variables at the snapshot k are stored in a vector U_k with $\mathcal{N}$ entries ($\mathcal{N}$ being the number of nodes), here, U can represent one of variables u, v, w, p. The average of the ensemble of snapshots is defined as:

$$\bar{U}_i = \frac{1}{K} \sum_{k=1}^{K} U_{k,i}, \quad 1 \leq i \leq \mathcal{N}, \tag{2}$$

Taking the deviation from the mean of variables, forms

$$V_{k,i} = U_{k,i} - \bar{U}_i, \quad 1 \leq i \leq \mathcal{N}. \tag{3}$$

A collection of all $V_{k,i}$ constructs a rectangular $\mathcal{N}$ by K matrix A. The goal of Proper Orthogonal Decomposition (POD) is to find a set of orthogonal basis functions which can represent the most dynamic energy in the original flow system. The $N \times N$ eigenvalue problem is established

$$AA^T x_k = \lambda_k x_k; \quad 1 \leq k \leq K. \tag{4}$$

The order $\mathcal{N}$ for matrix AA^T is far larger than the order K for matrix $A^T A$ in realistic ocean cases. Therefore the $K \times K$ eigenvalue problem is solved

$$A^T A y_k = \lambda_k y_k; \quad 1 \leq k \leq K. \tag{5}$$

This procedure is equivalent to a Singular Value Decomposition (SVD). The eigenvalues λ_k are real and positive and should be sorted in an descending order. The POD basis vectors Φ_k associated with the eigenvalues λ_k are orthogonal and expressed as follows:

$$\Phi_k = Ay_k/\sigma_k = Ay_k/\sqrt{\lambda_k}, \tag{6}$$

where, the kth eigenvalue is a measure of the kinetic energy transferred within the kth basis mode. If the POD spectrum (energy) decays fast enough, practically all the support of the invariant measure is contained in a compact set. Roughly speaking, all the likely realisations in the ensemble can be found in a relatively small set of bounded extent. By neglecting modes corresponding to the small eigenvalues, the following formula is therefore defined to choose a low-dimensional basis of size M ($M << K$),

$$I(M) = \frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{K} \lambda_i}, \tag{7}$$

subject to

$$M = argmin\{I(M) : I(M) \geq \gamma\}, \tag{8}$$

where, $0 \leq \gamma \leq 1$ is the percentage of energy which is captured by the POD basis $\Phi_1, \ldots, \Phi_m, \ldots, \Phi_M$.

2.2 Reduced model

The variables in (1) can be expressed in an expansion of the POD basis functions $\{\Phi_1, \ldots, \Phi_M\}$, that is,

$$\mathbf{u}(t, x, y, z) = \bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_m(t)\Phi_m(\mathbf{x}), \tag{9}$$

where, $\bar{\mathbf{u}}$ is the mean of the ensemble of snapshots for the variables $\mathbf{u}(t)$, α_m ($1 \leq m \leq M$) are the time-dependent coefficients to be determined, and $\alpha_m(0)$ are the coefficients at the initial time level. Substituting (9) into (1) and taking the POD basis function as the test function, then integrating over the computational domain Ω,

$$\int_{\Omega} \Phi_m f\left((\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_m \Phi_m(\mathbf{x})), t, \mathbf{x}\right) d\Omega, \tag{10}$$

The POD reduced model is then obtained:

$$\frac{\partial \alpha_m}{\partial t} = \left\langle f\left((\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_m(t)\Phi_m(\mathbf{x})), t, \mathbf{x}\right), \Phi_m \right\rangle, \tag{11}$$

subject to the initial condition

$$\alpha_m(0) = ((\mathbf{u}(0, \mathbf{x}) - \bar{\mathbf{u}}(\mathbf{x})), \Phi_m). \tag{12}$$

In the finite element method, the POD basis $\Phi_m(\mathbf{x}) = \sum_{i=1}^{\mathcal{N}} N_i \Phi_{m,i}$, the equation (11) can therefore be written as:

$$\frac{\partial \alpha_m}{\partial t} = \left\langle f\left((\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_m(t) \sum_{i=1}^{\mathcal{N}} N_i \Phi_{m,i}), t, \mathbf{x}\right), \Phi_m \right\rangle. \tag{13}$$

where, N_i is the basis function in the finite element, $\mathcal{N}$ is the number of nodes in the computational domain.

3 Reduced order 4D-Var

The aim of 4D-Var is to determine optimal control variables (e.g., initial conditions, U_0), hence optimal solution in equation (1) by minimising the functional defined below:

$$\Im(U^0) = \frac{1}{2}(U^0 - U_b)^T \mathbf{B}^{-1}(U^0 - U_b) + \frac{1}{2}\sum_{n=1}^{NT}(\mathbf{H}U^n - y_o^n)^T W_o(\mathbf{H}U^n - y_o^n), \tag{14}$$

where, $\mathbf{B}$ is the background error covariance matrix, W_o is the observation error covariance matrix, $\mathbf{H}$ is the observation operator, U^0 is a vector containing the control variables (here, initial conditions), U^n is a vector containing the solution of variables from the model (the reduced order model) at the time level n (here, NT is the number of time levels), and y_o is the observation vector. In a POD reduced mode, the initial value U^0 and the reduced order solution U^n are expressed as:

$$U^0 = \bar{U} + \sum_{m=1}^{M} \alpha_m(0)\Phi_m(\mathbf{x}), \tag{15}$$

$$U^n = \bar{U} + \sum_{m=1}^{M} \alpha_m(t^n)\Phi_m(\mathbf{x}), \tag{16}$$

where, $\mathbf{x} = (x, y, z)^T$, $\bar{U}$ is the average of the ensemble of snapshots $\bar{U} = \frac{1}{K}\sum_{k=1}^{K} U_k$ (here, K is the number of snapshots). Equation (14) can be rewritten as:

$$\Im(\alpha(0)) = \frac{1}{2}\left(\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(0)\Phi_m(\mathbf{x})\right) - U_b\right)^T \mathbf{B}^{-1}\left(\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(0)\Phi_m(\mathbf{x})\right) - U_b\right)$$

$$+ \frac{1}{2}\sum_{n=1}^{NT}\left(\mathbf{H}\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(t^n)\Phi_m(\mathbf{x}),\right) - y_o^n\right)^T W^o\left(\mathbf{H}\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(t^n)\Phi_m(\mathbf{x}),\right) - y_o^n\right). \tag{17}$$

3.1 Discrete reduced order adjoint equations

The discrete forward model of equation (11) at the time level n can be written in a subpace:

$$\mathbf{A}^n \alpha^n = s^n, \tag{18}$$

where,

$$s^n = \mathbf{B}^n \alpha^{n-1} + f, \tag{19}$$

where, $\mathbf{A}^n$ and $\mathbf{B}^n$ ($\mathbf{A}^n, \mathbf{B}^n \in R^{M \times M}$, where, M is the number of POD bases) are the matrices at the time level n which include all the discretization of equation (11), $\alpha^n = (\alpha_1^n, \dots, \alpha_M^n)$ and $\alpha^{n-1} = (\alpha_1^{n-1}, \dots, \alpha_M^{n-1})$ are the vectors of variables to be solved at the time levels n and $n-1$ respectively, here including the coefficients related to the POD basis functions for state variables $\mathbf{u}$ in equation (1), n is the time index, s^n is a discretised source term at the time level n and f is a source term.

For a nonlinear simulation, the matrices $\mathbf{A}^n$ and $\mathbf{B}^n$ can be written as:

$$\mathbf{A}^n = \hat{\mathbf{A}}_0^n + \sum_{m=1}^M \alpha_m^{n-1} \hat{\mathbf{A}}_m^n, \tag{20}$$

$$\mathbf{B}^n = \hat{\mathbf{B}}_0^n + \sum_{m=1}^M \alpha_m^{n-1} \hat{\mathbf{B}}_m^n, \tag{21}$$

where, $\mathbf{A}^n, \mathbf{B}^n, \hat{\mathbf{A}}_0^n, \hat{\mathbf{B}}_0^n, \hat{\mathbf{A}}_m^n, \hat{\mathbf{B}}_m^n \in R^{M \times M}$ and the components can be expressed as:

$$\hat{\mathbf{A}}_{0,i,j}^n = \int_\Omega \Phi_i (\bar{\mathbf{u}} \, \Phi_j + \theta \bar{\mathbf{u}} \, \Phi_j) \, d\Omega, \quad 1 \leq i, j \leq M, \tag{22}$$

$$\hat{\mathbf{A}}_{m,i,j}^n = \int_\Omega \Phi_i \left(\Phi_j \, \Phi_m + \theta \Phi_j \, \Phi_m \right) d\Omega, \quad 1 \leq i, j \leq M, \tag{23}$$

$$\hat{\mathbf{B}}_{0,i,j}^n = \int_\Omega \Phi_i (\bar{\mathbf{u}} \, \Phi_j + (1-\theta) \bar{\mathbf{u}} \, \Phi_j) \, d\Omega, \quad 1 \leq i, j \leq M, \tag{24}$$

$$\hat{\mathbf{B}}_{m,i,j}^n = \int_\Omega \Phi_i \left(\Phi_j \, \Phi_m + (1-\theta) \Phi_j \, \Phi_m \right) d\Omega, \quad 1 \leq, i, j \leq M. \tag{25}$$

Equations (20) and (21) can be rewritten as:

$$\mathbf{A}^n = \hat{\mathbf{A}}_0^n + \hat{\mathbf{A}}^n \alpha^{n-1}, \tag{26}$$

$$\mathbf{B}^n = \hat{\mathbf{B}}_0^n + \hat{\mathbf{B}}^n \alpha^{n-1}, \tag{27}$$

where, $\hat{\mathbf{A}}^n = (\hat{\mathbf{A}}_1^n, \dots, \hat{\mathbf{A}}_M^n)$ and $\hat{\mathbf{B}} = (\hat{\mathbf{B}}_1^n, \dots, \hat{\mathbf{B}}_M^n)$.

Taking into account equation (18), the discrete forward equation during the simulation period $[t_1, \dots, t_{NT}]$ (where, NT is the number of time levels) can be written:

$$\mathbf{A}\alpha = s \tag{28}$$

where,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^{NT} & -\mathbf{B}^{NT} & & \\ & \mathbf{A}^{NT-1} & -\mathbf{B}^{NT-1} & \\ & & \ddots & \ddots & \\ & & & \mathbf{A}^1 \end{pmatrix},$$

$$\alpha = \begin{pmatrix} \alpha^{NT} \\ \alpha^{NT-1} \\ \vdots \\ \alpha^1 \end{pmatrix},$$

and

$$s = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \mathbf{B}^1 \alpha^0 \end{pmatrix} + f.$$

Differentiating equation (28) with respect to the control variables to be optimised, i.e., the initial coefficient, $\alpha^0 = \alpha(0)$, the tangent linear model is obtained

$$\bar{\mathbf{A}}\alpha + \mathbf{A}\bar{\alpha} = \bar{s}, \tag{29}$$

where, the bar is defined as the differentiation with respect to the control variables α^0

$$\bar{\mathbf{A}} = \frac{\partial \mathbf{A}}{\partial \alpha^0} = \begin{pmatrix} \bar{\mathbf{A}}^{NT} & -\bar{\mathbf{B}}^{NT} & & \\ & \bar{\mathbf{A}}^{NT-1} & -\bar{\mathbf{B}}^{NT-1} & \\ & & \ddots & \ddots & \\ & & & \bar{\mathbf{A}}^1 \end{pmatrix},$$

and

$$\bar{\alpha} = \frac{\partial \alpha}{\partial \alpha^0} = \begin{pmatrix} \bar{\alpha}^N \\ \bar{\alpha}^{N-1} \\ \vdots \\ \bar{\alpha}^1 \end{pmatrix},$$

$$\bar{s} = \frac{\partial s}{\partial \alpha^0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \bar{\mathbf{B}}^1 \alpha^0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \mathbf{B}^1 \end{pmatrix},$$

where, taking into account equations (26) and (27)

$$\bar{\mathbf{A}}^n = \frac{\partial \mathbf{A}^n}{\partial \alpha^0} = \hat{\mathbf{A}} \bar{\alpha}^{n-1}, \tag{30}$$

$$\bar{\mathbf{B}}^n = \frac{\partial \mathbf{B}^n}{\partial \alpha^0} = \hat{\mathbf{B}} \bar{\alpha}^{n-1}, \tag{31}$$

where, n donates time levels, $1 \leq n \leq NT$. Equation (29) can therefore rewritten

$$\begin{pmatrix} \hat{\mathbf{A}}\bar{\alpha}^{NT-1} & -\hat{\mathbf{B}}\bar{\alpha}^{NT-1} & & \\ & \hat{\mathbf{A}}\bar{\alpha}^{NT-2} & -\hat{\mathbf{B}}\bar{\alpha}^{NT-2} & \\ & & \ddots & \ddots \\ & & & \hat{\mathbf{A}}\bar{\alpha}^0 \end{pmatrix} \begin{pmatrix} \alpha^{NT} \\ \alpha^{NT-1} \\ \vdots \\ \alpha^1 \end{pmatrix} + \begin{pmatrix} \mathbf{A}^{NT} & -\mathbf{B}^{NT} & & \\ & \mathbf{A}^{NT-1} & -\mathbf{B}^{NT-1} & \\ & & \ddots & \ddots \\ & & & \mathbf{A}^1 \end{pmatrix} \begin{pmatrix} \bar{\alpha}^{NT} \\ \bar{\alpha}^{NT-1} \\ \vdots \\ \bar{\alpha}^1 \end{pmatrix} =$$

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ \hat{\mathbf{B}}^1 \alpha^0 \bar{\alpha}^0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \mathbf{B}^1 \bar{\alpha}^0 \end{pmatrix}$$

The tangent linear model is then derived

$$(\mathbf{A} + \mathbf{A}_{extra})\frac{\partial \alpha}{\partial \alpha^0} = \frac{\partial s}{\partial \alpha^0} \tag{32}$$

where, $\mathbf{A}$ is calculated in [28], $\mathbf{A}_{extra}$ originates from the actual linearision of the nonlinear

10

terms and is expressed as:

$$\mathbf{A}_{extra} = \begin{pmatrix} 0 & \hat{\mathbf{A}}\bar{\alpha}^{NT} - \hat{\mathbf{B}}\bar{\alpha}^{NT-1} & & & \\ & 0 & \hat{\mathbf{A}}\bar{\alpha}^{NT-1} - \hat{\mathbf{B}}\bar{\alpha}^{NT-2} & & \\ & & \ddots & & \ddots \\ & & & 0 & \end{pmatrix}$$

The variation of the objective function (14) with respect to the control variables α^0 is

$$\frac{\partial \Im}{\partial \alpha^0} = \left(\frac{\partial \alpha}{\partial \alpha^0}\right)^T \frac{\partial \Im}{\partial \alpha} \tag{33}$$

Taking into account equation (32), yields

$$\frac{\partial \Im}{\partial \alpha^0} = \left(\frac{\partial s}{\partial \alpha^0}\right)^T (\mathbf{A} + \mathbf{A}_{extra})^{-T} \frac{\partial \Im}{\partial \alpha} \tag{34}$$

The gradient of the objective function can then be written

$$\frac{\partial \Im}{\partial \alpha^0} = \left(\frac{\partial s}{\partial \alpha^0}\right)^T \alpha^* \tag{35}$$

here, α^* is the adjoint variable and can be calculated by solving the following adjoint equation:

$$(\mathbf{A} + \mathbf{A}_{extra})^T \alpha^* = \frac{\partial \Im}{\partial \alpha} \tag{36}$$

3.2   Adaptive POD

In this work, since the POD model is based on the solution of the original model for a specified control variables (e.g., initial and boundary conditions, etc), it could be a poor model when the updated control variables are significantly different from the one on which the POD model is based upon. An adaptive POD 4DVAR procedure is proposed to successively update the the POD basis, therefore the reduced direct and inverse models (details in Ravindran (2002, 2006) [40, 41]. The inversion procedure starts with an initial estimated control variables. An initial set of snapshots are obtained by running the full forward model, and the corresponding POD subspace and reduced model are constructed. The process of updating the reduced-order model is continued during each inversion iteration until convergence is achieved.

4 Mesh adaptivity in reduced models

4.1 Description of an adaptive ocean ICOM

In this work, a POD-based reduced model is developed for ICOM that can simultaneously resolve both small and large scale ocean flows whilst smoothly varying resolution and conforming to complex coastlines and bathymetry . With more appropriate focused numerical resolution (e.g. adaptive and anisotropic resolution of fronts and boundary layers, and optimal representation of vertical structures in the ocean) ocean dynamics may be accurately predicted during future climatic change. To accurately represent local flow around steep topography the hydrostatic assumption is not made in this work. Here, the pressure is split into the non-geostrophic and geostrophci parts which are solved separately. This allows the accurate representation of hydrostatic/geostrophic balance [42]. A dynamically adapting anisotropic mesh in 3-D is used here [43, 44] which optimises the size and shape of a tetrahedral mesh using edge collapsing and splitting, edge swapping, face to edge swapping, and smoothing in a fashion similar to Freitag and Ollivier- Gooch [?] and Buscaglia and Dari [6?]. Mesh adaptivity or optimisation relies on the derivation of appropriate error measures, which dictate how the mesh is to be modified [43]

4.2 Mesh adaptive technique in the ICOM ocean model

The mesh adaptivity technique used in this work first constructs a metric tensor which encodes local error estimates using information from the current solution fields, see [45] for an alternative approach. The metric tensor is used to calculate the edge lengths of mesh elements. It is constructed (see below) such that an ideal edge length is unity when measured in metric space. This therefore gives a guide as to those elements which have edges that are too large or too small. Since the metric is dependent on both location and direction it is able to exhibit locally anisotropic information if the solution fields dictate this. Thus inhomogeneous and anisotropic meshes result from this approach. By defining an objective functional which is formed from the size and shape of the 'worst' element in the mesh, an optimisation technique is used to improve the overall quality of the mesh. That is, local operations on the mesh connectivity and node positioning are performed which aim to minimise this functional. The operations are performed on a three-dimensional tetrahedral mesh by the adaptivity method and include: edge collapsing/splitting; face to edge and edge to face swapping; edge to edge swapping; and local node movement or mesh smoothing. Constraints are imposed on these operations so as to preserve the integrity of non-planar geometrical boundaries. For further details see [43, 44]. The interpolation error resulting from a piecewise linear approximation of a smooth function ψ may be written in terms of the functions second derivatives or Hessian $H \equiv \nabla^T \nabla \phi$. In particular,

over a tetrahedral element the quantity

$$\epsilon = \mathbf{v}^T |H(\mathbf{x})| \mathbf{v}, \tag{37}$$

gives a guide to this interpolation error at location $\mathbf{x}$ and direction $\mathbf{v}$. Note that here the notation $|\cdot|$ corresponds to the absolute value of the matrix, defined by diagonalising the matrix and taking the absolute value of the eigenvalues, rather than the determinant. Taking the maximum of (37) over all locations and all vectors $\mathbf{v}$ that lie within the element gives a bound on the magnitude of the interpolation error on this element. For more details see [44] and the references therein.

Given a user defined error tolerance, $\hat{\epsilon}$, a metric tensor may be defined by

$$\hat{M} = \frac{1}{\hat{\epsilon}} |H|. \tag{38}$$

To obtain this error tolerance everywhere, length scales in the mesh should therefore take the value unity when measured in metric space. The mesh optimisation technique described above is used to achieve this aim as closely as possible. In practise the Hessian must be approximated and here reconstructed from the piecewise linear solution fields. To simultaneously satisfy separate error bounds for separate solution fields a metric is obtained for each and these are then superimposed to yield a combined error metric. Finally, by suitably altering the definition of this metric, computational constraints on the total number of nodes and maximum/minimum allowed element sizes and aspect ratios may be imposed. For further details see [43].

4.3   Mesh adaptive technique in POD and error estimate

When adaptive meshes are employed in ocean models, the mesh resolution requirements vary spatially and temporally, as the meshes are adapted according to the flow features through the whole simulation. The dimensional size of the variable vectors is different at each time level since the number of nodes varies during the simulation. Snapshots can therefore be of different length at different time levels. This unavoidably brings difficulties in the implementation of a POD-based reduced model for an adaptive model. To overcome these difficulties, a standard reference fixed mesh is adopted for the reduced model. The solutions from the original full model are interpolated from their own mesh onto the same reference fixed mesh at each time level, and then stored in the snapshots. The information at the snapshots is used to find the optimal POD basis. This allows the same length of base modes to be obtained at each time level. The resolution of the reference mesh and the interpolation errors between the two meshes (the adaptive mesh and the fixed reference mesh) may affect the accuracy of the POD simulation. This will be explained and discussed in detail through the applications presented below. To reduce the interpolation error, a high order interpolation approach can be adopted. The feasibility and accuracy of 4D-var simulation can be evaluated by the comparison of the results obtained through using the adaptive and static meshes in the optimal procedure.

5 Application and discussion

The above reduced order inverse model developed for an adaptive mesh finite element model (ICOM) has been applied to 2D gyre flows. The initial conditions are optimised using the POD reduced adjoint model. The accuracy and validation of the reduced order POD adjoint model have been evaluated. Error estimate is undertaken through the comparison of the results obtained from the original(full) and POD reduced models. Furthermore, the adaptive POD approach is employed to update the POD bases and its efficiency is investigated.

5.1 Description of the case: Gyre

The POD reduced adjoint model is tested in a computational domain, $1000km$ by $1000km$ with a depth of $500\ m$. The wind forcing on the free surface is given

$$U_{wind} = \tau_0 cos(\pi y/L_y), V_{wind} = 0.0, \tag{39}$$

where, $\tau_0 = 0.1Nm^{-1}$, here, $L_y = 1000km$. The Reynolds number is $Re = 250$, $\beta = 1.8 \times 10^{-11}$, and the reference density $\rho_0 = 1000\ kgm^{-1}$.

No-slip boundary conditions are applied to the lateral boundaries. The spin-up period is $200\ days$. The simulation period is $[200, 400]\ days$. The time step is $6\ hrs$. The non-dimensionalization is applied by using the length scale $L_y = 1000\ km$.

5.2 Optimisation of initial conditions

The control variables in the case are the initial conditions on the 200th day. The data assimilation experiments are designed by an identical twin technique. The 'true' flow state is generated by running the full forward model (ICOM) with a zonal wind forcing on the free surface. The pseudo-observational data is taken on days 250, 300 and 350 over the computational domain. The guess values of the initial conditions are given by the background flow (here, taken from the 'true' flow fields on days either 215 or 275 in the experiments).

The POD bases are constructed by the snapshots which are taken from the numerical solutions by forcing the full forward model with the background flow. 40 snapshots with 35 POD bases for the velocity components u, v, w and pressure are chosen to capture more than 99.5% of energy.

To represent the geostrophic balance accurately, the geostrophic pressure is constructed by two sets of basis functions which are obtained from the basis functions for the velocity components u and v respectively. The numerical technique developed in [39] is adopted to accelerate the POD inversion, that is, the matrices in the discretised POD forward and adjoint equations can be constructed by a set of time-independent sub-matrices prior to running the reduced forward and adjoint models. This can reduce the computer cost significantly.

An adaptive mesh is adopted in the full model. The mesh for the full model adapts every 19 time steps with maximum and minimum mesh size of 0.2 and 0.001 (non-dimensional) respectively. To allow the same length of POD bases at the snapshots for both the reduced forward and adjoint models, a reference fixed mesh is chosen for the POD inversion (right panel in figure **??**). To build up the snapshots, the solutions from the full forward model are interpolated from the adaptive mesh (left panel in figure **??**) onto the reference mesh.

5.3   Discussion

The utility and accuracy of the POD reduced inverse (forward and adjoint) models are evaluated through error analysis. The error for the inversely modelled results includes: (1) the error of the POD reduced model; (2) the interpolation error if adaptive meshes are adopted; (3) the error for the optimised controls.

To estimate the accuracy of the POD forward model, the error for the controls is eliminated. The POD solution is obtained by running the POD reduced model with the true controls. The error (misfit between the POD solution and true flow state) at time levels ($t = 250, 300, 350 \; day$) is shown in figure **??**. The error is very small during the first half simulation period and increases as the simulated time accrues. Since adaptive meshes are employed in the full model, the error shown in figure **??** includes the interpolation error as well.

As discussed in section 3.2, an adaptive POD approach can be used to update the POD bases and reduced models. It is however found in this study that the updated POD bases could be worse than the previous ones if the optimised control variables (here, the initial conditions) are far from the true ones. As an example, the initial guess controls (background flow) is taken from the true flow field on the 275th day. By fitting the numerical solution to observationbnal data, the errors of inversely modelled velocity field at the time levels (when the observational data is assimilated) are reduced to small values (figure **??**). The cost function (14) is reduced by 73 % of its original value during the optimisation procedure. However, the optimised initial controls during this POD inversion iteration are not what we expect (top panel in figure **??**). A unexpected local minimum is reached in this casce. This inverse problem is ill-posed. Obviously the optimised initial conditions cannot be used to update the POD bases.

An appropriate of initial guess controls is needed to improve the inverse solution. The initial controls (background flows) are taken from the true flow state on the 225th day, The cost function is reduced by 50 % of its original value at the first POD inversion iteration (here, the error of the POD reduced model is taken out from the cost function). It is further reduced by 25 % at the second POD iteration, i.e., after the snapshots are updated. The error for the velocity field is shown in figure **??**. it can be seen that the error of inversely modelled velocity is reduced after updating the snapshots (right panel). The optimised velocity fields are drawn in figure **??** and exhibit an overall good agreement with the true ones.

[Fig. 1 about here.]

[Fig. 2 about here.]

[Fig. 3 about here.]

[Fig. 4 about here.]

[Fig. 5 about here.]

Acknowledgements

References

[1] R. W. Lardner. Optimal control of open boundary conditions for a numerical tidal model. *Computer Methods in Applied Mechanics and Engineering*, 102:367–87, 1993.

[2] A. F. Bennett, Chua B. S., D. E. Harrison, and M. J. McPhaden. Generalized inversion of Tropical Atmosphere-Ocean (TAO) data and a coupled model of the tropical Pacific. Part II: The 1995-96 La Nina and 1997-98 El Nino. *J. Climate*, 13(15):2770–2785, 2000.

[3] Y. Leredde, J. L. Devenon, and I. Dekeyser. Turbulent viscosty optimized by data assimilation. *Ann. Geophysicae*, 17:1463–1477, 1999.

[4] A. M. Moore. Data assimilation in a quasi-geostrophic open-ocean model of the Gulf Stream region using the adjoint method. *J. Phys. Oceanogr.*, 21(3):398–427, 1991.

[5] W. C. Thacker and R. B. Long. Fitting dynamics to data. *J. Geophys. Res.*, 93(C2):1227–1240, 1988.

[6] E. Tziperman, W. C. Thacker, R. B. Long, S. M. Hwang, and S. R. Rintoul. Oceanic data analysis using a general circulation model. Part II: A North Atlantic model. *J. Phys. Oceanog.*, 22(12):1458–1485, 1992.

[7] J. Zhu, M. Kamachi, and D. Wang. Estimation of air-sea heat flux from ocean measurements: An ill-posed problem. *Journal of Geophysical Research*, 107(10):3159–3159, 2002.

[8] X. Zou and Y. H. Kuo. Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Weath. Rev.*, 124(12):2859–2882, 1996.

[9] M. D. Gunzburger. *Perspectives in Flow Control and Optimization*. SIAM, 2003.

[10] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation (I). Theory. *Q. J. Roy. Meteor. Soc.*, 113(478):1311–1328, 1987.

[11] B. F. Farrell and A. M. Moore. An adjoint method for obtaining the most rapidly growing perturbation to oceanic flows. *J. Phys. Oceanogr.*, 22(4):338–349, 1992.

[12] D. G. Gacuci, M. I. Bujor, and I. M. Navon. *Sensitivity and Uncertainty Analysis: Applications to Large-scale Systems Vol 2*. CRC, 2005.

[13] A. M. Moore, N. S. Cooper, and D. L. T. Anderson. Initialization and data assimilation in models of the Indian Ocean. *J. Phys. Oceanogr.*, 17:1965–1977, 1987.

[14] D. N. Daescu and I. M. Navon. Adaptive observations in the context of 4D-Var data assimilation. *Meteorol. Atmos. Phys.*, 85(4):205–226, 2003.

[15] F. X. Le Dimet and I. M. Navon. Variational and optimization methods in meteorology: A review, 1988. Technical Report: Early review on variational data assimilation, SCRI report No 144.

[16] Y. Q. Zhu and I. M. Navon. Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint. *Mon. Weath. Rev.*, 127(7):1497–1517, 1999.

[17] P. Courtier, J. N. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteor. Soc.*, 120(519):1367–1387, 1994.

[18] J. Derber and A. Rosati. A global oceanic data assimilation system. *Meteorol. Atmos. Phys.*, 85(4):205–226, 2003.

[19] A. T. Weaver, J. Vialard, D. L. T. Anderson, and P. Delecluse. Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific Ocean. Part II: physical validation. *Mon. Wea. Rev.*, 131:1379–1395, 2003.

[20] M. Wenzel, J. Schröter, and D. Olbers. The annual cycle of the global ocean circulation as determined by 4D VAR data assimilation. *Prog. Oceanog.*, 48(1):73–119, 2001.

[21] P. Courtier, J. N. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120:1367–138, 1994.

[22] F. Rabier, H. Jarvinen, E. Klinker, JF. Mahfouf, and A. Simmons. The ecmwf operational implementation of fourdimensional variational data assimilation. part i: experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126:11431170, 2000.

[23] E. Blayo, J. Blum, and J. Verron. *Assimilation Variationnelle de Donnees en Oceanography et reduction de la dimension de l'espace de controle In Equations aux Derivees partielles et Applications.* Gauthiers Villars, 1998.

[24] C. Robert, S. Durbiano, E. Blayo, J. Verron, J. Blum, and F. X. le Dimet. A reduced-order strategy for 4D-VAR data assimilation. *Journal of Marine Systems*, 57(1-2):70–82, 2005.

[25] I. Hoteit and A. Kohl. Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. *J. of Oceanography*, 62(4):539–550, 2006.

[26] C. Robert, E. Blayo, J. Verron, J. Blum, and F. X. le Dimet. Reduced-order 4d-var: A preconditioner for the incremental 4d-var data assimilation method. *Geophysical Research Letters*, 33, L18609:1–4, 2006.

[27] P. T. M. Vermeulen and A. W. Heemink. Model-reduced variational data assimilation. *Monthly Weather Review*, 134:2888–2899, 2006.

[28] P. Holmes, J. L. Lumley, and G. Berkooz. *Turbulence, coherent structures, dynamical systems and symmetry.* Cambridge, UK: Cambridge University Press, 1998.

[29] Z. Luo, J. Chen, J. Zhu, R. Wang, and I. M. Navon. An optimizing reduced order FDS for the tropical Pacific ocean reduced gravity model. *International Journal for Numerical Methods in Fluids*, DOI: 10.1002/fld.1452, 2007.

[30] Z. Luo, J. Zhu, R. Wang, and I. M. Navon. Proper orthogonal decomposition approach and error estimation of mixed finite element methods for the tropical Pacific ocean reduced gravity model. *Computer methods in Applied Mechanics and Engineering*, in press, 2007.

[31] K. Fukunaga. *Introduction to Statistical Recognition Second Edition (Computer Science and Scientific Computing Series).* Boston: Academic Press, 1990.

[32] N. Aubry, P. Holmes, and J. L. Lumley. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *Journal of Fluid Dynamics*, 192:115–173, 1988.

[33] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal*, 40(11):2323–2330, 2002.

[34] M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[35] K. Willcox, O. Ghattas, B. van Bloemen Waanders, and B.Bader. An optimization framework for goal-oriented, model-based reduction of large-scale systems. *44th IEEE Conference on Decision and Control and European Control Conference*, 2005.

[36] T. Bui-Thanh, K. Willcox, O. Ghattas, and B. van Bloemen Waanders. Goal-oriented, model-constrained optimization for reduction of large-scale systems. *J. of Computational Physics*, 224(2):880–896, 2007.

[37] Y. Cao, J. Zhu, I. M. Navon, and Z. Luo. A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 53(10):1571–1583, 2006.

[38] D. N. Daescu and I. M. Navon. A dual-weighted approach to order reduction in 4d-var data assimilation. *In press, Monthly Weather Review.*

[39] F. Fang, C. C. Pain, I. M. Navon, M. D. Piggott, G. J. Gorman, P. Allison, and A. J. H. Goddard. Reduced order modelling of an adaptive mesh ocean model. *International Journal for Numerical Methods in Fluids*, sub-judice, 2007.

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

[41] S. S. Ravindran. Reduced-order controllers for control of flow past an airfoil. *International Journal for Numerical Methods in Fluids*, 50(5):531–554, 2002.

[42] R. Ford, C. C. Pain, M. D. Piggott, A. J. H. Goddard, C. R. E. de Oliveira, and A. P. Umbleby. A nonhydrostatic finite-element model for three-dimensional stratified oceanic flows. Part I: Model formulation. *Mon. Weath. Rev.*, 132(12):2816–2831, 2004.

[43] C. C. Pain, A. P. Umpleby, C. R. E. de Oliveira, and A. J. H. Goddard. Tetrahedral mesh optimisation and adaptivity for steady-state and transient finite element calculations. *Comput. Methods Appl. Mech. Engrg.*, 190:3771–3796, 2001.

[44] M. D. Piggott, C. C. Pain, G. J. Gorman, P. W. Power, and A. J. H. Goddard. h, r, and hr adaptivity with applications in numerical ocean modelling. *Ocean Modell.*, 10(1–2):95–113, 2005.

[45] P. W. Power, C. C. Pain, M. D. Piggott, G. J. Gorman, F. Fang, D. P. Marshall, and A. J. H. Goddard. Adjoint goal-based error norms for adaptive mesh ocean modelling. *Ocean Modell.*, 15:3 – 38, doi:10.1016/j.ocemod.2006.05.001, 2006.

List of Figures