INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2009; **60**:709–732
Published online 9 September 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1911

A POD reduced-order 4D-Var adaptive mesh ocean modelling approach

F. Fang[1,*,†], C. C. Pain[1], I. M. Navon[2], M. D. Piggott[1], G. J. Gorman[1], P. E. Farrell[1], P. A. Allison[1] and A. J. H. Goddard[1]

[1]*Applied Modelling and Computation Group, Department of Earth Science and Engineering, South Kensington Campus, Imperial College London, SW7 2AZ, U.K.*
[2]*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, U.S.A.*

SUMMARY

A novel proper orthogonal decomposition (POD) inverse model, developed for an adaptive mesh ocean model (the Imperial College Ocean Model, ICOM), is presented here. The new POD model is validated using the Munk gyre flow test case, where it inverts for initial conditions. The optimized velocity fields exhibit overall good agreement with those generated by the full model. The correlation between the inverted and the true velocity is 80–98% over the majority of the domain. Error estimation was used to judge the quality of reduced-order adaptive mesh models. The cost function is reduced by 20% of its original value, and further by 70% after the POD bases are updated.

In this study, the reduced adjoint model is derived directly from the discretized reduced forward model. The whole optimization procedure is undertaken completely in reduced space. The computational cost for the four-dimensional variational (4D-Var) data assimilation is significantly reduced (here a decrease of 70% in the test case) by decreasing the dimensional size of the control space, in both the forward and adjoint models. Computational efficiency is further enhanced since both the reduced forward and adjoint models are constructed by a series of time-independent sub-matrices. The reduced forward and adjoint models can be run repeatedly with negligible computational costs.

An adaptive POD 4D-Var is employed to update the POD bases as minimization advances and loses control, thus adaptive updating of the POD bases is necessary. Previously developed numerical approaches Fang *et al*. (*Int. J. Numer. Meth. Fluids* 2008) are employed to accurately represent the geostrophic balance and improve the efficiency of the POD simulation. Copyright © 2008 John Wiley & Sons, Ltd.

Received 14 December 2007; Revised 25 July 2008; Accepted 28 July 2008

*Correspondence to: F. Fang, Applied Modelling and Computation Group, Department of Earth Science and Engineering, South Kensington Campus, Imperial College London, SW7 2AZ, U.K.
†E-mail: f.fang@imperial.ac.uk

Contract/grant sponsor: UK's Natural Environment Research Council; contract/grant numbers: NER/A/S/2003/00595, NE/C52101X/1, NE/C51829X/1
Contract/grant sponsor: Engineering and Physical Sciences Research Council; contract/grant number: GR/R60898
Contract/grant sponsor: Leverhulme Trust; contract/grant number: F/07058/AB
Contract/grant sponsor: Imperial College High Performance Computing Service
Contract/grant sponsor: Grantham Institute for Climate Change
Contract/grant sponsor: NSF; contract/grant numbers: ATM-0201808, CCF-0635162

Copyright © 2008 John Wiley & Sons, Ltd.

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

1. INTRODUCTION

The threat of impending climate change highlights the importance of improving the predictive capabilities of models of the oceans, as well as other components of the Earth system. Data assimilation techniques are a critical component of ocean modelling. 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 parameters and wind stresses, can be optimized [1–8].

A variety of approaches have been used in the past to facilitate data assimilation and include statistical interpolation methods, nudging data assimilation and variational methods along with sequential estimation such as Kalman filter, extended Kalman filter and Ensemble Kalman filter. In particular, the four-dimensional variational (4D-Var) method has proved an efficient means of assimilating observed data into simulations [9, 10]. The 4D-Var method is capable of producing a best estimate model solution by fitting a numerical simulation to observational data over both space and time. The technique also facilitates the estimation of the error sources caused by uncertainties (boundary conditions, initial condition and parameters) in the model. The solution is derived by minimizing 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. 4D-Var data assimilation has been used widely in both atmospheric and oceanographic models over the past two decades [4, 6–9, 11–20]. However, the major difficulty in the implementation of 4D-Var data assimilation in an ocean model 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 the 4D-Var method incurs high memory and computational costs. However, the computational cost can be reduced by decreasing the dimensions of the control space thus ensuring that the minimization of the cost function (or error covariances) is carried out within a low-dimensional space. This can be achieved through an incremental 4D-Var [21] whereby a succession of quadratic problems are generated over increasing time periods. The successive quadratic minimization problems can then be solved by running tangent linear model and adjoint model approximations using a coarse resolution in the inner-loop of the minimization. Using this approach, the dimension of the minimization problem can be decreased by one or two orders of magnitude. Although such an incremental approach is currently used in operational atmospheric models, the dimensions of the control space remain quite large in realistic applications [22–24].

The use of empirical orthogonal functions (EOF) analysis has been advocated as an approach that can lead to reduced-order ocean modelling [25]. The implementation of this method additionally results in a drastic reduction of the dimension of the control space and thus the iterative minimization process [22, 26]. Reduced-order 4D-Var can also be used to precondition 4D-Var and reduce computational cost [27]. It has been further proposed [28] that efficiencies can be enhanced if the adjoint model can be directly implemented in a subspace of the reduced-order model (determined by the leading EOFs) and then used to approximate the gradient of the cost function. The minimization process can thus be solved completely in reduced space with negligible computational costs.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

Proper orthogonal decomposition (POD) methodologies, in combination with the Galerkin projection procedure, have additionally provided an efficient means of generating reduced-order models [29–31]. 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 diverse disciplines, including signal analysis and pattern recognition [32], fluid dynamics and coherent structures [29, 33, 34] and image reconstruction [35]. To improve the accuracy of reduced-order models, a goal-oriented approach has been used to optimize the POD bases [36, 37]. 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 [22, 24, 30, 38]. The POD-based 4D-Var not only reduces the dimension of control space, but also reduces the size of the dynamical model, both in dramatic ways [38–40].

Herein we describe a POD reduced-order 4D-Var for an adaptive mesh ocean model. A POD-based reduced-order forward model [41] has been developed for the Imperial College Ocean Model (ICOM) that can simultaneously resolve both small- and large-scale ocean flows while 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-order forward model is derived in a subspace (details in [41]). Once the reduced-order forward model is available, the reduced-order adjoint model can be directly obtained from the POD reduced-order forward model in the subspace, instead of the original forward model. The minimization 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, which include snapshots of varying length at time levels. To overcome these difficulties, a standard reference fixed mesh is adopted for both the forward and adjoint reduced-order 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-order forward and adjoint models. However, this also introduces an interpolation error into snapshots and the POD reduced-order model. To reduce the interpolation error, high-order interpolation approaches are employed. Here, a quantitative comparison of POD solutions with the use of the linear, quadratic and cubic interpolation schemes has been carried out.

In this work, an adaptive POD procedure is employed to improve the reduced-order model by updating the POD basis. The original reduced basis for inverse problems is calculated using a set of snapshots based on the results from the full forward model with the specified control variables. The re-calculation of the reduced basis is needed when the resulting control variables from the optimization procedure are significantly different from those that the POD model is based on. Ravindran [42, 43] proposed an adaptive procedure that successively updates the reduced-order model being used via a Sequential Quadratic Programming constrained optimization algorithm. Cao *et al*. [38] introduced an adaptive POD approach into POD inverse (adjoint) models. In this approach the reduced basis is re-calculated using a refreshed set of snapshots based on the latest results obtained from the full forward model using a restart criterion of the adaptive POD procedure based on convergence of the minimization process. One can also consider the trust region method for restart criteria [44].

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

2. POD REDUCED MODEL

POD (also known as the Karhunen–Loéve decomposition method, principal components analysis and EOF) is the most efficient choice among linear decompositions in the sense that it can capture the greatest possible energy. A three-dimensional (3D) dynamical flow model is generally expressed as

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

where f is a general function representing 3D nonlinear flow dynamics (here, the incompressible Navier–Stokes equations), $\mathbf{u}$ is a vector containing all variables to be solved (e.g. velocities, pressure and temperature), t is time and $\mathbf{x} = (x, y, z)^{\mathrm{T}}$ represents the Cartesian coordinate position.

2.1. Proper orthogonal decomposition

The model variables $\mathbf{u}$ are sampled at defined checkpoint times during the simulation period $[t_1, \ldots, t_K]$, also referred to as snapshots $U = (U_1, \ldots, U_K)^{\mathrm{T}}$ where K is the number of snapshots. The snapshots can be obtained either from a 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, $U_k = (U_{k,1}, \ldots, U_{k,i}, \ldots, U_{k,\mathcal{N}})$, where $\mathcal{N}$ is the number of nodes and U represents one of variables u, v, w, p, etc. The average of the ensemble of snapshots is defined as

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

where $U_{k,i}$ is the model variable value at the snapshot k and node i. Taking the deviation from the mean of variables yields

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

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

$$AA^{\mathrm{T}} x_k = \lambda_k x_k, \quad 1 \leqslant k \leqslant K \tag{4}$$

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

$$A^{\mathrm{T}} A y_k = \lambda_k y_k, \quad 1 \leqslant k \leqslant K \tag{5}$$

This procedure is equivalent to a singular value decomposition. The eigenvalues λ_k are real and positive and for reasons that will be apparent, should be sorted in descending order. The POD basis vectors Φ_k associated with the eigenvalues λ_k are orthogonal (normalized in the L_2 norm) 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 energy transferred within the kth basis mode. If the POD spectrum (energy) decays fast enough, practically all the support of the invariant measure is

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

contained in a compact set. Roughly speaking, all the likely realizations 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, where $M \ll K$:

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

where $I(M)$ represents the percentage of energy that is captured by the POD basis $\Phi_1, \ldots, \Phi_m, \ldots, \Phi_M$.

2.2. Reduced-order model

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

$$\mathbf{u}(t, x, y, z) = \bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_m(t) \Phi_m(\mathbf{x}) \tag{8}$$

where $\bar{\mathbf{u}}$ is the mean of the ensemble of snapshots for the variables $\mathbf{u}(t)$, α_m $(1 \leqslant m \leqslant M)$ are the time-dependent coefficients to be determined and $\alpha_m(0)$ are the coefficients at the initial time level. Substituting (8) into (1) and taking the POD basis function as the test function, then integrating over the computational domain Ω yields the POD reduced-order model

$$\frac{\partial \alpha_j}{\partial t} = \left\langle f\left(\left(\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_m(t) \Phi_m(\mathbf{x})\right), t, \mathbf{x}\right), \Phi_j \right\rangle, \quad 1 \leqslant j \leqslant M \tag{9}$$

subject to the initial condition

$$\alpha_j(0) = ((\mathbf{u}(0, \mathbf{x}) - \bar{\mathbf{u}}(\mathbf{x})), \Phi_j) \tag{10}$$

where $\langle \cdot, \cdot \rangle$ is the canonical inner product in the L^2 norm. In the finite element method, the POD basis $\Phi_m(\mathbf{x}) = \sum_{i=1}^{\mathcal{N}} N_i(x) \Phi_{m,i}$; hence, (9) can be expressed as

$$\frac{\partial \alpha_j}{\partial t} = \left\langle f\left(\left(\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_m(t) \sum_{i=1}^{\mathcal{N}} N_i \Phi_{m,i}\right), t, \mathbf{x}\right), \Phi_j \right\rangle \tag{11}$$

where N_i is the basis function in the finite element and $\mathcal{N}$ 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). The optimal solution for (1) is obtained by minimizing the functional $\Im(U^0)$:

$$\Im(U^0) = \frac{1}{2}(U^0 - U_b)^{\mathrm{T}} \mathbf{B}^{-1} (U^0 - U_b) + \frac{1}{2} \sum_{n=1}^{N_t} (\mathbf{H} U^n - y_o^n)^{\mathrm{T}} R^{-1} (\mathbf{H} U^n - y_o^n) \tag{12}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

where $\mathbf{B}$ is the background error covariance matrix, R^{-1} 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, N_t is the number of time levels) and y_o^n is the observation at time level n. In a POD reduced model, the initial value U^0 and the reduced-order solution U^n are expressed as

$$U^0 = \bar{U} + \sum_{m=1}^{M} \alpha_m(0)\Phi_m(\mathbf{x}) \tag{13}$$

$$U^n = \bar{U} + \sum_{m=1}^{M} \alpha_m(t^n)\Phi_m(\mathbf{x}) \tag{14}$$

Substituting (13) and (14) into (12) yields

$$\begin{aligned}\Im(\alpha(0)) = {} & \frac{1}{2}\left(\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(0)\Phi_m(\mathbf{x})\right) - U_b\right)^{\mathrm{T}} \mathbf{B}^{-1} \left(\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(0)\Phi_m(\mathbf{x})\right) - U_b\right) \\ & + \frac{1}{2}\sum_{n=1}^{N_t} \left(\mathbf{H}\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(t^n)\Phi_m(\mathbf{x}),\right) - y_o^n\right)^{\mathrm{T}} R^{-1} \left(\mathbf{H}\left(\bar{U} + \sum_{m=1}^{M} \alpha_m(t^n)\Phi_m(\mathbf{x}),\right) - y_o^n\right)\end{aligned} \tag{15}$$

where the snapshots are chosen at time intervals with a constant time interval between them during the simulation period.

3.1. Discrete reduced-order adjoint equations

Here, the POD reduced-order model (11) for the incompressible Navier–Stokes equations can be expressed as

$$\begin{aligned}& \frac{\partial \alpha_j(t)}{\partial t} + \left\langle \Phi_{\mathbf{u},j}, \left(\left(\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_{\mathbf{u},m}(t)\Phi_{\mathbf{u},m}(\mathbf{x})\right) \cdot \nabla \left(\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_{\mathbf{u},m}(t)\Phi_{\mathbf{u},m}\right)\right)\right\rangle \\ & + \left\langle \Phi_{\mathbf{u},j}, \left(f\mathbf{k} \times \left(\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_{\mathbf{u},m}(t)\Phi_{\mathbf{u},m}(\mathbf{x})\right)\right)\right\rangle \\ & + \left\langle \Phi_{\mathbf{u},j}, \left(\nabla p - \mu\nabla^2 \left(\bar{\mathbf{u}} + \sum_{m=1}^{M} \alpha_{\mathbf{u},m}(t)\Phi_{\mathbf{u},m}(\mathbf{x})\right)\right)\right\rangle = 0\end{aligned} \tag{16}$$

where $\bar{\mathbf{u}}$ is the mean of the ensemble of snapshots for velocity field $\mathbf{u} = (u, v, w)^{\mathrm{T}}$, p is the pressure, f represents the Coriolis inertial force, μ is the kinematic viscosity and $\mathbf{k} = (0, 0, 1)^{\mathrm{T}}$. Taking into account second-order Crank–Nicolson time stepping, for example, the above equation can be

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

re-expressed as

$$\frac{\partial \alpha_j(t)}{\partial t}+\left\langle \Phi_{\mathbf{u},j},\left(\left(\bar{\mathbf{u}}+\sum_{m=1}^{M}\alpha_{\mathbf{u},m}^{n-1}\Phi_{\mathbf{u},m}\right)\cdot\nabla\left(\bar{\mathbf{u}}+\sum_{m=1}^{M}(0.5\alpha_{\mathbf{u},m}^{n-1}+0.5\alpha_{\mathbf{u},m}^{n})\Phi_{\mathbf{u},m}\right)\right)\right\rangle$$
$$+\left\langle \Phi_{\mathbf{u},j},\left(f\mathbf{k}\times\left(\bar{\mathbf{u}}+\sum_{m=1}^{M}(0.5\alpha_{\mathbf{u},m}^{n-1}+0.5\alpha_{\mathbf{u},m}^{n})\Phi_{\mathbf{u},m}\right)\right)\right\rangle$$
$$+\left\langle \Phi_{\mathbf{u},j},\left(\nabla p-\mu\nabla^2\left(\bar{\mathbf{u}}+\sum_{m=1}^{M}(0.5\alpha_{\mathbf{u},m}^{n-1}+0.5\alpha_{\mathbf{u},m}^{n})\Phi_{\mathbf{u},m}\right)\right)\right\rangle=0 \tag{17}$$

The discrete model of (17) at the time level n can be expressed in a general form in a subspace

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

where

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

$\mathbf{A}^n$ 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 (17) or (9), $\alpha^n=(\alpha_1^n,\ldots,\alpha_M^n)$ and $\alpha^{n-1}=(\alpha_1^{n-1},\ldots,\alpha_M^{n-1})$ 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 (1), s^n is a discretized source term at the time level n and f_s is a source term including the forcing terms on the boundaries.

For a nonlinear simulation, the matrices $\mathbf{A}^n$ and $\mathbf{B}^n$ can be expressed 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 $\mathbf{A}^n,\mathbf{B}^n,\hat{\mathbf{A}}_0^n,\hat{\mathbf{B}}_0^n,\hat{\mathbf{A}}_m^n,\hat{\mathbf{B}}_m^n\in R^{M\times M}$ (dependent on the governing equations and corresponding discretization). The components of the matrices $\mathbf{A}^n$ and $\mathbf{B}^n$ in (17) can be expressed as

$$\hat{\mathbf{A}}_{0,i,j}^n=1+\langle\Phi_{\mathbf{u},i},0.5\Delta t(\bar{\mathbf{u}}\cdot\nabla\Phi_{\mathbf{u},j}+f\mathbf{k}\times\Phi_{\mathbf{u},j}+\nabla p-\mu\nabla^2\Phi_{\mathbf{u},j})\rangle,\quad 1\leqslant i,j\leqslant M \tag{22}$$

$$\hat{\mathbf{A}}_{m,i,j}^n=\langle\Phi_{\mathbf{u},i},0.5\Delta t(\Phi_{\mathbf{u},m}\cdot\nabla\Phi_{\mathbf{u},j})\rangle,\quad 1\leqslant i,j\leqslant M \tag{23}$$

$$\hat{\mathbf{B}}_{0,i,j}^n=1+\langle\Phi_{\mathbf{u},i},0.5\Delta t(\bar{\mathbf{u}}\cdot\nabla\Phi_{\mathbf{u},j}+f\mathbf{k}\times\Phi_{\mathbf{u},j}+\nabla p-\mu\nabla^2\Phi_{\mathbf{u},j})+\Delta t\Phi_{\mathbf{u},j}\cdot\nabla\bar{\mathbf{u}}\rangle$$
$$1\leqslant i,j\leqslant M \tag{24}$$

$$\hat{\mathbf{B}}_{m,i,j}^n=\langle\Phi_{\mathbf{u},i},0.5\Delta t(\Phi_{\mathbf{u},m}\cdot\nabla\Phi_{\mathbf{u},j})\rangle,\quad 1\leqslant i,j\leqslant M \tag{25}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

Equations (20) and (21) can be re-expressed 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 $\hat{\mathbf{A}}^n = (\hat{\mathbf{A}}_1^n, \ldots, \hat{\mathbf{A}}_M^n)$ and $\hat{\mathbf{B}} = (\hat{\mathbf{B}}_1^n, \ldots, \hat{\mathbf{B}}_M^n)$, see [41].

Taking into account (18), the discrete forward equation during the simulation period $[t_1, \ldots, t_{N_t}]$ can be expressed as

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

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^1 & & & \\ -\mathbf{B}^2 & \mathbf{A}^2 & & \\ & \ddots & \ddots & \\ & & -\mathbf{B}^{N_t} & \mathbf{A}^{N_t} \end{pmatrix} \tag{29}$$

and

$$\alpha = (\alpha^1, \alpha^2, \ldots, \alpha^{N_t})^{\mathrm{T}} \tag{30}$$

$$s = (\mathbf{B}^1 \alpha^0, 0, \ldots, 0)^{\mathrm{T}} + f_s \tag{31}$$

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

$$\bar{\mathbf{A}}\alpha + \mathbf{A}\bar{\alpha} = \bar{s} \tag{32}$$

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

$$\bar{\mathbf{A}} = \frac{\partial \mathbf{A}}{\partial \alpha^0} = \begin{pmatrix} \bar{\mathbf{A}}^1 & & & \\ -\bar{\mathbf{B}}^2 & \bar{\mathbf{A}}^2 & & \\ & \ddots & \ddots & \\ & & -\bar{\mathbf{B}}^{N_t} & \bar{\mathbf{A}}^{N_t} \end{pmatrix} \tag{33}$$

and

$$\bar{\alpha} = \frac{\partial \alpha}{\partial \alpha^0} = (\bar{\alpha}^1, \bar{\alpha}^2, \ldots, \bar{\alpha}^{N_t})^{\mathrm{T}} \tag{34}$$

$$\bar{s} = \frac{\partial s}{\partial \alpha^0} = (\bar{B}^1 \alpha^0, 0, \ldots, 0)^{\mathrm{T}} + (B^1 \bar{\alpha}^0, 0, \ldots, 0)^{\mathrm{T}} \tag{35}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

where taking into account (26) and (27)

$$\bar{\mathbf{A}}^n = \frac{\partial \mathbf{A}^n}{\partial \alpha^0} = \hat{\mathbf{A}} \bar{\alpha}^{n-1} \tag{36}$$

$$\bar{\mathbf{B}}^n = \frac{\partial \mathbf{B}^n}{\partial \alpha^0} = \hat{\mathbf{B}} \bar{\alpha}^{n-1} \tag{37}$$

where n describes time levels, $1 \leqslant n \leqslant N_\mathrm{t}$. Equation (32) can be therefore re-expressed as

$$\begin{pmatrix} \bar{\mathbf{A}}^1 & & & \\ -\bar{\mathbf{B}}^2 & \bar{\mathbf{A}}^2 & & \\ & \ddots & \ddots & \\ & & -\bar{\mathbf{B}}^{N_\mathrm{t}} & \bar{\mathbf{A}}^{N_\mathrm{t}} \end{pmatrix} \begin{pmatrix} \alpha^1 \\ \alpha^2 \\ \vdots \\ \alpha^{N_\mathrm{t}} \end{pmatrix} + \begin{pmatrix} \mathbf{A}^1 & & & \\ -\mathbf{B}^2 & \mathbf{A}^2 & & \\ & \ddots & \ddots & \\ & & -\mathbf{B}^{N_\mathrm{t}} & \mathbf{A}^{N_\mathrm{t}} \end{pmatrix} \begin{pmatrix} \bar{\alpha}^1 \\ \bar{\alpha}^2 \\ \vdots \\ \bar{\alpha}^{N_\mathrm{t}} \end{pmatrix}$$
$$= \begin{pmatrix} \hat{\mathbf{B}}^1 \alpha^0 \bar{\alpha}^0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{B}^1 \bar{\alpha}^0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{38}$$

The tangent linear model is then derived

$$(\mathbf{A} + \mathbf{A}_\mathrm{extra}) \frac{\partial \alpha}{\partial \alpha^0} = \frac{\partial s}{\partial \alpha^0} \tag{39}$$

where $\mathbf{A}$ is calculated in (29) and $\mathbf{A}_\mathrm{extra}$ originates from the nonlinear terms and is expressed as

$$\mathbf{A}_\mathrm{extra} = \begin{pmatrix} 0 & & & \\ \hat{\mathbf{A}}\alpha^2 - \hat{\mathbf{B}}\alpha^1 & 0 & & \\ & \ddots & \ddots & \\ & & \hat{\mathbf{A}}\alpha^{N_\mathrm{t}} - \hat{\mathbf{B}}\alpha^{N_\mathrm{t}-1} & 0 \end{pmatrix} \tag{40}$$

The variation of the objective function (12) with respect to the control variables α^0 is

$$\frac{\partial \Im}{\partial \alpha^0} = \left(\frac{\partial \alpha}{\partial \alpha^0} \right)^\mathrm{T} \frac{\partial \Im}{\partial \alpha} \tag{41}$$

Taking into account Equation (39), yields

$$\frac{\partial \Im}{\partial \alpha^0} = \left(\frac{\partial s}{\partial \alpha^0} \right)^\mathrm{T} (\mathbf{A} + \mathbf{A}_\mathrm{extra})^{-\mathrm{T}} \frac{\partial \Im}{\partial \alpha} \tag{42}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

The gradient of the objective function can then be expressed as

$$\frac{\partial \Im}{\partial \alpha^0} = \left(\frac{\partial s}{\partial \alpha^0} \right)^{\mathrm{T}} \alpha^* \tag{43}$$

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

$$(\mathbf{A} + \mathbf{A}_{\mathrm{extra}})^{\mathrm{T}} \alpha^* = \frac{\partial \Im}{\partial \alpha} \tag{44}$$

3.2. Adaptive POD

In this work, the POD model is based on the solution of the original model for specified control variables (e.g. initial and boundary conditions). It is therefore necessary to reconstruct the POD model when the resulting control variables from the latest optimization iteration are significantly different from the ones upon which the POD model is based. An adaptive POD 4D-Var procedure is used to periodically update the POD basis, and the reduced-order direct and inverse models (for detail see [42, 43]). The inversion procedure starts with the initial estimation of the control variables. An initial set of snapshots is obtained by running the full forward model, and the corresponding POD subspace and reduced-order model are constructed. The adaptive POD procedure proposed is as follows:

1. Set the POD iteration level it=1 and the initial guess controls c_{it}.
2. Set up the snapshots U_{it} from the solution of the full forward model with the controls c_{it}.
3. Calculate the POD bases (the number of POD bases is chosen to capture a prescribed energy level).
4. Project the controls c_{it} on the reduced space $\alpha_{\mathrm{it},jt}$ $(jt=1)$.
5. Optimize the initial controls $\alpha_{\mathrm{it},jt}$ (note: the optimization procedure is carried out completely on the reduced space. The Polak–Ribière nonlinear conjugate gradient (CG) technique is employed here and jt is the Nonlinear CG iteration level).
6. (a) check the value of cost function (12). If $|\Im_{jt}|<\varepsilon$ (where ε is the tolerance for the optimization), then go to step 7;
   (b) if $|\Im_{jt}|>\varepsilon$ and $|\Im_{jt}-\Im_{jt-1}|>10^{-3}$ (where $jt-1$ and jt are the consecutive optimization iteration levels), then set $jt=jt+1$ and go back step 5;
   (c) if $|\Im_{jt}|>\varepsilon$ and $|\Im_{jt}-\Im_{jt-1}|<10^{-3}$, then update the POD bases:
      (i) find the new controls $c_{\mathrm{it}+1}$ by projecting the optimization controls $\alpha_{\mathrm{it},j}$ onto the original flow domain and
      (ii) set it=it+1 and go back step 2.
7. The adaptive POD optimization procedure is completed.

4. MESH ADAPTIVITY IN REDUCED MODELS

4.1. Description of ICOM and anisotropic mesh adaptivity

The POD-based reduced model presented was implemented for ICOM. This unstructured adaptive mesh model can simultaneously resolve both small- and large-scale ocean flows while smoothly varying resolution and conforming to complex coastlines and bathymetry. With more appropriate

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

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. The underlying model equations consist of the 3D incompressible Navier–Stokes equations. To accurately represent local flow around steep topography, the hydrostatic assumption is not made. The pressure is split into the geostrophic and ageostrophic parts, which are solved for separately. This allows the accurate representation of hydrostatic/geostrophic balance [45, 46].

A dynamically adapting anisotropic mesh in 3D is used here [47, 48]. Mesh adaptivity or optimization relies on the derivation of appropriate error measures, which dictate how the mesh is to be modified [47]. A metric tensor is used to calculate the required edge lengths and orientation of the mesh elements to control solution errors. It is constructed so that an ideal edge length is unity when measured in metric space. Since the metric is dependent on both location and direction it is able to reflect locally anisotropic information within the solution. Thus inhomogeneous and anisotropic meshes result from this approach. By defining an objective functional, which is based on the element quality in this metric space, an optimization technique is used to improve the overall quality of the mesh. Local operations (based on a series of mesh connectivity and node position searches) are performed on a 3D tetrahedral mesh and include: edge collapsing/splitting; face to edge and edge to face swapping; edge to edge swapping; and local node movement or mesh smoothing in a fashion similar to Freitag and Ollivier-Gooch [49] and Buscaglia and Dari [50]. Constraints are imposed on these operations so as to preserve the integrity of non-planar geometrical boundaries [47, 48].

To reduce the interpolation error, high-order interpolation approaches are employed. For each node in the reference mesh, the element of the adapted mesh in which the node lies is identified. Then, a local higher-order polynomial is fitted with a least-squares approach on a patch of nodes around this element. In the work presented here, the polynomial is either quadratic or cubic. This polynomial is then evaluated at the location of the node in the reference mesh to determine the interpolated value. This scheme is exact for polynomials up to the degree of the fitted polynomial, and thus is more accurate than linear interpolation, but it comes at a higher cost as a dense matrix (10×10 for cubic in two-dimensional (2D), 20×20 in 3D), must be inverted for each element in the original mesh to compute the interpolated solution.

4.2. Adaptive mesh technique in POD and error estimate

When adaptive meshes are employed, 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 lengths at different time levels. This unavoidably brings difficulties in the implementation of a POD-based reduced model for an adaptive mesh model. To overcome these difficulties, a standard reference fixed mesh is adopted for the reduced-order 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, and thus the reference mesh must be chosen carefully so that it is at least as fine as the finest adapted mesh at a given location. To reduce the interpolation error, high-order interpolation approaches are employed.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

5. APPLICATION AND DISCUSSION

The new model has been applied to 2D gyre flows in a computational domain, 1000 km by 1000 km with a depth of $H=500\,\mathrm{m}$. The underlying model equations consist of the 3D incompressible Navier–Stokes equations (for 2D flow cases, one element in the vertical). The initial conditions are optimized using the POD reduced-order adjoint model. The accuracy and validation of the reduced-order POD adjoint model have been evaluated. Error estimation 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 (when the value of the cost function cannot be decreased by more than 10^{-3} between the consecutive iterations), and its effectiveness is discussed. The Polak–Ribière nonlinear CG multivariate unconstrained minimization is employed in the implementation of the inversion.

5.1. Description of the case: Gyre

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

$$\tau_y=\tau_0\cos(\pi y/L),\quad \tau_x=0.0 \tag{45}$$

where τ_x and τ_y are the wind stresses on the free surface along the x and y directions, respectively, and $L=1000\,\mathrm{km}$. A maximum zonal wind stress of $\tau_0=0.1\,\mathrm{N\,m^{-1}}$ is applied in the latitude (y) direction. The Coriolis terms are taken into account with the beta-plane approximation ($f=\beta y$) where $\beta=1.8\times10^{-11}$ and the reference density $\rho_0=1000\,\mathrm{kg\,m^{-1}}$.

The problem is non-dimensionalized with the maximum Sverdrup balance velocity

$$\beta H\rho_0 v=\frac{\partial\tau}{\partial y}\leqslant\frac{\tau_0\pi}{L}\Rightarrow v\leqslant3.5\times10^{-2}\,\mathrm{m\,s^{-1}} \tag{46}$$

(and so the velocity scale $U=3.5\times10^{-2}\mathrm{m\,s^{-1}}$ is used here), and the length scale $L=1000\,\mathrm{km}$. Time is non-dimensionalized with $T=L/U$. Incorporating the beta-plane approximation gives a non-dimensional $\beta^*=L^2\beta/U=514.286$. The non-dimensional wind stress (applied as a body force here averaged over the depth of the domain) takes the same cosine of latitude profile with $\tau_0^*=\tau_0L/(U^2\rho_0H)=163.2653$. The Reynolds number is defined as $Re=UL/\nu=250$ (here the kinematic viscosity is $140\,\mathrm{m^2\,s^{-1}}$). The time step is 3.78×10^{-4}, equivalent to $3h$. No-slip boundary conditions are applied to the lateral boundaries. The pseudo-observational data are taken on days 125, 150 and 175 over the computational domain. The starting guess values of the initial conditions are given by the background flow (here, taken from the 'true' flow fields on days either 107.5 or 137.5 in the experiments).

The POD bases are constructed by the snapshots that are obtained from the numerical solutions by forcing the full forward model with the background flow. Forty snapshots with 35 POD bases for the velocity field u, v, w and pressure are chosen, which capture more than 99.5% of energy (calculated by the first 35 leading eigenvalues (7)).

To accurately represent geostrophic pressure, its basis functions are split into two sets: Φ_{pgu} and Φ_{pgv}, which are associated with the u- and v-velocity components. Furthermore the geostrophic

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

pressure can be represented by a summation of the two sets of geostrophic basis functions:

$$p_g = \bar{p}_g + \sum_{m=1}^{M} \alpha_{pgu,m} \Phi_{u,m} + \sum_{m=1}^{M} \alpha_{v,m} \Phi_{pgv,m} \tag{47}$$

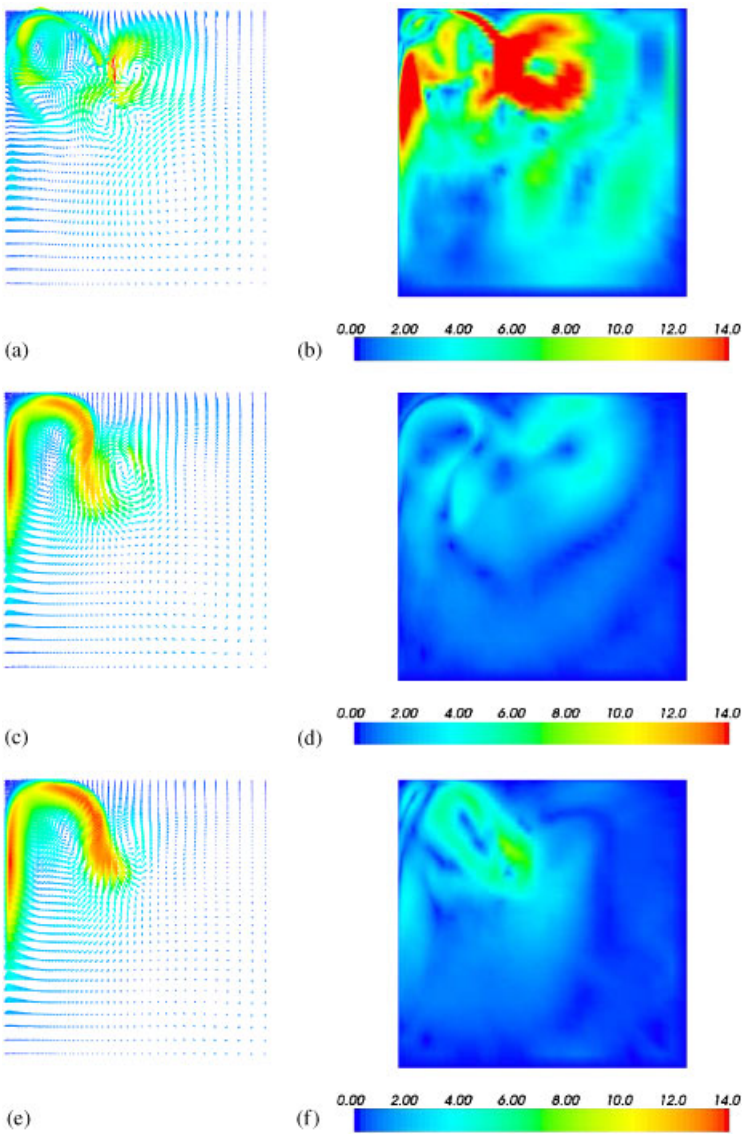

Figure 1. Optimization problem (local minimum) introduced by an unappropriate choice of initial guess controls (the initial guess controls is taken from the true flow field on the 137.5th day). Left panel: inverted model velocity field; right panel: error for the velocity field: (a) at the initial time level; (b) at the initial time level; (c) at time level $t = 125$ days; (d) at time level $t = 125$ days; (e) at time level $t = 175$ days; and (f) at time level $t = 175$ days.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

where $\bar{p}_g$ is the mean of the ensemble of snapshots for the variable p. The geostrophic basis functions are calculated by solving the elliptic equations (the geostrophic balance equations) using a CG iterative method:

$$-\nabla^2 \Phi_{pgu,m} = \frac{\partial(f\Phi_{u,m})}{\partial y}$$
$$-\nabla^2 \Phi_{pgv,m} = \frac{\partial(-f\Phi_{v,m})}{\partial x} \tag{48}$$

The geostrophic pressure has a quadratic finite element representation while linear finite element representations are used for the velocity components.

An adaptive mesh is adopted in the full model. The mesh for the full model adapts every 19 time steps with maximum and minimum allowed 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 9). To build up the snapshots, the solutions from the full forward model are interpolated from the adaptive mesh (left panel in Figure 9) onto the reference fixed mesh.

5.2. Issues with adaptive reduced 4D-Var

As discussed in Section 3.2, the POD reduced model based on the background flow can be improved by updating the POD bases. Here, the POD bases are re-calculated when the value of the cost function cannot be decreased by more than 10^{-3} between the consecutive optimization iterations. The resulting control variables from the latest optimization iteration are applied back to the full model to generate the new POD bases. The new POD bases then replace the previous ones to derive a new POD reduced-order model [42].

However, in the experiments the updated POD bases may be unacceptable if the control variables optimized during the latest POD iteration are far from the true values. As an example, the initial guess controls (background flow) are taken from the true flow field on the 137.5th day. By fitting the numerical solution to observational data, the errors of the inverted model velocity field at the time levels ($t = 125, 150, 175$ days) are reduced to small values (less than $8\,\mathrm{m\,s^{-1}}$ in Figure 1).

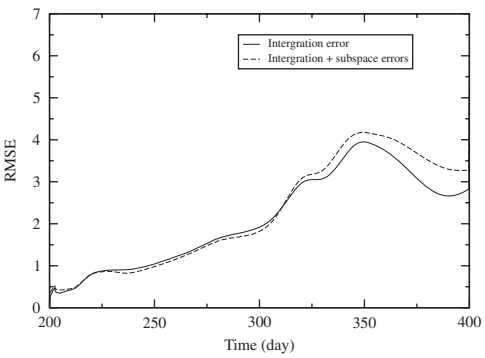

Figure 2. Error of the POD reduced forward model. Solid line: integration error (41 snapshots and 41 POD bases); dashed line: total reduced model error (integration error plus project error, 41 snapshots and 35 POD bases).

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

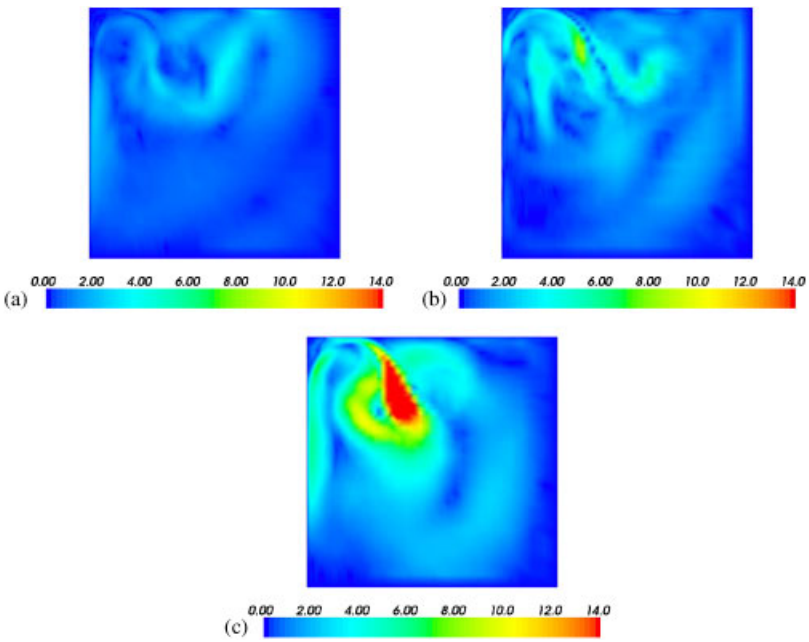

Figure 3. Error in the velocity field from the POD reduced model with the true initial conditions at time levels (a) $t=125$ days; (b) $t=150$ days; and (c) $t=175$ days.

The cost function (12) is reduced by 73% of its original value during the optimization procedure. However, the optimized initial controls during the current POD inversion iteration are far from what it is expected (inverted model velocity on the top panel in Figure 1 and the true value in Figure 7(a)). It is assumed that a local minimum was attained in this case. Obviously, the optimized initial conditions cannot be used to update the POD bases. An appropriate choice of initial guess controls is needed to improve the adaptive reduced 4D-Var solution (here, taken from the true flow field on the 107.5th day, see Section 6.3.2).

5.3. Error estimation and POD results

The total error of the inverted model results comprises: (a) the integration error of the POD reduced model and the projection error (including the interpolation error when adaptive meshes are adopted) and (b) the error introduced by the optimized controls (including the error of the POD reduced-order adjoint model).

5.3.1. Error estimation for the POD reduced-order model. The error of the POD reduced-order model is split into the projection error and the error from the integration in the subspace (see [51]). The two-norm of the projection error can be calculated by [51]

$$\|e_{\text{proj}}\|_{L_2}=\sqrt{\sum_{i=M+1}^{K}\lambda_i} \tag{49}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

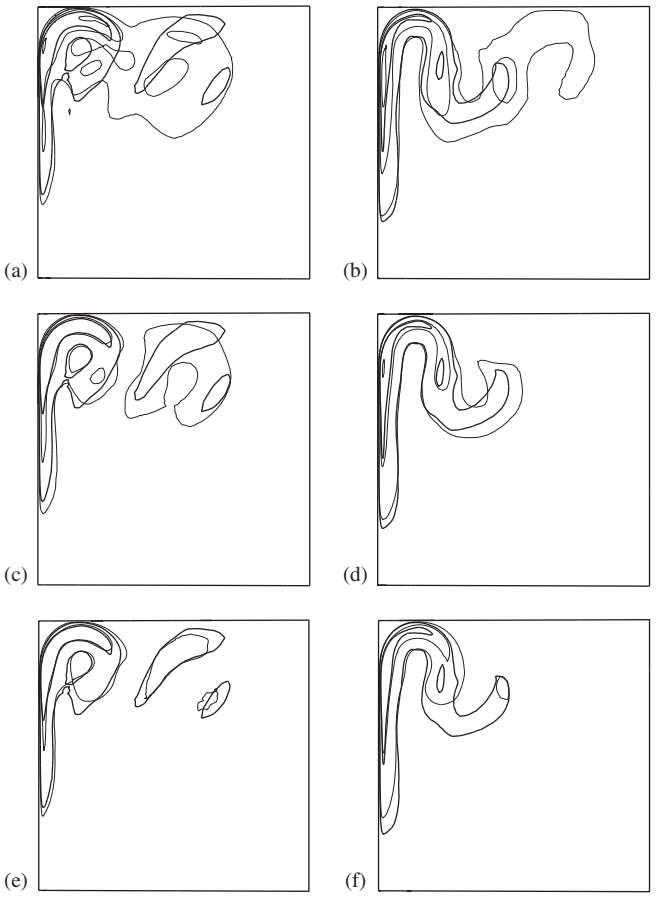

Figure 4. Contours of velocity at time levels $t=150$ days (left panel) and $t=200$ days (right panel). The thick lines represent the results from the full model while the thin lines from the POD reduced-order model with the use of the (a, b) linear mesh to mesh interpolation; (c, d) quadratic mesh to mesh interpolation; and (e, f) cubic mesh to mesh interpolation: (a) linear ($t=150$ days); (b) linear ($t=200$ days); (c) quadratic ($t=150$ days); (d) quadratic ($t=200$ days); (e) cubic ($t=150$ days); and (f) cubic ($t=200$ days).

where K is the number of snapshots and M is the number of POD bases (i.e. the subspace size). Here $M=35$, the two-norm of the projection error is therefore $\sqrt{\sum_{i=M+1}^{K}\lambda_i}=1.22, 1.3$ and 1.37 for the velocity components u, v and pressure p, respectively.

To isolate the error of the reduced model, the POD reduced model is driven by the true controls rather than the optimized ones. The results from the POD reduced model are compared with those from the full model. The root mean square error (RMSE) between the POD velocity solution and the true one at the time level n is used to estimate the error of the POD model:

$$\mathrm{RMSE}^n=\sqrt{\frac{\sum_{i=1}^{\mathcal{N}}(U_i^n-U_{0,i}^n)^2}{\mathcal{N}}} \tag{50}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

where U_i^n and $U_{o,i}^n$ are the vectors containing the POD velocity components and true ones at the node i, respectively, $\mathcal{N}$ is the total number of nodes over the domain. The RMSE during the simulation period is provided in Figure 2. The dashed line represents the total reduced model error (the integration error plus project error) when 41 snapshots with 35 POD bases are chosen, while the solid line shows the integration error isolated by eliminating the projection error (the number of POD bases being the same as that of snapshots). The total reduced model error (RMSE) remains small (less than $2\,\mathrm{m\,s^{-1}}$) when $t \leqslant 150$ days, and increases to $4\,\mathrm{m\,s^{-1}}$ at $t = 175$ days. Since adaptive meshes are employed in the full model, the RMSE shown in Figure 2 includes the interpolation error as well. The absolute error between the POD solution and true flow state over the domain at the different time levels ($t = 125, 150, 175$ days) is shown in Figure 3. The maximum error is less than $8\,\mathrm{m\,s^{-1}}$ during the first half simulation period and increases as the simulation time accrues.

When adaptive meshes are adopted in the original model, an interpolation error between the adaptive and reference meshes is introduced to snapshots and the POD reduced-order model. To reduce the interpolation error and improve the quality of snapshots, high-order interpolation schemes are employed. To evaluate the interpolation error, a comparison of POD results is carried out with linear and high-order (quadratic and cubic) interpolation schemes. Figure 4 shows the velocity contour at time levels ($t = 150$ and 200 days) where the thick and thin lines represent the solutions from the full and POD reduced-order models. It is shown that the accuracy of POD results is improved by introducing the high-order interpolation operators during the second half of the simulation period [150, 200] days where highly structured turbulent flows (eddies) develop. Compared with that, with the use of the linear interpolation, the RMSE of the velocity results (between the POD and full models) with the use of the quadratic and cubic interpolation schemes is reduced by half (the RMSE remains small—less than $2\,\mathrm{m\,s^{-1}}$) while the correlation of velocity results is increased to 95% (Figure 5).

5.3.2. Optimized results and error estimation. In this case, the guess values of the initial conditions (background flows) are taken from the true flow state on the 107.5th day. The POD reduced-order forward and adjoint models are used to optimize the initial conditions. The POD bases are updated once during the optimization procedure. The comparison between the optimized and true initial

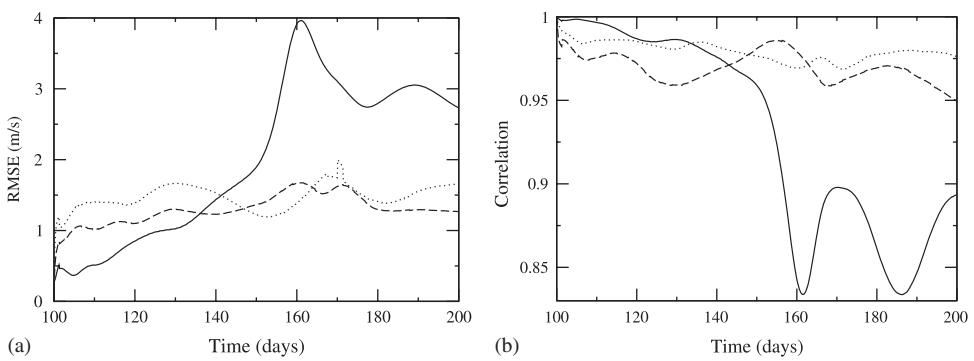

Figure 5. RMSE (a) and correlation coefficient (b) of velocity results between the POD model and the full model with the use of the linear (solid line), quadratic (dashed line) and cubic (dotted line) mesh to mesh interpolation schemes.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

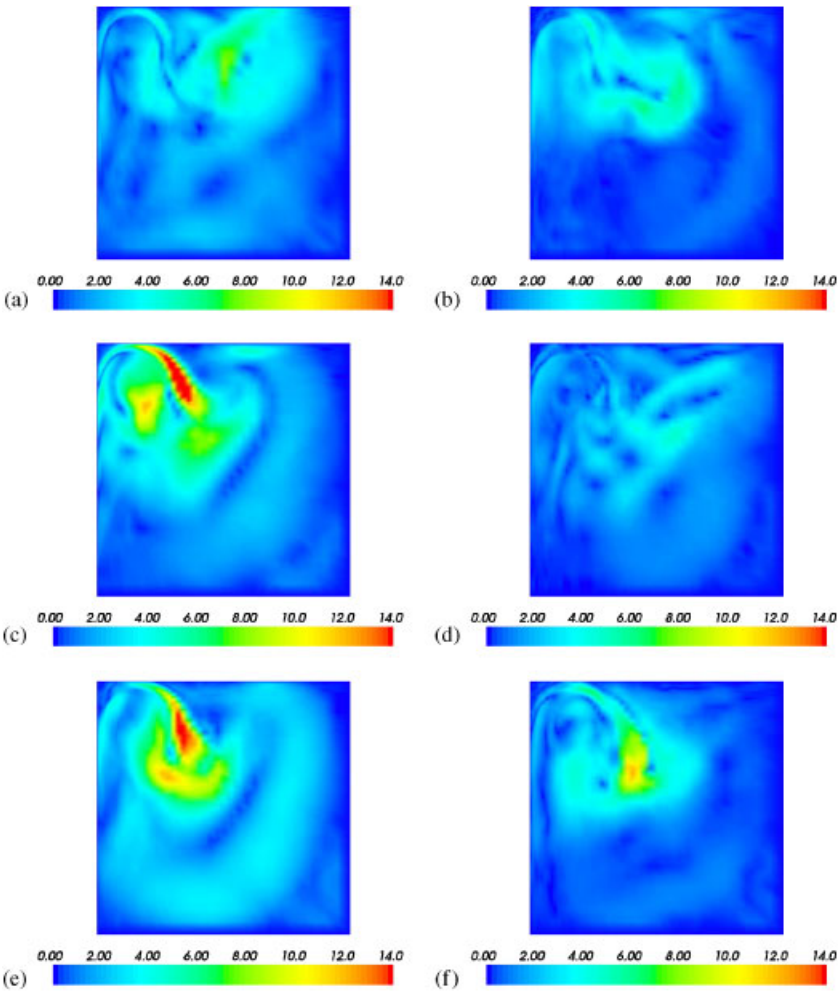

Figure 6. Error in the velocity of the POD forward model driven by the optimized initial conditions at time levels. Left panel: the first POD iteration; right panel: the second POD iteration (after updating the snapshots): (a) $t = 125$ days; (b) $t = 125$ days; (c) $t = 150$ days; (d) $t = 150$ days; (e) $t = 175$ days; and (f) $t = 175$ days.

velocity conditions is shown in Figure 7. The error between the inverted model velocity and the true value (Figure 6) decreases by 10–50% in the larger part of flow after updating the snapshots (right panel). The cost function (taking into account only the error introduced by the optimized controls) is reduced by 20% of its original value at the first POD inversion iteration. It is further reduced by 70% at the second adaptive POD iteration, i.e. after the snapshots are updated (Figure 7(d)).

The correlation cor defined below is also used to evaluate the quality of the inversion simulation

$$\mathrm{cor} = \frac{\mathrm{cov}_{12}}{\sigma_1 \sigma_2}$$

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

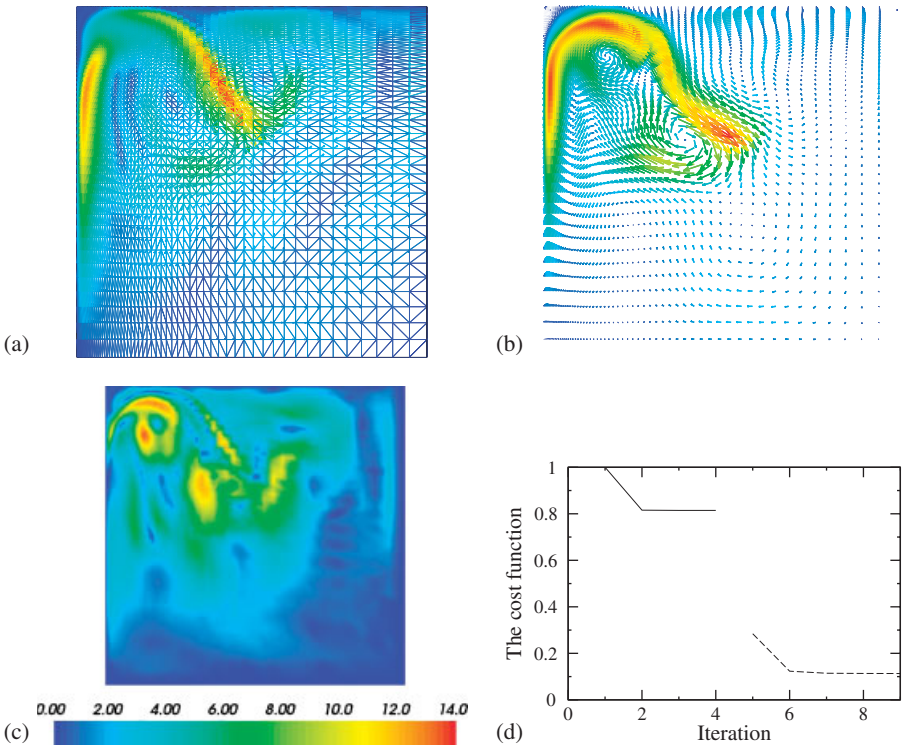

Figure 7. The comparison between the optimized and true initial velocity conditions: (a) the true initial conditions; (b) the optimized initial conditions; (c) the error between the optimized and true initial conditions; and (d) the cost functional (misfit between the optimized and true velocities (the solid line: the first adaptive POD iteration; the dashed line: the second adaptive POD iteration).

where

$$\sigma_1(\mathbf{x}) = \sum_{n=1}^{N_t} (U^n(\mathbf{x}) - \bar{U}(\mathbf{x}))^2, \quad \sigma_2(\mathbf{x}) = \sum_{n=1}^{N_t} (U_o^n(\mathbf{x}) - \bar{U}_o(\mathbf{x}))^2$$
$$\mathrm{cov}_{12}(\mathbf{x}) = \sum_{n=1}^{N_t} (U^n(\mathbf{x}) - \bar{U}(\mathbf{x}))(U_o(\mathbf{x})^n - \bar{U}_o(\mathbf{x})) \tag{51}$$

where U^n and U_o^n are the vectors containing the optimal and true velocity components (u, v) at the time level n over the domain, respectively, their respective means over the simulation period are $\bar{U}$ and $\bar{U}_o$, n is the time level and N_t is the total number of time levels, $\mathbf{x} = (x, y, z)$. The correlation between the true and modelled velocity in the case of running the POD model with the initial guess control (background flow) is low, mostly less than 0.5 over the domain (Figure 8(b)). It is improved after the initial conditions are optimized (Figure 8(c)), especially after updating the POD bases (at the second adaptive POD iteration, Figure 8(d)). The correlation between the inverted model and true velocity varies between 0.80 and 0.98 over the domain. The optimized velocity fields are drawn in Figure 9 and exhibit an overall good agreement with the true ones while the POD bases are updated only once for the entire optimization process.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

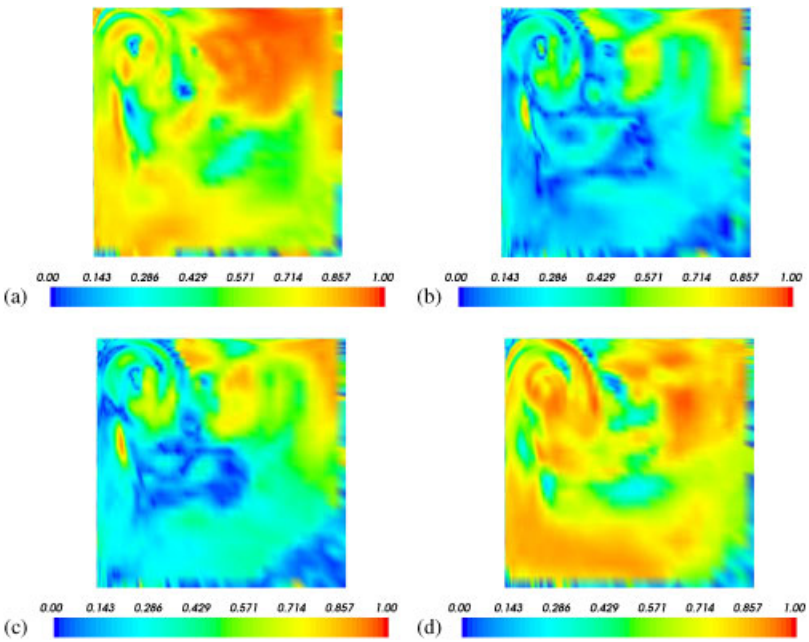

Figure 8. Correlation between the inverted model velocity and the true one. The inverted model velocity is obtained from the POD model driven by (a) the true control; (b) the initial guess controls; (c) the optimized controls at the first POD iteration; and (d) the optimized controls after the second POD iteration, i.e. after updating the POD bases.

5.4. Computational efficiency of reduced-order 4D-Var

In this test case, it takes 10 h to run the full model and 3 h to run the reduced-order model. Thus, running the reduced-order model results in a decrease of 70% in the CPU time. As a consequence, the computational time required for reduced 4D-Var is reduced by a factor of $3N$ (where N is the number of times needed to run the models until the optimality is satisfied). POD preconditions the minimization process, which results in less minimization iterations.

It is also noted that 99% of the CPU time required for the reduced-order model is used to calculate the discretized matrix. The numerical technique developed in [41] is adopted to accelerate the POD inversion, that is, the matrices in the discretized POD forward and adjoint equations can be constructed by sets of time-independent sub-matrices (see Equations (20) and (21)) prior to running the reduced forward and adjoint models. These sub-matrices remain the same until the POD bases are updated.

In total, the computer time required for the inverse simulation in this test case is given below:

- Running the full forward model to set up the snapshots and calculate the POD bases (10 h).
- Calculating the time-independent sub-matrices in preparation for running both the reduced order forward and adjoint models (2.5 h).
- Running the reduced-order forward and adjoint models during the optimization procedure (10 min, where the Polak–Ribière nonlinear CG approach for large-scale unconstrained minimization is employed).

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

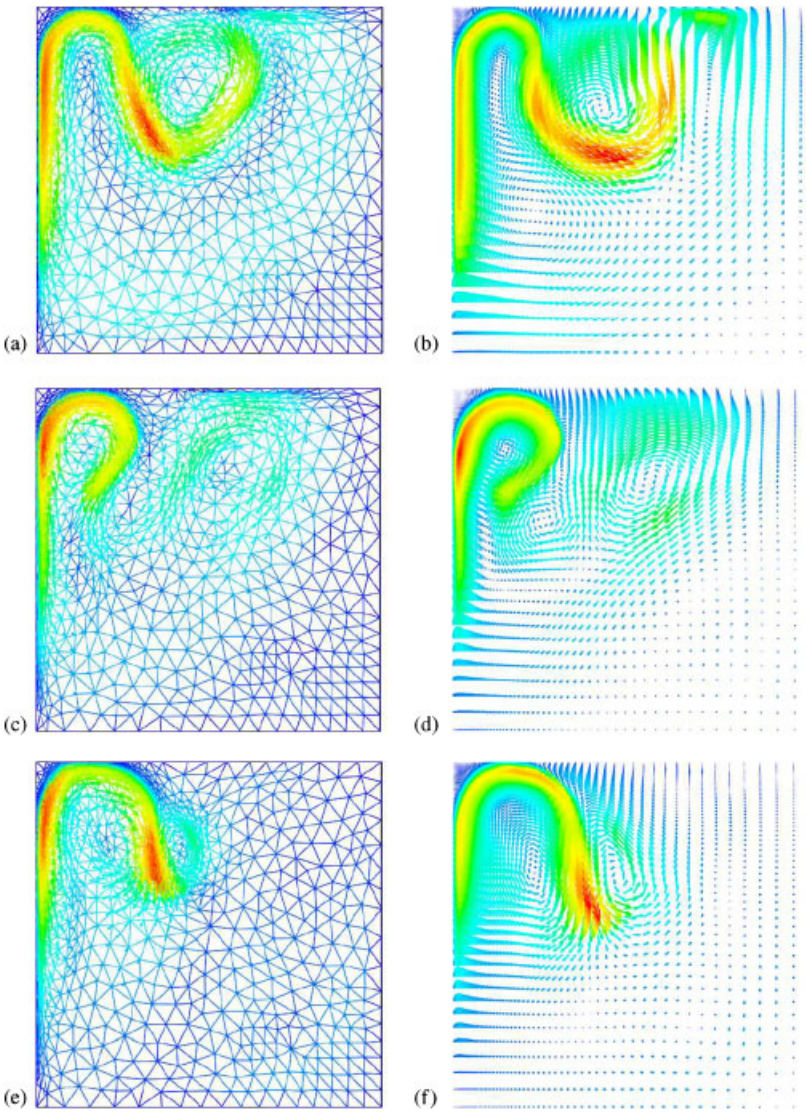

Figure 9. Comparison between the true velocity field and that from the POD reduced model (driven by the optimized initial conditions) at the time levels. Left panel: the true velocity field; right panel: the optimized velocity field: (a) $t=125$ days; (b) $t=125$ days; (c) $t=150$ days; (d) $t=150$ days; (e) $t=175$ days; and (f) $t=175$ days.

6. CONCLUSION

The development of the POD reduced-order 4D-Var model for an adaptive mesh, non-hydrostatic finite element ocean model is presented here. Using the POD and Galerkin projection approaches, the reduced-order forward model is derived in a reduced subspace (details in [41]). Once the

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

forward reduced-order model is available, the reduced-order adjoint model can be directly obtained from the POD reduced-order forward model in the subspace, instead of the original forward model. The minimization procedure is then carried out in the reduced space. The matrix for the discretized forward and adjoint models is constructed by a series of time-independent sub-matrices which remain unchanged until the POD bases are updated [41]. The reduced-order forward and adjoint models can thus be run repeatedly with negligible computational cost.

The performance of the POD 4D-Var model is demonstrated by inverting for the initial conditions of a wind driven gyre in an idealized geometry. The correlation between the inverted model and true velocity is of the order of 80–98% over the majority of the domain. The cost function (taking into account only the error introduced by the optimized controls) is reduced by 20% of its original value at the first POD inversion iteration. It is further reduced by 70% after the snapshots are updated.

In general, the advantages of the POD reduced-order inverse model developed here over existing POD approaches consist of the ability:

- to implement the reduced-order adjoint model from the discretized forward model easily;
- to use dynamically adaptive meshes in the reduced-order 4D-Var;
- to significantly reduce the computation cost for 4D-Var by carrying out the optimization in the reduced space;
- to accurately represent the geostrophic balance by two sets of POD bases for the velocity components u and v.

ACKNOWLEDGEMENTS

This work was carried out under funding from the UK's Natural Environment Research Council (projects NER/A/S/2003/00595, NE/C52101X/1 and NE/C51829X/1), the Engineering and Physical Sciences Research Council (GR/R60898) and the Leverhulme Trust (F/07058/AB), and with support from the Imperial College High Performance Computing Service and the Grantham Institute for Climate Change. Prof. I. M. Navon would like to acknowledge the support of NSF grants ATM-0201808 and CCF-0635162. Finally, the authors would like to thank the anonymous reviewer who assisted in substantially improving this paper.

REFERENCES

1. Lardner RW. Optimal control of open boundary conditions for a numerical tidal model. *Computer Methods in Applied Mechanics and Engineering* 1993; **102**:367–387.
2. Bennett AF, Chua BS, Harrison DE, McPhaden MJ. Generalised inversion of Tropical Atmosphere-Ocean (TAO) data and a coupled model of the tropical Pacific. Part II: the 1995–1996 La Nina and 1997–1998 El Nino. *Journal of Climate* 2000; **13**(15):2770–2785.
3. Leredde Y, Devenon JL, Dekeyser I. Turbulent viscosity optimised by data assimilation. *Annales Geophysicae* 1999; **17**:1463–1477.
4. Moore AM. Data assimilation in a quasi-geostrophic open-ocean model of the Gulf Stream region using the adjoint method. *Journal of Physical Oceanography* 1991; **21**(3):398–427.
5. Thacker WC, Long RB. Fitting dynamics to data. *Journal of Geophysical Research* 1988; **93**(C2):1227–1240.
6. Tziperman E, Thacker WC, Long RB, Hwang SM, Rintoul SR. Oceanic data analysis using a general circulation model. Part II: a North Atlantic model. *Journal of Physical Oceanography* 1992; **22**(12):1458–1485.
7. Zhu J, Kamachi M, Wang D. Estimation of air–sea heat flux from ocean measurements: an ill-posed problem. *Journal of Geophysical Research* 2002; **107**(10):3159.
8. Zou X, Kuo YH. Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Monthly Weather Review* 1996; **124**(12):2859–2882.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

9. Gunzburger MD. *Perspectives in Flow Control and Optimization*. SIAM: New York, 2003.
10. Talagrand O, Courtier P. Variational assimilation of meteorological observations with the adjoint vorticity equation (I). Theory. *Quarterly Journal of the Royal Meteorological Society* 1987; **113**(478):1311–1328.
11. Farrell BF, Moore AM. An adjoint method for obtaining the most rapidly growing perturbation to oceanic flows. *Journal of Physical Oceanography* 1992; **22**(4):338–349.
12. Cacuci DG, Bujor MI, Navon IM. *Sensitivity and Uncertainty Analysis*: *Applications to Large-scale Systems*, vol. 2. CRC: Boca Raton, FL, 2005.
13. Moore AM, Cooper NS, Anderson DLT. Initialisation and data assimilation in models of the Indian Ocean. *Journal of Physical Oceanography* 1987; **17**:1965–1977.
14. Daescu DN, Navon IM. Adaptive observations in the context of 4D-Var data assimilation. *Meteorology and Atmospheric Physics* 2004; **85**(4):205–226.
15. Le Dimet FX, Navon IM. Variational and optimization methods in meteorology: a review. *Technical Report*: *Early Review on Variational Data Assimilation*, *SCRI Report No. 144*, 1998; 88. Available from: http://people.scs.fsu.edu/navon/freqreq.html.
16. Zhu YQ, Navon IM. Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint. *Monthly Weather Review* 1999; **127**(7):1497–1517.
17. Courtier P, Thepaut JN, Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* 1994; **120**(519):1367–1387.
18. Derber J, Rosati A. A global oceanic data assimilation system. *Meteorology and Atmospheric Physics* 2003; **85**(4):205–226.
19. Weaver AT, Vialard J, Anderson DLT, Delecluse P. Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific Ocean. Part II: physical validation. *Monthly Weather Review* 2003; **131**:1379–1395.
20. Wenzel M, Schröter J, Olbers D. The annual cycle of the global ocean circulation as determined by 4D VAR data assimilation. *Progress in Oceanography* 2001; **48**(1):73–119.
21. Courtier P, Thepaut JN, Hollingsworth A. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* 1994; **120**:1367–1368.
22. Robert C, Durbiano S, Blayo E, Verron J, Blum J, le Dimet FX. A reduced-order strategy for 4D-VAR data assimilation. *Journal of Marine Systems* 2005; **57**(1–2):70–82.
23. Tremolet Y. Incremental 4D-Var convergence study. *Tellus Series A Dynamic Meteorology and Oceanography* 2007; **59**(5):706–718.
24. Tremolet Y. Diagnostics of linear and incremental approximations in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, *Part B* 2004; **130**(601):2233–2251.
25. Blayo E, Blum J, Verron J. *Assimilation Variationnelle de Donnees en Oceanography et reduction de la dimension de l'espace de controle In Equations aux Derivees partielles et Applications*. Gauthier-Villars: Paris, 1998.
26. Hoteit I, Kohl A. Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. *Journal of Oceanography* 2006; **62**(4):539–550.
27. Robert C, Blayo E, Verron J, Blum J, le Dimet FX. Reduced-order 4d-var: a preconditioner for the incremental 4d-var data assimilation method. *Geophysical Research Letters* 2006; **33**(L18609):1–4.
28. Vermeulen PTM, Heemink AW. Model-reduced variational data assimilation. *Monthly Weather Review* 2006; **134**:2888–2899.
29. Holmes P, Lumley JL, Berkooz G. *Turbulence*, *Coherent Structures*, *Dynamical Systems and Symmetry*. Cambridge University Press: Cambridge, U.K., 1998.
30. Luo Z, Chen J, Zhu J, Wang R, Navon IM. An optimising reduced order FDS for the tropical Pacific Ocean reduced gravity model. *International Journal for Numerical Methods in Fluids* 2007; **55**:143–161.
31. Luo Z, Zhu J, Wang R, Navon IM. 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* 2007; **196**(41–44):4184–4195.
32. Fukunaga K. *Introduction to Statistical Recognition* (2nd edn). Computer Science and Scientific Computing Series. Academic Press: Boston, 1990.
33. Aubry N, Holmes P, Lumley JL. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *Journal of Fluid Dynamics* 1988; **192**:115–173.
34. Willcox K, Peraire J. Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal* 2002; **40**(11):2323–2330.
35. Kirby M, Sirovich L. Application of the Karhunen–Loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1990; **12**(1):103–108.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/fld

36. Willcox K, Ghattas O, van Bloemen Waanders B, Bader B. An optimization framework for goal-oriented, model-based reduction of large-scale systems. *44th IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, December 2005.
37. Bui-Thanh T, Willcox K, Ghattas O, van Bloemen Waanders B. Goal-oriented, model-constrained optimization for reduction of large-scale systems. *Journal of Computational Physics* 2007; **224**(2):880–896.
38. Cao Y, Zhu J, Navon IM, Luo Z. A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids* 2007; **53**(10):1571–1583.
39. Daescu DN, Navon IM. A dual-weighted approach to order reduction in 4d-var data assimilation. *Monthly Weather Review* 2008; **136**(3):1026–1041.
40. Daescu DN, Navon IM. Efficiency of a POD-based reduced second order adjoint model in 4-D VAR data assimilation. *International Journal for Numerical Methods in Fluids* 2007; **53**:985–1004.
41. Fang F, Pain CC, Navon IM, Piggott MD, Gorman GJ, Allison P, Goddard AJH. Reduced order modelling of an adaptive mesh ocean model. *International Journal for Numerical Methods in Fluids* 2008; DOI: 10.1002/fld.1841.
42. Ravindran SS. Adaptive reduced-order controllers for a thermal flow system using proper orthogonal decomposition. *SIAM Journal on Scientific Computing* 2002; **23**(6):1924–1942.
43. Ravindran SS. Reduced-order controllers for control of flow past an airfoil. *International Journal for Numerical Methods in Fluids* 2002; **50**(5):531–554.
44. Fahl M. Trust-region method for flow control based on reduced order modelling. *Thesis*, dem Fachbereich IV der Universitat Trier, 2001; 140. Available from: http://ub-dok.uni-trier.de/diss/diss55/20010209/20010209.htm.
45. Ford R, Pain CC, Piggott MD, Goddard AJH, de Oliveira CRE, Umbleby AP. A nonhydrostatic finite-element model for three-dimensional stratified oceanic flows. Part I: model formulation. *Monthly Weather Review* 2004; **132**(12):2816–2831.
46. Piggott MD, Gorman GJ, Pain CC, Allison PA, Candy AS, Martin BT, Wells MR. A new computational framework for multi-scale ocean modelling based on adapting unstructured meshes. *International Journal for Numerical Methods in Fluids* 2008; **56**:1003–1015. DOI: 10.1002/fld.1663.
47. Pain CC, Umpleby AP, de Oliveira CRE, Goddard AJH. Tetrahedral mesh optimisation and adaptivity for steady-state and transient finite element calculations. *Computer Methods in Applied Mechanics and Engineering* 2001; **190**:3771–3796.
48. Piggott MD, Pain CC, Gorman GJ, Power PW, Goddard AJH. h, r, and hr adaptivity with applications in numerical ocean modelling. *Ocean Modelling* 2005; **10**(1–2):95–113.
49. Freitag LA, Ollivier-Gooch C. Tetrahedral mesh improvement using swapping and smoothing. *International Journal for Numerical Methods in Engineering* 1997; **40**:3979–4002.
50. Buscaglia GC, Dari EA. Anisotropic mesh optimization and its application in adaptivity. *International Journal for Numerical Methods in Engineering* 1997; **40**:4119–4136.
51. Homescu C, Petzold LR, Serban R. Error estimation for reduced-order models of dynamical systems. *SIAM Review* 2007; **49**(2):277–299.

Copyright © 2008 John Wiley & Sons, Ltd.