INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2011; **65**:520–541
Published online 22 December 2009 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/fld.2198

A dual-weighted trust-region adaptive POD 4D-VAR applied to a finite-element shallow-water equations model

X. Chen[1], I. M. Navon[*,†,2] and F. Fang[3]

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

SUMMARY

We consider a limited-area finite-element discretization of the shallow-water equations model. Our purpose in this paper is to solve an inverse problem for the above model controlling its initial conditions in presence of observations being assimilated in a time interval (window of assimilation). We then attempt to obtain a reduced-order model of the above inverse problem, based on proper orthogonal decomposition (POD), referred to as POD 4-D VAR. Different approaches of POD implementation of the reduced inverse problem are compared, including a dual-weighed method for snapshot selection coupled with a trust-region POD approach. Numerical results obtained point to an improved accuracy in all metrics tested when dual-weighing choice of snapshots is combined with POD adaptivity of the trust-region type. Results of *ad-hoc* adaptivity of the POD 4-D VAR turn out to yield less accurate results than trust-region POD when compared with high-fidelity model. Directions of future research are finally outlined. Copyright © 2009 John Wiley & Sons, Ltd.

Received 21 May 2009; Revised 26 August 2009; Accepted 27 August 2009

KEY WORDS: proper orthogonal decomposition; 4-D VAR; shallow water equations; dual weighting; trust-region method; inverse problem

1. INTRODUCTION

We address here the proper orthogonal decomposition (POD) model reduction along with inverse solution of a two-dimensional finite-element (FE) shallow-water equations model on a limited area domain. The shallow-water equations are frequently used to simulate the earth's atmosphere, which can be thought of as a thin (practically zero in height), semi-incompressible fluid that is flowing over the surface of a rotating globe (the earth). The shallow-water equations are the simplest form of the equations of motion that show how the fluid flow will evolve in response to rotational and gravitational accelerations of the earth, forming waves. Although there is a body of experience using POD model reduction for the shallow-water equations as well as for POD applied to 4-D VAR data assimilation of the shallow-water equations our intention is to draw on state-of-the-art methodologies for efficient POD implementation, i.e. combining efficient snapshot selection in the presence of data assimilation system namely merging dual weighting of snapshots with trust-region POD techniques.

*Correspondence to: I. M. Navon, Department of Scientific Computing, Florida State University, Tallahassee, FL, 32306-4120, U.S.A.
†E-mail: inavon@fsu.edu

Contract/grant sponsor: NSF; contract/grant numbers: ATM-0201808, CCF-0635162
Contract/grant sponsor: Natural Environment Research Council; contract/grant number: NE/C51829X/1

The trust-region proper orthogonal decomposition (TR-POD) was recently proposed in [1, 2] as a way to overcome difficulties related POD reduced-order model (ROM) use in solving the partial differential equation (PDE) constrained optimization problem. Combining POD technique with the concept of trust-region with general model functions (see Toint and Conn [3, 4] for a comprehensive survey or Nocedal and Wright [5] for an introduction to trust-region methods) presents a framework for decision as to when an update of the POD ROM is necessary during the optimization process. Moreover, from a theoretical point of view for TR-POD, we have a global convergence result [1] proving that the iterates produced by the optimization algorithm, started at an arbitrary initial iterate, will converge to a local optimizer for the original mode.

The novelty of this contribution is in assessing the combined effect of use of TR-POD in conjunction with dual weighting data assimilation system (DAS) snapshot selection in the framework of a relatively affordable, yet relevant model. One expects a beneficial cumulative effect from the combination of these two techniques. Comparisons to *ad-hoc* update adaptivity of the POD 4-D VAR and full 4-D VAR (high-fidelity model) are carried out for a variety of metrics to validate theoretical results in the light of numerical experiments. Indeed the combination of TR-POD and dual-weighted snapshots yields the best results in all metrics (see [6, 7]).

The plan of the paper is as follows: After the introduction we present in Section 2 the shallow-water equations model description followed by a brief presentation of the POD model reduction method. Section 3 provides the description of the generation of POD using a FE formulation. Section 4 provides the framework of POD for reduced-order 4-D VAR data assimilation and is composed of several subsections detailing dual weighting of snapshots and implementation of reduced order 4-D VAR. Section 5 addresses in some detail the trust-region POD (TR-POD) methodology. Section 6 details the numerical experiments carried out in order to validate accuracy of the POD reduced-order model and the POD 4-D VAR approach for the various numerical methods enumerated above. For recent work on POD 4-D VAR, see [8–14]. In particular we compare ad-hoc adaptivity for POD 4-D VAR with trust-region adaptivity in combination with dual-weighted snapshots. Section 7 provides error analysis of dual-weighted trust-region POD 4-D VAR compared with the high-fidelity model. A discussion of numerical results thus obtained ensues. Finally, the paper concludes with a conclusion section.

2. DESCRIPTION OF PROBLEMS

2.1 Shallow-water equations model on an f plane

The shallow-water equations (also called Saint Venant equations) are a set of hyperbolic PDEs that describe the evolution of an incompressible fluid (sometimes, but not necessarily, a free surface) in response to gravitational and rotational accelerations (see [15–17]).

The equations are derived from depth-integrating the Navier–Stokes equations, in the case where the horizontal length scale is much greater than the vertical length scale. Under this condition, conservation of mass implies that the vertical velocity of the fluid is small. It can be shown from the momentum equation that vertical pressure gradients are nearly hydrostatic, and that horizontal pressure gradients are due to the displacement of the pressure surface, implying that the velocity field is nearly constant throughout the depth of the fluid. Taking the vertical velocity and variations throughout the depth of the fluid to be exactly zero in the Navier–Stokes equations, the shallow-water equations are derived.

The shallow-water equations on an f plane can be written as (see [18]):

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+\frac{\partial \phi}{\partial x}-fv=0$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+\frac{\partial \phi}{\partial y}+fu=0$$

DOI: 10.1002/fld

$$\frac{\partial \phi}{\partial t}+u\frac{\partial \phi}{\partial x}+v\frac{\partial \phi}{\partial y}+\phi\frac{\partial u}{\partial x}+\phi\frac{\partial v}{\partial y}=0 \qquad (x,y)\in[0,L]\times[0,D], t\geqslant 0 \tag{1}$$

where L and D are the dimensions of a rectangular domain of integration, u and v are the velocity components in the x- and y-axis, respectively, $\phi=gh$ is the geopotential height, h is the depth of the fluid, and g is the acceleration of gravity.

The scalar function f is the Coriolis parameter. An approximation whereby the Coriolis parameter f is set to vary linearly in space is called a β-plane approximation with $\beta=\mathrm{d}f/\mathrm{d}y$ defined as the linear coefficient of variation.

$$f=\hat{f}+\beta\left(y-\frac{D}{2}\right) \tag{2}$$

The Coriolis frequency $\hat{f}$ is defined as

$$\hat{f}=2\Omega\sin\theta \tag{3}$$

where Ω is the angular velocity of the earth's rotation and θ is the latitude.

A well-posed initial and boundary condition [19] for the shallow-water equations model above can be imposed as

$$w(x,y,0)=\varphi(x,y) \tag{4}$$

where state variables are $w=w(x,y,t)=(\vec{v}(x,y,t),\phi(x,y,t))$ with periodic boundary conditions assumed in the x-direction:

$$w(0,y,t)=w(L,y,t) \tag{5}$$

whereas solid wall boundary conditions are used in the y-direction:

$$v(x,0,t)=v(x,D,t)=0 \tag{6}$$

The geopotential $\varphi(x,y)$ will be specified later in the numerical experiments.

3. PROPER ORTHOGONAL DECOMPOSITION

The POD identifies basis functions or modes that optimally captures the average energy content from numerical or experimental data. POD was introduced in the context of analysis of turbulent flow by Lumley [20], Berkooz *et al.* [21]. Sirovich *et al.* [22] introduced the idea of snapshots. See also the book of Holmes *et al.* [23].

Let Ω be a bounded domain in $\mathbb{R}^n$, the $L_2(\Omega)$ is defined as

$$L_2(\Omega)=\left\{f(x), x\in\Omega: \int_\Omega f^2\,\mathrm{d}\Omega<\infty\right\} \tag{7}$$

with inner product

$$\langle f,g\rangle=\int_\Omega fg\,\mathrm{d}\Omega \quad \forall f,g\in L_2(\Omega) \tag{8}$$

Given a set of sampled data

$$\mathbf{Y}^h=\{y^{h,1},y^{h,2},\ldots,y^{h,n}\} \tag{9}$$

where $y^{h,i}\in L_2(\Omega)$ and $V=\mathrm{span}(\mathbf{Y}^h)\subseteq\mathbb{R}^n$.

Let $\mathbf{K}$ be the correlation matrix of the data defined by

$$\mathbf{K}=\mathbf{Y^h}(\mathbf{Y^h})^{\mathrm{T}} \tag{10}$$

where $\mathbf{K}=(k_{ij})_{n\times n}$, $k_{ij}=\langle y^{h,i},y^{h,j}\rangle$, $i,j=1,\ldots,n$.

DOI: 10.1002/fld

Then from all the subspaces $V_M \subset V$ with a fixed dimension $M = \dim(V_M) < \dim(V)$,

$$\min_{V_M} \|\mathbf{Y} - \Pi_M \mathbf{Y}\| = \sum_{i=M+1}^{n} \lambda_i \tag{11}$$

where $\{\lambda_i\}_{i=1}^n$ are the non-negative ordered eigenvalues of symmetric matrix $\mathbf{K}$ and $\mathbf{\Psi^h} = \{\psi_i^h\}_{i=1}^n$ are the corresponding eigenvectors.

Such that

$$\langle \psi^{h,i}, \psi^{h,j} \rangle = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{12}$$

Thus, the optimal subspace is given by

$$V_M = \text{span}\{\psi_1^h, \psi_2^h, \ldots, \psi_M^h\} \tag{13}$$

in the sense that such type of POD bases describes more energy on average of the ensemble than any other linear basis of the same dimension, with optimal orthogonal projection $\Pi_M : V \to V_M$, where $\Pi_M^2 = \Pi_M$ given by

$$\Pi_M = \sum_{i=1}^{M} \psi_i^h (\psi_i^h)^{\mathrm{T}} \tag{14}$$

Historically, in other disciplines, the same procedure goes by the names of Karhunen–Loeve decomposition (KLD) (see [24, 25]) or principal components analysis (PCA) and before them it was discovered by Kosambi [26]. Historically, the method originated in the work of Pearson [27], who invented the PCA which involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. It was also put forward in statistical framework by Hotelling [28].

4. GENERATION OF POD USING FINITE-ELEMENT FORMULATION

The POD bases are applied with the Galerkin weak-form FE method to create a reduced-order numerical model with reduced computational cost. It is well known that under some circumstances, Galerkin projections can produce unstable equilibrium points and limit cycles where the full system possesses stable equilibrium points and limit cycles. If energy-based inner product is used, then Galerkin projection preserves the stability of an equilibrium point at the origin [29, 30]. Snapshots bases consist of the flow solution for several flow solutions corresponding to different sets of parameter values evaluated at different time instants of the model evolution. This involves solving the fully discretized model and saving states at various time instants in the time interval under consideration [23].

An ensemble of nodal-value represented snapshots chosen in the analysis time interval $[0, T]$ can be written as

$$\{y^1, y^2, \ldots, y^n\} \tag{15}$$

where $y^i \in \mathbb{R}^N$, $i = 1, \ldots, n$, n is the number of snapshots and N is the number of nodes in the mesh.

Define the weighted ensemble average of the FE represented data as

$$\bar{y} = \sum_{i=1}^{i=n} w_i y^i \tag{16}$$

where the snapshots weights w_i are such that $0 < w_i < 1$ and $\sum_{i=1}^n w_i = 1$, and they are used to assign a degree of importance to each member of the ensemble. Time weighting is usually considered, and in the standard approach $w_i = 1/n$.

DOI: 10.1002/fld

Hence, the FE represented POD solution can be expressed as

$$y^{\mathrm{POD}} = \bar{y} + \sum_{i=1}^{i=M} \alpha_i(t)\psi_i \tag{17}$$

where

$$\mathbf{\Psi} = \{\psi^1, \psi^2, \ldots, \psi^M\} \tag{18}$$

The nodal-value represented POD bases vectors $\mathbf{\Psi}$ and number of POD basis M are judiciously chosen to capture the dynamics of the flow as follows in the procedure described below:

1. The first step in creating a POD basis is to obtain a set of possible solution fields over the domain of the given problem. These fields will be generated through FE analysis as described above, and are referred to as snapshots. The snapshot selection is crucial to the generalization capabilities of the POD basis, and a strategy to create the set of snapshots is vital.
2. Compute mean value of snapshots

$$\bar{y} = \sum_{i=1}^{i=n} w_i y^i \tag{19}$$

3. Subtract the mean from each snapshot and we obtain

$$\mathbf{Y} = \{y^1 - \bar{y}, y^2 - \bar{y}, \ldots, y^n - \bar{y}\} \tag{20}$$

4. Denote the FE basis [31] by

$$[\mathbf{V}] = [V_1, \ldots, V_{\mathrm{n}}] \tag{21}$$

Compute the symmetric positive definite matrix

$$\mathbf{A} = \mathbf{V}^{\mathrm{T}}\mathbf{V} \tag{22}$$

and introduce a general form of inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{y} \tag{23}$$

The POD basis of order $M \leqslant n$ provides an optimal representation of the ensemble data in M-dimensional state subspace by minimizing the averaged projection error

$$\begin{aligned} &\min_{\{\psi^1, \psi^2, \ldots, \psi^M\}} \sum_{i=1}^{n} w_i \| (y^i - \bar{y}) - \Pi_{\Psi, M}(y^i - \bar{y}) \|_{\mathbf{A}}^2 \\ &\text{s.t. } \langle \psi^i, \psi^j \rangle_{\mathbf{A}} = \delta_{ij} \end{aligned} \tag{24}$$

where $\Pi_{\Psi, M}$ is the projection operator onto the M-dimensional space Span $\{\psi^1, \psi^2, \ldots, \psi^M\}$

$$\Pi_{\Psi, M} = \sum_{i=1}^{M} \langle y, \psi_i \rangle_{\mathbf{A}} \psi_i$$

5. Build the weighted spatial correlation matrix

$$\mathbf{C} = \mathbf{Y}\mathbf{W}\mathbf{Y}^{\mathrm{T}} \tag{25}$$

The POD modes $\psi^i \in \mathbb{R}^N$ are eigenvectors to the N-dimensional eigenvalue problem

$$\mathbf{C}\mathbf{A}\psi_i = \lambda_i \psi_i$$

As in practice the number of snapshots is much less than the the state dimension, $n \ll N$, an efficient way to compute the reduced basis is to introduce a n-dimensional matrix as follows

$$\mathbf{K}^{n \times n} = \mathbf{W}^{1/2}\mathbf{K}\mathbf{W}^{1/2} = \mathbf{W}^{1/2}\mathbf{Y}^{\mathrm{T}}\mathbf{A}\mathbf{Y}\mathbf{W}^{1/2} \tag{26}$$

DOI: 10.1002/fld

and compute the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ of $\mathbf{K}^{n\times n}$ with its corresponding eigenvectors $\xi_1, \ldots, \xi_n$

6. The nodal-value represented POD basis vectors are obtained by defining

$$\psi_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{Y}\mathbf{W}^{1/2}\xi_i, \quad i = 1, \ldots, M \tag{27}$$

and corresponding FE represented continuous POD basis can be expressed as

$$\{\psi^{h,1}, \psi^{h,2}, \ldots, \psi^{h,M}\} = \{\mathbf{V}\psi^1, \mathbf{V}\psi^2, \ldots, \mathbf{V}\psi^M\} \tag{28}$$

where

$$\langle \psi^{h,i}, \psi^{h,j} \rangle = \langle \psi^i, \psi^j \rangle_{\mathbf{A}} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{29}$$

One can define a relative information content to choose a low-dimensional basis of size $M \ll n$ by neglecting modes corresponding to the small eigenvalues. We define

$$I(m) = \frac{\sum_{i=1}^{i=m} \lambda_i}{\sum_{i=1}^{i=n} \lambda_i} \tag{30}$$

and choose M such that

$$M = \arg\min\{I(m) : I(m) > \gamma\} \tag{31}$$

where $0 \leqslant \gamma \leqslant 1$ is the percentage of total information by the reduced space and the tolerance γ must be chosen to be close to the value one in order to capture most of the energy of the snapshots basis.

For an atmospheric or oceanic flow defined in time interval $[0, T]$

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}t} &= F(y, t) \\ y(x, 0) &= y_0(x) \end{aligned} \tag{32}$$

To obtain a reduced model, we can first solve the PDE to obtain an ensemble of snapshots, then use a Galerkin projection scheme of the model equations onto the space spanned by the POD basis elements. We obtain the system of ODE as follows:

$$\frac{\mathrm{d}\alpha_i}{\mathrm{d}t} = \left\langle F\left(\bar{y}^h + \sum_{i=1}^{i=M} \alpha_i \psi_i^h, t\right), \psi_i^h \right\rangle \tag{33}$$

along with the initial conditions:

$$\alpha_i(0) = \langle y^h(x, 0) - \bar{y}^h, \psi_i^h \rangle = \langle y_0 - \bar{y}, \psi_i \rangle_{\mathbf{A}}, \quad i = 1, \ldots, M \tag{34}$$

5. POD FRAMEWORK TO REDUCED-ORDER 4-D VAR DATA ASSIMILATION

5.1. The dual-weighted POD basis

The aim of 4-D VAR data assimilation is that of fusing observational data and model predictions to obtain an optimal representation of the state of the atmosphere. In the full nonlinear 4-D VAR [32], this process is implemented by minimizing the cost functional in the following:

$$J(y_0) = \frac{1}{2}(y_0 - y^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(y_0 - y^{\mathrm{b}}) + \frac{1}{2}\sum_{k=0}^{k=n}(H_k y_k - y_k^{\mathrm{o}})^{\mathrm{T}}\mathbf{R}_k^{-1}(H_k y_k - y_k^{\mathrm{o}}) \tag{35}$$

where y^{b} is the background prior state estimation and $\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observational error covariance matrix, H is the observation operator, y_0 is a vector

DOI: 10.1002/fld

containing control variables such as initial conditions, y_k is a vector containing the solution of variables from the model at the time level k, y_k^o is the observation at time level k, and n is the number of time levels.

The snapshots are essentially a set of instantaneous flow solutions, obtained from experimental data or from a CFD simulation. They are then used to compute the POD basis vectors to yield an optimal representation of the data so that for any given basis vector size, the L_2 norm of the error between the original and reconstructed snapshot is minimized.

The construction of POD basis vectors depends not only on the features of model dynamics itself, but it also requires to properly account for the features from the 4-D VAR data assimilation. Furthermore, these two features may be quite different from each other. A recent method to avoid this problem is referred to as optimality system POD [33]. By implementing a dual-weighted proper orthogonal decomposition (DWPOD) method [7], we can incorporate the information from the 4-D VAR into the POD reduced-order modeling.

The specification of dual weights w_k associated with the snapshots may have a significant impact on which modes are selected as dominant and thus included into the POD basis. The dual-weighted approach makes use of the time-varying sensitivities of the 4-D VAR cost functional with respect to perturbations in the state at each time level where the snapshots are taken.

Assuming the cost functional $J(y(t))$ is defined explicitly in terms of each state $y(t)$ at time step t. For any fixed time step $\tau<t$, the model can be written as

$$\forall \tau<t, \quad y(t)=M_{\tau\to t}(y(\tau))=M_{\tau,t}(y(\tau)) \tag{36}$$

such that implicitly, the cost functional J can be viewed as a function of the previous state $y(\tau)$, to first-order approximation, the impact of small errors/perturbations δy_i in the state error at a snapshot time $t_i \leqslant t$ on J may be estimated using the tangent linear model $\mathbf{M}(t_i, t)$ and its adjoint model $\mathbf{M}^{\mathrm{T}}(t, t_i)$:

$$\begin{aligned}\delta J &\approx \langle \nabla J_{y(t)}(y(t)), \delta y(t)\rangle = \langle \nabla J_{y(t)}(y(t))\mathbf{M}(t_i, t)\delta y(t_i)\rangle \\ &= \langle \mathbf{M}^{\mathrm{T}}(t, t_i)\nabla J_{y(t)}(y(t)), \delta y(t_i)\rangle = \langle y_{t_i}^*, \delta y(t_i)\rangle\end{aligned} \tag{37}$$

where $y_{t_i}^* = \mathbf{M}^{\mathrm{T}}(t, t_i)\nabla J_{y(t)}(y(t))$ are the adjoint variables at time step t_i.

In particular, the model can be written as

$$\forall k, \quad y_k = M_{k-1\to k}(y_{k-1}) = M_k(y_{k-1}) \tag{38}$$

where $M_{k-1\to k}$ is defined as the model forecast operator from time $k-1$ to k.

In order to derive the algorithm for the computation of dual weighs by using the adjoint model, we explicitly choose $\tau = t_i = k-1$ and $t = k$, to the first-order approximation, the impact of perturbations δy_{k-1} in state vectors on cost functional J_k may be estimated using tangent linear model $\mathbf{M}_k$ and its adjoint model $\mathbf{M}_k^{\mathrm{T}}$:

$$\delta J_k \approx \langle \nabla J_k, \delta y_k\rangle = \langle \nabla J_k, \mathbf{M}_k \delta y_{k-1}\rangle = \langle \mathbf{M}_k^{\mathrm{T}} \nabla J_k, \delta y_{k-1}\rangle = \langle y_{k-1}^*, \delta y_{k-1}\rangle \tag{39}$$

where $y_{k-1}^* = \mathbf{M}_k^{\mathrm{T}} \nabla J_k$ are the adjoint variables at time step t_{k-1}.

Hence, it follows (see Equation (23)) that

$$\begin{aligned}|\delta J_k| &\approx \langle y_{k-1}^*, \delta y_{k-1}\rangle = |\langle (\mathbf{A})^{-1} y_{k-1}^*, \delta y_{k-1}\rangle_{\mathbf{A}}| \\ &\leqslant \|(\mathbf{A})^{-1} y_{k-1}^*\|_{\mathbf{A}} \|\delta y_{k-1}\|_{\mathbf{A}}\end{aligned} \tag{40}$$

Hence, the dual weights w_k associated with the snapshots selection are thus defined as normalized values in the following:

$$\begin{aligned} c_k &= \|(\mathbf{A})^{-1} y_k^*\|_{\mathbf{A}} \\ w_k &= c_k \Big/ \sum_{j=1}^{j=n} c_j, \quad k = 1, \ldots, n \end{aligned} \tag{41}$$

DOI: 10.1002/fld

and provide a measure of the relative impact of the perturbations of state variables on the cost functional. A large value of weight w_k indicates that state errors at time step t_k play an important role in the optimization. In other words, the dual weights are chosen in order that information from DAS is incorporated directly into the optimality criteria that determines the POD basis functions. Hence, the dual-weighted POD incorporates not only information from the dynamical system, but also information from the DAS. The traditional POD basis aims at capturing the most energetic modes of the dynamical system, whereas the dual-weighted approach may also capture lower energy modes that can be significant for the successful implementation of 4-D VAR.

From an implementation point of view, the evaluation of all dual weights requires only one adjoint model integration.

1. Initialize the adjoint variables y^* at final time to zero: $y_n^* = 0$
2. For each step $k-1$ the adjoint variables y_{k-1}^* are obtained by $y_{k-1}^* = \mathbf{M}_k^{\mathrm{T}} y_k^* + \mathbf{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} (\mathbf{H}_k y_k - y_k^{\mathrm{o}})$
3. We obtain $y_0^* = y_0^* + \mathbf{B}^{-1}(y_0 - y^{\mathrm{b}})$ where y^{b} is the background prior state estimation.
4. Compute $c_k = \|(\mathbf{A})^{-1} y_k^*\|_{\mathbf{A}}$ and $w_k = c_k / \sum_{j=1}^{j=n} c_j, k = 1, \ldots, n$

where $\mathbf{M}_k$ is the tangent linear model and $\mathbf{H}_k$ is the linearized observation operator at time step k.

Hence, the evaluation of the dual weights requires only the integration of the adjoint model backward in time. As the adjoint model is available during the implementation of 4-D VAR data assimilation, no additional cost is required for the development of DWPOD 4-D VAR over the classic POD 4-D VAR.

In the numerical experiments section, we will consider a total energy norm defined as

$$\begin{aligned} \|\mathbf{y}\|^2_{\mathbf{A}^{\mathrm{FEM}}} &= \frac{1}{2}(\|u\|^2_{L_2(\Omega)} + \|v\|^2_{L_2(\Omega)} + \frac{g}{\overline{h}}\|u\|^2_{L_2(\Omega)}) \\ &= \frac{1}{2}\left(u^{\mathrm{T}}\mathbf{A}u + v^{\mathrm{T}}\mathbf{A}v + \frac{g}{\overline{h}}h^{\mathrm{T}}\mathbf{A}h\right) = \mathbf{y}^{\mathrm{T}}\mathbf{A}^{\mathrm{FEM}}\mathbf{y} \end{aligned} \tag{42}$$

where $\mathbf{A} = \mathbf{V}^{\mathrm{T}}\mathbf{V}$ is a symmetric positive-definite matrix and $[\mathbf{V}] = [V_1, \ldots, V_N]$ is the FE basis, $\overline{h}$ is the mean height of the reference data at the initial time. Hence $\mathbf{A}^{\mathbf{FEM}}$ can be viewed as a symmetric positive-definite block-wise diagonal matrix:

$$\mathbf{A}^{\mathbf{FEM}} = \mathrm{diag}\left(\frac{1}{2}\mathbf{A} \quad \frac{1}{2}\mathbf{A} \quad \frac{g}{2\overline{h}}\mathbf{A}\right) \tag{43}$$

5.2. Reduced-order POD 4-D VAR

In order to reduce the computational cost of 4-D VAR data assimilation we consider minimization of the cost functional in a space whose dimension is much smaller than that of the original one. A way to drastically decrease the dimension of the control space without significantly compromising the quality of the final solution but sizably decreasing the cost in memory and CPU time of 4-D VAR motivates us to choose to project the control variable on a basis of characteristic vectors capturing most of the energy and the main directions of variability of the model, i.e. SVD. One would then attempt to control the vector of initial conditions in the reduced space model.

The reduced-order cost functional can be expressed as

$$\begin{aligned} J^{\mathrm{POD}}(y_0^{\mathrm{POD}}) &= \frac{1}{2}(y_0^{\mathrm{POD}} - y^b)^{\mathrm{T}}\mathbf{B}^{-1}(y_0^{\mathrm{POD}} - y^{\mathrm{b}}) \\ &\quad + \frac{1}{2}\sum_{k=0}^{k=n}(H_k y_k^{\mathrm{POD}} - y_k^o)^{\mathrm{T}}\mathbf{R}_k^{-1}(H_k y_k^{\mathrm{POD}} - y_k^{\mathrm{o}}) \end{aligned} \tag{44}$$

where $\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}_k$ is the observation error covariance matrix at time level k, H_k is the observation operator at time level k. y^{b} is the background prior state estimation. y_0^{POD} is a vector containing the control variables (here, initial conditions) represented

DOI: 10.1002/fld

by the POD basis. y_k^{POD} is a vector containing the solution of variables obtained from the reduced-order model at the time level k.

In a POD reduced-order model, the initial value y_0^{POD} and the reduced-order model solution y_k^{POD} can be expressed as

$$\begin{aligned} y_0^{\mathrm{POD}} &= \bar{y} + \sum_{i=1}^{i=M} \alpha_i(0)\psi_i = \bar{y} + \Psi\alpha_0 \\ y_k^{\mathrm{POD}} &= \bar{y} + \sum_{i=1}^{i=M} \alpha_i(t^k)\psi_i = \bar{y} + \Psi\alpha_k \end{aligned} \tag{45}$$

where an ensemble of POD basis is

$$\Psi = \{\psi^1, \psi^2, \ldots, \psi^M\} \tag{46}$$

Hence, we can rewrite the reduced-order cost functional $J^{\mathrm{POD}}(y_0^{\mathrm{POD}})$ dependent on y_0^{POD} as an explicit cost functional $J_\alpha^{\mathrm{POD}}(\alpha_0)$ dependent on α_0 that is the coefficient in the POD basis vectors Ψ. Once we find the minimizer of $\alpha_0^{\min} = \min_{\alpha_0} J_\alpha^{\mathrm{POD}}(\alpha_0)$, we can express the retrieved initial condition $y_0^{\mathrm{POD}} = \bar{y} + \Psi\alpha_0$ in the POD reduced-order model cost functional

$$\begin{aligned} J_\alpha^{\mathrm{POD}}(\alpha_0) = {} & \frac{1}{2}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}}) \\ & + \frac{1}{2}\sum_{k=0}^{k=n}(H_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}})^{\mathrm{T}}\mathbf{R}_k^{-1}(H_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}}) \end{aligned} \tag{47}$$

The reduced model can be written as:

$$\forall k, \quad \alpha_k = M_{0\to k}^{\mathrm{POD}}(\alpha_0) \tag{48}$$

By denoting

$$\forall k, \quad \alpha_k = M_{k-1\to k}^{\mathrm{POD}}(\alpha_{k-1}) = M_k^{\mathrm{POD}}(\alpha_{k-1}) \tag{49}$$

and by recurrence we obtain that

$$\alpha_k = M_k^{\mathrm{POD}} \cdots M_1^{\mathrm{POD}}\alpha_0 \tag{50}$$

The reduced-order cost functional $J_\alpha^{\mathrm{POD}}(\alpha_0)$ that is dependent on α_0 can be divided into two components

$$J_\alpha^{\mathrm{POD}} = J_\alpha^{\mathrm{POD},b} + J_\alpha^{\mathrm{POD,o}} \tag{51}$$

where background cost functional that is dependent on α_0 is written as

$$J_\alpha^{\mathrm{POD,b}} = \tfrac{1}{2}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}}) \tag{52}$$

and the observational cost functional that is dependent on α_0 is written as

$$J_\alpha^{\mathrm{POD,o}} = \frac{1}{2}\sum_{k=0}^{k=n}(H_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}})^{\mathrm{T}}\mathbf{R}_k^{-1}(H_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}}) \tag{53}$$

Denoting 'normalized departures'

$$d_k = \mathbf{R}_k^{-1}(H_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}}) \tag{54}$$

and the contributions to the observational cost functional that is dependent on α_0 can be written as

$$J_{\alpha,k}^{\mathrm{POD,o}} = (H_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}})^{\mathrm{T}}d_k \tag{55}$$

DOI: 10.1002/fld

Hence the reduced-order cost functional that is dependent on α_0 can be rewritten as

$$J_\alpha^{\mathrm{POD}} = J_\alpha^{\mathrm{POD,b}} + \sum_{k=0}^{n} J_{\alpha,k}^{\mathrm{POD,o}} \tag{56}$$

Therefore, the gradient of the reduced-order cost functional with respect to the α_0 can be derived as

$$\nabla_{\alpha_0} J_\alpha^{\mathrm{POD}} = \nabla_{\alpha_0} J_\alpha^{\mathrm{POD,b}} + \sum_{k=0}^{n} \nabla_{\alpha_0} J_{\alpha,k}^{\mathrm{POD,o}} \tag{57}$$

$$\nabla_{\alpha_0} J_\alpha^{\mathrm{POD}} = \mathbf{\Psi}^{\mathrm{T}} \mathbf{B}^{-1}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}}) + \sum_{k=0}^{n} (\mathbf{M}_1^{\mathrm{POD}})^{\mathrm{T}} \cdots (\mathbf{M}_{\mathbf{k}}^{\mathrm{POD}})^{\mathrm{T}} \mathbf{\Psi}^{\mathrm{T}} \mathbf{H}_{\mathbf{k}}^{\mathrm{T}} d_k \tag{58}$$

where $(\mathbf{M}_{\mathbf{k}}^{\mathrm{POD}})^{\mathrm{T}}$ is the POD reduced-order adjoint model at time step k.

From an implementation point of view, we can compute the gradient $\nabla_{\alpha_0} J_\alpha^{\mathrm{POD}}$ in the following steps(see [34–36])

1. Initialize the *reduced-order adjoint variable*s α^* at final time to zero: $\alpha_n^* = 0$
2. For each step $k-1$ the adjoint variables α_{k-1}^* is obtained by adding the *reduced-order adjoint forcing term* $\mathbf{\Psi}^{\mathrm{T}} \mathbf{H}_{\mathbf{k}}^{\mathrm{T}} d_k$ to α_k^* and by performing the *reduced-order adjoint integration* of reduced-order model by multiplying the result by $(\mathbf{M}_{\mathbf{k}}^{\mathrm{POD}})^{\mathrm{T}}$, i.e. $\alpha_{k-1}^* = (\mathbf{M}_{\mathbf{k}}^{\mathrm{POD}})^{\mathrm{T}}(\alpha_k^* + \mathbf{\Psi}^{\mathrm{T}} \mathbf{H}_{\mathbf{k}}^{\mathrm{T}} d_k)$
3. At the end of recurrence, the value of the adjoint variable $\alpha_0^* = J_{\alpha_0}^{\mathrm{o}}$ yields the gradient of the observational cost functional
4. Compute $\nabla_{\alpha_0} J_\alpha^{\mathrm{POD,b}} = \mathbf{\Psi}^{\mathrm{T}} \mathbf{B}^{-1}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}})$ and we obtain $\nabla_{\alpha_0} J_\alpha^{\mathrm{POD}} = \nabla_{\alpha_0} J_\alpha^{\mathrm{POD,b}} + \nabla_{\alpha_0} J_\alpha^{\mathrm{POD,o}}$

5.3. Trust-region POD optimal control approach

Classical trust-region method. Historically the trust-region method goes back to [37–39]. See also the work [40] followed by an important work [41, 42]. Finally, the terminology of trust region and Cauchy point was put forward by Dennis [43] and systematized by More and Sorensen [44].

The classical trust-region method [45] aims to define a region around the current iterate within which it trusts the model to be an adequate representation of the objective function f, and then choose the step to be the approximate minimizer of the model in the trust region, i.e. choosing direction and length of the step simultaneously. The algorithm approximates only a certain region (the so-called trust region) of the objective function with a model function (often a quadratic). It is assumed that the first two terms of the quadratic model function m_k at each iterate x_k are identical with the first two terms of the Taylor-series expansion of f around x_k in the following:

$$m_k(p) = f_k + \nabla f_k^{\mathrm{T}} + \frac{1}{2} p^{\mathrm{T}} \mathbf{B}_k p \tag{59}$$

where $f_k = f(x_k)$ and $\nabla f_k = \nabla f(x_k)$ and $\mathbf{B}_k$ is an approximation to the Hessian and more generally $\mathbf{B}_k$ is some symmetric matrix.

To obtain each step, we seek a solution of the following sub-problem for which we need only an approximate solution to obtain convergence and good practical behavior [5]

$$\min \quad m_k(p) = f_k + \nabla f_k^{\mathrm{T}} + \frac{1}{2} p^{\mathrm{T}} \mathbf{B}_k p \tag{60}$$

$$\text{s. t.} \quad \|p\| \leqslant \delta_k \tag{61}$$

where $\delta_k > 0$ is the trust-region radius.

In the strategy for choosing the trust-region radius δ_k at each iteration, we define the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)} \tag{62}$$

DOI: 10.1002/fld

where the numerator is called the actual reduction, and the denominator is called the predicted reduction. We measure agreement between the model function m_k and the objective function $f(x_k)$ as a criterion for choosing trust-region radius $\delta_k > 0$. If the ratio ρ_k is negative, the new objective value is greater than the current value so that the step must be rejected. On the other hand, if ρ_k is close to 1, there is a good agreement between the approximate model m_k and the object function f_k over this step, so it is safe to expand the trust-region radius for the next iteration. If ρ_k is positive but not close to 1, we do not alter the trust-region radius, but if it is close to zero or negative, we shrink the trust-region radius.

Trust-region POD method. In this work, the POD reduced-order 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 reduced-order model when the resulting control variables from the latest optimization iteration are significantly different from the ones on which the POD model is based. Hence, it is natural to improve the POD reduced-order control model successively by updating the snapshots which are used to generate the POD basis in the process of reduced-order 4-D VAR.

For the reduced-order cost functional [46, 47]

$$J^{\mathrm{POD}}(y_0^{\mathrm{POD}}) = \frac{1}{2}(y_0^{\mathrm{POD}} - y^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(y_0^{\mathrm{POD}} - y^{\mathrm{b}}) + \frac{1}{2}\sum_{k=0}^{k=n}(\mathbf{H}_k y_k^{\mathrm{POD}} - y_k^{\mathrm{o}})^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{H}_k y_k^{\mathrm{POD}} - y_k^{\mathrm{o}}) \tag{63}$$

or its explicit version

$$J_\alpha^{\mathrm{POD}}(\alpha_0) = \frac{1}{2}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\bar{y} + \Psi\alpha_0 - y^{\mathrm{b}}) + \frac{1}{2}\sum_{k=0}^{k=n}(\mathbf{H}_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}})^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{H}_k(\bar{y} + \Psi\alpha_k) - y_k^{\mathrm{o}}) \tag{64}$$

defined above, we first start with a random perturbation of given initial condition $y_0^{(0)}$ and compute the snapshots that correspond to the flow behavior forced by $y_0^{(0)}$. We then use these snapshots to compute the first POD basis $\mathbf{\Psi}^{(0)}$ and build up the corresponding POD-based control model forced by applying inner projection $\alpha_0^{(0)} = \langle y_0^{(0)} - \bar{y}, \mathbf{\Psi}^{(0)} \rangle$. We now implement the inner minimization iteration based on $\mathbf{\Psi}^{(0)}$ to obtain the new control variable $\alpha_0^{(1)}$ in the reduced-order space. When we carry out an outer iteration, we obtain $y_0^{(1)} = \bar{y} + \mathbf{\Psi}^{(0)}\alpha_0^{(1)}$. If we use $y_0^{(1)}$ for the computation of new snapshots and a new POD basis $\mathbf{\Psi}^{(1)}$, we can improve the initial condition of the PDE and thus improve the POD-based model. However, the outer projection computing new snapshots and corresponding new POD basis is computationally expensive and should only occur at rare instances controlled by appropriate criteria. One criterion for adaptivity consists of an *ad hoc* rule that an outer projection should occur whenever the value of the objective function cannot be decreased beyond a prescribed tolerance between two consecutive inner minimization iterations. Also, this criterion will abort the outer iteration cycle when the value of the objective function is less than a given tolerance. The trust-region POD approach for adaptivity is both efficient and mathematically correct, being based on the trust-region globalization properties derived from optimization theory [4].

Therefore, to find a new step s^k, we minimize with respect to s

$$\min \quad m_k(\alpha_0^{(k)} + s) := J_\alpha^{\mathrm{POD}}(\alpha_0^{(k)} + s) \tag{65}$$

$$\text{s. t.} \quad \|s\| \leqslant \delta_k \tag{66}$$

DOI: 10.1002/fld

Based on trust-region strategy from optimization [2, 47] , we can decide to increase or decrease the trust-region radius by comparing the actual(for the full-order model)

$$J(\bar{y}+\mathbf{\Psi}^{(k-1)}\alpha_0^{(k)})-J(\bar{y}+\mathbf{\Psi}^{(k-1)}(\alpha_0^{(k)}+s_k)) \tag{67}$$

with the predicted decrease (for the reduced-order model)

$$m_k(\alpha_0^{(k)})-m_k(\alpha_0^{(k)}+s_k) \tag{68}$$

Outline of trust-region POD algorithm: Let $0<\eta_1<\eta_2<1$, $0<\gamma_1<\gamma_2<1\leqslant\gamma_3$ and $y_0^{(0)}$, δ_0 be given, set $k=0$

1. Compute snapshot set $\mathscr{Y}_k^{\mathrm{SNAP}}$ based on initial condition $y_0^{(k)}$
2. Compute the POD basis $\mathbf{\Psi}^{(k)}$ and build up the corresponding POD-based control model based on the initial condition $\alpha_0^{(0)}=\langle y_0^{(0)}-\bar{y}, \mathbf{\Psi}^{(0)}\rangle$
3. Compute the minimizer s^k of

$$\begin{aligned} \min \quad & m_k(\alpha_0^{(k)}+s) \\ \text{s. t.} \quad & \|s\|\leqslant\delta_k \end{aligned}$$

4. Compute the new $J(\bar{y}+\mathbf{\Psi}^{(k-1)}(\alpha_0^{(k)}+s_k))$ of the full model and

$$\rho_k=\frac{J(\bar{y}+\mathbf{\Psi}^{(k-1)}\alpha_0^{(k)})-J(\bar{y}+\mathbf{\Psi}^{(k-1)}(\alpha_0^{(k)}+s_k))}{m_k(\alpha_0^{(k)})-m_k(\alpha_0^{(k)}+s_k)} \tag{69}$$

5. Update the trust-region radius:
   - If $\rho_k\geqslant\eta_2$: implement outer projection $y_0^{(k+1)}=\bar{y}+\mathbf{\Psi}^{(k-1)}(\alpha_0^{(k)}+s_k)$ and increase trust-region radius $\delta_{k+1}=\gamma_3\delta_k$ and GOTO 1
   - If $\eta_1<\rho_k<\eta_2$: implement outer iteration $y_0^{(k+1)}=\bar{y}+\mathbf{\Psi}^{(k-1)}(\alpha_0^{(k)}+s_k)$ and decrease trust-region radius $\delta_{k+1}=\gamma_2\delta_k$ and GOTO 1
   - If $\rho_k\leqslant\eta_1$: set $y_0^{(k+1)}=y_0^{(k)}$ and decrease trust-region radius $\delta_{k+1}=\gamma_1\delta_k$ and GOTO 3

In the trust-region POD optimal control algorithm above, the gradient of $m_k(\alpha_0^{(k)}+s)$ with respect to s plays an important role in the constrained minimization sub-problem

$$\begin{aligned} \min \quad & m_k(\alpha_0^{(k)}+s) \\ \text{s. t.} \quad & \|s\|\leqslant\delta_k \end{aligned}$$

On the one hand, if δ_k is large enough and the norm constraint is not active, then s_k is just in the vicinity of the unconstrained minimum. On the other hand, if δ_k is small, then the higher order terms in s play a less important role than the linear term, i.e. for some constant β_k it holds $s_k\approx-\beta_k\nabla J_\alpha^{\mathrm{POD}}(\alpha_0^{(k)})$. As δ_k is increasing we obtain a continuous change from the direction of steepest descent to the direction of the minimum of $J_\alpha^{\mathrm{POD}}(\alpha_0^{(k)})$. Therefore good gradient information is required, which can be obtained by performing the reduced-order adjoint backward in time integration (Figure 1).

Following the trust-region philosophy, it is not necessary to determine the exact step solution of the constrained problem above. It is sufficient to compute a trial step s_k that achieves only a certain amount of decrease for the full model. We can use a backtracking approach to find the sufficient decrease. For recent work on stable Galerkin reduced-order models, see Barone *et al.* [29].

5.4. Dual-weighted TR-POD approach

A new methodology combining the dual-weighted snapshots and trust-region POD adaptivity is put forward, allowing us to enhance the benefits already provided by using DWPOD. The combined algorithm proceeds as follows illustrated in the algorithm flowchart (see Figure 2).

DOI: 10.1002/fld

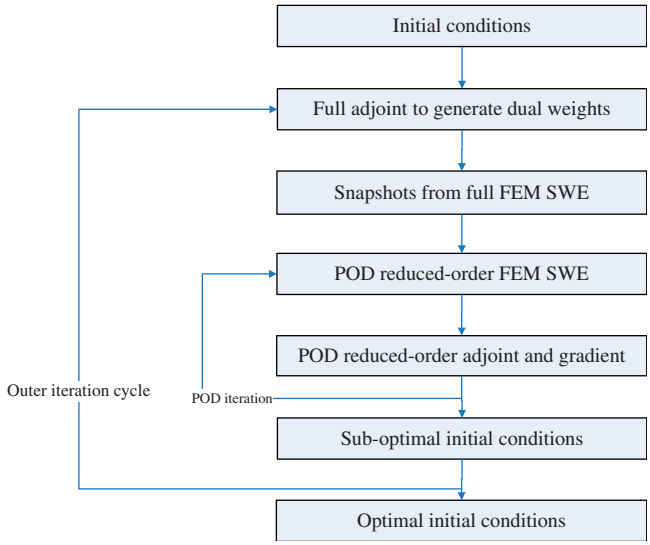

Figure 1. Flowchart of the methodology using adaptive POD reduced-order model for dual-weighted snapshots of the full model.

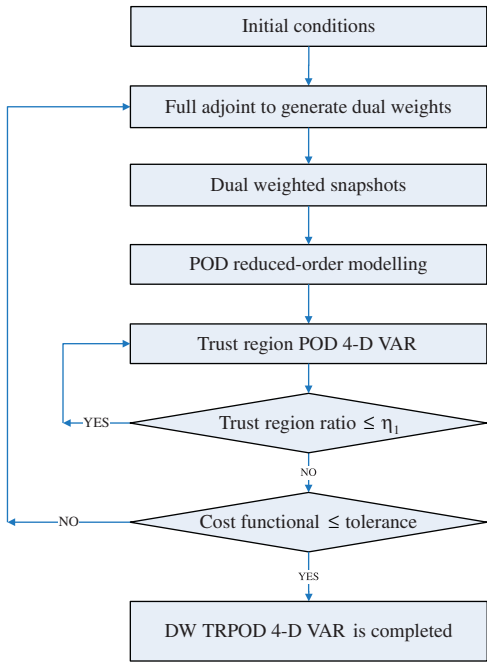

Figure 2. Flowchart of the methodology combining dual-weighed snapshots and TR-POD adaptivity.

6. NUMERICAL EXPERIMENTS

6.1. *Description of problem*

The model test problem used here adopts the following initial conditions (Figure 3) from the initial height field condition No.1 of Grammeltvedt [48]:

$$h(x,y)=H_0+H_1\tanh\left(\frac{9(D/2-y)}{2D}\right)+H_2\left(1/\cosh^2\left(\frac{9(D/2-y)}{D}\right)\right)\sin\left(\frac{2\pi x}{L}\right) \tag{70}$$

where this initial condition has energy in wave number one in the x-direction.

DOI: 10.1002/fld

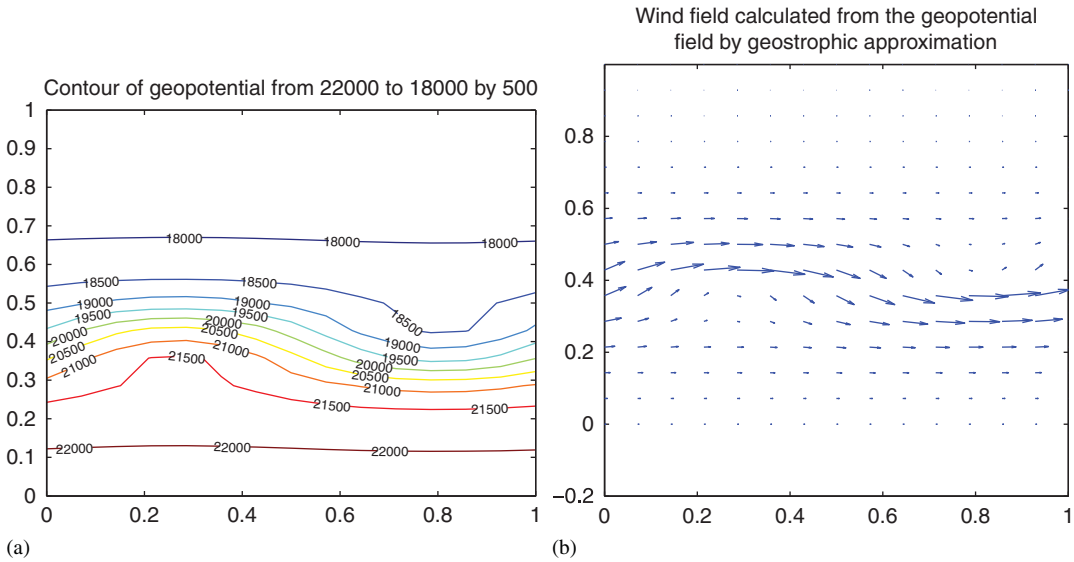

Figure 3. Grammeltvedt I initial condition: (a) geopotential field and (b) wind field by geostrophic approximation.

The initial velocity fields were derived from the initial height field using the geostrophic relationship:

$$u = -\left(\frac{g}{f}\right)\frac{\partial h}{\partial y} v = \left(\frac{g}{f}\right)\frac{\partial h}{\partial x} \tag{71}$$

The dimensional constants used here are:

$$\begin{aligned} &L = 4400\,\text{km}, \quad D = 6000\,\text{km}, \quad \bar{f} = 10^{-4}\,\text{s}^{-1}, \quad \beta = 1.5 \times 10^{-11}\,\text{s}^{-1}\,\text{m}^{-1}, \\ &g = 10\,\text{ms}^{-1}, \quad H_0 = 2000\,\text{m}, \quad H_1 = 220\,\text{m}, \quad H_2 = 133\,\text{m}. \end{aligned} \tag{72}$$

and the space increments used here are

$$\Delta x = \Delta y = 200\,\text{km}, \quad \Delta t = 1800\,\text{s} \tag{73}$$

We employed linear piecewise polynomials on triangular elements in the formulation of Galerkin FE shallow-water equations model [49], in which the global matrix was stored into a compact matrix (see [50]). A time-extrapolated Crank–Nicholson time differencing scheme was applied for integrating in time the system of ordinary differential equations resulting from the application of the Galerkin FE method and the Courant–Friedrichs–Levy (CFL) criterion was $\sqrt{gH_0}(\frac{\Delta t}{\Delta x}) < \frac{\sqrt{2}}{2}$ (see [51, 52]), based on which the shallow-water equations system was then coupled at every time step so that the equations are quasi-linearized (see [53]).

In order to implement boundary conditions in the Galerkin FE model, we have adopted the approach suggested by Payne and Irons [54] and mentioned by Huebner *et al.* [55]. This approach consists in modifying the diagonal terms of the global matrix associated with the nodal variables by multiplying them by a large number, say 10^{16}, while the corresponding term in the right-hand vector is replaced by the specified boundary nodal variable multiplied by the same large factor times the corresponding diagonal term. This procedure is repeated until all prescribed boundary nodal variables have been treated (see [51]).

6.2. Discussion of numerical results obtained by trust-region POD 4-D VAR combined with dual-weighted snapshots selection

In the numerical experiment, we applied a 1% uniform random perturbations on the initial conditions in order to provide twin-experiment 'observations'. We also computed the errors between the

DOI: 10.1002/fld

retrieved initial conditions related to 5% uniform random perturbations of the true initial conditions as the initial guess of the reduced-order 4-D VAR (Figure 6). The data assimilation was carried on a 48-h window using the $\Delta t = 1800$ s in time and a mesh of 30×24 grid points in space, thus we generated 96 snapshots by integrating the full finite-element shallow-water equations model forward in time, from which we choose 10 POD bases for each of the $(u(x, y), v(x, y), \phi(x, y))$ to capture over 99.9% of the energy. The dimension of control variables vector for the reduced-order 4-D VAR thus is $10 \times 3 = 30$ (Figure 4).

In the process of POD 4-D VAR, the resulting control variables from the latest optimization iteration are projected to the full model to generate new POD bases. The new POD bases then replace the previous ones resulting in a new POD reduced-order model. We found that both the root mean square error (RMSE) and the correlation error metrics between the full model solutions and the reduced-order solutions were improved after each outer projection was carried out.

The Polak Ribiere nonlinear conjugate gradient (CG) technique [56] was employed for high-fidelity full model 4-D VAR and all variants of ad-hoc POD 4-D VAR, whereas the steepest-descent strategy was employed for the trust-region POD 4-D VAR within the trust-region radius and provides a sufficient reduction in the high-fidelity model quantified in terms of the Cauchy point [5]. In the *ad hoc* POD 4-D VAR, the POD bases are re-calculated when the value of the cost function cannot be decreased by more than 10^{-1} for *ad hoc* POD 4-D VAR and 10^{-2} for *ad hoc* DWPOD 4-D VAR between the consecutive minimization iterations. In the trust-region 4-D VAR, the POD bases are re-calculated when the ratio ρ_k is larger than the trust-region parameter η_1 in the process of updating the trust-region radius.

The unweighted *ad hoc* POD 4-D VAR as a reduced-order approach required a smaller computation cost but could not achieve the same cost functional reduction as the high-fidelity model 4-D VAR. The dual-weighted *ad hoc* POD 4-D VAR achieves a better reduction in the cost functional. However, neither of the above mentioned methods can attain the minimum of the high-fidelity 4-D VAR model cost functional. Furthermore, the unweighted snapshots trust-region POD 4-D VAR yields an additional cost functional reduction compared with the *ad hoc* approach, albeit at a higher computational cost. Finally, the dual-weighted trust-region POD 4-D VAR achieves almost exactly the same cost functional reduction as the full high-fidelity 4-D VAR model, resulting in an additional decrease of four orders of magnitude compared with the minimization of the cost functional obtained by applying the unweighted *ad hoc* POD 4-D VAR (see Table I), showing

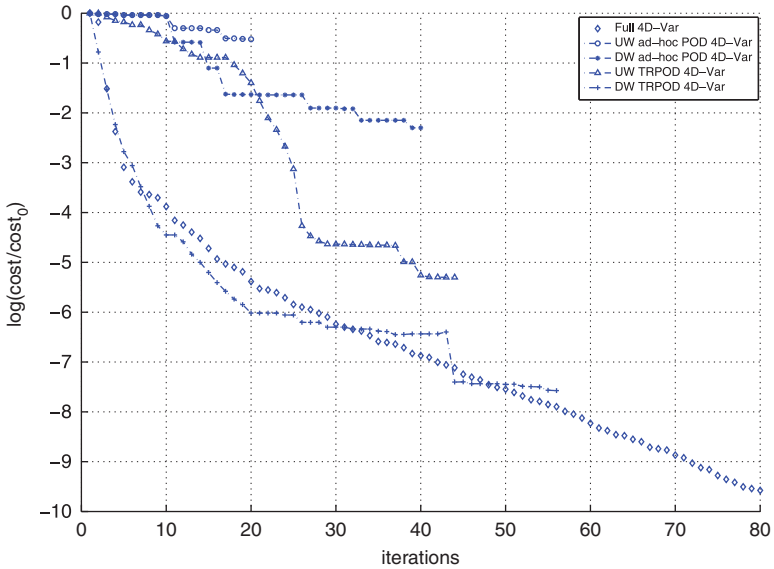

Figure 4. Comparison of the performance of minimization of cost functional in terms of number of iterations for *ad hoc* POD 4-D VAR, ad-hoc dual-weighed POD 4-D VAR, trust-region POD 4-D VAR, trust-region dual weighed POD 4-D VAR, and full model 4-D VAR.

DOI: 10.1002/fld

that the combination of the dual-weighted approach and trust-region method to model reduction is significantly beneficial in the achievement of a local minimum of optimization almost identical to one obtained by the high-fidelity full 4-D VAR.

In Figure 4, it is noticed that the dual-weighed 4-D VAR absorbs the information from the full 4-D VAR model and mimics the behavior of the full model 4-D VAR thus being able to achieve better reduction in the cost functional. It is also noticed that in the dual-weighted approach, the reduced basis is adjusted to according to the norm of the full adjoint variable. The dual weights are decreasing in time for the snapshots without sharp transients (Figure 5) due to the fact that observations are available in each time step in our experiments. Furthermore, the dual weights on the snapshot data are distinct from one outer projection to the next. The importance of snapshots for longer windows of assimilation may assume a preponderant importance after each outer iteration. However, it should be emphasized that the benefit obtained for POD 4-D VAR using the dual-weighted procedure diminishes as the dimension of the reduced space increases.

Once the retrieved initial condition is obtained by implementing the dual-weighted trust-region 4-D VAR, we can compare the results from the POD reduced model with those from the full model. To quantify the performance of the dual-weighted trust-region 4-D VAR, we use two metrics, namely the RMSE and correlation of the difference between the POD reduced-order simulation and high-fidelity model.

In particular, the RMSE (Figure 10) between variants of the POD reduced-model solution and the true one at the time level i is used to estimate the error of the POD model.

$$\mathrm{RMSE}^i = \sqrt{\frac{\sum_{j=1}^{j=N} (U_{i,j} - U_{i,j}^{\mathrm{POD}})^2}{n}}, \quad i = 1, \ldots, n \tag{74}$$

Table I. Comparison of iterations, outer projections, error and CPU time for *ad hoc* POD 4-D VAR, *ad hoc* dual-weighed POD 4-D VAR, trust-region POD 4-D VAR, trust-region dual-weighed POD 4-D VAR and the full model 4-D VAR.

POD 4-D VAR	AD-POD	DW-AD-POD	TR-POD	DW-TR-POD	Full
Iterations	22	42	46	57	80
Outer projections	2	6	10	12	NA
Error	10^{-1}	10^{-2}	10^{-5}	10^{-8}	10^{-10}
CPU time (s)	15.2	38.7	121.2	142.8	222.6

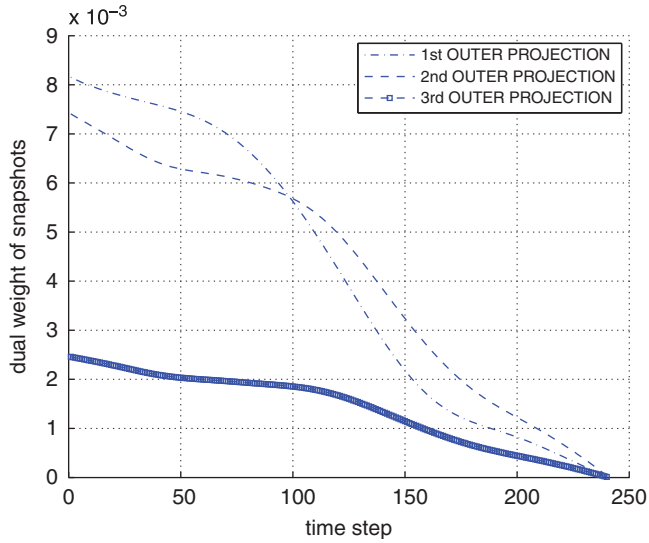

Figure 5. The dual weights of the snapshots data determined by the full adjoint variable for the trust-region POD 4-D VAR.

DOI: 10.1002/fld

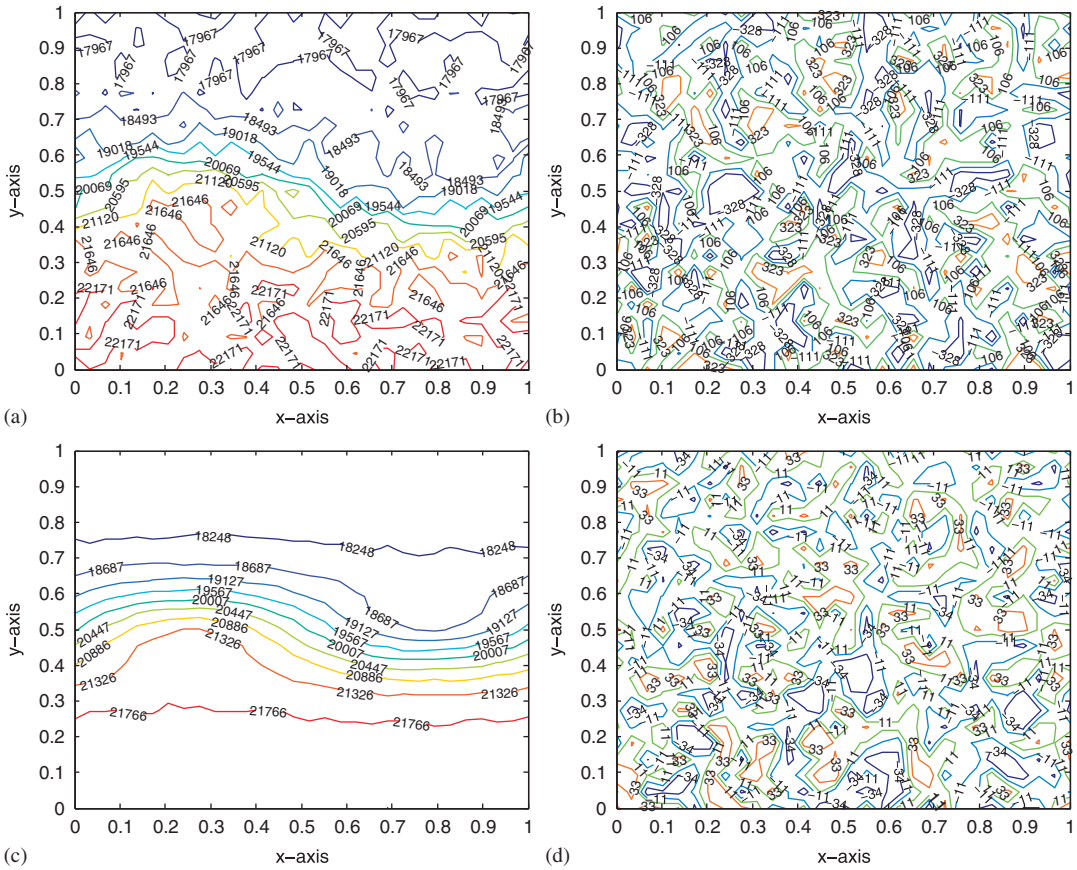

Figure 6. Errors between the retrieved initial geopotential and the true initial geopotential applying dual-weighted trust-region POD 4-D VAR to the 5% uniform random perturbations of the true initial conditions taken as the initial guess. (a) Shows the contour of 5% perturbation of true initial geopotential; (b) shows the contour of difference between 5% perturbation of true initial geopotential; (c) shows the contour of retrieved initial geopotential after 2 days with $dt = 1800$ s; (d) shows the contour of difference between retrieved initial geopotential and true initial geopotential.

where $U_{i,j}$ and $U_{i,j}^{\mathrm{POD}}$ are the state variables obtained by the full model and ones obtained by optimal POD reduced-order model of time level i at node j, respectively, and N is the total number of nodes over the domain. U and U^{POD} are used to either denote the geopotential or the velocity of the full model and the POD reduced-order model, respectively.

In Figure 11, the correlation r defined below is used as an additional metric to evaluate quality of the inversion simulation

$$r_i = \frac{cov_{12}^i}{\sigma_1^i \sigma_2^i} \tag{75}$$

where

$$\sigma_1^i = \sum_{j=1}^{j=N} (U_{i,j} - \overline{U}_j)^2, \quad \sigma_2 = \sum_{j=1}^{j=N} (U_{i,j}^{\mathrm{POD}} - \overline{U^{\mathrm{POD}}}_j)^2, \quad i = 1, \ldots, n \tag{76}$$

$$\mathrm{cov}_{12} = \sum_{j=1}^{j=N} (U_{i,j} - \overline{U}_j)(U_{i,j}^{\mathrm{POD}} - \overline{U^{\mathrm{POD}}}_j), \quad i = 1, \ldots, n \tag{77}$$

where $\overline{U}_j$ and $\overline{U^{\mathrm{POD}}}_j$ are the means over the simulation period $[0, T]$ obtained by the full model and ones obtained by optimal POD reduced-order model at node j, respectively.

DOI: 10.1002/fld

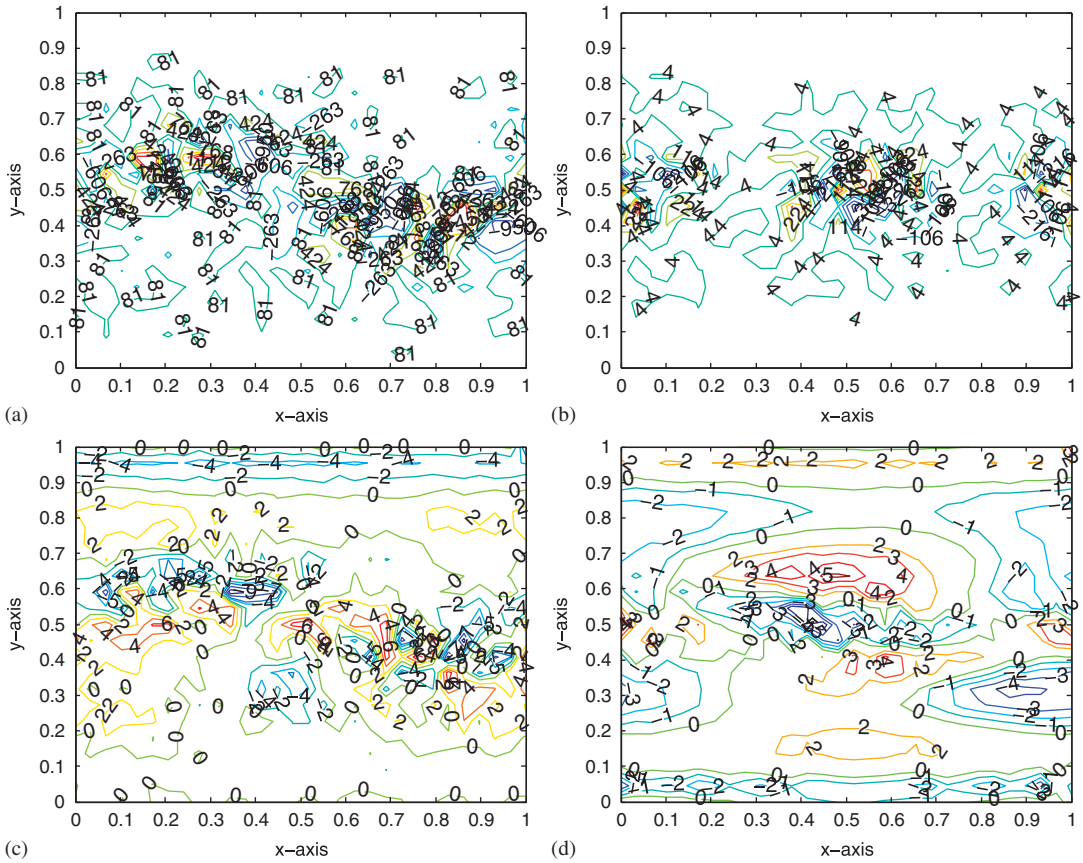

Figure 7. Errors scaled by 100 between the retrieved initial velocity field and the true initial velocity field applying dual-weighted trust-region POD 4-D VAR to the 5% uniform random perturbations of the true initial conditions taken as the initial guess. (a) Shows the contour of difference between true initial *u*-velocity and perturbed initial *u*-velocity; (b) shows the contour of difference between true initial *v*-velocity and perturbed initial *v*-velocity; (c) shows the contour of difference between retrieved initial u-velocity and true initial *u*-velocity; (d) shows the contour of difference between retrieved initial *v*-velocity and true initial *v*-velocity.

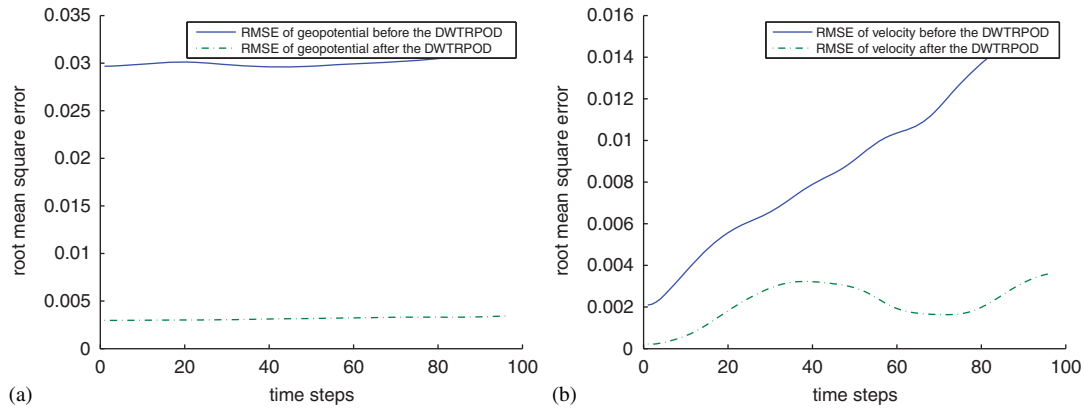

Figure 8. Comparison of the RMSE between the full model and the ROM before and after the data assimilation applying dual-weighted trust-region POD 4-D VAR to the 5% uniform random perturbations of the true initial conditions taken as the initial guess: (a) geopotential and (b) wind field.

DOI: 10.1002/fld

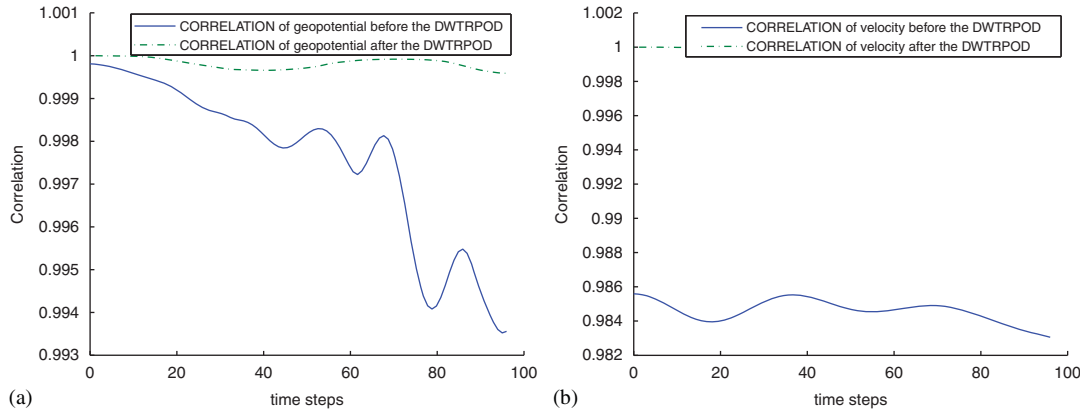

Figure 9. Comparison of the correlation between the full model and the ROM before and after the data assimilation applying dual-weighted trust-region POD 4-D VAR to the 5% uniform random perturbations of the true initial conditions serving as initial guess: (a) geopotential and (b) wind field.

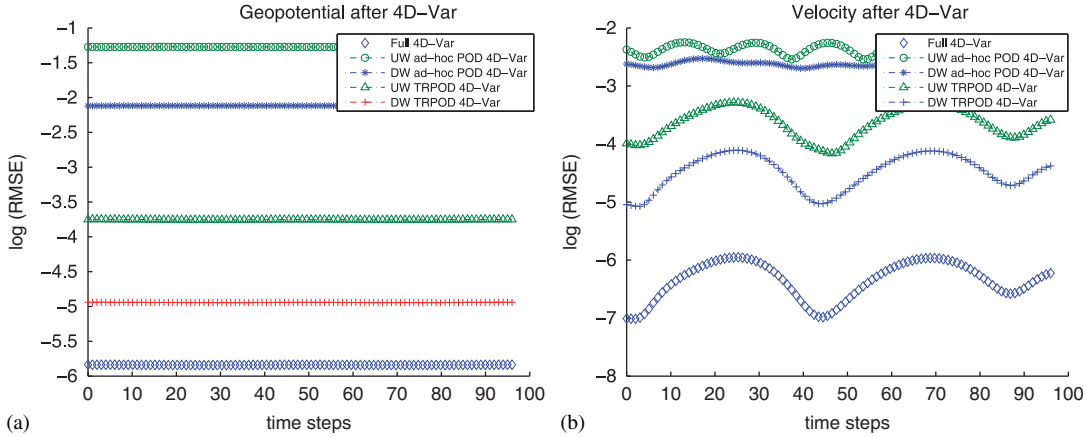

Figure 10. Comparison of the RMSE of between *ad hoc* POD 4-D VAR, ad-hoc dual weighed POD 4-D VAR, trust-region POD 4-D VAR, trust-region dual-weighed POD 4-D VAR, and the full model 4-D VAR: (a) geopotential and (b) wind field.

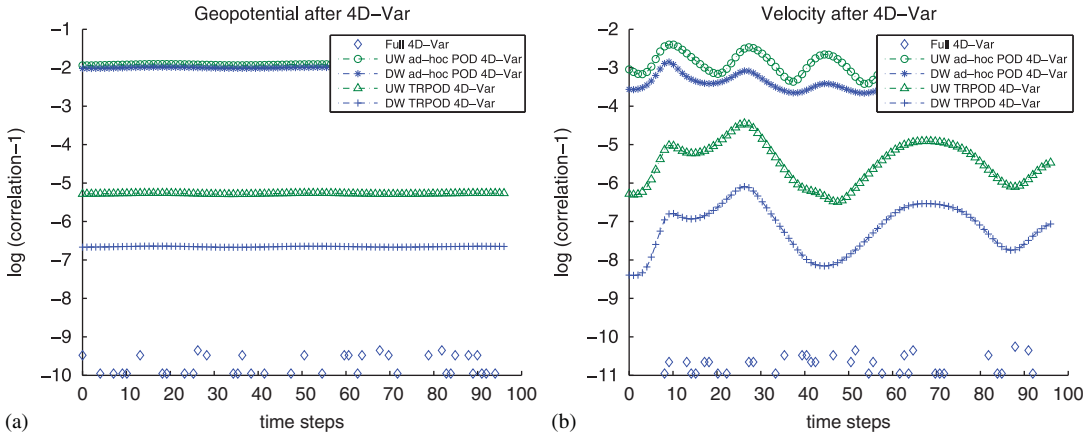

Figure 11. Comparison of correlation between ad-hoc POD 4-D VAR, *ad hoc* dual weighed POD 4-D VAR, trust-region POD 4-D VAR, trust-region dual-weighed POD 4-D VAR and the full model 4-D VAR: (a) geopotential and (b) wind field.

DOI: 10.1002/fld

Even though it turned out to be advantageous to combine the dual-weighed approach with the trust-region POD 4-D VAR, it should be emphasized that this advantage diminishes when we increase the number of POD bases for each component of the $(u(x, y), v(x, y), \phi(x, y))$ from 10 to 20 by applying both metrics mentioned above. However, increasing the dimension of the POD reduced-order space from 30 to 60 can increase the computational cost of POD reduced-order 4-D VAR. This agrees with results obtained in [7] for practical applications, the dual-weighted procedure may be of particular benefit for use only with small dimensional bases in the context of adaptive order reduction as the minimization approaches the optimal solution. For other beneficial effects of POD 4-D VAR related to its use in the framework of second order adjoint of a global shallow-water equations model, see Daescu and Navon (2007) [57] (Figures 6–11).

7. CONCLUSION

In this paper, we solved an inverse problem for the POD reduced-order shallow-water equations model using a FE formulation, controlling its initial conditions in presence of observations being assimilated in a time window. In this POD 4-D VAR, we developed the full adjoint of the FE shallow-water equations model and the reduced-order adjoint for POD reduced-order model. We integrated the full adjoint model backward in time to compute the time-varying sensitivities of the full 4-D VAR cost functional with respect to time-varying model states, from which we derived the dual weights of the ensemble of snapshots. Also, we integrated the reduced-order adjoint model backward in time to compute gradient of reduced-order cost functional.

In the numerical experiments, we compared several variants of POD 4-D VAR, namely unweighted *ad hoc* POD 4-D VAR, dual-weighed *ad hoc* POD 4-D VAR, unweighted trust-region POD 4-D VAR, and dual-weighed trust-region POD 4-D VAR, respectively. We found that the ad-hoc POD 4-D VAR version yielded the least reduction in the cost functional compared with the trust-region 4-D VAR . We assume that this result may be attributed to lack of feedbacks from the high-fidelity model . On the other hand, the trust-region POD 4-D VAR version yielded a sizably better reduction of the cost functional, due to inherent properties of TR-POD allowing local minimizer of the full problem to be attained by minimizing the TR-POD sub-problem. Thus, trust-region 4-D VAR resulted in global convergence to the high-fidelity local minimum starting from any initial iterates.

The dual-weighted POD selection of snapshots allows propagation of information from the DAS onto the reduced order model, possibly capturing lower energy modes that may play significant role in successful implementation of 4-D VAR data assimilation. Combining the dual-weighted approach with the trust-region POD approach to model reduction results in a significant enhanced benefit achieving a local minimum of reduced cost function optimization almost identical to the one obtained by the high-fidelity full 4-D VAR model. Hence we achieve a double benefit while running a reduced-order inversion at an acceptable computational cost, at least for the shallow-water equations model in a two-dimensional spatial domain.

In future research work we will consider a combination of the balanced truncation technique with the dual-weighted trust-region POD-reduced 4-D VAR (see work of Rowley [30, 58]) For novel original approaches to reduced-order Galerkin for fluid flows. (see work of Noack *et al.* [59, 60] and work of Tadmor *et al.* [61]).

ACKNOWLEDGEMENTS

Prof I. M. Navon acknowledges support of NSF grants ATM-0201808 and CCF-0635162. Dr F. Fang acknowledges support of the U.K.'s Natural Environment Research Council NE/C51829X/1.

DOI: 10.1002/fld

REFERENCES

1. Fahl M. Trust-region methods for flow control based on reduced order modeling. *Ph.D. Thesis*, Trier University, 2000.
2. Arian E, Fahl M, Sachs EW. Trust-region proper orthogonal decomposition for flow control. *Technical Report*: *TR-2000-25*, Institute for Computer Applications in Science and Engineering, 2000.
3. Conn AR, Gould NIM, Toint PL. *Trust-region Methods*. SIAM: Philadelphia, 2000.
4. Toint PL. Global convergence of a class of trust-region methods for nonconvex minimization in Hilbert space. *IMA Journal of Numerical Analysis* 1988; **8**(2):231–252.
5. Nocedal J, Wright SJ. *Numerical Optimization* (2nd edn). Springer Series in Operations Research and Financial Engineering. Springer: New York, 2006.
6. Bui-Thanh T, Willcox K, Ghattas O, Van Bloemen Waander B. Goal-oriented model constrained for reduction of large-scale systems. *Journal of Computational Physics* 2007; **224**(2):880–896.
7. Daescu DN, Navon IM. A dual-weighted approach to order reduction in 4DVAR data assimilation. *Monthly Weather Review* 2008; **136**(3):1026–1041.
8. Cao Y, Zhu J, Luo Z, Navon IM. Reduced order modeling of the upper tropical pacific ocean model using proper orthogonal decomposition. *Computers and Mathematics with Applications* 2006; **52**(8–9):1373–1386.
9. 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.
10. Fang F, Pain CC, Piggott MD, Gorman GJ, Goddard AJH. An adaptive mesh adjoint data assimilation method applied to free surface flows. *International Journal for Numerical Methods in Fluids* 2005; **47**(8–9):995–1001.
11. Fang F, Piggott MD, Pain CC, Gorman GJ, Goddard AJH. Adjoint data assimilation into a 3d unstructured mesh coastal finite element model. *Ocean Modelling* 2006; **15**(1–2):3–38.
12. 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; **59**(8):827–851.
13. Fang F, Pain CC, Navon IM, Piggott MD, Gorman GJ, Allison P, Goddard AJH. A POD reduced order 4d-var adaptive mesh ocean modelling approach. *International Journal for Numerical Methods in Fluids* 2009; **60**(7):709–732.
14. Altaf MU, Heemink AW, Verlaan M. Inverse shallow water flow modeling using model reduction. *International Journal for Multiscale Computational Engineering* 2009; **7**(6):577–594.
15. Tan WY. *Shallow-water Hydrodynamics*: *Mathematical Theory and Numerical Solution for a Two-dimensional System of Shallow-water Equations*. Elsevier Oceanography Series: Beijing, 1992.
16. Vreugdenhil CB. *Numerical Methods for Shallow-water Flow*. Kluwer Academic Publishers: Boston, 1994.
17. Galewsky J, Cott RK, Polvani LM. An initial-value problem for testing numerical models of the global shallow-water equations. *Tellus* 2004; **56**(5):429–440.
18. Chen X, Navon IM. Optimal control of a finite-element limited-area shallow-water equations model. *Studies in Informatics and Control* 2009; **18**(1):41–62.
19. Kreiss HO, Oliger J. Comparison of accurate methods for the integration of hyperbolic equations. *Tellus* 1972; **24**:199–215.
20. Lumley JL. The structure of inhomogeneous turbulent flows. In *Atmospheric Turbulence and Radio Wave Propagation*, Yaglom AM, Tatarski VI (eds). 1967; Nauka: Moscow, 166–178.
21. Berkooz G, Holmes PJ, Lumley JL. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics* 1993; **25**(1):539–575.
22. Sirovich L, Lumley JL, Berkooz G. Turbulence and the dynamics of coherent structures, part III: dynamics and scaling. *Quarterly of Applied Mathematics* 1987; **45**(3):583–590.
23. Holmes P, Lumley JL, Berkooz G. *Turbulence*, *Coherent Structures*, *Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics, Cambridge: London, 1996.
24. Karhunen K. Zur Spektraltheorie stochastischer Prozesse. *Annales Academiae Scientiarum Fennicae Series A1—Mathematica Physica* 1946; 34.
25. Loeve M. Fonctions Aleatoires de Second Ordre. *Comptes Rendus de l'Academie des Sciences*, *Paris* 1945; 220, 469.
26. Kosambi DD. Statistics in function space. *Journal of the Indian Mathematical Society* 1943; **7**(1):76–88.
27. Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 1901; **2**(1):559–572.
28. Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 1933; **24**(1):417–441, 498–520.
29. Barone MF, Kalashnikova I, Segalman DJ, Thornquist HK. Stable Galerkin reduced order models for linearized compressible flow. *Journal of Computational Physics* 2009; **228**(6):1932–1946.
30. Rowley CW, Colonius T, Murray RM. Model reduction for compressible flows using POD and Galerkin projection. *Physica D* 2004; **189**(1–2):115–129.
31. Aquino W, Brigham JC, Earls CJ, Sukumak N. Generalized finite element method using proper orthogonal decomposition. *International Journal for Numerical Methods in Engineering* 2009; **79**(7):887–906.
32. Navon IM, Zou X, Derber J, Sela J. Variational data assimilation with an adiabatic version of the NMC spectral mode. *Monthly Weather Review* 1992; **120**(7):1435–1443.

DOI: 10.1002/fld

33. Kunish K, Volkwein S. Proper orthogonal decomposition for optimality systems. *Mathematical Modelling and Numerical Analysis* 2008; **42**(1):1–23.
34. Vermeulen PTM, Heemink AW. Model-reduced variational data assimilation. *Monthly Weather Review* 2006; **134**(10):2888–2899.
35. Kunisch K, Volkwein S. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische Mathematik* 2001; **90**(1):117–148.
36. Kunisch K, Volkwein S. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical Analysis* 2002; **40**(2):492–515.
37. Levenberg K. A method for the solution of certain problems in least squares. *Quarterly Journal on Applied Mathematics* 1944; **2**:164–168.
38. Morrison DD. *Proceedings of the Seminar on Tracking Programs and Orbit Determination*. Jet Propulsion Laboratory: U.S.A., 1960.
39. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 1963; **11**:431–441.
40. Winfield D. *Function and Functional Optimization by Interpolation in Data Tables. Ph.D. Thesis*, Harvard University, Cambridge, MA, U.S.A., 1969; 146.
41. Powell MJD. A FORTRAN subroutine for solving systems of nonlinear algebraic equations. In *Numerical Methods for Nonlinear Algebraic Equations*, Rabinowitz P (ed.). Gordon & Breach: London, 1970; 115–161.
42. Powell MJD. A FORTRAN subroutine for unconstrained minimization, requiring first derivatives of the objective function. *Report AERE-R.6469*, A.E.R.E. Harwell, Oxon., England, 1970.
43. Dennis JE. A brief introduction to quasi-Newton methods. In *Numerical Analysis*, Golub GH, Oliger J (eds). Proceedings of Symposia in Applied Mathematics Series, vol. 22. AMS: Providence, RI, U.S.A., 1978; 19–52.
44. More JJ, Sorensen DC. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing* 1983; **4**(3):553–572.
45. Celis M, Dennis JE, Tapia RA. A trust region strategy for nonlinear equality constrained optimization. In *Numerical Optimization*, Boggs P, Byrd R, Schnabel R (eds), 1984. SIAM: Philadelphia, 1985; 71–82.
46. Bergmann M, Cordier L. Optimal control of the cylinder wake in the laminar regime by trust-region methods and pod reduced-order models. *Journal of Computational Physics* 2008; **227**(16):7813–7840.
47. Bergmann M, Bruneau C, Iollo A. Enablers for robust pod models. *Journal of Computational Physics* 2009; **228**(2):516–538.
48. Grammeltvedt A. A survey of finite-difference schemes for the primitive equations for a barotropic fluid. *Monthly Weather Review* 1969; **97**(5):384–404.
49. Zienkiewicz OC, Taylor RL, Zhu JZ, Nithiarasu P. *The Finite Element Method for a Barotropic Fluid*. Butterworth-Heinemann: London, 2005.
50. Zhu K, Navon IM, Zou X. Variational data assimilation with a variable resolution finite-element shallow-water equations model. *Monthly Weather Review* 1994; **122**(5):946–965.
51. Navon IM. Finite-element simulation of the shallow-water equations model on a limited area domain. *Applied Mathematical Modelling* 1979; **3**(1):337–348.
52. Navon IM. FEUDX: a two-stage, high-accuracy, finite-element FORTRAN program for solving shallow-water equations. *Computers and Geosciences* 1987; **13**(3):255–285.
53. Wang HH, Halpern P, Douglas J, Dupont I. Numerical solutions of the one-dimensional primitive equations using Galerkin approximation with localized basic functions. *Monthly Weather Review* 1972; **100**(10):738–746.
54. Payne NA, Irons BM. Private communication to O. Zienkiewicz, 1963.
55. Huebner KH, Dewhirst DL, Smith DE, Byrom TG. *The Finite-element Method for Engineers*. Wiley: New York, 2001.
56. Polak E, Ribiere G. Note sur la convergence de directions conjugues. *Revue Francaise Informat. Recherche Operationelle* 1969; **3**(16):35–43.
57. Daescu DN, Navon IM. Efficiency of a POD-based reduced second-order adjoint model in 4d-var data assimilation. *International Journal for Numerical Methods in Fluids* 2007; **53**(10):985–1004.
58. Rowley CW. Model reduction for fluids using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos* 2005; **15**(3):997–1013.
59. Noack BR, Michael S, Morzynski M, Tadmor G. System reduction strategy for Galerkin models of fluid flows. *International Journal for Numerical Methods in Engineering* 2009; DOI: 10.1002/fld.2049.
60. Noack BR, Papas P, Monkewitz PA. The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows. *Journal of Fluid Mechanics* 2005; **523**:339–365.
61. Tadmor G, Lehmann O, Noack BR, Morzynski M. System reduction mean field Galerkin models for the natural and actuated cylinder wake flow. *Physics of Fluids* 2009; in press.